ADVANCES IN SOFT COMPUTING 48

Didier Dubois
M. Asunción Lubiano
Henri Prade
María Ángeles Gil
Przemysław Grzegorzewski
Olgierd Hryniewicz (Eds.)

Soft Methods for Handling Variability and Imprecision

Springer

cajAstur

Advances in Soft Computing 48

Editor-in-Chief: J. Kacprzyk

Advances in Soft Computing

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, Kryzysztof Trojanowski (Eds.)
Intelligent Information Processing and Web Mining, 2005
ISBN 978-3-540-25056-2

Abraham Ajith, Bernard de Bacts, Mario Köppen, Bertram Nickolay (Eds.)
Applied Soft Computing Technologies: The Challenge of Complexity, 2006
ISBN 978-3-540-31649-7

Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, Kryzysztof Trojanowski (Eds.)
Intelligent Information Processing and Web Mining, 2006
ISBN 978-3-540-33520-7

Ashutosh Tiwari, Joshua Knowles, Erel Avineri, Keshav Dahal, Rajkumar Roy (Eds.)
Applications and Soft Computing, 2006
ISBN 978-3-540-29123-7

Bernd Reusch, (Ed.)
Computational Intelligence, Theory and Applications, 2006
ISBN 978-3-540-34780-4

Jonathan Lawry, Enrique Miranda, Alberto Bugarín Shoumei Li, María Á. Gil, Przemysław Grzegorzewski, Olgierd Hryniewicz,
Soft Methods for Integrated Uncertainty Modelling, 2006
ISBN 978-3-540-34776-7

Ashraf Saad, Erel Avineri, Keshav Dahal, Muhammad Sarfraz, Rajkumar Roy (Eds.)
Soft Computing in Industrial Applications, 2007
ISBN 978-3-540-70704-2

Bing-Yuan Cao (Ed.)
Fuzzy Information and Engineering, 2007
ISBN 978-3-540-71440-8

Patricia Melin, Oscar Castillo, Eduardo Gómez Ramírez, Janusz Kacprzyk, Witold Pedrycz (Eds.)
Analysis and Design of Intelligent Systems Using Soft Computing Techniques, 2007
ISBN 978-3-540-72431-5

Oscar Castillo, Patricia Melin, Oscar Montiel Ross, Roberto Sepúlveda Cruz, Witold Pedrycz, Janusz Kacprzyk (Eds.)
Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, 2007
ISBN 978-3-540-72433-9

Katarzyna M. Węgrzyn-Wolska, Piotr S. Szczepaniak (Eds.)
Advances in Intelligent Web Mastering, 2007
ISBN 978-3-540-72574-9

Emilio Corchado, Juan M. Corchado, Ajith Abraham (Eds.)
Innovations in Hybrid Intelligent Systems, 2007
ISBN 978-3-540-74971-4

Marek Kurzynski, Edward Puchala, Michal Wozniak, Andrzej Zolnierek (Eds.)
Computer Recognition Systems 2, 2007
ISBN 978-3-540-75174-8

Van-Nam Huynh, Yoshiteru Nakamori, Hiroakira Ono, Jonathan Lawry, Vladik Kreinovich, Hung T. Nguyen (Eds.)
Interval / Probabilistic Uncertainty and Non-classical Logics, 2008
ISBN 978-3-540-77663-5

Ewa Pietka, Jacek Kawa (Eds.)
Information Technologies in Biomedicine, 2008
ISBN 978-3-540-68167-0

Didier Dubois, M. Asunción Lubiano, Henri Prade, María Ángeles Gil, Przemysław Grzegorzewski, Olgierd Hryniewicz (Eds.)
Soft Methods for Handling Variability and Imprecision, 2008
ISBN 978-3-540-85026-7

Didier Dubois, M. Asunción Lubiano,
Henri Prade, María Ángeles Gil,
Przemysław Grzegorzewski,
Olgierd Hryniewicz (Eds.)

Soft Methods for Handling Variability and Imprecision

Springer cajAstur

Editors

Dr. Didier Dubois
Université Paul Sabatier
IRIT- Bureau 308
118 Route de Narbonne
31062 Toulouse Cedex 09
France
E-mail: dubois@irit.fr

M. Asunción Lubiano
Departamento de Estadística e I.O. y D.M.
Universidad de Oviedo
Facultad de Ciencias
C/ Calvo Sotelo, s/n
33007 Oviedo
Spain
E-mail: lubiano@uniovi.es

Dr. Henri Prade
Université Paul Sabatier
IRIT- Bureau 307
118 Route de Narbonne
31062 Toulouse Cedex 09
France
E-mail: prade@irit.fr

María Ángeles Gil
Departamento de Estadistica e I.O. y D.M.
Universidad de Oviedo
C/ Calvo Sotelo, s/n
33007 Oviedo
Spain
E-mail: angelesgil@telecable.es,
magil@uniovi.es

Dr. Przemysław Grzegorzewski
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland, and
Faculty of Mathematics and Information Science
Warsaw University of Technology
Plac Politechniki 1, 00-661 Warsaw
Poland
E-mail: pgrzeg@ibspan.waw.pl

Dr. Olgierd Hryniewicz
Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw
Poland
E-mail: hryniewi@ibspan.waw.pl

ISBN 978-3-540-85026-7 e-ISBN 978-3-540-85027-4

DOI 10.1007/978-3-540-85027-4

Advances in Soft Computing ISSN 1615-3871

Library of Congress Control Number: 2008931051

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

5 4 3 2 1 0

springer.com

Preface

Probability theory has been the only well-founded theory of uncertainty for a long time. It was viewed either as a powerful tool for modelling random phenomena, or as a rational approach to the notion of degree of belief. During the last thirty years, in areas centered around decision theory, artificial intelligence and information processing, numerous approaches extending or orthogonal to the existing theory of probability and mathematical statistics have come to the front. The common feature of those attempts is to allow for softer or wider frameworks for representing uncertain information. Various approaches have appeared, either on their own like fuzzy set theory, possibility theory, rough sets, or having their origin in probability theory itself, like imprecise probability, belief functions, fuzzy random variables. Many of the latter come down to blending interval or fuzzy interval analysis with probabilistic methods. These methods are softer than the traditional theories and techniques because they are less Procrustean, they more easily adapt to the actual nature of information.

Nearly ten years ago, the need was felt to establish a recurrent forum for discussing such new trends that enlarge the statistical and uncertainty modelling traditions, towards a flexible and more specific handling of incomplete or subjective information. This idea resulted in the launching of the International Conference on Soft Methods in Probability and Statistics (SMPS), organized for the first time in Warsaw in 2002. Subsequent events in this series took place in Oviedo in 2004 and then in Bristol in 2006.

This volume is a collection of selected papers presented at the 4th International Conference on Soft Methods in Probability and Statistics (SMPS'2008) held in Toulouse, France, on September 8–10, 2008. It was organized by the RPDMP team (Raisonnements Plausibles, Décision, Méthodes de Preuve) at the Institut de Recherche en Informatique de Toulouse (IRIT), on the precincts of Université Paul Sabatier.

The volume contains five sections. The first one is dedicated to papers based on invited talks. Two of them are devoted to generalizations (or soft versions) of Bayesian inference: Jean-Marc Bernard from Paris 5 University (France) considers imprecise predictions in the framework of the Dirichlet model; Reinhard

Viertl from the Technical University of Vienna (Austria) extends the Bayesian setting to the handling of fuzzy data and fuzzy prior probabilities. The third paper by Dominique Guyonnet, Senior Scientist at BRGM, Orléans (France's leading public institution in geoscience) considers the applications of soft methods to risk analysis in climate change problems. Finally some pages are devoted to the pioneering works of Robert Féron on fuzzy random variables invented by him in 1976.

Part II is devoted to contributions to the foundations of uncertainty theories such as imprecise probability representations, possibility theory, the bridge to linguistic information. Two contributions deal with the concepts of independence and belief revision, respectively.

Part III contains numerous papers devoted to soft statistical methods, ranging from the principles of statistical inference to detailed problems connected with statistical tests (of independence, of the mean, of the variance, etc.) and estimation. Part IV focuses on mathematical aspects of soft methods applied to probability and statistics. Various contributions address issues in measure theory, stochastic differential equations, convergence issues, the formalization of variance, but also some discrete mathematical problems. It includes the contributions to a special invited session on fuzzy set-valued analysis organized by Luis J. Rodríguez-Muñiz. Papers on aggregation functions and algebraic issues can also be found.

Part V is the application section, devoted to engineering. It ranges from computational methods for uncertainty propagation to regression, learning data-mining and decision analysis. Applications include expert opinion fusion, structural analysis, and design optimization.

The editors are grateful to contributing authors, invited speakers, and all Programme Committee members and additional referees who made it possible to put together an attractive program for the conference. Thanks go to Janusz Kacprzyk for his everlasting support to SMPS, to the Editorial staff of Springer for producing the volume, carefully put together by M. Asunción Lubiano.

This conference has also benefited from the financial support of several organizations, without which the meeting could not have taken place. We are grateful to the "Obra Social y Cultural" of the main Savings Bank in Asturias, CajAstur, for generously supporting the production costs of the proceedings. This conference was also sponsored by Institut de Radioprotection et de Sûreté Nucléaire (IRSN, Cadarache) and Université Paul Sabatier. It is placed under the auspices of EUSFLAT (the European Society for Fuzzy Logic and Technology).

Toulouse, May 2008

Didier Dubois
M. Asunción Lubiano
Henri Prade
María Ángeles Gil
Przemysław Grzegorzewski
Olgierg Hryniewicz

Members of Committees

Advisory Committee (Core SMPS Group)

María Ángeles Gil (University of Oviedo, Spain)
Przemysław Grzegorzewski (Polish Academy of Sciences, Warsaw, Poland)
Olgierg Hryniewicz (Polish Academy of Sciences, Warsaw, Poland)

Programme Committee

Salem Benferhat (Lens, France)
Christian Borgelt (Mieres, Spain)
Eric Chojnacki (IRSN, Cadarache, France)
Giulianella Coletti (Perugia, Italy)
Ana Colubi (Oviedo, Spain)
Frank P.A. Coolen (Durham, United Kingdom)
Ines Couso (Gijón, Spain)
Bernard De Baets (Gent, Belgium)
Gert De Cooman (Gent, Belgium)
Thierry Denœux (Compiègne, France)
Didier Dubois (Toulouse, France), Chairman
Pedro Gil (Oviedo, Spain)
Angelo Gilio (Roma, Italy)
Luis Godo (Barcelona, Spain)
Gil González-Rodríguez (Mieres, Spain)
Michel Grabisch (Paris, France)
Dominique Guyonnet (BRGM, Orléans, France)
Eyke Hüllermeier (Marburg, Germany)
Janusz Kacprzyk (Warsaw, Poland)
Uzay Kaymak (Rotterdam, The Nertherlands)
Etienne Kerre (Gent, Belgium)
Vladik Kreinovich (El Paso, Texas, USA)
Volker Krätschmer (Berlin,Germany)

Rudolph Kruse (Magdeburg, Germany)
Jonathan Lawry (Bristol, United Kingdom)
Shoumei Li (Beijing, China)
Weiru Liu (Belfast, Ulster, United Kingdom)
Miguel López (Oviedo, Spain)
Gilles Mauris (Annecy, France)
Enrique Miranda (Madrid, Spain)
Serafin Moral (Granada, Spain)
Wolfgang Näther (Freiberg, Germany)
Detlef Nauck (BT, Ipswich, United Kingdom)
Mirko Navara (Praha, Czech)
Hung T. Nguyen (Las Cruces, NM, USA)
Henri Prade (Toulouse, France)
Yukio Ogura (Saga, Japan)
Anca Ralescu (Cincinnati, Ohio, USA)
Qiang Shen (Aberyswyth, United Kingdom)
Olivier Strauss (Montpellier, France)
Pedro Terán (Zaragoza, Spain)
Reinhard Viertl (Wien, Austria)
Marco Zaffalon (Lugano, Swiss)

Additional Referees

Sébastien Destercke (IRIT Toulouse; IRSN Cadarache, France)
Tomáš Kroupa (Prague, Czech Republic)
Kevin Loquin (Montpellier, France)
M. Asunción Lubiano (Oviedo, Spain)
Luis J. Rodríguez-Muñiz (Gijón, Spain)
Mathieu Serrurier (Toulouse, France)
Luciano Stefanini (Urbino, Italy)
Matthias C.M. Troffaes (Durham, United Kingdom)

Local Organization

Florence Boué (IRIT, Université Paul Sabatier, Toulouse)
Sébastien Destercke (IRIT, Toulouse; IRSN, Cadarache)
Didier Dubois (IRIT, CNRS/Université Paul Sabatier, Toulouse)
Josette Pastor (INSERM/Université Paul Sabatier, Toulouse)
Henri Prade (IRIT, CNRS/Université Paul Sabatier, Toulouse)
Mathieu Serrurier (IRIT, Université Paul Sabatier, Toulouse)

Proceedings

M. Asunción Lubiano (University of Oviedo, Spain)

Contents

Part V: Engineering

List of Contributors

Isabel Aguiló
Departament de Ciències Matemàtiques i Informàtica,
Universitat
de les Illes Balears, 07122 Palma de Mallorca, Spain
isabel.aguilo@uib.es

Patrice Aknin
French National Institute for Transport and
Safety Research (INRETS) - LTN, 2 av. Malleret-Joinville, 94114
Arcueil Cedex, France
aknin@inrets.fr

Jean-Paul Auray
LIRIS-MA2D, CNRS and Université Lyon 1.
Bâtiment Nautibus, Campus de la Doua
8, Bd Niels Bohr, 69622 Villeurbanne Cedex, France
jean-paul.auray@liris.cnrs.fr

Andrés D. Báez-Sánchez
Instituto de
Matemática Estatística e Computação Científica,
Universidad Estadual de Campinas, CP 6065,
13083-970, Campinas - SP, Brazil
ra066805@ime.unicamp.br

Gaël Bellenfant
BRGM, 3 avenue Claude Guillemin, BP 36009, 45060
Orléans Cédex 2, France
g.bellenfant@brgm.fr

Jean-Marc Bernard
Université Paris Descartes,
Laboratoire de Psychologie Environnementale (CNRS, UMR 8069),
71 avenue Edouard Vaillant, F-92774
Boulogne-Billancourt, France
Jean-Marc.Bernard@univ-paris5.fr

Ángela Blanco
Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain, and

European
Centre for Soft Computing, Edificio Científico-Tecnológico, 33600 Mieres, Spain
blancoangela.uo@uniovi.es

Olivier Bouc
BRGM, 3 avenue Claude Guillemin, BP 36009, 45060

Orléans Cédex 2, France
o.bouc@brgm.fr

María Rosa Casals
Departamento de
Estadística e I.O. y D.M.,
Universidad de Oviedo, 33007 Oviedo,
Spain
rmcasals@uniovi.es

Marco E.G.V. Cattaneo
Institut für Statistik,
Ludwig-Maximilians-Universität
München, Ludwigstr. 33, 80539
München, Germany
cattaneo@stat.uni-muenchen.de

Giulianella Coletti
Dip. Matematica e Informatica,
Università di
Perugia, via Vanvitelli, 06123 -
Perugia, Italy
coletti@dipmat.unipg.it

Ana Colubi
Departamento de Estadística e I.O. y
D.M.,
Universidad de Oviedo, 33007 Oviedo,
Spain
colubi@uniovi.es

Etienne Côme
French National Institute for
Transport
and Safety Research (INRETS) - LTN
2 av. Malleret-Joinville, 94114 Arcueil
Cedex, France
come@inrets.fr

Frank P.A. Coolen
Dept. of Mathematical Sciences,
Durham University, Durham DH1
3LE, United
Kingdom
Frank.Coolen@durham.ac.uk

Norberto Corral
Departamento de Estadística e
I.O. y D.M.,
Universidad de Oviedo, 33007 Oviedo,
Spain
norbert@uniovi.es

Inés Couso
Dept. of Statistics and O.R.,
University
of Oviedo, Spain
couso@uniovi.es

Bernard De Baets
Department of Applied Mathematics,
Biometrics and
Process Control,
Ghent University, Coupure links 653,
B-9000 Gent,
Belgium
bernard.debaets@ugent.be

Gert de Cooman
SYSTeMS Research Group, Ghent
University,
Technologiepark–Zwijnaarde 914,
9052 Zwijnaarde, Belgium
gert.decooman@UGent.be

Karel De Loof
Department of Applied Mathematics
and
Computer Science,
Ghent University, Krijgslaan 281 (S9),
B-9000 Gent,
Belgium
karel.deloof@ugent.be

Hans De Meyer
Department of Applied Mathematics
and
Computer Science,
Ghent University, Krijgslaan 281 (S9),
B-9000 Gent,
Belgium
hans.demeyer@ugent.be

Thierry Denœux
HEUDIASYC, Université de Technologie
de Compiègne, CNRS - Centre de Recherches de Royallieu, B.P. 20529, 60205 Compiègne Cedex, France
ijar@hds.utc.fr

Sébastien Destercke
Institut de Radioprotection et
de Sureté Nucléaire (IRSN), Bat 702, Cadarache, 13115
St-Paul lez
Durance, France
sebastien.destercke@irsn.fr

Serena Doria
Department of Geotechnologies, University G.d'Annunzio, Via dei Vestini, 31, Chieti, Italy
s.doria@dst.unich.it

Fabrizio Durante
Department of Knowledge-Based Mathematical Systems, Johannes Kepler
University Linz, A-4040 Linz, Austria
fabrizio.durante@jku.at

Arturo Erdely
Universidad Anáhuac México Norte, Escuela de
Actuaría, Av. Lomas Anáhuac s/n, C.P. 52786,
Huixquilucan, Estado de México, México
aerdely@anahuac.mx

Steffen Freitag
Institute for Structural Analysis, Technische Universität Dresden, 01062 Dresden, Germany
steffen.freitag@tu-dresden.de

Martin Fuchs
University of Vienna, Faculty of Mathematics, Nordbergstr. 15, 1090 Wien, Austria
martin.fuchs@univie.ac.at

María Ángeles Gil
Departamento de
Estadística e
I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain
magil@uniovi.es

José M. González-Barrios
Departamento de
Probabilidad y Estadística, Instituto de Investigaciones en
Matemáticas Aplicadas y en Sistemas, Universidad Nacional
Autónoma de México, Circuito Escolar s/n, Ciudad
Universitaria, C.P. 04510, México D.F., México
gonzaba@sigma.iimas.unam.mx

Gil González-Rodríguez
European Centre for
Soft
Computing, Edificio Científico-Tecnológico, 33600 Mieres, Spain
gil.gonzalez@softcomputing.es

Inés González-Rodríguez
Dep. of Mathematics, Statistics and Computing, University of
Cantabria; Los Castros s/n, 39005 Santander, Spain
ines.gonzalez@unican.es

Wolfgang Graf
Institute for Structural Analysis, Technische Universität Dresden, 01062 Dresden, Germany
wolfgang.graf@tu-dresden.de

Przemysław Grzegorzewski
Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland, and Faculty of Mathematics and Information Science, Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, Poland
pgrzeg@ibspan.waw.pl

Dominique Guyonnet
BRGM, 3 avenue Claude Guillemin, BP 36009, 45060
Orléans Cédex 2, France
d.guyonnet@brgm.fr

Minh Ha-Duong
CIRED, CNRS. Campus du Jardin Tropical, 45 bis avenue de la belle Gabrielle, 94736 Nogent-sur-Marne Cedex, France
haduong@centre-cired.fr
Web page: http://minh.haduong.com/index.html

Olgierd Hryniewicz
Systems Research Institute, Polish Academy of Science, ul. Newelska 6, 01-447 Warsaw, Poland
Olgierd.Hryniewicz@ibspan.waw.pl

Nathan Huntley
Durham University, Department of Mathematical Sciences, Science Laboratories, South Road, Durham, DH1 3LE, United Kingdom
nathan.huntley@durham.ac.uk

Janusz Kacprzyk
Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01-447 Warszawa, Poland
kacprzyk@ibspan.waw.pl

Tomáš Kroupa
Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4,
182 08 Prague, Czech Republic, and Department of Mathematics, Faculty of Electrical Engineering, Czech Technical University in Prague, 166 27 Prague, Czech Republic
kroupa@utia.cas.cz

Rudolf Kruse
Department of Knowledge and Language
Engineering, University of Magdeburg, Universitätsplatz 2,
D-39106 Magdeburg, Germany
kruse@iws.cs.uni-magdeburg.de

Jonathan Lawry
Department of Engineering Mathematics, University of Bristol; University Walk, Bristol, BS8 1TR, United
Kingdom
j.lawry@bris.ac.uk

Jungang Li
Department of Applied Mathematics, Beijing University of Technology, 100 Pingleyuan, Chaoyang
District, Beijing, 100022, P.R. China
jgl001@emails.bjut.edu.cn

Shoumei Li
Department of Applied Mathematics, Beijing University of Technology, 100 Pingleyuan, Chaoyang
District, Beijing, 100022, P.R. China
lisma@bjut.edu.cn

María Teresa López
Departamento de Estadística e
I.O. y D.M., Universidad de Oviedo, 33007 Oviedo, Spain
mtlopez@uniovi.es

Kevin Loquin
Laboratoire d'Informatique, de Robotique et de Microéléctronique de Montpellier (LIRMM), 161 rue Ada, 34392 Montpellier
Cedex 5, France
Kevin.Loquin@lirmm.fr

M. Asunción Lubiano
Departamento de
Estadística e I.O.
y D.M., Universidad de Oviedo, 33007 Oviedo, Spain
lubiano@uniovi.es

Javier Martín
Departament de Ciències Matemàtiques i Informàtica,
Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain
javier.martin@uib.es

Gilles Mauris
Laboratoire d'Informatique, Systémes, Traitement de l'Information et de la Connaissance (LISTIC), Polytech'Savoie, Domaine Universitaire, BP 80439 - F74944 Annecy Le Vieux Cedex, France
mauris@univ-savoie.fr

Gaspar Mayor
Departament de Ciències Matemàtiques i Informàtica,
Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain
gmayor@uib.es

Enrique Miranda
Rey Juan Carlos University,
Department of
Statistics and O.R., C-Tulipán, s/n, 28933 Móstoles, Spain
enrique.miranda@urjc.es

Christian Moewes
Department of Knowledge and Language Engineering, University of Magdeburg,
Universitätsplatz 2,
D-39106 Magdeburg, Germany
cmoewes@iws.cs.uni-magdeburg.de

Manuel Montenegro
Departamento de Estadística e I.O. y D.M.,
Universidad de Oviedo, 33007 Oviedo, Spain
mmontenegro@uniovi.es

José María Moreno-Jiménez
Grupo Decisión
Multicriterio Zaragoza, Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza, Gran Vía, 2, 50005 Zaragoza, Spain
moreno@unizar.es

Arnold Neumaier
University of Vienna, Faculty of Mathematics, Nordbergstr. 15, 1090 Wien, Austria
Arnold.Neumaier@univie.ac.at

Yukio Ogura
Department of Mathematics, Saga University, Saga City, Saga, 840-8502, Japan
ogura@cc.saga-u.ac.jp

Clemens Otte
Siemens AG, Corporate Technology, Otto-Hahn-Ring 6, 81739
Munich, Germany
clemens.otte@siemens.com

Latifa Oukhellou
Université Paris XII - CERTES, 61 av. du Général de Gaulle, 94100 Créteil, France
oukhellou@univ-paris12.fr

Stephan Pannier
Institute for Structural Analysis, Technische Universität Dresden, 01062 Dresden, Germany
stephan.pannier@tu-dresden.de

Henri Prade
IRIT, CNRS and Université Toulouse 3. IRIT- Bureau 307, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France
prade@irit.fr

Erik Quaeghebeur
SYSTeMS Research Group, Ghent University, Technologiepark 914, 9052 Zwijnaarde, Belgium
Erik.Quaeghebeur@UGent.be

Ana Belén Ramos
European Centre for Soft Computing, Edificio Científico-Tecnológico, 33600 Mieres, Spain
anabelen.ramos@softcomputing.es

Uwe Reuter
Department of Civil Engineering, Technische Universität Dresden, 01062 Dresden, Germany
uwe.reuter@tu-dresden.de

Rosana Rodríguez-López
Departamento de Análisis Matemático, Facultad de Matemáticas, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain
amrosana@usc.es

Luis J. Rodríguez-Muñiz
Department of Statistics and O.R., University of Oviedo, E.U.I.T. Industrial, Campus de Viesques s/n, 33271 Gijón, Spain
luisj@uniovi.es

Marko A. Rojas-Medar
Departamento de Ciencias Básicas, Facultad de Ciencias, Campus Fernando May, Casilla 447, Universidad del Bío-Bío, Chillán, Chile
marko@ueubiobio.cl

Susanne Saminger-Platz
Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, A-4040 Linz, Austria
susanne.saminger-platz@jku.at

Luciano Sánchez
Dept. of Computer Sciences, University of Oviedo, Spain
luciano@uniovi.es

Peter Sarkoci
Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, A-4040 Linz, Austria
peter.sarkoci@jku.at

Andrew Schumann
Department of Philosophy and Science Methodology, Belarusian State University, K.Marx St. 31, 220030 Minsk, Belarus
Andrew.Schumann@gmail.com

Romano Scozzafava
Dip. Metodi e Modelli Matematici, Università “La Sapienza”, via Scarpa 16, 00161 - Roma, Italy
romscozz@dmmm.uniroma1.it

Jan-Uwe Sickert
Institute for Structural Analysis, Technische Universität Dresden, 01062 Dresden, Germany
jan-uwe.sickert@tu-dresden.de

Luciano Stefanini
Department of Economics and Quantitative Methods (DEMQ) and Faculty of Economics, University of Urbino, Urbino, Italy
luciano.stefanini@uniurb.it

Olivier Strauss
Laboratoire d'Informatique, de Robotique et de Microéléctronique de Montpellier (LIRMM), 161 rue Ada, 34392 Montpellier
Cedex 5, France
Olivier.Strauss@lirmm.fr

Jaume Suñer
Departament de Ciències Matemàtiques i Informàtica,
Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain
jaume.sunyer@uib.es

Anna Szediw
Systems Research Institute, Polish Academy of Science, ul. Newelska 6, 01-447 Warsaw, Poland
Anna.Szediw@ibspan.waw.pl

Yongchuan Tang
College of Computer Science, Zhejiang University, Hangzhou 310027, P.R. China
tyongchuan@gmail.com

Pedro Terán
Grupo Decisión Multicriterio Zaragoza, Facultad de Ciencias Económicas y Empresariales.
Universidad de Zaragoza, Gran Vía, 2, 50005 Zaragoza, Spain
teran@unizar.es

Matthias C.M. Troffaes
Durham University,
Department of Mathematical Sciences, Science Laboratories, South Road, Durham DH1 3LE, United Kingdom
matthias.troffaes@gmail.com

Barbara Vantaggi
Dip. Metodi e Modelli Matematici, Università "La Sapienza", via Scarpa 16, 00161 - Roma,
Italy
vantaggi@dmmm.uniroma1.it

Reinhard Viertl
Department of Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstraße 8-10, 1040 Wien, Austria
R.Viertl@tuwien.ac.at

Carl G. Wagner
Department of Mathematics, The University of Tennessee, Knoxville, TN 37996-1300, USA
wagner@math.utk.edu

Dabuxilatu Wang
School of Mathematics and Information Sciences, Guangzhou University, No. 230 WaiHuan XiLu, University Town, Guangzhou, 510006, P.R. China
dbxlt0@yahoo.com

Anna Wilbik
Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01-447 Warszawa, Poland
wilbik@ibspan.waw.pl

Ralph L. Wojtowicz
Metron Inc., 11911 Freedom Drive, Suite 800,

Reston, VA 20190, USA
wojtowicz@metsci.com
Web page:
http://www.adjoint-functors.net

Jinping Zhang
Department of Mathematics, Saga University, Saga 840-8502, Japan
zhangjinping@bjut.edu.cn

Part I

Invited Papers

Imprecise Probabilistic Prediction for Categorical Data: From Bayesian Inference to the Imprecise Dirichlet-Multinomial Model

Jean-Marc Bernard

Université Paris Descartes, Laboratoire de Psychologie Environnementale (CNRS, UMR 8069), Boulogne-Billancourt, France

Abstract. From n categorical observations, what can be predicted about the next n' ones? We present a generalization of the Bayesian approach, the imprecise Dirichlet-multinomial model (`IDMM`), which uses sets of Dirichlet-multinomial distributions to model prior ignorance. The `IDMM` satisfies coherence, symmetry and several desirable invariance properties.

Keywords: Predictive inference, Rule of succession, Imprecise Dirichlet model, Prior ignorance.

1 Introduction

1.1 A Story with (Too) Many Rabbits

Some years ago, a friend proposed me the following puzzle: "A man takes you to a room. There, on a table, stand two top-hats. The man asks you to pick one hat at random and to raise the hat, and to your great surprise, you discover a rabbit under the hat. The man then asks you: *What is the probability that the other hat contains a rabbit too?*" At this stage, my friend looked at me, checking that the desired effect was reached. In effect, I felt embarrassed because my probabilistic intuition told me that I could not reasonably answer the question. The question even sounded somehow meaningless.

The puzzle went on: "Suppose now that there are 100 hats on the table and that the man asks you to draw one hat, then another, and so on, and that, under each hat, you find a rabbit, until there is only one hat left. Now the question is: *What is the probability that the last hat also contains a rabbit?*" Now I felt somehow more embarrassed, because it seemed to me that there was an answer to the question, and that, with some calculations, I could put a figure on the requested probability, e.g. around $\frac{99}{100}$.

But then came the final stroke: "When moving from 2 hats to 100 hats, a question which initially sounded meaningless has changed to a relevant question. After how many hats does the question become meaningful?"

I had then realized that the Bayesian theory proposed answers to the rabbits story. For instance, assuming that there are only $K = 2$ possibilities from the set $C_1 = \{rabbit, no-rabbit\}$, and that one is in a state of prior ignorance, Laplace's famous *rule of succession* provides the answers $\frac{2}{3}$ and $\frac{100}{101}$ to the 2-hats and 100-hats cases respectively. But these probabilities become $\frac{4}{5}$ and $\frac{102}{103}$ if the set of possibilities is taken as $C_2 = \{rabbit, pigeon, other, nothing\}$. And other ways to define prior ignorance yield

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 3–9, 2008.
springerlink.com

a probability of 1 for the two cases indistinctly. Hence, in the 2-hats case, the answer appears to depend too much on the prior and on the number of basic categories, and this could explain the reluctance to give an answer in that case. In contrast, the 100-hats case appears less problematic because these dependencies are less strong.

It now seems to me that the rabbits puzzle illustrates the advantages of using *imprecise probability models* to express uncertainty, especially in situations of prior ignorance. This paper presents an imprecise probability model, the imprecise Dirichlet-multinomial model (IDMM) which possess several interesting properties: it generalizes Bayesian inference, it yields probability intervals for events of interest, it proposes a new definition of prior ignorance, it is invariant w.r.t. the possibility space C, and it satisfies several other important properties including coherence. The IDMM has been studied in great detail by Walley & Bernard [15] and can be seen as the predictive counterpart of Walley's *Imprecise Dirichlet Model* (IDM), see [13].

For example, under the IDMM, before observing any hat, the prior probability interval of finding a rabbit is $[0,1]$, a probability interval which expresses a maximal uncertainty in Walley's theory of imprecise probabilities (see [12]). After observing the first hat, the IDMM yields the probability interval $[\frac{1}{2},1]$, and, after observing 99 hats, the interval $[\frac{99}{100},1]$.[1]

1.2 The General Problem

The general problem of *predictive inference* can be stated as follows. Each observable unit is classified into one of K categories or types from a set C, labeled $1,2,\ldots,K$. We observe the types of n distinct units. Let $\boldsymbol{a}=(a_1,\ldots,a_K)$ and $\boldsymbol{f}=\boldsymbol{a}/n$ denote the counts and (relative) frequencies of each type in the observed sample of size n, with $\sum_k a_k = n$.[2] The problem is to make inferences about $\boldsymbol{a'}=(a'_1,\ldots,a'_K)$ or $\boldsymbol{f'}=\boldsymbol{a'}/n'$, the counts or frequencies of each type in a future sample of n' new units, where $\sum_k a'_k = n'$. The rabbits examples refers to the special case $n'=1$ of *immediate prediction*, in which the prediction bears on the next observation only.

The problem of probabilistic prediction for categorical data has been of great historical importance in Statistics. In the 18th century, Bayes discussed the problem in the case of two categories, and Laplace proposed a generalization to the case of multiple categories, The general problem, which Karl Pearson called "the fundamental problem of practical statistics", and the "rules of succession" proposed to answer the problem of immediate prediction have been widely discussed since then, see [8, 15] and references therein. Geisser [8] stresses that predictive inference is the most natural approach to inference since it attempts at modeling observables (past and future) only, without referring to an underlying population.

1.3 Laplace's and Other Bayesian Rules of Succession

A *rule of succession* is a solution to the problem of immediate prediction, i.e. the case $n'=1$. Laplace's *principle of indifference* leads to the famous rule of succession $Prob(a'_k = 1 \,|\, \boldsymbol{a}) = \frac{a_k+1}{n+K}$.

[1] For an IDMM with hyperparameter $s=1$, see further.

[2] In the sequel, all sums and products with index k run from 1 to K.

More generally, Bayesian conjugate analysis (based on a Dirichlet-multinomial prior, as we shall see) leads to the following rule,

$$Prob(a'_k = 1 \,|\, \boldsymbol{a}) = \frac{a_k + \alpha_k}{n + s}, \tag{1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ are fixed positive reals, with $\sum_k \alpha_k = s$. Each α_k can be thought of as a *prior strength* allocated to each category k.

Within the Bayesian framework, several approaches have been proposed for the purpose of *objective inference*, i.e. for making inferences from a state of prior ignorance. Each one leads to different choices for the prior strengths $\boldsymbol{\alpha}$, typically symmetrical, i.e. $\alpha_k = s/K$, with a small value for s, either $s = 0$, 1, $K/2$ or K. See e.g. [7, 15]. The problem with these objective Bayesian methods is not only that their answers differ, but that they can differ substantially, especially when n is small compared to K. Another problem is the dependence on the set C and the number of types K, whereas the way observations are categorized can be partly arbitrary.

The `IDMM` described in this paper answers these difficulties by using a set of prior distributions to characterize prior uncertainty, instead of a single prior distribution as in the Bayesian approach. As a result, the predictive inferences produced by the `IDMM` encompass several objective Bayesian methods, and also are invariant w.r.t. refinements or coarsenings of the possibility space C.

2 Bayesian Predictions

2.1 Link between Past and Future Observations

In order to make inferences about the unknown $\boldsymbol{a'}$ from the observed $\boldsymbol{a}$, past and future observations must be linked in some way. Intuitively, the idea is to think that past and future observations are "homogeneous", in other words that the observed counts $\boldsymbol{a}$ constitute a "fair" representative of the *combined* (past + future) counts $\boldsymbol{a} + \boldsymbol{a'}$. We denote by an asterisk the characteristics of the combined observations: $n^* = n + n'$, $\boldsymbol{a^*} = \boldsymbol{a} + \boldsymbol{a'}$, and $\boldsymbol{f^*} = (n\boldsymbol{f} + n'\boldsymbol{f'})/(n + n')$.

For instance, assume that the observed data were obtained by *multiple hypergeometric sampling*, i.e. n observations taken at random from the $n^* = n + n'$ combined ones. The probability of obtaining the counts $\boldsymbol{a}$ conditional on the combined counts $\boldsymbol{a^*}$ is then

$$Prob(\boldsymbol{a} \,|\, \boldsymbol{a^*}) = \prod_k \binom{a^*_k}{a_k} \Big/ \binom{n^*}{n} \propto \prod_k \binom{a_k + a'_k}{a_k}. \tag{2}$$

More generally, the intuitive idea of homogeneity can be operationalized by an assumption of *exchangeability* (or order-invariance): conditional on the counts of each type in the $n + n'$ observations, each possible ordering of the $n + n'$ observations is equally probable a priori. The exchangeability assumption is satisfied whether the size of the sampled population is finite (multiple hypergeometric process), of size at least $n^* = n + n'$, or even infinite (multinomial process). Under that assumption, the probability of $\boldsymbol{a}$, conditional on $\boldsymbol{a^*}$, is always proportional to (2), and hence the probability in (2) represents the *likelihood* of $\boldsymbol{a^*}$ given $\boldsymbol{a}$.

2.2 Dirichlet-Multinomial (*DiMn*) Prior and Posterior

The conjugate family associated to the likelihood (2) is the Dirichlet-multinomial (*DiMn*) family. The $DiMn(\boldsymbol{\alpha}, n^*)$ prior on $\boldsymbol{a}^*$ with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$, that is,

$$Prob_{\boldsymbol{\alpha}}(\boldsymbol{a}^*) = \binom{n^*}{\boldsymbol{a}^*} \frac{\alpha_1^{[a_1^*]} \cdots \alpha_K^{[a_K^*]}}{s^{[n^*]}}, \tag{3}$$

when combined with the likelihood (2), yields a $DiMn(\boldsymbol{a} + \boldsymbol{\alpha}, n')$ posterior on $\boldsymbol{a}'$ conditionally on $\boldsymbol{a}$ that is,

$$Prob_{\boldsymbol{\alpha}}(\boldsymbol{a}' \,|\, \boldsymbol{a}) = \binom{n+n'}{\boldsymbol{a}+\boldsymbol{a}'} \frac{(a_1+\alpha_1)^{[a_1']} \cdots (a_K+\alpha_K)^{[a_K']}}{(n+s)^{[n']}}. \tag{4}$$

In the above equations, the formulae for probabilities are expressed in terms of standard multinomial coefficients and of the ascending factorial function $y^{[u]}$, defined for all integer $u \geq 0$ and real y by[3]

$$y^{[u]} = y(y+1)\cdots(y+u-1), \quad \text{for } u > 0 \text{ and } \quad y^{[0]} = 1. \tag{5}$$

Note that the parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ act as prior strengths allocated to each category and that updating is done by incrementing them with the observed counts $\boldsymbol{a} = (a_1, \ldots, a_K)$. We shall also use an alternative parameterization in terms of s and the relative prior strengths $\boldsymbol{t} = (t_1, \ldots, t_K) = \boldsymbol{\alpha}/s$.

2.3 Some Properties of *DiMn* Distributions

Below, we list a few properties of the *DiMn*. Other properties and further references may be found in [15] and [9, pp. 80–83 & 202–211].

- *Expectations*: The prior (posterior) expectations are given by the prior (posterior) relative strengths, $E(f_k^*) = t_k$, and $E(f_k' \,|\, \mathbf{a}) = (a_k + st_k)/(n+s)$.
- *Pooling*: When pooling two categories k and l into a single one, the *DiMn* form is conserved (with $K-1$ categories instead of K), and the strengths associated with the compound category $k+l$ are obtained by summation: $\alpha_k + \alpha_l$ for prior strengths, and $a_k + a_l$ for observed counts. Note that s is preserved by pooling.
- *Links with the Dirichlet distribution*: Suppose that the data are obtained by multinomial sampling from an infinite population with true frequencies $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$, and that prior uncertainty is described by a Dirichlet distribution, $\boldsymbol{\theta} \sim Diri(\boldsymbol{\alpha})$. Then, the induced prior and posterior predictive distributions are exactly the ones given in (3) and (4). As a consequence, as $n' \to \infty$ (and hence n^* too), the prior *DiMn* on $\boldsymbol{f'}$ (and $\boldsymbol{f^*}$) tends to a $Diri(\boldsymbol{\alpha})$ and the posterior *DiMn* on $f' \,|\, \boldsymbol{a}$ tends to a $Diri(\boldsymbol{a} + \boldsymbol{\alpha})$.

[3] There are alternative expressions in terms of generalized binomial coefficients and/or Gamma functions, see [15].

3 The Imprecise Dirichlet-Multinomial Model (IDMM)

3.1 The Idea of an Ignorance Zone

Each of the various objective Bayesian or frequentist methods that were proposed for the predictive problem can in fact be seen or re-interpreted as derived from a *DiMn* with some specific choice for the prior strengths $\boldsymbol{\alpha}$, all α_k's being always within $[0,1]$. The most usual objective Bayesian solutions use symmetrical *DiMn* priors: $\alpha_k = 0$ for Haldane, $\alpha_k = 1/K$ for Perks, $\alpha_k = 1/2$ for Jeffreys, $\alpha_k = 1$ for Bayes-Laplace (see e.g. [2]). And frequentist solutions correspond to asymmetrical priors, but again with small values for the α_k's [11]. From this closeness of these various ways to define prior ignorance, has emerged the idea that ignorance could be defined by an *ignorance zone* for vector $\boldsymbol{\alpha}$, rather by than a single $\boldsymbol{\alpha}$ value (see [1, 2]). See also [10, 11] who came close to this idea. This idea was formalized by Walley, see [12] for the general theory, and [13] for the case of categorical data.

3.2 Definition of the IDMM

To model prior ignorance about the counts $\boldsymbol{a}^*$, we use the set of *DiMn* distributions, parameterized in terms of s and $\boldsymbol{t} = \boldsymbol{\alpha}/s$:

$$\{Prob_{s\boldsymbol{t}} : t_k > 0 \text{ for } k = 1,\ldots,K,\ \textstyle\sum_k t_k = 1\}, \tag{6}$$

where s and n^* are fixed. In (6), $\boldsymbol{t}$ ranges over the interior of the unit simplex. We call the model (6) the *imprecise Dirichlet-multinomial model* (IDMM) with hyperparameter s, which we write as IDMM(s).

Hence, the prior IDMM(s) is the set of all $DiMn(\boldsymbol{\alpha}, n^*)$ prior distributions which satisfy $\boldsymbol{\alpha} = s\boldsymbol{t}$, $t_k > 0$ and $\sum_k t_k = 1$. After observing the frequencies $\boldsymbol{a}$, the IDMM is updated to the set of all $DiMn(\boldsymbol{a} + s\boldsymbol{t}, n')$ posterior distributions on $\boldsymbol{a'}$ with the same constraints on $\boldsymbol{t}$, i.e.,

$$\{Prob_{\boldsymbol{a}+s\boldsymbol{t}} : t_k > 0 \text{ for } i = 1,\ldots,K,\ \textstyle\sum_k t_k = 1\}. \tag{7}$$

The posterior set in (7) defines the posterior IDMM and models the uncertainty about the future observations $\boldsymbol{a'}$ after observing the data $\boldsymbol{a}$. We can make inferences from the IDMM by calculating posterior lower and upper (L&U) probabilities of any event B or expectations of any function $V = V(\boldsymbol{a'})$, which are denoted by $\underline{Prob}(B\,|\,\boldsymbol{a})$, $\overline{Prob}(B\,|\,\boldsymbol{a})$, $\underline{E}(V\,|\,\boldsymbol{a})$, and $\overline{E}(V\,|\,\boldsymbol{a})$, by maximizing and minimizing $Prob_{s\boldsymbol{t}}(B\,|\,\boldsymbol{a}) = \sum_{\boldsymbol{a'}\in B} Prob_{s\boldsymbol{t}}(\boldsymbol{a'}\,|\,\boldsymbol{a})$ or $E_{s\boldsymbol{t}}(V\,|\,\boldsymbol{a}) = \sum_{\boldsymbol{a'}} V(\boldsymbol{a'}) Prob_{s\boldsymbol{t}}(\boldsymbol{a'}\,|\,\boldsymbol{a})$ with respect to $\boldsymbol{t}$, where $Prob_{s\boldsymbol{t}}(\boldsymbol{a'}\,|\,\boldsymbol{a})$ is given by (4).

The size of the IDMM prior and posterior sets of distributions are governed by the constant s. Values within $[1;2]$ for s have been proposed, in the context of the IDM as a good compromise between (i) having not too weak inferences and (ii) encompassing alternative Bayesian and frequentist models, see [13, 3] for detailed arguments.

3.3 Properties of Inferences from the IDMM

In the development of statistical methods, there has been a considerable attention on the properties inferences should satisfy, especially for an *objective inference*: what do

we learn about the future counts $\boldsymbol{a'}$ or frequencies $\boldsymbol{f'} = \boldsymbol{a'}/n'$ from the observed data $\boldsymbol{a}$, regardless of any possible prior knowledge. Statisticians have proposed several principles or properties that seem desirable for such inferences (see [4, Sect. 2.3 and 2.4], and [12, 13, 14] for formal definitions, detailed discussions of why they seem desirable or compelling, and other references). We list below the major properties that the IDMM satisfies.

- *IDMM rule of succession*: For immediate prediction, i.e. $n' = 1$, the IDMM produces the imprecise rule of succession,

$$\underline{Prob}(a'_k = 1 \,|\, \boldsymbol{a}) = \frac{a_k}{n+s} \quad \text{and} \quad \overline{Prob}(a'_k = 1 \,|\, \boldsymbol{a}) = \frac{a_k + s}{n+s} \tag{8}$$

  In the rabbits example of Section 1.1, under the IDMM($s = 1$), the predictive L&U probabilities of finding another rabbit are $[\frac{1}{2}; 1]$ for the 2-hats case, and $[\frac{99}{100}; 1]$ for the 100-hats case.
- *Symmetry principle* (SP): Prior uncertainty about any event relative to $\boldsymbol{f'}$ is invariant w.r.t. permutations of the K categories of set C. This property is similar to Laplace's principle of indifference.
- *Prior ignorance*: The prior L&U probabilities of simple events, such as $B = (f'_k = 1, n' = 1)$, and L&U expectations $E(f'_k)$ are both $[0; 1]$, i.e. maximally imprecise or *vacuous*. Both become non-vacuous as soon as $n \geq 1$. This is in contrast with the Bayesian approach which always yields a precise value for such events or expectations.
- *Likelihood principle* (LP): Posterior inferences from the IDMM depend on the data through the likelihood function (2) only. Formally, they satisfy the "likelihood principle". In particular, and in contrast with some alternative methods, they do not depend on data that might have been observed but were not.
- *Coherence principle* (CP): This principle is typically put forward in a Bayesian or generalized Bayesian context, in which uncertainty is described by personal probabilistic assessments (possibly imprecise) and are behaviourally interpreted as defining acceptable betting rates. Coherence is a rationality criterion which ensures that several bets or decisions induced by the overall model are mutually consistent. Inferences from the IDMM do satisfy coherence, in the strongest sense of Walley [12, Sect. 2.5 & 7.1]. It generalizes, and in some way strengthens, the Bayesian concept of coherence, see de Finetti [6, Chap. 3].
- *Representation invariance principle* (RIP): Inferences do not depend on what categories are distinguished, nor even on the number of categories, K. Formally they satisfy the RIP proposed by [13]. The RIP states that posterior uncertainty about any event B relative to $\boldsymbol{f'}$ should not depend on refinements or coarsenings of categories, provided that B remains unchanged. In effect, we saw that, when categories are pooled, both the form of a *DiMn* distribution and the value of s are preserved, so that the set of posterior *DiMn* distributions produced by the IDMM(s) is essentially unchanged. For the rabbits example, this implies that the inferences remain the same whether we take $C_1 = \{rabbit, no-rabbit\}$ or $C_2 = \{rabbit, pigeon, other, nothing\}$ as our set of categories.
- *Specificity property*: De Cooman et al. [5] recently studied immediate predictions assuming exchangeability and representation invariance. They showed that, among

such predictive systems, the IDMM is characterized by an additional property of *specificity*: when conditioning on a subset $C^* \subset C$ of the categories, inferences depend only on the counts in C^*.

The IDMM satisfies all these properties or principles jointly. In contrast, the SP and the RIP are mutually exclusive for Bayesian models using proper priors [12, Sect. 5.5], and frequentist methods typically violate the LP and the CP.

However imprecise models such as the IDMM or the related IDM provide a way to reconcile all these alternative objective models [14]. If s is taken large enough, typically $s = 1$ or $s = 2$, then the probability intervals produced by IDMM contain the corresponding probabilities obtained from all alternative objective models either Bayesian or frequentist, for the case $K = 2$, and from the most reasonable of these models for general K (see [3, 13]).

References

1. Bernard, J.-M.: Bayesian interpretation of frequentist procedures for a Bernoulli process. Amer. Statist. 50, 7–13 (1996)
2. Bernard, J.-M.: Bayesian inference for categorized data. In: Rouanet, H., Bernard, J.-M., Bert, M.-C., Lecoutre, B., Lecoutre, M.-P., Le Roux, B. (eds.) New Ways in Statistical Methodology: From Significance Tests to Bayesian Inference. European University Studies, Psychology. Peter Lang, Bern., vol. 6, pp. 159–226 (1998)
3. Bernard, J.-M.: An introduction to the imprecise Dirichlet model for multinomial data. Internat. J. Approx. Reason 39, 123–150 (2005)
4. Cox, D.R., Hinkley, D.V.: Theoretical Statistics. Chapman and Hall, London (1974)
5. De Cooman, G., Miranda, E., Quaeghebeur, E.: Representation insensitivity in immediate prediction under exchangeability. Internat. J. Approx. Reason (in press, 2008)
6. De Finetti, B.: Theory of Probability, vol. 1. John Wiley & Sons, Chichester (1974)
7. Geisser, S.: On prior distributions for binary trials (with discussion). Amer. Statist. 38, 244–251 (1984)
8. Geisser, S.: Predictive inference: An introduction. In: Monographs on Statistics and Applied Probability, vol. 55, Chapman and Hall, New York (1993)
9. Johnson, N.L., Kotz, S., Balakrishnan, N.: Discrete Multivariate Distributions. Wiley, New York (1997)
10. Perks, W.: Some observations on inverse probability including a new indifference rule (with discussion). J. Inst. Actuar. 73, 285–334 (1947)
11. Thatcher, A.R.: Relationships between Bayesian and confidence limits for predictions (with discussion). J. Roy. Statist. Soc. Ser. B 26, 176–210 (1964)
12. Walley, P.: Statistical reasoning with imprecise probabilities. In: Monographs on Statistics and Applied Probability, vol. 42. Chapman and Hall, London (1991)
13. Walley, P.: Inferences from multinomial data: learning about a bag of marbles. J. Roy. Statist. Soc. Ser. B 58, 3–57 (1996)
14. Walley, P.: Reconciling frequentist properties with the likelihood principle. J. Statist. Plann. Inference 105, 35–65 (2002)
15. Walley, P., Bernard, J.-M.: Imprecise probabilistic prediction for categorical data. Technical Report CAF-9901, Laboratoire Cognition et Activités finalisées, Université Paris 8, Saint-Denis, France (1999)

Fuzzy Bayesian Inference

Reinhard Viertl

Department of Statistics and Probability Theory, Vienna University of Technology,
Wien, Austria

Abstract. Fuzziness of data is usually neglected is statistics. But real data are frequently not precise numbers, but more or less imprecise. This imprecision is different from errors. Imprecision of data can be modelled by special fuzzy subsets of the set of real numbers, and statistical methods have to be generalized to fuzzy data.

Another kind of fuzziness is the fuzziness of a-priori information in Bayesian inference. It is possible to apply so-called fuzzy probability distributions as a-priori distributions. The corresponding generalization of Bayes' theorem is basic for what is called fuzzy Bayesian inference.

Keywords: Bayesian inference, Bayes' theorem, Decision analysis, Fuzzy Bayesian inference, Fuzzy data, Fuzzy information, Fuzzy intervals, Fuzzy probability distributions, Fuzzy utility, Non-precise numbers.

1 Introduction

In applications of Bayesian statistical methods frequently data are not precise numbers or vectors, and the a-priori distributions are not exact probability distributions in the standard sense. Therefore it is necessary to model real data in a suitable way to incorporate the fuzziness of data before they are analyzed by statistical methods. This is possible by using special fuzzy subsets of the set of real numbers $\mathbb{R}$, so-called *non-precise numbers*. The mathematical description of non-classical a-priori distributions is possible by so-called *fuzzy probability distributions*. Based on this Bayes' theorem can be adapted by generalizing the likelihood function to the situation of fuzzy data, based on the extension principle of fuzzy set theory.

2 Fuzzy Data

One dimensional data obtained by measurement of continuous quantities are not precise real numbers but more or less non-precise. This imprecision is different from errors and is also called fuzziness. The best up to date mathematical model for measurement data are so-called *non-precise numbers*.

Definition 1. *A non-precise number $x^\star$ is a fuzzy subset of $\mathbb{R}$ whose membership function $\xi(\cdot)$ obeys the following:*

$$\forall\, \delta \in (0;1] \quad \textit{the } \delta\textit{-cut } C_\delta(x^\star) := \{x \in \mathbb{R} \colon \xi(x) \geq \delta\}$$

is non-empty and a finite union of compact intervals $\left[a_{\delta,j}; b_{\delta,j}\right]$, i. e.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 10–15, 2008.
springerlink.com

$$C_\delta(x^\star) = \bigcup_{j=1}^{k_j} \left[a_{\delta,j}; b_{\delta,j}\right].$$

Remark 1. If all δ-cuts are compact intervals, then $x^\star$ is called *fuzzy interval*.

In order to generalize statistical inference procedures to the situation of fuzzy data it is necessary to propagate the imprecision of individual measurement results into the sample space.

For stochastic quantity X the set of all possible values of X is denoted by M_X and called *observation space*. For samples $X_1, \cdots, X_n$ of X the set of all possible values $(x_1, \cdots, x_n)$ with $x_i \in M_X$ is given by the Cartesian product $M_X \times \cdots \times M_X = M_X^n$ which is called *sample space* of X.

For a fuzzy sample consisting of n non-precise numbers $x_1^\star, \cdots, x_n^\star$ with corresponding characterizing functions $\xi_1(\cdot), \cdots, \xi_n(\cdot)$ these n characterizing functions are combined by the minimum t-norm in order to obtain the so-called *vector-characterizing function* $\zeta : \mathbb{R}^n \to [0;1]$ of a fuzzy element $\underline{x}^\star$ in the sample space M_X^n. The values $\zeta(x_1, \cdots, x_n)$ are obtained from the individual characterizing function $\xi_i(\cdot)$ by

$$\zeta(x_1, \cdots, x_n) := \min\{\xi_1(x_1), \cdots, \xi_n(x_n)\} \quad \forall\, (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

The fuzzy element $\underline{x}^\star$ is called *fuzzy combined sample*. It is the basis for the generalization of Bayes' theorem.

Data fuzziness has to be combined with the fuzziness of a-priori knowledge. Therefore it is necessary to generalize a-priori distributions. This is done in the next section.

3 Fuzzy Probability Distributions

Standard probability distributions as a-priori distributions for parameters θ in Bayesian inference are a topic of critics. Therefore so-called *soft models* are more suitable to express a-priori information. In the context of probabilistic a-priori information soft models are so-called *fuzzy probability distributions*.

Definition 2. *A fuzzy probability distribution $P^\star$ on a measurable space $(M, \mathscr{A})$ is a function defined on the event system $\mathscr{A}$ which assigns to every $A \in \mathscr{A}$ a fuzzy interval $P^\star(A)$ with δ-cuts $\left[\underline{P}_\delta(A); \overline{P}_\delta(A)\right]$ obeying the following:*

(1) $P^\star(\emptyset) = 0$ (characterizing function $I_{\{0\}}(\cdot)$)
(2) $P^\star(M) = 1$ (characterizing function $I_{\{1\}}(\cdot)$)
(3) For all $\delta \in (0;1]$ and all pairwise disjoint events $A_1, \cdots, A_n$ from $\mathscr{A}$

$$\overline{P}_\delta\left(\bigcup_{i=1}^n A_i\right) \le \sum_{i=1}^n \overline{P}_\delta(A_i)$$

and

$$\underline{P}_\delta\left(\bigcup_{i=1}^n A_i\right) \ge \sum_{i=1}^n \underline{P}_\delta(A_i)$$

A special case of fuzzy probability distributions is obtained from so-called fuzzy probability densities. *These are special fuzzy valued real functions $f^\star(\cdot)$ whose values $f^\star(x)$ are fuzzy intervals. The δ-cuts of $f^\star(x)$ are denoted by*

$$C_\delta\left[f^\star(x)\right] = \left[\underline{f}_\delta(x); \overline{f}_\delta(x)\right] \quad \forall\, \delta \in (0;1].$$

Definition 3. *Let $(M, \mathscr{A}, \mu)$ be a measure space and $f^\star(\cdot)$ be a fuzzy valued function defined on M. If $f^\star(\cdot)$ is fulfilling the following conditions*

(1) $\forall\, \delta \in (0;1]$ the δ-level functions $\underline{f}_\delta(\cdot)$ and $\overline{f}_\delta(\cdot)$ are integrable with finite integrals

$$\underline{I}_\delta = \int_M \underline{f}_\delta(x)\, d\mu(x) \quad \textit{and} \quad \overline{I}_\delta = \int_M \overline{f}_\delta(x)\, d\mu(x)$$

(2) $1 \in \left[\int_M \underline{f}_1(x)\, d\mu(x); \int_M \overline{f}_1(x)\, d\mu(x)\right]$

(3) there exists a classical probability density $f : M \to [0;\infty)$ with

$$\underline{f}_1(x) \le f(x) \le \overline{f}_1(x) \quad \forall\, x \in M$$

then $f^\star(\cdot)$ is called fuzzy probability density.

Remark 2. Based on a fuzzy probability density $f^\star(\cdot)$ on a measure space $(M, \mathscr{A}, \mu)$ a fuzzy probability distribution $P^\star$ on $(M, \mathscr{A})$ is generated in the following way:

Defining the set $\mathscr{S}_\delta$ of all classical probability densities $f(\cdot)$ on $(M, \mathscr{A}, \mu)$ obeying

$$\underline{f}_\delta(x) \le f(x) \le \overline{f}_\delta(x) \quad \forall\, x \in M$$

the fuzzy probability of $A \in \mathscr{A}$ is defined by its δ-cuts

$$C_\delta\left[P^\star(A)\right] := \left[\underline{P}_\delta(A); \overline{P}_\delta(A)\right] \quad \forall\, \delta \in (0;1]$$

with

$$\overline{P}_\delta(A) := \sup\left\{\int_A f(x)\, d\mu(x) : f(\cdot) \in \mathscr{S}_\delta\right\}$$

$$\underline{P}_\delta(A) := \inf\left\{\int_A f(x)\, d\mu(x) : f(\cdot) \in \mathscr{S}_\delta\right\}.$$

The characterizing function $\psi_{P^\star(A)}(\cdot)$ of the fuzzy interval $P^\star(A)$ is given by the representation lemma:

$$\psi_{P^\star(A)}(x) = \max\left\{\delta \cdot I_{C_\delta[P^\star(A)]}(x) : \delta \in [0;1]\right\} \quad \forall\, x \in \mathbb{R}$$

4 Bayes' Theorem for Fuzzy A-Priori Density and Fuzzy Data

For continuous stochastic model $X \sim f(\cdot \mid \theta)$, $\theta \in \Theta$, where $f(\cdot \mid \theta)$ denotes the probability density of X, M_X the observation space, continuous parameter space Θ, a-priori density $\pi(\cdot)$ on Θ, and observed complete sample $x_1, \cdots, x_n$, the standard Bayes' theorem gives the a-posteriori density $\pi(\cdot \mid x_1, \cdots, x_n)$ by its values

$$\pi(\theta \mid x_1,\cdots,x_n) = \frac{\pi(\theta)\cdot\prod_{i=1}^{n} f(x_i \mid \theta)}{\int\limits_{\Theta} \pi(\theta)\cdot\prod_{i=1}^{n} f(x_i \mid \theta)\, d\theta} \qquad \forall\, \theta \in \Theta.$$

Remark 3. The function whose values are $\prod_{i=1}^{n} f(x_i \mid \theta)$ is the *likelihood function* for complete data denoted by $\ell(\,\cdot\,; x_1,\cdots,x_n)$, considered as a function of the variable θ.

For fuzzy data $x_1^\star,\cdots,x_n^\star$ the likelihood function becomes a fuzzy valued function $\ell^\star(\cdot,\underline{x}^\star)$ whose values $\ell^\star(\theta;\underline{x}^\star)$ are assumed to be fuzzy intervals with characterizing functions $\psi_{\ell^\star(\theta;\underline{x}^\star)}(\cdot)$. The values of this characterizing function are obtained by application of the so-called *extension principle* from fuzzy set theory:

$$\psi_{\ell^\star(\theta;\underline{x}^\star)}(y) = \left\{ \begin{array}{ll} \sup\{\zeta(\underline{x}) : \ell(\theta;\underline{x}) = y\} & \text{if } \exists\ \ \underline{x} \in M_X^n : \ell(\theta;\underline{x}) = y \\ 0 & \text{if } \not\exists\ \ \underline{x} \in M_X^n : \ell(\theta;\underline{x}) = y \end{array} \right\} \ \forall\, y \in \mathbb{R}$$

Taking the δ-cuts of $\psi_{\ell^\star(\theta;\underline{x}^\star)}(\cdot)$ and denoting them by $\left[\underline{\ell}_\delta(\theta;\underline{x}^\star);\overline{\ell}_\delta(\theta;\underline{x}^\star)\right]$, for variable θ we obtain two classical real valued functions $\underline{\ell}_\delta(\cdot;\underline{x}^\star)$ and $\overline{\ell}_\delta(\cdot;\underline{x}^\star)$, these are the δ-level functions.

Based on these functions the fuzzy a-posteriori density $\pi^\star(\cdot \mid x_1^\star,\cdots,x_n^\star) = \pi^\star(\cdot \mid \underline{x}^\star)$ is defined by its δ-level functions $\underline{\pi}_\delta(\cdot \mid \underline{x}^\star)$ and $\overline{\pi}_\delta(\cdot \mid \underline{x}^\star)$ respectively, using the δ-level functions $\underline{\pi}_\delta(\cdot)$ and $\overline{\pi}_\delta(\cdot)$ of the fuzzy a-priori density $\pi^\star(\cdot)$ for all $\delta \in (0;1]$:

$$\left.\begin{array}{l} \overline{\pi}_\delta(\theta \mid \underline{x}^\star) = \dfrac{\overline{\pi}_\delta(\theta)\cdot\overline{\ell}_\delta(\theta;\underline{x}^\star)}{\int\limits_{\Theta} \frac{1}{2}\left[\underline{\pi}_\delta(\theta)\cdot\underline{\ell}_\delta(\theta;\underline{x}^\star) + \overline{\pi}_\delta(\theta)\cdot\overline{\ell}_\delta(\theta;\underline{x}^\star)\right] d\theta} \\ \qquad \text{and} \\ \underline{\pi}_\delta(\theta \mid \underline{x}^\star) = \dfrac{\underline{\pi}_\delta(\theta)\cdot\underline{\ell}_\delta(\theta;\underline{x}^\star)}{\int\limits_{\Theta} \frac{1}{2}\left[\underline{\pi}_\delta(\theta)\cdot\underline{\ell}_\delta(\theta;\underline{x}^\star) + \overline{\pi}_\delta(\theta)\cdot\overline{\ell}_\delta(\theta;\underline{x}^\star)\right] d\theta} \end{array}\right\} \forall\, \theta \in \Theta$$

This fuzzy a-posteriori density is the basis for Bayesian inference in case of fuzzy information.

Remark 4. The above definition of the δ-level functions of the a-posteriori density keeps the sequential nature of the updating procedure from standard Bayes' theorem.

5 Applications of Fuzzy A-Posteriori Densities

In standard Bayesian inference the a-posteriori density can be used for different statistical procedures: Confidence regions, predictive densities, calculation of probabilities of parameter hypotheses, construction of Bayesian decisions based on utility functions and others.

For fuzzy a-posteriori densities generalized procedures are available and described below.

5.1 Fuzzy Confidence Regions

Standard confidence regions for the parameter θ of a stochastic model $X \sim f(\cdot \mid \theta)$, $\theta \in \Theta$ in the Bayesian context are based on the a-posteriori density $\pi(\cdot \mid D)$ of $\tilde{\theta}$. A confidence region $C_{1-\alpha}$ for confidence level $1-\alpha$ is defined by

$$\int\limits_{C_{1-\alpha}} \pi(\theta \mid D) d\theta = 1 - \alpha.$$

For fuzzy data $D^\star = (x_1^\star, \cdots, x_n^\star)$ the combined fuzzy sample $\underline{x}^\star$ is the basis for the construction of generalized confidence regions, which are fuzzy subsets of the parameter space Θ.

Let $\zeta(\cdot, \cdots, \cdot)$ be the vector-characterizing function of $\underline{x}^\star$. For $\underline{x} \in \mathrm{supp}(\underline{x}^\star)$, a-posteriori density $\pi(\cdot \mid \underline{x})$, and confidence level $1-\alpha$ a corresponding confidence region is denoted by $C_{\underline{x},1-\alpha}$. The confidence set $C^\star_{1-\alpha}$ based on the fuzzy combined sample $\underline{x}^\star$ is the fuzzy subset of Θ whose membership function $\varphi(\cdot)$ is given by its values $\varphi(\theta)$ in the following way:

$$\varphi(\theta) := \left\{ \begin{array}{ll} \sup\left\{\zeta(\underline{x}) : \theta \in C_{\underline{x},1-\alpha}\right\} & \text{if } \exists\ \underline{x} \in M_X^n : \theta \in C_{\underline{x},1-\alpha} \\ 0 & \text{if } \nexists\ \underline{x} \in M_X^n : \theta \in C_{\underline{x},1-\alpha} \end{array} \right\} \forall\, \theta \in \Theta$$

Remark 5. For classical samples $\underline{x} = (x_1, \cdots, x_n)$ the membership function $\varphi(\cdot)$ is the indicator function of the standard confidence region $C_{\underline{x},1-\alpha}$.

5.2 Predictive Densities

Standard predictive densities for $X \sim f(\cdot \mid \theta)$, $\theta \in \Theta$ based on $\pi(\cdot \mid D)$ are given by the marginal density

$$p(x \mid D) = \int\limits_{\Theta} f(x \mid \theta)\, \pi(\theta \mid D) d\theta \quad \forall\, x \in M.$$

For fuzzy a-posteriori densities $\pi^\star(\cdot \mid D^\star)$ the integration of a fuzzy valued function is necessary. This is possible by the generalized integral from Section 3:

$$p^\star(x \mid D^\star) := \oint\limits_{\Theta} f(x \mid \theta) \cdot \pi^\star(\theta \mid D^\star) d\theta$$

Remark 6. The fuzzy valued function $p^\star(\cdot \mid D^\star)$ is a fuzzy density in the sense of Section 3. This fuzzy density can be graphically displayed by several δ-level functions $\overline{p}_\delta(\cdot \mid D^\star)$ and $\underline{p}_\delta(\cdot \mid D^\star)$.

5.3 Probabilities of Parameter Hypotheses

For parametric hypothesis $\mathscr{H} : \theta \in \Theta_0 \subset \Theta$ it is possible to calculate a-posteriori probabilities based on fuzzy a-posteriori densities $\pi^\star(\cdot | D^\star)$. This is an application of the definition of fuzzy probabilities in Remark 2 in Section 3.

The a-posteriori probability of the hypothesis $\mathscr{H}$ is

$$Pr(\mathscr{H} \mid D^\star) = \int\limits_{\Theta_0} \pi^\star(\theta \mid D^\star) d\theta$$

which is a fuzzy interval whose support is a subset of $[0; 1]$.

5.4 Bayesian Decisions Based on Fuzzy Information

Standard Bayesian decisions based on classical probability distributions on the state space Θ and utility functions $U(\cdot,\cdot)$, where $U(\theta,d)$ is the utility of the decision $d \in \mathscr{D}$ if the considered system is in state $\theta \in \Theta$, are defined by maximization of the expected utility $\mathbb{E}U(\theta,d)$ where the Bayesian decision d_B obeys

$$\mathbb{E}U(\theta,d_B) = \max_{d\in\mathscr{D}} \mathbb{E}U(\theta,d).$$

In case of continuous state space Θ and probability density $\pi(\cdot)$ on Θ, the expected utility of the decision d is given by

$$\mathbb{E}U(\theta,d) = \int_{\Theta} U(\theta,d)\pi(\theta)\,d\theta.$$

If both, the utility function $U^\star(\theta,d)$ and the probability density $\pi^\star(\cdot)$ on Θ are fuzzy, we obtain a generalized (fuzzy) expected utility

$$\mathbb{E}U^\star(\theta,d) = \int\!\!\!\!\!\!\int_{\Theta} U^\star(\theta,d)\pi^\star(\theta)\,d\theta.$$

This is again the integral of a fuzzy valued function $g^\star(\cdot)$ as defined in Definition 3. The result is a fuzzy interval $\mathbb{E}U^\star(\theta,d)$ whose characterizing function $\varepsilon(\cdot)$ is obtained from its δ-cuts $C_\delta[\varepsilon(\cdot)]$ with

$$\begin{aligned} C_\delta[\varepsilon(\cdot)] &= \left[\int_{\Theta} \underline{g}_\delta(\theta)\underline{\pi}_\delta(\theta)\,d\theta; \int_{\Theta} \overline{g}_\delta(\theta)\overline{\pi}_\delta(\theta)d\theta\right] \quad \forall\,\delta\in(0;1]. \\ &= \left[\underline{\varepsilon}_\delta;\overline{\varepsilon}_\delta\right]. \end{aligned}$$

The characterizing function $\varepsilon(\cdot)$ is obtained by the Representation lemma for membership functions, i. e.

$$\varepsilon(x) = \max\left\{\delta\cdot I_{[\underline{\varepsilon}_\delta;\overline{\varepsilon}_\delta]}(x) : \delta\in[0;1]\right\} \quad \forall\, x\in\mathbb{R}.$$

References

1. Bandemer, H.: Mathematics of Uncertainty - Ideas, Methods, Application Problems. Springer, Berlin (2006)
2. Klir, G., Yuan, B.: Fuzzy Sets and Fuzzy Logic - Theory and Applications. Prentice Hall, Upper Saddle River (1995)
3. Viertl, R.: Foundations of fuzzy Bayesian inference. J. Uncert. Syst. 2(3) (to appear, 2008)
4. Viertl, R.: Fuzzy models for precision measurements. Math Comput Simulat (in press, 2008) doi:10.1016/j.matcom.2008.02.013

Soft Methods for Treating Uncertainties: Applications in the Field of Environmental Risks

Dominique Guyonnet, Gaël Bellenfant, and Olivier Bouc

BRGM, Orléans, France

Abstract. In recent years, soft methods for treating uncertainties have reached the field of environmental risks. This paper presents some specific characteristics of this field and how they can affect the choice of methods and the way of representing results. Example applications illustrate the use of such methods in several areas of environmental risks: soil and groundwater contamination, health risks and risks related to the underground storage of carbon dioxide for climate change mitigation. A potential drawback of these methods in a decision-making framework is the fact that decision-makers do not like to work with multiple indicators, i.e. upper and lower bounds on the probability of exceeding a certain risk threshold. A way to avoid this difficulty is proposed, based on a weighted average of optimistic and pessimistic bounds on risk focal elements.

1 Introduction

Since a few decades, public awareness regarding the reality of environmental risks has grown considerably, as practically all natural environments are being adversely impacted by human activity. The most notable changes in public perception have occurred in recent years as a result of tangible evidence that human activity is significantly altering the Earth's climate, with potentially drastic consequences for life on the planet. Most industrialized countries have now enforced legislation aimed at regulating human activity so that it can be compatible with a more sustainable development (see for example Community legislation on water management: [15]; waste management: [16], or soil management: [4]). Legislation regarding environmental issues relies largely on the notion of "risk", which is defined here as the degree of "likelihood" that adverse effects might result from a given human activity. Existing legislation typically propose thresholds, e.g. for pollutant concentrations in water fit for human consumption, that should not be exceeded.

The field of environmental risks is characterized by at least two important features: (i) the considerable complexity of mechanisms involved and hence related uncertainties and (ii) a general attitude of "aversion to risk". Environmental issues are multidisciplinary by nature. A simple schematic of a typical framework relevant to environmental issues is depicted in Figure 1. The source could be for example a chimney on an industrial site sending pollutants into the atmosphere. The target could be a group of persons living at some distance from the chimney. The vector is the natural mechanism through which the targets may come into contact (become exposed) with the pollutants emitted by the source. In this example it could be windborne pollutants inhaled by the targets or else deposited pollutant uptake by vegetables grown and consumed by the targets,

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 16–26, 2008.
springerlink.com

Fig. 1. Schematic of a typical framework in a context of environmental risk

etc. In general, the mechanisms and parameters that influence how targets may become exposed to the source are complex and/or poorly understood, resulting in significant uncertainties when assessing environmental risks.

The attitude of "aversion to risk" has some important consequences with respect to the way uncertainties are treated in environmental risk assessments. In a many cases the main question is whether or not an "acceptable" threshold might be exceeded at any time in the future. For this reason the treatment of uncertainties in environmental risk assessments often relies on limiting scenario calculations. Ranges of possible values are defined for the parameters that are considered uncertain and interval-type analysis (possibly using optimization techniques) is performed to examine consequences on the estimation of risk. This may lead to the three possible outcomes depicted schematically in Figure 2. Either (Figure 2a) the threshold lies above the maximum calculated "possible" risk, in which case the risk is considered totally acceptable, or (Figure 2b) it lies below the smallest calculated risk, in which case the risk is considered as totally unacceptable. But in many cases (Figure 2c), the range of possible outcomes is so wide that the threshold lies within this range, in which case it is necessary to be able to discriminate between the relative likelihoods of possible outcomes.

In order to discriminate between levels of likelihood, researchers and decision-makers in the field of environmental risks often rely on the use of unique probability distributions (PDFs). The problem of course is that, in this field, we are typically confronted with information that is incomplete/imprecise and therefore the choice of unique PDFs is arbitrary. For example, in an analysis of uncertainties related to the risk of propagation of chlorinated organic pollutants in groundwater, MacNab et al. [14] defined twelve probability distributions for their model parameters, nine of which were qualified as "postulated". As shown by previous researchers (e.g. [7]), this approach, combined with the frequent hypothesis of parameter independence, may lead to very unconservative conclusions with respect to risk.

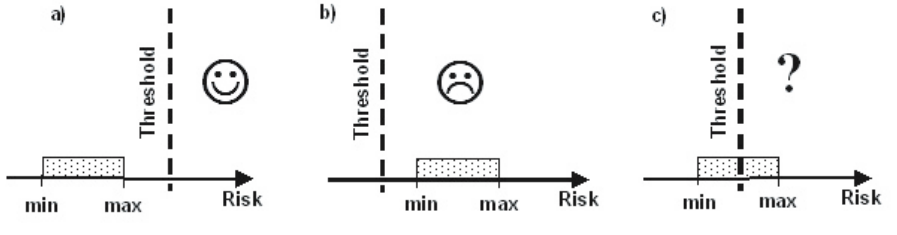


Fig. 2. Schematic of alternative responses to a risk threshold in environmental applications

In recent years, a better understanding of the fundamental differences between stochastic and epistemic uncertainties has pervaded to the field of environmental risks. As a result, researchers in the field of environmental risks have drawn on alternative methods for representing and propagating information relative to uncertain risk model parameters (see for example, [8, 9, 1, 2, 3]). Of course this evolution does not imply that "soft" methods for addressing uncertainties are becoming routine practice, for example by consulting companies involved in environmental risk assessment studies. Barriers with respect to a more widespread use of such methods in a decision-making framework and possible ways to overcome such barriers are discussed at the end of this paper.

2 Different Types of Information Warrant Different Methods of Information Representation and Propagation

A fuzzy set Possibly one of the most important reasons for using alternative methods for representing uncertainties in environmental risk assessments is: consistency with available information. While the complexity of processes and mechanisms involved are such that it is not possible to claim "validity" of proposed risk model predictions, investigators can at least hope to claim methodological "consistency". When an investigator, faced with incomplete/imprecise information, chooses to overlook this basic information character and to "disguise" it in the form of stochastic variability by postulating a unique PDF, he/she is misrepresenting reality, usually for reasons of methodological comfort. The message that the authors of this paper have been trying to bring forth to the environmental risk community is: first look at your data, and then choose the most appropriate conceptual framework to represent and propagate that data.

The types of information that are typically available in a context of environmental risks can be of a very different nature. In ideal situations, there may be abundant information regarding for example time-series of rainfall or temperature, in which case a stochastic representation may be the appropriate choice, relying on classical probability theory. But in many cases, information is scarce and/or imprecise, and alternative information theories can be used instead (e.g. possibility theory: [19, 5]; evidence theory: [17]; random set theory: [13], etc.).

In order to promote a systematic treatment of uncertainties that takes into account the basic nature of available information, we have attempted to develop a flowchart destined to guide investigators. The entry point to this flowchart (Figure 3) is whether or not the investigator wishes to represent a given risk model parameter by a single value? There may be various reasons for choosing to use a single parameter value; objective or subjective. For example, the investigator may know that the parameter is indeed a single value (e.g. the height of a chimney stack), or he may know that he will never have information regarding the parameter's variability (whether spatial or temporal) and therefore choose to use a single, albeit imprecise, value. Once the user of the flowchart has chosen whether he wishes to use a single parameter value or not, he is guided through a series of questions which help him select the most appropriate tool for representing the information.

The list of tools, which is by no means exhaustive, is drawn from the information theories cited above and try to cover the variety of "degrees of precision" typically encountered in the field of environmental risks. For example, if the user has a sufficient number of precise measurements relative to his parameter, and the parameter variability is temporal, then a single probability distribution function would seem the prime choice. If this variability is spatial, then geostatistical techniques can be used. But if the user must rely on incomplete information such as expert judgement, imprecise and scarce measurements, etc., then intervals, fuzzy intervals, random sets, parametric probability families, etc. may be more appropriate.

While inherently incomplete, the main benefit of the proposed flowchart is to bring the user to realize that there is no one-all-fit-all method for representing uncertain information. All depends on the type of information. Once an appropriate method of information representation has been selected for all uncertain risk model parameters, this information can be propagated using various techniques, the choice of which depends not only on the information representation tools, but also on possible dependencies between model parameters. Methods have been developed that are able to accommodate both stochastic and epistemic uncertainties in a single computation of risk. A robust method, that has been shown to be a systematically conservative counterpart of classical Monte Carlo calculations performed under hypotheses of parameter independence, is the so-called "Independent Random Set; IRS" method ([2, 3]). This method combines the random sampling of both PDFs (representing random variability) and fuzzy sets (representing incomplete/imprecise information) with optimisation techniques in order to estimate minimum and maximum values of risk, thus defining focal elements. Repeated iteration leads to a random set that can be expressed in terms of a minimum (optimistic) and maximum (pessimistic) level of probability of exceeding a given threshold. Such methods are illustrated below by several applications in the field of environmental risks.

3 Example Applications

The applications presented below are illustrated schematically in order to provide the reader with some insight into some typical problems addressed in the field of environmental risks and also the types of uncertainties involved.

3.1 Soil Contamination

In Guyonnet et al. [9], soft methods were used to assess risks of exposure through the consumption of vegetables, related to emissions from a chimney on a metallurgical industrial site. Deposition of cadmium on the soils surrounding the site, and the consumption of vegetables grown on these soils, were identified as a possible means of exposure of local populations to this metal, which is known to have toxic effects on the kidney ([18]). Primary uncertainties in this study were related to (i) the distribution of cadmium in the soils surrounding the industrial site, (ii) the amount of cadmium uptake by plants grown on these soils, (iii) the amount of vegetables consumed by local populations living around the site.

Regarding soil cadmium concentrations, a significant number of measurements were analyzed using geostatistical methods in order to provide information on spatial vari-

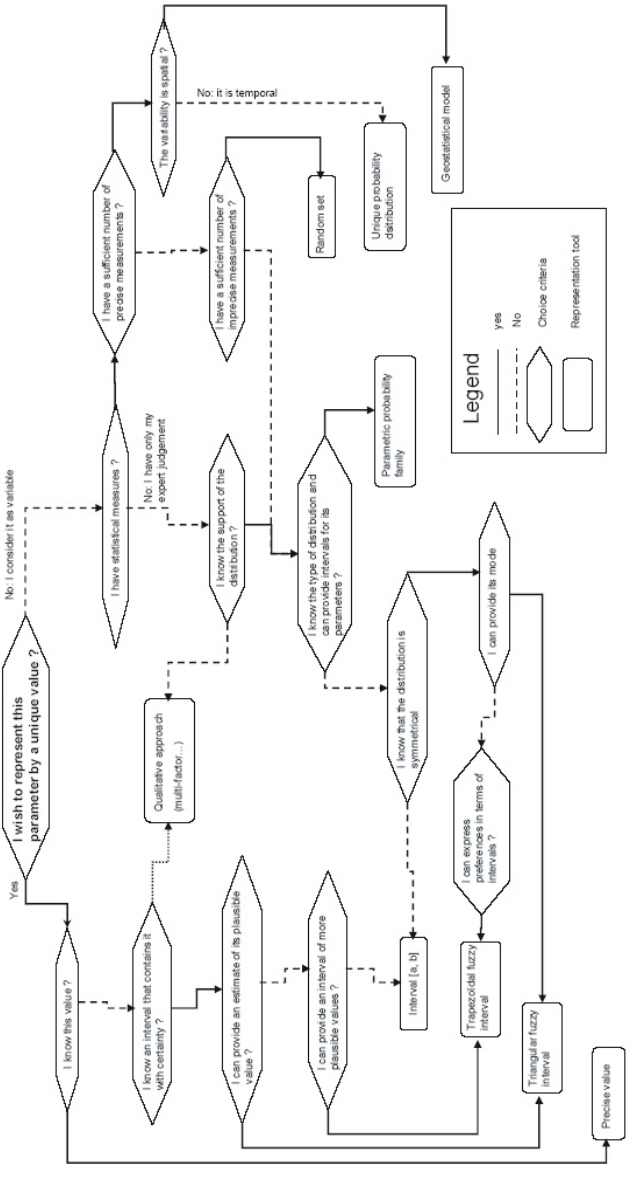


Fig. 3. Flowchart for helping choose the appropriate mode of information representation as a function of the nature of the information

ability. Information regarding (ii) and (iii) was of a more epistemic nature and possibility distributions were used. A so-called "hybrid" method was developed during this study in order to propagate the different sources of information. This method combines Monte Carlo sampling of PDFs (in this case related to soil cadmium concentrations)

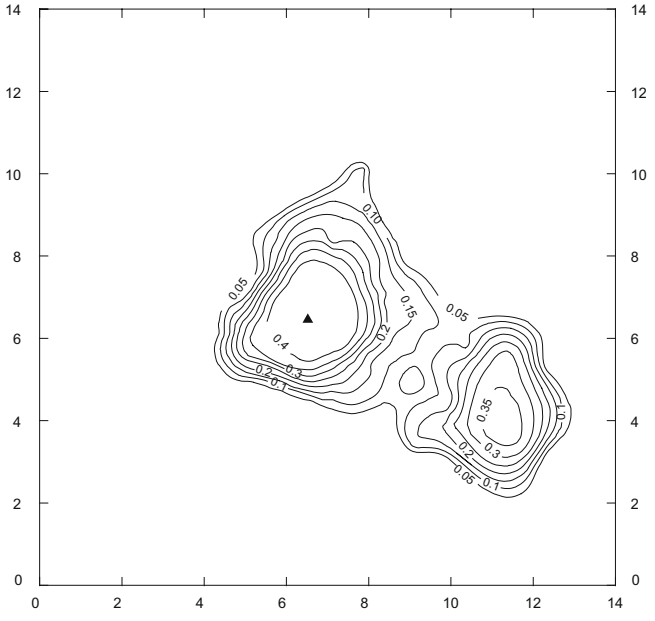


Fig. 4. Map of the "possibility" that the absorbed dose should exceed the reference dose (1 μg/d kg^{-1}). Graduation in km; Triangle = chimney location (from [9]).

with fuzzy calculus on possibility distributions. Results are expressed in terms of a family of distributions of the probability of exceeding a threshold, which in this case was ([18]) 1 microgram of cadmium per kilogram of person bodyweight. The method is slightly different from the IRS method mentioned previously in that it assumes dependence between possibilistic variables ([2]): a metadependence between information sources attached to the variables and also a dependence between variables themselves. However, results are very similar to those obtained with the IRS method.

Figure 4 presents some results of the analysis expressed as the upper level of probability (possibility; [6]) that the dose absorbed by a person consuming home-grown vegetables and living in the vicinity of the site should exceed the maximum threshold specified by the health authority. Possibilities of 0.45 of exceeding the reference dose are found in the close vicinity of the chimney, and decrease below 0.1 at a certain distance from the chimney. Such a map could be used in a decision-making framework to impose restrictions on the consumption of home-grown vegetables in this area.

3.2 Groundwater Contamination

In this example soft methods were used to assess the risk of exceeding a concentration threshold in groundwater located down-gradient from a contaminant spill ([3]). The contaminant is trichloroethylene, which is frequently a problem for groundwater because it is both persistent in the sub-surface (it does not degrade easily) and it is toxic (carcinogenic) at very low concentrations. Primary uncertainties in this study were (i)

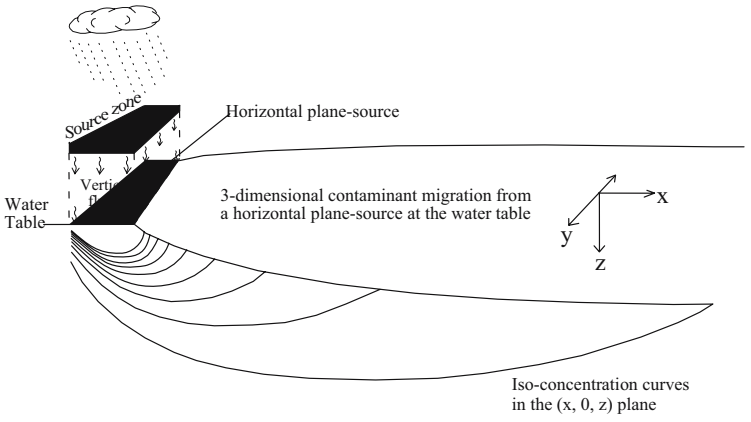


Fig. 5. Conceptual model of the groundwater risk model (Baudrit et al. [3])

the intensity of rainfall transporting contaminants down to the groundwater, (ii) dissolved trichloroethylene concentration in the infiltrating water, (iii) hydraulic conductivity of the aquifer and (iv) dispersion coefficients, which control pollutant dispersion in the groundwater.

The variability of rainfall is readily known from meteorological records and this parameter could be adequately represented by a single probability distribution function. On the other hand, it was preferred to represent the other three parameters using possibility distributions. One interesting aspect of the analysis was the fact that the "model" used to calculate concentrations down-gradient from the source, was more complex than those used to-date for hybrid-type propagation. As such methods involve both iterative sampling and optimization, calculation times may rapidly become restrictive. In that respect, the IRS method is faster than the "hybrid" method as there are only $2 \times n$ optimization problems due to the application of Monte Carlo sampling to the focal sets (n being the number of random sampling iterations).

The conceptual model of the calculation tool is depicted in Figure 5. The actual calculation tool is an analytical model that involves an integral and error functions. Results in Figure 6 illustrate the range of uncertainty in groundwater concentration at a point located down-gradient from the source resulting from uncertainty in risk model input parameters. Also shown is the result obtained when unique PDFs are assumed, despite the epistemic uncertainties, and applying the Monte Carlo method. The advantage of the classical Monte Carlo method is that a unique value of probability of exceeding the threshold is obtained. But this uniqueness is the result of the arbitrary selection of PDFs in presence of incomplete/imprecise information. This point will be further discussed below.

3.3 Risks Related to CCS: A Climate Change Mitigation Technology

Since a few years, BRGM has become strongly involved in the development of CCS; carbon capture and storage. The objective of CCS is to capture the carbon dioxide

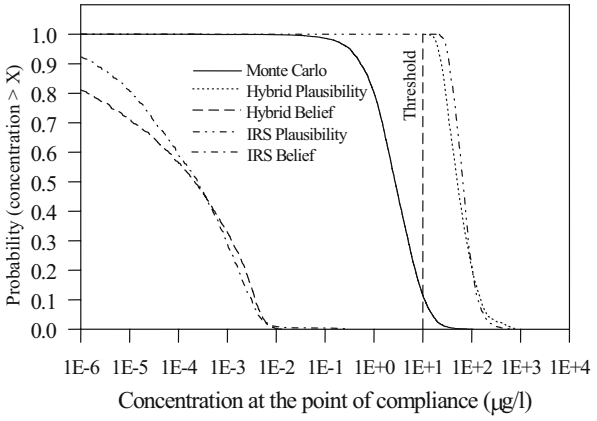


Fig. 6. Comparison between various uncertainty propagation methods (from Baudrit et al. [3])

emitted by large emittors (e.g. the steel industry) and to inject it into deep geological traps. Such deep injection has already been practiced since many years by the oil industry to enhance the productivity of oil fields, by pushing out residual oil. But in France, primary targets are deep saline aquifers such as the Dogger reservoir of the Paris Basin. The groundwater in this reservoir is so saline that it will never be considered as a

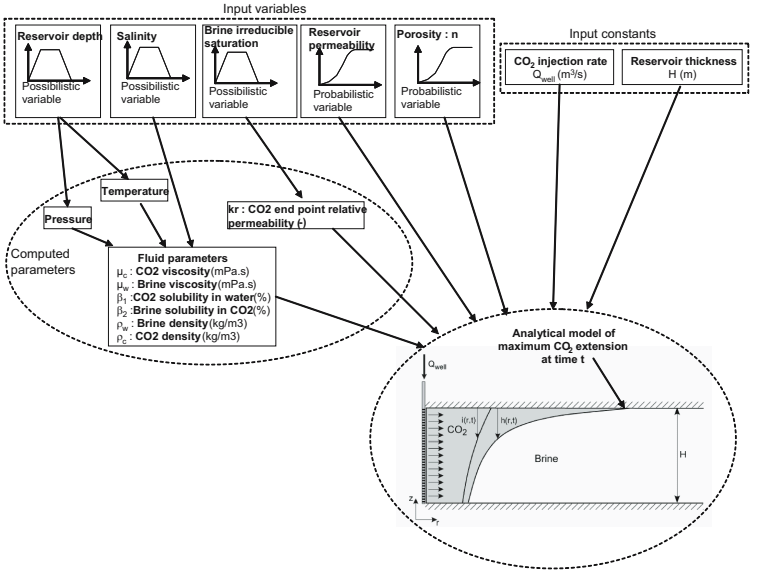


Fig. 7. Schematic of some uncertainties involved in the assessment of risks related to the deep injection of CO_2

potential groundwater resource. However, there are risks of CO_2 leakage through a variety of "features"; geological or man-made. Geological features include faults and other discontinuities while man-made features are essentially abandoned wells which may provide preferential pathways for CO_2 migration up to the ground surface. Such migration represents a health hazard because leaked CO_2, being heavier than air, might accumulate at ground surface and provoke asphyxia of nearby populations, as was the case with the infamous natural CO_2 eruption at Lake Nyos in the North-West of Cameroun, where over 1700 people died as well as large numbers of cattle.

Figure 7 illustrates the types of uncertainties involved in these risk calculations that are currently ongoing. Calculations to-date have related primarily to the extent of the CO_2 "bubble" around the injection point. But results will serve to calculate leakage fluxes through abandoned wells located in the vicinity of the injection well.

4 Discussion and Conclusions

As was stated previously, the significant advantage of arbitrarily selecting PDFs in the presence of incomplete/imprecise information and applying the Monte Carlo method is that a single value for the probability of exceeding a given threshold can be obtained. Decision-makers dealing with environmental risks often have difficulties let alone with the notion of probability, without having also to deal with upper and lower bounds on probability. In order to promote increased acceptance of soft methods in the field of environmental risks, it is necessary to introduce an additional treatment step and to provide a result that can be more easily "digested" by potential users.

In a context of "aversion to risk" one might suggest that the pessimistic bound on probability be used as the unique indicator of the acceptability of risk. This approach, while being conservative, presents the disadvantage of ignoring all the information leading to less pessimistic estimates of risk. It is proposed instead to adopt the approach of Jaffray [11, 12] in order to obtain a "reasonably conservative" estimation of risk. This approach, based on earlier work by Hurwicz [10] proposes to compute a single indicator as a weighted average of focal element bounds. The proposed probability measure is ([11]):

$$P_\alpha = \alpha e_m + (1-\alpha) e_M \tag{1}$$

where e_m and e_M are the minimum and maximum bounds of the focal elements.

The choice of weight α is subjective and reflects the attitude of the decision-maker with respect to risk. The concept is illustrated in Figure 8 which presents the results of a health risk calculation using a "hybrid" approach and a purely Monte Carlo approach. An additional curve is presented and calculated according to Equation (1) using a value $\alpha = 1/3$. This means that a weight of 1/3 is applied to the optimistic limit values of the focal elements, while a weight of $1 - 1/3 = 2/3$ is applied to the pessimistic values. In a context of aversion to risk, it would seem normal to privilege the pessimistic values, but without completely obliterating the optimistic one.

The curve is indicated in Figure 8 as a "Confidence Index". This term is borrowed from the field of meteorology. The meteorological community has extensive experience with respect to predicting natural events and also of communicating on these predictions with the general public. It is therefore significant that meteorologists should

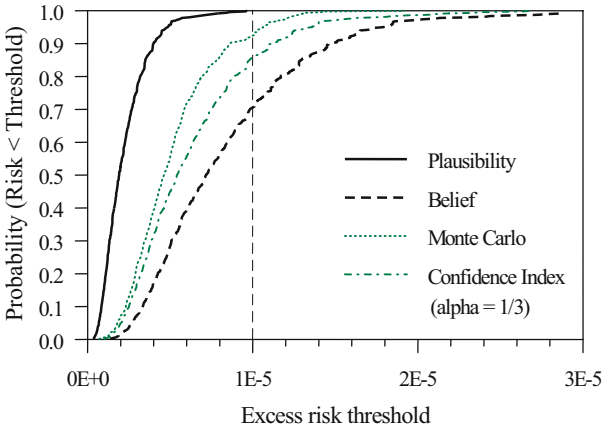


Fig. 8. Indicators of the probability of lying below a certain threshold of risk

have adopted the term "Confidence Index" to communicate on the uncertainty related to their predictions. In our view, the term holds value both from both a scientific and a sociological viewpoint. Scientific, because it avoids referring to any particular uncertainty paradigm (probabilistic, possibilistic, etc.). Sociological, because the notion of "confidence" has positive connotations. Referring to Figure 8, one would state that the computed risk is lower than the threshold (10^{-5}), with a Confidence Index of 85%.

The subjectivity introduced by such an approach can easily be justified in a decision-making framework as it does not attempt to "disguise" epistemic uncertainty in the form of stochastic variability, but offers a practical way for the decision-maker to express his level of aversion to risk. The risk assessor should try to faithfully transmit the available information, so that the range of possible outcomes should be known. If this range is judged too large, then measures might be taken in order to reduce uncertainties in model input parameters (e.g. through measurement campaigns). Such an outcome never ensues from a Monte Carlo analysis performed using postulated PDFs, as there is no way of distinguishing, in the computed results, variability resulting from true stochastic randomness from variability due to arbitrary assumptions.

Acknowledgement. The material presented in this paper is the result of a long collaboration between BRGM and IRIT at the Paul Sabatier University of Toulouse. Special thanks are due to Didier Dubois, who has provided the primary guidance for these developments.

References

1. Baudrit, C., Guyonnet, D., Dubois, D.: Post-processing the hybrid method for addressing uncertainty in risk assessments. J. Envir. Eng. 131(12), 1750–1754 (2005)
2. Baudrit, C., Guyonnet, D., Dubois, D.: Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment models. IEEE Trans. Fuzzy Syst. 14(5), 593–608 (2006)

3. Baudrit, C., Guyonnet, D., Dubois, D.: Joint propagation of variability and partial ignorance in a groundwater risk assessment. J. Contamin. Hydrol. 93, 72–84 (2007)
4. CEC, Proposal for a Directive of the European Parliament and of the Council establishing a framework for the protection of soil and amending. Directive 2004/35/EC, COM 2006/0086 of the Commission of the European Communities, Brussels (2006)
5. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertaint. Plenum Press, New York (1988)
6. Dubois, D., Prade, H.: When upper probabilities are possibility measures. Fuzzy Sets Syst. 49, 65–74 (1992)
7. Ferson, S.: What Monte Carlo methods cannot do. Human Ecol. Risk Assess 2, 990–1007 (1996)
8. Ferson, S., Ginzburg, L.R.: Different methods are needed to propagate ignorance and variability. Reliab. Eng. Syst. Saf. 54, 133–144 (1996)
9. Guyonnet, D., Bourgine, B., Dubois, D., Fargier, H., Côme, B., Chilès, J.P.: Hybrid approach for addressing uncertainty in risk assessments. J. Envir. Eng. 129, 68–78 (2003)
10. Hurwicz, L.: Optimality criteria for decision making under ignorance. Cowles Commission discussion paper, Statistics No. 370 (1951)
11. Jaffray, J.-Y.: Application of linear utility theory for belief functions. In: Bouchon-Meunier, B., Saitta, L., Yager, R.R. (eds.) IPMU 1988. LNCS, vol. 313, pp. 1–8. Springer, Heidelberg (1988)
12. Jaffray, J.-Y.: Dynamic decision making with belief functions. In: Yager, R.R., Kacprzyk, J., Fedrizzi, M. (eds.) Advances in the Dempster-Shafer theory of evidence, John Wiley & Sons, New York (1994)
13. Kendall, D.G.: Foundations of a theory of random sets. In: Harding, E.F., Kendall, D.G. (eds.) Stochastic Geometry, pp. 322–376. Wiley, New York (1974)
14. McNab, W., Rice, D., Bear, J., Ragaini, R., Tuckfield, C., Oldenburg, C.: Historical case analysis of chlorinated volatile organic compound plumes. Report UCRL-AR-133361, Lawrence Livermore National Laboratory, California, USA (1999)
15. OJEC, Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000, establishing a framework for community action in the field of water policy. Official Journal of the European Communities, 22 December 2000, L327 (2000)
16. OJEC, Directive 2006/12/EC of the European Parliament and of the Council of 5 April 2006 on waste. Official Journal of the European Communities, 27 April 2006, L114 (2006)
17. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
18. WHO, Quality Directives for Drinking Water, Vol 1: Recommendations, 2nd edn. World Health Organisation, Geneva, Switzerland (1994)
19. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1, 3–28 (1978)

Robert Féron: A Pioneer in Soft Methods for Probability and Statistics

Jean-Paul Auray[1] and Henri Prade[2]

[1] LIRIS-MA2D, CNRS and Université Lyon 1, Villeurbanne, France
[2] IRIT, CNRS and Université Toulouse 3, Toulouse, France

Abstract. Robert Féron invented fuzzy random variables in the mid-seventies. As such, his works deserve due recognition among specialists of soft methods in probability and statistics. This short paper surveys his contributions to information theory, generalized distances, and the joint use of probability and fuzzy set theories. An extensive bibliography of his publications is provided.

1 Robert Féron's Career

Robert Féron was 33 years old when he obtained the Doctorat ès Sciences degree in mathematics. He defended it on December 21, 1954, at the Faculté des Sciences de l'Université of Paris, in front of a committee chaired by Maurice Fréchet and where Robert Fortet and Daniel Dugué were the examiners.

The main part of this thesis, entitled "Information, Régression, Corrélation" [4], was the result of about ten years of research, in an area of mathematics considered as applied, where interesting progress was appearing. It is within the framework of the young CNRS (National Centre for Scientific Research) that this research was taking place, since Robert Féron had been accepted as "Stagiaire de Recherche" in 1945, and then as "Attaché de Recherche". After obtaining his "Doctorat d'Etat", he was found worth becoming a "Chargé de Recherche" in 1957, and was then admitted on the very restrictive list of people having the recognized capability to be appointed as "Maître de Recherche" (currently equivalent to 2nd-class "Directeur de Recherche"). But his career at CNRS ended in 1958, when he became a "Maître de Conférence" (currently equivalent to Associate Professor) at the Faculté of Sciences of Université of Lyon. Among his new colleagues were some members of the highly regarded SMR (Société Mathématique de France), to which he had been belonging since 1946. He remained a Professor at this University (now named "Université Lyon 1") until his retirement in 1986.

2 Contributions to Probability, Information, Correlation and Generalized Distances

His thesis contains original developments in the area of the functional characteristics of random variables. He first defined the concept of uncertainty in gathering different

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 27–32, 2008.
springerlink.com

approaches already developed, by giving some coherent axiomatic foundations based on the notion of cumulative probability distributions. It is noticeable that, to achieve this purpose, Robert Féron used the concepts of probability space and of Borel sets of $\mathbb{R}$; this deserves to be pointed out if we remember that, in those times, some probability scholars were still reluctant to use this approach, considered by them as being too much oriented towards set-theoretic tools for their taste.

On such a basis, he then developed an original work on the concept of information gain, studied in the case of pairs of random variables (X,Y), the starting idea being to analyze to what extent the knowledge of information on the values taken by X entails a reduction of the uncertainty pervading the values taken by Y. In this perspective, the use of the probabilistic approach, based on measure theory, enables a unified, homogeneous and consistent formalization of concepts previously defined in particular settings. Taking as an example the concept of regression, especially of regression lines, the advantages become clear, since the proposed model, always based on a probabilistic approach, encompasses different types of uncertainty measures, including the so-called *typical values* introduced by Fréchet, some years before. The section devoted to the concept of correlation unifies, without particular restrictions on the nature of the variables, Pearson correlation ratio, the classical R^2, but also ratios obtained by taking Fréchet's typical deviation as a measure of uncertainty. This was naturally generalized to n-tuples of variables.

His thesis ended with two independent chapters. One was devoted to the study of indices liable to characterize the fact that two variables are linked by a functional; two Gini connexion indices, simple or quadratic, and a Jordan index were examined. The other chapter corresponds to a requirement, compulsory for obtaining this kind of thesis at the time, namely treating a topic chosen by the committee. In this case, an attempt at generalizing some notions of processes with independent random variables, of simple Markov chains, of stochastic processes that are stationary or with independent increases.

Several of the results presented in his thesis led to publications either in the "Journal de la Société Statistique de Paris", or in the "Publications de l'Institut Statistique de Paris", alone or in collaboration with Claude Fourgeaud, but also as "Notes" in the "Comptes Rendus de l'Académie des Sciences (CRAS)" (3 such Notes were published in years 1951-1952 [1, 2, 17]).

Once in Lyon University, Robert Féron, without forgetting his scientific origins, extended his field of research. He launched a small research group, that grew with years, and became a team officially associated to CNRS in the seventies. In particular, he started at that time to be interested in problems coming from economics. Especially he noticed that, quite often, the available mathematical tools were not always appropriate for dealing with the economical, or more generally social sciences contexts, and he tried to create specific mathematical tools accordingly.

He first devoted some time to a generalization of metric spaces, towards a weakening of the axiomatics, in order to make them more appropriate to the handling of problems that are no longer coming from "hard" sciences such as physics. He then took up again one of Fréchet's ideas, that amounts to no longer requiring the triangle inequality, and developed it. The subject-matter of a 1966 CRAS note [6], this work led to the building of a new concept of space, initially named F-topological space, as a tribute to Fréchet, but maybe also in consonance with the author's name. This work was then further developed

within Féron's research team, especially by Marcel Brissaud, who, in order to nicely couple the concepts of non transitive graphs and of topology, systematically investigated the weakening of Kuratowski's axiomatics. This led to the definition of different kinds of pre-topological spaces that generalize topological spaces. On such spaces, the classical concepts of separability, connexity, compacity, continuity, subspace, product space, ... are redefined sometimes with great subtleties and an unexpected diversity with respect to classical topology.

3 Contributions to Fuzzy Set Theory and Fuzzy Random Variables

Fuzzy set theory and fuzzy logic originated in Zadeh's 1965 seminal paper. After a few years, several researchers, coming from various scientific areas, became interested in this new idea. In France, among them, some were mathematicians, working in logic as Daniel Ponasse (also in Lyon), or working on probabilities as Robert Féron. They both examined the concept of fuzzy membership, the former studying the case where membership values belong to a lattice, the later looking at the differences and similarities with probability calculus. The few other researchers that were already working on fuzzy sets in France around 1975 were Arnold Kaufmann, who made huge efforts to develop and popularize the basic concepts of the new theory, while Claude Ponsard, prematurely deceased, studied its first applications to spatial economics, and Elie Sanchez introduced a fuzzy relational calculus and its applications in medicine.

Robert Féron has published several pieces of work on fuzzy sets since 1976. Several of them appeared in the "Publications Econométriques", a French journal he founded and edited from 1967 to 1986, and others were published as CRAS notes. In summary, in his work he considers a non empty set $\mathscr{X}$, equipped with a topological structure, and a closed lattice $\mathscr{S}$. A fuzzy set A is characterized by a mapping $\mu_A(.)$ from $\mathscr{X}$ to $\mathscr{S}$. A fuzzy set is regular if for all s in $\mathscr{S}$, the sets $\{x \in \mathscr{X}\,;\, \mu_A(x) \geq s\}$ and $\{x \in \mathscr{X}\,;\, \mu_A(x) \not\leq s\}$ are closed sets in $\mathscr{X}$. $\mathscr{Y}$ denotes the set of regular fuzzy sets and $\mathscr{B}$ a σ-algebra on $\mathscr{Y}$. Lastly, $(\Omega, \mathscr{A})$ is a measurable space. Then a random fuzzy set is a measurable mapping from $(\Omega, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$. From this construction, it is possible to get a generalization of the usual fuzzy concepts. Several particular cases are considered, as, e.g., one where the topology of $\mathscr{X}$ is based on a metric, one where $\mathscr{S}$ is finite, and one where $\mathscr{S} = [0,1]$.

It is worth pointing out that such a generalization was motivated by practical modeling issues, in order to go beyond the simple framework of standard fuzzy sets. For instance, Robert Féron wrote in [12, p. 84][1]:

> Similarly, a generalization of the theory of random sets with closed values has suggested us a generalization of the notion of random set and a theory that

[1] French text: De même, une généralisation de la théorie des ensembles aléatoires à valeurs fermées nous a suggéré une généralisation de la notion d'ensemble aléatoire et une théorie qui permet de tirer encore des conclusions dans le cas où on a un ensemble aléatoire flou. L'ingénieur a ainsi un outil mathématique puissant pour opérer sur autre chose que des champs aléatoires, puisque ici l'ensemble S n'est plus un sous-ensemble de $\mathbb{R}^n$, mais un treillis distributif fermé.

> still enables us to draw conclusions in the case where we have a random fuzzy set. The engineer has thus a powerful mathematical tool for dealing with other structures than random fields, since here the set S is no longer a subset of $\mathbb{R}^n$, but a closed distributive lattice.

We see here the interest of tools that offer the highest possible generalization. Maybe another illustration of similar views and attempts, is the famous integral developed at the same time by Sugeno.

Robert Féron did not restrict his research area to the design of mathematical tools, but he also thought of possible applications to economics, an area he had always been interested in. It is in the seventies, two decades after the seminal works of Arrow and Debreu, that one could see a blossoming of publications aiming at modeling the concept of *economy*, involving a finite number of goods, agents having some information at their disposal, as well as initial amounts, and who have to share out some wealth by possibly forming coalitions; in this context the objective was, thanks to fixed point theorems, to look for repartitions that no agent or group of agents could dispute, i.e. for "the core of an economy". Féron then pointed out that in a real context, coalitions possess every reason to be random fuzzy sets[2], and he proposed an original model leading to the concept of core of a fuzzy economy [8, 9], which could be for instance exploited in dynamical models, for forecasting purposes; however, being realistic, he remained cautious in his conclusion[3]:

> We must however notice that the construction of such a dynamical model requires a very difficult experimental study to be conducted on the actual behavior of economic agents.

Robert Féron continued to work until his retirement and later as well since in 1988 he published a paper [16], in collaboration with his son Marc, where he applied his results on random fuzzy sets to the evaluation, according to a finite number of criteria, of industrial products, such as cars.

Unsurprisingly, in his works from years 1970-1980, Robert Féron did not forget his previous research in the area of topology and pretopological structures, which led him to open some new research directions that are still partly to investigate. Let us mention along this line, his work on pre-uniform fuzzy structures. The concept of uniform structure, developed by Bourbaki should logically lead to such generalizations. This is what he did along two directions simultaneously, by, on the one hand, weakening the uniformity axioms (for instance only requiring that the family of *entourages* be a pre-filter - we then speak of *pre-entourages*), and on the other hand by introducing the concept of fuzzy uniform structure, allowing the entourages or the pre-entourages to be fuzzy sets [15].

Retrospectively, Robert Féron can be considered as a pioneer who always tried to follow the difficult way that consists in proposing to adapt existing modeling tools for

[2] The idea of fuzzy core in game theory had been introduced a few years before by Jean-Pierre Aubin.

[3] French text: Il convient toutefois de remarquer que la construction d'un tel modèle dynamique implique une étude expérimentale très difficile à effectuer du comportement réel des agents économiques.

dealing with application domains, while the temptation was often to develop theories only for their own sake, implicitly assuming that reality has to be adapted to the model! Robert Féron can be regarded as one of the very first researchers in soft methods in probability and statistics (using the name of the conference where the present homage is presented), especially for having been the first to consider random fuzzy sets, while at about the same time Fortet and Kambouzia, or Kampé de Fériet pointed out that the contour function of a random set may be viewed as a fuzzy set. The scientific community in fuzzy logic, and more specifically in soft methods in probability and statistics is thus indebted to Professor Robert Féron for his early contribution to the theory of random fuzzy sets, and their applications.

In fact, there is yet another important reason why the fuzzy set community has to pay a tribute to Professor Féron. Indeed, in 1980, Robert Féron took the risk of organizing a CNRS Round Table, entitled "*Quelques applications concrètes utilisant les derniers perfectionnements de la théorie du flou*" (in English: "Some concrete applications using the most recent advances in fuzzy set theory"), in Lyon (on June 23-24, 1980). Most of the active researchers in fuzzy sets at that time participated to this meeting[4]. This was a time when fuzzy sets were severely criticized by many scientists, in spite of the efforts made by a few others, and this was especially the case in France. Professor Féron nevertheless managed to put together what turned to be a very important meeting for the further development of the field. But what is perhaps still another remarkable indication of how broad was Professor Féron's understanding of the directions in which fuzzy sets should be developed, is the fact that he invited in the organization committee (which met on January 25, 1980) not only Kampé de Fériet, but also Gustave Choquet (1915-2006), (together with several much younger fuzzy set researchers) at a time where almost nobody had still a clear view of the importance of Choquet capacities and integrals in relation with fuzzy sets and other non-classical approaches to the handling of uncertainty and the modeling of preference.[5]

[4] The list of participants included: Jean-Paul Auray, Bernadette Bouchon, Marcel Brissaud, Christer Carlsson, Didier Dubois, Bernard Dubuisson, Alain Dussauchoy, Gérard Duru, Hubert Emptoz, Siegfried Gottwald, Arnold Kaufmann, Joseph Kampé de Fériet, Peter Klement, Abraham Kandel, Michel Lamure, Ramon Lopez de Mantaras, Robert Lowen, Noël Malvache, Constantin Negoita, Hung Nguyen, Serge Oppenchaim, Henri Prade, Daniel Ponasse, Michel Prévot, Dan Ralescu, Enrique Ruspini, Elie Sanchez, Philippe Smets, Robert Vallée, Didier Willeys, Lotfi Zadeh, Hans Zimmermann. Sorry for most probably forgetting many people. Unfortunately, the proceedings of this meeting were never published, although the papers of the main presentations were distributed to the participants.

[5] As a last minute member of this committee (due to the support of Professor C. V. Negoita), the last author of this note can however testify that apparently Professor Féron was not completely successful in communicating his enthusiasm for fuzzy set theory to Professor Choquet. To be fair, at that time, fuzzy sets were still in infancy, and had not much to offer, especially to mathematics, and only a few mathematicians like Joseph Kampé de Fériet, or Robert Fortet who had been working in information theory for many years, could foresee some interest in such a risky, and still incompletely formalized topic. It also took some more years before the parallel and the differences between Sugeno integral and Choquet integral could be more deeply understood.

4 Conclusion

The name of Robert Féron may not be familiar to a number of younger researchers in fuzzy sets or imprecise probabilities, since many of his papers are not easily accessible and most of them are in French, if we except [13, 14, 16]. This note is a modest tribute to a talented and very open-minded mathematician, with a broad vision of problems, and a great generosity towards his colleagues. We hope that it will help enlarging the recognition that he truly deserves.

Acknowledgement. This note was finalized with the kind help of Agnès Rico and Didier Dubois.

References

1. Féron, R.: Information et corrélation. Note CRAS 234, 1344–1345 (1952)
2. Féron, R.: Convexité et information. Note CRAS 234, 1840–1841 (1952)
3. Féron, R.: De la régression. Note CRAS 234, 2143–2145 (1952)
4. Féron, R.: Information, Régression, Corrélation. Thèse pour l'obtention du grade de Docteur es Sciences mathématiques (Thèse d'Etat), Pub Inst Stat De l'Univ Paris (1956)
5. Féron, R.: De l'information. Note CRAS 240, 1945–1947 (1958)
6. Féron, R.: Espaces à écart de Fréchet. Note CRAS 262, 278–280 (1966)
7. Féron, R.: Ensembles aléatoires flous. Note CRAS, Ser. A 282, 903–906 (1976)
8. Féron, R.: Economie d'échange aléatoire floue. Note CRAS, Ser. A 282, 1379–1382 (1976)
9. Féron, R.: Ensembles flous, ensembles aléatoires flous, et économie aléatoire floue. Publications Econométriques IX(1), 25–64 (1976)
10. Féron, R.: Ensembles flous attachés un ensemble aléatoire flou. Publications Econométriques IX(2), 51–65 (1976)
11. Féron, R.: Sur les notions de distance et d'écart dans une structure floue et leurs applications aux ensembles aléatoires flous. Cas où le référentiel n'est pas métrique. Note CRAS, Ser. A 289, 35–38 (1979)
12. Féron, R.: Ensembles aléatoires flous dont la fonction d'appartenance prend ses valeurs dans un treillis distributif fermé. Publications Econométriques XII(2), 81–118 (1979); Bibliographical addendum XII(2), 63–67
13. Féron, R.: Probabilistic and statistical study of random fuzzy sets whose referential is $\mathbb{R}^N$. In: Lasker, G.E. (ed.) Proceedings of the International Congress on Applied Systems Research and Cybernetics (Acapulco, 1980) Applied Systems and Cybernetics. Fuzzy Sets and Systems, Possibility Theory and Special Topics in Systems Research, vol. VI, pp. 2831–2836. Pergamon Press, New York (1981)
14. Féron, R.: On the notions of distance and deviation in a fuzzy structure. In: Trapple, R., Klir, G., Pichler, F. (eds.) Proceedings of the 5th European Meeting on Cybernetics and Systems Research (Vienna, 1980) Progress in Cybernetics and Systems Research, vol. 8, pp. 443–447. Hemisphere Pub. Corporation, Washington (1982)
15. Féron, R.: Structure pré-uniforme classique et structure pré-uniforme floue. Note de travail, ERA 639. CNRS-Université Lyon 1 (1987)
16. Féron, R., Féron, M.: Fuzzy specifications and random fuzzy events considered as basic tools for statistical prediction. Fuzzy Sets Syst 28, 285–293 (1988)
17. Féron, R., Fourgeaud, C.: Information et régression. Note CRAS 232, 1636–1638 (1951)

Part II

Foundations

Relating Prototype Theory and Label Semantics

Jonathan Lawry[1] and Yongchuan Tang[2]

[1] Department of Engineering Mathematics, University of Bristol, Bristol, United Kingdom
[2] College of Computer Science, Zhejiang University, Hangzhou, P.R. China

Abstract. An interpretation of the label semantics framework is introduced based on prototype theory. Within this interpretation it is shown that the appropriateness of an expression is characterised by an interval constraints on a parameter ε. Here ε is an uncertain distance threshold according to which an element x is sufficiently close to the prototype p_i of a label L_i for L_i to be deemed appropriate to describe x, if the distance between x and p_i is less than or equal to ε. Appropriateness measures and mass functions are then defined in terms of an underlying probability density function δ on ε.

1 Introduction

In classical logic a concept label L is defined by the set of elements from an underlying universe which satisfies L (the extension of L) and more generally in Kripke semantics [5] as a mapping from a set of possible worlds into sets of elements (an interpretation of L). Such an approach fails to capture certain aspects of our intuitive understanding of concepts in natural language, in particular the role of similarity in establishing the meaning of concept labels. Furthermore, a possible worlds model seems to overlook our natural focus on understanding reality as represented by one particular possible world (see [3] for discussion).

Prototype theory (Rosch [9]) is an alternative approach to concept representation according to which decisions regarding the applicability of a concept label to a particular instance are made on the basis of the similarity of that instance to a (set of) prototypical element(s) for that concept. Prototypes may not correspond to actual perceptions of objects or experiences but instead may identify a particular point or region of conceptual space [3] which is in some way representative of the concept. From this perspective the human ability to rank elements in terms of the degree to which they satisfy a concept L can be explained in terms of a comparison of their relative similarity (or distance) from the prototype(s) for L.

Prototype theory has been proposed as the basis for a possible interpretation of membership functions in fuzzy set theory ([1, 2]), where the membership of an element x in a concept L is taken to be a scaled version of the similarity between x and the prototype(s) for L [10]. This rather intuitive approach has the drawback that the prototype similarity interpretation of membership does not naturally result in a truth-functional calculus when concepts are combined (See Lawry [7] chapter 2 for a discussion). Consequently a prototype based model of membership does not seem to capture the underlying calculus of fuzzy set theory.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 35–42, 2008.
springerlink.com

Label semantics (Lawry [6, 7, 8]) is an uncertainty theory for vague concepts which encodes the meaning of linguistic labels according to how they are used by a population of communicating agents to convey information. From this perspective, the focus is on the decision making process an intelligent agent must go through in order to identify which labels or expressions can actually be used to describe an object or value. In other words, in order to make an assertion describing an object in terms of some set of linguistic labels, an agent must first identify which of these labels are appropriate or assertible in this context. Given the way that individuals learn language through an ongoing process of interaction with the other communicating agents and with the environment, then we can expect there to be considerable uncertainty associated with any decisions of this kind. In label semantics we quantify this uncertainty in terms of appropriateness measures, linked to an associated mass function through a calculus which, while not truth-function, can be functional in a weaker sense (See Lawry [6] and [7]). In the sequel we will propose a prototype theory interpretation of label semantics which relates both appropriateness measures and mass functions to distance from prototypes and naturally captures the label semantics calculus.

2 An Overview of Label Semantics

The underlying philosophy of label semantics [8] is very close to the epistemic view of vagueness as expounded by Timothy Williamson [12]. Williamson assumes that for the extension of a vague concept there is a precise but unknown dividing boundary between it and the extension of the negation of that concept. However, while there are marked similarities between the epistemic theory and the label semantics view, there are also some subtle differences. For instance, the epistemic view would seem to assume the existence of some objectively correct, but unknown, definition of a vague concept. Instead of this we argue that individuals when faced with decision problems regarding assertions find it useful as part of a decision making strategy to assume that there is a clear dividing line between those labels which are and those which are not appropriate to describe a given instance. We refer to this strategic assumption across a population of communicating agents as the *epistemic stance*, a concise statement of which is as follows:

> *Each individual agent in the population assumes the existence of a set of labelling conventions, valid across the whole population, governing what linguistic labels and expressions can be appropriately used to describe particular instances.*

The idea is that the learning processes of individual agents, all sharing the fundamental aim of understanding how words can be appropriately used to communicate information, will eventually converge to some degree on a set of shared conventions. The very process of convergence then to some extent vindicates the epistemic stance from the perspective of individual agents.

Label semantics proposes two fundamental and inter-related measures of the appropriateness of labels as descriptions of an object or value. We begin by assuming that for all agents there is a fixed shared vocabulary in the form of a finite set of basic labels

LA for describing elements from the underlying universe Ω. A countably infinite set of expressions *LE* can then be generated through recursive applications of logical connectives to the basic labels in *LA*. The measure of appropriateness of an expression $\theta \in LE$ as a description of instance x is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective probability that θ can be appropriately used to describe x. From an alternative perspective, when faced with describing instance x, an agent may consider each label in *LA* and attempt to identify the subset of labels that are appropriate to use. This is a totally meaningful endeavour for agents who adopt the epistemic stance. Let this complete set of appropriate labels for x be denote by $\mathscr{D}_x$. In the face of their uncertainty regarding labelling conventions agents will also be uncertain as to the composition of $\mathscr{D}_x$, and we represent this uncertainty with a probability mass function $m_x : 2^{LA} \to [0,1]$ defined on subsets of labels. We now provide formal definitions for the set of expressions *LE* and for mass functions m_x, following which we will propose a link between the two measures $\mu_\theta(x)$ and m_x for expression $\theta \in LE$.

Definition 1 (Label Expressions)
The set of label expressions LE generated from LA, is defined recursively as follows: If $L \in LA$ *then* $L \in LE$*; If* $\theta, \varphi \in LE$ *then* $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi \in LE$.

Definition 2 (Mass Function on Labels)
$\forall x \in \Omega$ *a mass function on labels is a function* $m_x : 2^{LA} \to [0,1]$ *such that* $\sum_{S \subseteq LA} m_x(S) = 1$.

Note that there is no requirement for the mass associated with the empty set to be zero. Instead, $m_x(\emptyset)$ quantifies the agent's belief that none of the labels are appropriate to describe x. We might observe that this phenomena occurs frequently in natural language, especially when labelling perceptions generated along some continuum. For example, we occasionally encounter colours for which none of our available colour descriptors seem appropriate. Hence, the value $m_x(\emptyset)$ is an indicator of the describability of x in terms of the labels *LA*.

The link between the mass function m_x and the appropriateness measures $\mu_\theta(x)$ is motivated by the intuition that the assertion 'x is θ' directly provides information dependent on θ, as to what are the possible values for $\mathscr{D}_x$. For example, the assertion 'x is *blue*' would mean that *blue* is an appropriate label for x, from which we can infer that $blue \in \mathscr{D}_x$. Similarly, the assertion 'x is *green and not blue*' would mean that *green* is an appropriate label for x while *blue* is not, so that we can infer $green \in \mathscr{D}_x$ and $blue \notin \mathscr{D}_x$. Another way of expressing this information is to say that $\mathscr{D}_x$ must be a member of the set of sets of labels which contain *green* but do not contain *blue* i.e. $\mathscr{D}_x \in \{S \subseteq LA : green \in S,\ blue \notin S\}$. More generally, we can define a functional mapping λ from *LE* into $2^{2^{LA}}$ (i.e. the set containing all possible sets of label sets) for which the assertion 'x is θ' enables us to infer that $\mathscr{D}_x \in \lambda(\theta)$. This mapping is defined recursively as follows:

Definition 3 (λ-mapping)
$\lambda : LE \to 2^{\mathscr{F}}$ *is defined recursively as follows:* $\forall L \in LA$, $\forall \theta,\ \varphi \in LE$; $\lambda(L) = \{S \in \mathscr{F} : L \in S\}$; $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$; $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$; $\lambda(\neg\theta) = \lambda(\theta)^c$.

The λ-mapping then provides us with a means of evaluating the appropriateness measure of an expression θ directly from m_x, as corresponding to the subjective belief that $\mathscr{D}_x \in \lambda(\theta)$ so that:

Definition 4 (Appropriateness Measures)
For any expression $\theta \in LE$ and $x \in \Omega$, the appropriateness measure $\mu_\theta(x)$ can be determined from the mass function m_x according to:

$$\forall \theta \in LE \quad \mu_\theta(x) = \sum_{S \in \lambda(\theta)} m_x(S)$$

From this relationship the following list of general properties hold for expressions θ and φ in *LE* [6]:

Theorem 1 (Lawry [6, 7])

- *If $\theta \models \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) \leq \mu_\varphi(x)$*
- *If $\theta \equiv \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) = \mu_\varphi(x)$*
- *If θ is a tautology then $\forall x \in \Omega\ \mu_\theta(x) = 1$*
- *If θ is a contradiction then $\forall x \in \Omega\ \mu_\theta(x) = 0$*
- $\forall x \in \Omega\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$

Notice, here that the laws of excluded middle, non-contradiction and idempotence are all preserved.

In practice an agent's estimation of both m_x and $\mu_\theta(x)$ should depend on their experience of language use involving examples similar to x. Clearly the form of this knowledge is likely to be both varied and complex. However, one natural type of assessment for an agent to make would be to order or rank label in terms of their estimated appropriateness for x. This order information could then be combined with estimates of appropriateness measure values for the basic labels (i.e. elements of *LA*) in order to provide estimates of values for compound expressions (i.e. elements of *LE*).

Definition 5 (Ordering on Labels)
For $x \in \Omega$ let $\preceq_x$ be an ordering on LA such that for $L, L' \in LA$, $L' \preceq_x L$ means that L is at least as appropriate as a label for x as L'.

For any labels $L_i, L_j \in LA$ if $L_i \preceq_x L_j$ it follows that if $L_j \in \mathscr{D}_x$ then $L_i \in \mathscr{D}_x$ and consequently when $\preceq_x$ is a total ordering then the mass function m_x must be nested. In that case the following theorem shows that the min and max rules for conjunction and disjunction hold for a restricted class of expressions:

Theorem 2 ([6, 11])
Let $LE^{\wedge,\vee} \subseteq LE$ denote those expressions generated recursively from LA using only the connectives $\wedge$ and $\vee$. If the appropriateness of the basic labels as descriptions for x is ranked according to a total ordering $\preceq_x$ on LA then $\forall \theta, \varphi \in LE^{\wedge,\vee}$ it holds that $\mu_{\theta\wedge\varphi}(x) = \min\left(\mu_\theta(x), \mu_\varphi(x)\right)$, $\mu_{\theta\vee\varphi}(x) = \max\left(\mu_\theta(x), \mu_\varphi(x)\right)$.

3 A Prototype Theory Interpretation of Label Semantics

Suppose that a distance metric d is defined on Ω such that $d : \Omega^2 \to [0, \infty)$ and satisfies $d(x,x) = 0$ and $d(x,y) = d(y,x)$ for all elements $x, y \in \Omega$. For each label $L_i \in LA$ let

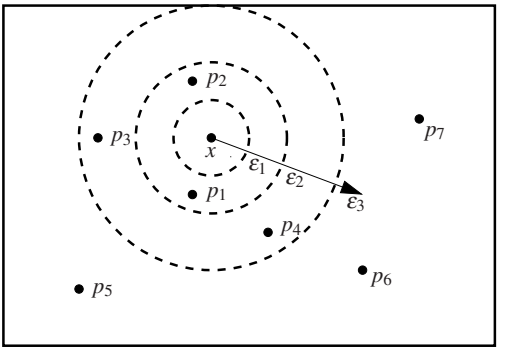


Fig. 1. Identifying $\mathscr{D}_x^\varepsilon$ as ε varies; For ε_1, ε_2 and ε_3 shown in the diagram $\mathscr{D}_x^{\varepsilon_1} = \emptyset$, $\mathscr{D}_x^{\varepsilon_2} = \{L_1, L_2\}$, $\mathscr{D}_x^{\varepsilon_3} = \{L_1, L_2, L_3, L_4\}$

there be a single element $p_i \in \Omega$[1] corresponding to a prototypical case for which L_i is certainly an appropriate description. Within this framework L_i is deemed to be appropriate to describe an element $x \in \Omega$ provided x is sufficiently close or similar to the prototypical element p_i. This is formalized by the requirement that x is within a maximal distance threshold ε of p_i. i.e. L_i is appropriate to describe x if $d(x, p_i) \leq \varepsilon$ where $\varepsilon > 0$. From this perspective an agent's uncertainty regarding the appropriateness of a label to describe a value x is characterised by his or her uncertainty regarding the distance threshold ε. Here we assume that this uncertainty is represented by a probability density function δ for ε defined on $[0, \infty)$. Within this interpretation a natural definition of the description of an element $\mathscr{D}_x$ and the associated mass function m_x can be given as follows:

Definition 6. *For* $\varepsilon \in [0, \infty)$ $\mathscr{D}_x^\varepsilon = \{L_i \in LA : d(x, p_i) \leq \varepsilon\}$ *and* $m_x(F) = \delta(\{\varepsilon : \mathscr{D}_x^\varepsilon = F\})$[2].

Intuitively speaking $\mathscr{D}_x^\varepsilon$ identifies the set of labels with prototypes lying within ε of x. Figure 1 shows $\mathscr{D}_x^\varepsilon$ in a hypothetical conceptual space as ε varies. Notice that the sequence $\mathscr{D}_x^\varepsilon$ as ε varies generates a nested hierarchy of label sets. Furthermore, the distance metric d naturally generates a total ordering on the appropriateness of labels for any element x, according to which label L_j is as least as appropriate to describe x as label L_i if x is closer (or equidistant) to p_j than to p_i i.e. $L_i \preceq_x L_j$ iff $d(x, p_i) \geq d(x, p_j)$. The following theorem shows that this ordering constrains the labels contained in $\mathscr{D}_x^\varepsilon$ as suggested in Section 2:

Theorem 3. *If* $L_i \preceq_x L_j$ *(as defined above) then* $\forall \varepsilon \geq 0$ $L_i \in \mathscr{D}_x^\varepsilon$ *implies that* $L_j \in \mathscr{D}_x^\varepsilon$.

[1] For simplicity of notation we assume that each label has a single prototype. However, the case where there is a set of prototypes P_i for L_i can easily be accommodated by extending the distance metric d such that $d(x, P_i) = \inf\{d(x, p_i) : p_i \in P_i\}$.

[2] For Lesbegue measurable set I, we denote $\delta(I) = \int_I \delta(\varepsilon) d\varepsilon$.

Proof. Suppose $\exists x \in \Omega$ for which $L_i \preceq_x L_j$ and $\exists \varepsilon \geq 0$ such that $L_i \in \mathscr{D}_x^\varepsilon$ and $L_j \notin \mathscr{D}_x^\varepsilon$. From this it follows that $d(x,p_i) \leq \varepsilon$ and $d(x,p_j) > \varepsilon$ and hence $L_i \not\preceq_x L_j$ which is a contradiction as required. □

Also notice from Definition 6, that for $L_i \in LA$ the appropriateness measure $\mu_{L_i}(x)$ is given by $\delta(\{\varepsilon : L_i \in \mathscr{D}_x^\varepsilon\})$. Consequently, if we view $\mathscr{D}_x^\varepsilon$ as a random set from $[0,\infty)$ into 2^{LA} then $\mu_{L_i}(x)$ corresponds to the single point coverage function of $\mathscr{D}_x^\varepsilon$. This provides us with a link to the random set interpretation of fuzzy sets (See [1], [2] or [4] for an exposition) except that in this case the random set maps to sets of labels rather than sets of elements. Hence, the interpretation of label semantics as proposed above provides a link between random set theory and prototype theory.

The following results show how the appropriateness of an expression $\theta \in LE$ to describe an element x is equivalent to a constraint $\varepsilon \in I(\theta,x)$, for some measurable subset $I(\theta,x)$ of $[0,\infty)$.

Definition 7. *$\forall x \in \Omega$ and $\theta \in LE$, $I(\theta,x) \subseteq [0,\infty)$ is defined recursively as follows: $\forall L_i \in LA$, $\forall \theta,\varphi \in LE$; $I(L_i,x) = [d(x,p_i),\infty)$; $I(\theta \wedge \varphi,x) = I(\theta,x) \cap I(\varphi,x)$; $I(\theta \vee \varphi,x) = I(\theta,x) \cup I(\varphi,x)$; $I(\neg\theta,x) = I(\theta,x)^c$.*

Theorem 4. $\forall \theta \in LE, \forall x \in \Omega\ I(\theta,x) = \{\varepsilon : \mathscr{D}_x^\varepsilon \in \lambda(\theta)\}$.

Corollary 1. $\forall \theta \in LE,\ \forall x \in \Omega\ \mu_\theta(x) = \delta(I(\theta,x))$.

Definition 8. *We define $k : LE^{\wedge,\vee} \times \Omega \to [0,\infty)$ recursively as follows: $\forall x \in \Omega$, $\forall L_i \in LA$, $\forall \theta,\varphi \in LE^{\wedge,\vee}$; $k(L_i,x) = d(x,p_i)$; $k(\theta \wedge \varphi,x) = \max(k(\theta,x),k(\varphi,x))$ and $k(\theta \vee \varphi,x) = \min(k(\theta,x),k(\varphi,x))$.*

Theorem 5. *$\forall x \in \Omega$, $\forall x \in LE^{\wedge,\vee}$, then $I(\theta,x) = [k(\theta,x),\infty)$.*

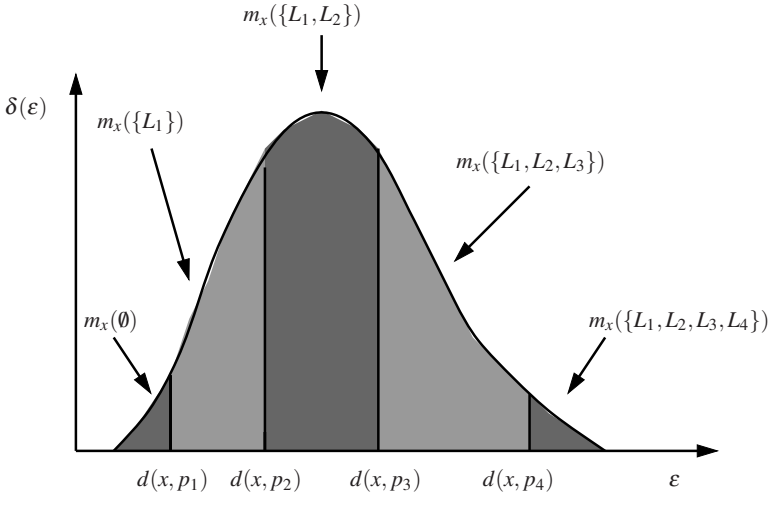


Fig. 2. Let $LA = \{L_1,L_2,L_3,L_4\}$ and $L_4 \preceq_x L_3 \preceq_x L_2 \preceq_x L_1$. This figure shows the values of m_x as areas under δ.

From Theorem 5 we have that

$$\mu_{\theta\vee\varphi}(x) = \delta([k(\theta\vee\varphi,x),\infty)) = \delta([\min(k(\theta,x),k(\varphi,x)),\infty))$$
$$= \max(\delta([k(\theta,x),\infty)),\delta([k(\varphi,x),\infty))) = \max(\mu_\theta(x),\mu_\varphi(x)).$$

Similarly, $\mu_{\theta\wedge\varphi}(x) = \min(\mu_\theta(x),\mu_\varphi(x))$ as is consistent with Theorem 2.

Example 1. $I(L_i,x) = [d(x,p_i),\infty)$, $I(\neg L_i,x) = [0,d(x,p_i))$, $I(L_i \wedge L_j,x) = [\max(d(x,p_i),d(x,p_j)),\infty)$, $I(L_i \vee L_j,x) = [\min(d(x,p_i),d(x,p_j)),\infty)$. Also $I(L_i\wedge\neg L_j,x) = [d(x,p_i),d(x,p_j))$ provided $d(x,p_i) < d(x,p_j)$ and $=\emptyset$ otherwise.

From Lawry [7] we have that for $F \subseteq LA$ $m_x(F) = \mu_{\alpha_F}(x)$ where $\alpha_F = (\bigwedge_{L\in F} L)\wedge(\bigwedge_{L\notin F}\neg L)$. Hence, $m_x(F) = \delta(I(\alpha_F,x))$ where $I(\alpha_F,x) = [\max\{d(x,p_i) : L_i \in F\},\min\{d(x,p_i) : L_i \notin F\})$ provided that $\max\{d(x,p_i) : L_i \in F\} < \min\{d(x,p_i) : L_i \notin F\}$ and $=\emptyset$ otherwise.

Figure 2 shows the areas under δ corresponding to the values of the mass function m_x.

4 Conclusions

Label semantics is an epistemic theory of uncertainty for vague concepts based on appropriateness measures and mass functions. The underlying calculus is not truth-functional but can be functional in a weaker sense, with the min and max rules for conjunction and disjunction being preserved for a restricted class of expressions.

Appropriateness measures and mass functions can be interpreted, within prototype theory, as the probability that a distance threshold ε lies in a measurable subset of $[0,\infty)$ as determined by the relevant label or expression. Here ε represents an upper-bound on the distance that an element x can be from the prototype p_i for a label L_i, in order that L_i is still deemed an appropriate description of x.

Acknowledgement. Yongchuan Tang is funded by the National Natural Science Foundation of China (NSFC) under Grant 60604034, the joint funding of NSFC and MSRA under Grant 60776798, and the Science and Technology Program of Zhejiang Province under Grant 2007C23061.

References

1. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets Syst. 90(2), 141–150 (1997)
2. Dubois, D., Godo, L., Prade, H., Esteva, F.: An information-based discussion of vagueness. In: Cohen, H., Lefebvre, C. (eds.) Handbook of Categorization in Cognitive Science, pp. 892–913 (2005)
3. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge (2000)
4. Goodman, I.R., Nguyen, H.T.: Uncertainty Models for Knowledge-Based Systems; A Unified Approach to the Measurement of Uncertainty. Elsevier Science Inc., New York (1985)
5. Kripke, S.: A completeness theorem for modal logic. J. Symbolic. Logic 24, 1–24 (1959)
6. Lawry, J.: A framework for linguistic modelling. Artificial Intelligence 155, 1–39 (2004)

7. Lawry, J.: Modelling and Reasoning with Vague Concepts, vol. 12. Springer, New York (2006)
8. Lawry, J.: Appropriateness measures: An uncertainty model for vague concepts. Synthese 161(2), 255–269 (2008)
9. Rosch, E.: Cognitive representation of semantic categories. J. Exp. Psychol. Gen. 104, 192–233 (1975)
10. Ruspini, E.H.: On the semantics of fuzzy logic. Internat. J. Approx. Reason 5, 45–88 (1991)
11. Tang, Y., Zheng, J.: Linguistic modelling based on semantic similarity relation among linguistic labels. Fuzzy Sets Syst. 157, 1662–1673 (2006)
12. Williamson, T.: Vagueness. Routledge, London (1994)

Fuzzy Probabilities Based on the Likelihood Function

Marco E.G.V. Cattaneo

Institut für Statistik, Ludwig-Maximilians-Universität München, München, Germany

Abstract. If we interpret the statistical likelihood function as a measure of the relative plausibility of the probabilistic models considered, then we obtain a hierarchical description of uncertain knowledge, offering a unified approach to the combination of probabilistic and possibilistic uncertainty. The fundamental advantage of the resulting fuzzy probabilities with respect to imprecise probabilities is the ability of using all the information provided by the data.

Keywords: Likelihood function, Hierarchical model, Fuzzy probabilities, Imprecise probabilities, Statistical inconsistency.

1 Introduction

This paper presents a probabilistic-possibilistic hierarchical model based on the likelihood function. Thanks to the intuitivity and asymptotic properties of the likelihood function, the hierarchical model is an ideal basis for inference and decision making: this aspect is analyzed in [2]. The hierarchical model can be interpreted as a fuzzy probability measure, and offers a unified approach to the combination of probabilistic and possibilistic uncertainty.

Fuzzy probabilities generalize imprecise probabilities by additionally considering the relative plausibility of different values in the probability intervals (imprecise probabilities correspond to the crisp case of fuzzy probabilities). By abandoning the crispness of imprecise probabilities, the hierarchical model solves a fundamental problem of the imprecise probability approach: its statistical inconsistency.

2 Hierarchical Model

Let $\mathscr{P}$ be a set of probability measures on a measurable space $(\Omega, \mathscr{A})$ such that $\mathscr{A}$ contains all singletons of Ω. Each $P \in \mathscr{P}$ is interpreted as a probabilistic model of the reality under consideration. The interpretation of probability is not important: for instance the elements of $\mathscr{P}$ can be statistical models, or describe the forecasts of a group of experts.

When an event $A \in \mathscr{A}$ is observed, the *likelihood function*

$$lik : P \longmapsto P(A)$$

on $\mathscr{P}$ describes the relative ability of the probabilistic models in $\mathscr{P}$ to forecast the observed data. Spurious modifications of the situation considered or of its mathematical

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 43–50, 2008.
springerlink.com

representation can lead to likelihood functions proportional to lik. Therefore, proportional likelihood functions are considered equivalent; in fact, Fisher [8] defined the likelihood of a statistical model as a quantity *proportional* to the probability of the observed data. Hence, only ratios $lik(P)/lik(P')$ of the values of lik for different $P, P' \in \mathscr{P}$ have meaning: Kullback and Leibler [11] interpreted $\log[lik(P)/lik(P')]$ as the information in A for discrimination in favor of P against P'. When the realization of a continuous random object is observed, the usual definition of likelihood function in terms of density can be seen as an approximation of lik (see [2, Section 1.2]).

The likelihood function can thus be interpreted as a measure of the relative plausibility of the probabilistic models in the light of the observed data alone. Under each probabilistic model $P \in \mathscr{P}$, the likelihood ratio $lik(P)/lik(P')$ of P against a different probabilistic model $P' \in \mathscr{P}$ almost surely increases without bound when more and more data are observed, and consequently lik tends to concentrate around P, if some regularity conditions are satisfied. Thanks to this asymptotic property and to its intuitivity, the likelihood function is an ideal basis for statistical inference and decision making (see [13] for an introduction to the likelihood approach to statistics).

Example 1. Let $\mathscr{P} = \{P_p : p \in [0.1, 0.6]\}$ be a set of probability measures on a measurable space $(\Omega, \mathscr{A})$, such that for each $P_p \in \mathscr{P}$ the random variables $X_0, \ldots, X_{100} : \Omega \to \{0,1\}$ satisfy the following conditions: $P_p\{X_0 = 0\} = \frac{1}{2}$, and conditional on the realization of X_0 the random variables $X_1, \ldots, X_{100}$ are independent with $P_p\{X_i = 1 \,|\, X_0 = 0\} = \frac{1}{2}$ and $P_p\{X_i = 1 \,|\, X_0 = 1\} = p$ for all $i \in \{1, \ldots, 100\}$.

The realizations of $X_1, \ldots, X_{100}$ are observed: 20 of them take the value 1. The resulting likelihood function

$$lik : P_p \longmapsto \tfrac{1}{2}\left(\tfrac{1}{2}\right)^{100} + \tfrac{1}{2}\, p^{20}(1-p)^{80}$$

on $\mathscr{P}$ is concentrated around $P_{0.2}$, which is the most plausible element of $\mathscr{P}$ in the light of the observed data alone. The case with $X_0 = 0$ has almost no influence on the form of the likelihood function, and in fact this case is extremely implausible in the light of the observed data and of the probabilistic models considered.

The likelihood function lik measures the relative plausibility of the elements of $\mathscr{P}$, but a measure of the relative plausibility of the subsets of $\mathscr{P}$ is often needed. A simple and effective way to obtain it consists in defining the plausibility of a set of probabilistic models as the plausibility of its best element: the result is the set function

$$\mathscr{H} \longmapsto \sup_{P \in \mathscr{H}} lik(P)$$

on the power set $2^{\mathscr{P}}$ of $\mathscr{P}$ (in this paper, $\sup \varnothing = 0$). Proportional set functions of this form are equivalent, since they correspond to equivalent likelihood functions: to underline this relative meaning, the expression "relative plausibility measure" is used in [2] to denote an equivalence class of proportional set functions of this form. Their normalized version LR associates to each $\mathscr{H} \subseteq \mathscr{P}$ the corresponding likelihood ratio statistic

$$LR(\mathscr{H}) = \frac{\sup_{P \in \mathscr{H}} lik(P)}{\sup_{P \in \mathscr{P}} lik(P)}.$$

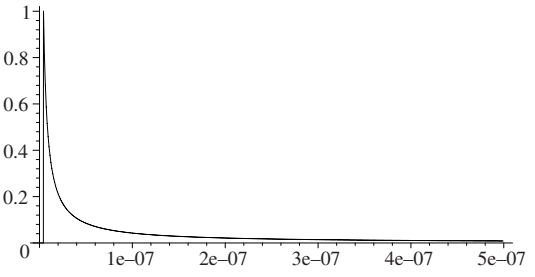


Fig. 1. Profile likelihood function from Example 2 and membership function of fuzzy probability from Example 3

The *likelihood ratio test* discards the hypothesis that the data were generated by some $P \in \mathscr{H}$ if $LR(\mathscr{H})$ is sufficiently small.

Let $g : \mathscr{P} \to \mathscr{G}$ be a function. The likelihood function lik on $\mathscr{P}$ induces the (normalized) *profile likelihood function*

$$lik_g : \gamma \longmapsto LR(g^{-1}\{\gamma\}) \propto \sup_{P \in \mathscr{P} : g(P) = \gamma} lik(P)$$

on $\mathscr{G}$ (in this paper, g^{-1} denotes the set function associating to each subset of $\mathscr{G}$ its inverse image under g). The profile likelihood function lik_g measures the relative plausibility of the values of g, on the basis of the above definition of plausibility for a set of probabilistic models. The *maximum likelihood estimate* $\hat{\gamma}_{ML}$ of $g(P)$ is the $\gamma \in \mathscr{G}$ maximizing $lik_g(\gamma)$ (that is, $lik_g(\hat{\gamma}_{ML}) = 1$), when such a γ exists and is unique. The *likelihood-based confidence region* for $g(P)$ with cutoff point $\alpha \in (0,1)$ is the set $\{\gamma \in \mathscr{G} : lik_g(\gamma) > \alpha\}$: it is the smallest $G \subseteq \mathscr{G}$ such that $LR\{P \in \mathscr{P} : g(P) \notin G\} \leq \alpha$.

Example 2. Consider the situation of Example 1, and let $g : \mathscr{P} \to [0,1]$ associate to each probabilistic model in $\mathscr{P}$ the probability of $X_0 = 0$ conditional on the observed realizations of $X_1, \ldots, X_{100}$:

$$g : P_p \longmapsto \frac{\left(\frac{1}{2}\right)^{100}}{\left(\frac{1}{2}\right)^{100} + p^{20}(1-p)^{80}}.$$

Figure 1 shows the graph of the profile likelihood function lik_g on $[0, 5 \cdot 10^{-7}]$: as expected, lik_g is extremely concentrated near 0, because $X_0 = 1$ is compatible with the observed data, while $X_0 = 0$ is not. In fact, the maximum likelihood estimate of $g(P_p)$ is $\hat{\gamma}_{ML} \approx 0.04 \cdot 10^{-7}$, and the likelihood-based confidence region for $g(P_p)$ with cutoff point $\alpha = 0.01$ corresponds approximately to the interval $(0.04 \cdot 10^{-7}, 4.26 \cdot 10^{-7})$.

The probabilistic models in $\mathscr{P}$ and the likelihood function lik on $\mathscr{P}$ can be interpreted as the two levels of a *hierarchical model* of the reality under consideration. The two levels describe different kinds of uncertain knowledge: in the first level the uncertainty is stochastic, while in the second one it is about which of the probabilistic models in $\mathscr{P}$ is the best representation of the reality. It is important to underline that no probabilistic

model in $\mathscr{P}$ is assumed to be in some sense "true": the elements of $\mathscr{P}$ are simply interpreted as more or less plausible representations of the reality (this interpretation of the hierarchical model is shared by Edwards [7]). By contrast, the use of a probability measure on $\mathscr{P}$, suggested by the Bayesian approach, carries the implicit assumption that exactly one of the probabilistic models in $\mathscr{P}$ is "true" (see [2, Section 3.1]).

The definition of likelihood function implies that when an event $A \in \mathscr{A}$ is observed, the two levels $\mathscr{P}$ and lik of the hierarchical model are updated to

$$\mathscr{P}' = \{P(\cdot|A) : P \in \mathscr{P}, P(A) > 0\} \tag{1}$$

$$\text{and to} \quad lik' : P' \longmapsto \sup_{P \in \mathscr{P} : P(\cdot|A) = P'} lik(P)\, P(A),$$

respectively. When A is the first observed event, the *prior* likelihood function lik can be interpreted as a (subjective) measure of the relative plausibility of the probabilistic models in $\mathscr{P}$ according to the prior information. The choice of a prior likelihood function on $\mathscr{P}$ seems to be better supported by intuition than the choice of a prior probability measure on $\mathscr{P}$: in particular, a constant likelihood function describes *complete ignorance* (in the sense of absence of information for discrimination between the probabilistic models). In fact, if lik is constant, then lik' is proportional to the profile likelihood function on $\mathscr{P}'$ induced by the observation A and the conditioning $P \mapsto P(\cdot|A)$. Moreover, the choice of a prior likelihood function can be based on analogies with the likelihood functions induced by hypothetical data (see also [3]).

3 Fuzzy Probabilities

A *possibility distribution* on a set $\mathscr{G}$ is a function $\pi : \mathscr{G} \to [0,1]$. The possibility measure on $\mathscr{G}$ with possibility distribution π is the set function

$$G \longmapsto \sup_{\gamma \in G} \pi(\gamma)$$

on $2^{\mathscr{G}}$. A possibility distribution π on $\mathscr{G}$ can also be considered as the *membership function* of a fuzzy subset of $\mathscr{G}$ (see [17]); when π is *crisp* (that is, π can take only the values 0 and 1), the subset is not fuzzy and π is its indicator function on $\mathscr{G}$. The likelihood ratio statistic LR is a possibility measure on $\mathscr{P}$ with possibility distribution proportional to the likelihood function lik on $\mathscr{P}$. In fact, the membership function of a fuzzy set has often been interpreted as a likelihood function (see for example [10, 5]), even though proportional membership functions were not always considered equivalent (see for instance [6]). In the present paper, membership functions and possibility distributions are interpreted as *proportional* to likelihood functions. Hence, it suffices to consider normalized fuzzy sets and normalized possibility measures (that is, $\sup_{\gamma \in \mathscr{G}} \pi(\gamma) = 1$ is assumed), but grades of membership and degrees of possibility have only a relative meaning.

The hierarchical model considered in the previous section can thus be interpreted as consisting of a probabilistic level (described by $\mathscr{P}$) and a possibilistic level (described by LR). That is, it can be interpreted as a probabilistic-possibilistic hierarchical

description of uncertain knowledge about $\omega \in \Omega$. Both the purely probabilistic and the purely possibilistic descriptions of uncertain knowledge about $\omega \in \Omega$ appear as special cases. In fact, when $\mathscr{P}$ is a singleton, the uncertainty about $\omega \in \Omega$ is purely probabilistic (*LR* on $\mathscr{P} = \{P\}$ contains no information, since its meaning is only relative). By contrast, when $\mathscr{P}$ consists of all the Dirac measures (that is, $\mathscr{P} = \{\delta_\omega : \omega \in \Omega\}$ with $\delta_\omega\{\omega\} = 1$), the uncertainty about $\omega \in \Omega$ is purely possibilistic (*LR* can be considered as a possibility measure on Ω, since each $\delta_\omega \in \mathscr{P}$ can be identified with the corresponding $\omega \in \Omega$).

The hierarchical model can also be interpreted as a fuzzy probability measure on $(\Omega, \mathscr{A})$, in the sense that it is a fuzzy subset of the set of all probability measures on $(\Omega, \mathscr{A})$, with membership function proportional to *lik* on $\mathscr{P}$ and constant equal to 0 outside $\mathscr{P}$. More generally, the uncertain knowledge about the value $g(P)$ of a function $g : \mathscr{P} \to \mathscr{G}$ is described by the induced possibility measure $LR \circ g^{-1}$ on $\mathscr{G}$; that is, by the fuzzy subset of $\mathscr{G}$ with membership function lik_g. In particular, when $g : \mathscr{P} \to \mathbb{R}$, the uncertain knowledge about $g(P)$ is described by a fuzzy number (that is, a fuzzy subset of $\mathbb{R}$). For example, g can associate to each probabilistic model P the expectation $g(P) = E_P(X)$ of a random variable X, or the probability $g(P) = P(A)$ of an event $A \in \mathscr{A}$: the membership function lik_g describes then the *fuzzy expectation* of X, or the *fuzzy probability* of A, respectively. Sometimes a fuzzy number can be a satisfactory conclusion about the value of $g(P)$, but it is often necessary to evaluate the fuzzy number by a single real number (such as the maximum likelihood estimate $\hat{\gamma}_{ML}$) or by a couple of real numbers (such as the infimum and the supremum of a likelihood-based confidence region $\{\gamma \in \mathbb{R} : lik_g(\gamma) > \alpha\}$). The discussion on how to evaluate a fuzzy number by one or more real numbers goes beyond the scope of the present paper, but see [2, Section 4.1] for some interesting results (to each evaluation method corresponds a likelihood-based decision criterion).

Example 3. The prior fuzzy probability measure on $(\Omega, \mathscr{A})$ considered in Examples 1 and 2 is crisp, in the sense that its membership function on the set of all probability measures on $(\Omega, \mathscr{A})$ is crisp. In fact, the only prior (non-stochastic) uncertainty is about the value of the probability of $X_i = 1$ conditional on $X_0 = 1$ (with $i \in \{1, \ldots, 100\}$), and the only prior information about this value is that it lies in the interval $[0.1, 0.6]$. But the updated fuzzy probability measure on $(\Omega, \mathscr{A})$ obtained after having observed the realizations of $X_1, \ldots, X_{100}$ is not crisp anymore: the fuzzy (conditional) probability of $X_0 = 0$ has membership function lik_g (plotted in Figure 1). Hence, any reasonable evaluation of the fuzzy (conditional) probability of $X_0 = 0$ by a real number (such as the maximum likelihood estimate $\hat{\gamma}_{ML} \approx 0.04 \cdot 10^{-7}$, or the lower and upper evaluations $0.04 \cdot 10^{-7}$ and $4.26 \cdot 10^{-7}$ considered at the end of Example 2) would be approximately 0.

The hierarchical model offers a unified approach to the combination of probabilistic and possibilistic uncertainty (in particular, fuzzy data would pose no problem). Since membership functions and possibility distributions are interpreted as proportional to likelihood functions, the rules for manipulating fuzzy probabilities are implied by the well-established theories of probability and likelihood (the same holds for the approach of De Cooman [4], which uses a different interpretation of possibility measures). By contrast, approaches based on the arithmetic of fuzzy numbers (see for example [14, 1])

face the problem of choosing and justifying such rules: the choice of a consistent way of updating the fuzzy probability models in the light of data seems to be particularly difficult.

4 Imprecise Probabilities

The mathematical representations of reality used in the classical and Bayesian approaches to statistics can be considered as special cases of the hierarchical model (see [2, Section 3.2]). By contrast, the imprecise probability model cannot be considered as a special case of the hierarchical model, because the updating rules are different. The mathematical representation of reality used in the imprecise probability approach to statistics can be described as a (convex) set $\mathscr{P}$ of probabilistic models, without information for discrimination between them. This corresponds to a hierarchical model with constant likelihood function on $\mathscr{P}$, but the imprecise probability model is usually updated by *regular extension* (see [15, Appendix J]): that is, by conditioning each $P \in \mathscr{P}$ on the observed data, without considering the information provided by the likelihood function on $\mathscr{P}$. More precisely, when an event $A \in \mathscr{A}$ is observed, the set $\mathscr{P}$ is updated to the set $\mathscr{P}'$ as in (1), but the constant likelihood function on $\mathscr{P}$ is not updated: the likelihood function on $\mathscr{P}'$ is still constant; that is, the information in A for discrimination between the elements of $\mathscr{P}$ is disregarded.

For instance, if the probabilistic models in $\mathscr{P}$ describe the opinions of a group of Bayesian experts, then the updating by regular extension corresponds to update the opinion of each expert without reconsidering her/his credibility, independently of how bad her/his forecasts were when compared to the forecasts of the other experts. This is not very reasonable, and in fact the updating by regular extension can lead to *inconsistency*, in the statistical sense of not tending to the correct conclusion, even when the amount of information provided by the data tends to infinity.

Example 4. The set $\mathscr{P}$ of probabilistic models considered in Examples 1, 2, and 3 can be interpreted as an imprecise probability measure on $(\Omega, \mathscr{A})$. If it is updated by regular extension, when the realizations of $X_1, \ldots, X_{100}$ are observed, then the resulting imprecise probability measure is described by the set $\mathscr{P}'$. In particular, the resulting uncertain knowledge about the (conditional) probability of $X_0 = 0$ is described by the lower and upper probabilities

$$\inf_{P' \in \mathscr{P}'} P'\{X_0 = 0\} \approx 4.26 \cdot 10^{-9} \quad \text{and} \quad \sup_{P' \in \mathscr{P}'} P'\{X_0 = 0\} \approx 1 - 6.77 \cdot 10^{-7}.$$

That is, despite the overwhelming information in favor of $X_0 = 1$ against $X_0 = 0$, almost complete ignorance about the (conditional) probabilities of $X_0 = 0$ and $X_0 = 1$ is obtained when the imprecise probability model is updated by regular extension (it is important to note that these results do not change when $\mathscr{P}$ is replaced by its convex hull). In fact, the resulting interval probability of $X_0 = 0$ is the support $\{\gamma \in [0,1] : lik_g(\gamma) > 0\}$ of the membership function lik_g of the fuzzy (conditional) probability of $X_0 = 0$ (plotted in Figure 1): lik_g is extremely concentrated near 0, but this information is disregarded when updating the imprecise probability model by regular extension (the present example was proposed by Wilson [16]).

The imprecise probability model can be seen as the crisp (and convex) case of the fuzzy probability model, but in general the crispness of the fuzzy probability model is lost when it is updated. Hence, from the point of view of the hierarchical model, the regular extension forces the crispness of the updated model by disregarding a part of the information provided by the data, and this can lead to statistical inconsistency. Many authors (see for example [16, 12]) have replaced, in particular problems, the regular extension with alternative updating rules making use of some information contained in the likelihood function on $\mathscr{P}$. But no alternative rule updating $\mathscr{P}$ to a subset of $\mathscr{P}'$ can assure the statistical consistency, because any discarded probabilistic model can become the most plausible one in the light of new data.

5 Conclusion

Statistical inconsistency is a fundamental problem of the theory of imprecise probabilities: a simple solution is to generalize imprecise probabilities to fuzzy probabilities, and use the probabilistic-possibilistic hierarchical model presented in this paper. In fact, fuzzy probabilities seem to be very intuitive: many authors (see for example [9, 4]) have studied models similar to the hierarchical one to accommodate the fact that usually not all the values in probability intervals are considered equally plausible.

References

1. Buckley, J.J.: Fuzzy Probability and Statistics. Studies in Fuzziness and Soft Computing, vol. 196. Springer, New York (2006)
2. Cattaneo, M.: Statistical Decisions Based Directly on the Likelihood Function. PhD Thesis, ETH Zurich (2007), http://e-collection.ethz.ch
3. Dahl, F.A.: Representing human uncertainty by subjective likelihood estimates. Internat. J. Approx. Reason 39(1), 85–95 (2005)
4. De Cooman, G.: A behavioural model for vague probability assessments. Fuzzy Sets Syst. 154, 305–358 (2005)
5. Dubois, D.: Possibility theory and statistical reasoning. Comput. Stat. Data Anal. 51, 47–69 (2006)
6. Dubois, D., Moral, S., Prade, H.: A semantics for possibility theory based on likelihoods. J. Math. Anal. Appl. 205(2), 359–380 (1997)
7. Edwards, A.W.F.: Likelihood, 2nd edn. Johns Hopkins University Press, Baltimore (1992)
8. Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond., Ser. A 222, 309–368 (1922)
9. Gärdenfors, P., Sahlin, N.E.: Unreliable probabilities, risk taking, and decision making. Synthese 53, 361–386 (1982)
10. Hisdal, E.: Are grades of membership probabilities? Fuzzy Sets Syst. 25, 325–348 (1988)
11. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. 22, 79–86 (1951)
12. Moral, S.: Calculating uncertainty intervals from conditional convex sets of probabilities. In: Dubois, D., Wellman, M.P. (eds.) Proceedings of Uncertainty in Artificial Intelligence (UAI 1992, Stanford, CA, USA), San Mateo, CA, USA, pp. 199–206. Morgan Kaufmann, San Francisco (1992)

13. Pawitan, Y.: In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press, New York (2001)
14. Viertl, R., Hareter, D.: Beschreibung und Analyse unscharfer Information. Springer, Wien (2006)
15. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Monographs on Statistics and Applied Probability, vol. 42. Chapman and Hall, Ltd., London (1991)
16. Wilson, N.: Modified upper and lower probabilities based on imprecise likelihoods. In: De Cooman, G., Fine, T.L., Seidenfeld, T. (eds.) Proceedings of the 2nd International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2001, Ithaca, USA), pp. 370–378. Shaker Publishing, Maastricht (2001)
17. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1(1), 3–28 (1978)

Possibility Measures in Probabilistic Inference

Giulianella Coletti[1], Romano Scozzafava[2], and Barbara Vantaggi[2]

[1] Dip. Matematica e Informatica, Università di Perugia, Perugia, Italy
[2] Dip. Metodi e Modelli Matematici, Università "La Sapienza", Roma, Italy

Abstract. By means of a logical condition between two partitions $\mathscr{L}$ and $\mathscr{L}'$ ("weak logical independence"), we find connections between probabilities and possibilities. We show that the upper envelope of the extensions of a probability on $\mathscr{L}$ is a possibility on the algebra generated by $\mathscr{L}'$. Moreover we characterize the set of possibilities obtained as extensions of a coherent probability on an arbitrary set: in particular, we find the two "extreme" (i.e., dominated and dominating) possibilities.

Keywords: Probabilistic inference, Weakly logical independence, Uncertainty measures, Coherence.

1 Introduction

The classic approaches to knowledge acquisition or decision processes start from a knowledge–base able to settle once for all the set of objects ruling the inferential process (states of nature, events, "rules", functions measuring uncertainty, etc.), requiring also further conditions (such as closure of the family of events with respect to Boolean operations). In particular, for semantic reasons a framework of reference (probability theory, Dempster-Shafer theory, possibility theory, default logic, fuzzy set theory and so on) is usually chosen once for all. Actually, often we need to manage uncertainty relative to a set of events while having information only for a *different* family of events. In fact, making inference essentially means extending a structured information (carried, for example, by a particular measure of uncertainty) to "new" events, and this is done by taking into account only the logical relations among the events of the two given families.

In general, in the extension processes, the enlargements can lead to uncertainty measures different from the initial ones. For instance, in [2] it has been proved that, if we start from a (coherent) assessment P on a set $\mathscr{L}$ of pairwise incompatible events, and consider any algebra of events $\mathscr{A}$, then the lower [upper] envelope of the class of coherent probabilities extending P to $\mathscr{L} \cup \mathscr{A}$ is a belief [plausibility] function. Vice versa, for any belief function *Bel* on an algebra $\mathscr{A}$, there exists a partition $\mathscr{L}$ and a relevant probability P such that the lower bound of the class of probability extending P on $\mathscr{A}$ coincides with *Bel* (similarly for a plausibility function, and referring to the upper bound). This result is independent of any logical relation between the partition $\mathscr{L}$ and that $\mathscr{L}'$

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 51–58, 2008.
springerlink.com

of atoms of $\mathscr{A}$. Obviously, any logical constraint between the two partitions rules the numerical values of the belief (or plausibility) function.

In [3] we proved that under suitable logical conditions between the partitions, the upper envelope (i.e., plausibility) is a possibility and the lower envelope is a necessity. Moreover, any possibility measure on an algebra $\mathscr{A}$, can be obtained as an enlargement of a probability distribution on a partition satisfying the same logical condition. This logical condition between the partitions is a suitable weakening of logical independence (see Sect. 3). A particular case is that corresponding to the logical independence of the two aforementioned partitions, in which we get a plausibility equal to 1 on $\mathscr{A} \setminus \emptyset$ for any P (which is also a noninformative possibility). These results are based on the assumption that the initial information consists of a probability distribution on the elements of a *partition* of Ω. But this is not realistic in real problems, so we study what happens starting from a (coherent) probability on an *arbitrary set of events* $\mathscr{E}$ and enlarging this assessment to an other finite set $\mathscr{E}'$: we need to handle a class of probability $\mathbf{P}_0$ (all those consistent with the coherent assessment) on the partition $\mathscr{C}$ constituted by the set of atoms generated by $\mathscr{E}$. Clearly, for every distribution on $\mathbf{P}_0$, we obtain (as lower and upper envelope of the relevant extension on $\mathscr{E}'$) a coherent belief function and a plausibility respectively, and when $\mathscr{C}$ and $\mathscr{C}'$ are weakly logically independent ($\mathscr{C}'$ is the set of atoms generated by $\mathscr{E}'$) we obtain a coherent necessity and a possibility on $\mathscr{E}'$, respectively. Obviously, it is interesting to characterize the class of these measures and in particular to study whether there is a minimum and a maximum element: in general this characterization is not possible, since the upper [lower] envelope of plausibilities [belief function] is not a plausibility [belief]. On the contrary, we prove that a characterization is possible when $\mathscr{C}$ and $\mathscr{C}'$ are weakly logically independent, obtaining a class of possibilities such that both its upper and lower envelopes Π^* and Π_* are (respectively, the dominating and dominated) possibilities. This class contains all the possibilities weakly comonotone with Π_* and Π^* (equivalent results hold for necessities).

These results contribute to the deepening of hybrid models involving probability, plausibility and possibility, which have been studied in many papers, e.g. [7, 8, 9, 10, 11, 13]: our approach is essentially syntactic and emphasizes an inferential point of view.

2 Coherent Assessments and Their Enlargements

The axioms defining an uncertainty measure strictly refer to the assumption that its domain is a Boolean algebra. Then dealing with an *arbitrary set of events* requires to characterize assessments which are coherent (or consistent) with a specific measure on a Boolean algebra containing this set.

In probability theory it is well known the concept of coherence introduced by de Finetti [6] through a betting scheme, or its dual version based on the solvability of a linear system. An analogous notion of coherence for possibilities has been introduced in [4].

Definition 1. *Let $\mathscr{E} = \{E_1, ..., E_n\}$ be a finite set of events and denote by $\mathscr{A}$ the algebra generated by $\mathscr{E}$. An assessment φ on $\mathscr{E}$ is a coherent possibility [probability] if there exists a possibility [probability] Φ defined on $\mathscr{A}$ extending φ (i.e. $\Phi_{|\mathscr{E}} = \varphi$).*

The so–called *fundamental theorem of probability* assures that, given a *coherent* assessment P on an *arbitrary* finite family $\mathscr{E}$, it can be extended (possibly not in a unique way) to any set $\mathscr{E}' \supset \mathscr{E}$; moreover, for each event $E \in \mathscr{E}' \setminus \mathscr{E}$ there exist two events E_* and E^* (possibly $E_* = \emptyset$ and $E^* = \Omega$) that are, respectively, the "maximum" and the "minimum" union of atoms A_r (generated by the initial family $\mathscr{E}$) such that $E_* \subseteq E \subseteq E^*$. If E is logical dependent on $\mathscr{E}$, then $E_* = E = E^*$. Then, given the set $\{\tilde{P}\}$ of all possible extensions of P, coherent assessments of $\tilde{P}(E)$ are all real numbers of a closed interval $[p_*, p^*]$, with

$$p_* = \inf \tilde{P}(E_*) = \inf \sum_{A_r \subseteq E_*} \tilde{P}(A_r)\,, \qquad p^* = \sup \tilde{P}(E^*) = \sup \sum_{A_r \subseteq E^*} \tilde{P}(A_r)\,. \tag{1}$$

We proved in [4] for a possibility Π a similar result: coherence of a possibility assessment assures its extendibility to new events, and for any new event the coherent possibility values belong to an interval $[\pi_*, \pi^*]$ with

$$\pi_* = \min \left(\max_{A_r \subseteq E_*} \Pi'(A_r) \right)\,, \qquad \pi^* = \max \left(\max_{A_r \subseteq E^*} \Pi'(A_r) \right)\,,$$

where $\{\Pi'\}$ is the set of all possible extensions of Π.

It is well known that by computing for some "new" events the relevant coherence probability [possibility] intervals, not all the choices of values in these intervals lead to "an overall" coherent probability [possibility]. In the probabilistic framework, if we choose for any event the minimum [the maximum] value (which correspond essentially to natural extension, see [14]), we obtain a lower [upper] probability. Furthermore, in the possibilistic setting we get different results: in fact, the upper envelope of possibilities is still a possibility [4], while the lower envelope of possibilities is not necessarily a possibility.

3 Weakly Logically Independent Partitions

We recall that two partitions $\mathscr{L}, \mathscr{L}'$ of Ω are logically independent if for every $E_i \in \mathscr{L}$ and $E'_j \in \mathscr{L}'$ one has $E_i \wedge E'_j \neq \emptyset$ (or, equivalently, $\Omega = \bigvee_{E_i \wedge E'_j \neq \emptyset} E_i$ for any $E'_j \in \mathscr{L}'$). In [3] we introduced the following "weaker" condition: for any $E'_j \in \mathscr{L}'$, denote by A_j the minimal (with respect to the inclusion) event logically dependent on $\mathscr{L}$ containing E'_j, that is

$$A_j = \bigvee_{E_i \wedge E'_j \neq \emptyset} E_i\,.$$

(Obviously, A_j is an element of the algebra $\mathscr{A}$ spanned by $\mathscr{L}$). Given $\mathscr{L}, \mathscr{L}'$, for any $E'_j \in \mathscr{L}'$ we consider the corresponding $A_j \in \mathscr{A}$.

Definition 2. *The partition $\mathscr{L}'$ is weakly logically independent of the partition $\mathscr{L}$ (in symbols, $\mathscr{L}' \perp_w \mathscr{L}$) if, for any given $E'_i \in \mathscr{L}'$, every other $E'_k \in \mathscr{L}'$ $(k \neq i)$ satisfies al leat one of the following conditions*

- $E'_k \subseteq A_i$
- $E'_k \wedge E_j \neq \emptyset$ *for any* $E_j \subseteq A_i$.

Clearly, if $\mathscr{L}, \mathscr{L}'$ are logically independent, then $\mathscr{L}' \perp_w \mathscr{L}$, but the vice versa does not hold: let $\mathscr{L} = \{E, E^c\}$, $\mathscr{L}' = \{F, F^c\}$ with $F \subset E$, then $\mathscr{L}' \perp_w \mathscr{L}$, but $\mathscr{L}'$ and $\mathscr{L}$ are not logically independent. As proved in [3] the notion of weakly logically independent partitions is symmetric (i.e. $\mathscr{L}' \perp_w \mathscr{L} \Longrightarrow \mathscr{L} \perp_w \mathscr{L}'$).

We recall now some properties of weakly logically independent partitions.

Proposition 1. *Let* $\mathscr{L}, \mathscr{L}'$ *be two partitions of* Ω. *If* $\mathscr{L}' \perp_w \mathscr{L}$, *then the following statements hold:*

1. *for every* $E'_i, E'_j \in \mathscr{L}'$, $A_j \subseteq A_i$ *or* $A_i \subseteq A_j$;
2. *there exists* $E'_i \in \mathscr{L}'$ *such that* $E'_i \wedge E_j \neq \emptyset$ *for any* $E_j \in \mathscr{L}$;
3. *if there exist* $E'_i \in \mathscr{L}'$ *and* $E_j \in \mathscr{L}$ *such that* $E'_i \subseteq E_j$, *then, for every* $E'_r \in \mathscr{L}'$, *we have* $E'_r \wedge E_j \neq \emptyset$.
4. *there exists at most one* $E_k \in \mathscr{L}$ *such that* $E'_i \subseteq E_k$ *for some* $E'_i \in \mathscr{L}'$.

Proposition 1 easily implies that if $\mathscr{L}$ is a refinement of $\mathscr{L}'$, then $\mathscr{L}' \not\perp_w \mathscr{L}$.

Theorem 1. *Let* $\mathscr{L} = \{E_1, ..., E_i, ..., E_n\}$ *and* $\mathscr{L}' = \{E'_1, ..., E'_j, ..., E'_m\}$ *be two partitions of* Ω. *The following two conditions are equivalent:*

1. $\mathscr{L}' \perp_w \mathscr{L}$;
2. *there exists a permutation of the indices* $1, ..., m$ *such that the corresponding events* $A_1, ..., A_j, ...A_m$ *are completely ordered by inclusion.*

4 Possibility as Enlargement of a Coherent Probability

In [2, 5] it has been proved that, if $\mathscr{L}, \mathscr{L}'$ are two partitions of Ω and $\mathscr{A}'$ the algebra spanned by $\mathscr{L}'$, and P a probability distribution on $\mathscr{L}$, then, considering the family $\mathbf{P}$ of probabilities P_i extending P on $\mathscr{L} \cup \mathscr{A}'$, the lower bound of $\mathbf{P}$ on $\mathscr{A}'$ is a belief function (and the upper bound a plausibility function). Vice versa, for any belief function *Bel* on an algebra $\mathscr{A}'$ there exists a partition of Ω and a relevant probability distribution such that the lower bound of the class of probability extending P on $\mathscr{A}'$ coincides with *Bel* [2] (similarly for a plausibility function). This result is independent of any logical relation between the partition $\mathscr{L}$ and that of atoms of $\mathscr{A}'$. Obviously, the logical constraints rule the numerical values of the belief (or plausibility) function.

In [3] we proved that if two partitions are weakly logically independent, then the plausibility obtained as upper envelope of the class $\mathbf{P}$ is a possibility:

Theorem 2. *Let* $\mathscr{L}, \mathscr{L}'$ *be two partitions of* Ω *and* $\mathscr{A}'$ *the algebra spanned by* $\mathscr{L}'$. *Let* P *be a probability distribution on* $\mathscr{L}$ *and* $\overline{P}$ *the upper envelope of the class* $\mathbf{P} = \{P'\}$ *of all the probabilities extending* P *onto* $\mathscr{L} \cup \mathscr{A}'$. *If* $\mathscr{L}' \perp_w \mathscr{L}$, *then* $\overline{P}$ *is a possibility measure on* $\mathscr{A}'$.

This result is related to that given in [9]: any set of lower bounds on a nested class $A_1, ..., A_m$ induces an upper probability, that is a possibility. As shown in [3] a possibility

can be obtained also when $\mathscr{L}' \not\perp_w \mathscr{L}$ (but not if the probability distribution is strictly positive).

Theorem 3 shows how weakly logically independent partitions rule the transition from probability to possibility and also the other way round.

Theorem 3. *Consider a possibility measure Π on an algebra $\mathscr{A}$ and let $\mathscr{L}$ be the set of atoms of $\mathscr{A}$. Then, there exists a partition $\mathscr{L}'$ and a probability distribution on $\mathscr{L}'$ such that:*

1. *$\mathscr{L}' \perp_w \mathscr{L}$,*
2. *the upper envelope $\overline{P}$ of the class $\mathbf{P} = \{P'\}$ of all the probabilities extending P on $\mathscr{L}' \cup \mathscr{A}$ coincides on $\mathscr{A}$ with the possibility measure Π.*

Remark 1. In [3] we proved that, given two logically independent partitions $\mathscr{L}$ and $\mathscr{L}'$, the upper envelope of the extensions on $\mathscr{L} \cup \mathscr{A}'$ of a probability P on $\mathscr{L}$ is a possibility on $\mathscr{A}'$ and, for any $A \in \mathscr{A}' \setminus \emptyset$, $\overline{P}(A) = 1$. Thus, we get in this case the non informative possibility independently of the initial probability distribution.

5 From a Coherent Probability to the Upper Possibility

All the results of the previous Section are based on the assumption that the initial information is handled by a probability distribution on the elements of a partition of Ω. Now we start instead from a coherent probability on an arbitrary set of events $\mathscr{E}$. Then, we need to consider all the extensions on any other finite set $\mathscr{E}'$. Since coherence implies the existence of a class $\mathbf{P} = \{P_i\}$ of probabilities on the set $\mathscr{C}$ of atoms generated by $\mathscr{E}$, for any such probability distributions $P_i \in \mathbf{P}$ we have a plausibility [belief] as an upper [lower] bound of the probabilities extending P_i in $\mathscr{E}'$; moreover if $\mathscr{C} \perp_w \mathscr{C}'$ (with $\mathscr{C}'$ the set of atoms generated by $\mathscr{E}'$) for each $P_i \in \mathbf{P}$ we obtain a possibility.

In general it is not possible to characterize the set of plausibilities, since the upper envelope of plausibilities is not a plausibility. In this Section we prove instead that, when $\mathscr{C} \perp_w \mathscr{C}'$, we obtain a class of possibilities such that both their upper and lower envelopes are possibilities (i.e., that dominating and that dominated by all other possibilities, respectively).

Theorem 4. *Let $\mathscr{E}, \mathscr{E}'$ be two finite sets of events and $\mathscr{C}, \mathscr{C}'$ the corresponding sets of atoms generated by $\mathscr{E}$ and $\mathscr{E}'$. Moreover, let P be a coherent probability on $\mathscr{E}$, and $\mathbf{P}$ the set of coherent probability extensions of P on $\mathscr{E} \cup \mathscr{E}'$. If $\mathscr{C} \perp_w \mathscr{C}'$, then the upper envelope of $\mathbf{P}$ on $\mathscr{E}'$ is a coherent possibility.*

Proof. The coherent probability P on $\mathscr{E}$ can be extended on $\mathscr{E} \cup \mathscr{C}$ and let $\mathbf{P} = \{P'\}$ be the set of all the coherent probability extensions of P on $\mathscr{E} \cup \mathscr{C}$. Since $\mathscr{C}$ is finite [12] there exists a finite subset $\mathbf{P}_m$ of $\mathbf{P}$ such that

$$\overline{P}(C) = \sup_{P' \in \mathbf{P}} P'(C) = \sup_{P' \in \mathbf{P}_m} P'(C)$$

for any $C \in \mathscr{C}$. Since $\mathscr{C} \perp_w \mathscr{C}'$, the upper envelope of the extensions of a probability $P' \in \mathbf{P}_m$ is a possibility distribution on the algebra $\mathscr{A}'$ generated by $\mathscr{C}'$ by Theorem 2.

Then, we can consider the finite set $\{\Pi\}$ of possibilities on $\mathscr{A}'$ associated to $\mathbf{P}_m$. The upper envelope Π^* of $\{\Pi\}$ is a possibility and then the restriction of Π^* on $\mathscr{E}' \subseteq \mathscr{A}'$ is a coherent possibility. The coherent possibility Π^* on $\mathscr{E}'$ coincides with the upper envelope of P on $\mathscr{E}'$, in fact for any $E \in \mathscr{E}'$

$$\Pi^*(E) = \sup \Pi(E) = \sup_{P' \in \mathbf{P}_m} \sum_{C_r \wedge E \neq \emptyset} P'(C_r) = \sup_{P' \in \mathbf{P}} \sum_{C_r \wedge E \neq \emptyset} P'(C_r) = \overline{P}(E). \qquad \square$$

The coherent possibility Π^* of the above result is the less informative, in the sense that it dominates any possibility arising in the enlargement procedure. Now, we are interested also to look for the most informative one, in the sense that is dominated by any other one.

Theorem 5. *Let $\mathscr{E}, \mathscr{E}'$ be two finite sets of events, $\mathscr{C}, \mathscr{C}'$ the corresponding sets of atoms generated by $\mathscr{E}$ and $\mathscr{E}'$ and $\mathscr{A}, \mathscr{A}'$ the algebras spanned by $\mathscr{E}, \mathscr{E}'$, respectively. Given a coherent probability P on $\mathscr{E}$, consider the lower envelope $\underline{P}$ of the set $\mathbf{P} = \{P'\}$ of extensions of P on $\mathscr{A}$ and the function Π_* defined on $\mathscr{A}'$ as follows: for any $B \in \mathscr{A}'$*

$$\Pi_*(B) = \inf_{A \in \mathscr{A} : A \supseteq B} \underline{P}(A).$$

If $\mathscr{C} \perp_w \mathscr{C}'$, then Π_* is a coherent possibility on $\mathscr{E}'$. Moreover, the upper envelope Π_1 on $\mathscr{A}'$ of the extensions of any $P' \in \mathbf{P}$ dominates Π_*.

Proof. If $\mathscr{C}' \perp_w \mathscr{C}$, then by Theorem 1 there exists an ordering on the elements of $\mathscr{C}' = \{E'_1, \ldots, E'_m\}$ such that $A_i \subseteq A_{i+1}$ for $i = 1, \ldots, m-1$. Hence, for any $E'_i \in \mathscr{C}'$ one has

$$\Pi_*(E'_i) = \inf_{A \in \mathscr{A} : A \supseteq E'_i} \underline{P}(A) = \underline{P}(A_i).$$

In particular, since $A_m = \Omega$, it follows $\Pi_*(E'_m) = 1$. Consider any $F = \bigvee_{j \in J} E'_j$: there exists $\overline{j} \in J$ such that $j < \overline{j}$ for any $j \in J$ (with $j \neq \overline{j}$), then $A_j \subseteq A_{\overline{j}}$ and so

$$\Pi_*(F) = \inf_{A \in \mathscr{A} : A \supseteq F} \underline{P}(A) = \underline{P}(A_{\overline{j}}) = \max_{j \in J} \{\underline{P}(A_j)\} = \max_{E'_i \subseteq F} \Pi_*(E'_i)$$

then Π_* is a possibility on $\mathscr{A}'$ and so Π_* on $\mathscr{E}'$ is a coherent possibility. Now, given $P' \in \mathbf{P}$, since $\mathscr{C}' \perp_w \mathscr{C}$, by Theorem 2 the upper envelope Π_1 of the extensions of P' on $\mathscr{A}'$ is a possibility and for any $F \in \mathscr{A}'$, $F = \bigvee_{j \in J} E'_j$, there exists $\overline{j} \in J$ such that $A_j \subseteq A_{\overline{j}}$ for any $j \in J$, $F \subseteq A_{\overline{j}}$ and

$$\Pi_*(F) = \inf_{A \in \mathscr{A} : A \supseteq F} \underline{P}(A) = \underline{P}(A_{\overline{j}}) \leq P'(A_{\overline{j}}) = \Pi_1(F). \qquad \square$$

By the previous result we obtain a possibility Π_* that is dominated by any possibility obtained as the upper envelope of the extensions of a coherent probability (on the assumption that the two sets of atoms are weakly logically independent). Note that in general the minimum of a set of possibilities is not a possibility, while in the case that the possibilities are obtained through the inferential procedure shown in Section 4 their infimum, that coincides with Π_*, is still a possibility. Then, for any $F \in \mathscr{A}'$ we get two (possibly coincident) values $\Pi_*(F)$ and $\Pi^*(F)$. The following Theorem 6 shows that any possibility Π weakly comonotone with (Π_*, Π^*) can be obtained as the upper envelope of the extensions of a coherent probability, where weakly monotonicity is defined as follows:

Definition 3. *A possibility Π on $\mathscr{A}'$ is weakly comonotone with (Π_*, Π^*) if $\Pi_*(F) \leq \Pi(F) \leq \Pi^*(F)$ for any $F \in \mathscr{A}'$ and for any pair of atoms $E'_i, E'_j \in \mathscr{A}'$ the following conditions hold:*

- *if $\Pi_*(E'_i) < \Pi_*(E'_j)$ or $\Pi^*(E'_i) < \Pi^*(E'_j)$, then $\Pi(E'_i) \leq \Pi(E'_j)$,*
- *if $\Pi_*(E'_i) = \Pi_*(E'_j)$ and $\Pi^*(E'_i) = \Pi^*(E'_j)$, then $\Pi(E'_i) = \Pi(E'_j)$.*

Theorem 6. *Let $\mathscr{E}, \mathscr{E}', \mathscr{C}, \mathscr{C}', \mathscr{A}, \mathscr{A}'$, $\mathbf{P}$ and Π_* as in Theorem 5 and consider the upper envelope Π^* on $\mathscr{A}'$ of the coherent extensions of P. If $\mathscr{C} \perp_w \mathscr{C}'$, then Π_* and Π^* are possibilities on $\mathscr{A}'$. Moreover, Π_* is dominated by Π^* and, for any coherent possibility Π on $\mathscr{A}'$ weakly comonotone with (Π_*, Π^*), there exists a coherent probability P on $\mathscr{E}$ such that the upper envelope of the extensions of P on $\mathscr{A}'$ coincides with Π.*

Proof. By Theorem 5 and by Theorem 4 the functions Π^* and Π^* are possibilities and $\Pi_*(F) \leq \Pi^*(F)$ for any $F \in \mathscr{A}'$. Now, consider any possibility Π on $\mathscr{A}'$ satisfying the conditions in the hypothesis. Let us assume (without loss of generality, see Proposition 1) that the partition $\mathscr{C}' = \{E_1, ..., E_k\}$ is ordered in a way that for any $i < j$ one has $\Pi(E_i) \leq \Pi(E_j)$ for any $E_i, E_j \in \mathscr{C}'$. This order is compatible with that built starting from Π_* or Π^*, then the partition $\mathscr{C}' = \{F_1, ..., F_m\}$ is such that $\mathscr{C}' \perp_w \mathscr{L}$ and considering $E'_i, E'_j \in \mathscr{C}'$ if $i < j$, then $\Pi(E'_i) \leq \Pi(E'_j)$, since the associated A_i and A_j are such that $A_i \subseteq A_j$. Hence, there exists a probability on $\mathscr{A}$ such that $P(A_i) \leq P(A_j)$ for any $i < j$ and $P(A_j) = \Pi(E'_j) - \Pi(E'_{j-1})$ for any $j = 1, ..., k$ by putting $\Pi(E'_0) = \Pi(\emptyset)$. This probability on $\mathscr{A}$ generates Π through the inferential process. □

References

1. Coletti, G., Scozzafava, R.: Probabilistic Logic in a Coherent Setting. Trends in Logic, vol. 15. Kluwer Academic Publishers, Dordrecht (2002)
2. Coletti, G., Scozzafava, R.: Toward a general theory of conditional beliefs. Internat. J. Intelligent. Syst. 21, 229–259 (2006)
3. Coletti, G., Scozzafava, R., Vantaggi, B.: Possibility measures through a probabilistic inferential process. In: Barone, J., Tastle, B., Yager, R. (eds.) Proceedings of North America Fuzzy Information Processing Society 2008 (NAFIPS 2008, New York, USA), IEEE CN: CFP08750-CDR Omnipress (2008)
4. Coletti, G., Vantaggi, B.: T-conditional possibilities: coherence and inference. Fuzzy Sets Syst. (in press, 2008) doi:10.1016/j.fss.2008.04.006
5. De Cooman, G., Troffaes, M., Miranda, E.: n-Monotone lower previsions and lower integrals. In: Cozman, F.G., Nau, R., Seidenfeld, T. (eds.) Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2005, Pittsburgh, Pennsylvania, USA), pp. 145–154 (2005)
6. De Finetti, B.: Teoria della probabilitá. Einaudi, Torino. In: Theory of Probability: A Critical Introductory Treatment, John Wiley & Sons, Chichester (1970) (Engl. Transl. 1974)
7. Delgado, M., Moral, S.: On the concept of possibility-probability consistency. Fuzzy Sets Syst. 21(3), 311–318 (1987)
8. Dubois, D., Nguyen, H.T., Prade, H.: Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. In: Dubois, D., Prade, H. (eds.) Fundamentals of Fuzzy Sets. The Handbooks of Fuzzy Sets, vol. 7, pp. 343–438. Kluwer Academic, Dordrecht (2000)

9. Dubois, D., Prade, H.: When upper probabilities are possibility measures. Fuzzy Sets Syst. 49, 65–74 (1992)
10. Dubois, D., Prade, H.: Qualitative possibility theory and its probabilistic connections. In: Grzegorzewski, P., Hryniewicz, O., Gil, M.A. (eds.) Soft Methods in Probability, Statistics and Data Analysis, pp. 3–26. Physica Verlag, Heidelberg-Germany (2002)
11. Dubois, D., Prade, H., Smets, P.: A definition of subjective possibility. Operacyjne I Decyzje (Pologne) 4, 7–22 (2003)
12. Shapley, L.S.: Cores of convex games. Internat. J. Game Theory 1, 11–26 (1971)
13. Sudkamp, T.: On probability-possibility transformations. Fuzzy Sets Syst. 51(1), 311–318 (1992)
14. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Non-well-Founded Probabilities on Streams

Andrew Schumann

Department of Philosophy and Science Methodology, Belarusian State University, Minsk, Belarus

Abstract. In the paper we propose non-well-founded probabilities as a kind of fuzzy ones. They are defined on the set of streams. We also show that the set of p-adic numbers can be understood as a set of streams. In the set theory without the axiom of foundation, the powerset is not a Boolean algebra in the general case. Therefore, if we tried to define probabilities on non-well-founded data, i.e. on streams or p-adic numbers, then we couldn't use the Kolmogorovian approach and we should refer to non-Kolmogorovian models of probabilities. Probabilities on streams have a lot of unexpected properties. For instance, p-adic probabilities may be negative rational numbers as well as rational numbers that are larger than 1. Bayes' formula doesn't also hold in the general case for non-well-founded probabilities.

1 Introduction

A non-well-founded (NWF) set theory belongs to axiomatic set theories that violate the rule of well-foundedness and, as an example, allow sets to contain themselves: $X \in X$.

Recall that a relation R is called well-founded if for every set x there is no infinitely descending chain $\ldots Rx_2Rx_1Rx_0 = x$. The foundation axiom postulates that the set-membership relation $\in$ is well-founded: for every set x there exists no infinitely descending chain $\cdots \in x_2 \in x_1 \in x_0 = x$. Evidently, the statement $X \in X$ or $X = \{X\}$ doesn't satisfy the foundation axiom.

In NWF set theories, the foundation axiom of Zermelo-Fraenkel set theory is replaced by axioms implying its negation. The theory of NWF sets has been explicitly applied in diverse fields such as logical modeling non-terminating computational processes and behavior of interactive systems in computer science (process algebra, coalgebra, logical programming based on coinduction and corecursion), linguistics and natural language semantics (situation theory), logic (analysis of semantic paradoxes). NWF sets have been also implicitly used in non-standard (more precisely, non-Archimedean) analysis like infinitesimal and p-adic analysis.

Denying the foundation axiom in number systems implies setting the non-Archimedean ordering structure. In this paper we consider probabilities on streams and as well as on p-adic numbers and we show that these probabilities can be regarded only as fuzzy ones, because for NWF mathematical objects, the powerset is not a Boolean algebra in the general case. The conventional (Kolmogorov's) probability theory is built in the language of well-founded mathematics. It sets a framework of modern physics, taking into account that physical reality is regarded in modern science as reality of stable repetitive phenomena (phenomena that have probabilities, i.e. do not fluctuate in the

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 59–65, 2008.
springerlink.com

standard real metric). The main corollary of our paper is that *physical reality may be regarded as NWF in the framework of probabilities distributed on streams, in particular, distributed on p-adic numbers.*

2 Ordering Relations on Streams and p-adic Numbers

One of the most useful NWF mathematical object is a stream – a recursive data-type of the form $s = \langle a, s' \rangle$, where s' is another stream. The notion of *stream calculus* was introduced by Escardó and Pavlović [4] as a means to do symbolic computation using the coinduction principle instead of the induction one. Let A be any set. We define the set A^ω of all streams over A as $A^\omega = \{\sigma\colon \{0,1,2,\dots\} \to A\}$. For a stream σ, we call $\sigma(0)$ the initial value of σ. We define the *derivative* $\sigma(0)$ of a stream σ, for all $n \geq 0$, by $\sigma'(n) = \sigma(n+1)$. For any $n \geq 0$, $\sigma(n)$ is called the n-th element of σ. It can also be expressed in terms of higher-order stream derivatives, defined, for all $k \geq 0$, by $\sigma^{(0)} = \sigma$; $\sigma^{(k+1)} = (\sigma^{(k)})'$. In this case the n-th element of a stream σ is given by $\sigma(n) = \sigma^{(n)}(0)$. Also, the stream is understood as an infinite sequence of derivatives. It will be denoted by an infinite sequence of values or by an infinite tuple: $\sigma = \sigma(0) :: \sigma(1) :: \sigma(2) :: \cdots :: \sigma(n-1) :: \sigma^{(n)}$, $\sigma = \langle \sigma(0), \sigma(1), \sigma(2), \dots \rangle$.

Streams are defined by coinduction: two streams σ and τ in A^ω are equal if they are *bisimilar*: (i) $\sigma(0) = \tau(0)$ (they have the same *initial value*) and (ii) $\sigma' = \tau'$ (they have the same *differential equation*). To set addition and multiplication by coinduction, we should use the following facts about differentiation of sums and products by applying the basic operations: $(\sigma + \tau)' = \sigma' + \tau'$, $(\sigma \times \tau)' = (|\sigma(0)| \times \tau') + (\sigma' \times \tau)$, where $|\sigma(0)| = \langle \sigma(0), 0, 0, 0, \dots \rangle$. Now we can define them and as well as one another stream operation as follows:

We can embed the real numbers into the streams by defining the following constant stream. Let $r \in \mathbb{R}$. Then $|r| = \langle r, 0, 0, 0, \dots \rangle$ is defined so: its differential equation is $|r|' = [0]$, its initial value is $|r|(0) = r$. We are to rely on our intuitions that it would be natural to define the positive real numbers to be less than the positive streams.

Consider the set of streams $[0,1]^\omega$ and extend the standard order structure on $[0,1]$ to a partial order structure on $[0,1]^\omega$. Further define this order as follows:

$\mathscr{O}_{[0,1]^\omega}$ (1) For any streams $\sigma, \tau \in [0,1]^\omega$, we set $\sigma \leq \tau$ if $\sigma(n) \leq \tau(n)$ for every $n \in \mathbb{N}$. For any streams $\sigma, \tau \in [0,1]^\omega$, we set $\sigma = \tau$ if σ, τ are bisimilar. For any streams $\sigma, \tau \in [0,1]^\omega$, we set $\sigma < \tau$ if $\sigma(n) \leq \tau(n)$ for every $n \in \mathbb{N}$ and there exists n_0 such that $\sigma(n_0) \neq \tau(n_0)$. (2) Each stream of the form $|r| \in [0,1]^\omega$ (i.e. constant stream) is less than inconstant stream σ.

Table 1. Coinductive definitions of sum, product and inverse

Differential equation	Initial value	Name
$(\sigma + \tau)' = \sigma' + \tau'$	$(\sigma + \tau)(0) = \sigma(0) + \tau(0)$	Sum
$(\sigma \times \tau)' = (\lvert\sigma(0)\rvert \times \tau') + (\sigma' \times \tau)$	$(\sigma \times \tau)(0) = \sigma(0) \times \tau(0)$	Product
$(\sigma^{-1})' = \lvert -1\rvert \times \lvert\sigma(0)^{-1}\rvert \times \sigma' \times \sigma^{-1}$	$(\sigma^{-1})(0) = \sigma(0)^{-1}$	Inverse

This ordering relation is not linear, but partial, because there exist streams $\sigma, \tau \in [0,1]^\omega$, which are incompatible.

Introduce two operations sup, inf in the partial order structure $\mathscr{O}_{[0,1]^\omega}$. Assume that $\sigma, \tau \in [0,1]^\omega$ are either both constant streams or both inconstant streams. Then their supremum and infimum are defined by coinduction: the differential equation of supremum is $(\sup(\sigma,\tau))' = \sup(\sigma',\tau')$ and its initial value is $(\sup(\sigma,\tau))(0) = \sup(\sigma(0),\tau(0))$, the differential equation of infimum is $(\inf(\sigma,\tau))' = \inf(\sigma',\tau')$ and its initial value is $(\inf(\sigma,\tau))(0) = \inf(\sigma(0),\tau(0))$. Suppose now that one and only one of $\sigma, \tau \in [0,1]^\omega$ is constant, then an inconstant stream is greater than a constant one, therefore their supremum gives an inconstant stream, but their infimum gives a constant stream.

According to $\mathscr{O}_{[0,1]^\omega}$, there exist the maximal stream $[1] \in [0,1]^\omega$ and the minimal stream $[0] = |0| \in [0,1]^\omega$.

In 1897 the German mathematician Kurt Hensel presented an idea how to use an analogy of Taylor and Laurent series to study algebraic numbers by expressing them as an expansion in terms of powers of a prime number. He was mainly inspired by the work of Kummer. This approach by Hensel led him to introduce the p-adic numbers. There are many books which give a good introduction to the p-adic theory, see for instance Koblitz [3].

It can be easily shown that p-adic numbers may be represented as potentially infinite data structures such as streams. Each stream of the form $\sigma = \sigma(0) :: \sigma(1) :: \sigma(2) :: \cdots :: \sigma(n-1) :: \sigma^{(n)}$, where $\sigma(n) \in \{0,1,\ldots,p-1\}$ for every $n \in \mathbb{N}$, may be converted into a p-adic integer by the following rule:

$$\forall n \in \mathbb{N}, \quad \sigma(n) = \sum_{k=0}^{n} \sigma(k) \cdot p^k \wedge \sigma(n) = \sigma(0) :: \sigma(1) :: \cdots :: \sigma(n). \tag{1}$$

And vice versa, each p-dic integer may be converted into a stream taking rule (1). Such a stream is called *p-adic*.

Extend rule (1) as follows. Suppose that we have a stream of the form $\sigma = \sigma(0) :: \sigma(1) :: \sigma(2) :: \cdots :: \sigma(n-1) :: \sigma^{(n)}$, where $\sigma(n) \geq 0$ for every $n \in \mathbb{N}$. Then its p-adic representation is

$$\forall n \in \mathbb{N}, \sigma(n) = \sum_{k=0}^{m} \tau(k) \cdot p^k \wedge \sigma(n) = \tau(0) :: \tau(1) :: \cdots :: \tau(m), \tag{2}$$

where $\tau(i) \in \{0,1,\ldots,p-1\}$ for every $i = \overline{1,m}$ and $\sum_{k=0}^{m} \tau(k) \cdot p^k = \sum_{k=0}^{n} \sigma(k) \cdot p^k$. (In the case $\sigma(i) \in \{0,1,\ldots,p-1\}$ for every $i = \overline{1,n}$, we have $n = m$ and then $\sigma(i) = \tau(i)$ for every $i = \overline{1,n}$.) Such a stream is called *p-adic* too. Its *canonical form* is $\tau(0) :: \tau(1) :: \cdots :: \tau(m) :: \tau^{(m+1)}$, where $\tau(n) \in \{0,1,\ldots,p-1\}$ for every $n \in \mathbb{N}$.

Using (1), (2), we can show that sum, product and inverse have the same differential equations and initial values as in stream calculus. This proves that *p-adic numbers are one of the natural interpretations of streams*.

It is easily shown that the set A^ω of all p-adic streams includes the set of natural numbers. Let n be a natural number. It has a finite p-adic expantion $n = \sum_{k=0}^{m} \alpha_k \cdot p^k$. Then we can identify n with a p-adic stream $\sigma = \sigma(0) :: \sigma(1) :: \cdots :: \sigma(m) :: \sigma^{(m+1)}$, where $\sigma(i) = \alpha_i$ for $i = \overline{0,m}$ and $\sigma^{(m+1)} = [0]$.

Extend the standard order structure on $\mathbb{N}$ to a partial order structure on p-adic streams (i.e. on $\mathbb{Z}_p$).

1. For any p-adic streams $\sigma, \tau \in \mathbb{N}$ we have $\sigma \leq \tau$ in $\mathbb{N}$ iff $\sigma \leq \tau$ in $\mathbb{Z}_p$,
2. Each p-adic stream $\sigma = \sigma(0) :: \sigma(1) :: \cdots :: \sigma(m) :: \sigma^{(m+1)}$, where $\sigma^{(m+1)} = [0]$ (i.e. each finite natural number), is less than any infinite number τ, i.e. $\sigma < \tau$ for any $\sigma \in \mathbb{N}$ and $\tau \in \mathbb{Z}_p \backslash \mathbb{N}$.

Define this partial order structure on $\mathbb{Z}_p$ as follows:

$\mathscr{O}_{\mathbb{Z}_p}$ Let $\sigma = \sigma(0) :: \sigma(1) :: \cdots :: \sigma(n-1) :: \sigma^{(n)}$ and $\tau = \tau(0) :: \tau(1) :: \cdots :: \tau(n-1) :: \tau^{(n)}$ be p-adic streams. (1) We set $\sigma < \tau$ if the following three conditions hold: (i) there exists n such that $\sigma(n) < \tau(n)$; (ii) $\sigma(k) \leq \tau(k)$ for all $k > n$; (iii) σ is a finite integer, i.e. there exists m such that $\sigma^{(m)} = [0]$. (2) We set $\sigma = \tau$ if σ and τ are bisimilar. (3) Suppose that σ, τ are infinite integers. We set $\sigma \leq \tau$ by coinduction: $\sigma \leq \tau$ iff $\sigma(n) \leq \tau(n)$ for every $n \in \mathbb{N}$. We set $\sigma < \tau$ if we have $\sigma \leq \tau$ and there exists $n_0 \in \mathbb{N}$ such that $\sigma(n_0) < \tau(n_0)$.

Now introduce two operations sup, inf in the partial order structure on $\mathbb{Z}_p$. Suppose that p-adic streams σ, τ represent infinite p-adic integers. Then their sup and inf may be defined by coinduction as follows: the differential equation of supremum is $(\sup(\sigma,\tau))' = \sup(\sigma',\tau')$ and its initial value is $(\sup(\sigma,\tau))(0) = \sup(\sigma(0),\tau(0))$, the differential equation of infimum is $(\inf(\sigma,\tau))' = \inf(\sigma',\tau')$ and its initial value is $(\inf(\sigma,\tau))(0) = \inf(\sigma(0),\tau(0))$. Now suppose that at most one of two streams σ, τ represents a finite p-adic integer. In this case $\sup(\sigma,\tau) = \tau$ if and only if $\sigma \leq \tau$ under condition $\mathscr{O}_{\mathbb{Z}_p}$ and $\inf(\sigma,\tau) = \sigma$ if and only if $\sigma \leq \tau$ under condition $\mathscr{O}_{\mathbb{Z}_p}$.

It is important to remark that there exists the maximal p-adic stream $N_{max} \in \mathbb{Z}_p$ under condition $\mathscr{O}_{\mathbb{Z}_p}$. It is easy to see: $N_{max} = [p-1] = -1 = (p-1) + (p-1)\cdot p + \ldots + (p-1)\cdot p^k + \ldots$

3 Non-well-Founded Probabilities

There is a problem how it is possible to define probabilities on stream structures if we have no opportunity to put them on an algebra of subsets, taking into account the following result:

Proposition 1. *Define union, intersection and complement in the standard way. The powerset $\mathscr{P}(A^\omega)$, where A^ω is the set of all streams over A, is not a Boolean algebra.*

Proof. Consider a counterexample on 7-adic streams. Let $A_1 = \{x\colon 0 \leq x \leq \ldots 11234321\}$ and $A_2 = \{x\colon \ldots 66532345 \leq x \leq \ldots 66666\}$ be subsets of $\mathbb{Z}_7$. It is readily seen that $\neg(A_1 \cap A_2) = \mathbb{Z}_7$, but $(\neg A_1 \cup \neg A_2) \subset \mathbb{Z}_7$, because $\neg A_1 = \{x\colon \ldots 11234321 < x \leq \ldots 66666\}$ and $\neg A_2 = \{x\colon 0 \leq x < \ldots 66532345\}$, therefore $\mathbb{Z}_7 \setminus (\neg A_1 \cup \neg A_2) = A_3 = \{x\colon x = \ldots y_5 y_4 3 y_2 y_1 y_0$, where $y_i \in \{0,1,\ldots,6\}$ for each $i \in \mathbb{N}\backslash\{3\}\}$. It is obvious that the set A_3 is infinite. As a result, we obtain that $\neg(A_1 \cap A_2) \neq \neg A_1 \cup \neg A_2$ in the general case. □

This proposition is a particular case of the following provable statement: *if A is a NWF set, then its powerset will not be a Boolean algebra in the general case*.

In stream calculus and p-adic calculus we have, evidently, a different partial ordering relation and obtain different powersets $\mathscr{P}^{[0,1]^\omega}(A^\omega)$, $\mathscr{P}^{\mathbb{Z}_p}(A^\omega)$, but in any case there is no Boolean algebra, because *the complement in them is not Boolean*. The powersets $\mathscr{P}^{[0,1]^\omega}(A^\omega)$, $\mathscr{P}^{\mathbb{Z}_p}(A^\omega)$ should be interpreted as a corresponding class $\mathscr{F}^V(A^\omega)$ of fuzzy subsets $Y \subset A^\omega$, where V is equal one of sets $[0,1]^\omega$, $\mathbb{Z}_p$.

We can try to get NWF probabilities on a NWF algebra $\mathscr{F}^V(A^\omega)$ of fuzzy subsets $Y \subset A^\omega$ that consists of the following: (1) union, intersection, and difference of two *NWF fuzzy* subsets of A^ω; (2) $\emptyset$ and A^ω. In this case a *finitely additive NWF probability measure* is a nonnegative set function $P(\cdot)$ defined for sets $Y \in \mathscr{F}^V(A^\omega)$ that runs the set V and satisfies the following properties: (1) $P(A) \geq [0]$ for all $A \in \mathscr{F}^V(A^\omega)$, (2) $P(A^\omega) = |1|$ and $P(\emptyset) = [0]$, (3) if $A \in \mathscr{F}^V(A^\omega)$ and $B \in \mathscr{F}^V(A^\omega)$ are disjoint, then $P(A \cup B) = P(A) + P(B)$, (4) $P(\neg A) = |1| + |-1| \times P(A)$ for all $A \in \mathscr{F}^V(A^\omega)$.

This probability measure is called *NWF probability*. Their main originality is that conditions (3), (4) are independent. As a result, in a probability space $\langle X, \mathscr{F}^V(X), P\rangle$ some Bayes' formulas do not hold in the general case.

As an example of *trivial NWF probability* we can introduce the following function defined on streams by coinduction: (1) $P(\sigma) = \inf(\sigma, [1]) \times [1]^{-1}$ for every $\sigma \in [0,1]^\omega$, (2) $P(\sigma) = \inf(\sigma, N_{max}) \times N_{max}^{-1}$ for every $\sigma \in \mathbb{Z}_p$.

Consider a random experiment $\mathscr{S}$ and by $L = \{s_1, \ldots, s_m\}$ denote the set of all possible results of this experiment. The set $\mathscr{S}$ is called the label set, or the set of attributes. Suppose there are N realizations of $\mathscr{S}$ and write a result x_j after each realization. Then we obtain the finite sample: $x = (x_1, \ldots, x_N), x_j \in L$. A collective is an infinite idealization of this finite sample: $x = (x_1, \ldots, x_N, \ldots), x_j \in L$. Let us compute frequencies $\nu_N(\alpha; x) = n_N(\alpha; x)/N$, where $n_N(\alpha; x)$ is the number of realizations of the attribute α in the first N tests.

There exists the statistical stabilization of relative frequencies: the frequency $\nu_N(\alpha; x)$ approaches a limit as N approaches infinity for every label $\alpha \in L$. This limit $P(\alpha) = \lim \nu_N(\alpha; x)$ is said to be the probability of the label α in the frequency theory of probability. Sometimes this probability is denoted by $P_x(\alpha)$ to show a dependence on the collective x. Notice that the limits of relative frequencies have to be stable with respect to a place selection (a choice of a subsequence) in the collective.

The statistical stabilization of relative frequencies $\nu_N(\alpha; x)$ can be considered not only in the real topology on the field of rational numbers $\mathbb{Q}$ but also in any other topology on $\mathbb{Q}$. For instance, it is possible to construct the frequency theory in which probabilities were defined as limits of relative frequencies $\nu_N(\alpha; x)$ in the p-adic topology. The frequency theory of p-adic probability was proposed in [2]. It is a kind of NWF probability.

Since stream calculus and as well as p-adic calculus contain infinitely large numbers, they give the possibility to consider statistical ensembles with an infinite number of elements.

Define a *NWF operation of cardinality* $/\cdot/$ as follows: suppose $X \subseteq A^\omega$ and $K(\cdot)$ is the conventional operation of cardinality. Represent X by a Cartesian product $\prod_{j=0}^{\infty} X_j$, where X_0 is the set of all values of the form $\sigma(0)$ belonging to all streams of X, X_1 is the

set of all values of the form $\sigma(1)$ belonging to all streams of X, etc. Then $/X/$ is defined by coinduction: its initial value is $K(X_0)$, its differential equation is $(/X/)' = /X'/$. The informal meaning of NWF operation of cardinality is that we obtain an infinite sequence of conventional cardinalities $K(X_0), K(X_1), \ldots, K(X_m), \ldots$ that coinductively calculates not the number of streams from X but the number of their possible values at every step. It is evident, therefore, that the values of $/\cdot/$ are streams.

We study now some ensembles $S = S_N$, which have a NWF volume N, i.e. $/S/ = N$, where N is the stream of $[0,1]^\omega$ or $\mathbb{Z}_p$. Consider a sequence of ensembles S_j having volumes $K(S_j)$, $j = 0, 1, \ldots$ Get $S = \prod_{j=0}^{\infty} S_j$. Then the cardinality $/S/ = N$. We may imagine an ensemble S as being the population of a tower $T = T_S$, which has an infinite number of floors with the following distribution of population through floors: population of j-th floor is S_j. Set $T_k = \prod_{j=0}^{k} S_j \times \prod_{m=k+1}^{\infty} \emptyset_m$. This is population of the first $k+1$ floors. Let $A \subset S$ and let there exists: $n(A) = \lim\limits_{k\to\infty} n_k(A)$, where $n_k(A) = /A \cap T_k/$. The quantity $n(A)$ is said to be a *NWF volume of the set A*.

We define the probability of A by the standard proportional relation:

$$P(A) := P_S(A) = n(A) \times N^{-1},$$

where $/S/ = N$, $n(A) = /A \cap S/$.

We denote the family of all $A \subset S$, for which $P(A)$ exists, by $\mathscr{G}_S$. The sets $A \in \mathscr{G}_S$ are said to be events. The ordered system $\langle S, \mathscr{G}_S, P_S \rangle$ is called a *NWF ensemble probability space for the ensemble S*.

Proposition 2. *Let $\mathscr{F}$ be the NWF algebra of fuzzy subsets. Then $\mathscr{F} \subseteq \mathscr{G}_S$.*

Proof. Let A be a set of streams. Then $n(A) = /A/$ and the probability of A has the form: $P(A) = /A/ \times /S/^{-1}$.

For instance, let $B = \neg A$. Then $/B \cap T_k/ = /T_k/ + |-1| \times /A \cap T_k/$. Hence there exists $\lim\limits_{k\to\infty} /B \cap T_k/ = N + |-1| \times /A/$. This equality implies the standard formula: $P(\neg A) = |1| + |-1| \times P(A)$.

In particular, we have: $P(S) = |1|$. □

Proposition 3. *Let $A_1, A_2 \in \mathscr{G}_S$ and $A_1 \cap A_2 = \emptyset$. Then $A_1 \cup A_2 \in \mathscr{G}_S$ and $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.*

Proposition 4. *Let $A \in \mathscr{G}_S$, $P_S(A) \neq 0$ and $B \in \mathscr{G}_A$. Then $B \in \mathscr{G}_S$ and the following Bayes formula holds:*

$$P_A(B) = P_S(B/A) = P_S(B) \times P_S(A)^{-1}.$$

Proposition 5. *Let $N \in \mathbb{Z}_p$, $N \neq 0$ and let the ensemble S_{-1} have the p-adic volume $-1 = N_{max}$ (it is the largest ensemble, because N_{max} is the largest p-adic integer in accordance with $\mathscr{O}_{\mathbb{Z}_p}$).*

1. *Then $S_N \in \mathscr{G}_{S_{-1}}$ and $P_{S_{-1}}(S_N) = /S_N/ \times /S_{-1}/^{-1} = -N$.*
2. *Then $\mathscr{G}_{S_N} \subset \mathscr{G}_{S_{-1}}$ and probabilities $P_{S_N}(A)$ are calculated as conditional probabilities with respect to the subensemble S_N of ensemble S_{-1}: $P_{S_N}(A) = P_{S_{-1}}(\frac{A}{S_N})$ $= P_{S_{-1}}(A) \times P_{S_{-1}}(S_N)^{-1}, A \in \mathscr{G}_{S_N}$.*

If we take the p-adic case of NWF probability theory, then we observe essentially new properties of relative frequencies that do not appear on real numbers. For example, consider two attributes α_1 and α_2. Suppose that in the first $N := N_k = (\sum_{j=0}^{k} 2^j)^2$ tests the label α_1 has $n_N(\alpha_1;x) = 2^k$ realizations, α_2 has $n_N(\alpha_2;x) = \sum_{j=0}^{k} 2^j$ realizations. According to our intuition, their probabilities should be different, but in real probability theory we obtain: $P_x(\alpha_1) = \lim n_N(\alpha_1;x)/N = P_x(\alpha_2) = \lim n_N(\alpha_2;x)/N = 0$. In 2-adic probability theory we have $P_x(\alpha_1) = 0 \neq P_x(\alpha_2) = -1$, because in $\mathbb{Q}_2$, $2^k \to 0$, $k \to 0$, and $-1 = 1 + 2 + 2^2 + \cdots + 2^n + \ldots$

This example shows that in p-adic probability theory there are statistical phenomena for that relative sequences of observed events have non-zero probabilities in the p-adic metric, but do not have positive probabilities in the standard real metric.

4 Conclusion

Real probabilities are obtained as a result of a limiting process for rational frequencies in real topology by means of the law of large numbers. Using these probabilities we accept only well-founded phenomena. But we can introduce other forms of stability in physical experiments, namely p-adic forms [2], because besides the usual real topology, there exist only the p-adic topologies $p = 2,3,5,\ldots$ (for more details see [3]). The main reason is that p-adic numbers are, in fact, a unique alternative to real numbers: there is no other possibility to complete the field of rational numbers and obtain a new number field (Ostrovski's theorem, see, for example, [3]). In p-adic physics and in p-adic probability theory we assume that reality is NWF. Since statistical stabilization (the limiting process) can be considered not only in the real topology on the field of rational numbers $\mathbb{Q}$ but also in p-adic topologies on $\mathbb{Q}$, we see that reality can be considered as NWF too.

References

1. Aczel, P.: Non-Well-Founded Sets. CSLI Lecture Notes, vol. 14. Stanford University, Center for the Study of Language and Information, Stanford (1988)
2. Yu, K.A.: Interpretations of Probability. VSP Int. Sc. Publishers, Utrecht/Tokyo (1999)
3. Koblitz, N.: p-adic Numbers, p-adic Analysis, and Zeta-Functions, 2nd edn. Graduate Texts in Mathematics, vol. 58. Springer, New York (1984)
4. Pavlović, D., Escardó, M.H.: Calculus in coinductive form. In: Proceedings of the 13th Annual IEEE Symposium on Logic in Computer Science, pp. 408–417 (1998)
5. Schumann, A.: p-adic multiple-validity and p-adic valued logical calculi. J. Mult-Valued Logic Soft. Comput. 13(1-2), 29–60 (2007)

Relating Epistemic Irrelevance to Event Trees

Sébastien Destercke[1] and Gert de Cooman[2]

[1] Institut de Radioprotection et de Sureté Nucléaire (IRSN), Cadarache, France
[2] SYSTeMS Research Group, Ghent University, Zwijnaarde, Belgium

Abstract. We relate the epistemic irrelevance in Walley's behavioural theory of imprecise probabilities to the event-tree independence due to Shafer. In particular, we show that *forward irrelevance* is equivalent to event-tree independence in particular event trees, suitably generalised to allow for the fact that imprecise rather than precise probability models are attached to the nodes in the tree. This allows us to argue that in a theory of uncertain processes, the asymmetrical notion of epistemic irrelevance has a more important role to play than its more involved and symmetrical counterpart called epistemic independence.

Keywords: Independence, Forward irrelevance, Event trees, Stochastic process.

1 Introduction

Assessments of independence between variables are very important and useful in modelling uncertainty, as they allow for a reduction of complexity in many problems (e.g., in building joint models from marginal information, making statistical inferences, etc.). Here, we are interested in the case where beliefs are modelled by lower and upper expectations for random variables or, equivalently [13], by closed convex sets of (finitely additive) probabilities, also called *credal sets* [2, 3, 8]. In this *imprecise probabilities* setting, there are many different notions of irrelevance and independence, each with a different interpretation, but which generally coincide for models involving only *precise probabilities*, i.e., classical Bayesian belief models; see Couso et al. [1] for a review. Starting from given imprecise marginals, these different types of irrelevance and independence assessments will generally lead to different joint belief models, whereas they all lead to the classical independent product when marginal beliefs are modelled by precise, or Bayesian, probabilities. A discussion of this phenomenon can also be found in De Cooman and Miranda [6].

As far as we know, there are currently two important approaches to probability theory that involve lower and upper *expectations* (also called *previsions* or *prices*, depending on the interpretation): Walley's [13] *behavioural* approach, and Shafer and Vovk's [12] *game-theoretic* framework, where event trees play a central role. De Cooman and Hermans [4, 5] have shown that these two approaches can be related to each other, and they have introduced imprecise probability trees as a bridge between them. By showing that many results can be imported from one theory into the other, they make some progress towards a more unified handling of uncertainty.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 66–73, 2008.
springerlink.com

Here, we take one more step towards such a unification, by studying, in Sect. 5, how Walley's epistemic irrelevance [13, Chap. 9] can be related to the notion of event-tree independence that is central in Shafer's discussion of causal reasoning [11]. We discuss the relevance of our findings in the Conclusions, where we also argue why in a theory of uncertain processes, (forward) epistemic irrelevance may be more useful than its symmetrical counterpart, epistemic independence. But let us first recall the basic ideas behind Walley's behavioural theory of coherent lower previsions [13] (Sect. 2), Shafer's event and probability trees [11] (Sect. 3), and the imprecise probability trees that form the connection between them [4, 5] (Sect. 4).

2 Coherent Lower and Upper Previsions

In Walley's theory, beliefs held by a *subject* about the actual value of a random variable X on a finite[1] space $\mathscr{X}$ are modelled by coherent lower and upper previsions. We call *gamble* a real-valued function f on $\mathscr{X}$, and denote by $\mathscr{L}(\mathscr{X})$ the set of all gambles on $\mathscr{X}$. $f(X)$ is interpreted as an uncertain reward. A *lower prevision* $\underline{P}$ is a real-valued map defined on some subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X})$. Its conjugate *upper prevision* $\overline{P}$ is then defined on the set of gambles $-\mathscr{K} := \{-f\colon f \in \mathscr{K}\}$ by $\overline{P}(f) := -\underline{P}(-f)$. $\underline{P}(f)$ is interpreted as the subject's supremum buying price for the uncertain reward $f(X)$, i.e., the smallest price s such that the subject accepts to buy $f(X)$ for any price $\mu < s$, meaning he accepts the uncertain transaction $f(X) - \mu$. Given an event $A \subseteq \mathscr{X}$, its *lower probability* $\underline{P}(A)$ is the lower prevision of its *indicator* I_A, a gamble that assumes the value one on A and zero elsewhere. The *upper probability* $\overline{P}(A)$ is defined likewise in terms of the upper prevision $\overline{P}(I_A)$. With a lower prevision $\underline{P}$ we can associate a closed convex set of (dominating) probability mass functions: $\mathscr{M}(\underline{P}) := \{p \in \Sigma_{\mathscr{X}}\colon (\forall f \in \mathscr{K})(E_p(f) \geq \underline{P}(f)\}$, where $\Sigma_{\mathscr{X}}$ is the set (simplex) of all probability mass functions on $\mathscr{X}$, and $E_p(f) := \sum_{x\in\mathscr{X}} f(x)p(x)$. We call $\mathscr{M}(\underline{P})$ the *credal set* induced by $\underline{P}$. A lower prevision $\underline{P}$ is said to be *coherent* if and only if $\mathscr{M}(\underline{P}) \neq \emptyset$ and $\underline{P}(f) = \min\{E_p(f)\colon p \in \mathscr{M}(\underline{P})\}$ for all f in $\mathscr{K}$, i.e., if $\underline{P}$ is the *lower envelope* of $\mathscr{M}(\underline{P})$.

3 Event Trees

An event tree is composed of situations linked together, and it represents what relevant events may possibly happen in what particular order in the world, according to a particular subject. It is formally equivalent to a rooted tree in graph theory. We restrict ourselves to trees with finite depth and width. The notions we are about to introduce are illustrated in Fig. 1. A *situation* is a node in the tree. The *initial situation* is the root of the tree. A *terminal situation* is a leaf of the tree; all other situations, including the initial one, are called *non-terminal*. A *path* is a sequence of situations from the initial to a terminal situation. A path *goes through* a situation s if s belongs to it. The set Ω of all possible paths, or equivalently, of all terminal situations, is called the *sample space*. Any set of terminal situations is an *event*. Situations immediately following a non-terminal

[1] To make this discussion as simple as possible, we restrict ourselves to finite spaces throughout, but it is straightforward to extend our results to infinite spaces.

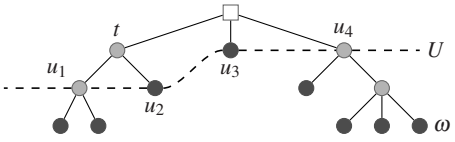


Fig. 1. Event tree with non-terminal situations (grey), terminal situations (black), and root $\Box$. $U = \{u_1, \ldots, u_4\}$ is a cut, $t < u_1$ and $d(t) = \{u_1, u_2\}$. Also, u_4 and t are disjoint, but not u_4 and ω.

situation s are called *daughters* of s, and the set of such daughters is denoted by $d(s)$. The link between a situation s and one of its daughters t is called a *move* from s to t. If a situation s is before a situation t in the tree, we say that s *strictly precedes* t, and denote this as $s < t$; and if a situation s is before or equal to a situation t, we say that s *precedes* t, and denote this as $s \leq t$. Two situations are called *disjoint* if there is no path they both belong to. A *cut* is a set of disjoint situations, such that every path goes through exactly one situation in the cut. If each situation in a cut V (strictly) precedes some situation in another cut U, then V is said to *(strictly) precede* U, and we denote this as $V \leq U$ ($V < U$).

4 Imprecise Probability Trees

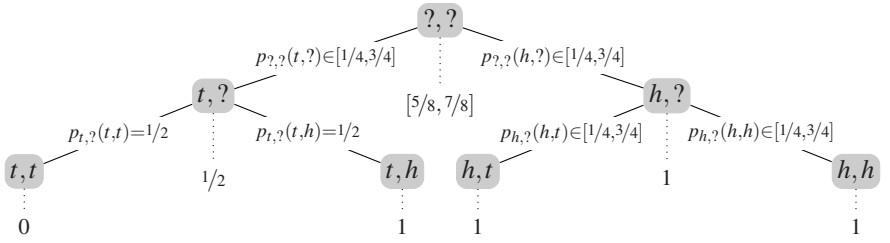
Branching probabilities p_s for a non-terminal situation s are non-negative numbers summing up to one, each of them attached to a different move originating in s: we denote by $p_s(t)$ the probability to go from s to its daughter t; p_s is a probability mass function on $d(s)$. A (precise) *probability tree* is an event tree for which every non-terminal situation has such branching probabilities.

An *imprecise probability tree*[2] is an event tree for which each non-terminal situation s has a closed convex set $\mathscr{M}_s$ of branching probabilities p_s, describing a subject's uncertainty about which move is going to be observed just after s. With an imprecise probability tree, we can associate coherent lower previsions. First of all, for any non-terminal situation s, and for any gamble h on $d(s)$, we can consider the lower prevision $\underline{P}_s(h) = \min\{E_{p_s}(h) \colon p_s \in \mathscr{M}_s\}$. $\underline{P}_s$ and $\mathscr{M}_s$ are equivalent *local* predictive models for what is going to be observed immediately after s. But we can also consider *global* predictive models: Let f be a gamble on the set of paths Ω. For every situation t, we consider the lower prevision $\underline{P}(f|t)$ conditional on t: the subject's supremum buying price for f, given that the actual path goes through t.

The global models $\underline{P}(\cdot|t)$ can be calculated from the local $\underline{P}_s$ by backwards recursion, using the *Concatenation Formula* [4, 5]: for any given situation t, $\underline{P}(f|t) = \underline{P}_t(\underline{P}(f|d(t)))$, where $\underline{P}(f|d(t))$ is the gamble on $d(t)$ that assumes the value $\underline{P}(f|s)$ in each $s \in d(t)$; and for a terminal situation $\omega \in \Omega$, we have $\underline{P}(f|\omega) = f(\omega)$.

Example 1. Let us illustrate this with the successive flipping of two coins. In the corresponding event tree:

[2] Shafer [11, Chap. 12] uses the term 'martingale tree'.



the labels for the situations are explicit, e.g., $h,?$ means that the first coin has landed 'heads', and the second still has to be flipped. As indicated on the edges of the tree, the subject's beliefs about the first coin are modelled by the imprecise probability assignments $p(h) \in [1/4, 3/4]$ and $p(t) \in [1/4, 3/4]$. If it lands 'heads', we keep the same coin, otherwise the second flip is made with a fair coin ($p(h) = p(t) = 1/2$). We have also indicated the different steps in the calculation of the lower and upper probability of getting 'heads' at least once, using the Concatenation Formula.

5 Forward Irrelevance in Event Trees

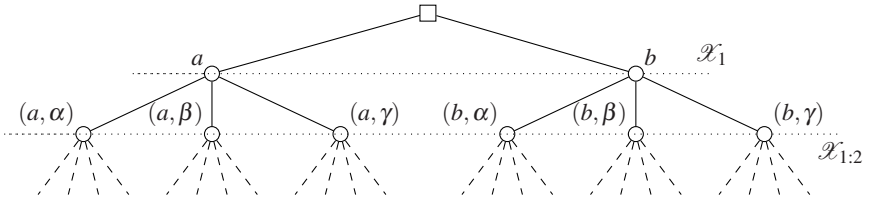
Let us briefly recall the notion of forward irrelevance, discussed in detail by De Cooman and Miranda [6], before relating it to independence in event trees.

For two random variables X_1 and X_2, if a subject says that X_1 is *epistemically irrelevant* to X_2, this means that he assesses that learning the actual value of X_1 won't change his beliefs about the value of X_2. For imprecise probability models, this notion is asymmetric: the epistemic irrelevance of X_1 to X_2 is not generally equivalent to the epistemic irrelevance of X_2 to X_1 [1, 6].

Assume that the uncertainty bears on random variables $X_1, \ldots, X_N$ that assume values in the respective finite sets $\mathscr{X}_1, \ldots, \mathscr{X}_N$. For $1 \leq k \leq \ell \leq N$, we denote by $\mathscr{X}_{\ell:k} := \times_{i=\ell}^{k} \mathscr{X}_i$ the Cartesian product of the $k-\ell+1$ sets $\mathscr{X}_\ell, \ldots, \mathscr{X}_k$, and by $X_{\ell:k} := (X_\ell, \ldots, X_k)$ the associated joint random variable taking values in $\mathscr{X}_{\ell:k}$. Similarly, $x_{\ell:k} := (x_\ell, \ldots, x_k) \in \mathscr{X}_{\ell:k}$ denotes a generic value of $X_{\ell:k}$. The random variables $X_1, \ldots, X_N$ are assumed to be *logically independent*, meaning that $X_{\ell:k}$ can assume all values in $\mathscr{X}_{\ell:k}$, for all $1 \leq \ell \leq k \leq N$. A gamble f defined on $\mathscr{X}_{1:N}$ is called $\mathscr{X}_{\ell:k}$-*measurable* if $f(x_{1:N}) = f(y_{1:N})$ for all $x_{1:N}$ and $y_{1:N}$ in $\mathscr{X}_{1:N}$ such that $x_{\ell:k} = y_{\ell:k}$. We denote by $\mathscr{L}(\mathscr{X}_{\ell:k})$ the set of all $\mathscr{X}_{\ell:k}$-*measurable* gambles, and by $f_{\ell:k}$ a generic gamble in this set. Of course, we identify the index '$k:k$' with 'k'.

An important problem is how to build joint belief models from partial ones. Let us consider the specific example where *the X_k constitute a stochastic process with time variable k*, implying in particular that the subject knows in advance that the value of random variable X_ℓ will be revealed to him before that of $X_{\ell+1}$, where $\ell = 1, 2, \ldots, N-1$. This leads to a special event tree (also called a *standard tree* [11, Chap. 2]) where the nodes s have the general form $x_{1:k} \in \mathscr{X}_{1:k}$, $k = 0, \ldots, N$. For $k = 0$ there is some abuse of notation, as we let $\mathscr{X}_{1:0} := \{\square\}$ and $x_{1:0} := \square$. The sets $\mathscr{X}_{1:k}$ constitute special cuts of the tree, where the value of X_k is revealed. We have $\mathscr{X}_{1:1} < \mathscr{X}_{1:2} < \cdots < \mathscr{X}_{1:N}$, and this sequence of cuts is also called a *standard filter* [11, Chap. 2]. It is clear that $d(x_{1:k}) = \{x_{1:k}\} \times \mathscr{X}_{k+1}$ for $k = 0, 1, \ldots, N-1$. The sample space of such a tree is $\Omega = \mathscr{X}_{1:N}$, and

with the variable X_k there corresponds a set $\mathscr{L}(\mathscr{X}_k)$ of $\mathscr{X}_k$-measurable gambles on this sample space. For instance, in the standard tree of Example 1, gambles characterising the second coin flip are such that $f(t,h) = f(h,h)$ and $f(t,t) = f(h,t)$. Below, we see the first two cuts of another standard tree, with $\mathscr{X}_1 = \{a,b\}$ and $\mathscr{X}_2 = \{\alpha,\beta,\gamma\}$.



A natural way to specify partial beliefs consists in attaching, as explained in the previous section, to each of the non-terminal nodes $x_{1:k}$ a (coherent) local predictive lower prevision $\underline{P}_{x_{1:k}}$ on $\mathscr{L}(d(x_{1:k}))$, i.e., on $\mathscr{L}(\mathscr{X}_{k+1})$, where $k = 0,1,\ldots,N-1$. This represents a subject's beliefs about the value of X_{k+1} given that the k previous variables $X_{1:k}$ assume the values $x_{1:k}$. For *standard imprecise probability trees*, the Concatenation Formula given above for deriving the global lower previsions $\underline{P}(\cdot|x_{1:\ell})$ on $\mathscr{L}(\mathscr{X}_{1:N})$ from the local models $\underline{P}_{x_{1:k}}$ completely coincides with the formulae for *Marginal Extension*, derived by Miranda and De Cooman [9].

A subject may make an assessment of *forward irrelevance*, meaning that for $1 \leq k \leq N-1$, his beliefs about the 'future' random variable X_{k+1} won't be changed by learning new information about the values of the 'past' random variables $X_{1:k}$: the past random variables $X_1,\ldots,X_k$ are *epistemically irrelevant* to the future random variable X_{k+1}, for $1 \leq k \leq N-1$. This is expressed by the following condition involving the local models: for all $0 \leq k \leq N-1$, any gamble f_{k+1} in $\mathscr{L}(\mathscr{X}_{k+1})$, and all $x_{1:k}$ in $\mathscr{X}_{1:k}$:

$$\underline{P}_{x_{1:k}}(f_{k+1}) = \underline{P}_{k+1}(f_{k+1}), \tag{1}$$

where $\underline{P}_{k+1}$ is the so-called *marginal* lower prevision on $\mathscr{L}(\mathscr{X}_{k+1})$, which expresses the subject's beliefs about the value of X_{k+1}, irrespective of the values assumed by the other random variables. Invoking the Concatenation Formula now leads to a very specific way of combining the marginal lower previsions $\underline{P}_1,\ldots,\underline{P}_N$ into a joint lower prevision, reflecting the assessment of forward irrelevance. This joint lower prevision, called the *forward irrelevant product*, is studied in detail by De Cooman and Miranda [6], who also use it to prove very general laws of large numbers [7].

We now proceed to show that *forward irrelevance* is exactly the same thing as Shafer's notion of event-tree independence, when applied to standard imprecise probability trees. In Shafer's [11] terminology, *a situation s influences a variable X* if there is at least one situation $t \in d(s)$ such that the subject's beliefs about the value of X are modified when moving from s to t; for imprecise probability trees, this means that there should be at least one gamble f whose value depends on the outcome of X for which $\underline{P}(f|s) \neq \underline{P}(f|t)$. Two variables X and Y are called *event-tree independent* if *there is no situation that influences both of them.*

In a standard imprecise probability tree, a situation $x_{1:k}$ influences a variable X_m if there is at least one situation $x_{1:k+1}$ in $d(x_{1:k})$ and at least one gamble f_m on $\mathscr{X}_m$ such that $\underline{P}(f_m|x_{1:k}) \neq \underline{P}(f_m|x_{1:k+1})$. The only situations $x_{1:k}$ that can influence X_m are such

that $k < m$, since in all other situations, the value of X_m has already been revealed 'for some time'. In addition, it is easy to check that X_m is always influenced by any situation $x_{1:m-1}$ in the cut $\mathscr{X}_{1:m-1}$ right before the value of X_m is revealed.

Theorem 1. *Let $X_1, \ldots, X_N$ be N random variables. Then there is forward irrelevance, or in other words, the random variables $X_{1:k}$ are epistemically irrelevant to X_{k+1} for $1 \leq k \leq N-1$ if and only if the random variables $X_1, \ldots, X_N$ are event-tree independent in the corresponding standard imprecise probability tree.*

Proof. We deal with the 'only if' part first. Suppose the random variables $X_{1:N}$ are forward irrelevant. Consider any X_k and $f_k \in \mathscr{L}(\mathscr{X}_k)$, where $1 \leq k \leq N$. Then it follows from the forward irrelevance condition (11) and the Concatenation Formula that $\underline{P}_k(f_k) = \underline{P}_{x_{1:k-1}}(f_k) = \underline{P}(f_k|x_{1:k-1})$ for all $x_{1:k-1}$ in $\mathscr{X}_{1:k-1}$. Applying the Concatenation Formula again leads to $\underline{P}(f_k|x_{1:k-2}) = \underline{P}_{x_{1:k-2}}(\underline{P}(f_k|x_{1:k-2},\cdot)) = \underline{P}_{x_{1:k-2}}(\underline{P}_k(f_k)) = \underline{P}_k(f_k)$, and if we continue the backwards recursion, we see that

$$\underline{P}_k(f_k) = \underline{P}(f_k|x_{1:k-1}) = \underline{P}(f_k|x_{1:k-2}) = \cdots = \underline{P}(f_k|x_{1:2}) = \underline{P}(f_k|x_1) = \underline{P}(f_k|\square).$$

This implies that the only situations that (may) influence X_k are the ones in the cut $\mathscr{X}_{1:k-1}$ immediately before X_k is revealed. Therefore, no situation can influence more than one variable, and there is event-tree independence.

Next, we turn to the 'if' part. Assume that all variables are event-tree independent in the standard tree. This implies that no variable X_k can be influenced by a situation $x_{1:\ell}$ corresponding to a time $\ell < k-1$ [If X_k were influenced by such a situation, then we know that this situation also always influences $X_{\ell+1}$, and $\ell+1 < k$, a contradiction]. So for all $x_{1:k-1} \in \mathscr{X}_{1:k-1}$ and all $f_k \in \mathscr{L}(\mathscr{X}_k)$:

$$\underline{P}(f_k|x_{1:k-1}) = \underline{P}(f_k|x_{1:k-2}) = \cdots = \underline{P}(f_k|x_{1:2}) = \underline{P}(f_k|x_1) = \underline{P}(f_k|\square).$$

Now of course $\underline{P}(f_k|\square) = \underline{P}(f_k) = \underline{P}_k(f_k)$, where $\underline{P}_k$ is the marginal lower prevision for X_k, and it follows from the Concatenation Formula that $\underline{P}(f_k|x_{1:k-1}) = \underline{P}_{x_{1:k-1}}(f_k)$. This shows that (11) is satisfied, so there is forward irrelevance. □

6 Conclusions

What is the message we want to convey in this paper? In the theory of coherent lower previsions [13], there are essentially two behavioural notions that generalise classical independence:[3] epistemic irrelevance and the derived notion of epistemic independence. Assessing that two random variables X_1 and X_2 are *epistemically independent* amounts to assessing that (i) X_1 is *epistemically irrelevant* to X_2, meaning that getting to know the value of X_1 doesn't change our subject's beliefs about X_2; and (ii) X_2 is epistemically irrelevant to X_1.

Suppose we want to consider a theory of uncertain processes where probabilities aren't necessarily precise. What will be the most useful or meaningful counterpart of the

[3] There are other generalisations, such as strong independence [1], but these have a sensitivity analysis interpretation, rather than a behavioural one; see also [13, Chap. 9]. Our comments below don't bear on such other types of independence.

important notion of independence in the classical theory of random processes? There are a number of reasons for preferring the asymmetric notion of epistemic irrelevance, and its generalisation to many variables, called *forward irrelevance*, to that of epistemic independence. We begin with arguments of perhaps less importance, and then go on to present the most compelling one.

First of all, when a notion that is (more or less) automatically symmetrical, breaks apart into two asymmetrical counterparts when using a more powerful language, symmetry becomes something that has to be justified: it can't be imposed without giving it another thought.

Secondly, an assessment of epistemic independence is stronger, and leads to higher joint lower previsions. As lower previsions represent supremum buying prices, higher values represent stronger commitments, and these may be unwarranted when it is only epistemic irrelevance that our subject really wants to model.

Thirdly, joint lower previsions based on an epistemic irrelevance assessment are generally speaking straightforward to calculate, as the discussion of the Concatenation Formula in Sect. 5 testifies. But calculating joint lower previsions from marginals based on an epistemic independence assessment is quite often a very complicated affair [13, Sect. 9.3.2].

Finally, and most importantly, when considering an uncertain process, the subject knows that the values of the random variables X_k will be revealed one after the other, and that the value of X_k *will be revealed before* that of X_{k+1}. If he states that X_k and X_{k+1} are epistemically independent, this amounts to his assessing that (i) getting to know the value of X_k won't change his beliefs about X_{k+1} [forward irrelevance]; and (ii) getting to know the value of X_{k+1} won't change his beliefs about X_k [backward irrelevance]. But since the subject knows that he will always know the value of X_k before that of X_{k+1}, (ii) is effectively a counter-factual statement for him: "if I got to the value of X_{k+1} first, then learning that value wouldn't affect my beliefs about X_k". It's not clear that making such an assessment has any real value, and we feel it is much more natural in such situations context to let go of (ii) and therefore to resort to epistemic (forward) irrelevance.

This line of reasoning can also be related to Shafer's [10] idea that conditioning is never automatic, and must always be associated with a *protocol*. A subject can only *meaningfully* condition a probability model on events that he envisages may happen (according to the established protocol). In the specific situation described above, conditioning the belief model about X_k on the variable X_{k+1} could only legitimately be done if it were possible to find out the value of X_{k+1} without getting to know that of X_k, *quod non*. Therefore, it isn't legitimate to consider the conditional lower prevision $\underline{P}_k(\cdot|X_{k+1})$ expressing the beliefs about X_k conditional on X_{k+1}, and we therefore can't meaningfully impose (ii), as it requires that $\underline{P}_k(\cdot|X_{k+1}) = \underline{P}_k$. Again, this leads to epistemic (forward) irrelevance, instead of epistemic independence.

In his book on causal reasoning [11], Shafer seems to propose the notion of an event tree in order to develop and formalise his ideas about protocols and conditioning. We have seen in Theorem 1 that for standard event trees, which correspond to uncertain processes, the general notion of *event-tree independence* that he develops in his book, is effectively equivalent to the notion of forward irrelevance.

References

1. Couso, I., Moral, S., Walley, P.: Examples of independence for imprecise probabilities. Risk Decis Policy 5, 165–181 (2000)
2. Cozman, F.G.: Credal networks. Artificial Intelligence 120, 199–233 (2000)
3. Cozman, F.G.: Graphical models for imprecise probabilities. Internat. J. Approx. Reason 39(2-3), 167–184 (2005)
4. De Cooman, G., Hermans, F.: On coherent immediate prediction: Connecting two theories of imprecise probability. In: De Cooman, G., Vejnarova, J., Zaffalon, M. (eds.) Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2007, Prague, Czech Republic), SIPTA, pp. 107–116 (2007)
5. De Cooman, G., Hermans, F.: Imprecise probability trees: Bridging two theories of imprecise probability. Artificial Intelligence (in press, 2008) doi:10.1016/j.artint.2008.03.001
6. De Cooman, G., Miranda, E.: Forward irrelevance. J. Statist. Plan Infer. (in press, 2008a) doi:10.1016/j.jspi.2008.01.012
7. De Cooman, G., Miranda, E.: Weak and strong laws of large numbers for coherent lower previsions. J. Statist. Plan Infer. (in press, 2008b)doi:10.1016/j.jspi.2007.10.020
8. Levi, I.: The Enterprise of Knowledge. MIT Press, London (1980)
9. Miranda, E., De Cooman, G.: Marginal extension in the theory of coherent lower previsions. Internat. J. Approx. Reason 46(1), 188–225 (2007)
10. Shafer, G.: Conditional probability. Internat. Statist. Rev. 53, 261–277 (1985)
11. Shafer, G.: The Art of Causal Conjecture. MIT Press, Cambridge (1996)
12. Shafer, G., Vovk, V.: Probability and Finance: It's Only a Game! Wiley, New York (2001)
13. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Probability Revision, the Uniformity Rule, and the Chan–Darwiche Metric

Carl G. Wagner

Department of Mathematics, The University of Tennessee, Knoxville, USA

Abstract. The author has proposed a rule of probability revision dictating that identical learning be reflected in identical ratios of new to old odds. Following this rule ensures that the final result of a sequence of probability revisions is undisturbed by an alteration in the temporal order of the learning prompting these revisions. There is also a close connection between this rule and an intriguing metric on probability measures introduced by Chan and Darwiche.

Keywords: Bayes factor, Chan-Darwiche metric, Probability revision.

1 The Commutativity Principle

Consider the following belief revision schema, representing two possible sequential revisions of the probability measure p:

$$p \longrightarrow q \longrightarrow r \qquad \text{and} \qquad p \longrightarrow s \longrightarrow t.$$

Suppose that the revisions of p to q, and of s to t, are prompted by identical learning, and that the revisions of q to r, and of p to s, are prompted by identical learning. It is then widely held that it ought to be the case that $r = t$. As van Fraassen [7] puts it, two persons who undergo identical learning experiences on the same day, but in a different order, ought to agree in the evening if they had exactly the same opinions in the morning. Call this the *Commutativity Principle.*

A simple rule of probability revision ensures that the Commutativity Principle is satisfied. This *Uniformity Rule,* occurring in particular cases in Wagner [8, 9, 10, 11], and given general formulation in Wagner [12], dictates that identical learning be reflected in identical ratios of new to old odds, also known as *Bayes factors.* This note explores the connection between the Uniformity Rule and an intriguing metric on probability measures introduced by Chan and Darwiche [1]. The upshot is that revisions of two different probability measures based on identical learning, when effected by the Uniformity Rule, move us the same Chan-Darwiche distance from the priors in question.

2 Terminology and Notation

A sigma algebra $\mathbf{A}$ of subsets of Ω is *purely atomic* if the family $\mathbf{A}^*$ of atomic events in $\mathbf{A}$ is countable, and constitutes a partition of Ω. Every finite algebra is purely atomic, whatever the cardinality of Ω, and if Ω is countable, then every sigma algebra on Ω is

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 74–79, 2008.
springerlink.com

purely atomic ([6, Theorems 1.6.1, 1.6.2]). If q is a revision of probability measure p, and A and B are events, then the *probability factor* (or *relevance quotient*) $\Pi_{q,p}(A)$ is the ratio

$$\Pi_{q,p}(A) := \frac{q(A)}{p(A)}$$

of new to old probabilities, and the *Bayes factor* $\beta_{q,p}(A:B)$ is to ratio

$$\beta_{q,p}(A:B) := \frac{\frac{q(A)}{q(B)}}{\frac{p(A)}{p(B)}} \tag{1}$$

of new to old odds. When $q(.) = p(.|E)$, then (1) is simply the likelihood ratio $\dfrac{p(E|A)}{p(E|B)}$. More generally,

$$\beta_{q,p}(A:B) = \frac{\Pi_{q,p}(A)}{\Pi_{q,p}(B)}, \tag{2}$$

a simple, but useful, identity.

In what follows we assume for simplicity that all probability measures are *strictly coherent,* i.e., that all nonempty events have positive probability. With the addition of certain technical conditions, however, Theorem 1 below holds for arbitrary probabilities.

3 Bayes Factors and Commutativity

The following theorem demonstrates that the Commutativity Principle is satisfied for purely atomic sigma algebras when identical learning is represented by identical Bayes factors at the level of atomic events.

Theorem 1. *Suppose that the probabilities in the revision schema*

$$\begin{array}{ccc} p & \longrightarrow & q \\ & & \downarrow \\ \downarrow & & r \\ \\ s & \longrightarrow & t \end{array}$$

are defined on a purely atomic sigma algebra $\mathbf{A}$*, with* $\mathbf{A}^*$ *denoting the set of atomic events in* $\mathbf{A}$*. If the Bayes factor identities*

$$\beta_{q,p}(A:B) = \beta_{t,s}(A:B), \quad \textit{for all } A,B \in \mathbf{A}^*, \tag{3}$$

and

$$\beta_{r,q}(A:B) = \beta_{s,p}(A:B), \quad \textit{for all } A,B \in \mathbf{A}^* \tag{4}$$

hold, then $r = t$*. Indeed, for all* $A \in \mathbf{A}^*$*, we have the explicit formula*

$$r(A) = t(A) = \frac{\left[\frac{q(A)s(A)}{p(A)}\right]}{\sum_{B\in\mathbf{A}^*}\frac{q(B)s(B)}{p(B)}} \tag{5}$$

Proof. The identity (3) is equivalent to

$$\frac{t(A)q(B)s(B)}{p(B)} = \frac{q(A)s(A)t(B)}{p(A)}, \text{ for all } A,B \in \mathbf{A}^*. \tag{6}$$

Fixing A in (6), and summing over all $B \in \mathbf{A}^*$ then yields (5) for $t(A)$, since

$$\sum_{B\in\mathbf{A}^*} t(B) = 1.$$

The proof of (5) for $r(A)$ follows from (4) in exactly analogous fashion. □

Remark 1. If p, q, r, s and t are well-defined and in place and (3) and (4) hold, then, necessarily, the sum in the denominator of the right-hand side of (5) converges. If only p, q, and s are in place at the outset and the aforementioned sum converges, then (5) defines probabilities r and t satisfying (3) and (4). So (5) furnishes a recipe for constructing a probability measure r that would be the appropriate revision of q if, in the probabilistic state q, one were to undergo learning identical to that which prompted the revision of p to s. Similarly, (5) furnishes a recipe for constructing a probability measure t that would be the appropriate revision of s if, in the probabilistic state s, one were to undergo learning identical to that which prompted the revision of p to q. However, it is easy to construct examples where the sum in the denominator of (5) fails to converge. Then there exists no probability measure t satisfying (3) and no probability r satisfying (4). Thus from the perspective of the Uniformity Rule, it is impossible in the conceptual state reflected in s (respectively, q) to experience learning identical to that which prompted the revision of p to q (respectively, of p to s).

4 The Chan-Darwiche Metric

When Ω is finite the Uniformity Rule has intriguing connections with a metric on probability measures introduced by Chan and Darwiche [1]. Assume for simplicity that all probabilities are strictly coherent[1], and defined on all subsets of Ω. Define the *Chan-Darwiche distance* $CD(p,q)$ by

$$CD(p,q) := \log(R) - \log(r), \tag{7}$$

where

$$R := \max_{\omega\in\Omega}\frac{q(\omega)}{p(\omega)} \quad \text{and} \quad r := \min_{\omega\in\Omega}\frac{q(\omega)}{p(\omega)}. \tag{8}$$

[1] On the set all probability measures on the power set of Ω, CD is, strictly speaking, no longer a metric, since it can take the extended real number ∞ as a value. Indeed, with the stipulation that $\frac{0}{0} = 1$, $CD(p,q) < \infty$ iff p and q have exactly the same support, i.e., iff $\{\omega \in \Omega : p(\omega) > 0\} = \{\omega \in \Omega : q(\omega) > 0\}$.

It is straightforward to show that CD is a *metric* on the set of all strictly coherent probability measure on the power set of Ω, i.e., that

$$CD(p,q) \geqslant 0, \quad \text{with} \quad CD(p,q) = 0 \quad \text{iff} \quad p = q.$$

$$CD(p,q) = CD(q,p), \text{ and}$$

$$CD(p,q) \leqslant CD(p,p') + CD(p',q).$$

$CD(p,q)$ yields uniform bounds on the Bayes factors $\beta_{q,p}(A:B)$:

Theorem 2. *For all nonempty events* $A, B \in 2^{\Omega}$,

$$\exp(-CD(p,q)) \leqslant \beta_{q,p}(A,B) \leqslant \exp(CD(p,q)). \tag{9}$$

Proof. Suppose that $\max \frac{q(\omega)}{p(\omega)}$ and $\min \frac{q(\omega)}{p(\omega)}$ are attained, respectively, at $\omega = \omega_2$ and $\omega = \omega_1$. Then

$$\frac{q(\omega_1)p(\omega)}{p(\omega_1)} \leqslant q(\omega) \leqslant \frac{q(\omega_2)p(\omega)}{p(\omega_2)}. \tag{10}$$

Summing (10) over all $\omega \in A$, and over all $\omega \in B$ yields

$$\frac{q(\omega_1)}{p(\omega_1)} \leqslant \frac{q(A)}{p(A)}, \qquad \frac{q(B)}{p(B)} \leqslant \frac{q(\omega_2)}{p(\omega_2)}$$

whence,

$$\frac{\left[\frac{q(\omega_1)}{p(\omega_1)}\right]}{\left[\frac{q(\omega_2)}{p(\omega_2)}\right]} \leqslant \frac{\Pi_{q,p}(A)}{\Pi_{q,p}(B)} \leqslant \frac{\left[\frac{q(\omega_2)}{p(\omega_2)}\right]}{\left[\frac{q(\omega_1)}{p(\omega_1)}\right]} \tag{11}$$

which is equivalent to (9) by (2) of Section 2 above, (7), and (8). □

Remark 2. Note that the bounds in (9) are sharp, the upper bound being attained when $A = \{\omega_2\}$ and $B = \{\omega_1\}$, and the lower bound when $A = \{\omega_1\}$ and $B = \{\omega_2\}$.

In view of (11) and the preceding remark, it is clear that $CD(p,q)$ may be equivalently defined by the formulas

$$CD(p,q) = \max_{\phi \neq A,B \subset \Omega} \log \beta_{q,p}(A:B) = \max_{\omega,\omega' \in \Omega} \log \beta_{q,p}(\{w\}:\{w'\}).^{2} \tag{12}$$

The number $\log \beta_{q,p}(A:B)$ has been termed the *weight of evidence* by I.J. Good [3]. According to Good, Alan Turing was an enthusiastic advocate of using weights of evidence to measure the gain or loss of plausibility of one hypothesis vis-á-vis another as a result of the receipt of new evidence. Such weights were routinely used in the code-breaking work at Bletchley Park, where Good and Turing were colleagues during World War II ([5]). Indeed, Turing coined the term *ban* (after the town of Banbury, where the

[2] Upon reading Chan-Darwiche [1], I communicated this result to the authors, who incorporated it in Chan-Darwiche [2].

sheets were printed on which weights of evidence were recorded) for the unit weight of evidence, with logarithms taken to the base 10. One-tenth of a ban was termed a *deciban* (abbreviated *db*, in obvious analogy with acoustic notation). See Jeffrey ([5, pp. 32–32]) and Good [4] for further details.

Formula (12) thus provides a particularly salient formulation of the Chan-Darwiche distance, as well as an attractive and evocative unit of measurement. Moreover, there is a hand-in-glove fit between the Uniformity Rule and the Chan-Darwiche distance: If p is revised to q, and p' is revised to q', based on identical learning, and we construct q' in accord with the Uniformity Rule, then $CD(p,q) = CD(p',q')$. *So revisions based on identical learning, carried out acording to the dictates of the Uniformity Rule, move us the same CD-distance (i.e., the same number of decibans) from the priors in question.* As can be seen from the elementary example,

$$\begin{array}{cccc} & \omega_1 & \omega_2 & \\ p: & \frac{1}{2} & \frac{1}{2} & \\ q: & \frac{4}{5} & \frac{1}{5} & \end{array} \qquad \begin{array}{cccc} & \omega_1 & \omega_2 & \\ p': & \frac{2}{5} & \frac{3}{5} & \\ q': & \frac{8}{11} & \frac{3}{11}, & \end{array}$$

where $CD(p,q) = CD(p',q') = 2\log 2$, this fails to be the case for other measures of distance, including the *Euclidean distance*

$$ED(p,q) := \left[\sum_{\omega}(p(\omega) - q(\omega))^2\right]^{\frac{1}{2}},$$

the *variation distance*

$$\begin{aligned} V(p,q) : &= \max\{|p(A) - q(A)| : A \subset \Omega\} \\ &= \frac{1}{2}\sum_{\omega}|p(\omega) - q(\omega)|, \end{aligned}$$

the *Hellinger distance*

$$H(p,q) := \sum_{\omega}\left[\sqrt{p(\omega)} - \sqrt{q(\omega)}\right]^2,$$

and the *Kullback-Leibler information number*

$$KL(p,q) := \sum_{\omega} q(\omega)\log\left(\frac{q(\omega)}{p(\omega)}\right).$$

References

1. Chan, H., Darwiche, A.: A distance measure for bounding probabilistic belief change. In: Proceedings of 18th International Conference on Artificial Intelligence (AAAI-2002, Menlo Park, CA), pp. 539–545. AAAI Press, Menlo Park (2002)
2. Chan, H., Darwiche, A.: A distance measure for bounding probabilistic belief change. Int. J. Approx. Reas. 38, 149–174 (2004)
3. Good, I.J.: Probability and the Weighing of Evidence. Charles Griffin, London (1950)

4. Good Good, I.J.: Studies in the history of probability and statistics. XXXVII. A. M. Turing's statistical work in World War II. Biometrika 66, 393–396 (1979)
5. Jeffrey, R.C.: Subjective Probability: The Real Thing. Cambridge University Press, Cambridge (2004)
6. Rényi, A.: Foundations of Probability. Holden-Day, San Francisco (1970)
7. Van Fraassen, B.C.: Laws and Symmetry. Clarendon Press, Oxford (1989)
8. Wagner, C.G.: Old evidence and new explanation. Philos. Sci. 64, 677–691 (1997)
9. Wagner, C.G.: Old evidence and new explanation II. Philos. Sci. 66, 283–288 (1999)
10. Wagner, C.G.: Old evidence and new explanation III. Philos. Sci. 68(3) Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association. Part I: Contributed Papers, pp. S165–S175 (2001)
11. Wagner, C.G.: Probability kinematics and commutativity. Philos. Sci. 69, 266–278 (2002)
12. Wagner, C.G.: Commuting probability revisions: the uniformity rule. Erkenntnis 59, 349–364 (2003)

Part III

Statistical Methods

Statistical Inference

On Nonparametric Predictive Inference for Bernoulli Quantities with Set-Valued Data

Frank P.A. Coolen

Dept. of Mathematical Sciences, Durham University, Durham, United Kingdom

Abstract. Coolen [3] introduced lower and upper probabilities for m future Bernoulli random quantities, based on the number of successes in n trials and adding few structural assumptions. These results form part of the statistical approach called 'Nonparametric Predictive Inference'. In this paper, we explore the generalization of these results for the case with data only available in the form of a set of values for the number of successes in the first n trials. A special case of such inferences occurs in applications to basic acceptance sampling problems in quality control.

1 Introduction

Statistical inference in situations with incomplete data has received much attention in the literature, reflecting its importance in many applications. Manski [12] presents a wide range of methods to partially identify probability distributions based on incomplete data, considering a wide range of reasons for data to be incomplete and presenting historical notes and further references. The manner in which inferential methods based on different foundations deal with incomplete data differs substantially, in particular when methods using precise probabilities are compared to methods in which uncertainty is quantified via lower and upper probabilities. For example, if some data are not reported precisely but only to belong in a particular range of values, precise Bayesian methods simply include the probability for all these possible data values in the likelihood function used in the updating calculations to derive the posterior distribution. Methods that allow imprecision to be taken into account, by the use of lower and upper probabilities, can deal with such data differently, namely by considering best- and worst-case data possibilities, within the range of values reported, and as such no further assumptions on values within this range need to be included in the inferences.

De Cooman and Zaffalon [8] present a detailed theory for updating imprecise probabilities based on incomplete (set-valued) data, within the framework of coherent imprecise probability presented by Walley [13]. Related to this work, Zaffalon [16] focuses specifically on predictive inference. Weichselberger [15] has been developing a novel interval-probabilistic statistical approach, called 'symmetrical probability', for which results in case of set-valued data have also been presented.

Coolen [3] presented an inferential approach for Bernoulli quantities, which uses lower and upper probabilities and differs from the generalized Bayes approach by Walley [13]. This approach, which fits in the more general framework of 'nonparametric predictive inference' (NPI) [1, 4], takes a fundamentally different view to learning

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 85–91, 2008.
springerlink.com

from data than generalized Bayes methods, as it only considers events of the form A is followed by B, without any assumptions on the single events A or B. An example of an application of this approach is comparison of proportions data for different groups [5, 6]. NPI for Bernoulli quantities [3] has, so far, been presented for events of the form $Y_1^n = s$ is followed by $Y_{n+1}^{n+m} \in R$, with Y_i^j the number of successes in trials i to j, and $R \subset \{0,1,\ldots,m\}$. So, it was assumed that the number of successes in the first n trials observed is known to be equal to s. This raises the question of how NPI for Bernoulli quantities deals with set-valued information, which occurs if the number of successes in the first n trials is not precisely known, but if only $Y_1^n \in S$ is known, with $S \subset \{0,1,\ldots,n\}$. This is explored in the current paper, general theory including a discussion of principles of such inference will be presented elsewhere.

NPI is different in nature to Walley's generalized Bayes approach, with a fundamentally different view on the manner in which data are used. In Walley's theory, in line with precise Bayesian methods, one updates prior lower and upper probabilities, using a generalized Bayes' rule, to derive posterior lower and upper probabilities, so learning from data is modelled via conditioning based on an all-encompassing parametric model and prior distributions. In NPI there are no non-trivial statements (i.e. non-vacuous lower and upper probabilities for non-trivial events) for the single events involving only Y_1^n or Y_{n+1}^{n+m}. As such, when generalizing to set-valued data of the form $Y_1^n \in S$, inference is not based on conditioning on this event, and the precise meaning of an event '$Y_1^n \in S$ is followed by $Y_{n+1}^{n+m} \in R$' must be considered with great care. It turns out that, for such events with set-valued data, the difference between NPI and other approaches becomes clearer than for precise data.

Section 2 of this paper is a brief summary of the NPI approach for Bernoulli quantities [3], with some details on the derivations of the main results as needed to generalize the method to set-valued data, Section 3 discusses key aspects of this generalization. In Section 4, we briefly consider a special case, with data $Y_1^n \geq y$, which is relevant for acceptance sampling. The paper ends with some concluding remarks in Section 5.

2 Nonparametric Predictive Inference for Bernoulli Random Quantities

Nonparametric prediction of Bernoulli random quantities uses Hill's assumption $A_{(n)}$ [10], and defines direct predictive lower and upper probabilities for future observations, based on available data. This fits in the framework of nonparametric predictive inference (NPI) with strong internal consistency and frequentist properties [1, 4].

Suppose that we have a sequence of $n+m$ exchangeable Bernoulli trials [9], each with success and failure as possible outcomes, and data consisting of s successes observed in n trials. A sufficient representation of the data for our inferences is $Y_1^n = s$, due to the exchangeability of all trials. We are interested in the number of successes in trials $n+1$ to $n+m$. Let $R = \{r_1,\ldots,r_t\}$, with $1 \leq t \leq m+1$ and $0 \leq r_1 < r_2 < \ldots < r_t \leq m$, and let $\binom{s+r_0}{s} = 0$. The NPI upper probability for the event $Y_{n+1}^{n+m} \in R$, given data $Y_1^n = s$, for $s \in \{0,\ldots,n\}$, is [3]

$$\overline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n = s) =$$
$$\binom{n+m}{n}^{-1} \sum_{j=1}^{t} \left[\binom{s+r_j}{s} - \binom{s+r_{j-1}}{s} \right] \binom{n-s+m-r_j}{n-s}$$

The lower probability can be derived via the conjugacy property,

$$\underline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n = s) = 1 - \overline{P}(Y_{n+1}^{n+m} \in \{0,1,\ldots,m\} \backslash R \mid Y_1^n = s)$$

This is justified in [3], and agrees with the fact that these lower and upper probabilities are F-probability in the theory of interval probability [1, 14]. The key aspects of this theory are explained below.

In NPI for Bernoulli random quantities [3], past observations are related to future random quantities via an assumed underlying latent variable representation, such that each value is represented by a point on the real line, with a threshold such that all points to one side of the threshold represent 'successes', and all points to the other side of the threshold represent 'failures'. No knowledge about this threshold is assumed. This representation is very similar to that used by Bayes [2], with the exception that Bayes made explicit assumptions on the threshold, which in the later development of Bayesian statistical methodology corresponded to the assumption of a prior distribution. In NPI, with the latent variable representation, past observations are related to future observations via Hill's $A_{(n)}$ [10]. Suppose that the ordered values of the latent variables corresponding to the n observations are $u_{(1)} < u_{(2)} < \ldots < u_{(n)}$. These n values define a partition of the real line, consisting of $n+1$ intervals. Hill's $A_{(n)}$ states that a future random quantity U_{n+1} has equal probability $1/(n+1)$ to be in each of these intervals, in our NPI setting this U_{n+1} is the latent variable corresponding to the first future observation, which will again be a success or failure, depending on which side of the threshold U_{n+1} is.

When interested in m future observations, the same assumption needs to be made for each future observation consecutively, so one needs to assume $A_{(n)}, \ldots, A_{(n+m-1)}$. In fact, assuming $A_{(n+m-1)}$ is sufficient, as Hill [10] shows that the assumption $A_{(n)}$ implies $A_{(k)}$ for all $k \leq n$. Under these assumptions, the following result holds [3, 10]. Suppose that there is no interest in precisely which of the first n observations are successes or failures, so that one considers the number of successes as a sufficient statistic, and the same is assumed for the m future observations of interest. Then, under the assumption $A_{(n+m-1)}$, all $\binom{n+m}{n}$ different orderings of the underlying latent variables on the real line, which represent the first n observations and the m future observations, have the same probability, also after information about the number of successes in the first n observations has become available. Denoting these $\binom{n+m}{n}$ different orderings by O_j for $j = 1, \ldots, \binom{n+m}{n}$, the above lower and upper probabilities are derived by counting orderings [3]: for the lower probability, only those orderings are included for which $Y_1^n = s$ *must* be followed by $Y_{n+1}^{n+m} \in R$, while for the upper probability all orderings are included for which $Y_1^n = s$ *can* be followed by $Y_{n+1}^{n+m} \in R$.

3 Set-Valued Data

We now explore the generalization of NPI for Bernoulli random quantities with set-valued data, so we assume that the information on the first n trials is only $Y_1^n \in S$, with

$S \subset \{0,1,\ldots,n\}$. For completeness of the NPI theory, it is important to derive general expressions for the lower and upper probabilities for the events that $Y_1^n \in S$ is followed by $Y_{n+1}^{n+m} \in R$, for any sets S,R (we assume throughout that these sets are not empty). The combinatorial problems involved in deriving general formulae have not yet been solved, we hope to present these elsewhere in the near future, together with a more detailed discussion of this theory and further comparison with other approaches. We now explore this generalization by discussing the derivation of such lower and upper probabilities and some of their properties, and by a basic example.

We consider again the $\binom{n+m}{n}$ different orderings O_j of the n latent variables representing the n observations on the real line together with the m latent variables representing the m future observations, all these orderings remain equally likely under the assumption $A_{(n+m-1)}$. With these set-valued data, the reasoning that leads to the lower and upper probabilities for the event $(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S)$ remains the same as discussed Section 2 [3]. The lower probability for this event is derived by counting all orderings O_j for which $Y_1^n \in S$ *must* be followed by $Y_{n+1}^{n+m} \in R$, while the upper probability is derived by counting all orderings O_j for which $Y_1^n \in S$ *can* be followed by $Y_{n+1}^{n+m} \in R$. It is important to emphasize that, for the lower probability, an O_j is only included in the count if for *each* $s \in S$, $Y_1^n = s$ *must* be followed by $Y_{n+1}^{n+m} \in R$, whereas for the upper probability an O_j is already included if there is *at least one* $s \in S$ for which $Y_1^n = s$ *can* be followed by $Y_{n+1}^{n+m} \in R$. Hence, the actual events that correspond to the lower and upper probabilities for $(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S)$ differ substantially, in a way that could be described as 'most conservative', and which also ensures that the conjugacy property remains valid, so

$$\underline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S) = 1 - \overline{P}(Y_{n+1}^{n+m} \in \{0,1,\ldots,m\} \backslash R \mid Y_1^n \in S)$$

Basic logic and set theory imply some important general properties for these lower and upper probabilities. Let $S_1 \subset S_2$, then for all R,

$$\underline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S_1) \geq \underline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S_2)$$
$$\overline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S_1) \leq \overline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S_2)$$

so one could say that this NPI lower (upper) probability decreases (increases) in S. For all R that are strict subsets of $\{0,1,\ldots,m\}$, we have $\underline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in \{0,1,\ldots,n\}) = 0$ and $\overline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in \{0,1,\ldots,n\}) = 1$, reflecting that with $S = \{0,1,\ldots,n\}$ no information is provided about the number of successes in the first n trials. In precise Bayesian statistics, this situation results in just the prior probability for the event $Y_{n+1}^{n+m} \in R$, which also implies that nothing has been learned to update this prior probability, but of course does require one to have such a prior probability in the first place. Obviously, for any given set S, these lower and upper probabilities are increasing in R. As part of our detailed study of NPI with set-valued data, we will consider more such properties of these lower and upper probabilities as functions of S and R.

An important issue for uncertainty quantification is the interpretation of (lower and upper) probabilities. As mentioned before, the lower probability $\underline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S)$ and upper probability $\overline{P}(Y_{n+1}^{n+m} \in R \mid Y_1^n \in S)$ are conservative, as is clear from the way they are derived. They can be used without further assumptions about the specific underlying reasons for reporting S instead of a specific unique value for Y_1^n, which is in

Table 1. NPI lower and upper probabilities, $n = 4$ and $m = 2$

$\times 1/15$	$R = \{0\}$	$\{1\}$	$\{2\}$	$\{0,1\}$	$\{0,2\}$	$\{1,2\}$
$S = \{0\}$	$(10,15)$	$(0,5)$	$(0,1)$	$(1,15)$	$(10,15)$	$(0,5)$
$\{1\}$	$(6,10)$	$(3,8)$	$(1,3)$	$(12,14)$	$(7,12)$	$(5,9)$
$\{2\}$	$(3,6)$	$(4,9)$	$(3,6)$	$(9,12)$	$(6,11)$	$(9,12)$
$\{3\}$	$(1,3)$	$(3,8)$	$(6,10)$	$(5,9)$	$(7,12)$	$(12,14)$
$\{4\}$	$(0,1)$	$(0,5)$	$(10,15)$	$(0,5)$	$(10,15)$	$(14,15)$
$\{1,2\}$	$(3,10)$	$(2,11)$	$(1,6)$	$(9,14)$	$(4,13)$	$(5,12)$
$\{1,3\}$	$(1,10)$	$(1,12)$	$(1,10)$	$(5,14)$	$(3,14)$	$(5,14)$
$\{2,3\}$	$(1,6)$	$(2,11)$	$(3,10)$	$(5,12)$	$(4,13)$	$(9,14)$
$\{1,2,3\}$	$(1,10)$	$(1,13)$	$(1,10)$	$(5,14)$	$(2,14)$	$(5,14)$
$\{0,1,2,3\}$	$(1,15)$	$(0,14)$	$(0,10)$	$(5,15)$	$(1,15)$	$(0,14)$
$\{1,2,3,4\}$	$(0,10)$	$(0,14)$	$(1,15)$	$(0,14)$	$(1,15)$	$(5,15)$
$\{0,1,3,4\}$	$(0,15)$	$(0,14)$	$(0,15)$	$(0,15)$	$(1,15)$	$(0,15)$

line with alternative approaches for dealing with set-valued data in imprecise probability theory, but which cannot be achieved with precise probabilities. We present a basic example to illustrate these NPI lower and upper probabilities, and discuss some more features.

Example 1. To illustrate NPI for Bernoulli quantities with set-valued data, we consider $n = 4$ available observations and $m = 2$ future observations. The underlying assumed data representation has $\binom{6}{2} = 15$ different orderings of past and future observations, each having probability $1/15$ under the inferential assumption in this paper. Table 1 gives lower and upper probabilities for a variety of events and set-valued data, represented as pairs $(15\underline{P}(Y_5^6 \in R \mid Y_1^4 \in S), 15\overline{P}(Y_5^6 \in R \mid Y_1^4 \in S))$, so as indicated the values of the lower and upper probabilities are those given in the table multiplied by $1/15$.

The final case, with $S = \{0,1,3,4\}$, only leads to non-vacuous NPI lower and upper probabilities for two sets R (due to conjugacy), which is due only to the specific ordering of the 4 past and 2 future observations in which the latter two are in between the second and third ordered past observation, in the underlying assumed data representation. Clearly, for that specific ordering none of the values in this S can be followed by precisely one future success, for all other orderings this is possible for at least one of the values in S. These results illustrate clearly the decreasing (increasing) nature of the lower (upper) probabilities if S becomes larger. Imprecision is pretty large, which is due to the specific manner in which the lower and upper probabilities are derived, and their conservative nature, yet it should not be too surprising. For example, if one gets information that, out of 4 trials, the number of successes was either 1 or 3, clearly this information does not reveal much, in particular if one has no idea why this specific information was given. Most remarkable, perhaps, are some of these inferences for $R = \{1\}$. For example, when one compares the values corresponding to $S = \{1\}$, $S = \{3\}$ and $S = \{1,3\}$, one might perhaps be surprised that, for the latter case, the lower and upper probabilities are not also equal to $3/15$ and $8/15$, respectively. This is an important feature of the NPI approach, where it differs fundamentally from other approaches, including imprecise probabilistic approaches which are Walley-coherent

[8, 13, 16] and Weichselberger's symmetrical probability [15]. Such behaviour of these NPI lower and upper probabilities as function of S are currently being studied in more detail, and will be reported on in a future paper.

4 Acceptance Sampling

A special case of the theory in Section 3, for which the NPI lower and upper probabilities are easily derived, occurs in basic problems of quality control, when decisions are required about acceptance of a batch of products on the basis of tests on a sample. Some results are briefly discussed in this section, a detailed account will be presented elsewhere [7].

For acceptance sampling, one is often interested in the event $(Y_{n+1}^{n+m} \geq r \mid Y_1^n \geq s)$, for example if one can test n products and has to set a minimum number of successful tests for these, if one wishes to have at least r successful products in the related batch of m further products. Using the reasoning with the orderings O_j as in Section 3, it follows easily that for NPI for Bernoulli random quantities, the following relations hold,

$$\begin{aligned} \underline{P}(Y_{n+1}^{n+m} \geq r \mid Y_1^n \geq s) &= \underline{P}(Y_{n+1}^{n+m} \geq r \mid Y_1^n = s) \\ &= \binom{n+m}{m}^{-1} \sum_{j=r}^{m} \binom{s-1+j}{j} \binom{n-s+m-j}{m-j} \end{aligned}$$

and

$$\overline{P}(Y_{n+1}^{n+m} \geq r \mid Y_1^n \geq s) = \overline{P}(Y_{n+1}^{n+m} \geq r \mid Y_1^n = n) = 1$$

These results clearly indicate the conservativeness of the NPI lower and upper probabilities with set-valued data, but the example in Section 3 made clear that similar reductions of the observation set to a single extreme value does not generally work.

One particularly nice result for such NPI-based acceptance sampling [7] occurs when non-destructive sampling is considered. This means that products can be tested without affecting their functionality for future use, so in our setting a total batch would consist of $t = n + m$ products of which n were to be tested. Suppose that one requires that all products in the batch must function with NPI lower probability p, then it turns out that the minimum required number of products, out of these t, that have to be tested (and of course must all function) should at least be pt. For example, for a total batch of size $t = 100$, one would have to test at least $100p$ products, none of which should fail, to achieve NPI lower probability of p for the event that all 100 products will function successfully. No matter one's judgement about the NPI approach, one cannot argue against the beauty of this result.

5 Concluding Remarks

Statistical inferential methods that fully utilize the richness of opportunities provided by lower and upper probabilities are still in their infancy when compared to precise probabilistic methods. Although robust Bayes-like methods have become popular, following

Walley [13], they do not appear to fully utilize these opportunities, in particular as updating remains a form of conditioning, hence it is assumed that all that can ever happen is taken into account in the model and assessments at the prior stage. NPI provides an interesting alternative, that has several advantages, for example its general agreement with empirical probabilities, its strong internal consistency [1], and its strong frequentist properties (the underlying latent variables are exactly calibrated in the sense of Lawless and Fredette [11]). The preliminary results reported in this paper make clear that more research is needed, not only into NPI but also more generally on properties of statistical inference with lower and upper probabilities, in particular the relation between inference, imprecision, and information.

References

1. Augustin, T., Coolen, F.P.A.: Nonparametric predictive inference and interval probability. J. Statist. Plann. Inference. 124, 251–272 (2004)
2. Bayes, T.: An essay towards solving a problem in the doctrine of chances. Philos. Trans. R. Soc. London 53, 370–418 (1963); 54, 296–325
3. Coolen, F.P.A.: Low structure imprecise predictive inference for Bayes' problem. Statist. Probab. Lett. 36, 349–357 (1998)
4. Coolen, F.P.A.: On nonparametric predictive inference and objective Bayesianism. J. Logic Lang. Inform. 15, 21–47 (2006)
5. Coolen, F.P.A., Coolen-Schrijner, P.: Comparing proportions data with few successes. In: Lawry, J., Miranda, E., Bugarin, A., Li, S., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing, vol. 37, pp. 241–248. Springer, Berlin (2006)
6. Coolen, F.P.A., Coolen-Schrijner, P.: Nonparametric predictive comparison of proportions. J. Statist. Plann. Inference 137, 23–33 (2007)
7. Coolen, F.P.A., Elsaeiti, M.A.: Nonparametric predictive methods for acceptance sampling (in preparation, 2008)
8. De Cooman, G., Zaffalon, M.: Updating beliefs with incomplete observations. Artificial Intelligence 159, 75–125 (2004)
9. De Finetti, B.: Theory of Probability. Wiley, Chichester (1974)
10. Hill, B.M.: Posterior distribution of percentiles: Bayes' theorem for sampling from a population. J. Amer. Statist. Assoc. 63, 677–691 (1968)
11. Lawless, J.F., Fredette, M.: Frequentist prediction intervals and predictive distributions. Biometrika 92, 529–542 (2005)
12. Manski, C.F.: Partial Identification of Probability Distributions. Springer, New York (2003)
13. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
14. Weichselberger, K.: Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept. Physika, Heidelberg (2001)
15. Weichselberger, K.: Applying symmetric probability theory. In: Proceedings 56th Session of the International Statistical Institute (ISI 2007, Lisbon, Portugal) CD-Rom:848.pdf (2007)
16. Zaffalon, M.: Conservative rules for predictive inference with incomplete data. In: Cozman, F.G., Nau, R., Seidenfeld, T. (eds.) Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2005, Pittsburgh, Pennsylvania, USA) (2005)

Inferring a Possibility Distribution from Very Few Measurements

Gilles Mauris

Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance (LISTIC), Polytech'Savoie, Annecy Le Vieux, France

Abstract. This paper considers the problem of the possibility representation of measurement uncertainty in the cases of information shortage: very few measurements, poor knowledge about the underlying probability distribution. After having related possibility distribution to probability confidence intervals, we present a procedure to build a possibility distribution for one measurement issued from an unimodal probability distribution. We consider then the addition of other measurements and more knowledge about the probability distribution. The key role of the uniform distribution as the probability distribution leading to the least specific possibility distribution is highlighted. The approach is compared and discussed versus the conventional one based on the Student distribution.

Keywords: Possibility theory, Probability theory, Uncertainty, Scarce measurements.

1 Introduction

In many application domains, it is important to take the measurement uncertainties into account [11], [16], especially in order to define around the measurement result an interval which will contain the real value of the considered entity with specified confidence [8], that is, a confidence interval [9]. Such an interval allows to define decision risks later, as for example the risk to exceed an alarm threshold, etc. In practice, two main theories are considered to deal with measurement uncertainty: interval calculus [14] and probability theory [9]. As interval calculus only supplies the confidence interval with 100% confidence, probability theory seems to be required to supply the other confidence intervals. But to handle the whole set of confidence intervals (with all the confidence levels) is quite complex by a probability approach. And choosing a particular confidence level (e.g. 95% which means a .05 probability for the value to be out of the interval) is rather arbitrary. Thus a possibility approach has been proposed in [5, 6, 15] and further developed by a few authors in a measurement context [2, 7, 12, 13, 17].

This paper further explores the connection between possibility distribution and confidence intervals and addresses the possibility expression of measurement uncertainty for situations where only very limited knowledge is available: very few measurements, unknown unimodal probability density. In Section 2, we recall how a possibility distribution can be built from confidence intervals. In the third section, we present the main contribution of the paper, i.e. how to define confidence intervals where only limited knowledge is available about the underlying probability density (unimodal bounded/non-bounded,

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 92–99, 2008.
springerlink.com

symmetric or not). The results are then applied to expression of uncertainty when only very few measurements are available. The key role of the uniform distribution as the probability distribution leading to the least specific possibility distribution is highlighted. The approach is compared and discussed versus the conventional one based on the Student distribution. Some concluding remarks point out the interest of the approach and some future developments.

2 Possibility Distribution Versus Confidence Intervals

2.1 Basics of the Possibility Theory

The possibility theory is one of the modern theories available to represent uncertainty when information is scarce and/or imprecise [18]. The basic notion is the possibility distribution, denoted π. Here, we consider possibility distributions defined on the real line, i.e. π is an upper semi-continuous mapping from the real line to the unit interval. Thus π is a fuzzy subset but with specific semantics for the membership function. Indeed, a possibility distribution describes the more or less plausible values of some uncertain variable X. The possibility theory provides two evaluations of the likelihood of an event, for instance whether the value of a real variable X does lie within a certain interval: the possibility Π and the necessity N. The normalized measures of possibility Π and necessity N are defined from the possibility distribution $\pi : \mathbb{R} \rightarrow [0, 1]$ such that $\sup_{x\in\mathbb{R}} \pi(x) = 1$ as follows:

$$\forall A \subset \mathbb{R},\ \Pi(A) = \sup_{x\in A} \pi(x) \quad \text{and} \quad \forall A \subset \mathbb{R},\ N(A) = 1 - \Pi(\bar{A}) = \inf_{x\in A}(1 - \pi(x)).$$

The possibility measure Π satisfies $\Pi(A \cup B) = \max(\Pi(A), \Pi(B)), \forall A, B \subset \mathbb{R}$.

The necessity measure N satisfies $N(A \cap B) = \min(N(A), N(B)), \forall A, B \subset \mathbb{R}$.

A possibility distribution π_1 is more specific than π_2 as soon as $\pi_1 \leq \pi_2$ (in the usual definition of inclusion of fuzzy sets), i.e. π_1 is more informative than π_2. In fact, possibility measures are set functions similar to probability measures, but they rely on axioms which involve the operations "maximum" and "minimum" instead of the operations "addition" and "product" (if the measures are decomposable [3]).

2.2 Possibility Representation of Confidence Intervals

Let us assume that the random variable associated to the measurement results is denoted X (a realization of X is denoted x), is continuous on the set of reals and is described by a probability density function p, F being its corresponding probability distribution function with F^{-1} its inverse function if it exists (otherwise the pseudo-inverse function can be considered [9]). For every possible confidence level $\beta \in [0, 1]$, the corresponding confidence interval is defined as an interval that contains the measurand (i.e. the physical entity to be determined denoted μ) with probability $\geq \beta$. In other words, a confidence interval of confidence level β (denoted I_β) is defined as an interval for which the probability P_{out} to be outside this interval I_β does not exceed $\alpha \stackrel{\text{def}}{=} 1 - \beta$, i.e. $P(\mu \notin I_\beta) = \alpha$.

It is possible to link confidence intervals and possibility distribution in the following way. A unimodal numerical possibility distribution may be viewed as a nested set of

confidence intervals, which are the α cuts $[\underline{x}_\alpha, \overline{x}_\alpha] = \{x,\ \pi(x) \geq \alpha\}$ of π. The degree of certainty that $[\underline{x}_\alpha, \overline{x}_\alpha]$ contains μ is $N([\underline{x}_\alpha, \overline{x}_\alpha])$ (if continuous). Obviously, the confidence intervals built around the same point x^* are nested. It has been proven in [12] that stacking confidence intervals of a probability distribution on top of one another leads to a possibility distribution (denoted π^* having x^* as modal value). In fact, in this way, the α-cuts of π^*, i.e. $A_\alpha = \{x, | \pi^*(x) \geq \alpha\}$ are identified with the confidence interval I^*_β of confidence level $\beta = 1 - \alpha$ around the nominal value x^*. Thus, the possibility distribution π^* encodes the whole set of confidence intervals in its membership function. Moreover, this possibility distribution satisfies $\Pi^*(A) \geq P(A)$, $\forall A \subset \mathbb{R}$, with Π^* and P the possibility and probability measures associated respectively to π^* and p (the underlying probability density function of the measurement results).

A closed form expression of the possibility distribution $\pi^M(x)$ induced by confidence intervals around the mode $x^* = M$ is obtained for unimodal continuous probability densities $p(x)$ strictly increasing on the left and decreasing on the right of M [4]:

$$\pi^M(x) = \int_{-\infty}^{x} p(y)dy + \int_{\phi(x)}^{+\infty} p(y)dy = F(x) + 1 - F(\phi(x)) = \pi^M(\phi(x)) \tag{1}$$

for all $x \in [-\infty, M]$, where ϕ is a decreasing mapping $\phi : [-\infty, M] \mapsto [M, \infty] \,|\, \phi(M) = M$. $\pi^M(x)$ is the probability that the measurand μ is outside the interval $[x, \phi(x)]$, i.e. $1 - \pi^M(x)$ is the confidence level of this interval.

3 Inferring a Possibility Distribution from a Small Sample

We will consider confidence intervals associated with an underlying probability density being unimodal (i.e. having only one maximum, both local and global) with different assumptions: bounded and non bounded, symmetric or not. Most of the following results are based on trivial properties of unimodal distribution described below.

Let us consider a unimodal probability density p with the mode M that will be identified to the measurand. Thus, p is non increasing for its argument values greater than M, and non decreasing for its argument values less than M. Therefore, for any values superior to M such that $x_3 \geq x_2 \geq x_1$, the average of p over $[x_2, x_3]$ must be less than or equal to its average over $[x_1, x_3]$:

$$\frac{\int_{x_2}^{x_3} p(x)dx}{x_3 - x_2} \leq \frac{\int_{x_1}^{x_3} p(x)dx}{x_3 - x_1} \tag{2}$$

Similarly, for any values less than M such that $x_1 \leq x_2 \leq x_3$:

$$\frac{\int_{x_1}^{x_2} p(x)dx}{x_2 - x_1} \leq \frac{\int_{x_1}^{x_3} p(x)dx}{x_3 - x_1} \tag{3}$$

Note that the equality in (2) and (3) holds if p is constant on the considered domain.

3.1 Bounded Probability Density

Let us consider that X is defined by a probability density, its mode is denoted M and its support $[M - a, M + b]$. Then the mode and the support of $X - M$ are respectively 0 and the interval $[-a, b]$. We have the following result:

Proposition 1. $\forall t \in [0, 1]$, $\Pr[X - ta \leq M \leq X + tb] \geq t$.

Proof. $\Pr[X-ta \le M \le X+tb] = \Pr[-ta \le X-M \le tb]$ and $\Pr[-ta \le X-M \le tb] = 1 - \int_{-a}^{-ta} p(x)dx - \int_{tb}^{b} p(x)dx$.

Then by applying (3) to $x_1 = -a, x_2 = -ta, x_3 = 0$ and (2) to $x_1 = 0, x_2 = tb, x_3 = b$, we obtain:

$$\int_{-a}^{-ta} p(x)dx \le (1-t)\int_{-a}^{0} p(x)dx \quad \text{and} \quad \int_{tb}^{b} p(x)dx \le (1-t)\int_{0}^{b} p(x)dx$$

Therefore:

$$\int_{-a}^{-ta} p(x)dx + \int_{tb}^{b} p(x)dx \le (1-t)\int_{-a}^{b} p(x)dx = 1-t$$

Then: $\forall t \in [0,1], \Pr[X-ta \le M \le X+tb] \ge 1-(1-t) = t$. □

Therefore the corresponding possibility distribution is defined by:

$$\forall x \in [M-a, M], \pi^M(x) \le \frac{x+a-M}{a} \quad \text{and} \quad \forall x \in [M, M+b], \pi^M(x) \le \frac{-x+M+b}{b}$$

Therefore, the possibility distribution defined by the triangular possibility distribution having for support $[M-a, M+b]$ is consistent with all the unimodal probability distributions (symmetric or not) having M as modal value and $[M-a, M+b]$ as support.

Note that the triangular possibility distribution is also the possibility distribution associated to the uniform probability density. Moreover, the triangular symmetric possibility distribution with support $[M-a, M+b]$ and mode M, is the least upper bound of all the possibility transforms of symmetric probability distributions having M for modal value and $[M-a, M+b]$ for support. This result has been previously stated in [4] but in another way.

3.2 Non Bounded Probability Density

As the support is known as infinite, the intervals have to be built from other information from the random variable. Thus, we will consider intervals of the form $X \pm t|X|$. In fact, instead of starting from the support as for bounded distributions, we propose to start from the mode.

The following result holds for any unimodal distribution [1, 10]:

Proposition 2

$$\Pr[X - t|X| \le M \le X + t|X|] \ge 1 - \frac{2}{1+t} \quad \textit{for } t \ge 1 \tag{4}$$

Proof

$$\Pr[|X-M| \le t|X|] = \Pr\left[\left|1 - \frac{M}{X}\right| \ge t\right] = \Pr\left[\frac{M}{X} \in 1 \pm t\right]$$
$$= \Pr\left[X \in M\frac{1}{1 \pm t}\right] = \Pr\left[X - M \in M\left(\frac{1}{1 \pm t} - 1\right)\right]$$

Thus

$$\Pr\left[X - M \in M\left(\frac{1}{1 \pm t} - 1\right)\right] = \begin{cases} F\left(M\frac{-t}{t+1}\right) - F\left(M\frac{-t}{t-1}\right) & \text{for } M \ge 0 \\ F\left(M\frac{-t}{t-1}\right) - F\left(M\frac{-t}{t+1}\right) & \text{for } M \le 0 \end{cases}$$

Then by applying (2) to $x_1 = \frac{-tM}{t-1}$, $x_2 = \frac{-tM}{t+1}$, $x_3 = M$ and (3) to $x_1 = \frac{-tM}{t+1}$, $x_2 = \frac{-tM}{t-1}$, $x_3 = M$, we obtain respectively:

$$F\left(M\frac{-t}{t+1}\right) - F\left(M\frac{-t}{t-1}\right) \leq \frac{2}{t+1} \text{ for } M \geq 0$$

$$F\left(M\frac{-t}{t-1}\right) - F\left(M\frac{-t}{t+1}\right) \leq \frac{2}{t+1} \text{ for } M \leq 0$$

Therefore: $\Pr[|X-M| \leq t|X|] \leq \frac{2}{t+1}$ and finally we obtain: $\Pr[X - t|X| \leq M \leq X + t|X|] \geq 1 - \frac{2}{1+t}$. □

By the same reasoning, we obtain for a symmetric unimodal probability density:

$$\Pr[X - t|X| \leq M \leq X + t|X|] \geq 1 - \frac{1}{1+t} \quad \text{for} t \geq 1 \tag{5}$$

Note that the equality (5) holds for p uniform, and thus this probability distribution is the least favourable in the sense that it gives the least specific possibility distribution (for $t \geq 1$). If the shape of the probability distribution is known, the inequality can be reduced for high values of t. For example, if it is Gaussian, the bound in (5) can be improved [1]:

$$\Pr[X - t|X| \leq M \leq X + t|X|] \geq 1 - \frac{0.484}{t-1} \quad \text{for } t \geq 1 \tag{6}$$

3.3 Case of One Measurement

Let us consider the case where only one single measurement is available. In this case it is natural to consider that the observed value corresponds to the mode of the underlying probability density. If the density is assumed to be non symmetric, we have from (4) $\pi(x_0 - tx_0) = \pi(x_0 + tx_0) = \frac{2}{1+t}$ for $t \geq 1$. If it is symmetric, we have from (5) $\pi(x_0 - tx_0) = \pi(x_0 + tx_0) = \frac{1}{1+t}$ for $t \geq 1$. If it is Gaussian, we have from (6) $\pi(x_0 - tx_0) = \pi(x_0 + tx_0) = \frac{0.484}{t-1}$ for $t \geq 0.484$.

Let us consider for example the case where a sensor provides a single value of 30°C the associated probability distribution is supposed to be unimodal. Figure 1a) highlights the reduction of confidence interval lengths according to the amount of available knowledge: when the distribution is non symmetric, when the distribution is symmetric. When it is Gaussian, the use of the equation (6) leads to a reduction of confidence interval lengths only for high values of t. For low values of t, the exact expression will also give reduced intervals but it has not yet been computed; the uniform distribution being the least favorable for $t \geq 1$.

3.4 Case of one Measurement and a Guess

By making the variable change of X into $X - A$, in (4), is replaced by $X - A$ and M by $M - A$, then the following result is deduced for any unimodal symmetric distribution:

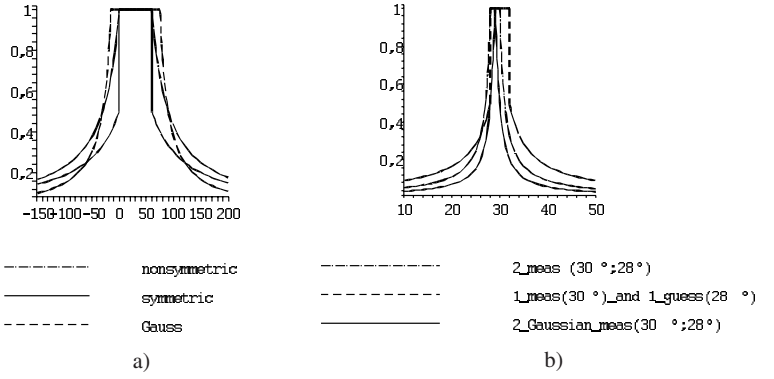


Fig. 1. Possibility distributions a) for one measurement b) for two measurements

$$\Pr[X - t|X - A| \leq M \leq X + t|X - A|] \geq 1 - \frac{1}{1+t} \text{ for } t \geq 1 \tag{7}$$

This result can be used to introduce via A some form of prior information (called a guess, coming for example from an expert) concerning the dispersion. In fact, $|x_1 - A|$ can be viewed as the equivalent of the sample standard deviation used classically (see Section 3.6). The introduction of A allows to reduce the lengths of confidence intervals obtained by one single measurement as it is illustrated in Figure 1b).

3.5 Case of Two Measurements

Let us now consider the case where a second measurement x_2, coming from the same probability distribution as x_1 and considered as being independent from it. We propose (in an equivalent way with classical propositions when two measurements are available) to consider the confidence intervals of the form:

$$\frac{X_1 + X_2}{2} - t\frac{|X_1 - X_2|}{2} \leq M \leq \frac{X_1 + X_2}{2} + t\frac{|X_1 - X_2|}{2}$$

In the case of symmetric unimodal distribution, we obtain by the same reasoning as the one used in Section 3.2:

$$\Pr\left[\frac{X_1 + X_2}{2} - \frac{t}{2}|X_1 - X_2| \leq M \leq \frac{X_1 + X_2}{2} + \frac{t}{2}|X_1 - X_2|\right] \geq 1 - \frac{1}{1+t} \tag{8}$$

The Fig. 1b) illustrates the case where the sensor provides the two measurements $x_1 = 30^{\circ}\text{C}$ and $x_2 = 28^{\circ}\text{C}$.

3.6 Discussion Versus the Conventional Probability Approach

The above mathematical derivations formalize the idea that without any appeal to other information (except unimodality), we can compute the actual length of the finite confidence interval. It is remarkable that the confidence intervals thus created have finite lengths, except for the 100% confidence level (see Fig. 1a). Indeed, this result seems to contradict the standard statistical intuition that at least two measurements are required

in order to have some idea about the dispersion (i.e. to have an estimation of the standard deviation σ). Indeed, the conventional probability recommendation to deal with a small number n (but $n > 1$) of measurement consists in using confidence intervals of the form [8]:

$$\bar{X} - tS/\sqrt{n} \leq M \leq \bar{X} + tS/\sqrt{n} \tag{9}$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$ is the sample mean, and $S = [\sum_{i=1}^{n}(X_i - \bar{X})/(n-1)]^{1/2}$ the sample standard deviation.

If the underlying probability distribution is Gaussian, the t value is the one given by the Student distribution for a given confidence level. An interesting remark is that for $n = 2$, (9) has the same form as (8). Indeed, in this case of two measurement, (8) is equivalent to (9) for a Gaussian distribution. The Fig. 1a) gives an example of the effect on the possibility distribution specificity (for high values of t) of making the Gaussian assumption.

4 Conclusion

A possibility distribution can encode a family of probability distributions. This fact has been used as a basis for a transformation of a probability distribution into a possibility distribution by using the notion of confidence intervals. Thus the possibility distribution has been related to probability inequalities, especially for unimodal bounded (or not) symmetric (or not) probability distributions. The obtained results have been used for a possibility expression of measurement uncertainty in situations where only a very limited knowledge is available: one or two measurements, unknown unimodal probability density. In fact, the proposed approach extends the conventional probability approach of Student to the case of one single measurement and to the case of non Gaussian distribution for two measurements. The results highlight the key role of the uniform probability distribution that leads to the least specific possibility distribution at least for high confidence levels. Further developments will consider how having more measurements allows to shorten the confidence intervals and thus to increase the specificity of the corresponding possibility distribution.

References

1. Anděl, J.: Confidence intervals for the mode based on one observation. Zb Rad. 5, 87–96 (1991)
2. Baudrit, C., Dubois, D.: Practical representation of incomplete probabilistic information. Comput. Statist. Data Anal. 51(1), 86–108 (2006)
3. Dubois, D., Fodor, J.C., Prade, H., Roubens, M.: Aggregation of decomposable measures with applications to utility theory. Theory and Decision 41(1), 195–223 (1996)
4. Dubois, D., Foulloy, L., Mauris, G., Prade, H.: Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities. Reliab. Comput. 10(4), 273–297 (2004)
5. Dubois, D., Prade, H.: On several representations of an uncertain body of evidence. In: Gupta, M.M., Sanchez, E. (eds.) Fuzzy Information and Decision Processes, pp. 167–181. North-Holland, Amsterdam (1982)

6. Dubois, D., Prade, H., Sandri, S.: On possibility/probability transformations. In: Lowen, R., Roubens, M. (eds.) Fuzzy Logic, pp. 103–112 (1993)
7. Ferrero, A., Salicone, S.: The random-fuzzy variables: a new approach to the expression of uncertainty. IEEE Trans. Instrum. Meas. 53(5), 1370–1377 (2004)
8. ISO-IEC-OIML-BIPM, Guide to the Expression of Uncertainty in Measurement, 2nd edn. (1995)
9. Kendall, M., Stuart, A.: The Advanced Theory of Statistics, vol. 1. Charles Griffin, London (1977)
10. Machol, R.E., Rosenblatt, J.: Confidence interval based on single observation. Proceedings of IEEE 54(8), 1087–1088 (1966)
11. Mari, L., Zingales, G.: Uncertainty in measurement science. In: Proceedings of the Int. Workshop on Advances of Measurement Science (Kyoto, Japan), pp. 177–189 (1999)
12. Mauris, G., Lasserre, V., Foulloy, L.: Fuzzy modeling of measurement data acquired from physical sensors. IEEE Trans. Instrum. Meas. 49(6), 1201–1205 (2000)
13. Mauris, G., Lasserre, V., Foulloy, L.: A fuzzy approach for the expression of uncertainty in measurement. Measurement 29(3), 165–177 (2001)
14. Moore, R.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
15. Moral, M.: Constructing of possibility distribution from a probability distribution. In: Jones, A., Kauffmann, A., Zimmermann, H.J. (eds.) Fuzzy Sets Theory and Applications, pp. 51–60 (1986)
16. Reznik, L., Dabke, K.P.: Evaluation of uncertainty in measurement: a proposal for application of intelligent methods. In: Proceedings of the XV IMEKO World Congress (Osaka, Japan), vol. II, pp. 93–100 (1999)
17. Urbanski, M., Wasowski, J.: Fuzzy approach to the theory of measurement inexactness. Measurement 34, 67–74 (2003)
18. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1(1), 3–28 (1978)

Statistics with Vague Data and the Robustness to Data Representation

Przemysław Grzegorzewski[1,2]

[1] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
[2] Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

Abstract. Robustness is an important problem in statistics. However, robustness of statistical procedures for vague data cannot be limited to insensitivity to departures from assumptions on the underlying distributions. Since the shapes of membership functions applied for modelling vague data are generally strongly subjective one may ask about the influence of these shapes on further decisions.Thus the robustness of the statistical procedures to data representation is also of interest.

Keywords: Membership function, Robustness, Vague data.

1 Introduction

As stringent assumptions on distributions lead sometimes to serious difficulties in statistics, the problem becomes much more serious in the presence of imprecise data, where we still do not have satisfactory goodness-of-fit techniques. A remedy for this problem might be the use of distribution-free methods (e.g. some nonparametric tests for fuzzy data were suggested in [2], [5] or [7]). However, if we process fuzzy data another aspect of robustness appears. Statistical procedures depend strongly on the shapes of membership function utilized for modelling data. Moreover, different persons may assign distinct membership functions to the same vague objects since modelling vagueness cannot be completely free from subjectivity. There we are faced with a kind of paradox especially that using fuzzy modelling we make every endeavor to be flexible yet we are still very restricted by the very choice of the precise form of the membership functions.

Since the shape of membership functions applied for modelling vague data is generally strongly subjective one should ask about the possible influence of that shape on further decisions. Therefore, we need some tools to evaluate the robustness of the statistical procedures to data representation. In the present paper we propose a few measures for evaluating and comparing this kind of robustness.

2 Robustness in Statistics - The General Idea

The notion of robustness has been introduced to statistics by Box and Anderson [1] as follows: "To fulfill the needs of the experimenter, statistical criteria should be: (i)

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 100–107, 2008.
springerlink.com

sensitive to change in the specific factor tested, (ii) insensitive to change, of a magnitude likely to occur in practice, in extraneous factors". A statistical procedure which satisfies the first requirement is said to be powerful or efficient, while the procedure that satisfies the second is called *robust*.

A robust procedure performs well not only under assumptions that have been postulated but also under departures from the ideal model. This ideal model will be called a *core model*. To get a satisfying quantitative description of robustness we have to specify two other objects: an extension of the core model corresponding to departures from the ideal that are likely to occur and a measure quantifying the behavior of the discussed property under deviation from the core model described by the extended model.

Robust estimation when gross errors occur has been studied primarily and estimators that are relatively unaffected by the presence of outliers were of interest (see, e.g., Huber [9] and Hampel [8]). Later various statistical techniques which admit departures from the assumptions on the underlying distribution were proposed. Another concept of robustness can be found in the Bayesian analysis to denote the inference that will be unaffected by any possible mis-specification of the prior distribution. However, here we will present another concept for quantifying robustness which seems to be the closest to the original definition of robustness given by Box and Anderson. This approach was suggested by Zieliński [12] .

3 A Quantitative Approach to Robustness

Let $M_0 = (\mathscr{X}, \mathscr{A}, \mathscr{P}_0)$ denote a core model, where $\mathscr{P}_0 \subset \mathscr{P}$ is a subfamily of the set $\mathscr{P}$ of all possible probability measures on $\mathscr{A}$. Facing any statistical decision problem (estimation, hypothesis testing, etc.) we choose a suitable statistic T. Its distribution that corresponds to $P \in \mathscr{P}$ will be denoted by P^T. Moreover, let $\mathscr{P}^T = \{P^T : P \in \mathscr{P}\}$, $\mathscr{P}_0^T = \{P^T : P \in \mathscr{P}_0\}$, etc.

Let $\pi : \mathscr{P}_0 \to 2^{\mathscr{P}}$ be a function such that $P \in \pi(P)$. It means that for each distribution $P \in \mathscr{P}_0$ we get a neighborhood $\pi(P) \subset \mathscr{P}$. Moreover, let $\mathscr{P}_1 = \bigcup_{P \in \mathscr{P}_0} \pi(P)$. Then the statistical structure $M_0 = (\mathscr{X}, \mathscr{A}, \mathscr{P}_1)$ will be called the extension of the core model M_0 (or, more precisely, π-extension of the core model). Now, let ρ denote a real valued function on $\mathscr{P}_1^T$. Then we get a following definition.

Definition 1. *A function* $r_T : \mathscr{P}_0 \to \mathbb{R}$ *defined as*

$$r_T(P) = \sup\{\rho(Q^T) : Q \in \pi(P)\} - \inf\{\rho(Q^T) : Q \in \pi(P)\} \tag{1}$$

is called ρ*-robustness of statistic* T *with respect to the* π*-extension* M_1 *of the core model* M_0.

If ρ used in given statistical problem has its traditional name (like power, bias, width of the confidence interval, risk, etc.) this very name is also used for function r_T (i.e. we consider the power-robustness, the bias-robustness, etc.).

Using function r_T we can compare the robustness of different statistical procedures. We say that statistic T is more robust than statistic S if $r_T(P) \leq r_S(P)$ for each $P \in \mathscr{P}_0$ and $r_T(P) < r_S(P)$ for some $P \in \mathscr{P}_0$. A statistic T is called uniformly most robust in a

given class $\mathscr{T}$ of statistics if $r_T(P) \leq r_U(P)$ for any $P \in \mathscr{P}_0$ and for any $U \in \mathscr{T}$. And we say that statistic T is absolutely robust if $r_T(P) = 0$ for each $P \in \mathscr{P}_0$.

The last approach to robustness is not only the most general on the ground of the classical statistical inference but it seems also to be the most suitable one for our purpose to extend the discussion on robustness to the statistical inference for imprecise data. Actually, here not only distributional robustness but insensitivity to the particular shape of the membership functions applied for modelling fuzzy data is of interest as well.

4 Modelling Vague Data

Consider a random experiment with vague outcomes described by fuzzy numbers. Let us recall that a fuzzy subset A of the real line $\mathbb{R}$, with the membership function $\mu_A : \mathbb{R} \to [0,1]$, is a fuzzy number if and only if A is normal (i.e. there exists an element x_0 such that $\mu_A(x_0) = 1$), A is fuzzy convex (i.e. $\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \mu_A(x_1) \wedge \mu_A(x_2)$, $\forall x_1, x_2 \in \mathbb{R}$, $\forall \lambda \in [0,1]$), μ_A is upper semicontinuous and supp$A$ is bounded, where supp$A = cl(\{x \in \mathbb{R} : \mu_A(x) > 0\})$, and cl is the closure operator.

The $\alpha-$cut of a fuzzy number A is a nonfuzzy set $A_\alpha = \{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$. It is easily seen that every α-cut of a fuzzy number is a closed interval $A_\alpha = [A_\alpha^L, A_\alpha^U]$, where $A_\alpha^L = \inf\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$ and $A_\alpha^U = \sup\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}$. A space of all fuzzy numbers will be denoted by $\mathbb{FN}(\mathbb{R})$.

We assume that the outcomes of our experiment are realizations of n-dimensional fuzzy random sample $Y_1, \ldots, Y_n$ which may be treated as a fuzzy perception of the usual random sample $X_1, \ldots, X_n$. There are several definitions of a fuzzy random variable ([10], [11]). Here we simply assume that a mapping $Y : \Omega \to \mathbb{FN}(\mathbb{R})$ is called a fuzzy random variable if $\{Y(\alpha, \omega) : \alpha \in (0,1]\}$ is a set representation of $X(\omega)$ for all $\omega \in \Omega$ and for each $\alpha \in (0,1]$ both $Y_\alpha^L = Y_\alpha^L(\omega)$ and $Y_\alpha^U = Y_\alpha^U(\omega)$, are usual real-valued random variables on a probability space $(\Omega, \mathbb{A}, P)$.

It is obvious that there is a significant subjectivity in attributing membership function to observations. Moreover, even for similar contexts, fuzzy sets representing the same concepts may vary considerably. Then the crucial problem is to check whether given application is not overly sensitive to variations in shape of the membership functions applied for modelling data. Thus we need tools that help in characterizing the sensitivity of the statistical procedures to the particular choice of the membership function. Below we suggest how to construct quantitative measures useful both for the comparison and for evaluating robustness of the procedures under study. To distinguish this kind of robustness and the distributional robustness discussed in the previous section, the insensitivity to shape of the membership functions will be called further on as *robustness to data representation*.

5 Robustness to Data Representation

Let $\mathscr{Y}_i$, $i = 1, \ldots, n$ denote a set of all possible fuzzy numbers that are realizations of Y_i ($\mathscr{Y}_i \subset \mathbb{FN}(\mathbb{R})$). Without loss of generality we will identify Y_i with its membership function μ_{Y_i}. Then $\mathscr{Y}_0 = \bigcup_{i=1}^{n} \mathscr{Y}_i$ creates a sample space of the core model.

Now, let $\kappa : \mathscr{Y}_0 \to 2^{\mathbb{FN}(\mathbb{R})}$ denote a function such that each $\mathscr{Y}_i \subset \kappa(\mathscr{Y}_i)$. It means that for each family of membership functions $\mathscr{Y}_i \subset \mathscr{Y}_0$ we obtain a neighborhood $\kappa(\mathscr{Y}_i) \subset \mathbb{FN}(\mathbb{R})$. Moreover, let $\mathscr{Y}_{ext} = \bigcup_{i=1}^{n} \kappa(\mathscr{Y}_i)$. This family of fuzzy numbers becomes the extension of our core model $\mathscr{Y}_0$ (or, more precisely, κ-extension of the core model).

To solve a statistical problem we use a suitable statistic $T = T(Y_1, \ldots, Y_n)$. Let $\mathscr{Y}_0^T$ denote a family of all possible values of T provided data come from the core model, while $\mathscr{Y}_{ext}^T$ denotes a family of all possible values of T in the extended model. Finally, let ζ denote a real valued function on $\mathscr{Y}_{ext}^T$.

Definition 2. *A function $SUS_T : \mathscr{Y}_0 \to \mathbb{R}$ defined as*

$$\begin{aligned} SUS_T(Y_1, \ldots, Y_n) = &\sup\{\zeta(T(Z_1, \ldots, Z_n)) : Z_i \in \kappa(\mathscr{Y}_i), i = 1, \ldots, n\} \\ &- \inf\{\zeta(T(Z_1, \ldots, Z_n)) : Z_i \in \kappa(\mathscr{Y}_i), i = 1, \ldots, n\} \end{aligned} \tag{2}$$

*is called a **total ζ-susceptibility to data representation** of statistic T with respect to the κ-extension $\mathscr{Y}_{ext}$ of the core model $\mathscr{Y}_0$.*

Using function SUS_T we can compare the robustness to data representation of different statistical procedures. So we say that statistic T is more robust to data representation than statistic S if it has smaller susceptibility, i.e.

$$SUS_T(Y_1, \ldots, Y_n) < SUS_S(Y_1, \ldots, Y_n). \tag{3}$$

We may also say that statistic T is uniformly most robust to data representation in a given class $\mathscr{T}$ of statistics if and only if $SUS_T(Y_1, \ldots, Y_n) \leq SUS_U(Y_1, \ldots, Y_n)$ for $U \in \mathscr{T}$. Similarly, we can say that statistic T is absolutely robust to data representation if $SUS_T(Y_1, \ldots, Y_n) = 0$.

The suggested measure of robustness to data representation tells us what can happen if we admit possible departures from the core model for all observations simultaneously. However, it might be also interesting to quantify how sensitive is given statistical procedure under departures from the core model on a single observation only.

Definition 3. *A function $AS_T : \mathscr{Y}_0 \to \mathbb{R}$ defined as*

$$AS_T(Y_1, \ldots, Y_n) = \frac{1}{n} \sum_{i=1}^{n} sus_i^T(Y_1, \ldots, Y_n), \tag{4}$$

where

$$\begin{aligned} sus_i^T(Y_1, \ldots, Y_n) = &\sup\{\zeta(T(Z_1, \ldots, Z_n)) : Z_1 \in \mathscr{Y}_1, \ldots, Z_{i-1} \in \mathscr{Y}_{i-1}, \\ &Z_i \in \kappa(\mathscr{Y}_i), Z_{i+1} \in \mathscr{Y}_{i+1}, , \ldots, Z_n \in \mathscr{Y}_n\} \\ &- \inf\{\zeta(T(Z_1, \ldots, Z_n)) : Z_1 \in \mathscr{Y}_1, \ldots, Z_{i-1} \in \mathscr{Y}_{i-1}, \\ &Z_i \in \kappa(\mathscr{Y}_i), Z_{i+1} \in \mathscr{Y}_{i+1}, , \ldots, Z_n \in \mathscr{Y}_n\} \end{aligned} \tag{5}$$

*is called an **average individual ζ-susceptibility to data representation** of statistic T with respect to the κ-extension $\mathscr{Y}_{ext}$ of the core model $\mathscr{Y}_0$.*

Definition 4. *A function $MIS_T : \mathscr{Y}_0 \to \mathbb{R}$ defined as*

$$MIS_T(Y_1, \ldots, Y_n) = \max_{i=1,\ldots,n} sus_i^T(Y_1, \ldots, Y_n), \tag{6}$$

where $sus_i^T(Y_1,\ldots,Y_n)$ *is given by* (5)*, is called a* ***maximal individual*** ζ***-susceptibility to data representation*** *of statistic* T *with respect to the* κ*-extension* $\mathscr{Y}_{ext}$ *of the core model* $\mathscr{Y}_0$.

The suggested measures of robustness to data representation are useful for comparing different procedures but looking on their value it is hardly to say whether the robustness of given procedure is high or weak. Thus we propose another measure which could is defined provided the susceptibility to data representation with respect to the maximal possible extension of the core model is finite. More precisely, let us assume that the supremum of the ζ-susceptibility to data representation of statistic T with respect to any κ-extension, is finite, i.e. if

$$MSUS = \sup_{\kappa\in\mathscr{K}} SUS_T(Y_1,\ldots,Y_n) < \infty, \tag{7}$$

where $\mathscr{K}$ is a family of all functions creating possible κ-extensions. Then we get a following measure.

Definition 5. *A function* $R_T : \mathscr{Y}_0 \to \mathbb{R}$ *defined as*

$$R_T(Y_1,\ldots,Y_n) = 1 - \frac{SUS_T(Y_1,\ldots,Y_n)}{MSUS} \tag{8}$$

is called a ***total*** ζ***-robustness to data representation*** *of statistic* T *with respect to the* κ*-extension* $\mathscr{Y}_{ext}$ *of the core model* $\mathscr{Y}_0$.

As it is easily seen $0 \le R_T \le 1$ and the bigger value of R_T the higher robustness of the procedure under study. Given statistical procedure is absolutely robust if and only if $R_T = 1$. Obviously, assuming that $MSUS$ is finite one can define an **average individual** ζ**-robustness to data representation** of statistic T, i.e.

$$AR_T(Y_1,\ldots,Y_n) = 1 - \frac{AS_T(Y_1,\ldots,Y_n)}{MSUS} \tag{9}$$

and a **maximal individual** ζ**-robustness to data representation** of statistic T with respect to the κ-extension $\mathscr{Y}_{ext}$ of the core model $\mathscr{Y}_0$, i.e.

$$MIR_T(Y_1,\ldots,Y_n) = 1 - \frac{MIS_T(Y_1,\ldots,Y_n)}{MSUS}. \tag{10}$$

6 Example

Suppose $Y_1,\ldots,Y_5$ denote a fuzzy perception of random sample $X_1,\ldots,X_5$ from the normal distribution $N(\theta,\sigma)$ described as follows: Y_1 =“about 1”, Y_2 =“about 2”, Y_3 =“about 3”, Y_4 =“about 4” and Y_5 =“about 5”. Suppose these data are modelled by triangular fuzzy numbers characterized by following α-cuts, respectively: $(Y_1)_\alpha = [\alpha, 2-\alpha]$, $(Y_2)_\alpha = [1+\alpha, 3-\alpha]$, $(Y_3)_\alpha = [2+\alpha, 4-\alpha]$, $(Y_4)_\alpha = [3+\alpha, 5-\alpha]$, $(Y_5)_\alpha = [4+\alpha, 6-\alpha]$, where $\alpha \in (0,1]$. Moreover, let us assume that the standard deviation σ is equal to 1 while the mean θ remains unknown. Our aim is to verify a null hypothesis $H : \theta = 4$ against the alternative $K : \theta \neq 4$ on the significance level $\gamma = 0.05$.

According to [6] for testing hypothesis on the mean $H : \theta = \theta_0$ against the alternative $K : \theta \neq \theta_0$ there exist a fuzzy test $\varphi : \mathscr{Y}_0 \to \mathbb{F}(\{0,1\})$ of a form

$$\varphi(Y_1,\ldots,Y_n) = \mu_{\Pi}(\theta_0)|0 + (1-\mu_{\Pi}(\theta_0))|1, \tag{11}$$

where $\Pi = \Pi(Y_1,\ldots,Y_n)$ is an appropriate fuzzy confidence interval for θ. In our statistical model Π is a fuzzy number with following α-cuts

$$\Pi_\alpha(Y_1,\ldots,Y_n) = [(\overline{Y})^L_\alpha - u_{1-\gamma/2}\frac{\sigma}{\sqrt{n}}, (\overline{Y})^L_\alpha + u_{1-\gamma/2}\frac{\sigma}{\sqrt{n}}], \tag{12}$$

where $\overline{Y}$ is a sample average and $u_{1-\gamma/2}$ is a quantile of order $1-\gamma/2$ from the standard normal distribution.

After simple calculations we obtain a fuzzy sample average $\overline{Y}$ given by α-cuts $(\overline{Y})_\alpha = [2+\alpha, 4-\alpha]$ and substituting $u_{1-\gamma/2} = u_{0.975} = 1.96$, $n=5$ and $\sigma = 1$ into (12) we conclude that $\Pi_\alpha(Y_1,\ldots,Y_5) = [1.1235+\alpha, 4.8765-\alpha]$. Combining this α-cut and $\theta_0 = 4$ with (11) we get $\varphi(Y_1,\ldots,Y_5) = 0.8765|0 + 0.1235|1$ which may be interpreted as "rather accept H". Now we will try to check how much is our conclusion robust to the particular shapes of the membership functions describing data.

Let us assume that the core model $\mathscr{Y}_0$ is given by these triangular membership functions while the extended model $\mathscr{Y}_{ext}$ would be given by a family of all possible fuzzy numbers having the same core and support as the original observations $Y_1,\ldots,Y_5$, i.e. the neighborhood $\kappa(\mathscr{Y}_i)$ is a family of all fuzzy numbers $\{Z \in \mathbb{FN}(\mathbb{R}) : \text{core}Z = \text{core}Y_i, \text{supp}Z = \text{supp}Y_i\}$.

Suppose ζ is the level of acceptance, i.e. $\zeta(\varphi(Y_1,\ldots,Y_5)) = \mu_{\Pi}(\theta_0)$, where Π is now evaluated for the extended model. It can be shown that the supremum of $\mu_{\Pi(Z_1,\ldots,Z_n)}(\theta_0)$ over all possible functions given by the extended model is obtained for the widest possible fuzzy confidence interval $\Pi' = [1.1235, 4.8765]$ while the infimum is reached for the most narrow possible fuzzy confidence interval $\Pi'' = [2.1235, 3.8765]$. Hence we get $SUS_\varphi(Y_1,\ldots,Y_5) = \mu_{\Pi'}(4) - \mu_{\Pi''}(4) = 1-0 = 1$. Since our extended model admits all possible membership functions we can also say that $MSUS_\varphi = 1$. Hence $R_\varphi = 0$, which means that our test turns out to be very sensitive to data representation.

Now let us consider how sensitive is our test under departures from the core model on a single observation only. The supremum of $\mu_{\Pi(Z_1,\ldots,Z_n)}(\theta_0)$ over single nontriangular membership function and four triangular is obtained for the fuzzy confidence interval Π' with α-cuts $\Pi'_\alpha = [1.1235+0.8\alpha, 4.8765-0.8\alpha]$ while the infimum is reached for the fuzzy confidence interval Π'' with α-cuts $\Pi''_\alpha = [1.3234+0.8\alpha, 4.6765-0.8\alpha]$. Hence we get $AS_\varphi(Y_1,\ldots,Y_5) = \mu_{\Pi'}(4) - \mu_{\Pi''}(4) = 1-0.8457 = 0.1543$ which produces $AR_\varphi = 0.8457$ and $MIR_\varphi = 0.8457$. Thus if we admit departures from the triangular membership function on a single observations only the conclusion proposed by our test is about 85% robust.

One may also ask whether robustness of the level of acceptance of our test depends on θ_0 in the null hypothesis $H: \theta = \theta_0$. It can be shown that our test is absolutely robust if $\theta_0 \in (2.1235, 3.8765)$ since then we always get $\mu_{\Pi}(\theta_0) = 1$ and if θ_0 is lower than 1.1235 or greater than 4.8765 for which we get $\mu_{\Pi}(\theta_0) = 0$. For other values of θ_0 we get $R_\varphi = 0$ provided we allow all possible departures from the triangular fuzzy numbers. If we consider departures from the core model on a single observation only for all possible values of the parameter θ_0 we obtain $0 \le AS_\varphi \le 0.25$.

It is worth noting that these seemingly so poor robustness of our test was obtained when all possible departures from the triangular membership functions were allowed. This way our core model has been enlarged too much because we have included into

$\mathscr{Y}_{ext}$ even such fuzzy sets which are quite far from the core model. Thus actually it is not so surprising that under such huge extended model we have obtained so strange results. Please note, that very similar situations also happen in classical statistics if we extend the core model too much (e.g., it was shown in [3], [4] and [13] that the size of any two-sample nonparametric test can come arbitrarily close to 1 when all kind of dependencies are allowed). It seems that for $\mathscr{Y}_{ext}$ closer of the core model the estimated robustness surely will not be so striking.

7 Conclusions

In the paper we have suggested some tools for describing and quantifying robustness to data representation. We do not claim the these tools are the most efficient ones. Actually the primary goal of that paper was rather to draw attention to the problem of robustness to data representation which seems to be very important in statistics for vague data. Presumably it would be difficult to construct such statistical procedures for fuzzy data that disregard completely the actual shape of the membership functions applied for modelling data. However, we may try to eliminate the impact of the particular form of membership functions as much as possible or even reduce it to the acceptable degree. Moreover, this aspect cannot substitute the traditional area of studies on robust statistics, i.e. distributional robustness, but should be considered in parallel. Therefore, the aim of the robust statistics for vague data is to derive statistical procedures which are both distribution-free and robust to the choice of the particular form of membership function describing data.

References

1. Box, G.E.P., Anderson, S.L.: Permutation theory in the derivation of robust criteria and study of departures from assumptions. J. Roy Statist. Soc. Ser. B 17, 1–34 (1955)
2. Denœux, T., Masson, M.H., Hébert, P.A.: Nonparametric rank-based statistics and significance tests for fuzzy data. Fuzzy Sets Syst. 153, 1–28 (2005)
3. Grzegorzewski, P.: The robustness against dependence of nonparametric tests for the two-sample location problem. Appl. Math. 22, 469–476 (1995)
4. Grzegorzewski, P.: Robustness against dependence of some nonparametric tests. Discuss. Math. Algebra Stochastic. Methods 15, 203–212 (1995)
5. Grzegorzewski, P.: Statistical inference about the median from vague data. Control Cybernet 27, 447–464 (1998)
6. Grzegorzewski, P.: Testing statistical hypotheses with vague data. Fuzzy Sets Syst. 112, 501–510 (2000)
7. Grzegorzewski, P.: Distribution-free tests for vague data. In: Lawry, J., Miranda, E., Bugarin, A., Li, S., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing, vol. 37, pp. 495–502. Springer, Berlin (2004)
8. Hampel, F.R.: A general qualitative definition of robustness. Ann. Math. Statist. 42, 1887–1896 (1971)
9. Huber, P.J.: Robust Statistics. Wiley Series in Probability and Statistics. Wiley, New York (1981)

10. Kwakernaak, H.: Fuzzy random variables, Part I: Definitions and theorems. Inform. Sci. 15, 1–15 (1978); Fuzzy random variables, Part II: Algorithms and examples for the discrete case. Inform. Sci. 17, 253–278 (1979)
11. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)
12. Zieliński, R.: Robustness: a quantitative approach. Bull. Acad. Polon. Sci. Ser. Math. Astr. Phi. 24, 1281–1286 (1977)
13. Zieliński, R.: Robustness of the one-sided Mann-Whitney-Wilcoxon test to dependency between samples. Statist. Probab. Lett. 10, 291–295 (1990)

Statistical Tests

On a Linear Independence Test for Interval-Valued Random Sets

Ángela Blanco[1,2], Ana Colubi[1], Norberto Corral[1], and Gil González-Rodríguez[2]

[1] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain
[2] European Centre for Soft Computing, Mieres, Spain

Abstract. The linear relationship between interval-valued random sets can arise in different ways. Recently, a linear model based on the natural arithmetic for intervals has been studied. In order to test whether the explanatory random set contributes significantly to explain the response random set through that linear model, an asymptotic testing procedure is here proposed. The empirical size of the test is illustrated by means of some simulations. The approach is also applied to a case-study.

1 Introduction

The linear regression problem between interval-valued random sets has been previously considered in the literature from different viewpoints (see, for instance, [4, 5, 6, 9], [8, 12]).

In [8] a linear regression model for compact and convex random sets based on a set-arithmetic approach has been established, and the estimators for the parameters have been obtained by applying the least-squares criterion based on a generalized L_2-type metric (see also [7]). In this communication we propose to complement those studies by proposing a linear independence test in the same context.

The organization of the paper is as follows. In Section 2 some preliminary concepts about interval-valued random sets and the considered linear regression model are presented. In Section 3 we suggest a test statistic for the linear independence. The asymptotic distribution of the statistic in some particular cases is used to state the asymptotic testing procedure. In Section 4 we show the results of some simulations in connection with the empirical significance level. The test is applied to a case-study in Section 5. Finally, in Section 6 some concluding remarks are commented.

2 Preliminaries

Let $\mathscr{K}_c(\mathbb{R})$ denote the class of nonempty compact intervals endowed with the natural interval-arithmetic induced by the Minkowski addition and the product by a scalar; namely, $A+B=\{a+b : a\in A, b\in B\}$ and $\lambda A=\{\lambda a : a\in A\}$, for all $A,B\in\mathscr{K}_c(\mathbb{R})$ and $\lambda\in\mathbb{R}$.

Due to the lack of symmetric element with respect to the addition, the space $(\mathscr{K}_c(\mathbb{R}),+,\cdot)$ is not linear, but semilinear, so it is useful to consider the *Hukuhara*

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 111–117, 2008.
springerlink.com

difference between A and B, defined as the interval C so that $A = B + C$ (if it exists) and denoted in this case as $C = A -_H B$ (see [11]). It is possible to assure the existence of $A -_H B$ if, and only if, $\inf A - \inf B \leq \sup A - \sup B$; moreover, in this case $A -_H B = [\inf A - \inf B, \sup A - \sup B]$.

The space $(\mathscr{K}_c(\mathbb{R}), +, \cdot)$ can be embedded onto a convex cone of the square integrable functions $\mathscr{L}(\mathbb{R})$ via the mapping $s : \mathscr{K}_c(\mathbb{R}) \longrightarrow \mathscr{L}(\mathbb{R})$ defined by $s(A) = s_A$ for all $A \in \mathscr{K}_c(\mathbb{R})$, where s_A denotes the *support function* of the interval A, namely, $s_A : \mathbb{R} \to \mathbb{R}$ such that $s_A(u) = \sup_{a \in A} \langle a, u \rangle$ for every $u \in \mathbb{R}$, $\langle \cdot, \cdot \rangle$ being the usual inner product on $\mathbb{R}$. The support function is semilinear, that is, $s_{A+B} = s_A + s_B$ and $s_{\lambda A} = \lambda s_A$, for $A, B \in \mathscr{K}_c(\mathbb{R})$ and $\lambda \geq 0$. Furthermore, if $A -_H B$ exists, then $s_{A -_H B} = s_A - s_B$. The function s allows us to deal with the space $\mathscr{L}(\mathbb{R})$, which can be endowed with an inner product which entails a Hilbertian structure.

The least square method considered in [8] for the estimation process is based on a generalized metric on $\mathscr{K}_c(\mathbb{R})$ via support functions (see [14]), which is defined for any $A, B \in \mathscr{K}_c(\mathbb{R})$ as

$$d_K(A,B) = \Big(\int_{\mathbb{S}^0} (s_A(u) - s_B(u))(s_A(v) - s_B(v)) dK(u,v) \Big)^{1/2}.$$

where $\mathbb{S}^0$ is the unit sphere in $\mathbb{R}$ and $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a positive definite and symmetric kernel such that $K(u,v) = K(-u,-v)$ for any $u, v \in \mathbb{S}^0$. The support function s is an isometry between $\mathscr{K}_c(\mathbb{R})$ and a cone of the Hilbert subspace $\mathscr{L}(\mathbb{S}^0) \subset \mathscr{L}(\mathbb{R})$ endowed with the generic L_2-type distance w.r.t. K. Thus, if $\langle \cdot, \cdot \rangle_K$ denotes the corresponding inner product, it is possible to express the d_K metric on $\mathscr{K}_c(\mathbb{R})$ as $d_K(A,B) = \langle s_A - s_B, s_A - s_B \rangle_K$.

Given a probability space $(\Omega, \mathscr{A}, P)$, a mapping $X : \Omega \to \mathscr{K}_c(\mathbb{R})$ is said to be an *interval-valued random set* associated with $(\Omega, \mathscr{A}, P)$ if the corresponding variables $\inf X$ and $\sup X$ are real random variables. It can be shown that this condition is equivalent to the $\mathscr{A}$-β_{d_H} measurability, where β_{d_H} denotes the σ-field generated by the topology induced by Hausdorff metric d_H on $\mathscr{K}_c(\mathbb{R})$. X can be also characterized by means of the random vector $(\mathrm{mid}X, \mathrm{spr}X)$ where $\mathrm{mid}X = (\sup X + \inf X)/2$ and $\mathrm{spr}X = (\sup X - \inf X)/2$ denote the mid-point and the spread of X, respectively.

If $E(|X|) < \infty$, where $|X|(\omega) = \sup\{|x| : x \in X(\omega)\}$ for any $\omega \in \Omega$, the *expected value of X in Kudō-Aumann's sense* (see [2]), is given by the expression

$$E(X) = \Big\{ E(f) | f : \Omega \to \mathbb{R}, f \in \mathscr{L}^1(\Omega), f \in X a.s.(P) \Big\}.$$

The expected value of an interval-valued random set is an element of $\mathscr{K}_c(\mathbb{R})$, that can be expressed in terms of the classical expectations of the real random variables $\inf X$ and $\sup X$ as $\big[E(\inf X), E(\sup X)\big]$. Furthermore, if $E(|X|^2) < \infty$, the *variance* of X is defined as $\sigma_X^2 = E\Big(\big(d_K(X, E[X])\big)^2 \Big)$ (see [10], [13]). It can be also expressed in terms of the inner product in $\mathscr{L}(\mathbb{S}^0)$ as $\sigma_X^2 = E\Big(\langle s_X - E(s_X), s_X - E(s_X) \rangle_K \Big)$. Finally, the covariance between two random sets X and Y can be defined via support functions as $\sigma_{X,Y} = E\Big(\big\langle s_X - E(s_X), s_Y - E(s_Y) \big\rangle_K \Big)$ whenever this expectation exists.

Let $X, Y : \Omega \longrightarrow \mathscr{K}_c(\mathbb{R})$ be two interval-valued random sets, and $\{X_i, Y_i\}_{i=1}^n$ a simple random sample obtained from (X, Y). The *sample mean* of X is defined by

$\overline{X} = (X_1 + X_2 + \ldots + X_n)/n$. It should be remarked that the Aumann expected value for a random set is coherent with the interval-arithmetic in the sense of the Strong Law of Large Numbers, which means that the preceding concept of sample mean converge a.s.-[P] to the Aumann expectation (see, for instance, [1]). The *sample variance* of X is given by $\hat{\sigma}_X^2 = \overline{d_K(X,\overline{X})^2}$ (analogously $\overline{Y}$ and $\hat{\sigma}_Y^2$). Finally, $\hat{\sigma}_{X,Y}$ denotes the *sample covariance* of X and Y, and it is defined as $\hat{\sigma}_{X,Y} = \overline{\langle s_X - s_{\overline{X}}, s_Y - s_{\overline{Y}} \rangle}_K$.

2.1 Simple Linear Regression Model

The *Simple Linear Regression Model* between X and Y on the basis of the interval-arithmetic approach is formalized as $Y = aX + \varepsilon$, where $a \in \mathbb{R}$ and $\varepsilon : \Omega \longrightarrow \mathscr{K}_c(\mathbb{R})$ is a random set such that $E(\varepsilon|X) = B \in \mathscr{K}_c(\mathbb{R})$ and $\sigma_{\varepsilon,X} = 0$ (see [9], [8]). The population *linear regression function* associated with this model is given by $E(Y|x) = ax + B$ for any $x \in \mathscr{K}_c(\mathbb{R})$.

The theoretical constants of the linear regression function can be expressed in terms of the moments of X and Y as $B = E(Y) -_H aE(X)$ and

$$a = \begin{cases} \dfrac{\sigma_{X,Y}}{\sigma_X^2} & \text{if } a \geq 0 \\ \dfrac{-\sigma_{-X,Y}}{\sigma_X^2} & \text{if } a \leq 0 \end{cases} \tag{1}$$

The estimates for the regression parameters have been obtained in [8]. In this communication we restrict ourselves to the case $a \geq 0$ as a first step. Note that in this way some of the difficulties that entail the lack of linearity of the space $\mathscr{K}_c(\mathbb{R})$ are avoided.

Following the ideas in [8] for the estimation process, we can obtain the corresponding estimates for the particular situation in which $a \geq 0$.

Let (X,Y) be two interval-valued random sets satisfying the considered linear model $Y = aX + \varepsilon$, with $a \geq 0$, and let $\{X_i, Y_i\}_{i=1}^n$ be a simple random sample obtained from (X,Y). Since $Y_i = aX_i + \varepsilon_i$, we have that $Y_i -_H aX_i$ exists for all $i = 1,\ldots,n$, then the estimator of a should be searched within the set

$$\widetilde{A} = \{c \geq 0 : \exists Y_i -_H cX_i, \text{ for all } i = 1 \ldots n\}. \tag{2}$$

The set of feasible solutions $\widetilde{A}$ can be represented by means of a non-empty compact real interval as $[0, \hat{a}^0]$, with $\hat{a}^0 \geq 0$.

The least squares estimation problem is expressed as

$$\text{Minimize } \tfrac{1}{n} \textstyle\sum_{i=1}^n d_K(Y_i, aX_i + B)^2$$

$$\text{subject to } a \in \widetilde{A}.$$

The solutions for this minimization problem, and then, the estimators for the regression model parameters, can be expressed in terms of moments of X and Y as

$$\widehat{a} = \min\left\{\hat{a}^0, \max\left\{0, \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}\right\}\right\} \tag{3}$$

and $\widehat{B} = \overline{Y} -_H \widehat{a}\overline{X}$.

3 Linear Independence Test

Let $X, Y : \Omega \to \mathscr{K}_c(\mathbb{R})$ be two interval-valued random sets such that $Y = aX + \varepsilon$, with $a \geq 0$ and $\varepsilon : \Omega \to \mathscr{K}_c(\mathbb{R})$ fulfilling $E(\varepsilon|X) = B \in \mathscr{K}_c(\mathbb{R})$ and $\sigma_{X,\varepsilon} = 0$.

The aim in this work is to develop a test to determine whether X contributes to explain Y through the linear model or not. Since we have assumed that $a \geq 0$, this is equivalent to test

$$\begin{array}{l} H_0 : a = 0 \\ H_1 : a > 0 \end{array} \tag{4}$$

In this work we propose testing H_0 by means of the statistic

$$T_n = \sqrt{n} \max\left\{0, \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}\right\} \tag{5}$$

Remark 1. From (3), the intuitive statistic for the test would be

$$\widetilde{T}_n = \sqrt{n} \min\left\{\hat{a}^0, \max\left\{0, \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}\right\}\right\}$$

because it uses the information given by the linear model. Unfortunately, the asymptotic behaviour of $\widetilde{T}_n$ is not easy to find, because the term $\hat{a}^0$ is difficult to handle. Nonetheless, given a significance level α and $k \geq 0$ such that $P(T_n > k|H_0) \to \alpha$ as $n \to \infty$, it is possible to check that $P(\widetilde{T}_n > k|H_0)$ is asymptotically lower or equal to α. Thus, the critical region $\{\widetilde{T}_n > k\}$ allows us to solve asymptotically the test (4) by using the statistic $\widetilde{T}_n$ with a significance level $\beta \leq \alpha$.

Remark 2. Both statistics $\widetilde{T}_n$ and T_n depend on $\hat{\sigma}_{X,Y}$, that converges almost-sure to zero under H_0. Indeed, since the random intervals X and Y are linear independent under H_0, then $\sigma_{X,Y} = 0$, and the strong consistency of the covariance guarantees the convergence.

If $0 < \sigma_X, \sigma_Y, \sigma_{X,Y} < \infty$, the asymptotic distribution of $\sqrt{n}\hat{\sigma}_{X,Y}$ under H_0 can be shown to be a normal distribution, with mean value 0 and variance σ_η, where η is the real-valued random variable defined as

$$\eta = \langle s_X - s_{E(X)}, s_\varepsilon - s_{E(\varepsilon)} \rangle_K.$$

Since the sample variance $\hat{\sigma}_X^2$ is consistent w.r.t. σ_X^2, by means of the Slutsky Theorem we obtain that $\sqrt{n}\hat{\sigma}_{X,Y}/\hat{\sigma}_X^2$ converges in law to a distribution $N(0, \sigma_\eta/\sigma_X^2)$.

Finally, since the function $\max\{0, \_\}$ is continuous, by means of the Continuous Function Theorem we can assure that T_n converges in law to the corresponding function of the normal distribution above, that is,

$$T_n \xrightarrow{\mathscr{L}} \max\left\{0, N(0, \sigma_\eta/\sigma_X^2)\right\}.$$

Remark 3. The population variance σ_X^2 is often unknown, so it would be necessary to estimate it by $\hat{\sigma}_X^2$ and then, the obtained asymptotic distribution corresponds to

$$\hat{\sigma}_X^2 T_n \xrightarrow{\mathscr{L}} \max\left\{0, N(0, \sigma_\eta)\right\}$$

For this reason, we could solve the test equivalently with the statistic

$$T_n' = \hat{\sigma}_X^2 T_n = \sqrt{n}\max\{0, \hat{\sigma}_{X,Y}\}$$

whose asymptotic distribution under H_0 does not depend on σ_X^2.

As a result, we can conclude that to test (4) at the nominal significance level α, H_0 should be asymptotically rejected whenever

$$T_n' > \max\{0, z_\alpha\}, \tag{6}$$

where z_α is the $100(1-\alpha)$ fractile of the normal distribution $N(0, \sigma_\eta)$.

Remark 4. In practice, the population variance σ_η^2 is usually unknown, so we should approximate this parameter by its estimator, $\hat{\sigma}_\eta^2$.

4 Simulation Studies

To illustrate the empirical behaviour of the asymptotic procedure suggested in Section 3, some simulations have been carried out. Let X and Y be two interval-valued random sets such that $\text{mid}X, \text{mid}Y \sim N(0,1)$, $\text{spr}X, \text{spr}Y \sim \chi_1^2$ are independent random variables.

Samples of intervals $\{(x_i, y_i)\}_{i=1}^n$ for different sizes n have been generated in order to apply the suggested test. We have developed two different tests. T_1' represents the theoretical test in which the variance of η is known, and T_2' denotes the test in which the population variance of η is aproximated by $\hat{\sigma}_\eta$. In Table 1 we present the percentage of rejections of H_0 at a significance level $\alpha = 0.05$ in 10,000 iterations for each different sample size and each test. The results indicate that the test T_2' is conservative. As expected, in both tests the empirical size is closer to the theoretical one as the sample size increases, although large sample sizes are required in order to obtain suitable results. In addition, T_1' seems to be more accurate than T_2', because T_1' uses the population information instead of the sample one.

Remark 5. In the case of dealing with small samples, asymptotic procedures do not apply. In these situations, alternative techniques should be developed in order to solve the linear independence test considered in this work. For instance, conditions to find the exact distribution of the statistic may be investigated. However, in general they mean

Table 1. Simulation results: empirical size at $\alpha = 0.05$

Sample size	T_1'	T_2'
100	5.25	4.56
200	5.24	4.6
500	5.18	4.62
1000	5.11	4.72
5000	5.03	4.75

the addition of important restrictions to the problem. Bootstrap procedures are another possible way to solve the test more widely applicable.

5 Case-Study: The Blood Pressure Data-Set

In order to show the application of the asymptotic procedure to test the linear independence, we have applied the suggested procedure to a real-life sample data set. Data have been previously used in some works (see, for instance, [6]). They have been supplied by the Hospital Valle del Nalón in Asturias (Spain), and correspond to the range of the systolic X and diastolic Y blood pressure over a day for 59 patients. In Table 2 some of the sample data are presented (full sample data set is available at [6]).

Table 2. Some data of the ranges of systolic (X) and diastolic (Y) blood pressure

X	11.8-17.3	10.4-16.1	13.1-18.6	10.5-15.7	12-17.9	10.1-19.4	...
Y	6.3-10.2	7.1-11.8	5.8-11.3	6.2-11.8	5.9-9.4	4.8-11.6	...

If we test the linear independence between X and Y by using the asymptotic test suggested in Section 3 at nominal significance level $\alpha = 0.05$, we obtain that the value of the typified statistic is $T^* = 6.027$, which is greater than $\max\{0, z_{0.05}\} = 1.645$. Thus, the null hypothesis should be rejected, and we conclude that there is a linear relationship between the fluctuation of the systolic and the diastolic blood pressure in terms of the model considered in this communication.

6 Concluding Remarks

In this communication, an asymptotic procedure for testing the linear independence between two interval-valued random sets by considering a particular case has been suggested. Furthermore, its suitability for large samples has been demonstrated by means of some simulations. It should be underlined that the results are not accurate for moderate and small sample sizes. We are analyzing currently other techniques, like bootstrap procedures, which are often better in these cases.

In the particular case we have analyzed, only positive coefficients for X have been considered. In this way some difficulties due to the lack of linearity of $\mathscr{K}_c(\mathbb{R})$ are avoided. We are also analyzing at present the general case.

Acknowledgement. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2006–07501 and by the Government of Asturias through the Innovation, Technology and Science Plan (PCTI) 2006–2009, PhD Grant BP06–010. Their financial support is gratefully acknowledged.

References

1. Arstein, Z., Vitale, R.A.: A strong law of large numbers for random compact sets. Ann. Probab. 5, 879–882 (1975)
2. Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. 12, 1–12 (1965)
3. Blanco, A., Colubi, A., González-Rodríguez, G., Corral, N.: On the consistency of the least-squares estimators for a linear regression model between interval-valued random sets. In: Abstracts of the 56th Session of the International Statistical Institute (ISI 2007, Lisbon, Portugal), p. 406 (2007)
4. De Carvalho, F.A.T., Lima Neto, E.A., Tenorio, C.P.: A new method to fit a linear regression model for interval-valued data. In: Biundo, S., Frühwirth, T., Palm, G. (eds.) KI 2004. LNCS (LNAI), vol. 3238, pp. 295–306. Springer, Heidelberg (2004)
5. Gil, M.A., López, M.T., Lubiano, M.A., Montenegro, M.: Regression and correlation analyses of a linear relation between random intervals. Test 10(1), 183–201 (2001)
6. Gil, M.A., Lubiano, M.A., Montenegro, M., López, M.T.: Least squares fitting of an affine function and strength of association for interval valued data. Metrika 56(2), 97–111 (2002)
7. González-Rodríguez, G., Blanco, A., Colubi, A., Lubiano, M.A.: Estimation of a simple linear regression model for fuzzy random variables. In: Abstracts of the 28th Linz Seminar on Fuzzy Set Theory: Fuzzy sets, Probabilities and Statistics - Gaps and Bridges (LINZ 2007, Linz, Austria), pp. 56–59 (2008)
8. González-Rodríguez, G., Blanco, A., Corral, N., Colubi, A.: Least squares estimation of linear regression models for convex compact random sets. Adv. Data Anal. Classif. 1(1), 67–81 (2007)
9. González-Rodríguez, G., Colubi, A., Coppi, R., Giordani, P.: On the estimation of linear models with interval-valued data. In: Proceedings of the 17th Conference of IASC-ERS (COMPSTAT 2006, Roma, Italy), pp. 697–704 (2006)
10. Körner, R., Näther, W.: On the variance of random fuzzy variables. In: Bertoluzza, C., Gil, M., Ralescu, D. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data, pp. 22–39. Physica-Verlag, Heildelberg (2002)
11. Hukuhara, M.: Intégration des applications measurable dont la valeur est un compact convexe. Funkcial Ekvac 10, 205–223 (1967)
12. Lima Neto, E.A., De Carvalho, F.A.T., Tenorio, C.P.: Univariate and multivariate linear regression methods to predict interval-valued features. In: Webb, G.I., Yu, X. (eds.) AI 2004. LNCS (LNAI), vol. 3339, pp. 526–537. Springer, Heidelberg (2004)
13. Lubiano, M.A., Gil, M.A., López-Díaz, M., López, M.T.: The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable. Fuzzy Sets Syst. 111(3), 307–317 (2000)
14. Näther, W.: On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. Metrika 51(3), 201–221 (2000)

A Small-Sample Nonparametric Independence Test for the Archimedean Family of Bivariate Copulas

Arturo Erdely[1] and José M. González-Barrios[2]

[1] Escuela de Actuaría, Universidad Anáhuac México Norte, Huixquilucan, México
[2] Departamento de Probabilidad y Estadística, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México D.F., México

Abstract. In this paper we study the problem of independence of two continuous random variables using the fact that there exists a unique copula that characterizes independence, and that such copula is of Archimedean type. We use properties of the empirical diagonal to build nonparametric independence tests for small samples, under the assumption that the underlying copula belongs to the Archimedean family, giving solution to an open problem proposed by Alsina et al. [2].

1 Introduction

A *bivariate copula* is a function $C:[0,1]^2 \to [0,1]$ with the following properties: For every u,v in $[0,1]$, $C(u,0)=0=C(0,v)$, $C(u,1)=u$ and $C(1,v)=v$, and for every u_1,u_2,v_1,v_2 in $[0,1]$ such that $u_1 \le u_2$ and $v_1 \le v_2$, $C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1) \ge 0$. Also, $W(u,v) \le C(u,v) \le M(u,v)$, where $W(u,v) := \max(u+v-1,0)$ and $M(u,v) := \min(u,v)$, where W and M are themselves copulas, known as the *Fréchet-Hoeffding lower and upper bounds,* respectively. The *diagonal section* of a bivariate copula, $\delta_C(u) := C(u,u)$, is a nondecreasing and uniformly continuous function on $[0,1]$ where: i) $\delta_C(0)=0$ and $\delta_C(1)=1$; ii) $0 \le \delta_C(u_2)-\delta_C(u_1) \le 2(u_2-u_1)$ for all u_1,u_2 in $[0,1]$ with $u_1 \le u_2$; iii) $\max(2u-1,0) \le \delta_C(u) \le u$. A copula C is said to be *Archimedean*, see [17], if $C(u,v)=\varphi^{[-1]}[\varphi(u)+\varphi(v)]$, where φ, called the *generator* of the copula, is a continuous, convex, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\varphi(1)=0$, and $\varphi^{[-1]}$ is the *pseudo-inverse* of φ given by: $\varphi^{[-1]}(t) := \varphi^{-1}(t)$ if $0 \le t \le \varphi(0)$, and $\varphi^{[-1]}(t) := 0$ if $\varphi(0) \le t \le \infty$. Its diagonal section is given by $\delta_C(u)=\varphi^{[-1]}[2\varphi(u)]$. One may ask, as observed in [6], given δ, what can be said about φ? The following result is part of what was proved in [9] and [3]:

Theorem 1. *If C is an Archimedean copula whose diagonal δ satisfies $\delta'(1-)=2$ then C is uniquely determined by its diagonal.*

From now on we will refer to the condition $\delta'(1-)=2$ as *Frank's condition.* An important example of an Archimedean copula that satisfies Frank's condition is the case of the product copula $\Pi(u,v)=uv$, which characterizes a couple of independent continuous random variables, via Sklar's Theorem [20], and so it is uniquely determined by its diagonal section $\delta_\Pi(u)=u^2$. Frank's condition is satisfied by 13 out of 22 copulas in the catalog of Archimedean copulas provided by [17].

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 118–125, 2008.
springerlink.com

2 The Empirical Diagonal and Some Properties

In the case of Archimedean bivariate copulas, the diagonal section contains all the information we need to build the copula, provided that Frank's condition $\delta'(1-)=2$ is satisfied, and in such case this leads us to concentrate in studying and estimating the diagonal. The main benefit of this fact is a reduction in the dimension of the estimation, from 2 to 1 in the case of bivariate copulas.

Let $S := \{(x_1,y_1),\ldots,(x_n,y_n)\}$ denote a sample of size n from a continuous random vector (X,Y). The *empirical copula* is the function C_n given by (see [17])

$$C_n\left(\frac{i}{n},\frac{j}{n}\right)=\frac{1}{n}\sum_{k=1}^{n}\mathbf{1}\left\{x_k\le x_{(i)},\,y_k\le y_{(j)}\right\},$$

where $x_{(i)}$ and $y_{(j)}$ denote the order statistics of the sample, for i and j in $\{1,\ldots,n\}$, and $C_n(\frac{i}{n},0)=0=C_n(0,\frac{j}{n})$. The domain of the empirical copula is the grid $\{0,1/n,\ldots(n-1)/n,1\}^2$ and its range is the set $\{0,1/n,\ldots,(n-1)/n,1\}$.

Remark 1. The domain of the empirical copula is just a rescaling of the set $\{0,1,\ldots,n\}$. Hence the empirical copula can be thought as equivalent to a discrete copula as noticed in [15] and [16]. Moreover, an empirical copula is an example of an irreducible discrete copula as defined in [13]. An empirical copula is not a copula, but a (two-dimensional) *subcopula*, for details of subcopulas see [17]. We should notice also the following relationship between the empirical copula and the empirical joint distribution function $H_n : C_n(\frac{i}{n},\frac{j}{n}) = H_n(x_{(i)},y_{(j)})$.

Definition 1. *The* empirical diagonal *is the function* $\delta_n(j/n) := C_n(j/n,j/n)$ *for* $j=0,1,\ldots,n$, *and* $\delta_n(0):=0$.

It is clear from above that δ_n is a nondecreasing function of j. Moreover, by Fréchet-Hoeffding bounds for subcopulas we have that $\max(2j/n-1\,,0)\le\delta_n(j/n)\le j/n$, and it is also straightforward to prove that the difference $\delta_n((j+1)/n)-\delta_n(j/n)$ equals one of the values $\{0,1/n,2/n\}$. These properties also follow from properties of the diagonal section in discrete copulas and quasi-copulas, see [1] or [14].

We will call an *admissible diagonal path* any path $\{\delta_n(j/n) : j=0,1,\ldots,n\}$ satisfying the Fréchet-Hoeffding bounds, that is any path between the paths $\{\max(2j/n-1,0) : j=0,1,\ldots,n\}$ and $\{j/n : j=0,1,\ldots,n\}$, with jumps of size $0,1/n$, or $2/n$ between consecutive steps. The proof of the following theorem is in [7]:

Theorem 2. *Let* $S=\{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ *be a random sample from the random vector of continuous random variables* (X,Y). *If* X *and* Y *are independent and if* $\mathbf{T}=(t_0=0,t_1,\ldots,t_{n-1},t_n=1)$ *is an admissible diagonal path, then*

$$\Pr\left[\mathbf{T}=(t_0=0,t_1,\ldots,t_{n-1},t_n=1)\right]=\frac{1}{n!}\prod_{j=1}^{n}f(j),$$

where, for $j=1,\ldots,n$: $f(j)=1$ *if* $n(t_j-t_{j-1})=0$; $f(j)=2(j-nt_{j-1})-1$ *if* $n(t_j-t_{j-1})=1$; *and* $f(j)=(j-1-nt_{j-1})^2$ *if* $n(t_j-t_{j-1})=2$.

3 A Nonparametric Test for Independence under the Archimedean Family of Bivariate Copulas

In this section we give solution to an open problem proposed in [2] and [3]:

> Can one design a test of statistical independence based on the assumptions that the copula in question is Archimedean and that its diagonal section is $\delta(u) = u^2$?

As a corollary of Sklar's Theorem, see [20, 19, 17], we know that if X and Y are continuous random variables, then X and Y are independent if and only if their corresponding copula is $C(u,v) = uv$. It is customary to use the notation $\Pi(u,v) := uv$, and to call it the *product* or *independence copula*. Recall that the product copula is Archimedean and it is characterized by the diagonal section $\delta_\Pi(u) = u^2$. If we are interested in analyzing independence of two continuous random variables, the previous results suggest to measure some kind of closeness between the empirical diagonal and the diagonal section of the product copula. Moreover, a nonparametric test of independence can be carried out, as suggested by [2, 21], using the diagonal section. Let (X,Y) be a random vector of continuous random variables with Archimedean copula C, then the following hypothesis are equivalent:

$$H_0 : X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad H_0^* : C = \Pi \quad \Leftrightarrow \quad H_0^{**} : \delta_C(u) = u^2. \tag{1}$$

Using the results of the previous sections, we wish to propose a statistical test based on the empirical diagonal because under H_0 we know the exact distribution of the empirical diagonal (Theorem 2) and so we could theoretically obtain the **exact distribution** of any test statistic based on it. A first idea would be to work with a Cramér-von Mises type test statistic based on the empirical diagonal:

$$CvM_n := \frac{1}{n-1}\sum_{j=1}^{n-1}\left(\delta_n\left(\frac{j}{n}\right) - \frac{j^2}{n^2}\right)^2, \tag{2}$$

rejecting H_0 whenever $CvM_n \geq k_\alpha$ for α a given test size. The performance of a test based on (2) will be analyzed later in a short simulation study. Under some Archimedean families, a test based on (2) can be improved under certain alternatives by the following idea: It is straightforward to verify that under H_0 the expectation $\mathbf{E}[\delta_n(j/n)] = \delta_\Pi(j/n) = j^2/n^2$ so we define for $j = 1,\ldots,n-1$ the quotient $\xi(j/n) := |\delta_n(j/n) - j^2/n^2| / (j/n - \max(2j/n - 1, 0))$ as a way of measuring pointwise closeness to independence, noticing that the denominator just standardizes dividing by the distance between the Fréchet-Hoeffding bounds at point j/n, in the spirit of a correction as in [4]. It is straightforward to verify that $0 \leq \xi(j/n) \leq \max(j/n, 1-j/n) \leq 1 - 1/n$. We propose as a test statistic

$$S_n := \frac{1}{n-1}\sum_{j=1}^{n-1}\xi\left(\frac{j}{n}\right), \tag{3}$$

rejecting H_0 whenever $S_n \geq k_1(\alpha)$, for α a given test size. Before we proceed, let us denote by $\delta_M(u) = u$ and $\delta_W(u) = \max(2u-1, 0)$ the upper and lower Fréchet-Hoeffding diagonal bounds, respectively. For u in $[0,1]$ the average distance between

$\delta_\Pi(u)$ and $\delta_M(u)$ is $1/6$ while the average distance between $\delta_\Pi(u)$ and $\delta_W(u)$ is $1/12$, this means that the diagonal that represents independence is, on average, twice closer to the lower than to the upper Fréchet-Hoeffding diagonal bound, thus independence is far from being in the middle of such bounds, and so we should consider the possibility of taking this into account in defining a test statistic. We define $h(j/n) := (j/n - j^2/n^2)\,/\,(j^2/n^2 - \max(2j/n-1,0))$ as a factor to be multiplied by $\xi(j/n)$ for those observations for which $\delta_n(j/n) < j^2/n^2$, in order to compensate somehow the non-equal closeness of the independence diagonal to the Fréchet-Hoeffding bounds. In other words, let us define $\nu(j/n) := h(j/n)\xi(j/n)$ if $\delta_n(j/n) < j^2/n^2$, and $\nu(j/n) := \xi(j/n)$ if $\delta_n(j/n) \geq j^2/n^2$.

We have that $h(j/n)$ is symmetric with respect to $1/2$ and that $1 \leq h(j/n) \leq h(1/n) = h(1-1/n) = n-1$. We now propose the following test statistic

$$A_n := \frac{1}{n-1} \sum_{j=1}^{n-1} \nu\left(\frac{j}{n}\right), \tag{4}$$

rejecting H_0 when $A_n \geq k_2(\alpha)$, for α a given test size. The test statistics (3) and (4) alone lead to biased tests of independence, but an appropriate combination of both leads to an approximately unbiased independence test, by using the decision rule

$$\text{reject } H_0 \text{ whenever } S_n \geq k_1 \text{ or } A_n \geq k_2\,, \tag{5}$$

where $\text{Prob}\left(\{S_n \geq k_1\} \cup \{A_n \geq k_2\}\,|\,H_0\right) \leq \alpha$, for k_1 and k_2 chosen appropriately, according to a given test size α. From their definitions it is immediate to verify that $0 < S_n \leq A_n \leq 3/4 - 1/4n$. Even though the election of (k_1,k_2) is not unique, in order to obtain an approximately unbiased test, a good choice for the alternative hypotheses we will consider is (k_1,k_2) such that $\alpha_1 = \Pr(S_n \geq k_1\,|\,H_0) \approx \Pr(A_n \geq k_2\,|\,H_0) = \alpha_2$. We cannot prove this in general for all possible alternatives since the power of the test for $\theta \neq \theta_0$ depends on the distribution under the alternative hypothesis, but it seems to work adequately in the following simulations for the given alternatives.

Since the main goal of the present work is to give solution to the open problem proposed by [2], building the required independence test, we include a short simulation study just to show that the proposed tests work, without pretending that they are extremely powerful, and we made some comparisons against a few well-known independence tests, without pretending that they constitute an exhaustive list of independence tests:

- Spearman's test, see [11].
- The modified Hoeffding test as introduced in [5].
- A test in [12].

The simulated power comparisons presented here were obtained with sample sizes $n = 15, 50$, $\alpha = 0.05$. Every Monte Carlo experiment reported here has been simulated 10,000 times, using some one-parameter Archimedean and Non-Archimedean copulas as alternatives. In both cases we will consider families of copulas $\{C_\theta\}$ with one-dimensional parameter θ such that there exists a unique θ_0 such that $C_{\theta_0} = \Pi$ or $\lim_{\theta\to\theta_0} C_\theta = \Pi$. The null hypothesis (1) becomes $H_0 : \theta = \theta_0$ versus the alternative $H_1 : \theta \neq \theta_0$.

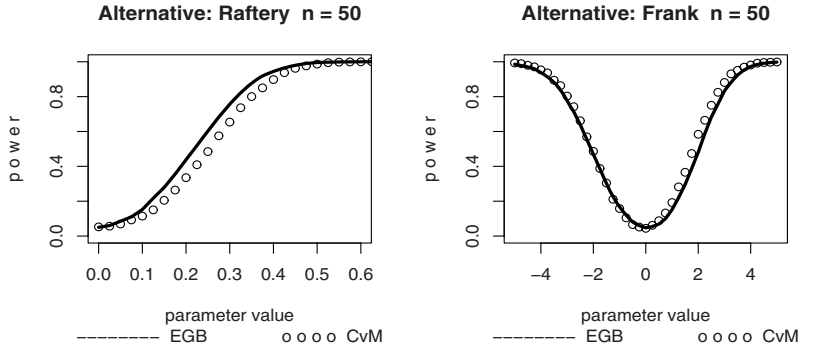


Fig. 1. Left: EGB vs CvM under Raftery. Right: EGB vs CvM under Frank

We will denote by CvM and EGB the tests proposed by the authors in (2) and (5), respectively. Under some families of copulas, there is a clear outperformance of EGB over CvM, for example, with the Raftery family as alternative; but under some other families it is almost the opposite, for example, with the Frank family as alternative, see Fig. 1. The proposed tests EGB and CvM will be compared against the already mentioned tests: R (Spearman), B ([5]), and V ([12]).

Archimedean alternatives. We compared the test powers for $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ under the following alternative families of Archimedean copulas, for details see [17]: Clayton, Frank, Nelsen's catalog number 4.2.7, Ali-Mikhail-Haq, and Gumbel-Barnett. In all cases these copulas satisfy $C_\theta = \Pi$ if and only if $\theta = 0$, or $\lim_{\theta \to 0} C_\theta = \Pi$, and satisfy Frank's condition $\delta'(1-) = 2$. For example, for the Clayton family see Fig. 2.

Non-Archimedean alternatives. An obvious question is what happens with the proposed EGB and CvM tests outside the Archimedean world. As proved in [10] it is possible to build copulas different from the product (or independence) copula $\Pi(u,v) = uv$

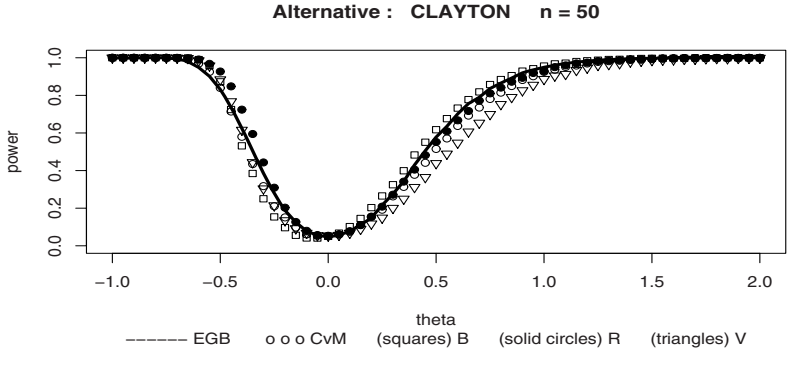


Fig. 2. All tests under Clayton

with the same diagonal as Π, but they are singular, and such copulas rarely appear in real problems. What really might be an issue for the proposed EGB and CvM tests is the fact that there are absolutely continuous non-Archimedean copulas which have the same diagonal as Π, as proved in [8], or as a consequence of the results in [18], so outside the Archimedean world the proposed EGB and CvM tests may face dependence structures that they will not be able to detect. Anyway, we performed similar simulation studies under some well-known non-Archimedean families of copulas, with surprising results. We compared the test powers for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ under the following alternative non-Archimedean families of copulas: Raftery, Cuadras-Augé, Farlie-Gumbel-Morgenstern, and Plackett (for details of these families see [17]). In all cases these copulas satisfy $C_\theta = \Pi$ if and only if $\theta = \theta_0$, or $\lim_{\theta \to 0} C_\theta = \Pi$, with $\theta_0 = 0$ for the first three families, and $\theta_0 = 1$ for the last one.

Summary of results. We made a summary of the power comparisons in the format suggested by [12]: For each test statistic, we have calculated the difference between the power of the test and the maximal power of the tests under consideration at the given alternative. For each graph this difference is maximized over the alternatives in the graph. This number can be seen as a summary for the behavior of the test in that graph, although of course some information of the graph is lost. In Table 1 we present percentage differences in maximal power of the five tests under comparison at various alternatives, so that the lower the difference number in the table, the better is the relative performance of the test.

Table 1. Relative power performance

n = 15	Alternative Copula	EGB	CvM	R	B	V
	Clayton	31	43	35	78	50
	Frank	40	37	34	75	54
	Nelsen 4.2.7	36	49	5	77	9
	Ali-Mikhail-Haq	43	37	33	76	55
	Gumbel-Barnett	24	45	13	78	44
	Raftery	19	29	29	5	31
	Cuadras-Augé	25	25	37	0	41
	Farlie-Gumbel-Morgenstern	48	37	32	77	57
	Plackett	42	38	33	73	53

n = 50	Alternative Copula	EGB	CvM	R	B	V
	Clayton	27	32	24	56	44
	Frank	42	27	24	50	52
	Nelsen 4.2.7	28	49	22	70	15
	Ali-Mikhail-Haq	40	28	24	50	53
	Gumbel-Barnett	20	33	8	58	42
	Raftery	4	31	32	20	34
	Cuadras-Augé	12	16	32	8	37
	Farlie-Gumbel-Morgenstern	44	26	25	51	53
	Plackett	40	26	18	43	49

In practice, when using a nonparametric test for independence we usually do not know what alternative we are dealing with, so what is valuable about a test is its ability to maintain an acceptable performance under different alternatives, rather than being the best under specific ones. In this sense, it seems that in general terms, the R test would be the best choice among the tests considered, followed by the EGB and CvM proposed tests.

4 Final Remark

If the underlying copula of a random vector (X,Y) is of the Archimedean type, independence tests can be carried out by defining appropriate test statistics based on the empirical diagonal. Such statistics are discrete random variables and their **exact distribution** may be obtained using Theorem 2, so no asymptotic approximations are required, which may be specially helpful with small samples.

References

1. Aguiló, I., Suñer, J., Torrens, J.: Matrix representation of discrete quasi-copulas. Fuzzy Sets Syst. (in press, 2008) doi:10.1016/j.fss2007.10.004
2. Alsina, C., Frank, M.J., Schweizer, B.: Problems on associative functions. Aequationes Math. 66, 128–140 (2003)
3. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific Publishing Co, Singapore (2006)
4. Anderson, T.W., Darling, D.A.: A test of goodness of fit. J. Amer. Statist. Assoc. 49, 765–769 (1954)
5. Blum, J.R., Kiefer, J., Rosenblatt, M.: Distribution-free tests of independence based on the sample distribution function. Ann. Math. Statist. 32, 485–498 (1961)
6. Darsow, W.F., Frank, M.J.: Associative functions and Abel-Schröder systems. Publ. Math. Debrecen. 30, 253–272 (1983)
7. Erdely, A.: Diagonal Properties of the Empirical Copula and Applications. Construction of Families of Copulas with Given Restrictions. Ph.D. Dissertation, Universidad Nacional Autónoma de México (2007)
8. Erdely, A., González-Barrios, J.M.: On the construction of families of absolutely continuous copulas with given restrictions. Comm. Statist. Theory Methods 35, 649–659 (2006)
9. Frank, M.J.: Diagonals of copulas and Schröder's equation. Aequationes Math. 51, 150 (1996)
10. Fredricks, G.A., Nelsen, R.B.: Copulas constructed from diagonal sections. In: Beneš, V., Štěpán, J. (eds.) Distributions with Given Marginals and Moment Problems, pp. 129–136. Kluwer Academic Publishers, Dordrecht (1997)
11. Hollander, M., Wolfe, D.A.: Nonparametric Statistical Methods. Wiley-Interscience, New York (1999)
12. Kallenberg, W.C.M., Ledwina, T.: Data-driven Rank tests for independence. J. Amer. Stat. Assoc. 94, 285–301 (1999)
13. Kolesárová, A., Mesiar, R., Mordelová, J., Sempi, C.: Discrete Copulas. IEEE Trans. Fuzzy Syst. 14, 698–705 (2006)
14. Kolesárová, A., Mordelová, J.: Quasi-copulas and copulas on a discrete scale. Soft Computing 10, 495–501 (2006)

15. Mayor, G., Suñer, J., Torrens, J.: Copula-like operations on finite settings. IEEE Trans. Fuzzy Syst. 13, 468–477 (2005)
16. Mesiar, R.: Discrete copulas - what they are. In: Montseny, E., Sobrevilla, P. (eds.) Joint EUSFLAT-LFA 2005 Conference Proceedings. Universitat Politecnica de Catalunya, Barcelona, pp. 927-930 (2005)
17. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
18. Nelsen, R.B., Quesada-Molina, J.J., Rodríguez-Lallena, J.A., Úbeda-Flores, M.: On the construction of copulas and quasi-copulas with given diagonal sections. Insurance Math. Econom. 42, 473–483 (2008)
19. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. North-Holland, New York (1983)
20. Sklar, A.: Fonctions de répartition à *n* dimensions et leurs marges. Publ. Inst.Statist. Univ. Paris 8, 229–231 (1959)
21. Sungur, E.A., Yang, Y.: Diagonal copulas of Archimedean class. Comm. Statist. Theory Methods 25, 1659–1676 (1996)

Defuzzification of Fuzzy p-Values

Inés Couso[1] and Luciano Sánchez[2]

[1] Dept. of Statistics and O.R., University of Oviedo, Oviedo, Spain
[2] Dept. of Computer Sciences, University of Oviedo, Gijón, Spain

Abstract. We provide a new description of the notion of fuzzy p-value, within the context of the theory of imprecise probabilities. The fuzzy p-value is viewed as a representation of a certain second-order possibility measure. According to Walley, any second-order possibility measure can be converted into a pair of lower and upper probabilities. Thus, we can convert the fuzzy p-value into an interval in the real line. We derive a construction of imprecise (but non fuzzy) tests, which are formally similar to recent tests used to manage with set-valued data.

Keywords: Imprecise probabilities, Hypothesis testing, Fuzzy p-value, Second-order possibility measure.

1 Introduction

Uncertainty about measurements arises naturally in a variety of circumstances (see [7] for a detailed description). This is the reason why the development of procedures for hypothesis testing with imprecise observations has recently gained increasing attention. When the data set contains intervals rather than points, we are not always able to take a clear decision about the null hypothesis. In the recent literature, imprecise tests are proposed to deal with such situations (see [7], for instance). According to this approach, an interval of upper and lower bounds of the critical value can be computed from the data set. When both bounds are on one side of the significance level, the decision (reject or accept) is clear. But when that interval and the significance threshold do overlap, we are not allowed to take a decision. In such situations, multi-valued test functions are defined. They can take the values $\{1\}$ (reject), $\{0\}$ (accept) and $\{0,1\}$ (undecided). This idea has been extended to the case of fuzzy-valued samples, under different approaches. Specifically, Filtzmoser & Viertl [8] and Denœux et al. [6] independently introduce the concept of fuzzy *p*-value. The concept of fuzzy test is then derived in a natural way by Denœux et al. [6]. But what should we do when a crisp decision is needed? They propose a particular defuzzification of the test output, in order to take a decision. Here we will propose an alternative construction, based on an interval-valued assignation for the critical level. We will justify why such defuzzification of the fuzzy p-value makes sense. We will show that it is in accordance with the possibilistic interpretation of fuzzy random variables developed in [3].

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 126–132, 2008.
springerlink.com

2 Fuzzy p-Values and Fuzzy Tests

2.1 Fuzzy p-Value Associated to a Fuzzy Random Sample

Let $X^* : \Omega \to \mathbb{R}$ be a random variable with distribution function F^* and let $\mathbf{X}^* = (X_1^*, \ldots, X_n^*) : \Omega^n \to \mathbb{R}^n$ be a simple random sample of size n from F^* (a collection of n iid random variables with common distribution F^*. They represent n independent observations of X^*.) Let now the Borel-measurable mapping $\varphi : \mathbb{R}^n \to \{0,1\}$ represent a non-randomized test for

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1.$$

Both hypotheses refer to a certain parameter of the df F^*. We will denote by R the critical region of φ, i.e., $R = \{\mathbf{x} \in \mathbb{R}^n : \varphi(\mathbf{x}) = 1\}$. Let $\sup_{\theta \in \Theta_0} E_\theta(\varphi(\mathbf{X})) = \sup_{\theta \in \Theta_0} P_\theta(\text{Reject } H_0)$ denote the size of the test φ. Suppose that for every $\alpha \in (0,1)$ we have a size α test φ_α with rejection region R_α and let $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)$ a realization of the sample. The p-value of $\mathbf{x}^*$ is defined as the quantity $\mathrm{p}_{\mathrm{val}}(\mathbf{x}^*) = \inf\{\alpha : \mathbf{x}^* \in R_\alpha\}$.

Let us now assume that we have got imprecise information about $\mathbf{x}^*$, and such imprecise information is given by means of a fuzzy subset of $\mathbb{R}^n$, $\tilde{\mathbf{x}} \in \mathscr{F}(\mathbb{R}^n)$. According to the possibilistic interpretation of fuzzy sets[1], $\tilde{\mathbf{x}}(\mathbf{x})$ represents the possibility grade that the "true" realization $\mathbf{x}^*$ coincides with the vector $\mathbf{x}$. Denœux et al. [6] and Filzmoser & Viertl [8] independently extend the concept of p-value, introducing the notion of fuzzy p-value. Each of those papers deals with a specific problem, but both definitions lead to the same general notion. We will call the *fuzzy p-value* of the fuzzy sample $\tilde{\mathbf{x}}$ to the fuzzy set $\widetilde{\mathrm{ext}}(\mathrm{p}_{\mathrm{val}})(\tilde{\mathbf{x}})$ determined by the membership function:

$$\widetilde{\mathrm{ext}}(\mathrm{p}_{\mathrm{val}})(\tilde{\mathbf{x}})(p) = \sup\{\tilde{\mathbf{x}}(\mathbf{x}) : \exists \mathbf{x} \in \mathbb{R}^n, \text{ with } \mathrm{p}_{\mathrm{val}}(\mathbf{x}) = p\}, \ \forall p \in [0,1]. \tag{1}$$

According to the possibilistic interpretation of fuzzy sets, the membership $\widetilde{\mathrm{ext}}(\mathrm{p}_{\mathrm{val}})(\tilde{\mathbf{x}})(p)$ represents the possibility grade of the equality $\mathrm{p}_{\mathrm{val}}(\mathbf{x}^*) = p$, according to the imprecise information we have about $\mathbf{x}^*$ described by $\tilde{\mathbf{x}}$. The last fuzzy set is closely related to the nested family of sets $(\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}}_\delta))_{\delta \in [0,1]}$ defined as follows:

$$\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}}_\delta) = \{\mathrm{p}_{\mathrm{val}}(\mathbf{x}) : \mathbf{x} \in \tilde{\mathbf{x}}_\delta\}, \ \forall \delta \in [0,1].$$

For some particular situations studied in [6] and [8], it is the family of $\delta-$cuts of $\widetilde{\mathrm{ext}}(\mathrm{p}_{\mathrm{val}})(\tilde{\mathbf{x}})$. In the general case, it is just a gradual representation of the fuzzy p-value. In other words, the membership function of $\widetilde{\mathrm{ext}}(\mathrm{p}_{\mathrm{val}})(\tilde{\mathbf{x}})$ can be derived from such nested family as follows:

$$\widetilde{\mathrm{ext}}(\mathrm{p}_{\mathrm{val}})(\tilde{\mathbf{x}})(p) = \sup\{\delta : p \in \mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}}_\delta)\}.$$

But we should assume some continuity properties to assure that $(\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}}_\delta))_{\delta \in [0,1]}$ is the family of $\delta-$cuts. In general, only the following relation holds:

$$[\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}})]_{\overline{\delta}} \subseteq \mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}}_\delta) \subseteq [\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}})]_\delta, \ \forall \delta,$$

where $[\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}})]_{\overline{\delta}}$ and $[\mathrm{p}_{\mathrm{val}}(\tilde{\mathbf{x}})]_\delta$ respectively denote the strong and the weak $\delta-$cut.

[1] We show in [2, 3] some specific situations where such a membership function is derived from an imprecise perception of some $\mathbf{x}^*$.

2.2 Fuzzy Test Associated to the Fuzzy p-Value

First of all, let us specify the meaning of the expression "fuzzy test" in our context: The null and the alternative hypotheses are referred to the distribution of the original random variable, F^*, so they are customary hypotheses in usual statistical problems. But the test is a fuzzy-valued function, i.e., it is a mapping that assigns, to each possible fuzzy sample $\tilde{\mathbf{x}} \in \mathscr{F}(\mathbb{R}^n)$, a fuzzy subset of $\{0,1\}$. That fuzzy subset reflects the possibility grades of rejection and acceptance of the null hypothesis, in accordance with the information provided by the fuzzy random sample. Some recent papers in the literature about statistics with imprecise data fit this formulation (see [6], for instance.) Let the reader notice that this approach is not related to other different works in the fuzzy statistics literature (see [9] for a detailed description), where the test functions are crisp, but they are referred to a certain parameter of the probability distribution induced by a fuzzy random variable on a certain σ-algebra of fuzzy events. This approach would not be useful in our context, where the frv represents the imprecise description of an otherwise standard random variable (see [1, 3, 4] for more detailed comments.)

In this paper, we will follow Denœux et al. [6] to construct a fuzzy test from a fuzzy p-value function. They specify the calculations for the Kendall and the Mann-Whitey-Wilcoxon tests. We will give here a more general description.

Let $(\varphi_\alpha)_{\alpha\in(0,1)}$ be a family of tests for H_0 against H_1, where $\varphi_\alpha : \mathbb{R}^n \to \{0,1\}$ is a test of size α, for each $\alpha \in (0,1)$. Let $\mathrm{p_{val}} : \mathbb{R}^n \to [0,1]$ and $\widetilde{\mathrm{ext}}(\mathrm{p_{val}}) : \mathscr{F}(\mathbb{R}^n) \to \mathscr{F}([0,1])$ respectively denote the crisp and the fuzzy p-value functions, in accordance with the formulae given in the last section. We can construct the fuzzy test $\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}$ from $\widetilde{\mathrm{ext}}(\mathrm{p_{val}})$ as follows:

$$\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})(1) = \sup\{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})(p) : p \le \alpha\}, \text{ and}$$
$$\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})(0) = \sup\{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})(p) : p > \alpha\}.$$

According to the interpretation of $\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})(p)$, the membership value $\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})(1)$ represents the possibility grade that $\mathrm{p_{val}}(\mathbf{x}^*)$ is less than or equal to α or, in other words, the possibility that $\mathbf{x}^*$ belongs to the rejection region. Similarly, $\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})(0)$ represents the possibility of accepting (no rejecting) the null hypothesis. Thus, $\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})$ represents a fuzzy decision. In the cases where a crisp decision is needed, this fuzzy subset may be defuzzified. Denœux et al. [6] suggest the following rule: rejecting the null hypothesis whenever $\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})(1) > \varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}(\tilde{\mathbf{x}})(\tilde{x})(0)$ and accepting (no rejecting) it otherwise. In Section 3, we will propose a different rule based on the theory of imprecise probabilities. First, we need to give an alternative description of the fuzzy p-value.

2.3 An Alternative Approach to the Concept of Fuzzy p-Value

Let us now give an alternative approach to the notion of fuzzy p-value. Let us first consider, for each particular realization $\mathbf{x} \in \mathbb{R}^n$, the Borel measurable mapping $D(\mathbf{x}) : \mathbb{R}^n \to \{0,1\}$ defined by:

$$D(\mathbf{x})(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathrm{p}_{\mathrm{val}}(\mathbf{y}) < \mathrm{p}_{\mathrm{val}}(\mathbf{x}) \\ 0 & \text{otherwise.} \end{cases}$$

$D(\mathbf{x})(\mathbf{y})$ takes the value 1 when the sample $\mathbf{y}$ is "less compatible" with the null hypothesis than $\mathbf{x}$ is. Thus, for a fixed $\mathbf{x} \in \mathbb{R}^n$, we have:

$$\sup_{\theta \in \Theta_0} P_\theta(D(\mathbf{x}) = 1) = \sup_{\theta \in \Theta_0} P_\theta(\{\mathbf{y} \in \mathbb{R}^n : \mathrm{p}_{\mathrm{val}}(\mathbf{y}) < \mathrm{p}_{\mathrm{val}}(\mathbf{x})\}).$$

Let us now remind that φ_α is assumed to be a test of size α, i.e.,

$$\sup_{\theta \in \Theta_0} E_\theta(\varphi_\alpha(\mathbf{X})) = \sup_{\theta \in \Theta_0} P_\theta(R_\alpha) = \alpha.$$

Hence, we can prove that $D(\mathbf{x})$ satisfies the equality:

$$\sup_{\theta \in \Theta_0} P_\theta(D(\mathbf{x}) = 1) = \mathrm{p}_{\mathrm{val}}(\mathbf{x}).$$

For the sake of simplicity, let us assume that the sizes of the $\alpha-$tests are associated to a certain value of the parameter $\theta_0 \in \Theta_0$, i.e., let us assume that:

$$\sup_{\theta \in \Theta_0} P_\theta(R_\alpha) = P_{\theta_0}(R_\alpha) = \alpha, \ \forall \alpha \in (0,1).$$

(The above condition holds, for instance, when the null hypothesis is simple and also for the most common unilateral and bilateral tests.) In that case, $D(\mathbf{x})$ is a Bernoulli random variable with parameter $\mathrm{p}_{\mathrm{val}}(\mathbf{x})$, under the distribution F_{θ_0}. In other words, $\mathrm{p}_{\mathrm{val}}(\mathbf{x}) = P_{\theta_0}(\{D(\mathbf{x}) = 1\})$, $\forall \mathbf{x} \in \mathbb{R}^n$. (The p-value of $\mathbf{x}$ represents the probability, under the null hypothesis, of getting a sample which is "less compatible" with H_0 than $\mathbf{x}$ is.) Let $\mathscr{X}$ represent the class of binary random variables that can be defined on $\mathbb{R}^n$ and let us now use the extension principle to extend $D : \mathbb{R}^n \to \mathscr{X}$ to $\mathscr{F}(\mathbb{R}^n)$. I.e., let us define the mapping $\widetilde{\mathrm{ext}}(D) : \mathscr{F}(\mathbb{R}^n) \to \mathscr{F}(\mathscr{X})$ as follows:

$$\widetilde{\mathrm{ext}}(D)(\tilde{\mathbf{x}})(Z) = \sup\{\tilde{\mathbf{x}}(\mathbf{x}) : D(\mathbf{x}) = Z\}, \ \forall Z \in \mathscr{X}.$$

Let us note that $\widetilde{\mathrm{ext}}(D)(\tilde{\mathbf{x}})$ is a possibility distribution over $\mathscr{X}$ and represents our imprecise information about $D(\mathbf{x}^*)$, according to our imprecise perception of the realization $\mathbf{x}^*$, represented by $\tilde{\mathbf{x}}$. More specifically, for each binary random variable $Z \in \mathscr{X}$, $\widetilde{\mathrm{ext}}(D)(\tilde{\mathbf{x}})(Z)$ represents the possibility grade that $D(\mathbf{x}^*)$ coincides with Z. Each binary random variable induces a Bernoulli distribution, $B(p)$. Thus, according to [3], we can derive a possibility distribution on the class of the Bernoulli measures. From now on, we will denote the class of all Bernoulli distributions by $\mathscr{P}_{\wp(\{0,1\})}$, since it is the class of probability measures that can be defined over $\wp(\{0,1\})$. This possibility measure, $\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}$, is determined by the possibility distribution $\boldsymbol{\pi}_{\tilde{\mathbf{x}}} : \mathscr{P}_{\wp(\{0,1\})} \to [0,1]$:

$$\boldsymbol{\pi}_{\tilde{\mathbf{x}}}(B(p)) = \sup\{D(\tilde{\mathbf{x}})(Z) : P_Z \equiv B(p)\}, \ \forall p \in [0,1].$$

In words, $\boldsymbol{\pi}_{\tilde{\mathbf{x}}}(B(p))$ represents the degree of possibility that the probability measure $B(\mathrm{p}_{\mathrm{val}}(\mathbf{x}^*))$ induced by $D(\mathbf{x}^*)$ coincides with $B(p)$. In other words, $\boldsymbol{\pi}_{\tilde{\mathbf{x}}}(B(p))$ represents the degree of possibility of the equality $\mathrm{p}_{\mathrm{val}}(\mathbf{x}^*) = p$. Mathematically,

$$\boldsymbol{\pi}_{\tilde{\mathbf{x}}}(B(p)) = \sup\{D(\tilde{\mathbf{x}})(Z) : P_Z \equiv B(p)\} = \sup\{D(\tilde{\mathbf{x}})(Z) : P(Z=1) = p\}$$
$$= \sup\{\tilde{\mathbf{x}}(\mathbf{x}) : P(D(\mathbf{x})=1) = p\} = \widetilde{\text{ext}}(\text{p}_{\text{val}})(\tilde{\mathbf{x}})(p),\ \forall p \in [0,1].$$

Summarizing, the fuzzy p-value is closely related to a certain second-order possibility measure [5]. Section 3 will be based on this alternative description of the fuzzy p-value.

3 Defuzzification of the Fuzzy p-Value

In Section 2.1 we have shown how the fuzzy p-value can be interpreted in terms of a second order possibility measure. In fact, $\widetilde{\text{ext}}(\text{p}_{\text{val}})(\tilde{\mathbf{x}})$ represents a possibility distribution over the class of possible values of the parameter of a Bernoulli random variable, and we have identified it with a second-order possibility measure $\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}$ defined over the class of all Bernoulli distributions. According to Section 2.1, $\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}$ and $\widetilde{\text{ext}}(\text{p}_{\text{val}})(\tilde{\mathbf{x}})$ are connected by the formula:

$$\widetilde{\text{ext}}(\text{p}_{\text{val}})(\tilde{\mathbf{x}})(p) = \boldsymbol{\pi}_{\tilde{\mathbf{x}}}(B(p)) = \boldsymbol{\Pi}_{\tilde{\mathbf{x}}}(\{B(p)\}) \tag{2}$$

According to Walley [10], any second-order possibility measure (which is an upper probability over the class of standard probabilities) can be reduced into a pair of upper and lower probabilities. Let us briefly describe Walley's procedure in our particular situation. We will consider the product space $\mathscr{P}_{\wp(\{0,1\})} \times \wp(\{0,1\})$ and:

- The possibility measure $\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}$ on $\mathscr{P}_{\wp(\{0,1\})}$. (In our particular problem, it represents our imprecise knowledge about the probability distribution of the random variable $D(\mathbf{x}^*)$.)
- The "transition probability" $\mathbb{P}_2^1 : \mathscr{P}_{\wp(\{0,1\})} \times \wp(\{0,1\}) \to [0,1]$ given by the formula:
$$\mathbb{P}_2^1(A,P) := P(A),\ \forall A \in \wp(\{0,1\}), P \in \mathscr{P}_{\wp(\{0,1\})}.$$
(It represents the following conditional probability information: if P were the true Bernoulli distribution associated to $D(\mathbf{x}^*)$, then the probability of occurrence of the event $D(\mathbf{x}^*) \in A$ should be $P(A)$. In particular, for $A = \{1\}$, and $P = B(p)$, the quantity $\mathbb{P}_2^1(\{1\}, B(p)) = p$ represents the probability of occurrence of the event $D(\mathbf{x}^*) = 1$ according to the conditional information "$D(\mathbf{x}^*)$ induces the probability measure $B(p)$".)

In this setting, Walley constructs, by means of natural extension techniques, an upper-lower joint model. Thus, the available information about the marginal distribution on the second space $\wp(\{0,1\})$ is described, in a natural way, by a pair of lower and upper probabilities, $\underline{P}_W$ and $\overline{P}_W$. In particular, $\underline{P}_W(\{1\})$ and $\overline{P}_W(\{1\})$ will represent the tightest bounds for the probability of the event $D(\mathbf{x}^*) = 1$ or, in other words, the tightest bounds for the p-value, $\text{p}_{\text{val}}(\mathbf{x}^*)$. To specify how this reduction is made, let us first recall that the second-order possibility measure $\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}$ can be identified with the class of second-order probability measures $\{\mathbb{P} : \mathbb{P} \leq \boldsymbol{\Pi}_{\tilde{\mathbf{x}}}\}$. If $\mathbb{P}$ were the "true" second-order probability that governs the "random"[2] experiment associated to the choice of the "true" Bernoulli

[2] Note that we are here interpreting the uncertainty associated to the perception of $\mathbf{x}^*$ as "randomness", since this imprecise perception is described by a possibility measure, which is, in turn, an upper probability.

distribution, then the probability of occurrence of the event $\{1\}$ (i.e., the "true" p-value) should be computed as follows (if we combine degrees of belief about events and about probabilities of events into the same model):

$$\int \mathbb{P}_2^1(\{1\},P)\, d\mathbb{P}(P) = \int P(\{1\})\, d\mathbb{P}(P).$$

Since all we know about $\mathbb{P}$ is that it is dominated by the possibility measure $\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}$, the lowest upper bound for the probability of occurrence of the event $D(\mathbf{x}^*) = \{1\}$ is determined by

$$\overline{P}_W(\{1\}) = \sup_{\mathbb{P}\leq\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}} \int \mathbb{P}_2^1(\{1\},P)\, d\mathbb{P}(P) = \sup_{\mathbb{P}\leq\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}} \int P(\{1\})\, d\mathbb{P}(P).$$

Similar arguments lead us to represent the highest lower bound of the probability by:

$$\underline{P}_W(\{1\}) = \inf_{\mathbb{P}\leq\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}} \int \mathbb{P}_2^1(\{1\},P)\, d\mathbb{P}(P) = \inf_{\mathbb{P}\leq\boldsymbol{\Pi}_{\tilde{\mathbf{x}}}} \int P(\{1\})\, d\mathbb{P}(P).$$

Thus, the Walley reduction allows us to convert the fuzzy p-value into the crisp interval $[\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}}), \overline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})] = [\underline{P}_W(\{1\}), \overline{P}_W(\{1\})]$. Furthermore, according to Walley [10], these upper and lower bounds can be alternatively computed as follows:

$$\overline{P}_W(\{1\}) = \int_0^1 \overline{P}_\delta(\{1\})\, d\delta, \quad \underline{P}_W(\{1\}) = \int_0^1 \underline{P}_\delta(\{1\})\, d\delta,$$

where, for each index, $\delta \in [0,1]$, $\overline{P}_\delta$ and $\underline{P}_\delta$ are defined as follows:

$$\overline{P}_\delta(\{1\}) = \sup\{Q(\{1\}) : Q \in \mathscr{P}_{\wp(\{0,1\})}, \boldsymbol{\Pi}_{\tilde{\mathbf{x}}}(\{Q\}) \geq \delta\} \text{ and}$$
$$\underline{P}_\delta(\{1\}) = \inf\{Q(\{1\}) : Q \in \mathscr{P}_{\wp(\{0,1\})}, \boldsymbol{\Pi}_{\tilde{\mathbf{x}}}(\{Q\}) \geq \delta\}.$$

Theorem 1

$$\overline{P}_\delta(\{1\}) = \sup[\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})]_\delta \;\; \textit{and} \; \underline{P}_\delta(\{1\}) = \inf[\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})]_\delta, \; \forall \delta \in [0,1].$$

According to the last theorem, the combination of first and second-order probabilities into the same model converts the fuzzy p-value, $\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})$ into the interval:

$$\overline{\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})} = [\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}}), \overline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})] = \left[\int_0^1 \inf[\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})]_\delta\, d\delta, \int_0^1 \sup[\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})]_\delta\, d\delta\right]. \quad (3)$$

The extreme points of such interval represent the most accurate bounds for the true p-value, $\mathrm{p_{val}}(\mathbf{x}^*)$, based on our imprecise knowledge of $\mathbf{x}^*$. Let us denote by $\varphi_{\overline{\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})}}$ the multi-valued $\alpha-$test associated to such interval

$$\varphi_{\overline{\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})}}(\tilde{\mathbf{x}}) = \begin{cases} \{0\} & \text{if } \underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}}) = \int_0^1 \inf[\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})]_\delta\, d\delta > \alpha \\ \{1\} & \text{if } \overline{\mathrm{p_{val}}}(\tilde{\mathbf{x}}) = \int_0^1 \sup[\widetilde{\mathrm{ext}}(\mathrm{p_{val}})(\tilde{\mathbf{x}})]_\delta\, d\delta \leq \alpha \\ \{0,1\} & \text{otherwise.} \end{cases}$$

The following relation between $\varphi_{\overline{\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})}}$ and the Denœux et al. [6] defuzzification of $\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}$ holds:

Theorem 2. $\mathrm{defuz_{DMH}}(\varphi_{\widetilde{\mathrm{ext}}(\mathrm{p_{val}})}) \subseteq \varphi_{\overline{\underline{\mathrm{p_{val}}}(\tilde{\mathbf{x}})}}$.

According to this result, the multi-valued test proposed in this paper is more times inconclusive than the Denœux et al. defuzzification is. I.e., whenever $\varphi_{\underline{p_{val}(\tilde{x})}}$ leads us to a clear decision (reject or accept the null hypothesis), $\text{defuz}(\varphi_{\widetilde{ext}(p_{val})})$ also leads to the same decision. But, for some fuzzy samples $\text{defuz}_{DMH}(\varphi_{\widetilde{ext}(p_{val})})$ is conclusive and $\varphi_{\underline{p_{val}(\tilde{x})}}$ is not. This could be viewed as an argument against the use of $\varphi_{\underline{p_{val}(\tilde{x})}}$. Nevertheless, it is not clear whether a higher number of inconclusive tests is a disadvantage or an improvement. The dependence between the degree of imprecision of the data-set and how many times a given test is inconclusive is not clear, and should be further studied in future works.

4 Concluding Remarks

We have proposed a new construction of crisp tests from fuzzy data, based on the theory of imprecise probabilities. The new tests are obtained as functions of the fuzzy p-values associated to the fuzzy samples, but they cannot be obtained as direct defuzzifications of the initial fuzzy tests.

Acknowledgement. This work has been supported by the FEDER-MEC Grants TIN2007-67418-C03-03, TIN2005-08386-C05-05 and MTM2007-61193.

References

1. Couso, I., Dubois, D., Montes, S., Sánchez, L.: On various definitions of the variance of a fuzzy random variable. In: De Cooman, G., Vejnarova, J., Zaffalon, M. (eds.) Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2007, Prague, Czech Republic), pp. 135–144. Action M Agency, SIPTA, Prague (2007)
2. Couso, I., Montes, S., Gil, P.: The necessity of the strong $\alpha-$cuts of a fuzzy set. Int. J. Uncertain Fuzziness Knowl-Based Syst. 9, 249–262 (2001)
3. Couso, I., Sánchez, L.: Higher order models for fuzzy random variables. Fuzzy Sets Syst 159, 237–258 (2008)
4. Couso, I., Sánchez, L.: Mark-recapture techniques in statistical tests for imprecise data (submitted for publication) (2008)
5. De Cooman, G.: A behavioural model for vague probability assessments. Fuzzy Sets Syst. 154, 305–358 (2005)
6. Denœux, T., Masson, M.H., Hébert, P.A.: Nonparametric rank-based statistics and significance test for fuzzy data. Fuzzy Sets Syst. 153, 1–28 (2005)
7. Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., Ginzburg, L.: Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. Technical Report SAND2007-0939. Sandia National Laboratories, Albuquerque, New Mexico (2007)
8. Filzmoser, P., Viertl, R.: Testing hypotheses with fuzzy data: The fuzzy p-value. Metrika 59, 21–29 (2004)
9. Körner, R.: Linear Models with Random Fuzzy Variables. PhD Dissertation, Technische Universität Bergakademie Freiberg (1997)
10. Walley, P.: Statistical inferences based on a second-order possibility distribution. Int. J. Gen. Syst. 26, 337–384 (1997)

Testing 'Two-Sided' Hypothesis about the Mean of an Interval-Valued Random Set

Manuel Montenegro, María Rosa Casals, Ana Colubi, and María Ángeles Gil

Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain

Abstract. Interval-valued observations arise in several real-life situations, and it is convenient to develop statistical methods to deal with them. In the literature on Statistical Inference with single-valued observations one can find different studies on drawing conclusions about the population mean on the basis of the information supplied by the available observations. In this paper we present a bootstrap method of testing a 'two-sided' hypothesis about the (interval-valued) mean value of an interval-valued random set based on an extension of the t statistic for single-valued data. The method is illustrated by means of a real-life example.

Keywords: Random interval, Interval mean, Hypothesis testing.

1 Introduction

In previous papers it has been pointed out that in many real-life situations observations are essentially (or customary) interval-valued rather than single-valued. For instance, some observations correspond to ranges or fluctuations (like price fluctuations, blood pressure fluctuations, income ranges, and so on), or they are engineering/physical data (as descriptions of amount, bounds, and limits, speed, mass, etc.), or interval-censoring times, or simply incomplete data which are treated as grouped ones.

In the last decade the interest for the statistical analysis of interval-valued data has increased, especially in which concerns descriptive aspects. In 2000 Billard and Diday [3] and Gil et al. [7] (see also [8], [9], [16], for a more detailed study) have considered different approaches for the regression (and also the correlation in the second one) analysis of interval-valued data: the symbolic data analysis and the random sets approach. The last approach has been also considered to deal with other descriptive problems (see, for instance, [13]).

An approach, which has been shown to be certainly valuable for the statistical management of interval-valued data, is the one based on the mid-spread (or centre-radium) approach and the use of interval arithmetic (see, for instance, Gil et al. [8], [9], and Marino and Palumbo [14]).

In some recent papers (cf. Montenegro et al. [17], Gil et al. [6], González-Rodríguez et al. [10]) we have developed some inferential procedures on the problems of least-squares regression and correlation between interval-valued random elements. The statistical analysis of these random elements has been developed by modelling them as particular random sets, using the set-valued arithmetic and a suitable metric between

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 133–139, 2008.
springerlink.com

interval-valued observations. This view allows us to capture the whole information on the considered problem and, hence, the approach means a sound and operational way to handle these data.

On the other hand, tests of the hypothesis that a population mean equals a particular value have been often studied when they refer to single-valued data. The Student's t statistics is the best known one when we handle data coming from a normal (or a nearly normal) population. In case of dealing with interval-valued data the assumption of normality does not make sense; more precisely, the few models for normally distributed random intervals become quite restrictive in practice, so it would be more realistic to look for asymptotic distribution free results or, even better, to develop bootstrap techniques.

In this paper, we introduce a bootstrap approach to testing about the interval-valued mean of an interval-valued random set in a population on the basis of a sample of interval-valued observations. The introduced test involves a statistic extending the well-known t, and particularizing to the interval case the test we have developed in previous papers for fuzzy-valued data (see [11], [15]). In fact, we have paid attention to express the new statistic in terms of the mids and spreads of intervals, and we conclude that the statistic can be viewed as a special convex linear combination of the squared Student statistic of the mids and the squared Student statistic of the spreads.

The result will be finally illustrated with a real-life example.

2 Preliminaries

In this paper we will assume that interval-valued observations are considered to be obtained from random mechanisms which are modelled by means of certain convex compact random sets.

Let $\mathcal{K}_c(\mathbb{R})$ be the class of nonempty compact intervals. $\mathcal{K}_c(\mathbb{R})$ can be endowed with a semilinear structure induced by the product by a scalar and the Minkowski addition from the usual interval arithmetic, that is,

$$I+I' = \left[\inf I + \inf I', \sup I + \sup I'\right], \quad \lambda \cdot I = \begin{cases} [\lambda \cdot \inf I, \lambda \cdot \sup I] \text{ if } \lambda \geq 0 \\ [\lambda \cdot \sup I, \lambda \cdot \inf I] \text{ if } \lambda < 0 \end{cases}$$

for $I, I' \in \mathcal{K}_c(\mathbb{R})$ and any $\lambda \in \mathbb{R}$.

To quantify the deviation/dissimilarity between the hypothetical interval-valued mean and the true one, we will make use of a metric on $\mathcal{K}_c(\mathbb{R})$ extending the Euclidean one, and being easy-to-use and interpret. More precisely, we will consider the *W-distance* on $\mathcal{K}_c(\mathbb{R})$ which is defined for two intervals $I, I' \in \mathcal{K}_c(\mathbb{R})$ as follows:

$$d_W(I,I') = \sqrt{\int_{[0,1]} [\mathrm{f}_I(\lambda) - \mathrm{f}_{I'}(\lambda)]^2 \, dW(\lambda)}$$

with $\mathrm{f}_I(\lambda) = \lambda \sup I + (1-\lambda)\inf I$ for all $\lambda \in [0,1]$, and W being formalized by means of a probability measure on the measurable space $([0,1], \mathscr{B}_{[0,1]})$ associated with a non-degenerate symmetric probability distribution on $[0,1]$ ($\mathscr{B}_{[0,1]}$ being the Borel σ-field on $[0,1]$).

The W-distance is a particularization of a metric recently introduced (although in a more general space) by Körner and Näther (see [12], and also [9]). The generalized distance d_W is equivalent to the generalized metric $d_{\overrightarrow{\lambda}}$ by Bertoluzza, Corral, and Salas [2], with $\overrightarrow{\lambda} = (\lambda_1, \lambda_2, \lambda_1)$, $\lambda_1 > 0, \lambda_2 = 1 - 2\lambda_1 \geq 0$,

$$d_{\overrightarrow{\lambda}}(I, I') = \sqrt{\int_{[0,1]} \left[\lambda_1 \left(\inf I - \inf I'\right)^2 + \lambda_2 \left(\operatorname{mid} I - \operatorname{mid} I'\right)^2 + \lambda_1 \left(\sup I - \sup I'\right)^2\right] dW(\lambda)}.$$

More precisely, the last generalized metric $d_{\overrightarrow{\lambda}}$ is indeed a particular case of d_W in which $W(0) = W(1) = \lambda_1$ and $W(.5) = \lambda_2$. Conversely, by choosing $\lambda_1 = 2\int_{[0,1]} \lambda^2 dW(\lambda) - .5$ and $\lambda_2 = 2 - 4\int_{[0,1]} \lambda^2 dW(\lambda)$ we can conclude that $d_{\overrightarrow{\lambda}} = d_W$, so that they correspond in fact to the same family of distances. It should be remarked that frequently choosing W on $[0,1]$ is more intuitive in practice than choosing $\overrightarrow{\lambda}$, whereas handling $\overrightarrow{\lambda}$ becomes easier and simpler than handling W, especially when we deal with simulation studies.

The measure W has no stochastic meaning, although we can formally deal with it in a probabilistic context and hence we can work with the probability space $([0,1], \mathscr{B}_{[0,1]}, W)$ and define $\sigma_W^2 = \int_{[0,1]} (2\lambda - 1)^2 dW(\lambda) = 2\lambda_1 \in (0,1]$. Due to the symmetry assumed for W, we can easily prove for arbitrary $I, I' \in \mathscr{K}_c(\mathbb{R})$ that

$$[d_W(I, I')]^2 = [\operatorname{mid} I - \operatorname{mid} I']^2 + \sigma_W^2 [\operatorname{spr} I - \operatorname{spr} I']^2$$

(with the mid and spread corresponding to the centre and radium of each interval, respectively). It can be concluded that the greater σ_W^2 the greater the influence of the Euclidean distance between the spreads of I and I' on $d_W(I, I')$, this influence attaining the maximum value 1 at a discrete W with $W(0) = W(1) = .5$, $= 0$ otherwise.

Given a probability space $(\Omega, \mathscr{A}, P)$, a mapping $X : \Omega \to \mathscr{K}_c(\mathbb{R})$ being d_W-Borel measurable is said to be an *interval-valued random set* (IVRS for short) associated with $(\Omega, \mathscr{A}, P)$. The concept of IVRS can be equivalently formalized in terms of the Borel σ-field generated by the topology induced by the well-known Hausdorff metric d_H on $\mathscr{K}_c(\mathbb{R})$. Borel-measurability guarantees that one can properly refer to concepts like statistical independence of IVRSs, distribution induced by an IVRS, and so on.

If $X : \Omega \to \mathscr{K}_c(\mathbb{R})$ is an interval-valued random set associated with $(\Omega, \mathscr{A}, P)$, and $E(\max\{|\inf X|, |\sup X|\}) < \infty$, the *mean value of* X (in Aumann's sense [1]) is defined as the compact interval

$$E^A[X] = [E(\inf X), E(\sup X)].$$

It should be emphasized that several arguments support considering d_W in the setting of this paper in contrast to the better known metric d_H. Thus,

- d_W is usually more operational and easy to compute and interpret than d_H in developing simple statistics (see, for instance, Blanco et al. [5]);
- Since d_W is an L^2-type metric, when one considers the extension of the least squares approach this metric is especially well-adapted; furthermore, $E^A[X]$ is the Fréchet-expectation when one considers the d_W metric, that is,

$$\min_{U\in\mathscr{K}_c(\mathbb{R})} E\left(\left[d_W(X,U)\right]^2\right) = E\left(\left[d_W(X,E^A[X])\right]^2\right),$$

whereas

$$\min_{U\in\mathscr{K}_c(\mathbb{R})} E\left(\left[d_H(X,U)\right]^2\right) \neq E\left(\left[d_H(X,E^A[X])\right]^2\right);$$

- Although W has not been fixed, one usually considers W to be the Lebesgue measure on $[0,1]$; however, the possibility of choosing different W will allow us to discuss whether such a choice affects the power of the test in Section 3, that is, to perform a sensitivity analysis with respect to the choice of W and look for the most suitable choices.

3 Bootstrap One-Sample Test about the Mean Value of an Interval-Valued Random Set

The aim of this section is introducing a statistic to test the null hypothesis that the mean value of an IVRS equals a specified compact interval on the basis of a sample of (interval-valued) observations from it. For this purpose we will particularize the bootstrap approach in [11].

Let $X:\Omega\to\mathscr{K}_c(\mathbb{R})$ be an IVRS associated with the probability space $(\Omega,\mathscr{A},P)$ and such that $\max\{|\inf X|,|\sup X|\}\in L^2(\Omega,\mathscr{A},P)$. Let $X_1,\ldots,X_n$ be IVRSs which are independent and identically distributed as X (i.e., a simple random sample from X), and let $X_1^*,\ldots,X_n^*$ be a bootstrap sample obtained from $X_1,\ldots,X_n$. On the basis of the result in [11], we can state that.

In testing the null hypothesis $H_0: E^A[X]=I\in\mathscr{K}_c(\mathbb{R})$ (which is equivalent to testing $H_0: d_W\left(E^A[X],I\right)=0$) at the nominal significance level $\alpha\in[0,1]$, H_0 should be rejected whenever

$$\pi_W\, t_{\text{mid}}^2 + (1-\pi_W)\, t_{\text{spr}}^2 > z_\alpha,$$

where z_α is the $100(1-\alpha)$ fractile of the bootstrap distribution of

$$T_n^* = \pi_W^*\,(t_{\text{mid}}^*)^2 + (1-\pi_W^*)\,(t_{\text{spr}}^*)^2$$

and with

$$t_{\text{mid}}^2 = \frac{\left(\overline{(\text{mid}\,X)}_n - \text{mid}\,I\right)^2}{\widehat{S_{\text{mid}}^2}},\quad t_{\text{spr}}^2 = \frac{\left(\overline{(\text{spr}\,X)}_n - \text{spr}\,I\right)^2}{\widehat{S_{\text{spr}}^2}},\quad \pi_W = \frac{\widehat{S_{\text{mid}}^2}}{\widehat{S_{\text{mid}}^2} + \sigma_W^2\,\widehat{S_{\text{spr}}^2}},$$

where

$$\overline{(\text{mid}\,X)}_n = \sum_{i=1}^n \text{mid}\,X_i/n,\quad \widehat{S_{\text{mid}}^2} = \sum_{i=1}^n \left[\text{mid}\,X_i - \overline{(\text{mid}\,X)}_n\right]^2/(n-1),$$

$$\overline{(\text{spr}\,X)}_n = \sum_{i=1}^n \text{spr}\,X_i/n,\quad \widehat{S_{\text{spr}}^2} = \sum_{i=1}^n \left[\text{spr}\,X_i - \overline{(\text{spr}\,X)}_n\right]^2/(n-1),$$

and the associated bootstrap estimates

$$(t^*_{\text{mid}})^2 = \frac{\left(\overline{(\text{mid}X)}_n - \overline{(\text{mid}X^*)}_n\right)^2}{\widehat{(S^2_{\text{mid}})^*}}, \quad (t^*_{\text{spr}})^2 = \frac{\left(\overline{(\text{spr}X)}_n - \overline{(\text{spr}X^*)}_n\right)^2}{\widehat{(S^2_{\text{spr}})^*}},$$

$$\pi^*_W = \frac{\widehat{(S^2_{\text{mid}})^*}}{\widehat{(S^2_{\text{mid}})^*} + \sigma^2_W \widehat{(S^2_{\text{spr}})^*}},$$

where

$$\overline{(\text{mid}X^*)}_n = \sum_{i=1}^{n} \text{mid}X^*_i/n, \quad \widehat{(S^2_{\text{mid}})^*} = \sum_{i=1}^{n} \left[\text{mid}X^*_i - \overline{(\text{mid}X^*)}_n\right]^2/(n-1),$$
$$\overline{(\text{spr}X^*)}_n = \sum_{i=1}^{n} \text{spr}X^*_i/n, \quad \widehat{(S^2_{\text{spr}})^*} = \sum_{i=1}^{n} \left[\text{spr}X^*_i - \overline{(\text{spr}X^*)}_n\right]^2/(n-1).$$

4 Illustrative Example

The following real-life example illustrates the application of the bootstrap test in Section 3 to data supplied by the Department of Nephrology of the Hospital Valle del Nalón in Langreo (Asturias, Spain). Data in Table 1 correspond to the "range of the pulse rate over a day" observed in a sample of 59 patients (suffering different types of illness) from a population of 3,000 who are hospitalized per year.

Values of X are obtained from several registers of the pulse rate of each patient measured at different moments (usually 60 to 70) over a concrete day. Pulse rate data are often collected by taking into account simply the lowest and highest registers during a day (actually, some devices used for this purpose only record and memorize these extreme values during a day); in these cases, the whole registers for a day and the associated variation can distort the information on the characteristic which is considered to be relevant: the range.

The interval $[60,100]$ is often assumed to be the adequate pulse rate fluctuation in a population of healthy adults. In testing the null hypothesis $H_0 : E^A(X) = [60,100]$ on the basis of the available sample information, we will apply the method in Section 3 (by considering 10,000 bootstrap iterations), and we conclude that:

- If we consider W to be the Lebesgue measure on $([0,1], \mathscr{B}_{[0,1]})$, the p-value of the test for the sample information is given by .0003,
- If we consider W to be the discrete measure weighting only the distances between the extreme points (i.e., $W(0) = W(1) = .5$) the p-value of the test for the sample information is given by .0010,
- If we consider W to be the discrete measure weighting uniformly the distances between the extreme points and the mid's (i.e., $W(0) = W(.5) = W(1) = 1/3$) the p-value of the test for the sample information is given by .0004,
- If we consider W to be the continuous measure associated with a beta distribution $\beta(2,2)$ (or, equivalently for purposes of defining d_W, W weighting 0 and 1 with weights equal to .1, and .5 with weight equal to .8), the p-value of the test for the sample information is given by .0004,

Table 1. Data on the ranges of pulse rate (X)

X		
58-90	54-78	56-133
47-68	53-103	37-75
32-114	47-86	61-94
61-110	70-132	44-110
62-89	63-115	46-83
63-119	47-83	52-98
51-95	56-103	56-84
49-78	71-121	54-92
43-67	68-91	53-120
55-102	62-100	49-88
64-107	52-78	75-124
54-84	55-84	58-99
47-95	61-101	59-78
56-90	65-92	55-89
44-108	38-66	55-80
63-109	48-73	70-105
62-95	59-98	40-80
48-107	59-87	56-97
26-109	49-82	37-86
61-108	48-77	

so that H_0 is scarcely sustainable (i.e., the range for the pulse rate of the people at the Nephrology Unit cannot be seriously claimed to coincide with that for healthy people), irrespectively of the considered measure W.

5 Concluding Remarks

In the above example we have developed a discussion on the p-value of the test for different choices of the measure W in the considered metric. This discussion suggests a more general one which could be developed in the future in connection with the effects of the choice of W on the conclusions of the test (that is a sensitivity analysis concerning the power of the test) as well as on the 'imprecision' (width) of the hypothetical interval (in this respect, the one considered in the example in Section 4 is rather wide).

On the other hand, another problem to be considered is that concerning one-sided hypothesis for the situation in this paper. The main problem should be that of formally stating the hypothesis, since there is no universally accepted total ordering on the space of intervals.

Acknowledgement. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2006–07501. This financial support is gratefully acknowledged.

References

1. Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. 12, 1–12 (1965)
2. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware Soft Comput. 2, 71–84 (1995)
3. Billard, L., Diday, E.: Regression analysis for interval-valued data. In: Kiers, H.A.L., Rasson, J.-P., Groenen, P.J.F., Schader, M. (eds.) Proceedings of the 7th Conference on the International Federation of Classification Societies (IFCS-2000, Namur, Belgium). Data Analysis, Classification and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization, pp. 369–374. Springer, Heidelberg (2000)
4. Billard, L., Diday, E.: Descriptive statistics for interval-valued observations in the presence of rules. Comput. Statist. 21(2), 187–210 (2006)
5. Blanco, A., Corral, N., González-Rodríguez, G., Lubiano, M.A.: Some properties of the d_K-variance for interval-valued random sets. In: Dubois, D., Lubiano, M.A., Prade, H., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Handling Variability and Imprecision, pp. 345–352. Springer, Heidelberg (2008)
6. Gil, M.A., González-Rodríguez, G., Colubi, A., Montenegro, M.: Testing linear independence in linear models with interval-valued data. Comput. Statist. Data Anal. 51(6), 3002–3015 (2007)
7. Gil, M.A., López, M.T., Lubiano, M.A., Montenegro, M.: A linear correlation coefficient for interval-valued random sets. In: Proceedings ot the 8th International Conference on Information Processing and Management of Uncertainty Conference (IPMU 2000, Madrid, Spain), pp. 104–110 (2000)
8. Gil, M.A., López, M.T., Lubiano, M.A., Montenegro, M.: Regression and correlation analyses of a linear relation between random intervals. Test 10(1), 183–201 (2001)
9. Gil, M.A., Lubiano, M.A., Montenegro, M., López, M.T.: Least squares fitting of an affine function and strength of association for interval-valued data. Metrika 56(2), 97–111 (2002)
10. González-Rodríguez, G., Blanco, A., Corral, N., Colubi, A.: Least squares estimation of linear regression models for convex compact random sets. Adv. Data Anal. Classif. 1(1), 67–81 (2007)
11. González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A.: Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst. 157, 2608–2613 (2006)
12. Körner, R., Näther, W.: On the variance of random fuzzy variables. In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data. Studies in Fuzziness and Soft Computing, vol. 87, pp. 22–39. Physica-Verlag, Heildelberg (2002)
13. López-García, H., López-Díaz, M., Gil, M.A.: Interval-valued quantification of the inequality associated with a random set. Statist. Probab. Lett. 46(2), 149–159 (2000)
14. Marino, M., Palumbo, F.: Interval arithmetic for the evaluation of imprecise data effects in least squares linear regression. Statistica. Applicata. 14(3), 277–291 (2002)
15. Montenegro, M., Colubi, A., Casals, M.R., Gil, M.A.: Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. Metrika 59, 31–49 (2004)
16. Montenegro, M., Gil, M.A., López, M.T., Lubiano, M.A.: Quantifying the linear correlation between two interval-valued random sets. In: Bouchon-Meunier, B., Gutiérrez-Ríos, J., Magdalena, l., Yager, R.R. (eds.) Technologies for Constructing Intelligent Systems 2 - Tools, pp. 99–108. Physica-Verlag, Heidelberg (2002)
17. Montenegro, M., González-Rodríguez, G., Colubi, A., Gil, M.A.: Bootstrap approach to test the linear independence between interval-valued random sets. In: López Díaz, M., Gil, M.A., Grzegorzewski, P., Hryniewicz, O., Lawry, J. (eds.) Soft Methodology and Random Information Systems. Advances in Soft Computing, pp. 431–438. Springer, Heidelberg (2004)

Asymptotic Tests for the Variance of a Fuzzy Random Variable Using the D_K-Metric

Ana Belén Ramos[1], Gil González-Rodríguez[1], Ana Colubi[2], and María Ángeles Gil[2]

[1] European Centre for Soft Computing, Mieres, Spain
[2] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain

Abstract. In this communication we present a procedure to test whether the variance of a fuzzy random variable (FRV) is a given value or not by using asymptotic techniques. The variance considered here is defined in terms of a generalized metric in order to quantify the variability of the fuzzy values of the FRV about its expected value. We present some simulations to show the empirical behavior of the test in different situations and an illustrative example to demonstrate its use in practice.

Keywords: Fuzzy random variable, Generalized metric between fuzzy numbers, Variance of a fuzzy random variable, Hypothesis testing.

1 Introduction

The concept of FRV, in Puri and Ralescu's sense [12], is an extension of the notion of random set. FRVs in this sense are models for random mechanisms associating fuzzy values with experimental outcomes. The fuzzy mean of a FRV has been introduced as a *fuzzy-valued* measure to summarize the "central tendency" of the variable (see [12]). In this communication, we consider a *real-valued* variance, defined by Körner and Näther in [9] and which is based on a generic $\mathscr{L}_2$-distance, to measure the dispersion or variability of the fuzzy values of a FRV about its mean.

Different statistical hypothesis testing problems based on fuzzy data have been studied previously in the literature. Regarding the one-sample test for the fuzzy mean, Körner [8] and Montenegro et al. [11] have developed the first asymptotic procedures. On the other hand, in [10] it has been analyzed the problem of testing hypotheses for the variance of a FRV in some particular cases.

Studies in [10] include results about hypothesis testing for the variance of simple FRVs (i.e., those taking on a finite number of different values). The techniques in [10] have been based on large samples theory and an operational metric on the space of fuzzy numbers with compact support introduced by Bertoluzza et al. [1]. In this communication we generalize these studies by considering a wider class of non-necesarily simple FRVs. The results are based on the D_K-metric defined by Körner and Näther (see [9]), and it should be noted that the generalization has required techniques completely different from those applied in [10].

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 140–146, 2008.
springerlink.com

In Section 2 we will introduce the concept of FRV, metric and variance that we will consider. In Section 3 the one-sample test for the variance will be stated and an asymptotic procedure will be proposed. The empirical size of the test will be illustrated by means of some simulations in Section 4. In Section 5 the approach will be exemplified through a case study and finally we will conclude with some remarks and future research directions.

2 Preliminaries

Consider the p-dimensional Euclidean space $\mathbb{R}^{\mathbf{p}}$ with the usual norm $\|\cdot\|$. Denote by $\mathscr{K}_c(\mathbb{R}^{\mathbf{p}})$ the class of the nonempty compact convex subsets of $\mathbb{R}^{\mathbf{p}}$ and by $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$ the class of the compact convex fuzzy sets on $\mathbb{R}^{\mathbf{p}}$, that is, $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}}) = \{U : \mathbb{R}^{\mathbf{p}} \to [0,1] \mid U_\alpha \in \mathscr{K}_c(\mathbb{R}^{\mathbf{p}}) \quad \forall \alpha \in [0,1]\}$ (where U_α denotes the α-level of the fuzzy set U for all $\alpha \in (0,1]$, and U_0 is the closure of the support of U).

The space $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$ can be endowed with a semilinear structure by means of the *sum* and the *product by a scalar* defined by applying Zadeh's extension principle [14]. This arithmetic agrees levelwise with the Minkowski sum and the product by a scalar (i.e., $(U \oplus V)_\alpha = U_\alpha + V_\alpha = \{u+v \,|\, u \in U_\alpha, v \in V_\alpha\}$ and $(\lambda U)_\alpha = \lambda U_\alpha = \{\lambda\, u \,|\, u \in U_\alpha\}$ for all $U, V \in \mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$, $\lambda \in \mathbb{R}$ and $\alpha \in [0,1]$). On the other hand, it will be useful to consider the *Hukuhara difference* of $U, V \in \mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$, $U -_H V$, which is defined (if it exists) as the element $W \in \mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$ such that $U = V \oplus W$.

A key element for the developments in [8] is the support function (see, for instance, Klement et al. [7]). The *support function mapping* can be defined as

$$s : \mathscr{F}_c(\mathbb{R}^{\mathbf{p}}) \longrightarrow \mathscr{L}(\mathbb{S}^{\mathbf{p-1}} \times [0,1])$$

in such a way that $s(U)$ is the support function of U, s_U, that is,

$$s_U(u,\alpha) = \sup_{w \in U_\alpha} \langle u, w \rangle, \quad u \in \mathbb{S}^{\mathbf{p-1}}, \alpha \in [0,1],$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^{\mathbf{p}}$ and $\mathbb{S}^{\mathbf{p-1}}$ denotes the unit sphere in $\mathbb{R}^{\mathbf{p}}$, that is, $\mathbb{S}^{\mathbf{p-1}} = \{u \in \mathbb{R}^{\mathbf{p}} \,|\, \|u\| = 1\}$.

The support function preserves the semi-linear structure of $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$, that is, if $U, V \in \mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$, $\lambda > \mathbf{0}$, then $s_{U+V} = s_U + s_V$, and $s_{\lambda U} = \lambda s_U$. Moreover, if $U, V \in \mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$ are so that the Hukuhara difference $U -_H V$ exists, it can be shown that $s_{U-_H V} = s_U - s_V$.

Several authors (see, for instance, [7] and [3]) have stated that isometries from $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$ onto a cone of the Lebesgue integrable functions $\mathscr{L}(\mathbb{S}^{\mathbf{p-1}} \times [0,1])$ can be built on the basis of several metrics by using the support function. In this context, Körner and Näther [9] consider a generalized family of metrics D_K on $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$, which are given by

$$D_K^2(U,V) = \int_{(\mathbb{S}^{(\mathbf{p-1})} \times [\mathbf{0},\mathbf{1}])^{\mathbf{2}}} (s_U(u,\alpha) - s_V(u,\alpha))(s_U(v,\beta) - s_V(v,\beta)) dK(u,\alpha,v,\beta),$$

where K is a definite positive and symmetric kernel.

The family of metrics D_K represents a generic L_2 distance on the Banach space $\mathscr{L}(\mathbb{S}^{\mathbf{p-1}} \times [0,1])$, whence each D_K induces an isometry between $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$ and the convex cone $s(\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})) \subset \mathscr{L}(\mathbb{S}^{\mathbf{p-1}} \times [0,1])$.

If $\langle\cdot,\cdot\rangle_K$ stands for the inner product in $\mathscr{L}(\mathbb{S}^{\mathbf{p-1}}\times[0,1])$ with respect to the kernel K and $\|\cdot\|_K$ is the norm associated with that inner product, the metrics D_K can be alternatively expressed as

$$D_K^2(U,V)=\langle s_U-s_V,s_U-s_V\rangle_K=\|s_U-s_V\|_K^2.$$

Given the probability space $(\Omega,\mathscr{A},\mathscr{P})$, a *fuzzy random variable* is a D_K-Borel measurable mapping $\mathscr{X}:\Omega\longrightarrow\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$. This definition is equivalent to the classical Puri & Ralescu's one [12] (see [2] and [9]).

Whenever $\|\mathscr{X}_0\|=\sup_{x\in\mathscr{X}_0}\|x\|\in\mathscr{L}^1(\Omega,\mathscr{A},\mathscr{P})$ we can define the *fuzzy mean* of $\mathscr{X}$, $E(\mathscr{X})$, or fuzzy expected value (see [12]), as the unique fuzzy set such that $(E(\mathscr{X}))_\alpha=$ Aumman's integral of the random set$(\mathscr{X})_\alpha$ for all $\alpha\in[0.1]$. If $p=1$ it can be shown that $(E(\mathscr{X}))_\alpha=[E(\inf(\mathscr{X}_\alpha),E(\sup(\mathscr{X}_\alpha)]$ for all $\alpha\in[0,1]$.

Moreover, if $E(\|\mathscr{X}_0\|^{\mathbf{2}})<\infty$, we can define the D_K-*variance* (or *variance* for short) of $\mathscr{X}$ (see [9]) as

$$\sigma^2_{\mathscr{X}}=E\Big(D_K^2(\mathscr{X},E(\mathscr{X}))\Big),$$

or, equivalently, in terms of the support function,

$$\sigma^2_{\mathscr{X}}=E\big(\langle s_{\mathscr{X}}-s_{E(\mathscr{X})},s_{\mathscr{X}}-s_{E(\mathscr{X})}\rangle_K\big).$$

Finally, to illustrate the empirical behavior of the test, *triangular fuzzy numbers* in $\mathbb{R}$ will be considered. This kind of fuzzy set is determined by 3 values: the center, the left spread and the right spread, and its alpha-cuts can be expressed as:

$$\big(T(l,c,r)\big)_\alpha=\big[c+l(\alpha-1),c+r(1-\alpha)\big]$$

where $c\in\mathbb{R}$ is the center, $l\in\mathbb{R}^+$ and $r\in\mathbb{R}^+$ are, respectively, the left and the right spread.

3 Stating the One-Sample Testing Problem for the Variance of a Fuzzy Random Variable

Given a simple random sample of n independent observations, $\mathscr{X}_1,\ldots,\mathscr{X}_n$, from a FRV $\mathscr{X}$, the aim of this communication is to test the null hypothesis $H_0:\sigma^2_{\mathscr{X}}=\sigma_0^2$ versus the alternative one $H_1:\sigma^2_{\mathscr{X}}\neq\sigma_0^2$ or, equivalently, testing $H_0:E\left(D_K^2(\mathscr{X},E(\mathscr{X}))\right)=\sigma_0^2$ versus $H_1:E\big(D_K^2(\mathscr{X},E(\mathscr{X}))\big)\neq\sigma_0^2$ for a given $\sigma_0\in\mathbb{R}^+$.

In the same way, we are interested in testing one-sided hypotheses for the variance of a FRV, that is, testing the null hypotheses $H_0:\sigma^2_{\mathscr{X}}\geq\sigma_0^2$ or $H_0:\sigma^2_{\mathscr{X}}\leq\sigma_0^2$.

In this setting the sample mean, $\overline{\mathscr{X}}_n=\frac{1}{n}(\mathscr{X}_1\oplus\cdots\oplus\mathscr{X}_n)$, will be a *fuzzy* estimator of $E(\mathscr{X})$ and the sample variance, $\widehat{\sigma}^2_{\mathscr{X}}=\dfrac{1}{n}\displaystyle\sum_{i=1}^n D_K^2(\mathscr{X}_i,\overline{\mathscr{X}}_n)$, will be the analogue *real-valued* estimator of $\sigma^2_{\mathscr{X}}$.

To test the considered null hypothesis the following statistic is proposed:

$$T_n=\frac{\sqrt{n}\left(\widehat{\sigma}^2_{\mathscr{X}}-\sigma_0^2\right)}{\sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^n\left(D_K^2(\mathscr{X}_i,\overline{\mathscr{X}}_n)-\widehat{\sigma}^2_{\mathscr{X}}\right)^2}}.$$

The properties of the support function and the metric D_K guarantee (see [13]) that

$$\left\{\sqrt{n}\left(\widehat{\sigma}^2_{\mathscr{X}}-\sigma_0^2\right)\right\}_n \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, E\left(\left[D_K^2(\mathscr{X}, E(\mathscr{X}))-\sigma_0^2\right]^2\right)\right).$$

In addition, it can be proved (see [13]) that

$$\frac{1}{n}\sum_{i=1}^{n}\left(D_K^2\left(\mathscr{X}_i, \overline{\mathscr{X}}_n\right)-\widehat{\sigma}^2_{\mathscr{X}}\right)^2 \xrightarrow{a.s.} E\left(\left[D_K^2(\mathscr{X}, E(\mathscr{X}))-\sigma_0^2\right]^2\right).$$

Therefore, $\{T_n\}_n \xrightarrow{\mathscr{L}} \mathscr{N}(0,1)$.

In this way, if the null hypothesis H_0 holds then the test statistic T_n is asymptotically normal.

On this basis, the testing procedure with asymptotic significance level α can be written as follows:

Asymptotic testing procedure:

a) Two-sided test
To test the null hypothesis $H_0 : \sigma^2_{\mathscr{X}} = \sigma_0^2$ against the alternative hypothesis $H_1 : \sigma^2_{\mathscr{X}} \neq \sigma_0^2$, H_0 should be rejected whenever $|T_n| > z_{(1-\alpha)/2}$, where $z_{(1-\alpha)/2}$ is the $[(1-\alpha)/2]$-quantile of the distribution $N(0,1)$. The p-value of this test is approximately given by $p = 2\left[1-\Phi(|T_n|)\right]$, where Φ is a c.d.f. of a r.v. $\mathscr{N}(0,1)$.

b) One-sided tests

(i) To test the null hypothesis $H_0 : \sigma^2_{\mathscr{X}} \geq \sigma_0^2$ against the alternative $H_1 : \sigma^2_{\mathscr{X}} < \sigma_0^2$, H_0 should be rejected whenever $T_n < z_\alpha$, where z_α is the α-quantile of the distribution $N(0,1)$. The p-value of this test is approximately given by $p = \Phi(T_n)$.

(ii) To test the null hypothesis $H_0 : \sigma^2_{\mathscr{X}} \leq \sigma_0^2$ against the alternative $H_1 : \sigma^2_{\mathscr{X}} > \sigma_0^2$, H_0 should be rejected whenever $T_n > z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution $N(0,1)$. The p-value of this test is approximately given by $p = 1-\Phi(T_n)$.

4 Simulation Studies

In this section we will show that in order to apply in practice the asymptotic one sample test for the variance of a FRV using the D_K metric we need at least moderate or large sample sizes.

We have considered a triangular FRV $\mathscr{X}$ with the left spread behaving as the χ^2_3 random variable, center varying as the $N(1,2)$ random variable, and the right spread behaving as the χ^2_8 random variable. In this section, the generalized distance has been chosen to be the Bertoluzza et al. one [1] with the Lebesgue measures on $[0,1]$. The variance of $\mathscr{X}$ is, approximately, $\sigma^2_{\mathscr{X}} = 6.4437$, so the null hypotheses we are going to consider are $H_0 : \sigma^2_{\mathscr{X}} = 6.4437$, $H_0 : \sigma^2_{\mathscr{X}} \geq 6.4437$ and $H_0 : \sigma^2_{\mathscr{X}} \leq 6.4437$. We have carried out 100000 simulations of the test (which implies a sample error of 0.002 at most with a confidence of 95%) using the previous asymptotic testing procedures for two-sided and one-sided tests at some significance levels (β). The results for different sample sizes n are gathered in Table 1.

Table 1. Empirical percentage of rejections under H_0

	$H_0 : \sigma^2_{\mathscr{X}} = 6.4437$			$H_0 : \sigma^2_{\mathscr{X}} \geq 6.4437$			$H_0 : \sigma^2_{\mathscr{X}} \leq 6.4437$		
$n \setminus 100 \cdot \beta$	1	5	10	1	5	10	1	5	10
50	0.13	1.84	5.30	6.65	13.72	19.56	5.12	10.58	15.63
100	0.20	2.39	6.18	4.30	10.84	16.44	3.18	7.84	12.97
500	0.45	3.50	8.08	2.11	7.17	12.45	1.50	5.53	10.79
1000	0.56	3.78	8.37	1.53	6.29	11.70	1.44	5.59	10.20
5000	0.80	4.67	9.26	1.27	5.67	10.69	1.07	5.02	10.05
10000	0.84	4.77	9.45	1.13	5.57	9.95	0.98	4.94	9.89

Table 1 shows that when $n \geq 1000$, the empirical percentage of rejections is quite close to the nominal significance level. It means that in order to apply the asymptotic procedure proposed in this communication, large sample sizes are required.

5 Illustrative Examples

The days of certain month are classified in accordance with their temperature (see [10]). The classes considered in [10] correspond to linguistic "values" like COLD, COOL, NORMAL, WARM and HOT. According to this classification, the type of day in a given area during July could be viewed as a fuzzy random variable $\mathscr{X}$ whose values are the preceding linguistic ones, which could be identified by means of some fuzzy numbers like those with support contained in [8,40] (measured in ^{o}C). We are going to described them in terms of triangular fuzzy sets represented in Fig. 1.

Example 1. Firstly, we consider the population Ω_1 of the days of July (31 days) during 30 years. Assume that in this period there was 151 COLD, 175 COOL, 199 NORMAL,

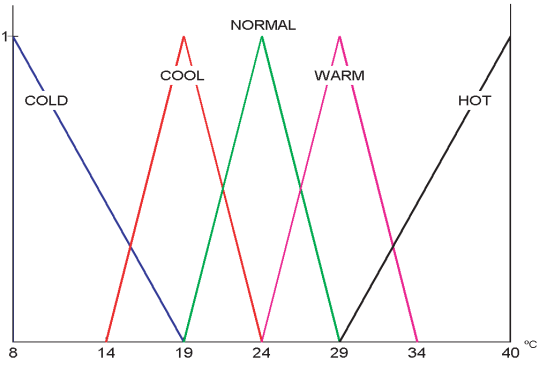


Fig. 1. The variable temperature at triangular fuzzy regions

224 WARM, and 181 HOT days and that a weatherman considers typical a standard deviation in the "temperature" lower than or equal to 8.66 degrees. Then, to test whether the standard deviation in Ω_1 satisfies such a condition, we can consider the variance and the test in Section 3. The estimate of the population variance of $\mathscr{X}$ in Ω_1 is $\widehat{\sigma}^2_{\mathscr{X}} = 73.8261$ and the corresponding p-value is 0.9922, whence the hypothesis of typical variability cannot be discarded at the usual significance levels.

Example 2. Assume now that the weatherman considers that the deviation in the "temperature" should be equal to 8.66 degrees. We can use again the corresponding test in Section 3 to obtain a p-value equal to 0.0156. Then the weatherman's hypothesis should be rejected at the significance levels 0.05 and 0.1, but it may be accepted at the significance level 0.01.

6 Concluding Remarks

The main advantage of the asymptotic tests for the variance introduced in this communication compared with previous works in the literature is that these techniques can be applied to non-necessary simple FRVs. Moreover, this test can be used for fuzzy data in $\mathscr{F}_c(\mathbb{R}^{\mathbf{p}})$, and not only for fuzzy numbers in $\mathscr{F}_c(\mathbb{R})$.

Although this procedure can be easily applied to large samples, the asymptotic results show that the use of bootstrap techniques for smaller or moderate sample sizes could be more efficient, as we have proposed in [13].

In addition, the statistic proposed in this communication can be compared with the classical one (established in terms of a quotient instead of a difference as we proposed here) in order to analyze similarities and differences.

The theoretical results developed previously are mainly focussed on the significance level (type I error), so an interesting open problem in connection with this subject is the study of the power function (type II error) associated with the test in order to establish its capability. We are currently working on this point as well as on the problem of testing the equality of the variances of two or more FRV's.

Acknowledgement. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2006–07501. Its financial support is gratefully acknowledged.

References

1. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware Soft. Comput. 2, 71–84 (1995)
2. Colubi, A., Domínguez-Menchero, J.S., López-Díaz, M., Ralescu, R.: A $D_E[0,1]$ representation of random upper semicontinuous functions. Proc. Amer. Math. Soc. 130, 3237–3242 (2002)
3. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc., River Edge (1994)
4. Gil, M.A., Montenegro, M., González-Rodríguez, G., Colubi, A., Casals, M.R.: Bootstrap approach to the multi-sample test of means with imprecise data. Comput. Statist. Data Anal. 51, 148–162 (2006)

5. Giné, E., Zinn, J.: Bootstraping general empirical measures. Ann. Probab. 18, 851–869 (1990)
6. González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A.: Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst. 157(19), 2608–2613 (2006)
7. Klement, E., Puri, M.L., Ralescu, D.A.: Limit theorems for fuzzy random variables. Proc. R. Soc. Lond A. 407, 171–182 (1986)
8. Körner, R.: An asymptotic α-test for the expectation of random fuzzy variables. J. Statist. Plann. Inference 83, 331–346 (2000)
9. Körner, R., Näther, W.: On the variance of random fuzzy variables. In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data. Studies in Fuzziness and Soft Computing, vol. 87, pp. 22–39. Physica-Verlag, Heildelberg (2002)
10. Lubiano, M.A.: Medidas de Variación para Elementos Aleatorios Imprecisos. PhD Thesis, University of Oviedo (1999)
11. Montenegro, M., Colubi, A., Casals, M.R., Gil, M.A.: Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. Metrika 59, 31–49 (2004)
12. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)
13. Ramos, A.B., González-Rodríguez, G., Gil, M.A., Colubi, A.: One-sample bootstrap tests for the variance of a fuzzy random variable. (submitted for publication) (2008)
14. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part 1. Inform. Sci. 8, 199–249 (1975); Part 2. Inform. Sci. 8, 301–353; Part 3. Inform. Sci. 9, 43–80

Empirical Results Concerning a Fuzzy-Based Measure of Symmetry of Real Random Variables

M. Asunción Lubiano[1], Ana Colubi[1], and Gil González-Rodríguez[2]

[1] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain
[2] European Centre for Soft Computing, Mieres, Spain

Abstract. A new measure of skewness for real-valued random variables based on fuzzy tools has been recently introduced. The measure is derived from certain fuzzy representations of real-valued random variables which can be used to characterize the distribution of the original variables through the *expected value* of the 'fuzzified' random variables. In this communication, the empirical behaviour of an asymptotic testing procedure for the symmetry of real random variables based on this approach will be examined by simulating different distributions. Some advantages of the proposed test will be illustrated by means of these simulations.

1 Introduction

In [8] and [3] it was shown the possibility of testing about the goodness-of-fit and the equality of distributions of real-valued random variables through some testing procedures about fuzzy expected values of certain fuzzifications. Specifically, it was shown that this approach presents some advantages; among them, a good average empirical behaviour in comparison with the usual traditional techniques (like the χ^2, Kolmogorov-Smirnov, and so on).

On the other hand, a new skewness measure based on the fuzzifications in [8] was proposed in [7]. In this communication, the aim is focused on the empirical analysis of an asymptotic testing procedure concerning this new measure.

The most commonly used tests for symmetry about a known value for continuous univariate distributions are based on either linear rank statistics (see [12], [6]), or some empirical distribution/density function estimators (see [2], [15]). On the other hand, some tests based on the likelihood ratio are available for discrete distributions (see [5]). It should be noted that most of these tests are not consistent against any nonsymmetric alternative, although they are asymptotically suitable for detecting alternatives in particular families of distributions.

The asymptotic test that will be empirically analyzed in this communication is valid for continuous and discrete distributions, can be used under very mild conditions and is consistent against any nonsymmetric alternative (see [11]). We will show that moderate/large sample sizes are required to obtain suitable empirical sizes and that the asymptotic power is similar or better than the corresponding to that of other methods in many cases.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 147–154, 2008.
springerlink.com

Since the approach is based on the expected value of a fuzzy random variable, in Section 2 we include some preliminaries concerning these random elements and the characterization of real distributions through a fuzzy set. In Section 3 we will present the measure of the symmetry of a random variable about a known value. The performance of the test will be illustrated with some simulation studies in Section 4. Finally, some relevant conclusions and future directions related to the study developed in this paper will be commented.

2 Preliminaries

We will consider the class $\mathscr{F}_c(\mathbb{R})$, which contains fuzzy sets $U:\mathbb{R}\to[0,1]$ whose α-levels are nonempty compact intervals of $\mathbb{R}$, that is, $U_\alpha\in\mathscr{K}_c(\mathbb{R})$ for all $\alpha\in[0,1]$, where $U_\alpha=\{x\in\mathbb{R}\,|\,U(x)\geq\alpha\})$ for all $\alpha\in(0,1]$, and U_0 is the closure of the support of U. Zadeh's extension principle [17] allows us to endow the space $\mathscr{F}_c(\mathbb{R})$ with a sum and a product by a scalar satisfying that

$$(U+V)_\alpha=U_\alpha+V_\alpha=\{u+v\,|\,u\in U_\alpha,v\in V_\alpha\},\quad (\lambda U)_\alpha=\lambda U_\alpha=\{\lambda u\,|\,u\in U_\alpha\}$$

for all $U,V\in\mathscr{F}_c(\mathbb{R})$, $\lambda\in\mathbb{R}$ and $\alpha\in[0,1]$. The space $(\mathscr{F}_c(\mathbb{R}),+,\cdot)$ is not linear.

$\mathscr{F}_c(\mathbb{R})$ can be embedded onto a convex cone of the squared integrable functions $\mathscr{L}(\{-1,1\}\times[0,1])$ by means of the support function (see [13], [4]). The support function of $U\in\mathscr{F}_c(\mathbb{R})$ is defined so that $s_U(u,\alpha)=\sup_{w\in U_\alpha}\langle u,w\rangle$ for any $u\in\{-1,1\}$ and $\alpha\in[0,1]$, where $\langle\cdot,\cdot\rangle$ denotes the usual inner product in $\mathbb{R}$.

For different statistical studies concerning imprecise random elements, the distance D_K introduced by [14] is especially valuable and easy to handle (see, for instance, [10]). The D_K-distance between two fuzzy numbers $U,V\in\mathscr{F}_c(\mathbb{R})$ is defined by

$$[D_K(U,V)]^2=\langle s_U-s_V,s_U-s_V\rangle_K$$

$$=\int_{\{-1,1\}^2\times[0,1]^2}\big(s_U(u,\alpha)-s_V(u,\alpha)\big)\big(s_U(v,\beta)-s_V(v,\beta)\big)\,dK(u,\alpha,v,\beta)$$

where K is a positive definite and symmetric kernel. Thus D_K is in fact a generic L_2 distance on the Banach space $\mathscr{L}(\{-1,1\}\times[0,1])$.

Fuzzy random variables (see [16]) were introduced to model random mechanisms leading to imprecise values which are modelled by means of fuzzy sets. In this setting, the fuzzy expected value plays the usual role of a central summary measure.

Given a probability space $(\Omega,\mathscr{A},P)$, $\mathscr{X}:\Omega\to\mathscr{F}_c(\mathbb{R})$ is a *Fuzzy Random Variable* (FRV) (in Puri & Ralecu's sense, 1986) if the α-level mappings $\mathscr{X}_\alpha:\Omega\to\mathscr{K}_c(\mathbb{R})$, are random sets for all $\alpha\in[0,1]$, that is, Borel-measurable mappings when the Hausdorff metric d_H is considered on $\mathscr{K}_c(\mathbb{R})$. This is equivalent to define an FRV as a D_K-Borel measurable mapping on $\mathscr{F}_c(\mathbb{R})$.

If an FRV $\mathscr{X}$ is integrably bounded (that is, $\max\{\inf\mathscr{X}_0,\sup\mathscr{X}_0\}\in L^1(\Omega,\mathscr{A},P)$), its *expected value (or mean)* is the unique $E(\mathscr{X})\in\mathscr{F}_c(\mathbb{R})$ such that $\big(E(\mathscr{X})\big)_\alpha=$ Aumman's integral of the random set $\mathscr{X}_\alpha=[\inf\mathscr{X}_\alpha,\sup\mathscr{X}_\alpha]$ for all $\alpha\in[0,1]$ (see [16]).

In [3] it is shown that when a real-valued random variable is transformed into a fuzzy-valued one by simply considering certain triangular numbers centered on the original values, the expected value of the obtained fuzzy random variable captures all the information of the original distribution whenever the random variable takes at most 4 different values. This result is used to make inferences about these random variables with good results. However, in order to establish the characterization for general distributions, more complex fuzzy sets than the triangular ones should be considered. On the basis of this idea, a family of interesting fuzzy representations is proposed in [8].

Consider the mapping $\gamma^{\mathcal{C}} : \mathbb{R} \to \mathscr{F}_c(\mathbb{R})$ which transforms each value $x \in \mathbb{R}$ into the fuzzy number whose α-level sets are $\left(\gamma^{\mathcal{C}}(x)\right)_\alpha =$

$$\left[f_L(x) - (1-\alpha)^{1/h_L(x)}, f_R(x) + (1-\alpha)^{1/h_R(x)}\right]$$

for all $\alpha \in [0,1]$, where $f_L : \mathbb{R} \to \mathbb{R}$, $f_R : \mathbb{R} \to \mathbb{R}$, $f_L(x) \leq f_R(x)$ for all $x \in \mathbb{R}$, and $h_L : \mathbb{R} \to (0,+\infty)$, $h_R : \mathbb{R} \to (0,+\infty)$ are continuous and bijective. The fuzzification $\gamma^{\mathcal{C}}$ can be seen as a modification of a triangular one in which different degrees of curvature for the infimum and supremum functions are allowed. The curvature will entail different shapes for the characterizing fuzzy set depending on the distributions.

In [8] it is proved that if $X : \Omega \to \mathbb{R}$ is a random variable and $f_L(X), f_R(X) \in L^1(\Omega, \mathscr{A}, P)$, then the function $G_X : \mathbb{R} \to [0,1]$ such that $G_X(t) = \widetilde{E}\left(\gamma^{\mathcal{C}} \circ X \,|\, P\right)(t)$ for all $t \in \mathbb{R}$ can be interpreted as a $[0,1]$-valued ***characteristic function*** associated with the distribution of the random variable X. In other words, $\widetilde{E}\left(\gamma^{\mathcal{C}} \circ X \,|\, P\right) = \widetilde{E}\left(\gamma^{\mathcal{C}} \circ Y \,|\, P\right)$ if and only if X and Y are identically distributed.

3 Statistical Inferences on a Skewness Measure of RVs

A random variable $X : \Omega \to \mathbb{R}$ is symmetric about a known center $\theta \in \mathbb{R}$ if, and only if, $X - \theta$ and $\theta - X$ are identically distributed, that is, $F(\theta - x) = 1 - F(\theta + x)$ where F denote the cumulative distribution function.

The above characterization of the symmetry of a random variable X can be expressed in terms of the characterizing fuzzy representation introduced by [8] as follows: if $f_L(X-\theta), f_L(\theta - X), f_R(X-\theta), f_R(\theta - X) \in L^1(\Omega, \mathscr{A}, P)$, we have that X is symmetric about θ if, and only if, $\widetilde{E}\left(\gamma^{\mathcal{C}} \circ (X-\theta)\right) = \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (\theta - X)\right)$ and hence $D_K\left[\widetilde{E}\left(\gamma^{\mathcal{C}} \circ (X-\theta)\right), \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (\theta - X)\right)\right] = 0$.

Intuitively, the greater this distance the lower the symmetry of X. Thus, in order to quantify the degree of skewness of X about θ we consider the ***$\gamma^{\mathcal{C}}$-skewness measure about*** θ defined in [7] as

$$k_{\gamma^{\mathcal{C}}}(X,\theta) = \left(D_K\left[\widetilde{E}\left(\gamma^{\mathcal{C}} \circ (X-\theta)\right), \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (\theta - X)\right)\right]\right)^2.$$

Under the above conditions, if we consider a simple random sample $(X_1, \ldots, X_n)$ obtained from X, the analogue estimator of $k_{\gamma^{\mathcal{C}}}(X,\theta)$ given by

$$\widehat{k}^n_{\gamma^{\mathcal{C}}}(X,\theta) = \left[D_K\left(\overline{[\gamma^{\mathcal{C}} \circ (X-\theta)]}_n, \overline{[\gamma^{\mathcal{C}} \circ (\theta - X)]}_n\right)\right]^2$$

where $\overline{[\gamma^{\mathcal{C}} \circ X]}_n = \frac{1}{n}\sum_{i=1}^{n} \gamma^{\mathcal{C}} \circ X_i$, has been proved to be asymptotically unbiased and consistent (see [11]).

On the other hand, testing the null hypothesis $H_0 : X$ is symmetric about θ is equivalent to testing whether the FRVs $\gamma^{\mathcal{C}} \circ (X-\theta)$ and $\gamma^{\mathcal{C}} \circ (\theta - X)$ have the same expected value or not, that is,

$$H_0 : \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (X-\theta)\right) = \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (\theta - X)\right) \text{ vs. } H_1 : \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (X-\theta)\right) \neq \widetilde{E}\left(\gamma^{\mathcal{C}} \circ (\theta - X)\right), \quad (1)$$

i.e.,

$$H_0 : k_{\gamma^{\mathcal{C}}}(X,\theta) = 0 \text{ vs. } H_1 : k_{\gamma^{\mathcal{C}}}(X,\theta) \neq 0$$

provided that $k_{\gamma}(X,\theta)$ exists.

In [11] the following asymptotic approach to test the symmetry of X about θ has been stated. Let $(X_1^*,\ldots,X_m^*)$ with $m \in \mathbb{N}$ large enough be a (re-)sample obtained from $(X_1,\ldots,X_n)$:

In testing (1) at the nominal significance level $\phi \in [0,1]$, H_0 should be rejected whenever

$$T_n = \sqrt{n\,\widehat{k}^n_{\gamma^{\mathcal{C}}}(X,\theta)} = \sqrt{n}\,D_K\left(\overline{[\gamma^{\mathcal{C}} \circ (X-\theta)]}_n, \overline{[\gamma^{\mathcal{C}} \circ (\theta - X)]}_n\right) > z_\phi,$$

where z_ϕ is the $100(1-\phi)$ fractile of the distribution of the statistic

$$T^*_{n,m} = \sqrt{m\left[\widehat{k}^m_{\gamma^{\mathcal{C}}}\right]^*} = \sqrt{m}\,D_K\left(\overline{[\gamma^{\mathcal{C}} \circ (X^*-\theta)]}_m + \overline{[\gamma^{\mathcal{C}} \circ (\theta - X)]}_n, \overline{[\gamma^{\mathcal{C}} \circ (\theta - X^*)]}_m + \overline{[\gamma^{\mathcal{C}} \circ (X-\theta)]}_n\right)$$

The distribution of the statistic $T^*_{n,m}$ can be approximated by MonteCarlo method in order to compute the p-value of the test.

Remark 1. Note that if $n = m$ the asymptotic test for symmetry is equivalent to use a bootstrap technique. The empirical conclusions obtained about the test of equality of (fuzzy) means values of two FRV measured on the same population in [9] indicate that the bootstrap approach may be more suitable than the asymptotic one for small and moderated sample sizes.

In the next section, we will examine the empirical behaviour of the proposed test with different discrete/continuous and symmetric/asymmetric distributions and we will compare this test with other approaches in the literature.

4 Simulation Results

Each simulation is the result of 10,000 iterations of the test at a nominal significance level .05. We have considered both the asymptotic and the bootstrap version of the test. The number of bootstrap replications has been 1000, and we have considered $m = 10,000$ in the asymptotic case.

The D_K-distance considered in the simulations is the (W,φ)-distance introduced by [11] where W and φ have been chosen to be the Lebesgue measure on $[0,1]$.

A useful choice of characterizing fuzzy representation is the one determined by $f_L(x) = f_R(x) = 0$,

$$h_L(x) = \begin{cases} \dfrac{1}{1+x} & \text{if } x \geq 0 \\ 1-x & \text{if } x < 0 \end{cases}$$

and

$$h_R(x) = \frac{1}{h_L(x)} \quad \text{for all } x \in \mathbb{R}$$

This choice provides us with fuzzy sets quite similar to triangular ones, the 1-level set is the singleton $\{0\}$, and it presents a convex curvature in the negative part and a concave one in the positive one.

Example 1. We have considered test for symmetry about the mean value of different symmetric distributions: Normal, Cauchy, t Student and binomial.

Table 1 summarizes the obtained results for the percentage of rejections at the nominal significance level .05 and sample sizes of $n = 20, 100$. On the basis of these simulations, we get that the behaviour is very similar for all distributions. The bootstrap technique is much more accurate for the small size and it is conservative. The results are more similar as the sample size increases.

Example 2. In Table 2 the evolution of the power of the test for different degrees of asymmetry is shown. We have considered the normal distribution $N(4,1)$ as a case of symmetric distribution, a χ^2_4 as a skewed distribution, and a mixture of the previous distributions with mixing proportion $q = .5$ as an intermediate situation. The expected value of the three variables is 4, thus we will focus on the symmetry about $\theta = 4$.

Table 2 shows that the power of test (i.e. the percentage of rejections at the nominal significance level .05) is higher for more asymmetric distributions and it increases as the sample size does.

Example 3. In this example we compare our approach to test the symmetry of random variables with the results for other methods given in [15]. The sign test (denoted by

Table 1. Empirical percentage of rejections under H_0 at significance level $\alpha = .05$

		$n = 20$		$n = 100$	
Distribution	H_0	Asymptotic	Bootstrap	Asymptotic	Bootstrap
$N(0,1)$	$\theta = 0$	6.98	4.34	5.24	4.77
Cauchy	$\theta = 0$	7.04	4.67	5.33	4.91
t_3	$\theta = 0$	7.11	4.56	5.06	4.76
$B(5,0.5)$	$\theta = 2.5$	7.46	4.28	5.15	4.89

Table 2. Empirical percentage of rejections under H_0 at significance level $\alpha = .05$

	$n = 20$		$n = 100$	
Distribution	Asymptotic	Bootstrap	Asymptotic	Bootstrap
$N(4,1)$	7.33	4.69	5.24	4.87
.5-mixture	10.03	6.47	15.16	14.09
χ^2_4	13.87	8.41	32.44	30.99

N), the signed rank test (T), and the test (I_n) based on density estimates studied in [15] which has good performance for some multimodal densities will be considered. We have focused on the normal distribution, the t-distribution with 3 degrees of freedom and some k-mixtures of normal distributions denoted by $M(k,\mu)$ whose densities are

$$f_m(x) = \frac{1}{k}\sum_{i=0}^{k-1} \phi(x - i\mu + (k-1)\mu/2 - m)$$

where ϕ denotes the standard normal density.

All the distributions considered here are symmetric about $\theta = 0$, however, in order to establish the comparisons w.r.t. the power, we have also tested the symmetry about $\theta = 0.1$ and $\theta = 0.5$.

Table 3 shows the obtained percentage of rejections at the nominal significance level .05 and sample size $n = 100$. The test considered in this communication has better power than the other ones in the case of normal distributions, and a similar power with other unimodal densities as t-distributions. However, for multimodal distributions, it seems that a better option can always be found, namely, the signed rank test for $M(3,3)$ and I_n for $M(3,5)$. It should be noted that all this preliminary results have been made

Table 3. Empirical percentage of rejections under H_0 at significance level $\alpha = .05$

	$N(0,1)$				t_3			
θ	N	T	I_n	k_{γ^C}	N	T	I_n	k_{γ^C}
0	5.61	4.9	5.02	5.24	5.79	5.11	4.85	5.06
0.1	13.72	15.94	11.22	16.89	12.65	12.73	10.45	13.24
0.5	97.90	99.75	98.38	99.83	95.17	97.01	94.44	96.72

	$M(3,3)$				$M(3,5)$			
θ	N	T	I_n	k_{γ^C}	N	T	I_n	k_{γ^C}
0	5.47	4.96	4.13	5.52	5.48	4.95	4.98	5.25
0.1	6.43	6.38	4.79	6.54	6.43	6.03	6.66	6.34
0.5	28.37	43.09	32.90	34.24	26.92	32.76	79.18	19.04

for a very particular choice of fuzzification, and other fuzzifications should be analyzed to verify its behaviour in these situations.

5 Concluding Remarks

In this communication some empirical results concerning a new test for symmetry based on fuzzy tools have been shown. The test can be applied for both continuous and discrete distributions under mild conditions and, on the contrary to what happens with most of the usual procedures, it is consistent against any nonsymmetric alternative. The statistic depends on the choice of a fuzzification within a family. A simple fuzzification has been chosen in this preliminary study. The simulations indicate that the test is suitable for the usual uni-modal distributions, however there are better options for some multimodal ones. Further studies concerning other fuzzifications, families of distributions, and power against particular alternatives are currently being developed.

Acknowledgement. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2006–07501. Its financial support is gratefully acknowledged.

References

1. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware Soft. Comput. 2, 71–84 (1995)
2. Cabaña, A., Cabaña, E.M.: Tests of symmetry based on transformed empirical processes. Canad. J. Statist. 28(4), 829–839 (2000)
3. Colubi, A., González-Rodríguez, G.: Triangular fuzzification of random variables and power of distribution tests: Empirical discussion. Comput. Statist. Data Anal. 51(9), 4742–4750 (2007)
4. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc., River Edge (1994)
5. Dykstra, R., Kochar, S., Robertson, T.: Likelihood ratio tests for symmetry against one-sided alternatives. Annals of the Institute of Statistical Mathematics 47(4), 719–730 (1995)
6. Gibbons, J.D., Chakraborti, S.: Nonparametric Statistical Inference, 4th edn. Statistics: Textbooks and Monographs, vol. 168. Marcel Dekker, Inc, New York (2003)
7. González-Rodríguez, G., Colubi, A., D'Urso, P., Giordani, P.: An asymptotic test about the asymmetry of random variables based on fuzzy tools. In: Lawry, J., Miranda, E., Bugarin, A., Li, S., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Integrated Uncertainty Modelling, pp. 87–94. Springer, Heidelberg (2006a)
8. González-Rodríguez, G., Colubi, A., Gil, M.A.: A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. Comput. Statist. Data Anal. 51(1), 163–176 (2006b)
9. González-Rodríguez, G., Colubi, A., Gil, M.A., D'Urso, P.: An asymptotic two dependent samples test of equality of means of fuzzy random variables. In: Proceedings of the 17th Conference of IASC-ERS (COMPSTAT 2006), pp. 689–695 (2006c)
10. González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A.: Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst. 157(19), 2608–2613 (2006d)

11. González-Rodríguez, G., Colubi, A., Gil, M.A., Lubiano, M.A.: A new way of quantifying the symmetry of a random variable: estimation and hypothesis testing (submitted for publication) (2008)
12. Hájek, J., Sidák, Z.: Theory of Rank Tests. Academic Press, New York (1967)
13. Klement, E.P., Puri, M.L., Ralescu, D.A.: Limit theorems for fuzzy random variables. Proc. R. Soc. Lond A. 407, 171–182 (1986)
14. Körner, R., Näther, W.: On the variance of random fuzzy variables. In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data. Studies in Fuzziness and Soft Computing, vol. 87, pp. 22–39. Physica-Verlag, Heildelberg (2002)
15. Mizushima, T., Nagao, H.: A test for symmetry based on density estimates. J. Japan Statist. Soc. 28(2), 205–225 (1998)
16. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)
17. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part 1. Inform. Sci. 8, 199–249; Part 2. Inform. Sci. 8, 301–353; Part 3. Inform. Sci. 9, 43–80 (1975)

Fuzzy Kendall τ Statistic for Autocorrelated Data

Olgierd Hryniewicz and Anna Szediw

Systems Research Institute, Polish Academy of Science, Warsaw, Poland

Abstract. Kendall's τ statistic has found many practical applications. Recently, it has been proposed by Hryniewicz and Szediw as the basis for the Kendall control chart for monitoring autocorrelated production processes. They have shown that this chart has good statistical characteristics only for large samples. Unfortunately, in such a case existing algorithms for the calculation of the fuzzy Kendall's τ statistic are not sufficiently effective. In the paper we investigate a simple heuristic algorithm for the calculation of fuzzy Kendall's τ that makes the implementation of the proposed chart applicable in statistical quality control.

Keywords: Time series, Fuzzy data, Kendall τ.

1 Introduction

Statistical data presented in a form of time series are usually either autocorrelated or interdependent in a more complicated way. The existence of such dependencies may make the statistical analysis of such data much more difficult. In some applications, however, such time-related dependencies are not frequent, and much simpler statistical tools are required. For example, in statistical quality control it is usually assumed that consecutive observations of monitored production processes are independent. When this assumption is true, simple statistical methods - which may be used even by workers - are sufficient for the control of a process. However, when data observed from a process are dependent, statistical analysis becomes very complicated. As a matter of fact, shop-floor practitioners are usually unable to work with autocorrelated data without an assistance of specialists. Moreover, in many cases it is necessary to use specialized software. Therefore, there is a practical need to detect autocorrelation in data as quickly as possible. Statistical tools available for such analysis are available, but generally they have been developed for dealing with normally distributed autoregression processes. In practice, however, we usually do not know whether the investigated process can be described by the normal autoregressive model. Thus, there is a need to develop a simple (for practitioners) non-parametric (distribution-free) tool that would be useful for the detection of autocorrelation in data. Such a tool - a Kendall control chart - has been proposed by Hryniewicz and Szediw [4]. The generalization of the Kendall control chart, when observed data are fuzzy, has been introduced recently by Hryniewicz and Szediw [5]. This statistical procedure is based on the fuzzy Kendall statistic which has been originally introduced by Hébert et al. [2] and Denœux et al. [1] who considered it as a statistic based on fuzzy ranks.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 155–162, 2008.
springerlink.com

The main problem with practical implementation of the fuzzy Kendall control chart is computational one. The algorithms proposed by Hébert et al. [2] (exact) and Denœux et al. [1] (approximate) are, according to those authors, effective only for relatively small samples (not larger than 20). However, investigations of the crisp version of the Kendall control chart presented in [4] have revealed that the required sample sizes should be much larger (at least 50 observations). Therefore, there is a need to find a fast approximate algorithm which may be useful for the analysis of such large samples. In the second section of the paper we propose a simple heuristic algorithm that might be useful for a fast approximate computation of the membership function of the fuzzy Kendall τ statistic. Some properties of this algorithm are investigated in the third section of the paper.

2 Kendall Test for Autocorrelated Fuzzy Data

Let $Z_1, Z_2, \ldots, Z_n$ denote a random sample of n consecutive process observations. These observations can be transformed into two-dimensional vector (X_i, Y_i), where $X_i = Z_i$ and $Y_i = Z_{i+1}$ for $i = 1, 2, \ldots, n-1$. Then, the Kendall's τ sample statistic which measures the association between random variables X and Y is given by the following formula

$$\tau_n = \frac{4}{n-1} \sum_{i=1}^{n-1} V_i - 1, \tag{1}$$

where

$$V_i = \frac{card\{(X_j, Y_j) : X_j < X_i, Y_j < Y_i\}}{n-2}, i = 1, \ldots, n-1. \tag{2}$$

In this paper we assume that our observations are imprecise, and may be modeled by fuzzy random variables, understood as fuzzy perceptions of ordinary random variables. It means that there exist non-fuzzy (crisp) original values of measured quantities, but due to the imprecise character of measurements they are perceived as imprecise fuzzy values. In such a case we can use a fuzzy version of the Kendall's τ statistic, originally introduced in the papers by Hébert et al. [2] and Denœux et al. [1], who also proposed useful methods for the calculation of its membership function.

In the case of statistical data given in a form of a time series we may use the fuzzy Kendall τ for an autocorrelated time series originally proposed in Hryniewicz [3] for the analysis of serial fuzzy binomial data. In [5] a more general case has been considered where the series of consecutive observations is described by a vector of fuzzy data $(\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_n)$. In order to compute the fuzzy version of the Kendall's τ statistic for the considered fuzzy time series let us assume that each fuzzy observation is described by a membership function $\mu_i(z)$, $i = 1, \ldots, n$.

Let us notice now that each fuzzy data point $\tilde{Z}_i$ is completely defined by the set of its α-cuts $[Z^{\alpha}_{i,L}, Z^{\alpha}_{i,U}]$, $0 < \alpha \leq 1$. Hence, the fuzzy equivalent of V_i given by (2), denoted by $\tilde{V}_i$, is defined as a convex hull of the set of its α-cuts $[V^{\alpha}_{i,L}, V^{\alpha}_{i,U}]$, $0 < \alpha \leq 1$, where

$$V^{\alpha}_{i,L} = \min_{\substack{z_i \in [Z^{\alpha}_{i,L}, Z^{\alpha}_{i,U}] \\ i=1,\ldots,n}} \frac{card_{j \neq i}\{(z_j, z_{j+1}) : z_j < z_i, z_{j+1} < z_{i+1}\}}{n-2} \tag{3}$$

and

$$V_{i,U}^{\alpha} = \max_{\substack{z_i \in [Z_{i,L}^{\alpha}, Z_{i,U}^{\alpha}] \\ i=1,\ldots,n}} \frac{card_{j \neq i}\{(z_j, z_{j+1}) : z_j < z_i, z_{j+1} < z_{i+1}\}}{n-2} \tag{4}$$

for $i = 1,\ldots,n-1$. Having the α-cuts $[V_{i,L}^{\alpha}, V_{i,U}^{\alpha}]$, $0 < \alpha \leq 1$ for all $i = 1,\ldots,n$ we can straightforwardly calculate the α-cuts of the fuzzy Kendall's τ statistic $[\tau_L^{\alpha}, \tau_U^{\alpha}]$, $0 < \alpha \leq 1$, and thus obtain its membership function.

The calculation of the membership function of the fuzzy Kendall's τ statistic may be, in a general case, a difficult and computationally intensive task. When the number of fuzzy observations in a whole set of observations is small we can use a general methodology proposed by Hébert et al. [2] for the exact calculations of the membership function of the fuzzy Kendall statistic. Another possibility is to use algorithms of stochastic optimization based on the Monte-Carlo simulations, such as an algorithm proposed in Denœux et al. [1]. However, for the particular cases, such as that of the fuzzy Kendall control chart, when the number of analyzed fuzzy observations may be quite large, these general methods suffer from the "curse of dimensionality", and cannot be efficiently used in practice. Therefore, there is a need to design a much faster approximate algorithm that might be used in such cases (especially in cases when there is a need to compute many α-cuts of the membership function of τ). First such algorithm has been proposed in [5]. However, the results of extensive simulations show that it has to be improved. In this paper we present the result of such improvement.

The construction of the optimization algorithm will be apparent if we consider the influence of the *pattern* of consecutive observations on the value of Kendall's τ. In Figure 1 we present a possible crisp sample with individual points belonging to respective α-cuts of fuzzy data points which leads to the maximum value of Kendall's τ, namely 1. The maximal value of the Kendall's τ, equal to 1, is attained when consecutive points form a monotonically increasing or decreasing series.

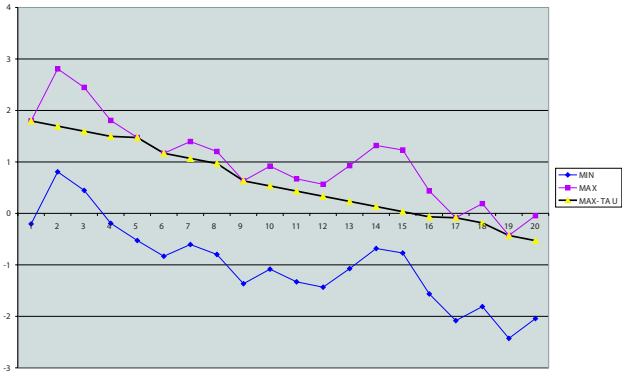


Fig. 1. Possible configuration of observations of a time series for the maximum value of the Kendall's τ

In general, for the given value of α the largest value of τ should be attained for a series of values $z_i^L \in [z_{i,L}^{\alpha}, z_{i,U}^{\alpha}]$, $i = 1,\ldots,n$, $0 < \alpha \leq 1$ that form a monotone (or nearly monotone) increasing (decreasing) series. To find such a series we can start with the series $z_i^* = z_{i,L}^{\alpha}$, $i = 1,\ldots,n$, $0 < \alpha \leq 1$. In the next steps we can increase certain values of this series (keeping in mind that they have to belong to their α-cuts) in order to arrive at a monotone (or nearly monotone) increasing series. The same procedure should be repeated in search of a monotone (or nearly monotone) decreasing series. In this case we can start with the series $z_i^* = z_{i,U}^{\alpha}$, $i = 1,\ldots,n$, $0 < \alpha \leq 1$, and in the next steps we should decrease certain values of this series in order to arrive at a monotone (or nearly monotone) decreasing series.

Let us describe this heuristic algorithm in a more formal way. For notational convenience we omit the symbol α which refers to a chosen α-cut. The upper limit of the α-cut for the fuzzy value of Kendall's τ is computed according to Algorithm 1.

Algorithm 1

begin
set ε *to a small value*
$k = 0$
$z_{k+1}^* = z_{k+1,L}$

*loop*1 : $k = k+1$
if $[(z_{k+1,L} \geq z_k^*)$ *or* $(z_{k+1,U} < z_k^*)]$ *then*
if $[(z_{k+1,L} \geq z_k^*)\,]$ *then* $z_{k+1}^* = z_{k+1,L}$
if $[(z_{k+1,U} < z_k^*)]$ *then* $z_{k+1}^* = z_{k+1,U}$
else $z_{k+1}^* = z_k^* + \varepsilon$
if $k < n-1$ *goto loop*1
use $(z_1^*,\ldots,z_n^*)$ *for the calculation of* $\tau_{U,1}$
$k = 0$
$z_{k+1}^* = z_{k+1,U}$

*loop*2 : $k = k+1$
if $[(z_{k+1,L} \geq z_k^*)$ *or* $(z_{k+1,U} < z_k^*)]$ *then*
if $[(z_{k+1,L} \geq z_k^*)\,]$ *then* $z_{k+1}^* = z_{k+1,U}$
if $[(z_{k+1,U} < z_k^*)]$ *then* $z_{k+1}^* = z_{k+1,L}$
else $z_{k+1}^* = z_k^* - \varepsilon$
if $k = n-1$ *stop*
*else goto loop*2
use $(z_1^*,\ldots,z_n^*)$ *for the calculation of* $\tau_{U,2}$
$\tau_U = \max(\tau_{U,1}, \tau_{U,2})$
end

Now let us consider the problem of finding the minimal value of τ. In Figure 2 we present a possible crisp sample with individual points belonging to respective α-cuts of fuzzy data points which leads to the minimum value of Kendall's τ, namely -1. As we can see, this minimal value of the Kendall's τ is attained when consecutive points

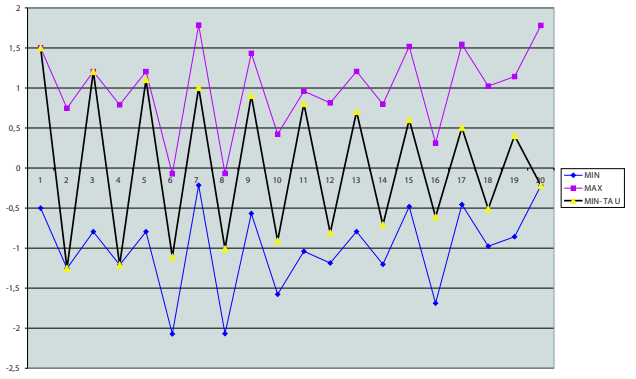


Fig. 2. Possible configuration of observations of a time series for the minimum value of the Kendall's τ

numbered by even numbers form a monotonically decreasing series and simultaneously consecutive points numbered by odd numbers form a monotonically increasing series, or vice versa. In both cases the increasing and decreasing series should not intersect.

In general, the lowest value of τ should be attained for a series of values $z_i^U \in [z_{i,L}^{\alpha}, z_{i,U}^{\alpha}]$, $i = 1, \ldots, n$, $0 < \alpha \leq 1$ that form an alternating series of values such that the odd (even) observations form a decreasing (or nearly decreasing) series, and the even (odd) observations form an increasing (or nearly increasing) series. To find such a series we can start with the series $z_{1,L}^{\alpha}, z_{2,U}^{\alpha}, z_{3,L}^{\alpha}, \ldots$ or with the series $z_{1,U}^{\alpha}, z_{2,L}^{\alpha}, z_{3,U}^{\alpha}, \ldots$. In the next step we can increase certain values initially defined by the lower limits of the α-cuts and decrease certain values initially defined by the upper limits of the α-cuts in order to arrive at an alternating (or nearly alternating) series. The lower limit of the α-cut for the fuzzy value of Kendall's τ is computed using an algorithm which is similar to the algorithm presented above. The formal description of this heuristic algorithm, which too long for the presentation in this paper, can be found in [5].

The application of both algorithms does not guarantee that the computed pair (τ_L, τ_U) is the true α-cut for the fuzzy value of the Kendall's τ. However, in case of large sample sizes it gives a quite good approximation. It may also serve for the generation of the starting sequence of ranks in the algorithm proposed by Denœux et al. [1]. In the next section we present some results of simulation experiments which support this claim.

3 Analysis of the Accuracy of the Calculated Minimal and Maximal Values of Kendall τ

The accuracy of the proposed heuristic algorithms can be precisely evaluated only in case of small samples. However, in certain applications of the Kendall's τ large and very large samples are required. In such a case we are not able, as for now, to calculate exact values for the limits of τ. Therefore, we have decided to compare

the results obtained by the application of our approximate algorithm with the results obtained in a simple simulation experiment. In our simulation experiment fuzzy data were represented by their α-cuts. For all simulated observations their α-cuts were given as intervals $[X-w, X+w]$, where random values X were generated from a normal autoregressive process characterized by the mean value equal to 0, standard deviation equal to 1, and the autocorrelation coefficient ρ. In the first of our experiments we have found that the intervals calculated using the proposed heuristic algorithm are on average better than the intervals calculated using the simulation algorithm proposed in [1] when we used the starting point representing the mid-points of observed intervals. In the next experiment we compared our algorithm with the algorithm based on random generation of original (crisp) observations. In our experiment for each fuzzy sample in the inner simulation loop we simulated 100 000 crisp samples in such a way, that each point of the simulated crisp sample was chosen randomly from the α-cut of the respective fuzzy observation. In every case the set of results simulated in the inner loop was extended with the result obtained using our heuristic algorithm, i.e. the result of 100 000 simulations entered the average computed in the outer loop of the simulation experiment only in this case when it has been better than our approximate solution. In the simulation experiment we varied the values of the autoregression coefficient ρ, sample size n, and the width of the α-cut. For each considered combination of these parameters we simulated 1000 fuzzy random samples.

From the results of those simulation experiments we have found that in case of the upper limits of α-cuts for sample sizes equal or larger than 50 items the random search for better solutions with the help of 100 000 simulations does not provide better results than the proposed heuristic algorithm. Even if we find a better solution, the difference between this solution and our approximate solution is very small. For sample sizes smaller than 50 this difference may be practically significant. Unfortunately, we have not obtained such good approximations for the lower limits of α-cuts.

Very promising results have been observed when we used vectors of ranks corresponding to the lower and upper limits of the calculated heuristic intervals as the initial vectors for the simulation of linear extensions of partial orders generated by observed

Table 1. Comparison of intervals computed using heuristic and 2 simulation methods

ρ	method	low. limit	up. limit	piMin	piMax
0,8	hint	0,2358	0,7501	-	-
	linext	0,1603	0,7518	95,2	18,8
	irawd	0,2193	0,7501	38,0	0,0
0,0	hint	-0,1607	0,2635	-	-
	linext	-0,2174	0,2956	98,8	97,6
	irawd	-0,2084	0,2641	80,3	3,3
-0,8	hint	-0,6919	-0,2503	-	-
	linext	-0,7087	-0,2193	76,5	93,0
	irawd	-0,7109	-0,2379	48,1	24,9

fuzzy ranks [1]. In Table 1 we present the comparison of intervals computed this way (linext) with the heuristically computed intervals (hint), and intervals computed using raw Monte Carlo data according to the algorithm described above (irawd). In the experiment we simulated 1000 fuzzy (interval-valued) vectors (with rather imprecise data), and in each case we performed 10^6 runs of the algorithm that simulates linear extension of ranks, and 500 000 runs of the algorithm that generates original values of observations. In the fifth (piMin) and sixth (piMax) column of Table 1 we present percentages of cases when we obtained better interval limits by combining our heuristic algorithm with either the simulation of linear extensions of ranks or the simulation of original values of observations for lower and upper limits, respectively.

From the results presented in Table 1 we see that neither of the compared algorithms is consistently better than the other one. We have to keep in mind, however, that these results are based on a relatively small number of simulations (1000).

4 Discussion

The fuzzy version of the Kendall's τ statistics considered in this paper seems to be a useful statistical tool for practitioners who have to look for dependencies in time series. However, due to the imprecise character of data necessary computations become prohibitively time consuming. The proposed heuristic algorithm makes necessary computations fast and sufficiently accurate in case of the upper limits of the α-cuts of the Kendall's τ statistic. These limits may be improved by combining our algorithm with additional simulations, but the improvement does not seem to be very significant, especially for positively correlated observations. In case of lower limits significantly better results have been obtained when we use the hybrid algorithm which combines the heuristic algorithm proposed by Hryniewicz and Szediw [5] and the Monte Carlo algorithm proposed by Denœux et al. [1]. The results presented in Table 1 show, however, that further investigations have to be done in order to improve the accuracy of approximations. In the applications of the fuzzy Kendall's τ in quality control this accuracy is needed for the investigation of an interesting and important problem which is still waiting for its solution, namely the influence of fuzziness of data on important characteristics of a control chart, such as e.g. the average run length ARL (the average time to signal).

References

1. Denœux, T., Masson, M.-H., Hébert, P.-A.: Nonparametric rank-based statistics and significance tests for fuzzy data. Fuzzy Sets Syst. 153, 1–28 (2005)
2. Hébert, P.-A., Masson, M.-H., Denœux, T.: Fuzzy rank correlation between fuzzy numbers. In: Bilgiç, T., De Baets, B., Kaynak, O. (eds.) IFSA 2003. LNCS, vol. 2715, pp. 224–227. Springer, Heidelberg (2003)
3. Hryniewicz, O.: Looking for dependencies in short time series using imprecise statistical data. In: Castillo, O., Melin, P., Ross, O.M., Cruz, R.S., Pedrycz, W., Kacprzyk, J. (eds.) Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, pp. 201–208. Springer, Berlin (2007)

4. Hryniewicz, O., Szediw, A.: Sequential signals on a control chart based on nonparametric statistical tests. In: Proceedings of the IXth International Workshop on Intelligent Statistical Quality Control (Beijing, China), pp. 111–126 (2007)
5. Hryniewicz, O., Szediw, A.: Kendall control chart for fuzzy data. In: Atanassov, K., Chountas, P., Kacprzyk, J., Krawczak, M., Melo-Pinto, P., Szmidt, E., Zadrozny, S. (eds.) Developments in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics: Applications. EXIT, Warszawa (in press, 2008)

Estimation

Mixture Model Estimation with Soft Labels

Etienne Côme[1], Latifa Oukhellou[2], Thierry Denœux[3], and Patrice Aknin[1]

[1] French National Institute for Transport and Safety Research (INRETS) - LTN, Arcueil, France
[2] Université Paris XII - CERTES, Créteil, France
[3] HEUDIASYC, Université de Technologie de Compiègne, CNRS - Centre de Recherches de Royallieu, Compiègne, France

Abstract. This paper addresses classification problems in which the class membership of training data is only partially known. Each learning sample is assumed to consist in a feature vector and an imprecise and/or uncertain "soft" label m_i defined as a Dempster-Shafer basic belief assignment over the set of classes. This framework thus generalizes many kinds of learning problems including supervised, unsupervised and semi-supervised learning. Here, it is assumed that the feature vectors are generated from a mixture model. Using the General Bayesian Theorem, we derive a criterion generalizing the likelihood function. A variant of the EM algorithm dedicated to the optimization of this criterion is proposed, allowing us to compute estimates of model parameters. Experimental results demonstrate the ability of this approach to exploit partial information about class labels.

Keywords: Dempster-Shafer theory, Transferable Belief Model, Mixture models, EM algorithm, Classification, Clustering, Partially supervised learning, Semi-supervised learning.

1 Introduction

Machine learning classically deals with two different problems: supervised learning (classification) and unsupervised learning (clustering). However, other paradigms exist such as *semi-supervised learning* [10], and *partially-supervised learning* [1, 5, 9, 11]. In the former approach, one use a mix of unlabelled and labelled examples, whereas in the latter, one can define constraints on the possible classes of the examples. The importance for such problems comes from the fact that labelled data are often difficult to obtain, while unlabelled or partially labelled data are easily available.

The investigations reported in this paper follow this path, in the context of belief functions. In this way, both the uncertainty and the imprecision of class labels may be handled. The considered training sets are of the form $\mathbf{X}^{iu} = \{(\mathbf{x}_1, m_1), \ldots, (\mathbf{x}_N, m_N)\}$, where m_i is a basic belief assignment, or Dempster-Shafer mass function [14] encoding our knowledge about the class of example i. The m_is (hereafter referred to as "soft labels") may represent different kinds of knowledge, from precise to imprecise and from certain to uncertain. Thus, previous problems are special cases of this general formulation. Other studies have already proposed solutions in which class labels are expressed by possibility distributions or belief functions [6, 8]. In this article, we present a new approach to solve learning problems of this type, which completes a preliminary study by Vannoorenberghe and Smets [21]. This solution is based on mixture models, and therefore assumes a generative model for the data.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 165–174, 2008.
springerlink.com

This article is organized as follows. Background material on belief functions and estimation of parameters in mixture models using the EM algorithm will first be recalled in Sections 2 and 3, respectively. The problem of learning from data with soft labels will then be addressed in Section 4, through the definition of a learning criterion, and of an EM type algorithm dedicated to its optimization. Finally we will presented some simulations results in Section 5.

2 Background on Belief Functions

2.1 Belief Functions on a Finite Frame

The theory of belief functions was introduced by Dempster [3] and Shafer [14]. The interpretation adopted throughout this paper will be that of the Transferable Belief Model (TBM) introduced by Smets [20]. The first building block of belief function theory is the *basic belief assignment* (bba), which models the beliefs held by an agent regarding the actual value of a given variable taking values in a finite domain (or *frame of discernment*) Ω, based on some body of evidence. A bba m^Ω is a mapping from 2^Ω to $[0,1]$ verifying $\sum_{\omega\subseteq\Omega} m^\Omega(\omega) = 1$. The subsets ω for which $m^\Omega(\omega) > 0$ are called the *focal sets*. Several kind of belief functions are defined according to the structure of focal sets. In particular, a bba is *Bayesian* if its focal sets are singletons, it is *consonant* if its focal sets are nested and a it is *categorical* if it has only one focal set. Bbas are in one to one correspondence with other representations of the agent's belief, including the plausibility function defined as:

$$pl^\Omega(\omega) \triangleq \sum_{\alpha\cap\omega\neq\emptyset} m^\Omega(\alpha), \quad \forall\omega\subseteq\Omega. \tag{1}$$

The quantity $pl^\Omega(\omega)$ is thus equal to the sum of the basic belief masses assigned to propositions that are not in contradiction with ω. The plausibility function associated to a Bayesian bba is a probability measure. If m^Ω is consonant, then pl^Ω is a possibility measure: it verifies $pl^\Omega(\alpha\cup\beta) = \max(pl^\Omega(\alpha), pl^\Omega(\beta))$, for all $\alpha,\beta\subseteq\Omega$.

2.2 Conditioning and Combination

Given two bbas m_1^Ω and m_2^Ω supported by two distinct bodies of evidence, we may build a new bba $m_{1\textcircled{\cap}2}^\Omega = m_1^\Omega \textcircled{\cap} m_2^\Omega$ that corresponds to the conjunction of these two bodies of evidence:

$$m_{1\textcircled{\cap}2}^\Omega(\omega) \triangleq \sum_{\alpha_1\cap\alpha_2=\omega} m_1^\Omega(\alpha_1) m_2^\Omega(\alpha_2), \quad \forall\omega\subseteq\Omega. \tag{2}$$

This operation is usually referred to as the *unnormalized Dempster's rule* or the TBM conjunctive rule. If the frame of discernment is supposed to be exhaustive, the mass of the empty set is usually reallocated to other subsets, leading to the definition of the normalized Demspter's rule $\oplus$ defined as:

$$m_{1\oplus2}^\Omega(\omega) = \begin{cases} 0 & \text{if } \omega = \emptyset \\ \frac{m_{1\textcircled{\cap}2}^\Omega(\omega)}{1-m_{1\textcircled{\cap}2}^\Omega(\emptyset)} & \text{if } \omega\subseteq\Omega, \omega\neq\emptyset, \end{cases} \tag{3}$$

which is well defined provided $m^{\Omega}_{1\textcircled{\cap}2}(\emptyset) \neq 1$. Note that, if m^{Ω}_1 (or m^{Ω}_2) is Bayesian, then $m^{\Omega}_{1\oplus 2}(\omega)$ is also Bayesian. The combination of a bba m^{Ω} with a categorical bba focused on $\alpha \subseteq \Omega$ using the TBM conjunctive rule is called (unnormalized) *conditioning*. The resulting bba is denoted $m^{\Omega}(\omega|\alpha)$. Probabilistic conditioning is recovered when m^{Ω} is Bayesian, and normalization is performed. Using this definition, we may rewrite the conjunctive combination rule: $m^{\Omega}_{1\textcircled{\cap}2}(\omega) = \sum_{\alpha\subseteq\Omega} m^{\Omega}_1(\alpha) m^{\Omega}_2(\omega|\alpha), \forall \omega \subseteq \Omega$, which is a counterpart of the total probability theorem in probability theory [7, 17]. This expression provides a shortcut to perform marginal calculations on a product space when conditional bbas are available [17]. Consider two frames Ω and Θ, and a set of conditional belief functions $m^{\Theta|\Omega}(\cdot|\omega)$ for all $\omega \subseteq \Omega$. Each conditional bba $m^{\Theta|\Omega}(\cdot|\omega)$ represents the agent's belief on Θ in a context where ω holds. The combination of these conditional bbas with a bba m^{Ω} on Ω yields the following plausibility on Θ:

$$pl^{\Theta}(\theta) = \sum_{\omega\subseteq\Omega} m^{\Omega}(\omega) pl^{\Theta|\Omega}(\theta|\omega), \quad \forall \theta \subseteq \Theta. \tag{4}$$

This property bears some resemblance with the total probability theorem, except that the sum is taken over the power set of Ω and not over Ω. We will name it the *total plausibility theorem*.

2.3 Independence, Continuous Belief Functions and Bayes Theorem

The usual independence concept of probability theory does not easily find a counterpart in belief function theory, where different notions must be used instead. The simplest form of independence defined in the context of belief functions is *cognitive independence* ([14], p. 149). Frames Ω and Θ are said to be cognitively independent with respect to $pl^{\Omega\times\Theta}$ iff we have $pl^{\Omega\times\Theta}(\omega\times\theta) = pl^{\Omega}(\omega)\, pl^{\Theta}(\theta), \forall \omega \subseteq \Omega, \forall \theta \subseteq \Theta$. Cognitive independence boils down to probabilistic independence when $pl^{\Omega\times\Theta}$ is a probability measure.

The TBM can be extended to continuous belief functions on the real line, assuming focal sets to be real intervals [19]. In this context, the concept of bba is replaced by that of *basic belief density* (bbd), defined as a mapping $m^{\mathbb{R}}$ from the set of closed real intervals to $[0,+\infty)$ such that $\int_{-\infty}^{+\infty}\int_{x}^{+\infty} m^{\mathbb{R}}([x,y])dydx \leq 1$. By convention, the one's complement of this integral is allocated to $\emptyset$. As in the discrete case, $pl^{\mathbb{R}}([a,b])$ is defined as an integral over all intervals whose intersection with $[a,b]$ is non-empty. Further extension of these definitions to $\mathbb{R}^d, d > 1$ is possible and it is also possible to define belief functions on mixed product spaces involving discrete and continuous frames.

The Bayes' theorem of probability theory is replaced in the framework of belief functions by the Generalized Bayesian Theorem (GBT), [18]. This theorem provides a way to reverse conditional belief functions without any prior knowledge. Let us suppose two spaces, $\mathcal{X}$ the observation space and Θ the parameter space. Assume that our knowledge is encoded by a set of conditional bbas $m^{\mathcal{X}|\Theta}(.|\theta_i), \theta_i \in \Theta$, which express our belief in future observations conditionally on each θ_i, and we observe a realization $x \subseteq \mathcal{X}$. The question is: given this observation and the set of conditional bbas, what is our belief on the value of Θ? The answer is given by the GBT and states that the resulting plausibility function on Θ has the following form:

$$pl^{\Theta|\mathscr{X}}(\theta|x) = pl^{\mathscr{X}|\Theta}(x|\theta) = 1 - \prod_{\theta_i \in \theta} (1 - pl^{\mathscr{X}|\Theta}(x|\theta_i)). \tag{5}$$

When a prior bba m_0^{Θ} on Θ is available, it should be combined conjunctively with the bba defined by (5). The classical Bayes' theorem is recovered when the conditional bbas $m^{\mathscr{X}|\Theta}(.|\theta_i)$ and the prior bba m_0^{Θ} are Bayesian.

3 Mixture Models and the EM Algorithm

After this review of some tools from belief functions theory, the next part is dedicated to the probabilistic formulation of the clustering problem in terms of mixture model. We will therefore present the data generation scheme underlying mixture models and the solution to parameter estimation in the unsupervised case.

3.1 Mixture Models

Mixture models suppose the following data generation scheme:

- The true class labels $\{y_1, \ldots, y_N\}$ of data points are realizations of independent and identically distributed (i.i.d) random variables $Y_1, \ldots, Y_N \sim Y$ taking their values in the set of all K classes $\mathscr{Y} = \{c_1, \ldots, c_K\}$ and distributed according to a multinomial distribution $\mathscr{M}(1, \pi_1, \ldots, \pi_K)$. The π_k are thus the class proportions and they verify $\sum_{k=1}^{K} \pi_k = 1$. The information on the true class labels of samples coming from such variables can also be expressed by a binary variable $\mathbf{z}_i \in \{0,1\}^K$, such that $z_{ik} = 1$ if $y_i = c_k$, and $z_{ik} = 0$ otherwise.
- The observed values $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ are drawn using the class conditional density in relation with the class label. More formally, $X_1, \ldots, X_N \sim X$ are continuous random variables taking values in $\mathscr{X}$, with conditional probability density functions $f(\mathbf{x}|Y = c_k) = f(\mathbf{x}; \theta_k), \quad \forall k \in \{1, \ldots, K\}$.

The parameters that need to be estimated are therefore the proportions $\pi = (\pi_1, \ldots, \pi_K)$ and the parameters of the class conditional densities $\theta_1, \ldots, \theta_K$. To simplify the notations, the vector of all model parameters is denoted $\Psi = (\pi_1, \ldots, \pi_K, \theta_1, \ldots, \theta_K)$. In unsupervised learning problems, the available data are only the i.i.d realizations of X, $\mathbf{X}^u = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, provided by the generative model. To learn the parameters and the associated clustering, the log-likelihood must be computed according to the marginal density $\sum_{k=1}^{K} \pi_k f(\mathbf{x}_i; \theta_k)$ of X_i. This leads to the unsupervised log-likelihood criterion:

$$L(\Psi; \mathbf{X}^u) = \sum_{i=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k f(\mathbf{x}_i; \theta_k) \right). \tag{6}$$

3.2 EM Algorithm

The log-likelihood function defined by (6) is difficult to optimize and may lead to a set of different local maxima. The EM algorithm [4] is nowadays the classical solution to this problem. The missing data of the clustering problem are the true class labels y_i of learning examples. The basis of the EM algorithm can be found in the decomposition of the likelihood function in two terms :

$$L(\Psi;\mathbf{X}^u) = \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{K} t_{ik}^{(q)} \ln\left(\pi_k f(\mathbf{x}_i;\theta_k)\right)}_{Q(\Psi,\Psi^{(q)})} - \underbrace{\sum_{i=1}^{N}\sum_{k=1}^{K} t_{ik}^{(q)} \ln\left(\frac{\pi_k f(\mathbf{x}_i;\theta_k)}{\sum_{k'=1}^{K}\pi_{k'} f(\mathbf{x}_i;\theta_{k'})}\right)}_{H(\Psi,\Psi^{(q)})}, \tag{7}$$

with:

$$t_{ik}^{(q)} = \mathbb{E}_{\Psi^{(q)}}[z_{ik}|\mathbf{x}_i] = \mathbb{P}(z_{ik}=1|\Psi^{(q)},\mathbf{x}_i) = \frac{\pi_k^{(q)} f(\mathbf{x}_i;\theta_k^{(q)})}{\sum_{k'=1}^{K}\pi_{k'}^{(q)} f(\mathbf{x}_i;\theta_{k'}^{(q)})}. \tag{8}$$

Such a decomposition is useful to define an iterative ascent strategy thanks to the form of H. As a consequence of Jensen's inequality we may write $H(\Psi^{(q)},\Psi^{(q)}) - H(\Psi,\Psi^{(q)}) \geq 0, \forall\Psi$. Consequently, the maximization of the auxiliary function $\Psi^{(q+1)} = \arg\max_\Psi Q(\Psi,\Psi^{(q)})$ is sufficient to improve the likelihood. Furthermore, because the sum over the classes is outside the logarithm in the Q function, the optimization problems are decoupled and the maximization is simpler. The EM algorithm can be described as follows. It starts with initial estimates $\Psi^{(0)}$ and alternates two steps : the E step where the t_{ik} are computed according to the current parameters estimates, defining a new Q function maximized during the M step. Thanks to (7), this defines a sequence of parameter estimates with increasing likelihood values. Finally, the mixture model setting and the EM algorithm can be adapted to handle specific learning problems such as the semi-supervised [10] and the partially supervised cases [1].

4 Extension to Imprecise and Uncertain Labels

4.1 Derivation of a Generalized Likelihood Criterion

Our method extends the approach described above to handle *imprecise* and *uncertain* class labels defined by belief functions. In this section, we shall assume the learning set to be of the form $\mathbf{X}^{iu} = \{(\mathbf{x}_1, m_1^{\mathscr{Y}}), \ldots, (\mathbf{x}_N, m_N^{\mathscr{Y}})\}$, where each $m_i^{\mathscr{Y}}$ is a bba on the set $\mathscr{Y}$ of classes, encoding all available information about the class of example i. As before, the $\mathbf{x}_i$ will be assumed to have been generated according to the mixture model defined in Section 3.1. Our goal is to extend the previous method to estimate the model parameters from such dataset. For that purpose, an objective function generalizing the likelihood function needs to be defined.

The concept of likelihood function has strong relations with that of possibility and, more generally, plausibility, as already noted by several authors [13, 15, 16]. Furthermore, selecting the simple hypothesis with highest plausibility given the observations $\mathbf{X}^{iu}$ is a natural decision strategy in the belief function framework [2]. We thus propose as an estimation principle to search for the value of parameter ψ with maximal conditional plausibility given the data: $\hat{\psi} = \arg\max_\psi pl^\Psi(\psi|\mathbf{X}^{iu})$. The correctness of the intuition leading to this choice of criterion as an estimation principle seems to be confirmed by the fact that the logarithm of $pl^\Psi(\psi|\mathbf{X}^{iu})$ is an immediate generalization of criterion (6), and the other likelihood criteria used for semi-supervised learning and partially supervised learning of mixture model, as shown by the following proposition.

Proposition 1. *If the samples* $\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}$ *are drawn independently according to the generative mixture model setting and if the soft labels* $\{m_1,\ldots,m_N\}$ *are independent*

from the parameters values, then the logarithm of the conditional plausibility of Ψ *given* $\mathbf{X}^{iu}$ *is given by*

$$\ln\left(pl^{\Psi}(\psi|\mathbf{X}^{iu})\right) = \sum_{i=1}^{N} \ln\left(\sum_{k=1}^{K} pl_{ik}.\pi_k f(\mathbf{x}_i;\theta_k)\right) + \nu, \tag{9}$$

where the pl_{ik} *are the plausibilities of each class* k *for each sample* i *according to soft labels* m_i *and* ν *is a constant independent of* ψ.

Proof. Using the GBT (5), the plausibility of parameters can be expressed from the plausibility of the observed values. By making the conditional independence assumption, this plausibility can be decomposed as a product over samples. Using the Total Plausibility Theorem (4), we may express the plausibility of an observed value as:

$$pl^{\mathscr{X}_i}(\mathbf{x}_i|\psi) = \sum_{C\subseteq\mathscr{Y}} m^{\mathscr{Y}_i}(C|\psi) pl^{\mathscr{X}_i|\mathscr{Y}_i}(\mathbf{x}_i|C,\psi), \tag{10}$$

where $m^{\mathscr{Y}_i}(.|\psi)$ is a bba representing our beliefs regarding the class of example i. This bba comes from the combination of two information sources: the "soft" label $m_i^{\mathscr{Y}}$ and the proportions π, which induce a Bayesian bba $m^{\mathscr{Y}}(\cdot|\pi)$. As these two sources are supposed to be distinct, they can be combined using the conjunctive rule (2). As $m^{\mathscr{Y}}(\cdot|\pi)$ is Bayesian, the same property holds for the result of the combination $m^{\mathscr{Y}_i}(.|\psi)$ and we have $m^{\mathscr{Y}_i}(\{c_k\}|\psi) = pl_{ik}\,\pi_k$. Therefore, in the right-hand side of (10), the only terms in the sum that need to be considered are those corresponding to the singletons. Consequently, we only need to express $pl^{\mathscr{X}_i|\mathscr{Y}_i}(\mathbf{x}_i|c_k,\psi)$ for all k. There is a difficulty at this stage, since $pl^{\mathscr{X}_i|\mathscr{Y}_i}(\cdot|c_k,\psi)$ is the continuous probability measure with density function $f(\mathbf{x};\theta_k)$: consequently, the plausibility of any single value would be null if observations $\mathbf{x}_i$ had an infinite precision. However, observations always have a finite precision, so that what we denote by $pl^{\mathscr{X}_i|\mathscr{Y}_i}(\mathbf{x}_i|c_k,\psi)$ is in fact the plausibility of a infinitesimal region around $\mathbf{x}_i$ with volume $\mathrm{d}x_{i1}\ldots\mathrm{d}x_{ip}$ (where p is the feature space dimension). We thus have $pl^{\mathscr{X}_i|\mathscr{Y}_i}(\mathbf{x}_i|c_k,\psi) = f(\mathbf{x}_i;\theta_k)\mathrm{d}x_{i1}\ldots\mathrm{d}x_{ip}$. Using all this results we obtain $pl^{\Psi}(\psi|\mathbf{X}^{iu}) = \prod_{i=1}^{N}\left[\left(\sum_{k=1}^{K} pl_{ik}\pi_k f(\mathbf{x}_i;\theta_k)\right)\mathrm{d}x_{i1}\ldots\mathrm{d}x_{ip}\right]$. The terms $\mathrm{d}x_{ij}$ can be considered as multiplicative constants that do not affect the optimization problem. By taking the logarithm we get (9), which completes the proof. □

Remark 1. Our approach can be shown to extend unsupervised, partially supervised and semi-supervised learning when the labels are, respectively, vacuous, categorical, and either vacuous or certain. This justifies denoting criterion, (9) as $L(\Psi,\mathbf{X}^{iu})$, as it generalizes the classical log-likelihood function.

4.2 EM Algorithm for Imprecise and Uncertain Labels

Once the criterion is defined, the remaining work concerns its optimization. This section presents a variant of the EM algorithm dedicated to this task. To build an EM algorithm able to optimize $L(\Psi;\mathbf{X}^{iu})$, we follow a path that parallels the one recalled in Section 3.2. At iteration q, our knowledge of the class of example i given the current parameter estimates comes from three sources: the class label $m_i^{\mathscr{Y}}$ of example i; the

current estimates $\pi^{(q)}$ of the proportions; the vector $\mathbf{x}_i$ and the current parameter estimate $\theta^{(q)}$, which, using the GBT (5), gives a plausibility over $\mathscr{Y}$. By combining these three items of evidence using Dempster's rule (3), we get a Bayesian bba. Let us denote by $t_{ik}^{(q)}$ the mass assigned to $\{c_k\}$ after combination. We have

$$t_{ik}^{(q)} = \frac{pl_{ik}\pi_k^{(q)} f(\mathbf{x}_i;\theta_k^{(q)})}{\sum_{k'=1}^{K} pl_{ik'}\pi_{k'}^{(q)} f(\mathbf{x}_i;\theta_{k'}^{(q)})}, \tag{11}$$

Using this expression, we may decompose the log-likelihood in two parts, as in (7).

$$L(\Psi;\mathbf{X}^{iu}) = \sum_{i=1}^{N}\sum_{k=1}^{K} t_{ik}^{(q)} \ln\left(\pi_k pl_{ik} f(\mathbf{x}_i;\theta_k)\right) - \sum_{i=1}^{N}\sum_{k=1}^{K} t_{ik}^{(q)} \ln\left(\frac{\pi_k pl_{ik} f(\mathbf{x}_i;\theta_k)}{\sum_{k'=1}^{K}\pi_{k'} pl_{ik'} f(\mathbf{x}_i;\theta_{k'})}\right) \tag{12}$$

This decomposition can be established thanks to basic properties of logarithmic functions and the fact that $\sum_{i=1}^{K} t_{ik}^{(q)} = 1$. Therefore, using the same argument as for the classical EM algorithm (Section 3.2), an algorithm which alternates between computing t_{ik} using (11) and maximization of the first term in the right hand side of (12) will increase our criterion. This algorithm is therefore the classical EM algorithm, except for the E step, where the posterior distributions t_{ik} are weighted by the plausibility of each class. During the M step the proportions are updated classically using $\pi_k^{(q+1)} = \frac{1}{N}\sum_{i=1}^{N} t_{ik}^{(q)}$. If multivariate normal densities functions are considered, $f(x;\theta_k) = \mathcal{N}(x;\mu_k,\Sigma_k)$, their parameters are updated using the following equations:

$$\mu_k^{(q+1)} = \frac{1}{\sum_{i=1}^{N} t_{ik}^{(q)}}\sum_{i=1}^{N} t_{ik}^{(q)}\mathbf{x}_i, \quad \Sigma_k^{(q+1)} = \frac{1}{\sum_{i=1}^{N} t_{ik}^{(q)}}\sum_{i=1}^{N} t_{ik}^{(q)}(\mathbf{x}_i-\mu_k^{(q+1)})(\mathbf{x}_i-\mu_k^{(q+1)})'. \tag{13}$$

4.3 Comparison with Previous Work

The idea of adapting the EM algorithm to handle soft labels can be traced back to the work of Vannoorenberghe and Smets [21], which was recently extended to categorical data by Jraidi et al. [12]. These authors proposed a variant of the EM algorithm called CrEM (Credal EM), based on a modification of the auxiliary function $Q(\Psi,\Psi^{(q)})$. However, our method differs from this previous approach in several respects. First, the CrEM algorithm was not derived as optimizing a generalized likelihood criterion such as (9); consequently, its interpretation was unclear, the relationship with related work (see Remark 1) could not be highlighted and, most importantly, the convergence of the algorithm was not proven. Furthermore, in our approach, the soft labels $m_i^{\mathscr{Y}}$ appear in the criterion and in the update formulas for posterior probabilities (11) only in the form of the plausibilities pl_{ik} of the singletons. In constrast, the CrEM algorithm uses the $2^{|\mathscr{Y}|}$ values in each bba $m_i^{\mathscr{Y}}$. This fact has an important consequence, as the computations involved in the E step of the CrEM algorithm have a complexity in $O(2^{|\mathscr{Y}|})$ whereas our solution only involves calculations which scale with the cardinality of the set of classes.

5 Simulations

The experiment presented in this section aimed at using information on class labels simulating expert opinions. As a reasonable setting, we assumed that the expert supplies, for each sample i, his/her more likely label c_k and a measure of doubt p_i. This doubt is represented by a number in $[0,1]$, which can be seen as the probability that the expert knows nothing about the true label. To handle this additional information in the belief function framework, it is natural to *discount* the categorical bba associated to the guessed label with a discount rate p_i ([14], p. 251). Thus, the imperfect labels built from expert opinions are simple bbas such that $m_i^{\mathscr{Y}}(\{c_{k^*}\}) = 1 - p_i$ for some k^*, and $m_i^{\mathscr{Y}}(\mathscr{Y}) = p_i$. The corresponding plausibilities are $pl_{ik^*} = 1$ and $pl_{ik} = p_i$ for all $k \neq k^*$.

Simulated data sets were build as follows. Two data sets of size $N \in \{2000, 4000\}$ were generated in a ten-dimensional feature space from a two component normal mixture with common identity covariance matrix and balanced proportions. The distance between the two centers was kept fixed at $\delta = 2$. For each training sample i, a number p_i was drawn from a specific probability distribution to define the doubt expressed by a hypothetical expert on the class of that sample. With probability $(1 - p_i)$, the true label of sample i was kept and with probability p_i the expert's label was drawn uniformly in the set of all class. The probability distribution used to draw the p_i specifies the expert's labelling error rate. For our experiments we used Beta distributions with expected value equal to $\{0.1, \ldots, 0.8\}$ and variance kept equal to 0.2.

The results of our approach were compared to *supervised learning* using the potentially wrong expert's labels; *unsupervised learning*, which does not use any information on class label coming from experts, and a strategy based on *semi-supervised learning* which takes into account the reliability of labels supplied by the p_i. This strategy considers each sample as labelled if the expert's doubt is moderate $(p_i \leq 0.5)$ and as unlabelled otherwise $(p_i > 0.5)$. Figure 1 shows the averaged performances of the different classifiers trained with one hundred independent training sets. As expected, when the expert's doubt increases, the error rate of supervised learning also increases.

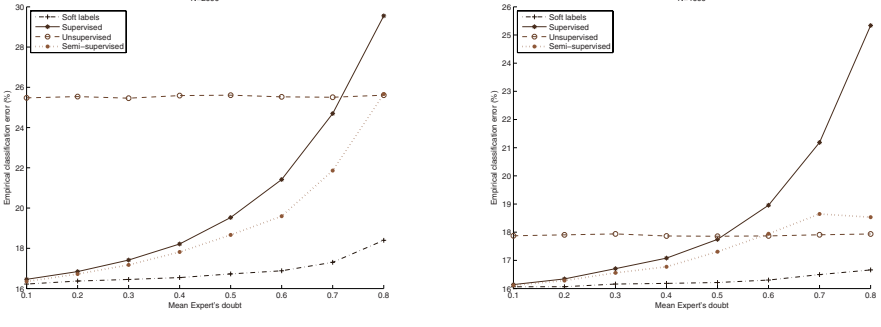


Fig. 1. Empirical classification error (%, estimated on a test set of 5000 observations) averaged over one hundred independent training sets, as a function of expert's mean doubt and for different sample size. For all methods, the EM algorithm was initialized with the true parameter values.

Our solution based on soft labels does not suffer as much as supervised learning and adaptive semi-supervised learning from label noise. Whatever the dataset size, our solution takes advantage of additional information on the reliability of labels to keep good performances. Finally, our approach clearly outperforms unsupervised learning, when the number of samples is low ($N = 2000$).

6 Conclusions

The approach presented in this paper, based on concepts coming from maximum likelihood estimation and belief function theory, offers an interesting way to deal with imperfect and imprecise labels. The proposed criterion has a natural expression that is closely related to previous solutions found in the context of probabilistic models, and has also a clear and justified origin in the context of belief functions. Moreover, the practical interest of imprecise and imperfect labels, as a solution to deal with label noise, has been highlighted by an experimental study using simulated data.

References

1. Ambroise, C., Govaert, G.: EM algorithm for partially known labels. In: Proc. of IFCS 2000 (Namur, Belgium), pp. 161–166. Springer, Heidelberg (2000)
2. Cobb, B.R., Shenoy, P.P.: On the plausibility transformation method for translating belief function models to probability models. Internat. J. Approx. Reason 41(3), 314–330 (2006)
3. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statist. 38, 325–339 (1967)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B. Stat. Methodol. 39, 1–38 (1977)
5. Denœux, T.: A *k*-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans. Systems Man Cybernet 25(05), 804–813 (1995)
6. Denœux, T., Zouhal, L.M.: Handling possibilistic labels in pattern classification using evidential reasoning. Fuzzy Sets Syst. 122(3), 47–62 (2001)
7. Dubois, D., Prade, H.: On the unicity of Dempster's rule of combination. Int. J. Intell. Syst. 1, 133–142 (1986)
8. Elouedi, Z., Mellouli, K., Smets, P.: Belief decision trees: Theoretical foundations. Internat. J. Approx. Reason 28, 91–124 (2001)
9. Grandvallet, Y.: Logistic regression for partial labels. In: Proc. of IPMU 2002 (Annecy, France), vol. III, pp. 1935–1941 (2002)
10. Hosmer, D.W.: A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. Biometrics 29, 761–770 (1973)
11. Hüllermeier, E., Beringer, J.: Learning from ambiguously labeled examples. In: Famili, A.F., Kok, J.N., Peña, J.M., Siebes, A., Feelders, A. (eds.) IDA 2005. LNCS, vol. 3646. Springer, Heidelberg (2005)
12. Jraidi, I., Elouedi, Z.: Belief classification approach based on generalized credal EM. In: Mellouli, K. (ed.) ECSQARU 2007. LNCS (LNAI), vol. 4724, pp. 524–535. Springer, Heidelberg (2007)
13. Monney, P.-A.: A Mathematical Theory of Arguments for Statistical Evidence. Contributions to Statistics. Physica-Verlag, Heidelberg (2003)

14. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
15. Shenoy, P.P., Giang, P.H.: Decision making on the sole basis of statistical likelihood. Artificial Intelligent 165(2), 137–163 (2005)
16. Smets, P.: Possibilistic inference from statistical data. In: Proc. of WCMSM 1982 (Las-Palmas, Spain), pp. 611–613 (1982)
17. Smets, P.: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Internat. J. Approx. Reason 9, 1–35 (1993)
18. Smets, P.: Quantifying beliefs by belief functions: An axiomatic justification. In: Proc. of IJCAI 1993 (Chambéry), vol. 1, pp. 598–603 (1993)
19. Smets, P.: Belief functions on real numbers. Internat. J. Approx. Reason 40(3), 181–223 (2005)
20. Smets, P., Kennes, R.: The Transferable Belief Model. Artificial Intelligence 66, 191–243 (1994)
21. Vannoorenberghe, P., Smets, P.: Partially supervised learning by a credal EM approach. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 956–967. Springer, Heidelberg (2005)

Imprecise Functional Estimation: The Cumulative Distribution Case

Kevin Loquin and Olivier Strauss

Laboratoire d'Informatique, de Robotique et de Microéléctronique de Montpellier (LIRMM), Montpellier, France

Abstract. In this paper, we propose an adaptation of the Parzen Rosenblatt cumulative distribution function estimator that uses maxitive kernels. The result of this estimator, on every point of the domain of F, the cumulative distribution to be estimated, is interval valued instead of punctual valued. We prove the consistency of our approach with the classical Parzen Rosenblatt estimator, since, according to consistency conditions between the maxitive kernel involved in the imprecise estimator and the summative kernel involved in the precise estimator, our imprecise estimate contains the precise Parzen Rosenblatt estimate.

Keywords: Parzen Rosenblatt, Cumulative distribution, Imprecise functional estimation, Possibility distribution, Choquet integral.

1 Introduction

The probability density function (pdf) f and the cumulative distribution function (cdf) F of a random variable X on $\Omega \subseteq \mathbb{R}$ are fundamental concepts for describing and representing real data in statistics. These representations are linked by $\forall \omega \in \Omega$, $F(\omega) = \int_{-\infty}^{\omega} f(u)du$. When they cannot be specified, estimates of these functions may be performed by using a sample of n observations independent and identically distributed $(X_1,...,X_n)$ of X. These observations are summarized by the empirical distribution defined by $e_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$, where δ_{X_i} is the Dirac distribution on X_i or by the empirical cumulative distribution function defined on Ω by $E_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[X_i \leq x]}$, where $\mathbb{1}_A$ is the characteristic function on A.

Different methods have been proposed in the literature for estimating or manipulating the pdf or the cdf underlying a sample of observations. The Parzen Rosenblatt method is one of the most efficient non-parametric techniques [10, 11]. It belongs to the class of functional estimation methods.

Generally speaking, functional estimation [1] consists of estimating, for all $x \in \Omega$, a function $h : \Omega \longrightarrow \mathbb{R}$ from another function $g : \Omega \longrightarrow \mathbb{R}$ related to h. The nature of this relation between h and g can take different form: g can be replaced by a sequence $(g_n)_{n>0}$, such that $g_n \longrightarrow h$, when $n \longrightarrow +\infty$, h can be a modification of g (or h is a filtered signal obtained from the signal g), or g can be a discretization of h that has to be recovered by interpolation. So, the estimate of h, at $x \in \Omega$, is function of g and x, which can be expressed as $\hat{h}(x) = \varphi(g,x)$.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 175–182, 2008.
springerlink.com

For the Parzen Rosenblatt pdf estimator, the function g is the empirical distribution e_n, h is the pdf to be estimated f. The estimator $\hat{h}$ is defined for all $x \in \Omega$, by:

$$\hat{h}(x) = f_{n\kappa_\Delta}(x) = \frac{1}{n}\sum_{i=1}^{n} \kappa_\Delta(x - X_i), \tag{1}$$

with κ the kernel used to perform this estimate and Δ the bandwidth. Note that $\kappa_\Delta(x) = \frac{1}{\Delta}\kappa(\frac{x}{\Delta})$. When g is the empirical cumulative distribution E_n, the cdf F is the function h to be estimated and the estimator $\hat{h}$ is defined for all $x \in \Omega$, by:

$$\hat{h}(x) = F_{n\kappa_\Delta}(x) = \int_{-\infty}^{x} f_{n\kappa_\Delta}(u)du. \tag{2}$$

In the Parzen Rosenblatt like methods, and more generally in all the functional estimation methods, the particular role of the kernel is to define a neighborhood that can be shifted to any location of Ω. The classical (precise) approach makes use of summative kernels. A summative kernel can be seen as a probability distribution, defining a probabilistic neighborhood around each location x of Ω.

This paper considers a new approach (imprecise) that makes use of maxitive kernels. A maxitive kernel can be seen as a possibility distribution, defining a possibilistic neighborhood around each location x of Ω. The main consequence of replacing a summative kernel by a maxitive kernel is that the estimated value is an interval $[\underline{h}(x), \overline{h}(x)]$, instead of a single value $\hat{h}(x)$. We are interested in the relation between the point estimate obtained with the classical approach and the interval estimate obtained with our approach.

The paper is organized as follows. In Section 2 we present the classical functional estimation using a summative kernel. In Section 3, functional estimation with maxitive kernels is exposed. In Section 4, the imprecise functional estimation is presented and mathematically justified. In Section 5, we apply our method to the Parzen Rosenblatt cdf estimator. Before concluding, we discuss in Section 6 of the choice of the involved maxitive kernel. The method is illustrated by an experiment.

2 Functional Estimation with Summative Neighborhoods

In functional estimation, a summative kernel can be considered as a weighted neighborhood of a given location, called its mode, formally similar to a probability distribution.

Definition 1. *Summative kernels are $\mathbb{R}^+$-valued functions κ defined on a domain Ω, verifying the summativity property: $\int_\Omega \kappa(x)dx = 1$.*

Note that any given monomodal summative kernel κ, can be the basis for a family of summative kernels tuned by a location-scale parameter $\theta = (u, \Delta)$, with u a translation factor and $\Delta > 0$ its bandwidth. Any element of this family is obtained, for $u \in \Omega$ and $\Delta > 0$, by

$$\kappa_\Delta^u(\omega) = \frac{1}{\Delta}\kappa(\frac{\omega - u}{\Delta}),\ \forall \omega \in \Omega. \tag{3}$$

When seen as a probability distribution, a summative kernel κ has a relevant meaning in the scope of uncertainty theories. It induces a probability measure given by

$P_\kappa(A) = \int_A \kappa(\omega)d\omega,\ \forall A \subseteq \Omega$. The value $P_\kappa(A)$ can be interpreted as the degree of probability for a realization of the underlying uncertain phenomenon to fall in A.

Estimation of a given function of $h : \Omega \to \mathbb{R}$ in a summative neighborhood κ_Δ^x of a given location x with bandwidth Δ is given by the expectation of its related function g according to the probability measure $P_{\kappa_\Delta^x}$:

$$\hat{h}(x) = \mathbb{E}_{\kappa_\Delta^x}(g). \tag{4}$$

This approach can be found in [1] for functional estimation in statistics. [7] presents digital signal processing methods that can be reformulated as functional estimators (4).

3 Functional Estimation with Maxitive Neighborhoods

A maxitive kernel is also a weighted neighborhood of a given location, called its mode, formally similar to a possibility distribution or membership function of a normalized fuzzy subset [3].

Definition 2. *A maxitive kernel is a* $[0,1]$*-valued function* π*, defined on a domain* Ω*, verifying the maxitivity property:* $\sup_{\omega\in\Omega} \pi(\omega) = 1$.

Note that any given monomodal maxitive kernel π, defined on Ω, can be the basis for a family of maxitive kernels tuned by a location-scale parameter $\theta = (u,\Delta)$, with u a translation factor and Δ its bandwidth. Any element of this family is obtained, for $u \in \Omega$ and $\Delta > 0$, by

$$\pi_\Delta^u(\omega) = \pi(\frac{\omega - u}{\Delta}),\ \forall \omega \in \Omega. \tag{5}$$

A possibility distribution π has a relevant meaning in the scope of uncertainty theories. π induces a possibility measure given by $\Pi_\pi(A) = \sup_{\omega\in A} \pi(\omega),\ \forall A \subseteq \Omega$. The value $\Pi_\pi(A)$ can be interpreted as the degree of possibility for a realization of the underlying uncertain phenomenon to fall in A.

Now, when the summative neighborhood κ_Δ^x is replaced by a maxitive neighborhood π_Δ^x of a given location x with bandwidth Δ, the Lebesgue integral in estimator (4) has to be replaced by the Choquet integral [2, 9] of g.

4 Imprecise Functional Estimation

A possibility measure is a special case of concave Choquet capacity ν [15]. The conjugate ν^c of such a capacity, defined by $\nu^c(A) = 1 - \nu(A^c), \forall A \subseteq \Omega$, is a convex capacity. A concave capacity ν can encode a special family of probability measures, noted $core(\nu)$ and defined by

$$core(\nu) = \{P_\kappa,\ |\ \forall A \subseteq \Omega, \nu^c(A) \leq P_\kappa(A) \leq \nu(A)\}. \tag{6}$$

David Schmeidler and Dieter Denneberg proved the following theorem ([12, Proposition 3] and [2, Proposition 10.3]) for capacities.

Theorem 1. *The capacity* ν *is concave if and only if for all* g *such that* $\mathbb{C}_\nu(|g|) < +\infty$*, then* $\forall \kappa | P_\kappa \in core(\nu),\ \mathbb{C}_{\nu^c}(g) \leq \mathbb{E}_\kappa(g) \leq \mathbb{C}_\nu(g)$.

From Theorem 1, since a maxitive kernel defines a possibility measure, a maxitive kernel-based estimation of h, generalizing expression (4) is interval valued. The upper and lower bounds are the Choquet integrals of g computed with respectively $\Pi_{\pi_\Delta^x}$ and $N_{\pi_\Delta^x}$, which are capacities (or non additive measures) associated to π_Δ^x a maxitive neighborhood of x, with bandwidth Δ. $N_{\pi_\Delta^x}$ is the conjugate of the possibility measure $\Pi_{\pi_\Delta^x}$, called a necessity measure. These remarks leads to the following corollary of Theorem 1.

Corollary 1. Imprecise functional estimation
Let π be a maxitive kernel, then $\forall x \in \Omega$ and $\forall \Delta > 0$,

$$\forall \kappa | P_\kappa \in core(\Pi_{\pi_\Delta^x}),\ \mathbb{C}_{N_{\pi_\Delta^x}}(g) \leq \mathbb{E}_\kappa(g) \leq \mathbb{C}_{\Pi_{\pi_\Delta^x}}(g). \tag{7}$$

Imprecise estimation of a given function of $h : \Omega \to \mathbb{R}$ in a maxitive neighborhood π_Δ^x of a given location x with bandwidth Δ is given by the Choquet integrals of its related function g according to the possibility and necessity measures $\Pi_{\pi_\Delta^x}$ and $N_{\pi_\Delta^x}$:

$$[\underline{h}(x), \overline{h}(x)] = [\mathbb{C}_{N_{\pi_\Delta^x}}(g), \mathbb{C}_{\Pi_{\pi_\Delta^x}}(g)]. \tag{8}$$

According to Corollary 1, an estimate $\hat{h}(x)$ of h obtained with a summative kernel κ, such that P_κ belongs to $core(\Pi_{\pi_\Delta^x})$, belongs to the estimated interval (8). Besides, the estimation bounds are attained, i.e. there exist two summative kernels η and μ, whose associated probability measures P_η and P_μ are in $core(\Pi_{\pi_\Delta^x})$, such that $\mathbb{E}_\eta(g) = \mathbb{C}_{N_{\pi_\Delta^x}}(g)$ and $\mathbb{E}_\mu(g) = \mathbb{C}_{\Pi_{\pi_\Delta^x}}(g)$.

Replacing a summative kernel by a maxitive kernel for estimating a function h aims at taking into account the imperfect knowledge of the modeler to choose a particular κ. The specificity [16, 8] of the maxitive kernel chosen by the modeler for performing this imprecise estimation reflects his knowledge. The most specific is the maxitive neighborhood, the smallest is the encoded set. Indeed, if π is more specific than π', some summative kernels encoded by π' will not be encoded by π. The smaller is the encoded set of summative neighborhoods, the closer are the estimation bounds with this method.

5 Imprecise Cumulative Distribution Function Estimation

The Parzen Rosenblatt density estimator (1) can be expressed as the estimation of the pdf f, with the empirical distribution $g = e_n$ (summarizing the observations) according to a summative neighborhood κ_Δ^x (see expression (3)):

$$f_{n\kappa_\Delta}(x) = \mathbb{E}_{\kappa_\Delta^x}(e_n). \tag{9}$$

Corollary 1, associated with expression (9) suggests that an imprecise estimation of the Parzen Rosenblatt pdf estimator should be performed by computing the Choquet integral of the empirical distribution e_n according to a maxitive kernel (encoding a family of summative kernels). This direct approach is however not applicable here, since the Choquet integral of the empirical distribution does not exist. Indeed, the computation of this integral only exists for bounded functions. The empirical distribution is not bounded. Actually, the Dirac delta functions, forming e_n, are not functions but mathematical constructions, called distributions.

Nevertheless, the Parzen Rosenblatt cdf estimator (2) involves the empirical cumulative distribution E_n, which is a bounded function. Theorem 2 expresses the Parzen Rosenblatt cdf estimate at a point x, as the estimation of the cdf F with the cumulative empirical distribution $g = E_n$ according to a summative neighborhood of x, κ_Δ^x.

Theorem 2. *Let κ be a summative kernel and $\Delta > 0$ and $n > 0$, then $\forall x \in \Omega$,*

$$F_{n\kappa_\Delta}(x) = \mathbb{E}_{\kappa_\Delta^x}(E_n). \tag{10}$$

Proof. First, note that $f_{n\kappa_\Delta}(x) = \int_\Omega \kappa_\Delta(\omega) e_n(x-\omega)d\omega$. Indeed, $\int_\Omega \kappa_\Delta(\omega)e_n(x-\omega)d\omega = \frac{1}{n}\sum_{i=1}^n \int_\Omega \kappa_\Delta(\omega)\delta_{X_i}(x-\omega)d\omega = \frac{1}{n}\sum_{i=1}^n \kappa_\Delta(x-X_i)$. Thus,

$$\begin{aligned} F_{n\kappa_\Delta}(x) &= \int_{-\infty}^x \Big(\int_\Omega \kappa_\Delta(\omega) e_n(u-\omega)d\omega \Big) du \\ &= \int_\Omega \Big(\int_{-\infty}^x e_n(u-\omega)du \Big) \kappa_\Delta(\omega)d\omega, \end{aligned}$$

E_n is the cumulative distribution associated to the empirical distribution, i.e. $E_n(\omega) = \int_{-\infty}^\omega e_n(u)du$. Then by successive changes of variable $v := u - \omega$ and $t := x - \omega$, we obtain:

$$\begin{aligned} F_{n\kappa_\Delta}(x) &= \int_\Omega E_n(x-\omega)\kappa_\Delta(\omega)d\omega \\ &= \int_\Omega E_n(t)\kappa_\Delta(x-t)dt \\ &= \mathbb{E}_{\kappa_\Delta^x}(E_n). \end{aligned}$$

□

Since E_n is bounded, an imprecise estimation of F at x can be obtained with a maxitive kernel π_Δ^x.

Theorem 3. *Let π be a maxitive kernel, then $\forall x \in \Omega$, $\forall n > 0$ and $\forall \Delta > 0$,*

$$\forall \kappa_{\Delta'} | P_{\kappa_{\Delta'}} \in core(\Pi_{\pi_\Delta^x}),\ \mathbb{C}_{N_{\pi_\Delta^x}}(E_n) \leq F_{n\kappa_{\Delta'}}(x) \leq \mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n). \tag{11}$$

We now present the computation of the imprecise Parzen Rosenblatt cdf estimate. First, observe that E_n is a simple function that can be expressed on Ω by $E_n(\omega) = \sum_{i=1}^n \frac{i}{n}\mathbb{1}_{[X_{(i)},X_{(i+1)}]}$, where $(.)$ indicates a permutation of the observations such that $X_{(i)} \leq X_{(i+1)}$. Thus, the Choquet integral of E_n can be rewritten as $\mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^n \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : E_n(\omega) \geq \frac{i}{n}\})$. It can easily be observed that $\{\omega \in \Omega : E_n(\omega) \geq \frac{i}{n}\} = \{\omega \in \Omega : \omega \geq X_{(i)}\}$. Since the summation does not depend on the order of the summed elements, $\mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^n \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : \omega \geq X_i\})$. With similar developments on $\mathbb{C}_{N_{\pi_\Delta^x}}(E_n)$, we obtain:

$$\begin{aligned} \mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) &= \frac{1}{n}\sum_{i=1}^n \Big(1 - N_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < X_i\})\Big), \\ \mathbb{C}_{N_{\pi_\Delta^x}}(E_n) &= \frac{1}{n}\sum_{i=1}^n \Big(1 - \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < X_i\})\Big). \end{aligned}$$

As exposed in [3, 5], $\underline{F_{\pi_\Delta^x}}(u) = N_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < u\})$ is the lower cdf of the set of cdf associated to the summative kernels of $core(\Pi_{\pi_\Delta^x})$. It is the lower cdf of a p-box [6], whose upper cdf is given by $\overline{F_{\pi_\Delta^x}}(u) = \Pi_{\pi_\Delta^x}(\{\omega \in \Omega : \omega < u\})$. As shown in [5], we have:

$$\underline{F_{\pi_\Delta^x}}(u) = \begin{cases} 0 & \text{if } u < x, \\ 1 - \pi_\Delta^x(u) & \text{otherwise,} \end{cases} \quad \text{and} \quad \overline{F_{\pi_\Delta^x}}(u) = \begin{cases} \pi_\Delta^x(u) & \text{if } u < x, \\ 1 & \text{otherwise.} \end{cases}$$

We thus obtain the imprecise cdf estimate:

$$\mathbb{C}_{\Pi_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^{n}\left(\pi_\Delta^x(X_i)\mathbb{1}_{[x \le X_i]} + \mathbb{1}_{[x > X_i]}\right), \tag{12}$$

$$\mathbb{C}_{N_{\pi_\Delta^x}}(E_n) = \frac{1}{n}\sum_{i=1}^{n}\left((1 - \pi_\Delta^x(X_i))\mathbb{1}_{[x \ge X_i]}\right). \tag{13}$$

6 Experiment and Choice of a Maxitive Kernel

As in the case of the summative kernel methods, the problem of the choice of a particular maxitive kernel for performing imprecise functional estimation can be discussed. The choice of the summative kernel shape κ is often considered as insignificant in the non-parametric statistics community. The main argument is that the asymptotic behavior (when $n \to +\infty$) of $F_{n\kappa_\Delta}$ and $f_{n\kappa_\Delta}$ depend more on Δ than on the choice of κ [14, 1]. However, the asymptotic conditions are barely fulfilled. In non-asymptotic conditions, the shape of the estimate strongly depend on the shape of κ. Moreover, the knowledge of the modeler is generally insufficient for choosing the appropriate kernel. Instead of choosing one particular summative kernel, we propose to the modeler to choose a family of summative kernels matching his knowledge via the choice of a maxitive kernel.

In such kernel methods, where a summative kernel is considered as a neighborhood, it seems sensible to assume that the chosen basic kernel to be shifted and dilated with expression (3) is centered, even and with a support included in $[-1, 1]$. Therefore, it naturally leads to choose a basic maxitive kernel π encoding these particular summative kernels. As shown in [4], the triangular maxitive kernel T is the most specific of such maxitive kernels. The triangular possibility distribution is defined on Ω by $T(\omega) = (1 - |\omega|)\mathbb{1}_{[|\omega| \le 1]}$. We now illustrate Theorem 3 by performing the summative and maxitive estimates of the cdf underlying a set of 107 observations of the duration in minutes of the eruptions of the Old Faithful geyser in Yellowstone National Park.[1] Each precise estimate has been performed by using four different summative kernels κ_Δ: uniform, Epanechnikov, triweight and cosine kernels, with $\Delta = 0.3$. The definitions of the used kernels can be found in [8]. The imprecise estimate is obtained by using a triangular maxitive kernel T with the same Δ. As illustrated on Figure 1, every precise estimates of the cumulative distribution are included in the imprecise estimation interval.

[1] This example, taken from [13], is a popular benchmark in nonparametric estimation.

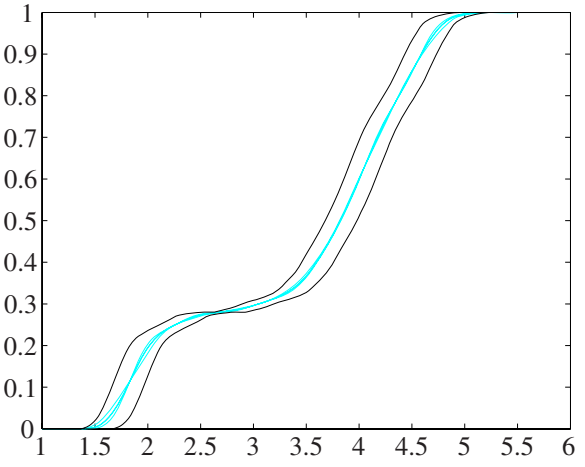


Fig. 1. Imprecise cumulative distribution estimate

7 Conclusion

In this paper, we proposed an extension of the Parzen Rosenblatt cdf estimate, which takes into account a possible lack of knowledge of the appropriate summative kernel to be involved. Compared to the classical method, our method results in an interval estimate instead of a point estimate. The imprecision of the obtained estimate consistently reflects the lack of knowledge of the modeler, quantified by the specificity of the involved maxitive kernel. We put this sensible imprecise cdf estimation into a wider framework of imprecise functional estimation. Now, the next significant step, in soft statistics, is likely to be the imprecise estimation of the pdf.

References

1. Bosq, D., Lecoutre, J.P.: Théorie de IEstimation Fonctionnelle. Economica, Paris (1987)
2. Denneberg, D.: Non Additive Measure and Integral. Kluwer Academic Publishers, Dordrecht (1994)
3. Dubois, D., Prade, H.: Théorie des Possibilités. Applications á la Représentation des Connaissances en Informatique. Collection Méthode + Programmes, 2. Masson, Paris (1985)
4. Dubois, D., Prade, H., Foulloy, L., Mauris, G.: Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. Reliab. Comput. 10, 273–297 (2004)
5. Dubois, D.: Possibility theory and statistical reasonning. Comput. Statist. Data Anal. 51, 47–69 (2006)
6. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Technical Report sand 2002-4015. Sandia National Laboratories, Albuquerque, New Mexico (2002)
7. Jan, J.: Digital Signal Filtering, Analysis and Restoration. Telecommunications Series. IET (2000)

8. Loquin, K., Strauss, O.: On the granularity of summative kernels. Fuzzy Sets Syst. (in press, 2008) doi:10.1016/j.fss.2008.02.021
9. Marichal, J.L.: Aggregation Operations for Multicriteria Decision Aid. PhD Thesis, Department of Mathematics, University of Liège, Liège, Belgium (1998)
10. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065–1076 (1962)
11. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, 832–837 (1956)
12. Schmeidler, D.: Integral representation without additivity. Proc. Amer. Math. Soc. 97, 255–261 (1986)
13. Silvermann, B.W.: Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability, vol. 26. Chapman and Hall, London (1986)
14. Simonoff, J.S.: Smoothing Methods in Statistics. Series in Statistics. Springer, New York (1996)
15. Walley, P.: Towards a unified theory of imprecise probability. Internat. J. Approx. Reason 24(2-3), 125–148 (2000)
16. Yager, R.R.: Entropy and specificity in a mathematical theory of evidence. Int. J. Gen. Syst. 9, 249–260 (1983)

Non-parametric Density Estimation Based on Label Semantics

Jonathan Lawry[1] and Inés González-Rodríguez[2]

[1] Department of Engineering Mathematics, University of Bristol, Bristol, United Kingdom
[2] Dept. of Mathematics, Statistics and Computing, University of Cantabria, Santander, Spain

Abstract. We propose a non-parametric density estimator based on label semantics, a framework for computing with words which allows to describe a numerical instance or set of instances in linguistic terms and to condition on a linguistic description. This will be the basis of the proposed density estimator, which is MSE consistent under certain regularity conditions. Experimental results illustrate the potential of the proposal.

Keywords: Density estimation, Computing with Words, Label semantics.

1 Introduction

Probability density estimation constitutes a classical approach to pattern classifier design as well as being useful in a broad range of applications. The approaches to density estimation are usually classified as *parametric* and *non-parametric* (cf. [10], [11]). In parametric estimation, it is assumed that the density f underlying the data belongs to a family of functions $f(\cdot;\theta)$ with parameters $\theta = (\theta_1, \ldots, \theta_k)$. A density estimate $\hat{f}$ is then obtained by computing from the data an estimate $\hat{\theta}$ of the parameters θ and having $\hat{f} = f(\cdot, \hat{\theta})$.

Non-parametric approaches do not assume a particular distribution shape. They are necessary when the distribution does not fit a known distribution model and are widely used in the field of pattern recognition and classification, neurocomputing, image processing and communications among others. Non-parametric methods comprise fixed and varying width histograms, naive estimator, kernel or Parzen estimator (perhaps, the most popular one), nearest neighbour methods, variable kernel method, orthogonal series estimators, maximum penalised likelihood estimators and general weight function estimators. The latter can be thought of as a unifying concept (histogram, kernel estimates or orthogonal series estimate can be seen as particular cases) and as a method to obtain estimators which do not fall into any of the other classes.

Despite the existing variety of methods, the problem of density estimation is far from being solved and is still subject to new research. For instance, a non-parametric density estimator is proposed in [1] that relies on topological self-organisation as a development of a nearest-neighbour density estimator; in [3] we find a new method of kernel density estimation with a varying adaptive window size; in [2] a semiparametric density estimator is obtained as the product of nonparametric and (conditionally) parametric

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 183–189, 2008.
springerlink.com

factors and this estimator is used to design a classifier, and in [7], the authors propose a histogram density estimator where the bins are fuzzy sets.

Label semantics [4] provides a framework where, for a given set of labels LA, it is possible to give a linguistic description of a single numerical value x or a whole set of data instances $x_1, \ldots, x_n$ and it is also possible to compute the probability of any value x given a linguistic description. In this paper we propose to use as a non-parametric density estimate the density that results from conditioning on the linguistic description of the whole sample. We see that the proposed estimate is MSE consistent and provide experimental results to illustrate the estimate's behaviour.

2 Label Semantics

Label semantics provide an alternative approach to the paradigm of computing with words and have been successfully applied to solving classification and prediction problems [4], [5]. In this section we briefly introduce the ideas most relevant to our work.

The fundamental notion underlying label semantics is that when individuals make assertions of the kind 'X is *tall*' they are essentially providing information about what labels are appropriate for the value of some underlying variable. For simplicity, we assume that for a given context only a finite set of words is available.

Let x be a variable into a domain of discourse Ω. Then, a finite set of words or labels $LA = \{L_1, \ldots, L_n\}$ are identified as possible descriptions of the elements of the universe Ω. For a specific value $x \in \Omega$, an individual I identifies a subset of LA, as the set of words with which it is appropriate to label x. This set is denoted $\mathscr{D}_x^I$, to stand for the description of x given by I.

Consider the expression 'Bill is *tall*', where Bill's height is represented by variable h, and suppose that there is a fixed finite set of possible labels for h, LA, both known and completely identical for any individual who will make or interpret a statement regarding Bill's height. Given these assumptions, the above statement as asserted by a particular individual I might be interpreted as meaning that according to I, *tall* is an appropriate label for the value of variable h. That is, suppose I knows that $h = H$ and that given this information he/she is able to identify a subset of LA consisting of those words appropriate as labels for the value H. This set is $\mathscr{D}_h^I$, the description of h given by I, and we have that $tall \in \mathscr{D}_h^I$.

If we allow I to vary across a population of individuals V, we naturally obtain a random set $\mathscr{D}_x$ from V into the power set of LA, where $\mathscr{D}_x(I) = \mathscr{D}_x^I$. A probability distribution or mass assignment can be defined, dependent on the prior distribution over the population V. We can view the random set $\mathscr{D}_x$ as a description of the variable x in terms of the labels LA.

Definition 1. *For LA a set of labels describing values in Ω, a* mass assignment on labels *is a function $m : 2^{LA} \to [0,1]$ such that*

$$\sum_{T \subseteq LA} m(T) = 1 \tag{1}$$

Notice that in Definition 1 there is no requirement for the mass associated with the empty set to be zero. In the context of label semantics, $m_x(\emptyset)$ quantifies the belief that no labels are appropriate to describe x.

Definition 2. *For $x \in \Omega$ the* label description *of x is a random set from V into the power set of LA, denoted $\mathscr{D}_x$, with associated distribution m_x given by*

$$\forall S \subseteq LA,\ m_x(S) = \Pr(\{I \in V : D_x^I = S\}) \tag{2}$$

Consider again the statement 'Bill is *tall*'. If we allow I to vary across the population of individuals V we can determine a probability distribution for the random set D_h by defining $\forall S \subseteq LA,\ m_h(S) = \Pr(\{I \in V : \mathscr{D}_h^I = S\})$.

Definition 3. *Given labels LA together with an associated mass assignment m_x $\forall x \in \Omega$, the* set of focal elements *for LA is given by:*

$$\mathscr{F} = \{S \in LA : \exists x \in \Omega, m_x(S) > 0\} \tag{3}$$

Another high level measure associated with m_x is the quantification of the degree of appropriateness of a particular word $L \in LA$ as a label of x.

Definition 4. *The* appropriateness degree *of a particular word $L \in LA$ as a label of x is defined as follows:*

$$\forall x \in \Omega\ \forall L \in LA,\ \mu_L(x) = \sum_{S \subseteq LA : L \in S} m_x(S) \tag{4}$$

Clearly, μ_L is a function from Ω into $[0,1]$ and therefore can technically be viewed as a fuzzy set. However, the term 'appropriateness degree' is used partly because it more accurately reflects the underlying semantics and partly to highlight the quite distinct calculus for these functions introduced in the label semantics framework [5].

We now make the additional *consonance assumption* that value descriptions are consonant random sets. In the current context consonance requires the restriction that individuals in V differ regarding what labels are appropriate for a value only in terms of generality or specificity. Certainly, given that the meaning of labels in LA must be sufficiently invariant across V to allow for effective communication then some strong restriction on $\mathscr{D}_x$ should be expected. The consonance restriction could be justified by the idea that all individuals share a common ordering on the appropriateness of labels for a value and that the composition of $\mathscr{D}_x^I$ is consistent with this ordering for each I. For further considerations on the consonance assumption, see [5], [6].

Proposition 1. *Given the consonance assumption, m_x can be completely determined from the values of $\mu_L(x)$ for $L \in LA$. Let $\{\mu_L(x) : L \in LA\} = \{y_1, \ldots, y_n\}$ ordered such that $y_i > y_{i+1}$ for $i = 1, \ldots, n-1$. Then, for $S_i = \{L \in LA : \mu_L(x) \geq y_i\}, i = 1, \ldots, n$ the value description is given by:*

$$m_x(\emptyset) = 1 - y_1,\ m_x(S_i) = y_i - y_{i+1},\ i = 1, \ldots, n-1,\ m_x(S_n) = y_n \tag{5}$$

The above has considerable practical advantages, since we no longer need to have any knowledge of the underlying population of individuals V in order to determine m_x. Rather, for reasoning with label semantics in practice we need only define appropriateness degrees μ_L for $L \in LA$ corresponding to the imprecise definition of each label.

For many types of data analysis it is useful to be able to estimate the distribution underlying variables given the information contained in a data set $D = \{X_1, \ldots, X_n\}$. In the current context, our knowledge of D is represented by a mass assignment m_D defined as follows:

Definition 5. *A* mass assignment conditional on the information provided by D *is given by:*

$$\forall S \in LA, m_D(S) = \sum_{i=1}^{n} P_D(X_i) m_{X_i}(S) \tag{6}$$

where $P_D(X_i)$ corresponds to the probability of X_i being chose at random from D and m_{X_i} is the mass assignment on $\mathscr{D}_{X_i}$, $i = 1, \ldots, n$.

The following definition provides a means of evaluating a distribution on the base variable x conditional on m_D.

Definition 6. *Let x be a variable into Ω with prior distribution $p(x)$, LA be a set of labels for x and m be a posterior mass assignment for the set of appropriate labels of x (i.e., $\mathscr{D}_x$). Then, the* posterior distribution of x conditional on m *is given by:*

$$\forall x \in \Omega, p(x|m) = p(x) \sum_{S \subseteq LA} \frac{m(S)}{pm(S)} m_x(S) \tag{7}$$

where pm is the prior mass assignment generated by the prior distribution p *according to*

$$pm(S) = \int_{\Omega} m_x(S)\, p(x)\, dx \tag{8}$$

This definition is motivated by the following argument. By the Theorem of Total Probability:

$$p(x|m) = \sum_{S \subseteq LA} p(x|\mathscr{D}_x = S) \Pr(\mathscr{D}_x = S) = \sum_{S \subseteq LA} p(x|\mathscr{D}_x = S) m(S) \tag{9}$$

Also,

$$p(x|\mathscr{D}_x = S) = \frac{\Pr(\mathscr{D}_x = S|x) p(x)}{\Pr(\mathscr{D}_x = S)} = \frac{m_x(S) p(x)}{pm(S)} \tag{10}$$

Making the relevant substitutions and then simplifying gives the expression in Definition 6.

Notice that in the case where $\forall S \subseteq LA, m(S) = pm(S)$ it follows that

$$\forall x \in \Omega, p(x|m_x) = p(x) \tag{11}$$

This is intuitive, since if the mass assignment m provides no new information, the conditional density $p(x|m)$ is not expected to differ from the prior $p(x)$.

3 Density Estimation from Data

Let $D = \{X_1, \ldots, X_n\}$ be a sample of data and let us assume that the underlying variable x takes values in a closed interval $\Omega = [l, u] \subseteq \mathbb{R}$ according to an unknown density

function f. Let us also assume that the appropriateness degrees μ_L have a trapezoidal shape for all $L \in LA$. These definitions may be obtained by uniformly partitioning the universe Ω, so the subinterval where $\mu_L(x) = 1$ has a constant width for all $L \in LA$. Alternatively, it is possible to use a *Percentile Method*, whereby each label covers approximately the same number of data elements. This method is quite intuitive and has been successfully used in several applications of label semantics to data mining and machine learning (for instance, in [8] and [9]). In both cases, appropriateness degrees form a pairwise overlapping full linguistic covering as follows:

Definition 7. *The labels in LA form a* full linguistic covering *of the universe Ω if for every value $x \in \Omega$ there exists a label such that its appropriateness degree as a descriptor of x is 1:*

$$\forall x \in \Omega \quad \exists L \in LA : \quad \mu_{L_i} = 1 \tag{12}$$

Additionally, if only two labels can overlap at a time, the covering is said to be pairwise overlapping*, that is, for every value x in the universe there exist at most two labels with nonzero appropriateness degrees:*

$$\forall x \in \Omega \quad \exists i, j \in \{1, \ldots, m\} : \{L \in LA : \mu_L(x) \neq 0\} = \{L_i, L_j\} \tag{13}$$

Given a full linguistic covering of universe Ω built from a sample D, the consonance assumption allows to completely determine m_x for any $x \in \Omega$ based on the appropriateness values $\mu_L(x)$ for $L \in LA$. It is then possible to determine the mass assignment conditional on D, m_D and use label semantics to provide the following density estimator.

Definition 8. *Let $D = \{X_1, \ldots, X_n\}$ be a sample of data where the underlying variable x takes values in a closed interval $\Omega = [l, u] \subseteq \mathbb{R}$ according to an unknown density function f. Let LA be a set of labels forming a full linguistic covering of the universe Ω. The* density estimate *$\hat{f}$ is obtained by conditioning on the mass obtained from D, m_D, assuming a uniform prior distribution on Ω, that is:*

$$\forall x \in \Omega, \hat{f}(x) = p(x|m_D) = C_\Omega \sum_{S \subseteq LA} \frac{m_D(S)}{pm(S)} m_x(S) \tag{14}$$

where C_Ω is the constant function of the uniform density in Ω, m_D is the mass assignment conditional on the information provided by D, pm is the prior mass assignment generated by the uniform distribution and m_x is the mass assignment for $x \in \Omega$ determined from $\{\mu_L(x) : L \in LA\}$.

Notice that there is certain formal similarity between (14) and the fuzzy histogram estimator [7]. However, the motivation here is clearly very different, as the estimator is obtained using label semantics calculus as a posterior distribution conditional on the information provided by the data. Indeed, it would be possible to drop the assumption of a uniform prior distribution in Ω to account for some prior knowledge of the data.

Definition 9. *A measure of the discrepancy of the density estimator $\hat{f}$ from the true density f at a single point is the* mean square error *[11]:*

$$MSE_x(\hat{f}) = E\{\hat{f}(x) - f(x)\}^2 \tag{15}$$

The proposed estimate is consistent with MSE under certain regularity conditions:

Theorem 1. *Let f be a C^2 function with bounded derivatives and let LA be a uniform full linguistic covering of Ω. Then $\forall x \in \Omega$, $\hat{f}$ is consistent in the MSE, that is, if m denotes the number of labels in LA and n the sample size:*

$$m \to \infty, \frac{n}{m} \to \infty \Rightarrow MSE_x(\hat{f}) \to 0. \tag{16}$$

Proof. (Sketch of proof) The proof is very similar to that of binned kernel estimators (cf. [10]), using the decomposition of MSE as a combination of bias and variance at x, $\text{MSE}_x(\hat{f}) = \{E\hat{f}(x) - f(x)\}^2 - \text{var}\hat{f}(x)$. Given the smoothness of f, both the bias and the variance can be rewritten using a Taylor series expansion. The resulting expressions can be seen to be bounded by functions that converge to 0 as the number of labels tends to infinity. □

4 Experimental Results

We now present experimental results for a toy problem where the density to be estimated is a normal mixture density given by:

$$f(x) = \frac{1}{2}\left(N(2,3) + N(8,0.5)\right) \tag{17}$$

We have generated a random sample of 100 data instances. From this sample, we have obtained trapezoidal definitions of five linguistic labels $\{vs, s, m, l, vl\}$ which could be seen as corresponding to "very small", "small", "medium", "large" and "very large"

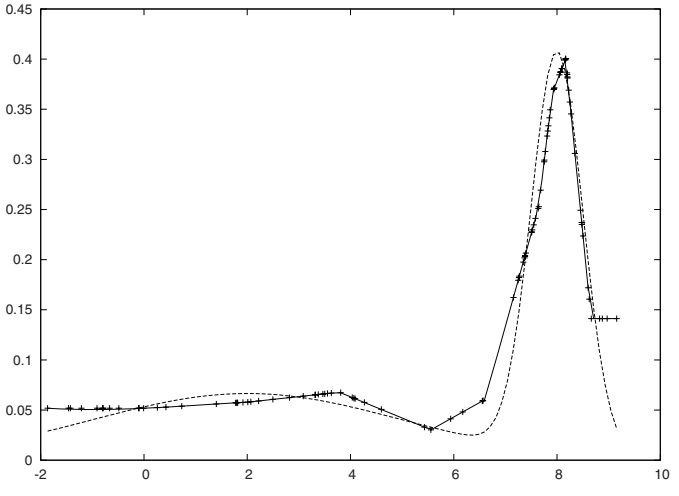


Fig. 1. Estimated density: f (dotted line) and $\hat{f}$ (solid line), with estimated density values for the sample D

using a percentile method. For these labels, the mass assignment conditional on the sample D is given by:

$$m_D = \{vs\} : 0.151,\ \{s\} : 0.12,\ \{s,vs\} : 0.107,\ \{m\} : 0.063,\ \{m,s\} : 0.043,$$
$$\{l\} : 0.095,\ \{l,m\} : 0.165,\ \{vl\} : 0.105,\ \{vl,l\} : 0.151. \quad (18)$$

Figure 1 shows the estimated density function $\hat{f}$ against the original one f as well as the estimated values of the sample data $\hat{f}(X_i)$, $i = 1,\ldots,n$. The average squared error $\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - \hat{f}(X_i))^2$ is $9.13 \cdot 10^{-7}$. If a uniform partition is used instead to obtain the label definitions, the average squared error is $9.704 \cdot 10^{-6}$.

5 Conclusions and Future Work

We have proposed a non-parametric density estimator in the framework of label semantics. It can be proved that this estimator is MSE consistent and we have illustrated its good behaviour with experimental results on a toy problem. In the future, the theoretical properties of the estimator should be further studied, contemplating the use of different definitions for appropriateness degrees or the possibility of introducing some prior knowledge about the underlying density function. Also, the approach should be further tested using data sets from the literature and comparing the results to those of other methods.

Acknowledgement. Inés González-Rodríguez is supported by MEC-Programa José Castillejo Grant JC2007-00152 and MEC-FEDER Grant TIN2007-67466-C02-01.

References

1. Browne, M.: A geometric approach to non-parametric density estimation. Pattern Recognition 40, 134–140 (2007)
2. Hoti, F., Holmström, L.: A semiparametric density estimation approach to pattern classification. Pattern Recognition 37, 409–419 (2004)
3. Katkovnik, V., Shmulevich, I.: Kernel density estimation with adaptive varying window size. Pattern Recognition Lett. 23, 1641–1648 (2002)
4. Lawry, J.: A framework for linguistic modelling. Artificial Intelligence 155, 1–39 (2004)
5. Lawry, J.: Modelling and Reasoning with Vague Concepts. Studies in Computational Intelligence, vol. 12. Springer, Berlin (2006)
6. Lawry, J.: Appropriateness measures: an uncertainty model for vague concepts. Synthese 161(2), 255–269 (2008)
7. Loquin, K., Strauss, O.: Fuzzy histograms and density estimation. In: Lawry, J., Miranda, E., Bugarin, A., Li, S., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing, vol. 37, pp. 45–52. Springer, Berlin (2006)
8. Qin, Z., Lawry, J.: Decision tree learning with fuzzy labels. Inform. Sci. 172, 91–129 (2005)
9. Randon, N.J., Lawry, J.: Classification and query evaluation using modelling with words. Inform. Sci. 176, 438–464 (2006)
10. Scott, D.W.: Multivariate Density Estimation: Theory, Practice and Visualization. Wiley Series in Probability and Statistics. Wiley, New York (1992)
11. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability. Chapman and Hall, London (1986)

Empirical Comparisons of Goodness-of-Fit Tests for Binomial Distributions Based on Fuzzy Representations

Ana Colubi[1], María Ángeles Gil[1], Gil González-Rodríguez[2], and María Teresa López[1]

[1] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain
[2] European Centre for Soft Computing, Mieres, Spain

Abstract. Fuzzy representations of a real-valued random variable have been introduced with the aim of capturing relevant information on the distribution of the variable, through the corresponding fuzzy-valued mean value. In particular, characterizing fuzzy representations of a random variable allow us to capture the whole information on its distribution. One of the implications from this fact is that tests about fuzzy means of fuzzy random variables can be applied to develop goodness-of-fit tests. In this paper we present empirical comparisons of goodness-of-fit tests based on some convenient fuzzy representations with well-known procedures in case the null hypothesis relates to some specified Binomial distributions.

Keywords: Fuzzy representation of a random variable, Fuzzy random variable, Fuzzy mean, Goodness-of-fit test.

1 Introduction

First of all we should clarify that this paper does not deal with statistics with fuzzy data but with statistics referred to real-valued ones. In this way, fuzzy random variables are not considered here to model mechanisms generating fuzzy data, but as an auxiliary tool: by considering an appropriate fuzzy transformation of a real-valued random variable one can obtain a fuzzy random variable whose mean value characterizes the distribution of the original one. A crucial point in this functional characterization is due to the fact that it corresponds to a functionally-valued "mean value", so we can make use of well-known results for the means of functional random elements.

In previous papers (see González-Rodríguez et al. [5], and Colubi et al. [3], [4]) a special family of fuzzy-valued functions defined on $\mathbb{R}$ has been introduced. The composition of each function in this family with any real-valued random variable leads to a fuzzy random variable in Puri and Ralescu's sense [10]: the so-called *fuzzy representation of the original random variable*. Moreover, functions in the family have been chosen to ensure that the functional mean value of the fuzzy representation of a random variable captures relevant information on the distribution of this variable. In some cases such a relevant information concerns the whole distribution. In these cases we will refer to *characterizing fuzzy representations*.

Goodness-of-fit tests are used to test whether a sample of data can be considered to come from a population with a specific distribution. Many goodness-of-fit methods can

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 190–197, 2008.
springerlink.com

be applied to any (or, at least, to a very wide class of) univariate distribution(s). The best known goodness-of-fit statistics can be viewed as certain random dissimilarities or distances between a characterizing functional (distribution function, probability or density function, etc.) of the distribution of the variable under the null hypothesis and an estimator of this functional.

On the other hand, the sample mean of a fuzzy random variable is a consistent estimator of the population one. As a consequence, the sample mean of a characterizing fuzzy representation of a sample from a random variable X will become a consistent estimate of the exact distribution of X (more precisely, a consistent estimate of the population mean value of the fuzzy representation of X).

If we consider a distance between fuzzy values, then we can immediately suggest the goodness-of-fit statistic given by the distance between the sample mean of a characterizing fuzzy representation of the random sample from X and the population mean value of the fuzzy representation of X. In the literature, one can find several studies devoted to the one-sample testing about the mean value of a fuzzy random variable (see Körner [7], Montenegro et al. [9], González-Rodríguez et al. [6]), which can be directly applied to carry out the proposed goodness-of-fit test.

In this paper, some preliminaries about fuzzy values, fuzzy random variables and the associated mean values, as well as a metric between fuzzy values are first recalled. Once two convenient characterizing fuzzy representations of a random variable are presented, we will develop some empirical statistical studies to compare the associated goodness-of-fit tests with classical omnibus tests for goodness-of-fit.

2 Preliminaries

In this section we recall some notions on fuzzy values and fuzzy means which are required to formalize the suggested characterization of random variables.

Let $\mathscr{F}_c(\mathbb{R})$ denote the space of fuzzy numbers, where a *fuzzy number* is a function $U : \mathbb{R} \to [0,1]$ such that the α-level of U (where $U_\alpha = \{x \in \mathbb{R} \,|\, U(x) \geq \alpha\}$ if $\alpha > 0$, $= \mathrm{cl}\{x \in \mathbb{R} \,|\, U(x) > 0\}$ otherwise) is a nonempty compact interval $[\inf U_\alpha, \sup U_\alpha]$ for each $\alpha \in [0,1]$.

Some basic operations between data will be later used, namely, the sum and the product by a 'scalar'. The application of Zadeh's extension principle [11] on $\mathscr{F}_c(\mathbb{R})$ is equivalent to consider the interval-valued arithmetic for the corresponding α-levels, so that for each $\alpha \in [0,1]$, if $U, V \in \mathscr{F}_c(\mathbb{R})$ and $\lambda \in \mathbb{R}$

$$(U+V)_\alpha = [\inf U_\alpha + \inf V_\alpha, \sup U_\alpha + \sup V_\alpha],$$

$$(\lambda \cdot U)_\alpha = \begin{cases} [\lambda \cdot \inf U_\alpha, \lambda \cdot \sup U_\alpha] \text{ if } \lambda \geq 0 \\ [\lambda \cdot \sup U_\alpha, \lambda \cdot \inf U_\alpha] \text{ if } \lambda < 0 \end{cases}$$

$(\mathscr{F}_c(\mathbb{R}), +, \cdot)$ is not a linear but a semilinear space (since there is no inverse element for the sum).

To compare fuzzy values it will be useful to consider a distance between fuzzy numbers. Bertoluzza et al. [1] has introduced the D_W^φ *metric* which has been shown to be valuable and operational in this setting. Given $U, V \in \mathscr{F}_c(\mathbb{R})$,

$$D_W^{\varphi}(U,V)=\sqrt{\int_0^1\left[\int_0^1\left[\lambda\left(\sup U_\alpha-\sup V_\alpha\right)+(1-\lambda)\left(\inf U_\alpha-\inf V_\alpha\right)\right]^2 dW(\lambda)\right]d\varphi(\alpha)},$$

where W and φ are normalized weighted measures on $[0,1]$ (formalized as probability measures on $([0,1],\mathscr{B}_{[0,1]})$), W being associated with a non-degenerate distribution, and φ being associated with a strictly increasing distribution function on $[0,1]$. It should be pointed out that W and φ have no stochastic meaning. To consider W is equivalent to consider a measure weighting points 0, 1 and a certain $t_0(W)\in(0,1)$. Metric D_W^{φ} is an L_2-distance on the cone of the image of $\mathscr{F}_c(\mathbb{R})$ through the support function (see, for instance, [8]).

Fuzzy random variables (FRVs for short) in Puri and Ralescu's sense [10] represent a well-formalized model in the probabilistic setting. Given a probability space $(\Omega,\mathscr{A},P)$, a mapping $\mathscr{X}:\Omega\to\mathscr{F}_c(\mathbb{R})$ is said to be a *fuzzy random variable* associated with the space, if it is Borel-measurable w.r.t. $\mathscr{B}_{D_W^{\varphi}}$ (σ-field generated by the topology induced by D_W^{φ} on $\mathscr{F}_c(\mathbb{R})$). Borel-measurability guarantees that one can properly refer to concepts like statistical independence of FRVs, distribution induced by an FRV, etc.

As a measure for the 'central tendency', Puri and Ralescu [10] have introduced the concept of (fuzzy) mean value of an FRV. If $\mathscr{X}:\Omega\to\mathscr{F}_c(\mathbb{R})$ is an FRV associated with the probability space $(\Omega,\mathscr{A},P)$ and such that $\max\left\{|\inf\mathscr{X}_0|,|\sup\mathscr{X}_0|\right\}$ is integrable, the *fuzzy expected value* (or *fuzzy mean*) of $\mathscr{X}$ is the fuzzy number $\widetilde{\mu}=\widetilde{E}(\mathscr{X})\in\mathscr{F}_c(\mathbb{R})$ such that for all $\alpha\in[0,1]$

$$\widetilde{\mu}_\alpha=\text{Aumann integral of }\mathscr{X}_\alpha=\left[E(\inf\mathscr{X}_\alpha|P),E(\sup\mathscr{X}_\alpha|P)\right].$$

Given n random observations from $\mathscr{X}$ (say $\mathscr{X}_1,\ldots,\mathscr{X}_n$), the *fuzzy sample mean* given by

$$\overline{\mathscr{X}}_n=\frac{1}{n}\cdot[\mathscr{X}_1+\ldots+\mathscr{X}n]$$

can be considered, on one hand, as an 'unbiased fuzzy-valued estimator' of $\widetilde{\mu}$. On the other hand, $\overline{\mathscr{X}}_n$ can be used to state the statistics for one-sample two-sided tests about the mean of an FRV. Thus, to test the null hypothesis $H_0:\widetilde{\mu}=U\in\mathscr{F}_c(\mathbb{R})$ (or, equivalently, to test $H_0:D_W^{\varphi}\left(\widetilde{\mu},U\right)=0$, the use of central limit theorems or their bootstrapped approaches for generalized spaces-valued random elements allows us to consider techniques in this context (see Körner [7], Montenegro et al. [9], González-Rodríguez et al. [6]).

3 Fuzzy Representations of Random Variables: Characterizing Representations

Let $X:\Omega\to\mathbb{R}$ be a real-valued random variable (RV for short) associated with the probability space $(\Omega,\mathscr{A},P)$. A *γ-fuzzy representation of X* (see [5]) is the output of the composition of a 'fuzzifying' measurable mapping $\gamma:\mathbb{R}\to\mathscr{F}_c(\mathbb{R})$ with X, so that $\gamma\circ X:\Omega\to\mathscr{F}_c(\mathbb{R})$ is an FRV.

The interest of fuzzy representations in statistics lies in the fact that γ can be defined so that the (fuzzy) mean value of the γ-fuzzy representation of RV X, $\gamma\circ X$, can capture information on either some relevant parameters or features of the distribution of X, or

the type of distribution (discrete or continuous), or even on the whole distribution of X, leading to the so-called *characterizing fuzzy representations*.

In [2] it has been stated that for RVs taking on a small number of different values (up to 4) a triangular representation of variable values (that is, a mapping transforming each variable value into a triangular fuzzy number) characterizes the distribution of the variable. However, this conclusion fails when there are more than 4 different variable values. The last assertion has motivated the introduction of alternate transformations in which either by incorporating curvatures or angles into fuzzy values it is possible to capture the whole information on the distribution of the variable. In previous papers (see [3], [4], [5]) characterizing fuzzy representations have been considered.

In this paper we are going to make use of two generalized fuzzy representations which will be studied in connection with the binomial distribution. Let $\gamma : \mathbb{R} \to \mathscr{F}_c(\mathbb{R})$ be the mapping transforming each value $x \in \mathbb{R}$

- either into the fuzzy number $\gamma^{sh}(x)$ such that

$$\left(\gamma^{sh}(x)\right)_\alpha = \left[f_L(x) - (1-\alpha)^{1/h_L(x)}, f_R(x) + (1-\alpha)^{1/h_R(x)}\right]$$

  for all $\alpha \in [0,1]$ where $f_L : \mathbb{R} \to \mathbb{R}, f_R : \mathbb{R} \to \mathbb{R}$, $f_L, f_R \in L^1(\Omega, \mathscr{A}, P)$, $f_L(x) \leq f_R(x)$ for all $x \in \mathbb{R}$, and $h_L : \mathbb{R} \to (0, +\infty), h_R : \mathbb{R} \to (0, +\infty)$ are continuous and bijective,

- or into the fuzzy number

$$\gamma^{pol}(x) = \mathbf{1}_{\{x\}} + \operatorname{sig}(x - x_0)\gamma_f\left(\left|\frac{x - x_0}{a}\right|\right)$$

  where the functional $\gamma_f : \mathbb{R} \to \mathscr{F}_c(\mathbb{R})$ is given for $\alpha \in [0,1]$ by

$$[\gamma_f(t)]_\alpha = \begin{cases} \left[0, t^2 + t^2\left(\dfrac{1-f(t)}{f(t)}\right)\left(\dfrac{f(t)-\alpha}{f(t)}\right)\right] & \text{if } 0 \leq \alpha \leq f(t) \\ \left[0, t^2\left(\dfrac{1-\alpha}{1-f(t)}\right)\right] & \text{if } f(t) < \alpha \leq 1 \end{cases}$$

  and $f : [0, +\infty) \to [0,1]$ is an injective function.

Examples of fuzzy representations of the above-described type for an RV taking on values 0,1,2,3,4 can be found in Figure 1.

These two fuzzy representations characterize the whole distribution of the original variable in case it is binomial, since if $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are two binomial RVs associated with $(\Omega, \mathscr{A}, P)$, $\widetilde{E}\left(\gamma \circ X \,|\, P\right) = \widetilde{E}\left(\gamma \circ Y \,|\, P\right)$ if, and only if, X and Y are identically distributed.

The above characterization result could be applied to both, discrete and continuous variables, although we constrain in this paper to binomial ones. Furthermore, a very relevant implication from the characterizing property is that probabilistic and statistical results for the 'mean values' of generalized space-valued random elements can be applied to develop statistical inferences or probabilistic results on the distribution of an RV.

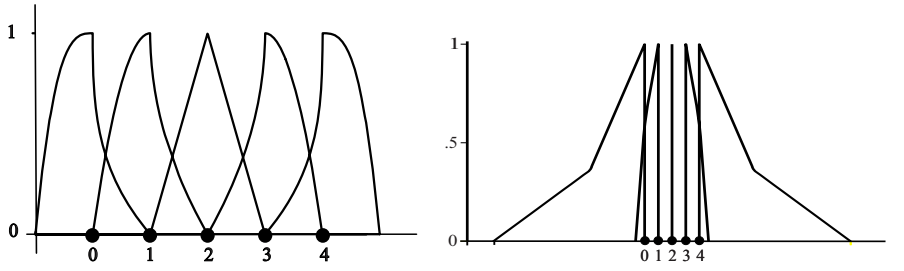


Fig. 1. Fuzzy representations γ^{sh} (on the left) and γ^{pol} (on the right) of an RV taking on values 0, 1, 2, 3 and 4

On the other hand, it should be remarked that there are more examples of fuzzy representations characterizing the distributions of random variables. The above recalled ones are just some examples we have considered in previous papers and showing suitable properties. Both representations, and especially the second one, are focussed on relevant parameters and features of the distribution of the variable. More precisely, the mean values of these fuzzy representations allow to easily identify and are very much influenced by the variable mean, variance, asymmetry, etc.

4 Goodness-of-Fit Tests for Binomial Distributions Based on Characterizing Fuzzy Representations: Method and Some Empirical Comparisons

In case we particularize the one-sample test about the mean value of an FRV (in Section 2) to a characterizing fuzzy representation of an RV, like those in Section 3, we clearly obtain a goodness-of-fit test. In case the hypothetical distribution is a specified binomial this test can be stated as follows:

Let $(\Omega, \mathscr{A}, P)$ be a probability space, and X be an RV associated with it. Consider the null hypothesis $H_0 : X \rightsquigarrow \mathrm{B}(n_0, p_0)$, which is equivalent to the null hypothesis $D_W^{\varphi}\left(\widetilde{E}(\gamma \circ X), \widetilde{E}(\gamma \circ \mathrm{B}(n_0, p_0))\right) = 0$ for a characterizing mapping $\gamma : \mathbb{R} \to \mathscr{F}_c(\mathbb{R})$ be like those in Section 3.

At the nominal significance level $\alpha \in [0,1]$, H_0 should be rejected whenever

$$T_n = \frac{\left[D_W^{\varphi}\left(\overline{(\gamma \circ X)}_n, \widetilde{E}(\gamma \circ \mathrm{B}(n_0, p_0))\right)\right]^2}{\left(\left[D_W^{\varphi}\left(\gamma \circ X, \overline{(\gamma \circ X)}_n\right)\right]^2\right)_n} > z_\alpha,$$

where $z_\alpha = 100(1-\alpha)$ fractile of the distribution of T_n under H_0 and this critical value can be obtained by means of Monte Carlo method.

The application of the above testing method can be made in an easy way, and it could be extended to any specified hypothetical distribution (not just binomial ones). Then, a key discussion to be made at this stage is that of how it performs in comparison with

Table 1. Testing $H_0 : X \rightsquigarrow \mathrm{B}(4, p = .5)$ at the significance level $\alpha = .05$, $\varphi = \mathscr{U}(0,1)$

true value of p	% reject. χ^2	% reject. KS	% reject. LR	% reject. CVM	% reject. γ^{sh}	% reject. γ^{pol}
0.5	4.918	3.049	4.64	4.987	4.729	4.64
0.6	16.007	14.997	13.83	23.949	23.333	23.011
0.7	54.303	56.301	51.15	71.554	72.547	72.244
0.4	15.931	15.173	13.913	19.401	23.346	22.868
0.3	54.134	56.181	51.128	64.367	72.64	72.241

well-known 'general' goodness-of-fit tests like chi-square (χ^2), Kolmogorov-Smirnov (KS), likelihood ratio (LR), Cramér-Von Mises (CVM), the test above when $\gamma = \gamma^{sh}$, and when $\gamma = \gamma^{pol}$.

Indeed, as for the traditional techniques, there is a need for developing simulation studies, since theoretical conclusions are generally unfeasible. A deep discussion on this point will be a very complex task that will be tackled in the future. In this paper, we present an introductory discussion for a very particular case: the hypothetical distribution being a $\mathrm{B}(4, p_0)$ and considering the transformations in Figure 1. For this purpose, we have examined several situations, that is, different values of p_0. For each of these situations we have first simulated by means of the Monte Carlo method 100,000 samples of size 10 from the $B(4, p_0)$ to approximate the critical value z_α. Later, to analyze the empirical achievement of the nominal significance level (.05) and the power of different tests, we have simulated by means of the Monte Carlo method 100,000 samples of size 10 from the $B(4, p)$ and compute the percentage of rejections of H_0. Each case has been studied for $W =$ Lebesgue measure, $\varphi = \mathscr{U}(0,1)$ and $\varphi = \beta(1,4)$ in D_W^φ.

Tables 1-4 show some of the results obtained in this empirical analysis. In all these tables the first row of numbers correspond to the accomplishment of the nominal significance level (5%), whereas the other rows are related to the power of the tests for different 'deviations' from the hypothetical distribution.

Conclusions we state below are drawn on the basis of a few simulation studies gathered in Tables 1-4. Nevertheless, we have developed some more simulations for other hypothetical values of n_0 and p_0, and other deviations from these values, although a deep discussion would require a much more exhaustive analysis. Anyway, we can state that introductory studies show a quite good behavior of the goodness-of-fit tests based

Table 2. Testing $H_0 : X \rightsquigarrow \mathrm{B}(4, p = .5)$ at the significance level $\alpha = .05$, $\varphi = \beta(1,4)$

true value of p	% reject. χ^2	% reject. KS	% reject. LR	% reject. CVM	% reject. γ^{sh}	% reject. γ^{pol}
.5	4.925	3.093	4.946	4.95	4.701	4.627
.6	15.913	15.078	14.271	23.729	23.093	22.788
.7	54.234	56.201	51.433	71.437	72.372	71.948
.4	16.019	14.995	14.234	19.395	23.382	23.07
.3	53.901	56.021	51.165	63.946	72.097	71.791

Table 3. Testing $H_0 : X \rightsquigarrow \mathrm{B}(4, p = .75)$ at the significance level $\alpha = .05$, $\varphi = \mathscr{U}(0,1)$

true value of p	% reject. χ^2	% reject. KS	% reject. LR	% reject. CVM	% reject. γ^{sh}	% reject. γ^{pol}
0.75	4.984	3.859	4.944	4.809	4.989	4.969
0.85	2.603	20.963	14.754	28.872	26.312	28.306
0.95	42.371	90.267	86.619	95.537	95.112	95.514
0.65	24.461	24.25	25.119	24.574	34.246	32.732
0.55	63.926	67.469	68.717	67.229	81.077	80.073

Table 4. Testing $H_0 : X \rightsquigarrow \mathrm{B}(4, p = .25)$ at the significance level $\alpha = .05$, $\varphi = \beta(1,4)$

true value of p	% reject. χ^2	% reject. KS	% reject. LR	% reject. CVM	% reject. γ^{sh}	% reject. γ^{pol}
0.25	4.991	4.067	4.946	4.853	4.687	4.97
0.35	23.911	24.166	24.124	28.493	32.571	32.466
0.45	62.037	67.501	67.559	74.025	79.913	79.801
0.15	2.027	21.018	14.788	26.986	26.26	28.256
0.05	42.04	90.304	86.555	95.302	95.165	95.555

on characterizing fuzzy representations in comparison with the best known ones. Although there is not a uniformly most powerful test, in most of the examined situations and for most of the deviations from the hypothesis, either the test based on the γ^{sh} or the one based on the γ^{pol} is the most powerful one.

In addition to the need for a deep comparative analysis of the goodness-of-fit techniques, a challenging open problem is the one related to the choice of the weighting measure φ as well as the γ^{sh} or γ^{pol} functions. That is, a sensitivity analysis should be also carried out.

Acknowledgement. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2006–07501. This financial support is gratefully acknowledged.

References

1. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware Soft. Comput. 2, 71–84 (1995)
2. Colubi, A., González-Rodríguez, G.: Triangular fuzzification of random variables and power of distribution tests: Empirical discussion. Comp. Statist. Data Anal. 51, 4742–4750 (2007)
3. Colubi, A., González-Rodríguez, G., Gil, M.A.: A new characterization of discrete random variables by means of fuzzy sets: Graphical features. In: Abstracts of the 56th Session of the International Statistical Institute (ISI 2007, Lisbon, Portugal), p. 165 (2007)

4. Colubi, A., González-Rodríguez, G., Lubiano, M.A., Montenegro, M.: Exploratory analysis of random variables based on fuzzification. In: Lawry, J., Miranda, E., Bugarin, A., Li, S., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Integrated Uncertainty Modelling. Advances in Soft Computing, vol. 37, pp. 95–102. Springer, Berlin (2006)
5. González-Rodríguez, G., Colubi, A., Gil, M.A.: A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. Comp. Statist. Data Anal. 51, 163–176 (2006)
6. González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A.: Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst. 157, 2608–2613 (2006)
7. Körner, R.: An asymptotic α-test for the expectation of random fuzzy variables. J. Statist. Plann. Inference 83, 331–346 (2000)
8. Körner, R., Näther, W.: On the variance of random fuzzy variables. In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data. Studies in Fuzziness and Soft Computing, vol. 87, pp. 22–39. Physica-Verlag, Heildelberg (2002)
9. Montenegro, M., Colubi, A., Casals, M.R., Gil, M.A.: Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. Metrika 59, 31–49 (2004)
10. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)
11. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part 1. Inform. Sci. 8, 199–249; Part 2. Inform. Sci. 8, 301–353; Part 3. Inform. Sci. 9, 43–80 (1975)

Part IV

Mathematical Aspects

Invited Session:
Fuzzy Set-Valued Analysis

A Generalization of Hukuhara Difference

Luciano Stefanini

Department of Economics and Quantitative Methods (DEMQ) and Faculty of conomics, University of Urbino, Urbino, Italy

Abstract. We propose a generalization of the Hukuhara difference. First, the case of compact convex sets is examined; then, the results are applied to generalize the Hukuhara difference of fuzzy numbers, using their compact and convex level-cuts. Finally, a similar approach is seggested to attempt a generalization of division for real intervals and fuzzy numbers.

Keywords: Hukuhara difference, Interval and fuzzy arithmetic, Fuzzy numbers, Invertible fuzzy operations, Interval and fuzzy calculus.

1 General Setting

We consider a metric vector space $\mathbb{X}$ with the induced topology and in particular the space $\mathbb{X}=\mathbb{R}^n$, $n\geq 1$, of real vectors equipped with standard addition and scalar multiplication operations. Following Diamond and Kloeden (see [3]), denote by $\mathscr{K}(\mathbb{X})$ and $\mathscr{K}_C(\mathbb{X})$ the spaces of nonempty compact and compact convex sets of $\mathbb{X}$. Given two subsets $A,B\subseteq\mathbb{X}$ and $k\in\mathbb{R}$, Minkowski addition and scalar multiplication are defined by $A+B=\{a+b|a\in A,b\in B\}$ and $kA=\{ka|a\in A\}$ and it is well known that addition is associative and commutative and with neutral element $\{0\}$. If $k=-1$, scalar multiplication gives the opposite $-A=(-1)A=\{-a|a\in A\}$ but, in general, $A+(-A)\neq\{0\}$, i.e. the opposite of A is not the inverse of A in Minkowski addition (unless $A=\{a\}$ is a singleton). Minkowski difference is $A-B=A+(-1)B=\{a-b|a\in A,b\in B\}$. A first implication of this fact is that, in general, even if it true that $(A+C=B+C)\iff A=B$, addition/subtraction simplification is not valid, i.e. $(A+B)-B\neq A$.

To partially overcome this situation, Hukuhara [4] introduced the following H-difference $A\ominus B=C\iff A=B+C$ and an important property of $\ominus$ is that $A\ominus A=\{0\}$, $\forall A\in\mathbb{R}^n$ and $(A+B)\ominus B=A$, $\forall A,B\in\mathbb{R}^n$; H-difference is unique, but a necessary condition for $A\ominus B$ to exist is that A contains a translate $\{c\}+B$ of B. In general, $A-B\neq A\ominus B$. From an algebraic point of view, the difference of two sets A and B may be interpreted both in terms of addition as in $\ominus$ or in terms of negative addition, i.e. $A\boxminus B=C\iff B=A+(-1)C$ where $(-1)C$ is the opposite set of C. Operations $\ominus$ and $\boxminus$ are compatible each other and this suggests a generalization of Hukuhara difference:

Definition 1. *Let $A,B\in\mathscr{K}(\mathbb{X})$; we define the generalized difference of A and B as the set $C\in\mathscr{K}(\mathbb{X})$ such that*

$$A\ominus_g B=C\iff\begin{cases}(i)\ \ A=B+C\\ or\ (ii)\ B=A+(-1)C\end{cases}. \tag{1}$$

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 203–210, 2008.
springerlink.com

Proposition 1. *(Unicity of $A \ominus_g B$)*
If $C = A \ominus_g B$ exists, it is unique and if also $A \ominus B$ exists then $A \ominus_g B = A \ominus B$.

Proof. See [7]. □

The generalized Hukuhara difference $A \ominus_g B$ will be called the gH-difference of A and B.

Proposition 2. *If $A \ominus_g B$ exists, it has the following properties:*

1. $A \ominus_g A = \{0\}$;
2. $(A + B) \ominus_g B = A$;
3. *If $A \ominus_g B$ exists then also $(-B) \ominus_g (-A)$ does and $-(A \ominus_g B) = (-B) \ominus_g (-A)$;*
4. $(A - B) + B = C \iff A - B = C \ominus_g B$;
5. *In general, $B - A = A - B$ does not imply $A = B$; but $(A \ominus_g B) = (B \ominus_g A) = C$ if and only if $C = \{0\}$ and $A = B$;*
6. *If $B \ominus_g A$ exists then either $A + (B \ominus_g A) = B$ or $B - (B \ominus_g A) = A$ and both equalities hold if and only if $B \ominus_g A$ is a singleton set.*

Proof. See [7]. □

If $\mathbb{X} = \mathbb{R}^n$, $n \geq 1$ is the real $n-$dimensional vector space with internal product $\langle x, y \rangle$ and corresponding norm $||x|| = \sqrt{\langle x, x \rangle}$, we denote by $\mathscr{K}^n$ and $\mathscr{K}_C^n$ the spaces of (nonempty) compact and compact convex sets of $\mathbb{R}^n$, respectively. If $A \subseteq \mathbb{R}^n$ and $\mathscr{S}^{n-1} = \{u | u \in \mathbb{R}^n, ||u|| = 1\}$ is the unit sphere, the support function associated to A is

$$s_A : \mathbb{R}^n \longrightarrow \mathbb{R} \text{ defined by } s_A(u) = \sup\{\langle u, a \rangle | a \in A\}, u \in \mathbb{R}^n.$$

If $A \neq \emptyset$ is compact, then $s_A(u) \in \mathbb{R}, \forall u \in \mathscr{S}^{n-1}$. For properties of the support functions see e.g. [3] or [5].

We can express the generalized Hukuhara difference (gH-difference) of compact convex sets $A, B \in \mathscr{K}_C^n$ by the use of the support functions. Consider $A, B, C \in \mathscr{K}_C^n$ with $C = A \ominus_g B$ as defined in (1); let s_A, s_B, s_C and $s_{(-1)C}$ be the support functions of A, B, C, and $(-1)C$ respectively. In case (i) we have $s_A = s_B + s_C$ and in case (ii) we have $s_B = s_A + s_{(-1)C}$. So, $\forall u \in \mathscr{S}^{n-1}$

$$s_C(u) = \begin{cases} s_A(u) - s_B(u) \\ s_B(-u) - s_A(-u) \end{cases} = \begin{cases} s_A(u) - s_B(u) & \text{in case } (i) \\ s_{(-1)B}(u) - s_{(-1)A}(u) & \text{in case } (ii) \end{cases}. \tag{2}$$

Now, s_C in (2) is a correct support function if it is continuous, positively homogeneous and subadditive and this requires that, in the corresponding cases (i) and (ii), $s_A - s_B$ and/or $s_{-B} - s_{-A}$ be support functions, assuming that s_A and s_B are.

Consider $s_1 = s_A - s_B$ and $s_2 = s_B - s_A$. Continuity of s_1 and s_2 is obvious. To see their positive homogeneity let $t \geq 0$; we have $s_1(tu) = s_A(tu) - s_B(tu) = ts_A(u) - ts_B(u) = ts_1(u)$ and similarly for s_2. But s_1 and/or s_2 may fail to be subadditive and the following four cases, related to the definition of gH-difference, are possible (for a proof see [7]).

Proposition 3. *Let s_A and s_B be the support functions of $A, B \in \mathscr{K}_C^n$ and consider $s_1 = s_A - s_B$, $s_2 = s_B - s_A$; the following four cases apply:*

1. *If s_1 and s_2 are both subadditive, then $A \ominus_g B$ exists; (i) and (ii) are satisfied simultaneously and $A \ominus_g B = \{c\}$;*
2. *If s_1 is subadditive and s_2 is not, then $C = A \ominus_g B$ exists, (i) is satisfied and $s_C = s_A - s_B$;*
3. *If s_1 is not subadditive and s_2 is, then $C = A \ominus_g B$ exists, (ii) is satisfied and $s_C = s_{-B} - s_{-A}$;*
4. *If s_1 and s_2 are both not subadditive, then $A \ominus_g B$ does not exist.*

Proposition 4. *If $C = A \ominus_g B$ exists, then $||C|| = \rho_2(A,B)$ and the Steiner points satisfy $\sigma_C = \sigma_A - \sigma_B$.*

2 The Case of Compact Intervals in $\mathbb{R}^n$

In this section we consider the gH-difference of compact intervals in $\mathbb{R}^n$. If $n = 1$, i.e. for unidimensional compact intervals, the gH-difference always exists. In fact, let $A = [a^-, a^+]$ and $B = [b^-, b^+]$ be two intervals; the gH-difference is

$$[a^-,a^+] \ominus_g [b^-,b^+] = [c^-,c^+] \iff \begin{cases} (i) \ \{a^- = b^- + c^-,\ a^+ = b^+ + c^+ \\ \text{or } (ii) \ \{b^- = a^- - c^+,\ b^+ = a^+ - c^- \end{cases}$$

so that $[a^-,a^+] \ominus_g [b^-,b^+] = [c^-,c^+]$ is always defined by $c^- = \min\{a^- - b^-, a^+ - b^+\}$, $c^+ = \max\{a^- - b^-, a^+ - b^+\}$. Conditions (i) and (ii) are satisfied simultaneously if and only if the two intervals have the same length and $c^- = c^+$. Also, the result is $\{0\}$ if and only if $a^- = b^-$ and $a^+ = b^+$.

Two simple examples on real compact intervals illustrate the generalization (from [3], p. 8); $[-1,1] \ominus [-1,0] = [0,1]$ as in fact *(i)* is $[-1,0] + [0,1] = [-1,1]$ but $[0,0] \ominus_g [0,1] = [-1,0]$ and $[0,1] \ominus_g [-\frac{1}{2}, 1] = [0, \frac{1}{2}]$ satisfy *(ii)*.

Let now $A = \times_{i=1}^n A_i$ and $B = \times_{i=1}^n B_i$ where $A_i = [a_i^-, a_i^+]$, $B_i = [b_i^-, b_i^+]$ are real compact intervals ($\times_{i=1}^n$ denotes the cartesian product).

In general, considering $D = \times_{i=1}^n (A_i \ominus_g B_i)$, we may have $A \ominus_g B \neq D$ e.g. $A \ominus_g B$ may not exist as for the example $A_1 = [3,6]$, $A_2 = [2,6]$, $B_1 = [5,10]$, $B_2 = [7,9]$ for which $(A_1 \ominus_g B_1) = [-4,-2]$, $(A_2 \ominus_g B_2) = [-5,-3]$, $D = [-4,-2] \times [-5,-3]$ and $B + D = [1,8] \times [2,6] \neq A$, $A + (-1)D = [5,10] \times [5,11] \neq B$.

But if $A \ominus_g B$ exists, then equality will hold. In fact, consider the support function of A (and similarly for B), defined by $s_A(u) = \max_x \{\langle u,x\rangle \,|\, a_i^- \le x_i \le a_i^+\}$, $u \in \mathscr{S}^{n-1}$; it can be obtained simply by $s_A(u) = \sum_{u_i>0} u_i a_i^+ + \sum_{u_i<0} u_i a_i^-$ as the box-constrained maxima of the linear objective functions $\langle u,x\rangle$ above are attained at vertices $\widehat{x}(u) = (\widehat{x}_1(u), ..., \widehat{x}_i(u), ..., \widehat{x}_n(u))$ of A, i.e. $\widehat{x}_i(u) \in \{a_i^-, a_i^+\}$, $i = 1,2,...,n$. Then, being $s_{-A}(u) = s_A(-u) = -\sum_{u_i<0} u_i a_i^+ - \sum_{u_i>0} u_i a_i^-$, one obtains $s_{-B}(u) - s_{-A}(u) = \sum_{u_i>0} u_i (a_i^- - b_i^-) + \sum_{u_i<0} u_i (a_i^+ - b_i^+)$.

From the relations above, we deduce that the gH-difference $A \ominus_g B$ exists if and only if one of the two conditions are satisfied:

$$A \ominus_g B = C \iff \begin{cases} (i) \begin{cases} C = \times_{i=1}^{n} [a_i^- - b_i^-, a_i^+ - b_i^+] \\ \text{provided that } a_i^- - b_i^- \le a_i^+ - b_i^+, \forall i \end{cases} \\ \text{or } (ii) \begin{cases} C = \times_{i=1}^{n} [a_i^+ - b_i^+, a_i^- - b_i^-] \\ \text{provided that } a_i^- - b_i^- \ge a_i^+ - b_i^+, \forall i \end{cases} \end{cases}.$$

Examples are given in [7]. We end this section with a comment on the simple interval equation

$$A + X = B \tag{3}$$

where $A = [a^-, a^+]$, $B = [b^-, b^+]$ are given intervals and $X = [x^-, x^+]$ is an interval to be determined satisfying (3). We have seen that, for unidimensional intervals, the gH-difference always exists. Denote by $l(A) = a^+ - a^-$ the length of interval A. It is well known from classical interval arithmetic that an interval X satisfying (3) exists only if $l(B) \ge l(A)$ (in Minkowski arithmetic we have $l(A+X) \ge \max\{l(A), l(X)\}$); in fact, no X exists with $x^- \le x^+$ if $l(B) < l(A)$ and we cannot solve (3) unless we interpret it as $B - X = A$. If we do so, we get

$$\text{case } l(B) \le l(A): \begin{cases} a^- + x^- = b^- \\ a^+ + x^+ = b^+ \end{cases} \text{ i.e. } \begin{array}{l} x^- = b^- - a^- \\ x^- = b^- - a^- \end{array}$$

$$\text{case } l(B) \ge l(A): \begin{cases} b^- - x^+ = a^- \\ b^+ - x^- = a^+ \end{cases} \text{ i.e. } \begin{array}{l} x^- = b^+ - a^+ \\ x^+ = b^- - a^- \end{array}.$$

We then obtain that $X = B \ominus_g A$ is the unique solution to (3) and it always exists, i.e.

Proposition 5. *Let $A, B \in \mathscr{K}_C(\mathbb{R})$; the gH-difference $X = B \ominus_g A$ always exists and either $A + (B \ominus_g A) = B$ or $B - (B \ominus_g A) = A$.*

From Property 6 of Proposition 2, a similar result is true for equation $A + X = B$ with $A, B \in \mathscr{K}_C(\mathbb{R}^n)$ but for $n > 1$ the gH-difference may non exist.

3 gH-Difference of Fuzzy Numbers

A general *fuzzy set* over a given set (or space) $\mathbb{X}$ of elements (the universe) is usually defined by its membership function $\mu : \mathbb{X} \longrightarrow \mathbb{T} \subseteq [0,1]$ and a fuzzy (sub)set u of $\mathbb{X}$ is uniquely characterized by the pairs $(x, \mu_u(x))$ for each $x \in \mathbb{X}$; the value $\mu_u(x) \in [0,1]$ is the membership grade of x to the fuzzy set u. We will consider particular fuzzy sets, called *fuzzy numbers*, defined over $\mathbb{X} = \mathbb{R}$ having a particular form of the membership function. Let μ_u be the membership function of a fuzzy set u over $\mathbb{X}$. The support of u is the (crisp) subset of points of $\mathbb{X}$ at which the membership grade $\mu_u(x)$ is positive: $supp(u) = \{x | x \in \mathbb{X},\ \mu_u(x) > 0\}$. For $\alpha \in]0,1]$, the $\alpha-$level cut of u (or simply the $\alpha - cut$) is defined by $[u]_\alpha = \{x | x \in \mathbb{X},\ \mu_u(x) \ge \alpha\}$ and for $\alpha = 0$ (or $\alpha \to +0$) by the closure of the support $[u]_0 = cl\{x | x \in \mathbb{X},\ \mu_u(x) > 0\}$.

A well-known property of the $level - cuts$ is $[u]_\alpha \subseteq [u]_\beta$ for $\alpha > \beta$ (i.e. they are nested).

A particular class of fuzzy sets u is when the support is a convex set and the membership function is quasi-concave i.e. $\mu_u((1-t)x' + tx'') \ge \min\{\mu_u(x'), \mu_u(x'')\}$ for every $x', x'' \in supp(u)$ and $t \in [0,1]$. Equivalently, μ_u is quasi-concave if the level sets $[u]_\alpha$ are

convex sets for all $\alpha \in [0,1]$. A third property of the fuzzy numbers is that the level-cuts $[u]_\alpha$ are closed sets for all $\alpha \in [0,1]$.

By using these properties, the space $\mathscr{F}$ of (real unidimensional) fuzzy numbers is structured by an addition and a scalar multiplication, defined either by the level sets or, equivalently, by the Zadeh extension principle. Let $u, v \in \mathscr{F}$ have membership functions μ_u, μ_v and $\alpha - cuts$ $[u]_\alpha$, $[v]_\alpha$, $\alpha \in [0,1]$ respectively. The addition $u+v \in \mathscr{F}$ and the scalar multiplication $ku \in \mathscr{F}$ have level cuts

$$[u+v]_\alpha = [u]_\alpha + [v]_\alpha \text{ and } [ku]_\alpha = k[u]_\alpha. \tag{4}$$

In the fuzzy or in the interval arithmetic contexts, equation $u = v+w$ is not equivalent to $w = u-v = u+(-1)v$ or to $v = u-w = u+(-1)w$ and this has motivated the introduction of the following Hukuhara difference ([3, 5]). The generalized Hukuhara difference is (implicitly) used by Bede and Gal (see [1]) in their definition of generalized differentiability of a fuzzy-valued function.

Definition 2. *Given $u, v \in \mathscr{F}$, the H-difference is defined by $u \ominus v = w \Longleftrightarrow u = v+w$; if $u \ominus v$ exists, it is unique and its $\alpha - cuts$ are $[u \ominus v]_\alpha = [u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+]$. Clearly, $u \ominus u = \{0\}$.*

The Hukuhara difference is also motivated by the problem of inverting the addition: if x, y are crisp numbers then $(x+y)-y = x$ but this is not true if x, y are fuzzy. It is possible to see that (see [2]), if u and v are fuzzy numbers (and not in general fuzzy sets), then $(u+v) \ominus v = u$ i.e. the H-difference inverts the addition of fuzzy numbers.

Definition 3. *Given $u, v \in \mathscr{F}$, the gH-difference is the fuzzy number w, if it exists, such that*

$$u \ominus_g v = w \Longleftrightarrow \begin{cases} (i) & u = v+w \\ or\ (ii) & v = u+(-1)w \end{cases}. \tag{5}$$

If $u \ominus_g v$ exists, its $\alpha - cuts$ are given by $[u \ominus_g v]_\alpha = [\min\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\}, \max\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\}]$ and $u \ominus v = u \ominus_g v$ if $u \ominus v$ exists. If (i) and (ii) are satisfied simultaneously, then w is a crisp number. Also, $u \ominus_g u = u \ominus u = \{0\}$.

A definition of $w = u \ominus_g v$ for multidimensional fuzzy numbers can be obtained in terms of support functions in a way similar to (2)

$$s_w(p;\alpha) = \begin{cases} s_u(p;\alpha) - s_v(p;\alpha) & \text{in case } (i) \\ s_{(-1)v}(p;\alpha) - s_{(-1)u}(p;\alpha) & \text{in case } (ii) \end{cases}, \ \alpha \in [0,1] \tag{6}$$

where, for a fuzzy number u, the support functions are considered for each $\alpha - cut$ and defined to characterize the (compact) $\alpha - cuts$ $[u]_\alpha$:

$$\begin{aligned} s_u &: \mathbb{R}^n \times [0,1] \longrightarrow \mathbb{R} \text{ defined by} \\ s_u(p;\alpha) &= \sup\{\langle p, x\rangle \,|\, x \in [u]_\alpha\} \text{ for each } p \in \mathbb{R}^n, \ \alpha \in [0,1]. \end{aligned}$$

In the unidimensional fuzzy numbers, the conditions for the definition of $w = u \ominus_g v$ are

$$[w]_\alpha = [w_\alpha^-, w_\alpha^+] = [u]_\alpha \ominus_g [v]_\alpha : \begin{cases} w_\alpha^- = \min\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\} \\ w_\alpha^+ = \max\{u_\alpha^- - v_\alpha^-, u_\alpha^+ - v_\alpha^+\} \end{cases}. \tag{7}$$

provided that w_α^- is nondecreasing, w_α^+ is nonincreasing and $w_\alpha^- \le w_\alpha^+$. If $u \ominus_g v$ is a proper fuzzy number, it has the same properties illustrated in Section 1 for intervals.

Proposition 6. *If $u \ominus_g v$ exists, it is unique and has the following properties:*

1. $u \ominus_g u = 0$;
2. $(u+v) \ominus_g v = u$;
3. *If $u \ominus_g v$ exists then also $(-v) \ominus_g (-u)$ does and $\{0\} \ominus_g (u \ominus_g v) = (-v) \ominus_g (-u)$;*
4. $(u-v)+v = w \iff u - v = w \ominus_g v$;
5. $(u \ominus_g v) = (v \ominus_g u) = w$ *if and only if* ($w = \{0\}$ *and* $u = v$);
6. *If $v \ominus_g u$ exists then either $u + (v \ominus_g u) = u$ or $v - (v \ominus_g u) = u$ and if both equalities hold then $v \ominus_g u$ is a crisp set.*

If the gH-differences $[u]_\alpha \ominus_g [v]_\alpha$ do not define a proper fuzzy number, we can use the nested property and obtain a proper fuzzy number by

$$[u \widetilde{\ominus}_g v]_\alpha := \bigcup_{\beta \geq \alpha} ([u]_\beta \ominus_g [v]_\beta); \tag{8}$$

As each gH-difference $[u]_\beta \ominus_g [v]_\beta$ exists for $\beta \in [0,1]$ and (8) defines a proper fuzzy number, it follows that $u \widetilde{\ominus}_g v$ can be considered as a generalization of Hukuhara difference for fuzzy numbers, existing for any u, v. A second possibility for a gH-difference of fuzzy numbers may be obtained following a suggestion by Diamond and Kloeden ([3]) and defining $z = u \widetilde{\ominus}_g v$ to be the fuzzy number whose $\alpha - cuts$ are as near as possible to the gH-differences $[u]_\alpha \ominus_g [v]_\alpha$, for example by minimizing the functional ($\omega_\alpha \geq 0$ and $\gamma_\alpha \geq 0$ are weighting functions)

$$G(z|u,v) = \int_0^1 (\omega_\alpha \left[z_\alpha^- - (u \ominus_g v)_\alpha^-\right]^2 + \gamma_\alpha \left[z_\alpha^+ - (u \ominus_g v)_\alpha^+\right]^2) d\alpha$$

such that $z_\alpha^- \uparrow, z_\alpha^+ \downarrow, z_\alpha^- \leq z_\alpha^+ \; \forall \alpha \in [0,1]$.

A discretized version of $G(z|u,v)$ can be obtained by choosing a partition $0 = \alpha_0 < \alpha_1 < ... < \alpha_N = 1$ of $[0,1]$ and defining the discretized $G(z|u,v)$ as

$$G_N(z|u,v) = \sum_{i=0}^{N} \omega_i \left[z_i^- - (u \ominus_g v)_i^-\right]^2 + \gamma_i \left[z_i^+ - (u \ominus_g v)_i^+\right]^2;$$

we minimize $G_N(z|u,v)$ with the given data $(u \ominus_g v)_i^- = \min\{u_{\alpha_i}^- - v_{\alpha_i}^-, u_{\alpha_i}^+ - v_{\alpha_i}^+\}$ and $(u \ominus_g v)_i^+ = \max\{u_{\alpha_i}^- - v_{\alpha_i}^-, u_{\alpha_i}^+ - v_{\alpha_i}^+\}$, subject to the constraints $z_0^- \leq z_1^- \leq ... \leq z_N^- \leq z_N^+ \leq z_{N-1}^+ \leq \ldots \leq z_0^+$. We obtain a linearly constrained least squares problem $\min_{z \in \mathbb{R}^{2N+2}} (z-w)^T D^2 (z-w)$ s.t. $Ez \geq 0$ where $D = diag\{\sqrt{\omega_0}, \ldots, \sqrt{\omega_N}, \sqrt{\gamma_N}, \ldots, \sqrt{\gamma_0}\}$, $z = (z_0^-, z_1^-, \ldots, z_N^-, z_N^+, z_{N-1}^+, \ldots, z_0^+)$, $w_i^- = (u \ominus_g v)_i^-$, $w_i^+ = (u \ominus_g v)_i^+$, $w = (w_0^-, w_1^-, \ldots, w_N^-, w_N^+, w_{N-1}^+, \ldots, w_0^+)$, and E is the $(N, N+1)$ matrix

$$E = \begin{bmatrix} -1 & 1 & 0 & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \ldots & -1 & 1 \end{bmatrix}$$

which can be solved by standard efficient procedures (see the classical book [6, Chap. 23]). If, at solution z^*, we have $z^* = w$, then we obtain the gH-difference as defined in (5).

4 Generalized Division

An idea silmilar to the gH-difference can be used to introduce a division of real intervals and fuzzy numbers. We consider here only the case of real compact intervals $A=[a^-,a^+]$ and $B=[b^-,b^+]$ with $b^->0$ or $b^+<0$ (i.e. $0\notin B$).

The interval $C=[c^-,c^+]$ defining the multiplication $C=AB$ is given by

$$c^-=\min\{a^-b^-,a^-b^+,a^+b^-,a^+b^+\}\ ,\ c^+=\max\{a^-b^-,a^-b^+,a^+b^-,a^+b^+\}$$

and the multiplicative "inverse" (it is not the inverse in the algebraic sense) of an interval B is defined by $B^{-1}=[\frac{1}{b^+},\frac{1}{b^-}]$; we define the generalized division (g-division) $\div_g$ as follows:

$$A\div_g B=C \iff (i)\ A=BC \text{ or } (ii)\ B=AC^{-1}\,.$$

If both cases *(i)* and *(ii)* are valid, we have $CC^{-1}=C^{-1}C=\{1\}$, i.e. $C=\{\widehat{c}\},C^{-1}=\{\frac{1}{\widehat{c}}\}$ with $\widehat{c}\neq 0$. It is easy to see that $A\div_g B$ always exists and is unique for given $A=[a^-,a^+]$ and $B=[b^-,b^+]$ with $0\notin B$. It is easy to see that it can be obtained by the following rules:

Case 1. If $(a^-\le a^+<0$ and $b^-\le b^+<0)$ or $(0<a^-\le a^+$ and $0<b^-\le b^+)$ then $c^-=\min\{\frac{a^-}{b^-},\frac{a^+}{b^+}\}\ge 0,\ c^+=\max\{\frac{a^-}{b^-},\frac{a^+}{b^+}\}\ge 0$;

Case 2. If $(a^-\le a^+<0$ and $0<b^-\le b^+)$ or $(0<a^-\le a^+$ and $b^-\le b^+<0)$ then $c^-=\min\{\frac{a^-}{b^+},\frac{a^+}{b^-}\}\le 0,\ c^+=\max\{\frac{a^-}{b^+},\frac{a^+}{b^-}\}\le 0$;

Case 3. If $(a^-\le 0, a^+\ge 0$ and $b^-\le b^+<0)$ then $c^-=\frac{a^-}{b^-}\le 0,\ c^+=\frac{a^+}{b^-}\ge 0$;

Case 4. If $(a^-\le 0, a^+\ge 0$ and $0<b^-\le b^+)$ then $c^-=\frac{a^-}{b^+}\le 0,\ c^+=\frac{a^+}{b^+}\ge 0$.

Remark 1. If $0\in]b^-,b^+[$ the g-division is undefined; for intervals $B=[0,b^+]$ or $B=[b^-,0]$ the division is possible but obtaining unbounded results C of the form $C=]-\infty,c^+]$ or $C=[c^-,+\infty[$: we work with $B=[\varepsilon,b^+]$ or $B=[b^-,\varepsilon]$ and we obtain the result by the limit for $\varepsilon\longrightarrow 0^+$. Example: for $[-2,-1]\div_g[0,3]$ we consider $[-2,-1]\div_g[\varepsilon,3]=[c^-_\varepsilon,c^+_\varepsilon]$ with (Case 2.) $c^-_\varepsilon=\min\{\frac{-2}{3},\frac{-1}{\varepsilon}\}$ and $c^+_\varepsilon=\max\{\frac{-2}{\varepsilon},\frac{-1}{3}\}$ and obtain the result $C=[-\infty,-\frac{1}{3}]$ at the limit $\varepsilon\longrightarrow 0^+$.

Proposition 7. *For any $A=[a^-,a^+]$ and $B=[b^-,b^+]$ with $0\notin B$, we have (here 1 is the same as $\{1\}$):*

1. $B\div_g B=1$, $B\div_g B^{-1}=\{b^-b^+\}$ $(=\{\widehat{b}^2\}$ *if* $b^-=b^+=\widehat{b})$;
2. $(AB)\div_g B=A$;
3. $1\div_g B=B^{-1}$ *and* $1\div_g B^{-1}=B$.

In the case of fuzzy numbers $u,v\in\mathscr{F}$ having membership functions μ_u, μ_v and $\alpha-cuts$ $[u]_\alpha=[u^-_\alpha,u^+_\alpha]$, $[v]_\alpha=[v^-_\alpha,v^+_\alpha]$, $0\notin[v]_\alpha\ \forall\alpha\in[0,1]$, the *g-division* $\div_g$ can be defined as the operation that calculates the fuzzy number $w=u\div_g v\in\mathscr{F}$ having level cuts $[w]_\alpha=[w^-_\alpha,w^+_\alpha]$ (here $[w]^{-1}_\alpha=[\frac{1}{w^+_\alpha},\frac{1}{w^-_\alpha}]$):

$$[u]_\alpha\div_g[v]_\alpha=[w]_\alpha \iff \begin{cases} (i)\ [u]_\alpha=[v]_\alpha[w]_\alpha \\ \text{or } (ii)\ [v]_\alpha=[u]_\alpha[w]^{-1}_\alpha \end{cases},$$

provided that w is a proper fuzzy number.

References

1. Bede, B., Gal, S.G.: Generalizations of the differentiability of fuzzy number valued functions with applications to fuzzy differential equations. Fuzzy Sets Syst. 151, 581–599 (2005)
2. Bouchon-Meunier, B., Kosheleva, O., Kreinovich, V., Nguyen, H.T.: Fuzzy numbers are the only fuzzy sets that keep invertible operations invertible. Fuzzy Sets Syst. 91, 155–163 (1997)
3. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc, River Edge (1994)
4. Hukuhara, M.: Integration des applications measurables dont la valeur est un compact convexe. Fukc Ekvacioj 10, 205–223 (1967)
5. Laksmikantham, V., Mohapatra, R.N.: Theory of Fuzzy Differential Equations and Inclusions. Taylor and Francis, New York (2003)
6. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs (1974)
7. Stefanini, L.: A generalization of Hukuhara difference for interval and fuzzy arithmetic. Working Paper EMS Series, University of Urbino (2008), www.repec.org

A Note about Bobylev's Differential

Luis J. Rodríguez-Muñiz

Dept. of Statistics and O.R., University of Oviedo, E.U.I.T. Industrial, Gijón, Spain

Abstract. In this work Bobylev's definition of differential of a fuzzy set-valued mapping is studied. Its connections with other common definitions of derivative and differential are analyzed.

Keywords: Bobylev differential, Fuzzy-valued mapping, Hukuhara derivative, s-Differential, Strong generalized differential.

1 Introduction

We study the relationships among Bobylev's differential ([4]) and other definitions simultaneously and posteriorly introduced. By developing this task, we will find that Bobylev's definition can generalize some of those definitions, as well as it is generalized by some others.

This topic on differentials can be used not only in a pure Mathematical Analysis framework, but also in connection with other disciplines like Probability or Statistics (see, for instance, [15, 16, 17]).

In Section 2 we introduce some notation and preliminaries. In Section 3 we recall Bobylev's definitions. In Section 4 we study the relationships among Bobylev's definitions and previous ones. Finally, some open problems are addressed.

2 Notation and Preliminaries

$\mathscr{K}(\mathbb{R}^n)$ will denote the class of compact subsets of $\mathbb{R}^n$ ($\mathscr{K}_c(\mathbb{R}^n)$ will stand for the convex case). $\mathscr{F}(\mathbb{R}^n)$ will denote the class of fuzzy subsets $A:\mathbb{R}^n\to[0,1]$ with α-cuts A_α in $\mathscr{K}(\mathbb{R}^n)$, for $\alpha\in[0,1]$ (being $A_0=\mathrm{cl}\{x\in\mathbb{R}^n : A(x)>0\}$). $\mathscr{F}_c(\mathbb{R}^n)$ will stand when $A_\alpha\in\mathscr{K}_c(\mathbb{R}^n)$, for $\alpha\in[0,1]$.

The class $\mathscr{F}_c(\mathbb{R}^n)$ is endowed (see [20] or [8]) with a semilinear structure, by defining $(A+B)_\alpha=A_\alpha+B_\alpha$ (Minkowski's addition) and $(\lambda A)_\alpha=\lambda A_\alpha$, being $A,B\in\mathscr{F}_c(\mathbb{R}^n)$, $\lambda\in\mathbb{R}$. We will also use the generalized Hausdorff distance ([10]), given by $d_\infty(A,B)=\sup_{\alpha\in[0,1]}d_H(A_\alpha,B_\alpha)$, for $A,B\in\mathscr{F}(\mathbb{R}^n)$, d_H being the Hausdorff distance.

For $A\in\mathscr{F}_c(\mathbb{R}^n)$, if the mapping $\alpha\to A_\alpha$ is continuous with respect to the Euclidean and the d_H metrics, we will say that $A\in\mathscr{F}_{cc}(\mathbb{R}^n)$. In [11] it is proved that $(\mathscr{F}_{cc}(\mathbb{R}^n),d_\infty)$ is complete and closed and it can be isometrically embedded into the Hilbert space of continuous real-valued functions from $[0,1]\times S^{n-1}$ with the usual $\|\cdot\|_\infty$ norm, by the embedding:

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 211–217, 2008.
springerlink.com

$$j : \mathscr{F}_{cc}(\mathbb{R}^n) \to \mathscr{C}([0,1] \times S^{n-1})$$

given by $j(A) = s_A$. Moreover, $\mathscr{F}_{cc}(\mathbb{R}^n)$ is the greatest subclass of $\mathscr{F}_c(\mathbb{R}^n)$ that can be isometrically embedded into $\mathscr{C}([0,1] \times S^{n-1})$.

Puri and Ralescu's definition of support function (see [10], [11]) for $A \in \mathscr{F}_{cc}(\mathbb{R}^n)$ and S^{n-1} the unit sphere in $\mathbb{R}^n$ is given by $s_A : [0,1] \times S^{n-1} \to \mathbb{R}$, with $s_A(\alpha, p) = \sup_{x \in A_\alpha} < p, x >$, being $< \cdot, \cdot >$ the inner product in $\mathbb{R}^n$. In the compact convex case, the supremum is attained. In [10] is also proved how the support function characterizes the fuzzy set. Other properties, mainly inheritated from the set-valued case can be found in [10], [6] and [11].

In [6] and [13] an overview of several definitions of differential can be found, including De Blasi's definition, the π-differential and the conical differential (introduced in [9] based on the set-valued case in [1]), and the s-differential (introduced in [14]), which is only defined for mappings going into $\mathscr{F}_{cc}(\mathbb{R}^n)$. Other definitions can also be found in, for instance, [2] or [12]. We can also consider the Fréchet differential of the support function, which is a special case of the s-differential. Of course, in those definitions of differential using Puri and Ralescu's support function, we should yield into the class $\mathscr{F}_{cc}(\mathbb{R}^n)$ to guarantee the well-definition, since in $\mathscr{F}_c(\mathbb{R}^n) \setminus \mathscr{F}_{cc}(\mathbb{R}^n)$ the isometry by j does not hold in general.

3 Bobylev's Definitions

Bobylev introduced ([3]) a concept of support function of a fuzzy set in $\mathscr{F}_c(\mathbb{R}^n)$, we will use φ for distinguishing it from the support function by Puri and Ralescu s. For $A \in \mathscr{F}_c(\mathbb{R}^n)$, the support function is given by $\varphi_A : B^n \to \mathbb{R}$, with $\varphi_A(k) = \sup_{\{x \in \mathbb{R}^n : A(x) \geq \|k\|\}} \{< k, x >\}$, B^n being the ball in $\mathbb{R}^n$ centered at $\mathbf{0} \in \mathbb{R}^n$ with radius 1.

In [3] it is demonstrated that φ_A is unique and its main properties are stated: (1) uppersemicontinuity, (2) positive homogeneity, (3) quasiadditiveness, (4) normality, (5) φ is a bounded operator, (6) $\varphi(\mathbf{0}) = 0$.

If we denote by Φ^n the family of all functions $\varphi : B^n \to \mathbb{R}$ satisfying above conditions (1) to (6), we have ([3]) that Φ^n is the set of all support functions of fuzzy sets in $\mathscr{F}_c(\mathbb{R}^n)$. If we denote by Ψ^n the set of all functions $\varphi : B^n \to \mathbb{R}$ satisfying above conditions (5) and (6), Ψ^n can be endowed with a linear structure by means of pointwise algebraic operations.

Then, a norm can be defined on Ψ^n by $\|\varphi\|_{\Psi^n} = \sup_{x \in B^n \setminus \{\mathbf{0}\}} \left\{ \frac{|\varphi(x)|}{\|x\|} \right\}$. Φ^n is a closed subset of Ψ^n and it is nonseparable in the sense of norm $\|\cdot\|_{\Psi^n}$. Bobylev defines the distance between two fuzzy subsets $A, B \in \mathscr{F}_c(\mathbb{R}^n)$ as $d_\varphi(A,B) = \|\varphi_A - \varphi_B\|_{\Psi^n}$. Bobylev proves that the space $(\mathscr{F}_c(\mathbb{R}^n), d_\varphi)$ is complete and nonseparable.

In [18] it is demonstrated a relationship between s_A and φ_A, and the equivalence between d_∞ and d_φ. Based on those results we are obtaining the results in the following section.

In [4] Bobylev introduced a concept of differential for fuzzy-valued mappings. The definition of the differential is based on the support function, since it yields on a Hilbert space.

Definition 1. *Given O an open subset of $\mathbb{R}^l$, and a fuzzy-valued mapping $F : O \to \mathscr{F}_c(\mathbb{R}^n)$, F is said to be Bobylev-differentiable at $t_0 \in O$ if the mapping $\varphi_{F(\cdot)} : O \to \Phi^n \subseteq \Psi^n$, given by $\varphi_{F(t)}$, is Fréchet differentiable at t_0 (with respect to the Euclidean norm and the norm $\|\cdot\|_{\Psi^n}$) (being $\varphi'_{F_{t_0}} : \mathbb{R}^l \to \Psi^n$ its differential) and, uniformly in $t \in \mathbb{R}^l$, there exist a fuzzy set $F^B_{t_0}(t) \in \mathscr{F}_c(\mathbb{R}^n)$ such that $\varphi'_{F_{t_0}}(t) = \varphi_{F^B_{t_0}(t)}$.*

4 Differentials

In this section we examine the relationships between Bobylev differential and others introduced in the literature, these are De Blasi differential ([5, 6]), Hukuhara derivative ([7, 9]), s-differential ([13, 14]) and strong generalized differential ([2]). In [6], [13] and [2] some of the properties of these definitions and relationships among them are studied.

Within the class $\mathscr{F}_{cc}(\mathbb{R}^n)$, the s-differential is the most general definition, thus, we are starting by proving what happens with the Bobylev differential of a mapping taking on values on $\mathscr{F}_{cc}(\mathbb{R}^n)$. Unfortunately we cannot guarantee, in general, that the Bobylev differential yields in the same class than the mapping. We can see it in the following example.

Example 1. Let us consider the mapping $F : (0.5, 2) \to \mathscr{F}_{cc}(\mathbb{R})$ where $F(t) : \mathbb{R} \to [0,1]$ is given by:

$$F(t)(x) = \begin{cases} 2x, & \text{if } x \in (0, 0.25); \\ 0.5, & \text{if } x \in [(0.25, t-0.25) \\ & \text{or } x \in [2t+0.25, 3t-0.25); \\ 2(x-t)+1, & \text{if } x \in [t-0.25, t); \\ 1, & \text{if } x \in [t, 2t); \\ 1-2(x-2t), & \text{if } x \in [2t, 2t+0.25); \\ 2(3t-x), & \text{if } x \in [3t-0.25, 3t); \\ 0, & \text{else.} \end{cases}$$

Next result states that within the class $\mathscr{F}_{cc}(\mathbb{R}^n)$ the s-differential generalizes the Bobylev differential.

Proposition 1. *Let O be an open subset of $\mathbb{R}^l$ and let $F : O \to \mathscr{F}_{cc}(\mathbb{R}^n)$ be a mapping. Let F be Bobylev differentiable at $t_0 \in O$ with $F^B_{t_0}(t) \in \mathscr{F}_{cc}(\mathbb{R}^n)$, for every $t \in \mathbb{R}^l$ (being $F^B_{t_0}(t)$ the set in the Bobylev differential, in Definition 1). Then F is also s-differentiable at t_0, and its s-differential is given by:*

$$F'_{t_0}(t) = s_{F^B_{t_0}(t)},$$

for every $t \in \mathbb{R}^l$.

Remark 1. It is obvious that, under hypothesis in Proposition 1, the Bobylev differentiability implies not only the s-differentiability but also the Fréchet differentiability of the associated support function, that is, $s_{F(\cdot)} : O \to \mathscr{C}([0,1] \times S^{n-1})$ associating $t \mapsto s_{F(t)}$.

The converse of Proposition 1 is not true in general. It is easy to find counterexamples, going from nonlinearity (required for the Bobylev differential but for the s-differential) to the existence of that set (we have denoted $F^B_{t_0}(\cdot)$) having as support function the Fréchet differential of the mapping $\varphi_{F(\cdot)}$. The following are two of them.

Example 2. Consider the mapping $F : (.5,2) \to \mathscr{F}_{cc}(\mathbb{R})$ where $F(t) : \mathbb{R} \to [0,1]$ is given by:

$$F(t)(x) = \begin{cases} 2(xt+1), & \text{if } x \in [\frac{-1}{t}, \frac{-1}{2t}); \\ 1, & \text{if } x \in [\frac{-1}{2t}, \frac{1}{2t}); \\ -2(xt-1), & \text{if } x \in [\frac{1}{2t}, \frac{1}{t}); \\ 0, & \text{else.} \end{cases}$$

F is not Bobylev differentiable at 1, because there does not exist a fuzzy set $F^B_1(t) \in \mathscr{F}_c(\mathbb{R})$ such that whose Bobylev support function is the Fréchet differential of $\varphi_{F(\cdot)}$ at 1 in Ψ^n (see Definition 1). Otherwise, let us suppose that F is Bobylev differentiable at 1, and, whence, there exist such a fuzzy set $F^B_1(t) \in \mathscr{F}_c(\mathbb{R})$ such that $\varphi'_{F_1}(t) = \varphi_{F^B_1(t)}$, for every $t \in \mathbb{R}$. Due to Proposition 1, it must hold $F'_1(t) = s_{F^B_1(t)}$. But it is impossible for $F'_1(t)$ to be the Puri and Ralescu support function of any fuzzy set in $\mathscr{F}_c(\mathbb{R})$ since $F'_1(t)(\cdot, p)$ is increasing in its first parameter (α), for fixed t and p, and support functions must be nonincreasing in α (see, for instance, [6] or [10]).

But even when assuming the existence of such a set $F^B_{t_0}(t)$ as in Counterexample 2, Bobylev differentiability can fail because of the Fréchet differentiability, as we show in the following counterexample.

Example 3. Consider the mapping $F : \mathbb{R} \to \mathscr{F}_{cc}(\mathbb{R}^2)$ given by $F(t) = \mathbf{1}_{tB^2}$. This mapping is s-differentiable at 0 (see, for instance, [13]). But it is not Bobylev differentiable at 0, since the support function associated with every $t \in \mathbb{R}$ is $\varphi_{F(t)} : B^2 \to \mathbb{R}$, given by:

$$\varphi_{F(t)}(k) = \sup_{\{x : \mathbf{1}_{tB^2}(x) \geq \|k\|\}} < k, x > .$$

From the definition of $< \cdot, \cdot >$ it follows that $\varphi_{F(t)}(k) = |t|\|k\|$. Thus, $\varphi_{F(t)}$, obviously cannot be Fréchet differentiable at 0.

When the mapping takes values not only on $\mathscr{F}_{cc}(\mathbb{R}^n)$ but on the general class $\mathscr{F}_c(\mathbb{R}^n)$, we can state the forthcoming results, starting with the Hukuhara derivative.

Proposition 2. *Let O be an open interval of $\mathbb{R}$ and let $F : O \to \mathscr{F}_c(\mathbb{R}^n)$ be a mapping. If F is Hukuhara derivable at $t_0 \in O$, then F is also Bobylev differentiable at $t_0 \in O$ and its differential is given by:*

$$\varphi'_{t_0}(t) = t\varphi_{F^h(t_0)},$$

for $t \in \mathbb{R}$, and $F^h(t_0)$ being the Hukuhara derivative of F at t_0.

The converse of Proposition 2 is not true in general. Counterexamples can be constructed from the set-valued analysis (see, for instance, [1]).

Example 4. Consider the mapping $F : (0, 2\pi) \to \mathscr{F}_c(\mathbb{R})$, given by:

$$F(t)(x) = \begin{cases} 1 + \frac{x}{2+\sin t}, & \text{if } x \in [-(2+\sin t), 0]; \\ 1 - \frac{x}{2+\sin t}, & \text{if } x \in (0, 2+\sin t]; \\ 0, & \text{else;} \end{cases}$$

which corresponds to the triangular fuzzy number with maximum at 0 and based on the interval $[-(2+\sin t), 2+\sin t]$. It is easy to see that this mapping is not Hukuhara derivable at any point $t_0 \in (0, 2\pi)$ (see [6]).

When we consider the more general case of the strongly generalized differential ([2]), based on the Hukuhara derivative, we can check that it is also a particular case of the Bobylev differential.

Proposition 3. *Let O be an open interval of $\mathbb{R}$ and let $F : O \to \mathscr{F}_c(\mathbb{R}^n)$ be a mapping. If F is strongly generalized differentiable at $t_0 \in O$, then F is also Bobylev differentiable at $t_0 \in O$ and its differential is given by:*

$$\varphi'_{t_0}(t) = t\varphi_{F^s(t_0)},$$

for $t \in \mathbb{R}$, and $F^s(t_0)$ being strong generalized differential of F at t_0.

The converse result of Proposition 3 is not true, in general. We can consider the function in Example 4, and it holds also as a counterexample for the strongly generalized differentiable case.

On the other hand, Bobylev differentiability is a particular case of De Blasi differentiability when working in the general class $\mathscr{F}_c(\mathbb{R}^n)$, as we can see in the following result:

Proposition 4. *Let O be an open subset of $\mathbb{R}^l$ and let $F : O \to \mathscr{F}_c(\mathbb{R}^n)$ be a mapping. If F is Bobylev differentiable at $t_0 \in O$, then F is also De Blasi differentiable at $t_0 \in O$ and its De Blasi differential is given by:*

$$DF_{t_0}(t) = F^B_{t_0}(t),$$

for every $t \in \mathbb{R}^l$, being $F^B_{t_0}(t)$ the set appearing in the Bobylev differential of F at t_0.

Converse result of Proposition 4 is not true, in general, since the lack of linearity and continuity of the De Blasi differential.

Example 5. Consider a fuzzy set $A \in \mathscr{F}_c(\mathbb{R}^n)$ and the mapping $F : \mathbb{R} \to \mathscr{F}_c(\mathbb{R}^n)$ given by $F(t) = |t|A$. Obviously, F is De Blasi differentiable at 0, since it is continuous and positively homogeneous, therefore F is its own De Blasi differential at 0. But F cannot be Bobylev differentiable at 0 since it would imply that $t \mapsto |t|$ would be Fréchet differentiable at 0.

5 Future Lines

Some open problems can be addressed.

- To define a relaxed-conditions differential based on Bobylev's one, as the *s*-differential is to the Puri and Ralescu's support function, and to study the properties of this new concept. This would keep good properties of Fréchet-type differentials but relaxing conditions within a more general space.
- To study other types of embeddings (like that in [19]), so that the continuity condition on the α-cuts can be omitted. Thus, *s*-differential and Bobylev differential could make easy to be compared.
- To study the problem of the Steiner point with Bobylev definition.
- To analyze the integral defined by Bobylev in [4] and its relationships with other concepts of integral for fuzzy-valued mappings, and, more precisely, with the concept of fuzzy expected value of a fuzzy random variable. Main problems to be studied here come from the non-separability of the space $(\mathscr{F}_c(\mathbb{R}^n), d_\infty)$, thus some other distances could be considered.
- To study the differential equations related to this concept of differential.

Acknowledgement. The research in this paper has been supported by the Spanish Ministry of Education and Science Grant MTM2005-02254. Its financial support is gratefully acknowledged.

References

1. Banks, H.T., Jacobs, M.Q.: A differential calculus for multifunctions. J. Math. Anal. Appl. 29, 246–272 (1970)
2. Bede, B., Gal, S.G.: Generalizations of the differentiability of fuzzy-number-valued functions with applications to fuzzy differential equations. Fuzzy Sets Syst. 151, 581–599 (2005)
3. Bobylev, V.N.: Support function for a fuzzy set and its characteristic properties. Akademiya Nauk SSSR Matematicheskie Zametki 37, 507–513 (1985)
4. Bobylev, V.N.: Cauchy problem under fuzzy control. Busefal 21, 117–126 (1985)
5. De Blasi, F.S.: On the differentiability of multifunctions. Pacific J. Math. 66, 67–81 (1976)
6. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc, River Edge (1994)
7. Hukuhara, M.: Intégration des applications mesurables dont la valeur est un compact convexe. Funkcial Ekvac 10, 205–223 (1967)
8. Puri, M.L., Ralescu, D.A.: Différentielle d'une fonction floue. C. R. Acad. Sci. Paris Ser. I. Math. 293, 237–239 (1981)
9. Puri, M.L., Ralescu, D.A.: Differentials of fuzzy functions. J. Math. Anal. Appl. 91, 552–558 (1983)
10. Puri, M.L., Ralescu, D.A.: The concept of normality for fuzzy random variables. Ann. Probab. 13, 1373–1379 (1985)
11. Rojas-Medar, M.A., Bassanezi, R.C., Román-Flores, H.: A generalization of the Minkowski embedding theorem and applications. Fuzzy Sets Syst. 102, 263–269 (1999)
12. Rojas-Medar, M.A., Jiménez-Gamero, M.D., Chalco-Cano, Y., Viera-Brandão, A.J.: Fuzzy quasilinear spaces and applications. Fuzzy Sets Syst 152, 173–190 (2005)

13. Rodríguez-Muñiz, L.J.: Several notions of differentiability for fuzzy set-valued mappings. In: Bertoluzza, C., Gil, M.A., Ralescu, D.A. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data. Studies in Fuzziness and Soft Computing, vol. 87, pp. 104–116. Physica-Verlag, Heildelberg (2002)
14. Rodríguez-Muñiz, L.J., López-Díaz, M., Gil, M.A., Ralescu, D.A.: The *s*-differentiability of a fuzzy-valued mapping. Inform. Sci. 151, 283–299 (2003)
15. Rodríguez-Muñiz, L.J., López-Díaz, M., Gil, M.A.: Differentiating random upper semicontinuous functions under the integral sign. Test 12, 241–258 (2003)
16. Rodríguez-Muñiz, L.J., López-Díaz, M.: Hukuhara derivative of the fuzzy expected value. Fuzzy Sets Syst. 138, 593–600 (2003)
17. Rodríguez-Muñiz, L.J., López-Díaz, M.: On the exchange of iterated expectations of random upper semicontinuous functions. Statist. Probab. Lett. 77, 1628–1635 (2007)
18. Rodríguez-Muñiz, L.J.: Revisiting Bobylev's definitions. (submitted for publication) (2008)
19. Terán, P.: An embedding theorem for convex fuzzy sets. Fuzzy Sets Syst. 152, 191–208 (2005)
20. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part 1. Inform. Sci. 8, 199–249 (1975)

On Boundary Value Problems for Fuzzy Differential Equations

Rosana Rodríguez-López

Departamento de Análisis Matemático, Facultad de Matemáticas, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

Abstract. In many real phenomena, it is interesting to study the periodic behavior of the magnitudes involved. If a certain natural process is subject to imprecise factors, its modelization can be made by using fuzzy differential equations or fuzzy dynamical systems. The special properties of the functions which are differentiable in the sense of Hukuhara (in particular, the solutions to fuzzy differential equations) make it difficult to handle periodic phenomena by means of fuzzy differential models. We include some considerations on the analysis of boundary value problems associated with fuzzy differential equations from the point of view of Hukuhara-differentiability.

Keywords: Fuzzy differential equations, Boundary value problems, Periodic solutions.

1 Introduction

We consider the space E^1 of one-dimensional fuzzy numbers, that is, elements $x:\mathbb{R}\longrightarrow [0,1]$ satisfying the following properties:

- x is normal: there exists $\hat{t}\in\mathbb{R}$ with $x(\hat{t})=1$,
- x is upper semicontinuous,
- x is fuzzy-convex: $x(\lambda t_1+(1-\lambda)t_2)\geq \min\{x(t_1),x(t_2)\}$, for all $t_1,t_2\in\mathbb{R}$, $\lambda\in[0,1]$,
- The support of x, $supp(x)=cl\,(\{t\in\mathbb{R}:x(t)>0\})$ is a bounded subset of $\mathbb{R}$,

equipped with the metric $d_\infty(x,y)=\sup_{a\in[0,1]}d_H([x]^a,[y]^a)$, $x,y\in E^1$, where d_H represents the Hausdorff distance in $\mathscr{K}_C^1$ (the set of nonempty compact convex subsets of $\mathbb{R}$). See [1, 5], for details.

We analyze the existence of solutions for a periodic boundary value problem associated to the fuzzy differential equation $u'(t)=f(t,u(t))$, $t\in I=[t_0,T]$, where $t_0\in\mathbb{R}$, $t_0<T$, and $f:I\times E^1\longrightarrow E^1$. We consider $t_0=0$, although an analogous reasoning can be followed for any fixed $t_0\in\mathbb{R}$.

For each $x\in E^1$, we denote the level sets of x by $[x]^a=[x_{al},x_{ar}]$, $\forall a\in[0,1]$.

We study the boundary value problem

$$u'(t)=f(t,u(t)),\ t\in I=[0,T],\qquad \lambda u(0)=u(T), \tag{1}$$

where $T>0$, $f:I\times E^1\longrightarrow E^1$, and $\lambda>0$. For the compact interval I, we consider the complete metric spaces $C(I,E^1)=\{x:I\longrightarrow E^1\,|\,x \text{ is continuous}\}$, and $C^1(I,E^1)=\{x:$

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 218–225, 2008.
springerlink.com

$I \longrightarrow E^1 \,|\, x, x'$ are continuous$\}$, where x', the derivative of x, is considered in the sense of Hukuhara.

A solution to (1) is a function $u \in C^1(I, E^1)$ satisfying conditions in (1).

2 Case $\lambda > 1$

Theorem 1. *Suppose that f is continuous and k-lipschitzian with respect to the second variable, that is, $d_\infty(f(t,x), f(t,y)) \leq k\, d_\infty(x,y)$, $\forall x, y \in E^1$, where $\dfrac{\lambda k T}{\lambda - 1} < 1$. Then problem (1) has a unique solution.*

Proof. Problem (1) can be written as the equivalent problem

$$u(t) = u(0) + \int_0^t f(s, u(s))\, ds,\ t \in [0,T], \qquad \lambda u(0) = u(T). \tag{2}$$

The boundary condition produces $\lambda u(0) = u(T) = u(0) + \int_0^T f(s,u(s))\, ds$, which, in the ordinary case, is reduced to $u(0) = \frac{1}{\lambda - 1} \int_0^T f(s,u(s))\, ds$. In the fuzzy case, passing to the level sets, we get $\lambda\,[u(0)_{al}, u(0)_{ar}] = [\lambda u(0)_{al}, \lambda u(0)_{ar}] = [u(T)]^a = [u(0)_{al}, u(0)_{ar}] + \left[\int_0^T f(s,u(s))\, ds\right]^a$. In consequence,

$$u(0)_{al} = \frac{1}{\lambda - 1} \left(\int_0^T f(s,u(s))\, ds \right)_{al}, \qquad u(0)_{ar} = \frac{1}{\lambda - 1} \left(\int_0^T f(s,u(s))\, ds \right)_{ar},$$

which makes sense since $\lambda > 1$, producing the fuzzy number $u(0) = \frac{1}{\lambda - 1} \int_0^T f(s,u(s))\, ds$. Hence, to find a solution to the periodic boundary value problem (2), we have to solve the integral equation

$$\begin{aligned} u(t) &= \frac{1}{\lambda - 1} \int_0^T f(s,u(s))\, ds + \int_0^t f(s,u(s))\, ds \\ &= \int_0^t \left(\frac{1}{\lambda - 1} + 1 \right) f(s,u(s))\, ds + \int_t^T \frac{1}{\lambda - 1} f(s,u(s))\, ds \\ &= \int_0^t \frac{\lambda}{\lambda - 1} f(s,u(s))\, ds + \int_t^T \frac{1}{\lambda - 1} f(s,u(s))\, ds = \int_0^T G(t,s) f(s,u(s))\, ds, \end{aligned}$$

where

$$G(t,s) = \begin{cases} \dfrac{\lambda}{\lambda - 1}, & \text{if } 0 \leq s \leq t \leq T, \\ \dfrac{1}{\lambda - 1}, & \text{if } 0 \leq t < s \leq T. \end{cases}$$

We define the operator $\mathscr{A}$ by $[\mathscr{A}u](t) = \int_0^T G(t,s) f(s,u(s))\, ds$. By hypotheses, $\mathscr{A} : C(I,E^1) \longrightarrow C(I,E^1)$. Now, we check that $\mathscr{A}$ has a unique fixed point. Indeed,

$$
\begin{aligned}
D(\mathscr{A}u,\mathscr{A}v) &= \sup_{t\in I} d_\infty(\mathscr{A}u(t),\mathscr{A}v(t)) \\
&= \sup_{t\in I} d_\infty\left(\int_0^T G(t,s)f(s,u(s))\,ds, \int_0^T G(t,s)f(s,v(s))\,ds\right) \\
&\le \sup_{t\in I}\left(\int_0^t \frac{\lambda}{\lambda-1} d_\infty\left(f(s,u(s)),f(s,v(s))\right)\,ds\right. \\
&\qquad \left. + \int_t^T \frac{1}{\lambda-1} d_\infty\left(f(s,u(s)),f(s,v(s))\right)\,ds\right) \\
&\le \sup_{t\in I}\left(\frac{\lambda}{\lambda-1}kt + \frac{1}{\lambda-1}k(T-t)\right) D(u,v) \\
&\le \frac{1}{\lambda-1}\sup_{t\in I}\left((\lambda-1)kt+kT\right) D(u,v) \\
&= \frac{1}{\lambda-1}\left((\lambda-1)kT+kT\right) D(u,v) = \frac{\lambda kT}{\lambda-1} D(u,v).
\end{aligned}
$$

The estimate on the constants and the Contractive Mapping Principle provide the existence of a unique fixed point u for $\mathscr{A}$. Note that, for such a function u, we get $\lambda u(0) = \int_0^T \frac{\lambda}{\lambda-1} f(s,u(s))\,ds = u(T)$, and the proof is finished. □

Next, we improve Theorem 1 by replacing the estimate $\frac{\lambda kT}{\lambda-1} < 1$ by the sharper one $\frac{kT}{\ln\lambda} < 1$.

Theorem 2. *Suppose that f is continuous and k-lipschitzian with respect to the second variable and $\dfrac{kT}{\ln\lambda} < 1$. Then problem (1) has a unique solution.*

Proof. We define again the operator $[\mathscr{A}u](t) = \int_0^T G(t,s)f(s,u(s))\,ds$, where $G(t,s)$ is defined in the proof of Theorem 1, and consider the complete distance in $C(I,E^1)$ given by $D_\rho(u,v) = \sup_{t\in I} d_\infty(u(t),v(t))e^{-\rho t}$, for $u, v \in C(I,E^1)$, where $\rho > 0$. Then

$$
\begin{aligned}
D_\rho(\mathscr{A}u,\mathscr{A}v) &= \sup_{t\in I} d_\infty(\mathscr{A}u(t),\mathscr{A}v(t))e^{-\rho t} \\
&\le \sup_{t\in I}\left(\int_0^t \frac{\lambda}{\lambda-1} k d_\infty\left(u(s),v(s)\right)\,ds + \int_t^T \frac{1}{\lambda-1} k d_\infty\left(u(s),v(s)\right)\,ds\right) e^{-\rho t} \\
&\le \sup_{t\in I}\left(\frac{\lambda k}{\lambda-1}\int_0^t e^{\rho s}\,ds + \frac{k}{\lambda-1}\int_t^T e^{\rho s}\,ds\right) e^{-\rho t} D_\rho\left(u,v\right) \\
&= \sup_{t\in I}\left(\frac{\lambda k}{\lambda-1}\frac{1-e^{-\rho t}}{\rho} + \frac{k}{\lambda-1}\frac{e^{\rho(T-t)}-1}{\rho}\right) D_\rho\left(u,v\right) \\
&= \frac{k}{(\lambda-1)\rho}\sup_{t\in I}\left(\lambda\left(1-e^{-\rho t}\right)+e^{\rho(T-t)}-1\right) D_\rho\left(u,v\right) \\
&= \frac{k}{(\lambda-1)\rho}\sup_{t\in I}\left(\lambda-1+(e^{\rho T}-\lambda)e^{-\rho t}\right) D_\rho\left(u,v\right).
\end{aligned}
$$

Now, taking $\rho = \frac{1}{T}\ln\lambda > 0$, we get $e^{\rho T} = \lambda$ and, in consequence,

$$D_\rho(\mathscr{A}u,\mathscr{A}v) \le \frac{k}{(\lambda-1)\rho}\sup_{t\in I}\left(\lambda-1+(e^{\rho T}-\lambda)e^{-\rho t}\right)D_\rho(u,v)$$
$$= \frac{(\lambda-1)k}{(\lambda-1)\frac{1}{T}\ln\lambda}D_\rho(u,v) = \frac{kT}{\ln\lambda}D_\rho(u,v).$$

It is easy to check that the value $\rho = \frac{1}{T}\ln\lambda > 0$ produces the best estimate on the constants. The Contractive Mapping Principle provides the existence of a unique fixed point for $\mathscr{A}$, and the proof is concluded. □

3 Case $0 \le \lambda \le 1$

Solutions u of problem (1) satisfy that $diam([u(t)]^a)$ is nondecreasing in the variable t, for each $a \in [0,1]$ fixed, therefore the boundary condition $\lambda u(0)_{al} = u(T)_{al}$, $\lambda u(0)_{ar} = u(T)_{ar}$ and $\lambda \in [0,1]$ imply that $diam([u(T)]^a) = u(T)_{ar} - u(T)_{al} = \lambda(u(0)_{ar} - u(0)_{al}) = \lambda diam([u(0)]^a) \le u(0)_{ar} - u(0)_{al} = diam([u(0)]^a)$.

If $0 < \lambda < 1$ and $u(0)$ is not crisp, then for some a, $diam([u(T)]^a) < diam([u(0)]^a)$, hence we can not find a solution to (1). For the existence of solution, it is necessary that $\lambda u(0) = u(T) = u(0) + \int_0^T f(s,u(s))\,ds$, hence

$$(\lambda-1)(u(0))_{al} = \int_0^T (f(s,u(s)))_{al}\,ds, \quad (\lambda-1)(u(0))_{ar} = \int_0^T (f(s,u(s)))_{ar}\,ds,$$

in consequence, $(\lambda-1)diam([u(0)]^a) = \int_0^T diam([f(s,u(s))]^a)\,ds \ge 0$, and $diam([u(0)]^a) > 0$ leads to a contradiction. Therefore, the unique possibility is $diam([u(0)]^0) = 0$.

If $\lambda = 1$, and $diam([u(0)]^a) > 0$, then the diameter has to be a constant function in the variable t, and $diam([u(T)]^a) = diam([u(0)]^a)$, for each $a \in [0,1]$. On the other hand, if $diam([u(0)]^a) = 0$, for every a, then the initial condition is crisp and the solution is also crisp.

If $\lambda \in (0,1)$, and u_0 is crisp, then the solution is crisp.

For a different approach to periodic boundary value problems for fuzzy differential equations, see [8], where the development of the monotone iterative technique is illustrated by considering an impulsive problem.

For $\lambda = 1$, the problem under consideration is

$$u'(t) = f(t,u(t)),\ t \in I = [0,T], \qquad u(0) = u(T). \tag{3}$$

We analyze some necessary conditions to obtain (periodic) solutions to problem (3). The equivalent integral expression and the boundary condition imply that $u(0) = u(T) = u(0) + \int_0^T f(s,u(s))\,ds$, that is, $\chi_{\{0\}} = u(0) -_H u(0) = \int_0^T f(s,u(s))\,ds$. This expression is equivalent to

$$0 = \int_0^T (f(s,u(s)))_{al}\,ds \le \int_0^T (f(s,u(s)))_{ar}\,ds = 0, \text{ for every } a \in [0,1].$$

Hence $\int_0^T \left((f(s,u(s)))_{ar} - (f(s,u(s)))_{al}\right)ds = 0$, for every $a \in [0,1]$ and, by continuity, $(f(s,u(s)))_{al} = (f(s,u(s)))_{ar}$, for every $a \in [0,1]$, $s \in I$, and $0 = \int_0^T (f(s,u(s)))_{al}\,ds$.

Next, we study some necessary conditions to obtain solutions with the property that the diameter of the a-level set is a constant function in the variable t, for every $a \in [0,1]$ fixed. Indeed, for each $a \in [0,1]$,

$$\begin{aligned} &diam([u(t)]^a) \\ &= diam\left(\left[(u(0))_{al} + \int_0^t (f(s,u(s)))_{al}\, ds, (u(0))_{ar} + \int_0^t (f(s,u(s)))_{ar}\, ds\right]\right) \\ &= diam([u(0)]^a) + \int_0^t diam([f(s,u(s))]^a)\, ds. \end{aligned}$$

For this function to be constant in the variable t, for each a fixed, it is necessary that $diam([f(s,u(s))]^a) = 0$, for every a, s. Assuming that f is continuous, the solution u has level sets with constant diameter if, for every $a \in [0,1]$, and every s, $diam([f(s,u(s))]^a) = 0$, that is, if $f(t,u(t))$ is crisp, for every $t \in I$.

In particular, if $f(t,x)$ is crisp, for every $t \in I$ and every $x \in E^1$, then the diameter of each level set for the solutions to the initial value problem associated to equation $u'(t) = f(t,u(t))$, $t \in I$, is constant. Note that this does not mean that the solutions are crisp, but $diam([u(t)]^a) = diam([u(0)]^a)$, for every $t \in I$ and $a \in [0,1]$, that is, the diameter of each level set of the solution is the diameter of the corresponding level set of the initial condition. Under this assumption, there could be fuzzy periodic solutions.

Example 1. Consider the fuzzy initial value problem

$$u'(t) = \chi_{\{3\}},\ t \in I = [0,T], \qquad u(0) = \chi_{[0,1]}. \tag{4}$$

Passing to the level sets, we get the equations $x' = y' = 3$, $x(0) = 0$, $y(0) = 1$, hence $x(t) = 3t$, $y(t) = 1 + 3t$, for every t, and the solution u to (4) is given by $[u(t)]^a = [3t, 1+3t]$, for every t and a, that is, $u(t) = \chi_{[3t,1+3t]} = \chi_{\{3t\}} + \chi_{[0,1]}$, $t \in I$. Note that $diam([u(t)]^a) = 1 = diam([u(0)]^a)$, for every $t \in I$ and $a \in [0,1]$.

Example 2. Now, consider the fuzzy initial value problem

$$u'(t) = \chi_{\{3\}},\ t \in I = [0,T], \qquad u(0) = u_0, \tag{5}$$

where $u_0 = (0;1,1)$ is the triangular fuzzy number given by

$$u_0(t) = \begin{cases} t+1, t \in [-1,0], \\ 1-t, t \in [0,1], \end{cases}$$

whose levelsets are $[u_0]^a = [-(1-a), 1-a]$, for every $a \in [0,1]$. Passing to the level sets, we get the equations $x' = y' = 3$, $x(0) = -(1-a)$, $y(0) = 1-a$, hence $x(t) = -(1-a) + 3t$, $y(t) = (1-a) + 3t$, $t \in I$, and the solution to (5) is the function u given by $[u(t)]^a = [-(1-a) + 3t, (1-a) + 3t]$, for every $t \in I$ and $a \in [0,1]$, that is, $u(t) = \chi_{\{3t\}} + u_0$, $t \in I$. Note that $diam([u(t)]^a) = 2(1-a) = diam([u(0)]^a)$, for every $t \in I$ and $a \in [0,1]$. We remark that $u(t)$ is also a triangular fuzzy number, for each t, that is, $u(t) = (3t;1,1)$.

Then, assuming that the right-hand side in the equation is a crisp function, we obtain solutions with constant diameter, and we obtain T-periodic solutions u to the fuzzy

differential equation in the sense of Hukuhara if $\int_0^T f(s,u(s))\,ds = \chi_{\{0\}}$. This is the situation if, for instance, there exists $c : I \longrightarrow \mathbb{R}$ such that $f(s,x) = \chi_{\{c(s)\}}$, for every $s \in I$ and $x \in E^1$, and $\int_0^T f(s,x)\,ds = \chi_{\{0\}}$, $\forall x \in E^1$ ($\int_0^T c(s)\,ds = 0$).

Example 3. Take $f(t,x) = -1 + \frac{2}{T}t$, for $t \in I = [0,T]$ and $x \in \mathbb{R}$, which is a continuous crisp function satisfying that $\int_0^T f(s,x)\,ds = \int_0^T \left(-1 + \frac{2}{T}s\right) ds = \left[-s + \frac{s^2}{T}\right]_0^T = 0$, for every $x \in \mathbb{R}$.

Consider the fuzzy initial value problem

$$u'(t) = \chi_{\{-1+\frac{2}{T}t\}}, t \in I = [0,T], \qquad u(0) = \chi_{[0,1]}, \tag{6}$$

which can be easily solved, obtaining that $[u(t)]^a = \left[-t + \frac{t^2}{T}, 1 - t + \frac{t^2}{T}\right]$, for every $t \in I$ and $a \in [0,1]$, that is, $u(t) = \chi_{\{-t+\frac{t^2}{T}\}} + \chi_{[0,1]}$, $t \in I$. Besides, $u(T) = \chi_{[0,1]}$. Hence $u(t) = \chi_{\left[-t+\frac{t^2}{T}, 1-t+\frac{t^2}{T}\right]}$, $t \in I$, is a T-periodic solution for (6), and $diam([u(t)]^a) = 1$, for every $t \in I$ and every $a \in [0,1]$.

If we take the triangular fuzzy number $u_0 = (0;1,1)$, then $[u(t)]^a = \left[-(1-a) - t + \frac{t^2}{T}, (1-a) - t + \frac{t^2}{T}\right]$, $\forall t$, $\forall a$, which defines a triangular fuzzy number $u(t) = \chi_{\{-t+\frac{t^2}{T}\}} + (0;1,1)$, for every $t \in I$. Besides, $diam([u(t)]^a) = 2(1-a) = diam([u_0]^a)$, $\forall t \in I$, $a \in [0,1]$. Note that $[u(T)]^a = [u(0)]^a$, $\forall a \in [0,1]$, then $u(0) = (0;1,1) = u(T)$.

Example 4. For the problem

$$u'(t) + \chi_{\{1\}} = \chi_{\{\frac{2}{T}\}} u(t), t \in I = [0,T], \qquad u(0) = u(T), \tag{7}$$

we have, at least, the periodic solution $u(t) = \chi_{\{\frac{T}{2}\}}$. If we start at a crisp initial condition, the periodic solutions are crisp, since $diam([u(t)]^a) = 0$, for every $t \in I$ and every $a \in [0,1]$. If the initial condition is not crisp, it is necessary for the diameter of the solution to be constant.

Remark 1. If $0 < \lambda < 1$, the boundary condition $\lambda u(0) = u(T)$ and the integral representation of the solution imply that $u(0)_{al} = \frac{1}{\lambda - 1} \int_0^T \left(f(s,u(s))\right)_{al} ds$, $u(0)_{ar} = \frac{1}{\lambda - 1} \int_0^T \left(f(s,u(s))\right)_{ar} ds$. However, if $\left(f(s,u(s))\right)_{al} < \left(f(s,u(s))\right)_{ar}$ for s in a set of positive measure, taking into account that $\lambda < 1$, then $u(0)_{al} > u(0)_{ar}$, and we do not obtain a fuzzy number. Hence, the unique possibility to obtain a solution to the boundary value problem is that $\left(f(s,u(s))\right)_{al} = \left(f(s,u(s))\right)_{ar}$, for almost every s (by continuity, for all s), and u_0 crisp. In this case, $u(0) = \frac{1}{\lambda - 1} \int_0^T f(s,u(s))\,ds$, and the solution is crisp.

Remark 2. If $f : I \times E^1 \longrightarrow E^1$ is such that $f(t,\chi_{\{x\}}) = \chi_{\{g(t,x)\}}$, for every $t \in I$ and $x \in \mathbb{R}$, where $g : I \times \mathbb{R} \longrightarrow \mathbb{R}$, and the crisp equation $y'(t) = g(t,y(t))$, $t \in I$, has a real solution y satisfying that $\lambda y(0) = y(T)$, then $u(t) = \chi_{\{y(t)\}}$, $t \in I$, is a solution to the boundary value problem (1).

Remark 3. Suppose that $f : I \times E^1 \longrightarrow E^1$ is such that $f(t,x) \in \mathbb{R}$, for every $t \in I$ and $x \in \mathbb{R}$, and, besides, assume that for all symmetric triangular fuzzy number x, $[f(t,x)]^a = [f(t,mp([x]^a)), f(t,mp([x]^a))]$, for every $t \in I$, $a \in [0,1]$, where $mp([x]^a)$ represents the midpoint of the interval $[x]^a$. Suppose also that the crisp equation $u'(t) = f(t,u(t))$ has a real solution u. Then the function given by the corresponding triangular fuzzy numbers $[\tilde{u}(t)]^a = [-(1-a) + u(t), 1-a+u(t)]$, $\forall a \in [0,1]$, and $t \in I$, is such that $[\tilde{u}'(t)]^a = [u'(t), u'(t)] = [f(t,u(t)), f(t,u(t))] = [f(t,\tilde{u}(t))]^a$, $\forall a \in [0,1]$, $t \in I$, hence $\tilde{u}$ is a solution to the fuzzy equation. Note that, for x a symmetric triangular fuzzy number, $\{mp([x]^a)\} = [x]^1$, for every $a \in [0,1]$. We remark that, if the solution u to the crisp equation satisfies that $\lambda u(0) = u(T)$, then $\lambda\tilde{u}(0) = \lambda(u(0);1,1) = (\lambda u(0);\lambda,\lambda) = (u(T);\lambda,\lambda)$, which is equal to $\tilde{u}(T)$ if $\lambda = 1$. Thus, this method is useful to finding periodic solutions.

Remark 4. Suppose that $\widetilde{f}(t,x)$ is a fuzzy function and that there exist $h_1(t,a)$, $h_2(t,a)$, such that $h_1(t,a) \leq 0 \leq h_2(t,a)$, $h_1(t,a)$ is nondecreasing in a, $h_2(t,a)$ is nonincreasing in a, for each t fixed, h_1, h_2 are left-continuous in a, $\frac{\partial h_1}{\partial t}(t,a) \leq 0$, $\frac{\partial h_2}{\partial t}(t,a) \geq 0$, $h_1(T,a) - \lambda h_1(0,a) = h_2(T,a) - \lambda h_2(0,a)$, for every $a \in [0,1]$, and $diam([\widetilde{f}(t,\widetilde{x})]^a) = h_2'(t,a) - h_1'(t,a)$, for every $t \in I$, $\widetilde{x} \in E^1$, and $a \in [0,1]$. For $t \in I$ and $x \in \mathbb{R}$, we define

$$f(t,x) = \widetilde{f}(t,\chi_{\{x\}} + \widetilde{r}(t))_{ar} - h_2'(t,a) = \widetilde{f}(t,\chi_{\{x\}} + \widetilde{r}(t))_{al} - h_1'(t,a),$$

where $[\widetilde{r}(t)]^a = [h_1(t,a), h_2(t,a)]$, for $t \in I$ and $a \in [0,1]$. Suppose that the real boundary value problem

$$u'(t) = f(t,u(t)),\ t \in I = [0,T],\ \lambda u(0) = u(T) + [h_1(T,a) - \lambda h_1(0,a)], \tag{8}$$

has a solution $u(t)$, which also satisfies that $\lambda u(0) = u(T) + [h_2(T,a) - \lambda h_2(0,a)]$. Then $\tilde{u}$ given by $[\tilde{u}(t)]^a = [u(t) + h_1(t,a), u(t) + h_2(t,a)]$, $\forall t \in I$, $a \in [0,1]$, is such that $[\tilde{u}'(t)]^a = [u'(t) + h_1'(t,a), u'(t) + h_2'(t,a)] = [f(t,u(t)) + h_1'(t,a), f(t,u(t)) + h_2'(t,a)] = [\widetilde{f}(t,\tilde{u}(t))]^a$, $\forall a,t$, where we have used that $\tilde{u}(t) = \chi_{\{u(t)\}} + \widetilde{r}(t)$, and $\tilde{u}$ is a solution to the fuzzy equation. Besides, $\lambda\tilde{u}(0) = \tilde{u}(T)$. It is clear that, if we assume the more restrictive hypothesis $h_1(T,a) - \lambda h_1(0,a) = h_2(T,a) - \lambda h_2(0,a) = 0$, for every $a \in [0,1]$, and the solution u to the crisp equation satisfies that $\lambda u(0) = u(T)$, then $\lambda\tilde{u}(0) = \tilde{u}(T)$.

References

1. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc, River Edge (1994)
2. Dubois, D., Prade, H.: Fuzzy sets and systems: Theory and applications. Academic Press, New York (1980)
3. Kaleva, O.: Fuzzy differential equations. Fuzzy Sets Syst. 24, 301–317 (1987)
4. Kaleva, O.: The Cauchy problem for fuzzy differential equations. Fuzzy Sets Syst. 35, 389–396 (1990)
5. Lakshmikantham, V., Mohapatra, R.N.: Theory of Fuzzy Differential Equations and Inclusions. Taylor & Francis, London (2003)
6. Nieto, J.J., Rodríguez-López, R.: Applications of contractive-like mapping principles to fuzzy and fuzzy differential equations. Rev. Mat. Complut. 19, 361–383 (2006)

7. Nieto, J.J., Rodríguez-López, R.: Complete lattices in fuzzy real line (preprint, 2007)
8. Rodríguez-López, R.: Periodic boundary value problems for impulsive fuzzy differential equations. Fuzzy Sets Syst. 159, 1384–1409 (2008)
9. Rodríguez-López, R.: Comparison results for fuzzy differential equations. Inform. Sci. 178, 1756–1779 (2008)
10. Rodríguez-López, R.: Monotone method for fuzzy differential equations. Fuzzy Sets Syst. (in press, 2008) doi:10.1016/j.fss.2007.12.020

On Fuzzy Sets Convolution, Fuzzy Lipschitz Sets and Triangular Fuzzy Sets of Rank p

Andrés D. Báez-Sánchez[1] and Marko A. Rojas-Medar[2]

[1] Instituto de Matemática Estatística e Computação Científica, Universidad Estadual de Campinas, Campinas - SP, Brazil
[2] Departamento de Ciencias Básicas, Facultad de Ciencias, Universidad del Bío-Bío, Chillán, Chile

Abstract. In this paper we present some counterexamples to a result related to fuzzy Lipschitz sets and fuzzy sets convolution. Using the concept of fuzzy triangular set of rank p is presented an alternative proof of a interesting density result over fuzzy sets which was previosly proved by using the result belied.

Keywords: Fuzzy sets convolution, Lipschitz fuzzy sets, Triangular fuzzy sets, Density, Hausdorff metric.

1 Introduction

There exist many situations where it is necessary to approximate an arbitrary normal upper semincontinuos fuzzy set with compact support by fuzzy sets with more convenient properties, for example, by continuous fuzzy sets or lipschitzians fuzzy sets (see [2]).

In this direction, Colling and Kloeden [1] shows that the normal compact- convex fuzzy sets with compact support on $\mathbb{R}^n$ can be approximate by continuous fuzzy sets in D-metric. Also, in [3] the authors prove that the space of level-lipschitzian fuzzy sets on $\mathbb{R}^n$ is a dense subspace of the normal compact- convex and level-continuous fuzzy sets with compact support in relation to D-metric. They generalized their work to Banach spaces with interesting a-pplication to the characterization of relative compactness in spaces of fuzzy sets and the existence of fuzzy differential equations (see [4]).

In [5] is established another density result over fuzzy sets where a fundamental step in the proof is given by result below (see notation in the next section).

Proposition 1. *Let* $u, v \in \mathscr{F}(\mathbb{R}^n)$. *If* $v \in \mathscr{L}(\mathbb{R}^n)$ *then* $u \nabla v \in \mathscr{L}(\mathbb{R}^n)$.

This article presents some counter-examples to proposition above, and later we present some concepts and results that allow us to prove properly the density result obtained in [5].

2 Peliminaries

Let $(\mathscr{F}(\mathbb{R}^n), D)$ be the metric space of compacts, upper semicontinuous and normal fuzzy sets of $\mathbb{R}^n$ with D the supremum metric, $\mathscr{L}(\mathbb{R}^n)$ is the set of elementes in $\mathscr{F}(\mathbb{R}^n)$

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 226–231, 2008.
springerlink.com

having membership funtion being Lipschitz on their support and $u\nabla v$ is the convolution between fuzzy sets u, v given by:

$$(u\nabla v)(x) = \sup_{y\in\mathbb{R}^n}\{u(y)\wedge v(x-y)\},$$

with $\wedge$ denoting minimum on $[0,1]$.

Using Proposition 1 with $v = \chi_{B_{\frac{1}{p}}(0)}$, membership function of the open ball with center at 0 and radio $\frac{1}{p}$, it can be proved that

Proposition 2. *$(\mathscr{L}(\mathbb{R}^n), D)$ is dense in $(\mathscr{F}(\mathbb{R}^n), D)$.*

Proof. For details see [5]. □

We are going to show some counter-example to Proposition 1 and in Section 3, an alternative proof of Proposition 2 is presented.

2.1 Counter-Example 1

Let u, v be (see Fig. 1 and 2) defined as:

$$[u]^\alpha = \begin{cases} [0,1], & \text{if} \quad 0 \le \alpha \le 1/4 \\ \{1/2\}, & \text{if} \quad 1/4 < \alpha \le 1 \end{cases}$$

and

$$v = \chi_{B_1(0)} = \chi_{[-1,1]}$$

Since $u, v \in \mathscr{F}(\mathbb{R})$, are both normals and have compact $\alpha-$cuts. Clearly v is fuzzy Lipschitz. Lets see what happens with $u\nabla v$. Because for $u, v \in \mathscr{F}(\mathbb{R})$ we have $[u\nabla v]^\alpha = [u]^\alpha + [v]^\alpha$ this implies that

$$[u\nabla v]^\alpha = [0,1] + [-1,1] = [-1,2], \quad \text{if} \quad 0 \le \alpha \le 1/4$$

and

$$[u\nabla v]^\alpha = \{1/2\} + [-1,1] = [-1/2, 3/2], \quad \text{if} \quad 1/4 < \alpha \le 1.$$

This contradicts Proposition 1 because the set $u\nabla v$ **is not fuzzy Lipschitz** as we can easily see from Figure 3. If exist a number $K \ge 0$ such that $|(u\nabla v)(x) - (u\nabla v)(z)| \le K|x-z|$

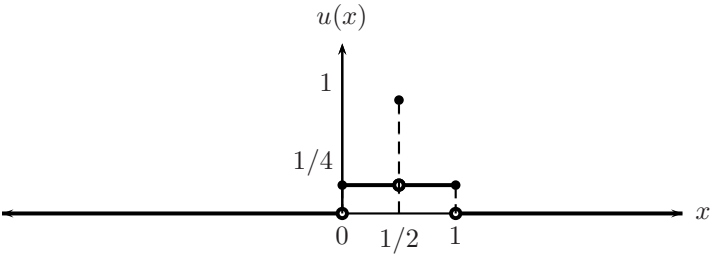


Fig. 1. $[u]^\alpha = [0,1], \quad \text{if} \quad 0 \le \alpha \le 1/4, [u]^\alpha = \{1/2\}, \quad \text{if} \quad 1/4 < \alpha \le 1$

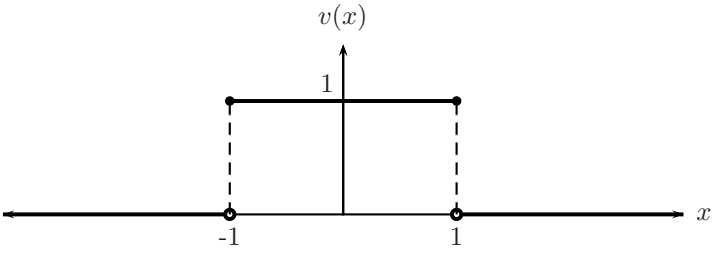


Fig. 2. $v = \chi_{B_1(0)} = \chi_{[-1,1]}$

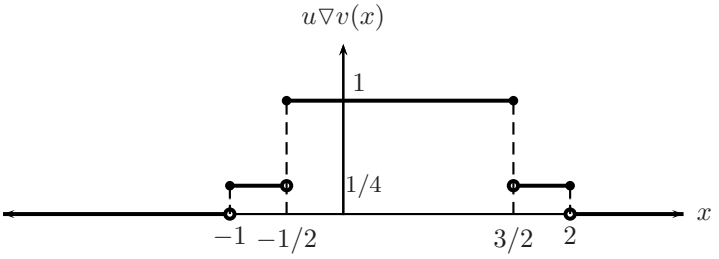


Fig. 3. Fuzzy number $u\nabla v$ is not fuzzy Lipschitz

for all $x, z \in [u\nabla v]^0 = [-1,2]$ then we would have particularly for $x = 3/2$ and any $z \in (3/2,2]$

$$\frac{|(u\nabla v)(3/2) - (u\nabla v)(z)|}{|3/2 - z|} \leq K$$

$$\frac{|1 - 1/4|}{|3/2 - z|} \leq K$$

$$\frac{|3/4|}{|3/2 - z|} \leq K$$

which is clearly absurd because z can be as close to $3/2$ as we want it.

The authors in [5] did not consider that $x - y$ and $z - y$ can not be simultaneosly in $supp(v) = [-1,1]$ for $x, z \in supp(u\nabla v)$ and $y \in supp(v)$.

Our counterexample also invalidates the argument in the proof of Proposition 2 founded in [5], since the proof uses Proposition 1 with $v = \chi_{B_{\frac{1}{p}}(0)}$.

Note that $u\nabla v$ is not fuzzy Lipschitz, because differential ratios $\frac{(u\nabla v)(x)-(u\nabla v)(z)}{x-z}$ are not bounded around $3/2$ or $-1/2$. Note also that u has a similar property around $1/2$ and in fact many other counterexamples can be constructed using fuzzy numbers u having this property.

2.2 Counter-Example 2

Previous counterexample have the characteristic that level set function of u, $[u]^{(\cdots)} : [0,1] \to K(\mathbb{R}^n)$ are discontinuous. So, if we restrict our attention to fuzzy sets u with

continuous level set function, can we get a valid result in the way of Proposition 1? The answer is again negative.

Let $u \in \mathscr{F}(\mathbb{R})$ defined by

$$u(x) = \begin{cases} \sqrt[3]{x-1}+1 & \text{if} \quad 0 \le x \le 1 \\ -\sqrt[3]{x-1}+1 & \text{if} \quad 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

which have level sets given by $[u]^\alpha = [1-(1-\alpha)^3, 1+(1-\alpha)^3]$ for $\alpha \in [0,1]$.

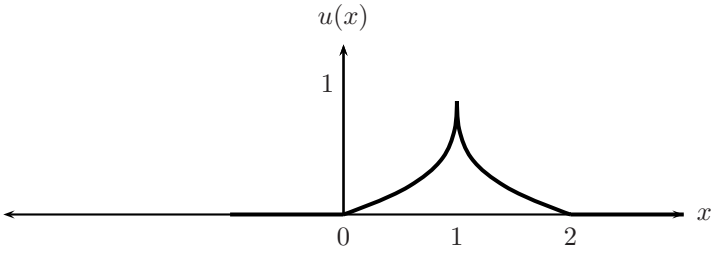


Fig. 4. $[u]^\alpha = [1-(1-\alpha)^3, 1+(1-\alpha)^3]$

So we have that u has continuous level set application but u is not Lipschitz on $\mathbb{R}^n$ not even fuzzy Lipschitz on their support, so obviously a trivial counterexample can be made taking $v = \chi_{\{0\}}$, but if we are looking for a less trivial counterexample it is enough consider $v = \chi_{B_1(0)} = \chi_{[-1,1]}$ as before.

In this situation, for any $\alpha \in [0,1]$

$$[u \nabla v]^\alpha = [1-(1-\alpha)^3, 1+(1-\alpha)^3] + [-1,1] = [-(1-\alpha)^3, 2+(1-\alpha)^3]$$

which give us by using Representation Theorem the next membership function

$$u \nabla v(x) = \begin{cases} 1+\sqrt[3]{x} & \text{if} \quad -1 \le x < 0 \\ 1 & \text{if} \quad 0 \le x \le 2 \\ 1-\sqrt[3]{x-2} & \text{if} \quad 2 < x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

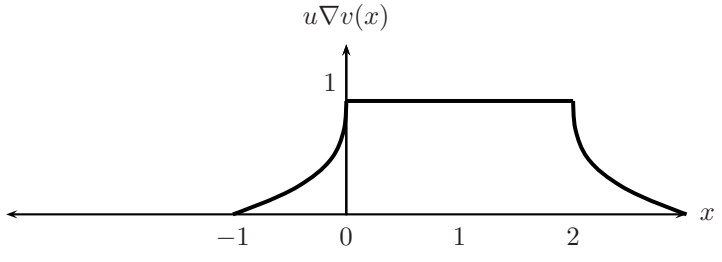


Fig. 5. $u \nabla v(x)$ is not Lipschitz

This fuzzy number is not Lipschitz and nor fuzzy Lipschitz, because again there is no bounded differential ratios, around 0 and 2 (See Figure 5).

3 Triangular Fuzzy Sets and Density

It is evident from arguments presented in [5], that if we assume that function v is Lipschitz over whole $\mathbb{R}^n$, then:

Proposition 3. *Let be $u, v \in \mathscr{F}(\mathbb{R}^n)$. If v is Lipschitz on $\mathbb{R}^n$ then $u\nabla v \in \mathscr{L}(\mathbb{R}^n)$.*

We are going to keep this result on mind and try to prove Proposition 2. The idea is to use triangular fuzzy sets instead characteristic function.

Definition 1. *For any $p \in \mathbb{R}$ $p > 0$ we define $T_p \in \mathscr{F}(\mathbb{R}^n)$ (triangular fuzzy set of rank p) as: $[T_p]^\alpha = B_{(1-\alpha)/p}(\mathbf{0})$, for all $\alpha \in [0,1]$.*

Note that from definition of T_p, we have that $T_p(x) = 0$ if and only if $\|x\| > 1/p$ and if $\|x\| \leq 1/p$ then for Representation Theorem que have

$$\begin{aligned} T_p(x) &= \sup z\{\alpha : x \in [T_p]^\alpha\} \\ &= \sup\{\alpha : x \in B[\mathbf{0}, (1-\alpha)/p]\} \\ &= \sup\{\alpha : \|x\| \leq (1-\alpha)/p\} \\ &= \sup\{\alpha : \alpha \leq 1 - p\|x\|\} \\ T_p(x) &= 1 - p\|x\| \end{aligned} \tag{1}$$

As is suggested by taking $n = 1$ (see Figure 6), T_p should be Lipschitz on whole $\mathbb{R}^n$, this actually can be proved.

Lemma 1. *For every $p > 0$, T_p is Lipschitz on $\mathbb{R}^n$ of rank p.*

Proof. Let $x, y \in \mathbb{R}^n$. If $T_p(x) = T_p(y)$ it is obvious that $|T_p(x) - T_p(y)| < K\|x - y\|$ for every $K > 0$. Lets suppose that $T_p(x) < T_p(y)$.

If $T_p(x) = 0$ then $\|x\| > 1/p$ and $0 < T_p(y)$ so $\|y\| \leq 1/p$, then $T_p(y) = 1 - p\|y\|$. In this case we have

$$|T_p(x) - T_p(y)| = T_p(y) = 1 - p\|y\| = p(1/p - \|y\|) < p(\|x\| - \|y\|) \leq p\|x - y\|$$

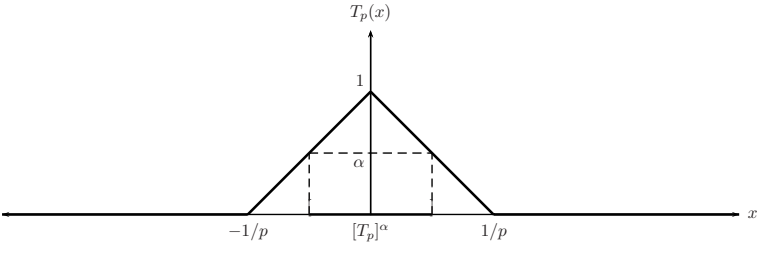


Fig. 6. T_p for $n = 1$

If $T_p(x) > 0$ then

$$|T_p(x) - T_p(y)| = T_p(y) - T_p(x) = 1 - p\|y\| - 1 + p\|x\| = p(\|x\| - \|y\|) \leq p\|x - y\|$$

□

Corollary 1. *Let $u \in \mathscr{F}(\mathbb{R}^n)$. If v is a triangular fuzzy set on $\mathbb{R}^n$ then $u\nabla v \in \mathscr{L}(\mathbb{R}^n)$.*

Proof. It follows immediately of Proposition 3 and Lemma 1. □

Using triangular fuzzy sets and last corollary it is possible to prove the density of fuzzy Lipschitz sets on $(\mathscr{F}(\mathbb{R}^n), D)$ with just a minor modification of arguments used in [5].

Proposition 4. *$(\mathscr{L}(\mathbb{R}^n), D)$ is dense in $(\mathscr{F}(\mathbb{R}^n), D)$.*

Proof. Let $u \in \mathscr{F}(\mathbb{R}^n)$ arbitrary and let T_p be as before, $p \in \mathbb{Z}^+$.
Setting $u_p = u\nabla T_p$, we have from Corollary 1 that $u_p \in \mathscr{L}(\mathbb{R}^n)$ for all $p \in \mathbb{Z}^+$.
Using properties of Hausdorff metric and α-cuts, we have for each $\alpha \in [0,1]$

$$\begin{aligned} H([u_p]^\alpha, [u]^\alpha) &= H([u\nabla v_p]^\alpha, [u]^\alpha) \\ &H([u]^\alpha + B[\mathbf{0}, (1-\alpha)/p], [u]^\alpha + \{\mathbf{0}\}) \\ &\leq H(B[\mathbf{0}, (1-\alpha)/p], \{\mathbf{0}\}) \\ &= (1-\alpha)/p. \end{aligned} \tag{2}$$

Taking sup over α we get $D(u_p, u) \leq 1/p$, for each $p \in Z^+$ and taking $p \to \infty$ we get the desired result. □

Remark 1. The section on convolution and Choquet integral established in [5] is correct when uses the results given in this work.

Acknowledgement. The first author (A.D. Báez-Sánchez) is supported by CNPq-Brazil. The second author (M.A. Rojas-Medar) is partially supported by Project MTM2007-063432, Spain and Fondecyt-Chile, Grant 1080628.

References

1. Colling, I.L., Loeden, P.: Continuous approximation of fuzzy sets. J. Fuzzy Math. 3, 449–453 (1995)
2. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc., River Edge (1994)
3. Rojas-Medar, M.A., Bassanezi, R.C., Román-Flores, H.: A generalization of the Minkowski embedding theorem and applications. Fuzzy Sets Syst. 102, 263–269 (1999)
4. Román-Flores, H., Rojas-Medar, M.A.: Embedding of level-continuous fuzzy sets on Banach spaces. Inform. Sci. 144, 227–247 (2002)
5. Román-Flores, H., Chalco-Cano, Y., Rojas-Medar, M.A.: Convolution of fuzzy sets and applications. Comput. Math. Appl. 46(8-9), 1245–1251 (2003)

Measure Theory

Generalised p-Boxes on Totally Ordered Spaces

Enrique Miranda[1], Matthias C.M. Troffaes[2], and Sébastien Destercke[3]

[1] Dept. of Statistics and O.R., Rey Juan Carlos University, Móstoles, Spain
[2] Department of Mathematical Sciences, Durham University, Durham, United Kingdom
[3] Institut de Radioprotection et de Sureté Nucléaire (IRSN), Cadarache, France

Abstract. Probability boxes are among the most simple and popular models used in imprecise probability theory, and many practical results concerning them exist in the literature. Nevertheless, little attention has been paid to their formal characterisation in the setting of Walley's behavioural theory of imprecise probabilities. This paper tries to remedy this situation by formalising, generalising and extending existing results as well as by giving new ones, within Walley's framework.

Keywords: Probability boxes, Ordered spaces, Limit representation, Characterization.

1 Introduction

Imprecise probability [7] is a generic term referring to uncertainty models where the available information does not allow singling out a unique probability measure. Unlike classical probability models, which are uniquely determined by their values on events, general imprecise probability models are determined by bounds on expectations of random variables [7, p. 82, §2.7.3]. This more advanced mathematical description allows more flexibility in the representation, but also implies more complexity when treating uncertainty.

For this reason, it is of interest to consider particular imprecise probability models that yield simpler mathematical descriptions, at the expense of generality, but gaining ease of use, elicitation, and graphical representation. One of such models is considered in this paper: pairs of lower and upper cumulative distribution functions, also called *probability boxes*, or briefly, p-boxes [3]. Practical aspects of this model have been extensively studied in the literature, but little attention has been given to their formal characterisation in terms of lower and upper expectations, or, equivalently, of coherent lower previsions (they are briefly studied in [6, 7], and in [4] cumulative distribution functions associated with a sequence of moments are considered).

This paper aims at such study, and considers a generalised version of p-boxes, defined on any (not necessarily finite) totally ordered space. In [2], a similar extension on total pre-ordered finite spaces is considered. This paper formulation covers generalised p-boxes defined on totally ordered finite spaces as well as on closed real intervals. More generally, such treatment also admits p-boxes on product spaces (by considering an appropriate order), and thus admits imprecise multivariate distributions through p-boxes as well.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 235–242, 2008.
springerlink.com

The paper is organised as follows: Section 2 provides a brief introduction to the theory of coherent lower previsions. Section 3 then introduces and studies the p-box model from the point of view of lower previsions. Section 4 provides a first expression for the natural extension of a p-box, and studies its main properties. In Section 5 we prove that any p-box can be approximated as a limit of discrete p-boxes, and that this limit holds into the natural extensions. Finally, we end in Section 6 with main conclusions and open problems. Due to limitations of space, proofs have been omitted.

2 Preliminaries

Let us briefly introduce coherent lower previsions; see [7] for more details. Let Ω be the possibility space. A subset of Ω is called an *event*. A *gamble* on Ω is a bounded real-valued function on Ω. The set of all gambles on Ω is denoted by $\mathscr{L}(\Omega)$, or simply by $\mathscr{L}$ if the possibility space is clear from the context. A particular type of gamble is the *indicator* of an event A, which is the gamble that takes the value 1 on elements of A and the value 0 elsewhere, and is denoted by I_A, or simply by A if no confusion is possible.

A *lower prevision* $\underline{P}$ is a real-valued functional defined on an arbitrary subset $\mathscr{K}$ of $\mathscr{L}$. If f is a gamble, $\underline{P}(f)$ is interpreted as the maximum buying price for the (uncertain) reward f. It can be argued that lower previsions model a subject's belief about the true state x in Ω. A lower prevision defined on a set of indicators of events is usually called a *lower probability*.

A lower prevision on $\mathscr{K}$ is called *coherent* when for all p in $\mathbb{N}$, all $f_0, f_1, \ldots, f_p$ in $\mathscr{K}$ and all $\lambda_0, \lambda_1, \ldots, \lambda_p$ in $\mathbb{R}^+$,

$$\sup_{x\in\Omega}\left[\sum_{i=1}^{p}\lambda_i(f_i-\underline{P}(f_i))-\lambda_0(f_0-\underline{P}(f_0))(x)\right]\geq 0.$$

A lower prevision on the set $\mathscr{L}$ of all gambles is coherent if and only if

(C1) $\underline{P}(f)\geq \inf f$,
(C2) $\underline{P}(\lambda f)=\lambda\underline{P}(f)$, and
(C3) $\underline{P}(f+g)\geq \underline{P}(f)+\underline{P}(g)$

for all gambles f, g and all non-negative real numbers λ. A lower prevision on $\mathscr{L}$ satisfying (C3) with equality for all gambles f and g is called a *linear prevision* on $\mathscr{L}$, and the set of all linear previsions on $\mathscr{L}$ is denoted by $\mathscr{P}$. A lower prevision $\underline{P}$ on $\mathscr{K}$ can also be characterised by the set

$$\mathscr{M}(\underline{P})=\{Q\in\mathscr{P}\colon (\forall f\in\mathscr{K})(Q(f)\geq\underline{P}(f))\}.$$

Then $\underline{P}$ is coherent if and only if $\underline{P}(f)=\min_{Q\in\mathscr{M}(\underline{P})}Q(f)$ for all $f\in\mathscr{K}$.

Given a coherent lower prevision $\underline{P}$ on $\mathscr{K}$, its *natural extension* to a larger set $\mathscr{K}_1\supseteq\mathscr{K}$ is the pointwise smallest coherent (i.e., least-committal) lower prevision on $\mathscr{K}_1$ that agrees with $\underline{P}$ on $\mathscr{K}$. The procedure of natural extension is transitive [6, p. 98]: if $\underline{E}_1$ is the natural extension of $\underline{P}$ to $\mathscr{K}_1$ and $\underline{E}_2$ is the natural extension of $\underline{E}_1$ to $\mathscr{K}_2\supseteq\mathscr{K}_1$, then $\underline{E}_2$ is also the natural extension of $\underline{P}$ to $\mathscr{K}_2$. The natural extension to all gambles is usually denoted by $\underline{E}$. It holds that $\underline{E}(f)=\min_{Q\in\mathscr{M}(\underline{P})}Q(f)$ for any $f\in\mathscr{L}$.

A particular class of coherent lower previsions of interest in this paper are *completely monotone lower previsions* [1]. A lower prevision $\underline{P}$ defined on a lattice of gambles $\mathscr{K}$ is called n-monotone when for all $p \in \mathbb{N}$, $p \leq n$, and all $f, f_1, \ldots, f_p$ in $\mathscr{K}$:

$$\Sigma_{I \subseteq \{1,\ldots,p\}} (-1)^{|I|} \underline{P}(f \wedge \textstyle\bigwedge_{i \in I} f_i) \geq 0,$$

and is called *completely monotone* when it is n-monotone for all $n \in \mathbb{N}$.

3 Characterising p-Boxes

Let $(\Omega, \leq)$ be an order complete chain. Let $x < y$ be a brief notation for $x \leq y$ and $x \not\geq y$. So $\leq$ is transitive, reflexive, and anti-symmetric, and for any two elements $x, y \in \Omega$ we have either $x < y$, $x = y$, or $x > y$. For simplicity, we assume that Ω has a smallest element 0_Ω and a largest element 1_Ω.

We call *cumulative distribution function* any non-decreasing function $F : \Omega \to [0,1]$ that satisfies $F(1_\Omega) = 1$. $F(x)$ provides information about the cumulative probability on the interval $[0_\Omega, x]$. Note that we do not need to impose $F(0_\Omega) = 0$. Also note that cumulative distribution functions are not assumed to be right-continuous. Given a cumulative distribution F on Ω and a value $x \in \Omega$, $F(x^+)$ is the right-limit and $F(x^-)$ is the left-limit,

$$F(x^+) = \inf_{y>x} F(y) = \lim_{y \to x,\, y>x} F(y) \qquad F(x^-) = \sup_{y<x} F(y) = \lim_{y \to x,\, y<x} F(y)$$

and $F(1_\Omega^+) = 1$ and $F(0_\Omega^-) = 0$.

Definition 1. *A* generalised probability box*, or* generalised p-box*, is a pair* $(\underline{F}, \overline{F})$ *of cumulative distribution functions from* Ω *to* $[0,1]$*, satisfying* $\underline{F} \leq \overline{F}$*. If* Ω *is a closed interval on* $\mathbb{R}$*, then we call the pair* $(\underline{F}, \overline{F})$ *a* p-box.

A generalised p-box is interpreted as a lower and an upper cumulative distribution function. In Walley's framework, this means that a generalised p-box is interpreted as a lower prevision (actually a lower probability) $\underline{P}_{\underline{F},\overline{F}}$ on the set of events

$$\mathscr{K} = \{[0_\Omega, x] \colon x \in \Omega\} \cup \{(y, 1_\Omega] \colon y \in \Omega\}$$

by

$$\underline{P}_{\underline{F},\overline{F}}([0_\Omega, x]) := \underline{F}(x) \text{ and } \underline{P}_{\underline{F},\overline{F}}((y, 1_\Omega]) = 1 - \overline{F}(y).$$

In the particular case of p-boxes it was mentioned by [7, Section 4.6.6] and proven by [6, p. 93] that $\underline{P}_{\underline{F},\overline{F}}$ is coherent. It is straightforward to show that generalised p-boxes are coherent as well.

Given a generalised p-box, we can consider the set of cumulative distribution functions that lie between $\underline{F}$ and $\overline{F}$,

$$\Phi(\underline{F}, \overline{F}) = \left\{F \colon \underline{F} \leq F \leq \overline{F}\right\}.$$

We can easily express the natural extension $\underline{E}_{\underline{F},\overline{F}}$ in terms of $\Phi(\underline{F}, \overline{F})$: $\underline{E}_{\underline{F},\overline{F}}$ is the lower envelope of the natural extensions of the F between $\underline{F}$ and $\overline{F}$:

$$\underline{E}_{\underline{F},\overline{F}}(f) = \inf_{F \in \Phi(\underline{F},\overline{F})} \underline{E}_F(f) \tag{1}$$

for all gambles f on Ω. A similar result for p-boxes in the unit interval can be found in [7, Section 4.6.6].

Next, we study the natural extension of a generalised p-box, that is, what information a generalised p-box provides about the buying prices for the gambles which are not in $\mathscr{K}$. For this, we shall regularly invoke the field of events $\mathscr{H}$ generated by the domain $\mathscr{K}$, i.e., events of the type

$$[0_\Omega, x_1] \cup (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$$

for $x_1 < x_2 < x_3 < \cdots < x_{2n+1}$ in Ω (if n is 0 then this is $[0_\Omega, x_1]$) and

$$(x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$$

for $x_2 < x_3 < \cdots < x_{2n+1}$ in Ω.

Since the procedure of natural extension is transitive, in order to calculate the natural extension of $\underline{P}_{\underline{F},\overline{F}}$ to all gambles we shall first consider the extension from $\mathscr{K}$ to $\mathscr{H}$, then the natural extension from $\mathscr{H}$ to the set of all events, and finally the natural extension from the set of all events to the set of all gambles. The first of these steps is achieved by the following proposition:

Proposition 1. *Given* $A = [0_\Omega, x_1] \cup (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$,

$$\underline{E}_{\underline{F},\overline{F}}(A) = \underline{F}(x_1) + \sum_{k=1}^{n} \max\{0, \underline{F}(x_{2k+1}) - \overline{F}(x_{2k})\}$$

and given $A = (x_2, x_3] \cup \cdots \cup (x_{2n}, x_{2n+1}]$,

$$\underline{E}_{\underline{F},\overline{F}}(A) = \sum_{k=1}^{n} \max\{0, \underline{F}(x_{2k+1}) - \overline{F}(x_{2k})\}.$$

We now describe the natural extension of a generalised p-box by a Choquet integral.

4 The Natural Extension as a Choquet Integral

As shown in [4, Section 3.1], the natural extension $\underline{E}_F$ of a cumulative distribution function F on $[0,1]$ is completely monotone. It is fairly easy to generalise this result to cumulative distribution functions on a totally ordered space Ω. In this section we establish this for generalised p-boxes.

Theorem 1. *The natural extension* $\underline{E}_{\underline{F},\overline{F}}$ *of* $\underline{P}_{\underline{F},\overline{F}}$ *to* $\mathscr{L}(\Omega)$ *is given by the Choquet integral* $(C)\int \cdot \, \mathrm{d}\underline{P}^{\mathscr{H}}_{\underline{F},\overline{F}_*}$, *where* $\underline{P}^{\mathscr{H}}_{\underline{F},\overline{F}_*}$ *is the inner measure of* $\underline{P}^{\mathscr{H}}_{\underline{F},\overline{F}}$,

$$\underline{P}^{\mathscr{H}}_{\underline{F},\overline{F}_*}(A) = \sup_{C \in \mathscr{H}, C \subseteq A} \underline{P}^{\mathscr{H}}_{\underline{F},\overline{F}}(C). \tag{2}$$

Moreover, $\underline{E}_{\underline{F},\overline{F}}$ *is a completely monotone lower prevision.*

The remainder of this section is devoted to the study of this natural extension, in order to provide more manageable expressions for it. We shall characterise $\underline{E}$ by the values it takes on intervals of the form $[0_\Omega, x], (x, y], [0_\Omega, x)$ and (x, y), for $x \leq y$ in Ω, through the lower oscillation of gambles and full components of events, as explained further on. For ease of notation, we shall denote $\underline{E}_{\underline{F},\overline{F}}$ by $\underline{E}$ when no confusion is possible.

Let us consider the *upper limit topology* on Ω which is the topology generated by the base $\tau := \{(x, y] : x, y \in \Omega, x < y\} \cup \{[0_\Omega, x] : x \in \Omega\}$. For any gamble f on Ω, let us define its *lower oscillation* as the gamble

$$\underline{osc}(f)(d) := \sup_{C \in \tau:\, d \in C} \inf_{x \in C} f(x);$$

given $A \subseteq \Omega$, the lower oscillation of I_A is the indicator function of

$$B := \{d \in A : \exists C \in \tau \text{ s.t. } d \in C \subseteq A\} = \bigcup_{C \in \tau:\, C \subseteq A} C = \text{int}(A); \tag{3}$$

note that B is the union of the elements of the base τ that are included in A, and is therefore the topological interior of A in the upper limit topology. It is not too difficult to show that the lower oscillation of f is the supremum of all continuous gambles (with respect to the upper limit topology) that are dominated by f.

Lemma 1. *For any subset A of Ω, $\underline{E}(A) = \underline{E}(B)$, where B is given by Eq. (3).*

This lemma allows us to deduce the following characterisation of $\underline{E}$:

Proposition 2. *For any gamble f on Ω, $\underline{E}(f) = \underline{E}(\underline{osc}(f))$.*

This result allows us to rewrite the Choquet integral of Theorem 1 as

$$\underline{E}(f) = \inf \underline{osc}(f) + \int_{\inf \underline{osc}(f)}^{\sup \underline{osc}(f)} \underline{E}(\{\underline{osc}(f) \geq x\}) dx = \underline{E}(\underline{osc}(f)), \tag{4}$$

which is indeed more manageable. Note that for any $t \in \mathbb{R}$, $\{\underline{osc}(f) > t\}$ is equal to $\underline{osc}(\{f > t\})$, and as consequence $\underline{osc}$ is a lower semi-continuous function if we consider the upper limit topology in the initial space. Hence, the natural extension of a generalised p-box is characterised by its restriction to lower semi-continuous gambles (and, because of Eq. (4), to open sets). Taking this into account, we are going to determine the expression of the natural extension $\underline{E}$ on the subsets of Ω which are open in the upper limit topology.

Let B be an open subset of Ω, and let us show that B is a union of pairwise disjoint open intervals of Ω. Recall that by open we are referring here to the upper limit topology, so the subinterval $(a, b]$ is also open for any a, b in Ω.

Definition 2. *[5] A set S is called* full *if $[a, b] \subseteq S$ for any $a \leq b$ in S. Given a set A and an element x of A, the* full component $C(x, A)$ *of x in A is the largest full set S which satisfies $x \in S \subseteq A$.*

The full components $\{C(x, A) : x \in A\}$ of a set $A \subseteq \Omega$ form a partition of A [5, §4.4(a)]. In the following lemma, we prove that the natural extension $\underline{E}$ is additive on full components.

Lemma 2. *Let A be an arbitrary subset of Ω, and let $(A_\lambda)_{\lambda\in\Lambda}$ be the full components of A. Then $\underline{E}(\cup_{\lambda\in\Lambda}A_\lambda)=\sum_{\lambda\in\Lambda}\underline{E}(A_\lambda)$. If moreover A is open, then A_λ is open for all $\lambda\in\Lambda$.*

So the natural extension $\underline{E}$ is characterised by the value it takes on the full components of open sets. By Lemma 2, these full components are open intervals of Ω, and are therefore of the form $[0_\Omega,x],(x,y],[0_\Omega,x)$ or (x,y), for $x\leq y$ in Ω. By Proposition 1 we have that $\underline{E}([0_\Omega,x])=\underline{F}(x)$ and $\underline{E}((x,y])=\max\{0,\underline{F}(y)-\overline{F}(x)\}$ for any $y\leq x$ in Ω, and by Eq. (2),

$$\underline{E}([0_\Omega,x))=\underline{F}(x-)\quad\text{and}\quad\underline{E}(x,y)=\max\{0,\underline{F}(y-)-\overline{F}(x)\}.$$

5 Limit Approximations of the Natural Extension

Next, we give an alternative expression of the natural extension of a generalised p-box as a limit of the natural extensions of discrete p-boxes. Consider a p-box $(\underline{F},\overline{F})$ on Ω. Let $(\underline{F}_n)_n,(\overline{F}_n)_n$ be increasing and decreasing sequences of cdfs converging point-wise to $\underline{F}$ and $\overline{F}$, respectively.

For ease of notation, denote by $\underline{P}_n$ the lower probability associated with $(\underline{F}_n,\overline{F}_n)$, that is, $\underline{P}_n=\underline{P}^{\mathscr{H}}_{\underline{F}_n,\overline{F}_n}$ and let $\underline{E}_n$ be natural extension of $\underline{P}_n$. Since $\underline{F}_n\leq\underline{F}$ and $\overline{F}_n\geq\overline{F}$, it follows that $\Phi(\underline{F},\overline{F})\subseteq\Phi(\underline{F}_n,\overline{F}_n)$, and Eq. (11) implies that $\underline{E}_n\leq\underline{E}$. Moreover, the same argument implies that $\underline{E}_n\leq\underline{E}_{n+1}$ for any $n\in\mathbb{N}$, so $\lim_n\underline{E}_n=\sup_n\underline{E}_n\leq\underline{E}$. The converse holds too:

Proposition 3. *$\underline{E}(f)=\lim_n\underline{E}_n(f)$ for any gamble f.*

Next, we use this Proposition to establish an expression for the natural extension of a generalised p-box in terms of discrete p-boxes. For any natural number $n\geq1$, and $i\in\{2,\ldots,n\}$, define the sets $A_1^n:=\overline{F}^{-1}\left([0,\frac{1}{n}]\right),A_i^n:=\overline{F}^{-1}\left((\frac{i-1}{n},\frac{i}{n}]\right),B_1^n:=\underline{F}^{-1}\left([0,\frac{1}{n}]\right)$ and $B_i^n:=\underline{F}^{-1}\left((\frac{i-1}{n},\frac{i}{n}]\right)$. Clearly, both $\{A_1^n,\ldots,A_n^n\}$ and $\{B_1^n,\ldots,B_n^n\}$ are partitions of Ω. Define $\underline{F}_n$ and $\overline{F}_n$ by

$$\overline{F}_n(x)=\tfrac{i}{n}\text{ if }x\in A_i^n,\quad\underline{F}_n(x)=\begin{cases}\frac{i-1}{n} & \text{if } x\in B_i^n \text{ and } x\neq 1_\Omega,\\ 1 & \text{if } x=1_\Omega.\end{cases}\tag{5}$$

Lemma 3. *The following statements hold for all $x\in\Omega$:*

(i) For any $n\in\mathbb{N}$, $\underline{F}_n$ and $\overline{F}_n$ are cdfs, $\underline{F}_n(x)\leq\underline{F}(x)$, and $\overline{F}(x)\leq\overline{F}_n(x)$.
(ii) $\lim_n\underline{F}_n(x)=\underline{F}(x)$ and $\lim_n\overline{F}_n(x)=\overline{F}(x)$.
(iii) $(\underline{F}_{2^n})_n$, $(\overline{F}_{2^n})_n$ are increasing and decreasing sequences of cdfs such that $\underline{F}(x)=\lim_n\underline{F}_{2^n}(x)$ and $\overline{F}(x)=\lim_n\overline{F}_{2^n}(x)$.

If we can find a simple expression for the natural extension of $\underline{P}_n$ for our particular choice of $\underline{F}_n$ and $\overline{F}_n$, then we also have a simple expression for $\underline{E}_{\underline{F},\overline{F}}$ via Proposition 3. Consider $\underline{G}_1,\ldots,\underline{G}_n$ and $\overline{G}_1,\ldots,\overline{G}_n$ defined by

$$\underline{G}_i(x)=\begin{cases}1 & \text{if } \underline{F}_n(x)\geq\frac{i}{n}\\ 0 & \text{otherwise}\end{cases}\qquad\overline{G}_i(x)=\begin{cases}1 & \text{if } \overline{F}_n(x)\geq\frac{i}{n}\\ 0 & \text{otherwise}\end{cases}$$

Proposition 4. *For each* $n \in \mathbb{N}$, $\underline{E}_n = \frac{1}{n}\sum_{i=1}^{n} \underline{E}_{\underline{G}_i,\overline{G}_i}$.

Hence, all we need to characterise the natural extension of $(\underline{F}_n, \overline{F}_n)$ is to determine the natural extension of a *degenerate p-box*, i.e. one where the lower and upper cdfs only assume the values 0 and 1. Note that a degenerate p-box $(\underline{G}, \overline{G})$ is uniquely determined by

$$I_{(\underline{G},\overline{G})} = \left\{x \in \Omega : \underline{G}(x) < \overline{G}(x)\right\} = \left\{x \in \Omega : \underline{G}(x) = 0 \text{ and } \overline{G}(x) = 1\right\}.$$

Proposition 5. *Let* $(\underline{G}, \overline{G})$ *be degenerate and* $f \in \mathscr{L}(\Omega)$. *If* $0_\Omega \notin I_{(\underline{G},\overline{G})}$,

(i) If $I_{(\underline{G},\overline{G})} = (a,b)$ *then* $\underline{E}_{\underline{G},\overline{G}}(f) = \inf_{z\in(a,b]} f(z)$.
(ii) If $I_{(\underline{G},\overline{G})} = (a,b]$ *then* $\underline{E}_{\underline{G},\overline{G}}(f) = \lim_{y \gtrdot b} \inf_{z\in(a,y]} f(z)$.
(iii) If $I_{(\underline{G},\overline{G})} = [a,b)$ *then* $\underline{E}_{\underline{G},\overline{G}}(f) = \lim_{x \lessdot a} \inf_{z\in(x,b]} f(z)$.
(iv) If $I_{(\underline{G},\overline{G})} = [a,b]$ *then* $\underline{E}_{\underline{G},\overline{G}}(f) = \lim_{x \lessdot a} \lim_{y \gtrdot b} \inf_{z\in(x,y]} f(z)$.

On the other hand, if $0_\Omega \in I_{(\underline{G},\overline{G})}$, *then*

(a) If $I_{(\underline{G},\overline{G})} = [0_\Omega, b)$ *then* $\underline{E}_{\underline{G},\overline{G}}(f) = \inf_{z\in[0_\Omega,b]} f(z)$.
(b) If $I_{(\underline{G},\overline{G})} = [0_\Omega, b]$ *then* $\underline{E}_{\underline{G},\overline{G}}(f) = \lim_{y \gtrdot b} \inf_{z\in[0_\Omega,y]} f(z)$.

Concluding, if we consider now the natural extension $\underline{E}'_n$ of $(\underline{F}_{2^n}, \overline{F}_{2^n})$ as defined in Eq. (5), it follows from Proposition 3 and Lemma 3 that $(\underline{E}'_n)_n$ is an increasing sequence of functionals that converges point-wise to $\underline{E}$. By Proposition 4, $\underline{E}'_n$ can be calculated as a convex combination of natural extensions of degenerate p-boxes, whose expressions follow from Proposition 5.

6 Conclusions

We have extended results concerning p-boxes from finite to infinite sets. In particular, we have proven that the natural extension of a p-box characerizing the coherent extensions to all gambles is a completely monotone lower prevision. Such lower previsions have interesting mathematical properties—i.e., they can be written as a Rieman integral, and are determined by their values on events—and relate to comonotone additive functionals, which are of interest in economics.

A convergence result for generalised p-boxes is given in Section 5: any generalised p-box can be expressed as a limit of a sequence of discrete p-boxes. This is interesting because discrete p-boxes are more manageable in practice, and are also related to earlier works [2, 3]. In particular, they can be related to belief functions and to finitely-valued random sets. Also of interest is that natural extension is preserved when taking point-wise limits of monotone sequences of p-boxes.

Acknowledgement. Work partially supported by the project TSI2007-66706-C04-01 and by a grant from the IRSN. We are also especially grateful to Gert de Cooman and Didier Dubois for the many fruitful discussions, suggestions, and for helping us with an earlier draft of this paper.

References

1. De Cooman, G., Troffaes, M.C.M., Miranda, E.: n-Monotone exact functionals (submitted for publication) (2008)
2. Destercke, S., Dubois, D., Chojnacki, E.: Relating imprecise representations of imprecise probabilities. In: Proceedings of 5th International Symposium on Imprecise Probability: Theories and Applications, pp. 155–164 (2007)
3. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories (2003)
4. Miranda, E., de Cooman, G., Quaeghebeur, E.: Finitely additive extensions of distribution functions and moment sequences: the coherent lower prevision approach. Internat. J. Approx. Reason 48(1), 132–155 (2008)
5. Schechter, E.: Handbook of Analysis and Its Foundations. Academic Press, San Diego (1997)
6. Troffaes, M.C.M.: Optimality, Uncertainty, and Dynamic Programming with Lower Previsions. PhD thesis, Ghent University, Ghent, Belgium (2005)
7. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

The F. Riesz Representation Theorem and Finite Additivity

Gert de Cooman[1] and Enrique Miranda[2]

[1] SYSTeMS Research Group, Ghent University, Zwijnaarde, Belgium
[2] Dept. of Statistics and Operations Research, Rey Juan Carlos University, Móstoles, Spain

Abstract. A positive and normalised real linear functional on the set of bounded continuous functions can be characterised as the integral of a σ-additive probability measure, by the F. Riesz Representation Theorem. In this paper, we look at the finitely additive extensions of such a functional to the set of all bounded random variables, and prove that they are determined by Riesz' extension to lower semi-continuous functions. In doing so, we establish links with Daniell's approach to integration, Walley's theory of coherent lower previsions, and de Finetti's Representation Theorem for exchangeable random variables.

Keywords: F. Riesz representation theorem, Lower semi-continuity, Coherent lower prevision, Natural extension, I-integral, Exchangeability.

1 Introduction

Let K be any compact metric space, and consider the linear space $\mathscr{L}(K)$ of all bounded real-valued maps on K. We provide this set with the topology of uniform convergence, which turns $\mathscr{L}(K)$ into a Banach space.

We call *gamble* any bounded real function on K, and *linear prevision* any positive, normalised (i.e., with operator norm 1) real linear functional on a linear subspace of $\mathscr{L}(K)$ that contains the constant gambles. We explain our reasons for this and other terminology in Sect. 3, which is intended to give background information and further discussion of the importance of the problem addressed here.

To set the stage, consider a positive, normalised real linear functional π on the set $\mathscr{C}(K)$ of all continuous bounded real functions on K. The F. Riesz Representation Theorem [14, Theorem 2.22] tells us that there is a *unique* (σ-additive) probability measure μ_π on the Borel sets of K such that for all continuous gambles f

$$\pi(f) = (L)\int f \,\mathrm{d}\mu_\pi,$$

where the integral is a Lebesgue integral associated with the probability measure μ_π. In other words, the linear prevision π on $\mathscr{C}(K)$ extends *uniquely* to a linear prevision L_π to the linear space $\mathscr{B}(K)$ of all Borel-measurable gambles on K that furthermore satisfies the *monotone convergence* requirement: if the increasing sequence of gambles $f_n, n \geq 0$ converges point-wise to a gamble f, then $L_\pi(f_n) \to L_\pi(f)$. The original linear prevision π on $\mathscr{C}(K)$ satisfies this extra monotone convergence condition automatically: any increasing sequence of continuous gambles that converges point-wise to some continuous

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 243–252, 2008.
springerlink.com

gamble, also converges uniformly to that gamble, by Dini's Convergence Theorem [18, Sect. 17.7], and any linear prevision is continuous with respect to uniform convergence [20, Sect. 2.6].

If we drop the monotone convergence requirement, then by the Hahn–Banach Theorem [18, Sect. 12.31], π can be extended to a linear prevision P on the linear space $\mathscr{L}(K)$ of *all* gambles, not just the Borel-measurable ones. But uniqueness is no longer guaranteed, and what is more, it seems that no manner for actually *constructing* such extensions can be given; see [18, Sect. 6.6] for a discussion of constructibility. Let us denote by $\mathscr{M}(\pi)$ the set of all linear previsions on $\mathscr{L}(K)$ that extend π, i.e., coincide with π on $\mathscr{C}(K)$.

Now consider the lower envelope $\underline{E}_\pi$ of this set $\mathscr{M}(\pi)$, also called the *natural extension* of π, which is defined on all gambles f by

$$\underline{E}_\pi(f) = \inf\{P(f)\colon P \in \mathscr{M}(\pi)\}.$$

This is a real functional on the linear space $\mathscr{L}(K)$ that is (C1) *super-additive*, meaning that $\underline{E}_\pi(f+g) \geq \underline{E}_\pi(f) + \underline{E}_\pi(g)$; (C2) *positively homogeneous*, meaning that $\underline{E}_\pi(\lambda f) = \lambda \underline{E}_\pi(f)$ for all real $\lambda \geq 0$; and (C3) *positive*, meaning that $\underline{E}_\pi(f) \geq \inf f$. In fact, it is the point-wise smallest such functional that extends π, and any linear prevision P extends π if and only if it point-wise dominates $\underline{E}_\pi$; this follows immediately from Theorems 2.5.5, 3.1.2, 3.3.3 and 3.4.1 in [20]. Moreover, for any gamble f and any real number a, there is a linear prevision P on $\mathscr{L}(K)$ that extends π such that $P(f) = a$ if and only if $a \in [\underline{E}_\pi(f), \overline{E}_\pi(f)]$, where $\overline{E}_\pi$ is the *conjugate* functional of $\underline{E}_\pi$, defined by $\overline{E}_\pi(g) = -\underline{E}_\pi(-g)$ for all gambles g. This is the essence of Bruno de Finetti's Fundamental Theorem of Probability [9, Vol. 1, Sect. 3.10], and a special case of [20, Corollary 3.4.3]. In this sense, the natural extension $\underline{E}_\pi$ characterises all the linear previsions that extend π.

But what makes this natural extension $\underline{E}_\pi$ especially interesting, is that it can be constructed explicitly. Indeed, it can be shown, invoking a general result by Walley [20, Theorem 3.1.4], that $\underline{E}_\pi$ coincides with the *inner extension* of π:

$$\underline{E}_\pi(f) = \sup\{\pi(g)\colon g \in \mathscr{C}(K) \text{ and } g \leq f\}. \tag{1}$$

In this paper, and in particular in Sect. 2, we intend to show that $\underline{E}_\pi$ is completely determined by the probability measure μ_π in a very specific way. Indeed, let us define the *lower oscillation* $\underline{\operatorname{osc}}(f)$ of f as the gamble on K that assumes the value

$$\underline{\operatorname{osc}}_x(f) = \sup_{N \in \mathscr{N}_x} \inf_{z \in N} f(z) \tag{2}$$

in any element x of K, where $\mathscr{N}_x$ is the set (filter) of all neighbourhoods of x. Then we prove in Theorem 1 that for all gambles f on K,

$$\underline{E}_\pi(f) = (L)\int \underline{\operatorname{osc}}(f)\,\mathrm{d}\mu_\pi. \tag{3}$$

Although we haven't come across Eq. (3) in the literature, to us the contribution of this paper does not lie in its mathematical derivation: it is proved easily enough using basic functional-analytic results. Rather, we want to draw attention, in Sect. 3, to the beauty that lies hidden in its interpretation, and in the connections it provides between fundamental results in probability theory and functional analysis.

2 Derivation

The functional $\underline{E}_\pi$ is completely determined by its values on the set $\underline{\mathscr{C}}(K)$ of lower semi-continuous gambles on K, and $\underline{\text{osc}}$ has a nice topological interpretation.

Proposition 1. *For any gamble f on K, $\underline{\text{osc}}(f)$ is the point-wise greatest lower semi-continuous gamble that is point-wise dominated by f, and $\underline{E}_\pi(f) = \underline{E}_\pi(\underline{\text{osc}}(f))$.*

Proof. It is clear from Eq. (2) that $\underline{\text{osc}}(f) \in \underline{\mathscr{C}}(K)$ and $\underline{\text{osc}}(f) \leq f$ for any gamble f. Given any $g \in \underline{\mathscr{C}}(K)$, then for any real t, $\{g > t\}$ is open, meaning that for any $x \in \{g > t\}$, there is some $N \in \mathscr{N}_x$ such that $N \subseteq \{g > t\}$. Consequently, $g(x) > t$ implies that $\underline{\text{osc}}_x(g) = \sup_{N \in \mathscr{N}_x} \inf_{z \in N} g(z) \geq t$, whence $\underline{\text{osc}}(g) \geq g$, and therefore $\underline{\text{osc}}(g) = g$ for any $g \in \underline{\mathscr{C}}(K)$. Now let $g \in \underline{\mathscr{C}}(K)$ be such that $g \leq f$. Then it follows from the definition of the lower oscillation that $\underline{\text{osc}}(g) \leq \underline{\text{osc}}(f)$, whence $g \leq \underline{\text{osc}}(f)$.

For the second part, note that for any $g \in \mathscr{C}(K)$ such that $g \leq f$ it follows that $g = \underline{\text{osc}}(g) \leq \underline{\text{osc}}(f) \leq f$, and since $\underline{E}_\pi$ is monotone as a lower envelope of (monotone) linear previsions, $\pi(g) = \underline{E}_\pi(g) \leq \underline{E}_\pi(\underline{\text{osc}}(f)) \leq \underline{E}_\pi(f)$. By (11), $\underline{E}_\pi(f) \leq \underline{E}_\pi(\underline{\text{osc}}(f)) \leq \underline{E}_\pi(f)$. □

$\underline{E}_\pi$ has continuity properties that are stronger than the mere continuity with respect to the topology of uniform convergence on $\mathscr{L}(K)$ that is implied by its definition as a lower envelope of (uniformly) continuous linear previsions.[1]

Proposition 2. *Let $(f_n)_{n\geq 0}$ be a non-decreasing sequence of non-negative elements of $\underline{\mathscr{C}}(K)$, such that for all x in K, $f(x) := \lim_{n\to\infty} f_n(x)$ is a real number, and such that the function f thus defined is bounded. Then $\underline{E}_\pi(f) = \lim_{n\to\infty} \underline{E}_\pi(f_n)$. Similarly, for any sequence $(B_n)_{n\geq 0}$ of open subsets of K, $\underline{E}_\pi(\bigcup_{n\geq 0} B_n) = \lim_{n\to\infty} \underline{E}_\pi(\bigcup_{i=1}^n B_i)$.*

Proof. It suffices to prove the first part. Fix $\varepsilon > 0$. Then it follows from (11) that there is $f_\varepsilon \in \mathscr{C}(K)$ such that $f_\varepsilon \leq f$ and $\underline{E}_\pi(f) - \pi(f_\varepsilon) < \frac{\varepsilon}{2}$. Consider, for any $n \geq 0$, the gamble $f_{\varepsilon,n} := \min\{f_\varepsilon, f_n\}$ on K. The sequence $f_{\varepsilon,n}$, $n \geq 0$ satisfies the following properties: (i) it converges point-wise to f_ε: indeed, for any $x \in K$,

$$\lim_{n\to\infty} \min\{f_\varepsilon(x), f_n(x)\} = \min\{f_\varepsilon(x), \lim_{n\to\infty} f_n(x)\} = \min\{f_\varepsilon(x), f(x)\} = f_\varepsilon(x);$$

(ii) $f_{\varepsilon,n} \in \underline{\mathscr{C}}(K)$ for any n, since it is the point-wise minimum of two elements of $\underline{\mathscr{C}}(K)$; and (iii) $f_{\varepsilon,n}$ is non-decreasing: $f_{\varepsilon,n+1}(x) \geq f_{\varepsilon,n}(x)$ for all $x \in K$ and all $n \geq 0$. By Dini's Convergence Theorem [18, Sect. 17.7], $(f_{\varepsilon,n})_{n\geq 0}$ converges uniformly to f_ε. Since $\underline{E}_\pi$ is continuous with respect to the uniform convergence and monotone, $\lim_{n\to\infty} \underline{E}_\pi(f_{\varepsilon,n}) = \sup_{n\geq 0} \underline{E}_\pi(f_{\varepsilon,n}) = \underline{E}_\pi(f_\varepsilon) = \pi(f_\varepsilon)$, whence there is $n_\varepsilon \geq 0$ such that $\pi(f_\varepsilon) - \underline{E}_\pi(f_{\varepsilon,n}) < \frac{\varepsilon}{2}$ for all $n \geq n_\varepsilon$. Since also $f_{\varepsilon,n} \leq f_n$ for all $n \geq 0$, we deduce that $\underline{E}_\pi(f_{\varepsilon,n}) \leq \underline{E}_\pi(f_n)$ for all $n \geq 0$, and so for all $n \geq n_\varepsilon$,

$$\underline{E}_\pi(f) - \underline{E}_\pi(f_n) = \underline{E}_\pi(f) - \pi(f_\varepsilon) + \pi(f_\varepsilon) - \underline{E}_\pi(f_{\varepsilon,n}) + \underline{E}_\pi(f_{\varepsilon,n}) - \underline{E}_\pi(f_n) < \varepsilon.$$

Hence $\underline{E}_\pi(f) = \sup_{n\geq 0} \underline{E}_\pi(f_n) = \lim_{n\to\infty} \underline{E}_\pi(f_n)$. □

[1] Here and in other places in this paper we use the identification between sets and their indicator functions, so $\underline{E}_\pi(A)$ refers to the value that $\underline{E}_\pi$ takes in the indicator I_A of the set A.

Theorem 1. *For any gamble f on K, $\underline{E}_\pi(f) = (L)\int \underline{\operatorname{osc}}(f)\,\mathrm{d}\mu_\pi$.*

Proof. By Proposition 1, it suffices to prove that for any $f \in \underline{\mathscr{C}}(K)$, $\underline{E}_\pi(f) = (L)\int f\,\mathrm{d}\mu_\pi$. [Observe that since f is lower semi-continuous, its cut sets $\{f > t\}$ are open for any real t, so f is Borel-measurable]. Since K is a metric space, any such f is the point-wise limit of a non-decreasing sequence of continuous gambles $(g_n)_{n\geq 0}$ [18, Theorem 16.16]. Assume first that $f \geq 0$. Then we may also assume that $g_n \geq 0$ for all n. Applying Proposition 2,

$$\underline{E}_\pi(f) = \lim_{n\to\infty} \underline{E}_\pi(g_n) = \lim_{n\to\infty} (L)\int g_n\,\mathrm{d}\mu_\pi = (L)\int \lim_{n\to\infty} g_n\,\mathrm{d}\mu_\pi = (L)\int f\,\mathrm{d}\mu_\pi,$$

also using the monotone convergence of the Lebesgue integral L_π associated with μ_π. Finally consider any $g \in \underline{\mathscr{C}}(K)$, and define $f := g - \inf g$. Then $f \geq 0$ also belongs to $\underline{\mathscr{C}}(K)$, so $\underline{E}_\pi(f) = (L)\int f\,\mathrm{d}\mu_\pi$. But the properties of lower envelopes of linear previsions and of the Lebesgue integral guarantee that

$$\underline{E}_\pi(g) = \underline{E}_\pi(f + \inf g) = \underline{E}_\pi(f) + \inf g = (L)\int f\,\mathrm{d}\mu_\pi + \inf g = (L)\int g\,\mathrm{d}\mu_\pi. \qquad \square$$

Corollary 1. *$\underline{E}_\pi$ is additive and satisfies monotone convergence on the convex cone $\underline{\mathscr{C}}(K)$, where it coincides with the linear functional L_π.*

This can also be derived using Eq. (1). Of further interest is the uniformly closed linear lattice $\mathscr{I}_\pi(K) = \{f \in \mathscr{L}(K) : \underline{E}_\pi(f) = \overline{E}_\pi(f)\}$ of all gambles to which the linear prevision π can be extended uniquely as a linear prevision. We call its elements *π-integrable*. We can characterise the π-integrability of a gamble by looking at its *oscillation* $\operatorname{osc}(f)$, defined by $\operatorname{osc}(f) := \overline{\operatorname{osc}}(f) - \underline{\operatorname{osc}}(f) = -\underline{\operatorname{osc}}(-f) - \underline{\operatorname{osc}}(f)$, i.e., by $\operatorname{osc}_x(f) = \inf_{N\in\mathscr{N}_x} \sup_{z_1,z_2\in N} |f(z_2) - f(z_1)|$ for all x in K. Observe that a gamble f is continuous at x if and only if $\operatorname{osc}_x(f) = 0$.

Corollary 2. *A gamble f on K belongs to $\mathscr{I}_\pi(K)$ if and only if its oscillation $\operatorname{osc}(f)$ is zero almost everywhere $[\mu_\pi]$, i.e., if f is continuous almost everywhere $[\mu_\pi]$.*

Proof. It is clear that f belongs to $\mathscr{I}_\pi(K)$ if and only if

$$0 = \overline{E}_\pi(f) - \underline{E}_\pi(f) = (L)\int [\overline{\operatorname{osc}}(f) - \underline{\operatorname{osc}}(f)]\,\mathrm{d}\mu_\pi = (L)\int f\,\mathrm{d}\mu_\pi. \qquad \square$$

In accordance with Choquet's [3] general definition for maps from an Abelian semigroup to an Abelian group, we can call a real functional Γ defined on a lattice of gambles $\mathscr{K}$ *n-monotone* if $\sum_{I\subseteq\{1,\ldots,p\}} (-1)^{|I|}\Gamma(f \wedge \bigwedge_{i\in I} f_i) \geq 0$, for all $p \leq n$, and all f, f_1, $\ldots$, f_p in $\mathscr{K}$. In this expression $|I|$ denotes the cardinality of a finite set I, and $\wedge$ denotes point-wise minimum. A real functional that is n-monotone for all natural numbers n is called *completely monotone*. We have shown elsewhere [6] that a linear prevision is always completely monotone, and that natural extension preserves complete monotonicity. This implies [6] that both π and $\underline{E}_\pi$ are completely monotone, and that for any $f \in \mathscr{L}(K)$,

$$\underline{E}_\pi(f) = (C)\int f\,\mathrm{d}\underline{E}_\pi := \inf f + \int_{\inf f}^{\sup f} \underline{E}_\pi(\{f \geq t\})\,\mathrm{d}t, \tag{4}$$

where the first integral is a *Choquet integral* [10], and the integral in its defining expression is a Riemann integral. As a consequence, a gamble f is π-integrable if and only if its cut sets $\{f \geq t\}$ are, for all but a countable number of t.

The restriction of $\underline{E}_\pi$ to $\mathscr{I}_\pi(K)$ is a linear prevision, which we denote by E_π. We easily derive from Eq. (1) that $\underline{E}_\pi$ is the inner extension of E_π, meaning that

$$\underline{E}_\pi(f) = \sup\{E_\pi(g)\colon g \leq f \text{ and } g \in \mathscr{I}_\pi(K)\}$$

for all $f \in \mathscr{L}(K)$. This means that $\underline{E}_\pi$ is completely determined by its values on π-integrable gambles. We mention an even stronger result: $\underline{E}_\pi$ is actually completely determined by the values that it assumes on finite unions of π-integrable open balls. A proof uses the following ingredients: (i) Eq. (4) together with Proposition 1 implies that $\underline{E}_\pi$ is uniquely determined by its restriction to the open subsets of K; (ii) because K is compact, any open set is a countable union of open balls; (iii) Proposition 2 tells us that $\underline{E}_\pi$ on open sets is therefore determined by its values on finite unions of open balls; and (iv) all but a countable number of open balls centred on any point x of K are π-integrable.

Corollary 3. *Let $\mathscr{B} = \{B_i\colon i \in I\}$ be a finite or countably infinite partition of K, consisting of Borel measurable π-integrable events. Then $\underline{E}_\pi(f) = \sum_{i\in I} \underline{E}_\pi(fI_{B_i})$ for any $f \in \mathscr{L}(K)$.*

Proof. First assume that $f \geq 0$. Then it follows from Theorem 1 that

$$\underline{E}_\pi(f) = (L)\int \underline{\mathrm{osc}}(f)\,\mathrm{d}\mu_\pi = (L)\int \sum_{i\in I} \underline{\mathrm{osc}}(f)I_{B_i}\,\mathrm{d}\mu_\pi = \sum_{i\in I}(L)\int \underline{\mathrm{osc}}(f)I_{B_i}\,\mathrm{d}\mu_\pi,$$

since each $\underline{\mathrm{osc}}(f)I_{B_i}$ is Borel measurable. Consider the topological interior $\mathrm{int}(B_i)$ and closure $\mathrm{cl}(B_i)$ of a set B_i. Since $\underline{\mathrm{osc}}(I_{B_i}) = I_{\mathrm{int}(B_i)}$, Proposition 1 implies that $\underline{E}_\pi(\mathrm{int}(B_i)) = \underline{E}_\pi(B_i) = \mu_\pi(B_i)$ and similarly $\overline{E}_\pi(\mathrm{cl}(B_i)) = \overline{E}_\pi(B_i) = \mu_\pi(B_i)$. But then we find for any Borel measurable set A such that $\mathrm{int}(B_i) \subseteq A \subseteq \mathrm{cl}(B_i)$ that

$$\mu_\pi(B_i) = \underline{E}_\pi(\mathrm{int}(B_i)) \leq \underline{E}_\pi(A) \leq \mu_\pi(A) \leq \overline{E}_\pi(A) \leq \overline{E}_\pi(\mathrm{cl}(B_i)) = \mu_\pi(B_i),$$

so $\mu_\pi(B_i) = \mu_\pi(A)$. In particular $\mu_\pi(B_i) = \mu_\pi(\mathrm{int}(B_i))$, and therefore

$$(L)\int \underline{\mathrm{osc}}(f)I_{B_i}\,\mathrm{d}\mu_\pi = (L)\int \underline{\mathrm{osc}}(f)I_{\mathrm{int}(B_i)}\,\mathrm{d}\mu_\pi = (L)\int \underline{\mathrm{osc}}(fI_{B_i})\,\mathrm{d}\mu_\pi = \underline{E}_\pi(fI_{B_i}),$$

since $\underline{\mathrm{osc}}(fI_{B_i}) = \underline{\mathrm{osc}}(f)I_{\mathrm{int}(B_i)}$ because $f \geq 0$. This means that the desired equality holds for non-negative gambles. Now for a general gamble g, let $f := g - \inf g$, then $f \geq 0$ and $g = f + \inf g$. Then since $\underline{E}_\pi$ is a lower envelope of linear previsions and $\inf g$ is a constant, $\underline{E}_\pi(g) = \underline{E}_\pi(f + \inf g) = \underline{E}_\pi(f) + \inf g$, and also $\underline{E}_\pi(fI_{B_i}) + \mu_\pi(B_i)\inf g \leq \underline{E}_\pi([f + \inf g]I_{B_i}) = \underline{E}_\pi(fI_{B_i}) + \mu_\pi(B_i)\inf g$, whence $\underline{E}_\pi(gI_{B_i}) = \underline{E}_\pi(fI_{B_i}) + \mu_\pi(B_i)\inf g$. Now using the result already proved for $f \geq 0$:

$$\sum_{i\in I}\underline{E}_\pi(gI_{B_i}) = \sum_{i\in I}\underline{E}_\pi(fI_{B_i}) + \inf g \sum_{i\in I}\mu_\pi(B_i) = \underline{E}_\pi(f) + \inf g. \qquad \square$$

3 Comments on Interpretation and Historical Background

De Finetti's previsions. Most of the terminology that we have been using, is borrowed from Walley's behavioural approach to decision making and probability [20], which has its roots in the work of Ramsey [17] and de Finetti [7, 9]. When a subject is uncertain about the actual value that a variable X assumes in a set of possible values $\mathscr{X}$, we can try and model his beliefs about this value by asking him whether he accepts to engage in certain risky transactions, called *gambles*, whose outcome depends on the actual value of X. Mathematically, a gamble is a bounded real-valued function on $\mathscr{X}$, and if a subject accepts a gamble f, this means that he accepts the transaction in which the value x of X is determined, and where he then receives the (possibly negative) amount of utility $f(x)$.

De Finetti [7, 9] defined the *fair price*, or the *prevision* of a gamble f as the unique fixed number $P(f)$ that to the subject is equivalent to the uncertain number f, i.e., such that the subject accepts the gambles $f-s$ (buying f for price s) and $t-f$ (selling f for price t) for all real numbers $s<P(f)<t$. If a subject gives fair prices $P(f)$ for a number of gambles f in some set $\mathscr{K}$, then this amounts to specifying a real functional P on $\mathscr{K}$. A subject is rational if he *avoids a sure loss* in specifying P, or in other words, if no Dutch book can be made against him. De Finetti showed [7, 9] that this requirement is equivalent to P being extendable to some positive and normalised real linear functional—a linear prevision—Q on all gambles.

Such linear previsions Q are mathematically equivalent to finitely additive probability measures: the restriction of Q to (indicators of) events is a finitely additive probability measure, and conversely, if we start with a finitely additive probability measure defined on all events, then it extends uniquely to a linear prevision on all gambles [2, Sect. 4.7]. De Finetti used his previsions to give a subjectivistic foundation to the theory of probability.

Walley's lower previsions. In de Finetti's definition of a fair price for a gamble f, it is implicitly assumed that a subject is always able to specify such a fair price, or in other words, is able to choose, for nearly every real price p, between buying f for price p (accepting $f-p$) and selling f for that price (accepting $p-f$). Arguably [20, 22] this may be asking too much of a subject's dispositions. For this reason Walley [20] (and Smith [19] and Williams [21, 22] before him) distinguish between the *lower prevision* $\underline{P}(f)$ of f, which is the supremum s such that the subject accepts $f-s$ (buying f for the price s), and the *upper prevision* $\overline{P}(f)$ of f, which is the infimum t such that the subject accepts $t-f$ (selling f for the price t). The lower and upper previsions of a subject are conjugate functionals, in the sense that $\overline{P}(f)=-\underline{P}(-f)$, and so if we establish the lower prevision for all gambles in some domain $\mathscr{K}$, we can derive immediately the upper prevision for all the gambles in $-\mathscr{K}$. When the gambles in the domain are indicators of events, our supremum buying prices for the indicator of an event are actually supremum betting rates on the event, and the infimum selling prices are one minus supremum betting rates for betting against the event. We then talk of *lower* and *upper probabilities*, respectively.

If a subject establishes a lower prevision $\underline{P}$ on some set of gambles $\mathscr{K}$, we say that he is rational when he cannot be made subject to a sure loss, and when furthermore

the supremum buying prices for any gamble in $\mathscr{K}$ cannot be raised by considering the implications of the supremum buying prices for finite collections of other gambles in $\mathscr{K}$. We say then that the lower prevision $\underline{P}$ is *coherent*. Walley [20] has shown that this is equivalent to $\underline{P}$ being extendable to a real functional on all gambles that is super-additive, positively homogeneous and positive; see the conditions (C1)–(C3) in the Introduction. For a coherent lower prevision $\underline{P}$ and its conjugate upper prevision $\overline{P}$ it holds that $\underline{P}(f) \leq \overline{P}(f)$ for all $f \in \mathscr{K} \cap -\mathscr{K}$. In the particular case that our subject establishes fair prices for all gambles f in $\mathscr{K}$, meaning that $\mathscr{K} = -\mathscr{K}$ and $\underline{P}(f) = \overline{P}(f)$ for all $f \in \mathscr{K}$, the coherence requirement is equivalent to de Finetti's notion of avoiding sure loss from the previous section. Moreover, a coherent lower prevision is always a lower envelope of linear previsions, and any such lower envelope is a coherent lower prevision.

Given a coherent lower prevision on some set of gambles $\mathscr{K}$, we can consider the smallest coherent lower prevision $\underline{E}_{\underline{P}}$ on the set of all gambles that coincides with $\underline{P}$ on its domain. It is called the *natural extension* of $\underline{P}$, and provides, for any gamble f, the supremum acceptable buying price $\underline{E}_{\underline{P}}(f)$ for f that can be derived from the supremum buying prices established for the gambles in $\mathscr{K}$, using coherence.

We immediately see that if a subject has in some way specified a (linear) prevision π on the set $\mathscr{C}(K)$ of all continuous gambles, then $\underline{E}_\pi$ represents the behavioural consequences of this assessment for *all* gambles f on K: $\underline{E}_\pi(f)$ is the supremum buying price for f that can be deduced from finite combinations of the assessments $\pi(g)$, $g \in \mathscr{C}(K)$. Indeed, consider the assessments $\pi(g)$ for any $g \leq f$. They mean in particular that our subject is willing to buy g for any price $p < \pi(g)$, i.e., willing to accept the uncertain transaction $g - p$. But then he should also be willing to accept the transaction $f - p \geq g - p$, because the resulting reward can never be lower. So we see that our subject should be willing to buy f for any price p such that there is a continuous gamble g for which $p < \pi(f)$. By Eq. (1), $\underline{E}_\pi(f)$ is the supremum of all such inferred acceptable buying prices p.

Lower oscillation as a probability model. Consider a lower probability that assumes only the values zero and one, so there is a set $\mathscr{F}$ of events on which the lower probability is one, meaning that the subject is practically certain (because prepared to bet at all odds on the fact) that these events occur. Then[2] this lower probability is coherent if and only if $\mathscr{F}$ is a proper set filter (i.e., a proper subset of the power set $\wp(\mathscr{X})$ that is closed under finite intersections, and increasing), and that it then has a *unique extension* to a coherent lower prevision on all gambles, given by

$$\underline{P}_{\mathscr{F}}(f) = \sup_{F \in \mathscr{F}} \inf_{z \in F} f(z).$$

Compare this to our definition (2) of the lower oscillation $\underline{\mathrm{osc}}_x(f)$ of a gamble f in an element x: we see that $\underline{\mathrm{osc}}_x$ is the unique coherent lower prevision associated with assigning (lower and therefore upper) probability one to any element of the neighbourhood filter $\mathscr{N}_x$, i.e., to any neighbourhood of x. In other words, $\underline{\mathrm{osc}}_x$ is the probability model that corresponds to the statement: *all probability mass is concentrated in any neighbourhood of* x, i.e., lies arbitrarily close to x. And Theorem 1 tells us that the

[2] See Sects. 2.9.8 and 3.2.6, as well as Endnote 4 on p. 502, in [20].

natural extension $\underline{E}_\pi$ is a σ-additive convex mixture of these elementary 'concentrated mass models' $\underline{\text{osc}}_x$, with weights characterised by μ_π.

Complete monotonicity and Choquet's Representation Theorem. There is yet another intriguing interpretation for Theorem 1, which is related to the notion of complete monotonicity, introduced in Sect. 2. It is a consequence of Choquet's Representation Theorem [3, Sect. 45] that any coherent and completely monotone lower prevision on $\mathscr{L}(K)$ can be written as a 'σ-additive convex mixture' of the extreme points of the set of all coherent and completely monotone lower previsions. But, as Choquet [3, Sect. 43.7] has essentially shown, the extreme coherent and completely monotone lower previsions are precisely the lower previsions $\underline{P}_{\mathscr{F}}$ associated with proper set filters $\mathscr{F}$. If we rewrite Theorem 1 as $\underline{E}_\pi(f) = (L)\int \underline{P}_{\mathscr{N}}(f)\,\mathrm{d}\mu_\pi$, we see that for the completely monotone and coherent natural extension $\underline{E}_\pi$ we can actually identify the 'σ-additive convex mixture' and the extreme points that participate in it: the mixture is precisely the one associated with the unique σ-probability measure μ_π induced by the linear prevision π, and the extreme points are the lower previsions $\underline{\text{osc}}_x$ associated with the neighbourhood filters $\mathscr{N}_x$, $x \in K$.

Daniell's approach to integration. There is an interesting connection between our results and Daniell's [4] treatment of the extension problem of I-integrals. Daniell's notion of I-integral refers to a real linear functional defined on a linear lattice of gambles $\mathscr{K}$, which is moreover continuous for monotone sequences of non-negative gambles decreasing to 0. Given such a functional, Daniell shows that it can be extended as an I-integral to the set of those functions which are limits of an increasing sequence of gambles in $\mathscr{K}$ by using a property of monotone convergence. From there, he considers inner and outer extensions to the set of all gambles.

In this paper, we started out with a linear prevision π on the set $\mathscr{C}(K)$ of continuous gambles on K. By the F. Riesz Representation Theorem, this linear prevision is the Lebesgue integral with respect to some σ-additive probability measure μ_π, and as such it satisfies monotone convergence. Hence, the linear prevision π is a particular instance of an I-integral. Using Daniell's results, we can extend it to an I-integral to the set of gambles which are limits of an increasing sequence of continuous gambles, i.e., to the set $\underline{\mathscr{C}}(K)$ of lower semi-continuous gambles on K.

What we have proved above in Corollary 1 is that Daniell's extension procedure coincides with Walley's notion of natural extension, at least as far as extension to the lower semi-continuous gambles is concerned. Moreover, the natural extension to all gambles coincides with the lower I-integral we would obtain on Daniell's approach.

We want to stress here that the linear prevision π does not have a *unique* extension to a linear prevision on the set of all lower semi-continuous gambles. In fact, we have already remarked that for any $a \in [\underline{E}_\pi(f), \overline{E}_\pi(f)]$, there is a linear prevision Q extending π to the set of all gambles and satisfying $Q(f) = a$. This is because we do not require monotone convergence for our finitely additive extensions. The surprising result is then that the lower envelope of all these finitely additive extensions to the set of lower semi-continuous gambles coincides with the *only* extension satisfying the monotone convergence, which is the Lebesgue integral with respect to the σ-additive

probability measure determined by Riesz' Representation Theorem, and which is also Daniell's extension as an I-integral.

De Finetti's Representation Theorem. A sequence of random variables $X_1, \ldots, X_n, \ldots$, all assuming values in the same finite set $\mathscr{X}$ is called *exchangeable* if for all natural numbers n, the mass function $p^n(x_1,\ldots,x_n)$ for the first n variables is invariant under permutation of its arguments, i.e., $p^n(x_1,\ldots,x_n) = p^n(x_{\pi(1)},\ldots,x_{\pi(n)})$ for all $(x_1,\ldots,x_n) \in \mathscr{X}^n$ and all permutations π of $\{1,\ldots,n\}$. For such exchangeable sequences, de Finetti [7] has proved a representation theorem that is considered to be of fundamental importance to Bayesian statistics.[3] Although it is usually formulated differently (see for instance [15, Theorem 1.1] for a fairly abstract, and [13] for a simple formulation), the essence of de Finetti's result can be stated as follows [5]:

The sequence X_n, $n \geq 1$ of random variables in $\mathscr{X}$ is exchangeable if and only if there is some linear prevision π on the set $\mathscr{C}(\Sigma_{\mathscr{X}})$ of all continuous gambles of the $\mathscr{X}$-simplex $\Sigma_{\mathscr{X}} = \left\{\boldsymbol{\theta} \in \mathbb{R}^{\mathscr{X}} : (\forall x \in \mathscr{X})(\theta_x \geq 0) \text{ and } \sum_{x\in\mathscr{X}} \theta_x = 1\right\}$ such that $p^n(\mathbf{x}) = \pi(B_{\mathbf{x}})$ for all $n \geq 1$ and $\mathbf{x} = (x_1,\ldots,x_n) \in \mathscr{X}^n$, where

$$B_{\mathbf{x}}(\boldsymbol{\theta}) = \frac{n!}{\prod_{z\in\mathscr{X}} T_z(\mathbf{x})!} \prod_{z\in\mathscr{X}} \theta_z^{T_z(\mathbf{x})} \text{ and } T_z(\mathbf{x}) = |\{k \in \{1,\ldots,n\} : x_k = z\}|.$$

The joint mass functions of any exchangeable sequence only determine a linear prevision π on the linear space of all polynomials on the compact metric space $\Sigma_{\mathscr{X}}$, or, what is equivalent, on the linear space $\mathscr{C}(\Sigma_{\mathscr{X}})$. It is only if σ-additivity (or equivalently monotone convergence) is required that the π leads to a unique probability measure μ_π, which can be defined on the Borel sets of $\Sigma_{\mathscr{X}}$. But it is well-known that de Finetti himself was strongly opposed to imposing σ-additivity as a general normative axiom for probability models; see [9, Vol. 1, Sect. 6.3]. If we therefore only require finite additivity, there are an infinity of linear previsions (or equivalently, finitely additive probability measures) that extend π to $\mathscr{L}(\Sigma_{\mathscr{X}})$. These are completely characterised by their lower envelope $\underline{E}_\pi$.

Acknowledgement. Research supported by research grant G.0139.01 of the Flemish Fund for Scientific Research (FWO), and by project TSI2007-66706-C04-01.

References

1. Bernardo, J.M., Smith, A.F.M.: Bayesian Theory. John Wiley & Sons, Chichester (1994)
2. Bhaskara Rao, K.P.S., Bhaskara Rao, M.: Theory of Charges. Academic Press, London (1983)
3. Choquet, G.: Theory of capacities. Ann. Inst. Fourier (Grenoble) 5, 131–295 (1953-1954)
4. Daniell, P.J.: A general form of integral. Ann. of Math (2) 19(4), 279–294 (1918)
5. De Cooman, G., Miranda, E., Quaeghebeur, E.: Exchangeable lower previsions (submitted for publication, 2008) (ArXiv:0801.1265v1)
6. De Cooman, G., Troffaes, M.C.M., Miranda, E.: n-Monotone exact functionals (submitted for publication, 2006) (ArXiv: 0801.1962v1)

[3] See for instance [1]. For a more critical discussion, see [20, Sects. 9.4 and 9.5].

7. De Finetti, B.: La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré 7, 1–68 (1937); (English translation in [16])
8. De Finetti, B.: Teoria delle Probabilità. Einaudi, Turin (1970)
9. De Finetti, B.: Theory of Probability: A Critical Introductory Treatment. John Wiley & Sons, Chichester (1974); (English translation of [8], two volumes)
10. Denneberg, D.: Non-Additive Measure and Integral. Theory and Decision Library B: Mathematical and Statistical Methods, vol. 27. Kluwer Academic Publishers Group, Dordrecht (1994)
11. Edwards, R.E. (1995) Functional Analysis: Theory and Applications. Dover Publications, New York, first published by Holt, Rinehart and Winston, New York (1965)
12. Gärdenfors, P., Sahlin, N.E.: Decision, Probability, and Utility. Cambridge University Press, Cambridge (1988)
13. Heath, D.C., Sudderth, W.D.: De Finetti's theorem on exchangeable variables. Amer. Statist. 30, 188–189 (1976)
14. Kallenberg, O.: Foundations of Modern Probability, 2nd edn. Springer, New York (2002)
15. Kallenberg, O.: Probabilistic Symmetries and Invariance Principles. Springer, New York (2005)
16. Kyburg Jr, H.E., Smokler, H.E. (eds.): Studies in Subjective Probability (2nd edn. (with new material) in 1980). Wiley, Chichester (1964)
17. Ramsey, F.P.: Truth and probability (1926). In: Braithwaite, R.B. (ed.) The Foundations of Mathematics and other Logical Essays, Kegan, Paul, Trench, ch. VII, pp. 156–198. Trubner & Co., London (1931); (reprinted in [16] and [12])
18. Schechter, E.: Handbook of Analysis and Its Foundations. Academic Press, San Diego (1997)
19. Smith, C.A.B.: Consistency in statistical inference and decision. J. Roy Statist. Soc. Ser. A. 23, 1–37 (1961)
20. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
21. Williams, P.M.: Notes on Conditional Previsions. Tech. Report. School of Mathematical and Physical Science, University of Sussex, UK (1975)
22. Williams, P.M.: Indeterminate probabilities. In: Przelecki, M., Szaniawski, K., Wojcicki, R. (eds.) Formal Methods in the Methodology of Empirical Sciences, Reidel, Dordrecht, pp. 229–246 (1976)

Set-Valued Stochastic Integrals with Respect to a Real Valued Martingale

Jinping Zhang*

Department of Mathematics, Saga University, Saga, Japan

Abstract. In a real separable martingale type 2 Banach space, first we give the definition of single valued stochastic integrals by the differential of a real valued continuous L^2-martingale, and then, consider the set-valued case. Submartingale property and Castaing representation of set-valued stochastic integrals are obtained.

Keywords: Martingale type 2 Banach space, Set-valued stochastic integral.

1 Introduction

Integration of set-valued functions is an important topic in set-valued analysis. There are several types of integration of set-valued functions. Aumann integral [1] is a popular one, which is the collection of all integrable selections's integrals with respect to a measure. Hukuhara in 1967 considered formal Riemann integration into the space of all convex and compact subsets. Debreu in 1967 used an embedding method to consider the Bochner integral in the embedded Banach space. Based on Aumann's sense, Hiai [3] studied the properties of the integrals of the set-valued functions, the conditional expectations w.r.t a σ-finite measure and then martingales of multivalued functions.

It is well known that the stochastic integral is the foundation of stochastic analysis. For set-valued stochastic analysis, it is necessary to set up theory of set-valued stochastic integrals. So far, only a few papers have been published concerning the set-valued stochastic integrals (e.g. [4, 5, 6, 8]), since this topic has been studied within last ten years and the theory is not complete until now. In [6] the set-valued integral may be not a set-valued stochastic process, which is not an analogue to the single case. In [4], the authors modified the definition of [6] in 1-dimensional Euclidean space $\mathbb{R}$ such that the set-valued integral is a set-valued process. Li and Ren [8] modified Jung and Kim's [4] definition by considering the predictable set-valued stochastic process as a set-valued random variable in the product space $(\mathbb{R}^n_+ \times \Omega)$. In a martingale type 2 Banach space, Zhang et al. [12] studied set-valued stochastic integrals with respect to a real valued Brownian motion.

In this paper, we construct a theory of stochastic integration of set-valued processes with respect to a real valued L^2-continuous martingale. The range space of the integrands is allowed to be a real separable martingale type 2 Banach space.

* This work is partially supported by the NSFC(10771010).

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 253–259, 2008.
springerlink.com

2 Preliminaries

Let $(\Omega,\mathscr{F},P)$ be a complete nonatomic probability space with filtration $\{\mathscr{F}_t\}_{t\geq 0}$ which satisfies the usual condition, $(\mathfrak{X},\|\cdot\|)$ a real separable Banach space, $\mathscr{M}(\Omega;\mathfrak{X})$ the family of all $\mathfrak{X}$-valued $\mathscr{F}$-measurable functions, $\mathbf{K}_{(c)}(\mathfrak{X})$ the family of all nonempty closed (convex) subsets of $\mathfrak{X}$, $L^p(\Omega,\mathscr{F},P;\mathfrak{X})(p\geq 1)$ the set of all $\mathfrak{X}$- valued p-integrable random variables, briefly by $L^p(\Omega,\mathfrak{X})$. A function $F:\Omega\to\mathbf{K}(\mathfrak{X})$ is measurable if for any open set $O\in\mathfrak{X}$, the inverse $F^{-1}(O):=\{\omega\in\Omega:F(\omega)\cap O\neq\emptyset\}\in\mathscr{F}$. Let $\mathscr{M}(\Omega,\mathscr{F},P;\mathbf{K}_{(c)}(\mathfrak{X}))$ be the family of all measurable $\mathbf{K}_{(c)}(\mathfrak{X})$- valued functions, briefly by $\mathscr{M}(\Omega;\mathbf{K}_{(c)}(\mathfrak{X}))$.

For $F\in\mathscr{M}(\Omega,\mathbf{K}(\mathfrak{X}))$, the family of all L^p- integrable selections is denoted by $S^p_F(\mathscr{F}):=\{f\in L^p(\Omega,\mathscr{F},P;\mathfrak{X}):f(\omega)\in F(\omega)\ a.s.\},p\geq 1$. $S^p_F(\mathscr{F})$ may be denoted briefly by S^p_F. A set-valued random variable F is said to be *integrable* if S^1_F is nonempty. F is called $L^p(p\geq 1)$-*integrably bounded* if there exits $h\in L^p(\Omega,\mathscr{F},P;\mathbb{R})$ s.t. for all $x\in F(\omega)$, $\|x\|\leq h(\omega)$ almost surely. The family of all $\mathbf{K}_{(c)}(\mathfrak{X})$-valued L^p-integrable random variables is denoted by $L^p(\Omega,\mathscr{F},P;\mathbf{K}_{(c)}(\mathfrak{X}))$. It may be written for brevity by $L^p(\Omega;\mathbf{K}_{(c)}(\mathfrak{X}))$.

Proposition 1. *([3]) Let Γ be a nonempty closed subset of $L^p(\Omega,\mathscr{F},P;\mathfrak{X})$ and $1\leq p<\infty$. Then there is an $F\in\mathscr{M}(\Omega;\mathbf{K}(\mathfrak{X}))$ such that $\Gamma=S^p_F$ if and only if Γ is decomposable with respect to $\mathscr{F}$.*

Lemma 1. *([3]) Let $F\in\mathscr{M}(\Omega,\mathscr{F},P;\mathbf{K}(\mathfrak{X}))$ and $1\leq p\leq\infty$. If S^p_F is nonempty, then there is a sequence $\{f^i:i\in\mathbb{N}\}\subset S^p_F$ such that $F(\omega)=cl\{f^i(\omega):i\in\mathbb{N}\}$ for all $\omega\in\Omega$, where the closure is taken in $\mathfrak{X}$.*

Lemma 2. *([3]) Let $F_1,F_2\in\mathscr{M}(\Omega,\mathscr{F},P;\mathbf{K}(\mathfrak{X}))$, $1\leq p\leq\infty$, $S^p_{F_1}\neq\emptyset$ and $S^p_{F_2}\neq\emptyset$ then $S^p_{F_1}=S^p_{F_2}$ if and only if $F_1(\omega)=F_2(\omega)$ a.s.; $S^p_{F_1}\subset S^p_{F_2}$ if and only if $F_1(\omega)\subset F_2(\omega)$ a.s..*

The *integral (or expectation)* of a set-valued random variable F was defined by Aumann in 1965: $E[F]:=\{E[f]:f\in S^1_F\}$. Since set-valued stochastic integrals (It will be studied in Section 3) are integrable but maybe unbounded almost surely (see [11]), in order to study martingale property of set-valued stochastic integrals later, here we need to use the extended definition of conditional expectation compared with that in [3].

Assume $\mathscr{B}$ is a sub-sigma algebra of $\mathscr{F}$, F is an L^1-integrable set-valued random variable, the *conditional expectation* of F with respect to $\mathscr{B}$ is defined as follows:

Lemma 3. *([10]) Let F be an L^1-integrable set-valued random variable. For each sub-sigma algebra $\mathscr{B}\subset\mathscr{F}$, there exists a unique integrable $\mathscr{B}$-measurable set-value random variable Y (denoted by $Y=\mathscr{E}[F|\mathscr{B}]$ and called the conditional expectation of F) such that*

$$S^1_Y(\mathscr{B})=cl\{E[f|\mathscr{B}]:f\in S^1_F\},$$

where the closure is taken in L^1.

$F=\{F_t:t\geq 0\}$ (or denoted by $F=\{F(t):t\geq 0\}$)is called a *set-valued stochastic process* if for every fixed $t\geq 0$, $F_t(\cdot)$ is a set-valued random variable. $F=\{F_t:t\geq 0\}$ is called L^p-*integrable* if every F_t is L^p-integrable. It is called *measurable* if it is $\mathscr{B}_+\times\mathscr{F}$-measurable, $\mathscr{F}_t$-*adapted* if for any fixed t, $F_t(\cdot)$ is $\mathscr{F}_t$-measurable.

By the Lemma 2, it is reasonable to give the following definitions on set-valued martingale, set-valued submartingale and supermartingale.

Definition 1. *An integrable convex set-valued $\mathscr{F}_t$-adapted stochastic process $\{F_t, \mathscr{F}_t : t \geq 0\}$ is called a set-valued $\mathscr{F}_t$-martingale if for any $0 \leq s \leq t$ it holds that $\mathscr{E}[F_t|\mathscr{F}_s] = F_s$ in the sense of $S^1_{\mathscr{E}[F_t|\mathscr{F}_s]}(\mathscr{F}_s) = S^1_{F_s}(\mathscr{F}_s)$.*

It is called a set-valued submartingale (supermartingale) if for any $0 \leq s \leq t$, $\mathscr{E}[F_t|\mathscr{F}_s] \supset F_s$ (resp. $\mathscr{E}[F_t|\mathscr{F}_s] \subset F_s$) in the sense of $S^1_{\mathscr{E}[F_t|\mathscr{F}_s]}(\mathscr{F}_s) \supset S^1_{F_s}(\mathscr{F}_s)$ (resp. $S^1_{\mathscr{E}[F_t|\mathscr{F}_s]}(\mathscr{F}_s) \subset S^1_{F_s}(\mathscr{F}_s)$).

3 Set-Valued Stochastic Integrals with Respect to a Real Valued Continuous L^2-Martingale

In this section, we assume $\mathfrak{X}$ is a separable martingale type 2 Banach space. Let $T \in \mathbb{R}_+$ and $\{M_t, \mathscr{F}_t : t \in [0,T]\}$ (or denoted by $\{M(t), \mathscr{F}_t : t \in [0,T]\}$ be a real valued continuous L^2-martingale with $M_0(\omega) = 0$ a.e., where we call $\{M_t, \mathscr{F}_t : t \in [0,T]\}$ a continuous L^2-martingale if it is an $\mathscr{F}_t$-adapted continuous martingale and for any $t \in \mathbb{R}_+$ $E[M_t^2] < +\infty$. Then by the Doob-Meyer decomposition theorem, there exists a unique predictable continuous increasing process $\langle M \rangle_t$ such that $M_t^2 - \langle M \rangle_t$ becomes a continuous martingale. Let $\mathscr{L}^p(\mathfrak{X})$ be the family of all predictable $\mathfrak{X}-$valued stochastic processes $f = \{f(t), \mathscr{F}_t : t \in [0,T]\}$ (or $f = \{f_t, \mathscr{F}_t : t \in [0,T]\}$) such that $E[\int_0^T \|f(s)\|^p d < M >_s] < \infty$, $\mathscr{L}^p(\mathbf{K}_{(c)}(\mathfrak{X}))$ the family of all predictable $\mathbf{K}_{(c)}(\mathfrak{X})$-valued processes $F = \{F_t, \mathscr{F}_t : t \in [0,T]\}$ (or $F = \{F(t), \mathscr{F}_t : t \in [0,T]\}$) such that $\{\|F(t)\|_{\mathbf{k}}\}_{t\in[0,T]} \in \mathscr{L}^p(\mathbb{R})$, where $\|A\|_{\mathbf{k}} = \sup_{a\in A} \|a\|$.

For a set-valued stochastic process $\{F_t, \mathscr{F}_t : t \in [0,T]\}$, a predictable selection $f = \{f(t), \mathscr{F}_t : t \in [0,T]\}$ is called $\mathscr{L}^p$*-selection* if $f = \{f(t), \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^p(\mathfrak{X})$. The family of all $\mathscr{L}^p$-selections is denoted by $S^p(F(\cdot))$.

Definition 2. *([2]) A Banach space $(\mathfrak{X}, \|\cdot\|)$ is called martingale type 2 if and only if there exists a constant $C > 0$ such that for any $\mathfrak{X}$-valued martingale $\{\mathbf{M}_k\}$ (*Note: it may be different from M_t given in the beginning of this section*), it holds that $\sup_k E\|\mathbf{M}_k\|^2 \leq C \sum_k E\|\mathbf{M}_k - \mathbf{M}_{k-1}\|^2$.*

Let $\mathscr{L}^p_{step}(\mathfrak{X})$ be the subspace of those $f \in \mathscr{L}^p(\mathfrak{X})$ for which there exists a partition $0 = t_0 < t_1 < \ldots < t_n = T$ such that $f(t) = f(t_k)$ for $t \in [t_k, t_{k+1}), 0 \leq k \leq n-1, n \in \mathbb{N}$.

For $f \in \mathscr{L}^2_{step}(\mathfrak{X})$, define a $\mathfrak{X}$-valued $\mathscr{F}$-measurable random variable

$$I_T(f) := \sum_{k=1}^{n-1} f(t_k)(M(t_{k+1}) - M(t_k)).$$

We have the following lemmas, which are crucial for defining the Itô integration successfully.

Lemma 4. *For $f \in \mathscr{L}^2_{step}(\mathfrak{X})$, $I_T(f) \in L^2(\Omega, \mathscr{F}, P; \mathfrak{X})$, $E[I_T(f)] = 0$ and*

$$E[\|I_T(f)\|^2] \leq CE[\int_0^T \|f(t)\|^2 d < M >_t], \tag{1}$$

where the constant C is the same one appearing in Definition 2.

Proof. The proof is similar to the Lemma 4.1 in [12] replacing Brownain motion by continuous L^2- martingale. □

Lemma 5. *([7]) $\mathscr{L}^2_{step}(\mathfrak{X})$ is dense in $\mathscr{L}^2(\mathfrak{X})$.*

Then by Lemmas 4 and 5, we can extend the integrands into a larger class $\mathscr{L}^2(\mathfrak{X})$. So for $f \in \mathscr{L}^2(\mathfrak{X})$, we can choose a sequence $\{f^n : n \in \mathbb{N}\} \subset \mathscr{L}^2_{step}(\mathfrak{X})$ such that

$$E\int_0^T ||f^n - f||^2 d < M >_t \to 0 (n \to \infty),$$

then define

$$I_T(f)(\omega) := \int_0^T f_t(\omega) dM_t(\omega) := \lim_{n\to\infty} \int_0^T f_t^n(\omega) dM_t(\omega),$$

where the limit is taken in L^2-sense. Then from Lemma 4 and property of limit, taking $f \in \mathscr{L}^2(\mathfrak{X})$, we have $E[||I_T(f)||^2] \le CE\int_0^T ||f_t||^2 d < M >_t$. For any interval $[s,t] \subset [0,T]$, the integral $\int_s^t f_u dM_u$ can be defined similarly. For any $f \in \mathscr{L}^2(\mathfrak{X})$, it is not difficult to prove the process $\{I_t(f) : t \in [0,T]\}$ is an $\mathfrak{X}$-valued $\mathscr{F}_t$ martingale.

Now we study the set-valued stochastic integration.

For a set-valued stochastic process $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}(\mathfrak{X}))$, define a set of functions

$$\Gamma_t := \left\{ \int_0^t f_s dM_s : (f(t))_{t\in[0,T]} \in S^2(F(\cdot)) \right\} \tag{2}$$

Theorem 1. *For any $t \in [0,T]$, Γ_t is a bounded subset of $L^2(\Omega, \mathscr{F}_t, P; \mathfrak{X})$. Furthermore, if $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}_c(\mathfrak{X}))$, then Γ_t is also convex. If $\mathfrak{X}$ is reflexive and F_t is convex valued, then Γ_t is weakly compact.*

Proof. Here we omit the proof since the limitation of pages. □

Let $de\Gamma_t$ denote the decomposable set of Γ_t with respect to $\mathscr{F}_t$, $\overline{de}\Gamma_t$ the decomposable closed hull of Γ_t with respect to $\mathscr{F}_t$, where the closure is taken in L^1.

Theorem 2. *Assume $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}(\mathfrak{X}))$, then for any $t \in [0,T]$, $\overline{de}\Gamma_t \subset L^1(\Omega, \mathscr{F}_t, P; \mathfrak{X})$. Moreover, there exists a set-valued random variable $I_t(F) \in \mathscr{M}(\Omega, \mathscr{F}_t, P; \mathbf{K}(\mathfrak{X}))$ such that $S^1_{I_t(F)}(\mathscr{F}_t) = \overline{de}\Gamma_t$.*

Definition 3. *The set-valued stochastic process $(I_t(F))_{t\in[0,T]}$ defined as above is called the stochastic integral of $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}(\mathfrak{X}))$ with respect to a real valued continuous L^2-martingale $\{M_t, \mathscr{F}_t; t \in [0,T]\}$. For each t, we denote $I_t(F) = \int_0^t F_s dM_s$. Similarly, for $0 \le s < t$, we also can define the set- valued random variable $\int_s^t F_u dM_u$.*

Remark 1. Although Γ_t is bounded in $L^2(\Omega; \mathfrak{X})$, unfortunately, the set $\overline{de}\Gamma_t$ may be not bounded in $L^2(\Omega; \mathfrak{X})$ even in $L^1(\Omega; \mathfrak{X})$. That implies the set-valued integral may be unbounded in $\mathfrak{X}$ almost surely.

Theorem 3. *Assume a sequence set-valued processes* $\{F^n : n \in \mathbb{N}\}$ $(F^n = \{F^n(t) : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}_c(\mathfrak{X}))$ *for* $n \in \mathbb{N}$*) is decreasing, that is* $F^1(t,\omega) \supset F^2(t,\omega) \supset \ldots \supset F(t,\omega)$*, where* $F(t,\omega) = \bigcap_{n=1}^{\infty} F^n(t,\omega)$*, a.e.* $(t,\omega) \in [0,T] \times \Omega$*, then*

$$\int_0^t F^1(s,\omega)dM_s(\omega) \supset \int_0^t F^2(s,\omega)dM_s(\omega) \supset \ldots \supset \int_0^t F(s,\omega)dM_s(\omega) \ a.s.$$

but in general, the following may not be true

$$\int_0^t F(s,\omega)dM_s(\omega) = \bigcap_{n=1}^{\infty} \int_0^t F^n(s,\omega)dM_s \ a.s.. \tag{3}$$

Proof. The first result is obviously by the Lemma 2. For the equality (3), Theorem 3.16 in [5] proved it is true when $\mathfrak{X}$ is a Hilbert space and the integrator is a Brownian motion. But now we will give a converse example to show it may not be true.

Let the space $\mathfrak{X} = \mathbb{R}$, the integrator be a real valued $\mathscr{F}_t$ Brownian motion $\{B_t, \mathscr{F}_t : t \in [0,T]\}$ with zero initial value, obvious $\{B_t, \mathscr{F}_t : t \in [0,T]\}$ is a continuous L^2-martingale.

Now consider the sequence of set-valued processes:

$$\left\{F^n(t,\omega) = [-\frac{1}{n}, \frac{1}{n}]\right\} \quad for\ all\ (t,\omega) \in [0,T] \times \Omega,$$

clearly

$$F^1(t,\omega) \supset F^2(t,\omega) \supset \ldots \supset F(t,\omega),$$

where $F(t,\omega) = \bigcap_{n=1}^{\infty} F^n(t,\omega) = \{0\}$. According to Theorem 1 in [11], it is easy to obtain that $\int_0^t F^n(s,\omega)dB_s(\omega) = \mathbb{R}$ a.s. for every n, then

$$\bigcap_{n=1}^{\infty} \int_0^t F^n(s,\omega)dB_s(\omega) = \mathbb{R} \ a.s.$$

but

$$\int_0^t F(s,\omega)dB_s(\omega) = \{0\},$$

that implies the equality (3) does not hold, further it is not available to define fuzzy set-valued stochastic integrals by considering the set-valued integral of every α-cut, which appears in [5]. □

Theorem 4. *Let* $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}_c(\mathfrak{X}))$*, then the stochastic integral* $\{I_t(F), \mathscr{F}_t : t \in [0,T]\}$ *is a set-valued submartingale.*

Proof. The proof is similar to that of Theorem 4.2 in [12], here we omit it since the limitation of pages. □

Lemma 6. *[12] Let* $(E, \mathscr{B}, \mu)$ *be a* σ*-finite measure space, if* $\mathscr{B}$ *is separable with respect to* μ *(i.e. there exists a countably generated sub-sigma algebra* $\mathscr{B}_0 \subset \mathscr{B}$ *such that for every* $A \in \mathscr{B}$*, there is* $B \in \mathscr{B}_0$ *satisfying* $\mu(A \triangle B) = 0$*), then space* $L^p(E;\mathfrak{X})(p \geq 1)$ *is separable in norm.*

Lemma 7. *If $\mathscr{F}$ is separable with respect to P, for $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}(\mathfrak{X}))$, there exists a sequence $\{f^n : n \in N\} \subset S^2(F(\cdot))$, such that for every $t \in [0,T]$,*

$$S^2_{I_t(F)} = \overline{de}\left\{\int_0^t f_s^n dM_s : n \in \mathbb{N}\right\}, \tag{4}$$

where the closure is taken in L^1, decomposability is with respect to $\mathscr{F}_t$.

Proof. By the Lemma 6, we can get the desired result. Here we omit the proof since the limitation of pages. □

Theorem 5. *(Castaing representation of set-valued stochastic integral)*
Assume $\mathscr{F}$ is separable with respect to P. Then for a set-valued stochastic process $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}(\mathfrak{X}))$, there exists a sequence $\{(f_t^i)_{t\in[0,T]} : i = 1,2,...\} \subset S^2(F(\cdot))$ such that for each $t \in [0,T]$, $F(t,\omega) = cl\{(f_t^i(\omega)) : i = 1,2,...\}$ a.s., and

$$I_t(F)(\omega) = cl\left\{\int_0^t f_s^i(\omega) dM_s(\omega) : i = 1,2,...\right\} \ a.s. \tag{5}$$

Proof. By using the Lemma 7, similar to the proof of Theorem 4.3 in [12], we can get the desired result. □

Theorem 6. *Assume $\mathscr{F}$ is separable with respect to P, a set-valued stochastic process $\{F_t, \mathscr{F}_t : t \in [0,T]\} \in \mathscr{L}^2(\mathbf{K}(\mathfrak{X}))$, and $\{M_t, \mathscr{F}_t : t \in [0,T]\}$ is a real valued continuous L^2-martingale, where $0 \le t_1 < t \le T$. Then the following holds*

$$I_t(F)(\omega) = cl\left\{I_{t_1}(F)(\omega) + \int_{t_1}^t F_s(\omega) dM_s(\omega)\right\}, \ a.s. \tag{6}$$

where the closure is taken in $\mathfrak{X}$.

Proof. From Theorem 5, it is not difficult to get it. □

Acknowledgement. The author is extremely grateful to Prof. Itaru Mitoma, Prof. Yoshiaki Okazaki in Japan and Prof. Shoumei Li in China. Also the author thanks the referees for their helpful comments and suggestions.

References

1. Aumann, R.: Integrals of set-valued functions. J. Math. Anal. Appl. 12, 1–12 (1965)
2. Brzeźniak, Z., Carroll, A.: Approximations of the Wong-Zakai differential equations in M-type 2 Banach spaces with applications to loop spaces. In: Séminaire de Probabilités XXXVII. Lecture Notes in Math, vol. 1832, pp. 251–289. Springer, Berlin (2003)
3. Hiai, F., Umegaki, H.: Integrals, conditional expectations and martingales of multivalued functions. J. Multivariate Anal. 7, 149–182 (1977)
4. Jung, E.J., Kim, J.H.: On set-valued stochastic integrals. Stochastic. Anal. Appl. 21(2), 401–418 (2003)

5. Kim, B.K., Kim, J.H.: Stochastic integrals of set-valued processes and fuzzy processes. J. Math. Anal. Appl. 236, 480–502 (1999)
6. Kisielewicz, M.: Set-valued stochastic integrals and stochastic inclusions. Stochastic. Anal. Appl. 15, 780–800 (1997)
7. Kunita, H.: Stochastic Differential Equations and Stochastic Flows of Diffeomorphisms. In: École d'été de probabilités de Saint-Flour XII—1982. Lecture Notes in Math, vol. 1097, pp. 143–303. Springer, Berlin (1984)
8. Li, S., Ren, A.: Representation theorems, set-valued and fuzzy set-valued Itô Integral. Fuzzy Sets Syst. 158, 949–962 (2007)
9. Megginson, R.E.: An Introduction to Banach Space Theory. Springer, New York (1998)
10. Molchanov, I.: Theory of Random Sets. Springer, London (2005)
11. Ogura, Y.: On stochastic differential equations with set coeffients and the Black-Scholes model. In: Proceedings of the Eighth International Conference on Intelligent Technologies (InTech 2007, Sydney, Australia) (2007)
12. Zhang, J., Li, S., Mitoma, I., Okazaki, Y.: On Set-Valued Stochastic Integrals in an M-type 2 Banach Space (submitted to publication) (2008)

Differential Equations

On Stochastic Differential Equations with Fuzzy Set Coefficients

Yukio Ogura*

Department of Mathematics, Saga University, Saga, Japan

Abstract. We define stochastic differential equations with fuzzy set coefficients and prove that their solutions are random fuzzy set processes. This is achieved by obtaining almost sure boundedness of solutions to stochastic differential equations with set coefficients. An example for Black-Scholes market model with expected return ratio being a fuzzy set is also given.

Keywords: Stochastic differential equation, Set coefficients, Fuzzy set coefficients, Random sets, Random fuzzy sets.

1 Introduction

Theory of stochastic differential equations (SDE for brief) which were introduced by K. Itô [4] founded the base of stochastic analysis and has been playing a most important role in it and its applications.

In this article, we define SDE's with fuzzy set coefficients of drift terms and prove that their solutions are random fuzzy set processes. For this purpose, we shall first study SDE's with set coefficients. Historically, those SDE's are sometimes regarded as stochastic inclusions and systematically developed (see e.g. [1], [5], [7], [8]). However, in most of the works, the solutions to those are given as L^2-valued (or L^p-valued) processes and this makes it hard to extend those to SDE's with fuzzy set coefficients, because monotone property of the solutions in L^2 space does not imply almost surely monotone property in the underlying space. In a previous work [11], we studied SDE's with set coefficients from the point of view of almost sure property, and showed that we can not expect almost surely bounded solutions if the coefficients of stochastic differential terms are set functions in typical cases, but can obtain almost surely bounded solutions if the coefficients of stochastic differential terms are usual functions. This study enables us to extend the results to those for SDE's with fuzzy set coefficients in the case where the coefficients of stochastic differential terms are usual functions.

The organization of this article is as follows. In the next Section 2, we review and refine the results for SDE's with set coefficients given in [11]. In Section 3, we make use of the results in Section 2 for the study SDE's with fuzzy set coefficients and prove that, if the coefficients of stochastic differential terms are usual functions, the solutions of SDE's with fuzzy set coefficients are random fuzzy set processes in most cases. As

* Partially supported by Grant-in-Aid for Scientific Research 19540140.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 263–270, 2008.
springerlink.com

an example, the Black-Scholes market model in mathematical finance is extended, so that the investors could presume the expected return ratio as a fuzzy set.

2 Stochastic Differential Equations with Set Coefficients

Before the study SDE's with fuzzy set coefficients, we will provide in this section some properties of SDE's with set coefficients.

Let W_T be the space of all continuous functions $w : [0, T] \to \mathbb{R}$ with the uniform norm $\|w\|_{[0,T]} = \max_{t\in[0,T]} |w(t)|$. The Wiener measure on $\{w \in W_T : w(0) = 0\}$ is denoted by P. Thus $(w(t), P)$ is a Brownian motion with respect to the increasing complete right continuous filtration derived from the natural filtration. We denote $B_t(w) = w(t)$ for $t \in [0, T]$ and $w \in W_T$ as usual.

Denote by $\mathscr{A}_T$ the set of all continuous functions $\alpha(t, x) : [0, T] \times \mathbb{R} \to \mathbb{R}$ which satisfy

$$|\alpha(t, x) - \alpha(t, y)| \le L|x - y|, \quad x, y \in \mathbb{R},\ t \in [0, T],$$
$$|\alpha(t, x)| \le L(1 + |x|), \quad x \in \mathbb{R},\ t \in [0, T],$$

for a positive constant L which may depend on α. It then follows that, for each $a, b \in \mathscr{A}_T$ and $x_0 \in \mathbb{R}$, the SDE

$$dx(t) = a(t, x(t))dB_t + b(t, x(t))dt,\ x(0) = x_0 \tag{1}$$

has a unique strong solution $x(\cdot) = x(\cdot, x_0; a, b)$, which is determined almost surely with respect to P (see e.g. [3]). However, in order to deal with SDE's with set coefficients, we must take the exceptional null set commonly for a class of SDE's.

For this purpose, we use approximation method. For each $\delta > 0$, let $\Delta_k w = w((k + 1)\delta \wedge T) - w(k\delta \wedge T)$, $k \in \mathbb{N}$, and

$$B_t^\delta = B_t^\delta(w) = ((t - [t/\delta]\,\delta)/\delta)\Delta_{[t/\delta]}w + w([t/\delta]\,\delta),$$
$$t \in [0, T],\ w \in W_T,$$

where $a \wedge b$ stands for $\min\{a, b\}$ and $[\ ]$ the Gauss symbol (that is, $[t/\delta]$ is the smallest integer which is equal to or greater than t/δ). The approximating equation for (1) is given by

$$\dot{x}^\delta(t) = a(t, x^\delta(t))\dot{B}_t^\delta + b(t, x^\delta(t)), \quad x^\delta(0) = x_0, \tag{2}$$

where $\dot{x}^\delta(t)$ and $\dot{B}_t^\delta$ stand for the right derivatives of $x^\delta(t)$ and $\dot{B}_t^\delta$ at t respectively. Then equation (2) has a unique solution $x^\delta(t, x_0; a, b)$, $t \in [0, T]$. Let $\mathscr{C}_T$ be a subset of $\mathscr{A}_T \times \mathscr{A}_T$ such that there exist $\delta_n > 0$, $n \in \mathbb{N}$ with $\lim_{n\to\infty} \delta_n = 0$ and a Borel set $\widetilde{W}_T$ in W_T which satisfy the following conditions.

(i) $P(\widetilde{W}_T) = 1$,

(ii) For each $(a, b) \in \mathscr{C}_T$, $x_0 \in \mathbb{R}$ and $w \in \widetilde{W}_T$, there exists an $x(\cdot, x_0; a, b)(w) \in W_T$ such that

$$\lim_{n\to\infty} \|x^{\delta_n}(\cdot, x_0; a, b)(w) - x(\cdot, x_0; a, b)(w)\|_{[0,T]} = 0, \quad w \in \widetilde{W}_T. \tag{3}$$

We define $x(t, x_0; a, b)(w) = 0$ for $t \in [0, T]$ and $w \notin \widetilde{W}_T$ for completeness. Note that each element $(a, b) \in \mathscr{A}_T \times \mathscr{A}_T$ satisfies (3) a.s. for some $x(\cdot, x_0; a, b)(w)$ (see e.g. [13]), so that any countable subset $\mathscr{C}_T$ of $\mathscr{A}_T \times \mathscr{A}_T$ satisfies the conditions above.

Let $\mathscr{K}_k = \mathscr{K}_k(\mathbb{R})$ be the family of all non-empty compact subsets of $\mathbb{R}$. Define the Hausdorff metric d_H by

$$d_H(A, B) = \max\left\{\max_{a\in A} d_{\mathbb{R}}(a, B), \max_{a\in B} d_{\mathbb{R}}(a, A)\right\},$$

for $A, B \in \mathscr{K}_k$, where $d_{\mathbb{R}}(a, A) = \min_{b\in A} |a - b|$ for $a \in \mathbb{R}$ and $A \in \mathscr{K}_k$. Then $\mathscr{K}_k$ is a complete separable metric space with respect to the Hausdorff metric d_H (see e.g. [9, Theorem 1.1.3]).

Now, for two mappings $A(t, x),\ B(t, x) : [0, T] \times \mathbb{R} \to \mathscr{K}_k$, we define the solution of SDE's with set coefficients

$$dX(t) = A(t, X(t))dB_t + B(t, X(t))dt,\ X(0) = X_0, \tag{4}$$

where $A(t, X(t))$ and $B(t, X(t))$ are understood as

$$A(t, X(t)) = \bigcup_{x\in X(t)} A(t, x), \quad B(t, X(t)) = \bigcup_{x\in X(t)} B(t, x).$$

Definition 1. *Let $\mathscr{C}_T \subset \mathscr{A}_T \times \mathscr{A}_T$ satisfy conditions* (i) *and* (ii). *Then a solution $X(t) = X(t; \mathscr{C}_T)$ of the SDE with set coefficients* (4) *is defined by*

$$X(t) = \mathrm{cl}\{x(t, x_0; a, b) : (a, b) \in \mathscr{C}_T,\ (a(s, x), b(s, x)) \in A(s, x) \times B(s, x),\\ \text{for all } (s, x) \in [0, T] \times \mathbb{R},\ x_0 \in X_0\}.$$

Note 1. In the proceeding works, they assume some measurability on $A(t, x)$ and $B(t, x)$. However, in our definition above, we need not to mention on it, because we only choose 'selections' $a(s, x)$ and $b(s, x)$. Of course, this is only for the definition, and measurability or even continuity are satisfied in most examples.

In the following, we restrict ourselves to the case when $A(t,x)$ and $B(t, x)$ are interval functions to make the situation clear. Take $a_i, b_i \in \mathscr{A}_T$, $x_0^i \in \mathbb{R}$ $(i = 1, 2)$ which satisfy

$$a_1(t, x) \le a_2(t, x),\ b_1(t, x) \le b_2(t, x),\ x_0^1 \le x_0^2,$$

and denote

$$A_{a_1,a_2}(t, x) = [a_1(t, x), a_2(t, x)], \quad B_{b_1,b_2} = [b_1(t, x), b_2(t, x)],$$

for $(t, x) \in [0, T] \times \mathbb{R}$. Then equation (4) is reduced to

$$dX(t) = A_{a_1,a_2}(t, X(t))dB_t + B_{b_1,b_2}(t, X(t))dt,\ X(0) = X_0. \tag{5}$$

We then have the following Theorem.

Theorem 1. *(1) Suppose that there exists a $t_e \in (0, T]$ such that $a_1(t, x) < a_2(t, x)$ for all $(t, x) \in [0, t_e] \times \mathbb{R}$. Then there is a $\mathscr{C}_T$ satisfying conditions* (i) *and* (ii) *such that $X(t_e, X_0) \notin \mathscr{K}_k$, a.s.*

(2) Suppose that $a_1(t,x) = a_2(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}$. Then, for each $\mathscr{C}_T$ satisfying conditions (i) *and* (ii)*, it holds that $X(t,X_0) \in \mathscr{K}_k$, a.s. for all $t \in [0,T]$. Further, if $b_1(t,x) < b_2(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}$ and $x_0^1 < x_0^2$, then $X(t,X_0)$ is non-degenerate for all $t \in [0,T]$.*

Proof (Outline of proof).

(1) Suppose that the assumption in (1) is satisfied. Then, for each $L > 0$, there are partitions $0 = t_0 < t_1 < \cdots < t_l = t_e$, $0 = x_0 < x_1 < \cdots < x_l = L$ and an $\eta > 0$ such that, for each $k = 1, \ldots, l$,

$$\begin{aligned} a_k &:= \max\{a_1(t,x) : (t,x) \in [t_{k-1}, t_k] \times [x_{k-1}, x_k]\} \\ &< \min\{a_2(t,x) : (t,x) \in [t_{k-1}, t_k] \times [x_{k-1}, x_k]\} - \eta. \end{aligned}$$

We fix a $k = 1, \ldots, l$ for a while, and choose a non-trivial nonnegative continuous function φ on $\mathbb{R}$ with the support in $(0,1)$ and $\max_{t \in \mathbb{R}} \varphi(t) \le \eta$. For each $n \in \mathbb{N}$, let $h_n = (t_k - t_{k-1})/n$ and $t_i^n = t_{k-1} + ih_n$, $i = 0, 1, 2, \ldots, n$. Also, for each $\varepsilon^n = (\varepsilon_1^n, \varepsilon_2^n, \ldots, \varepsilon_n^n) \in \{-1, 1\}^n$, define $g_{\varepsilon^n}^n = g_{\varepsilon^n}^n(t)$ by

$$g_{\varepsilon^n}^n(t) = \sum_{i=1}^n \varepsilon_i^n \varphi(h_n^{-1}(t - t_{i-1}^n)), \qquad t \in [t_{k-1}, t_k].$$

We then have

$$a_k + g_{\varepsilon^n}^n(t) \in [a_1(t,x), a_2(t,x)], \quad \text{for all } (t,x) \in [t_{k-1}, t_k] \times [x_{k-1}, x_k].$$

Hence we can define an $a_{\varepsilon^n}^n \in \mathscr{A}_T$ such that $a_{\varepsilon^n}^n(t,x) \in A_{a_1,a_2}(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}$ and $a_{\varepsilon^n}^n(t,x) = a_k + g_{\varepsilon^n}^n(t)$, for all $(t,x) \in [t_{k-1}, t_k] \times [x_{k-1}, x_k]$ and $k = 1, 2, \ldots, l$. We then take a $b \in \mathscr{A}_T$ such that $b(t,x) \in B_{b_1,b_2}(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}$. It then follows $(a_{\varepsilon^n}^n, b) \in \mathscr{A}_T \times \mathscr{A}_T$ and $(a_{\varepsilon^n}^n(t,x), b(t,x)) \in A_{a_1,a_2}(t,x) \times B_{b_1,b_2}(t,x)$, for all $(t,x) \in [0,T] \times \mathbb{R}$. Now let

$$\mathscr{C}_T = \bigcup_{n \in \mathbb{N}} \{(a_{\varepsilon^n}^n, b_{\varepsilon^n}^n) : \varepsilon^n \in \{-1,1\}^n\}.$$

Then through the same line as in the proof of [11, Theorem 1], we obtain the conclusion. In fact, for each $k = 1, 2, \ldots, l$ and $n \in \mathbb{N}$, the sequence

$$Y_i^n := \int_{t_{i-1}^n}^{t_i^n} \varphi(h_n^{-1}(t - t_{i-1}^n)) dB_t, \qquad i = 1, 2, \ldots, n$$

is an independent random variables. Further, due to the stationary property and scaling law of Brownian motion, the law of Y_i^n is same as that of

$$\int_0^{h_n} \varphi(h_n^{-1}t) dB_t = \int_0^1 \varphi(t) dB_{h_n t} = h_n^{1/2} \int_0^1 \varphi(t) d\widetilde{B}_t,$$

where $\widetilde{B}_t$ is another Brownian motion $h_n^{-1/2} B_{h_n t}$. This implies

$$\lim_{n \to \infty} \sum_{i=1}^n |Y_i^n| = \infty, \ a.s.,$$

which provides the key point of the proof.

(2) Denote $a(t,x) = a_1(t,x) = a_2(t,x)$ and take a $\mathscr{C}_T$ satisfying conditions (i) and (ii). Then, for each $(a,b) \in \mathscr{C}_T$ with $b(t,x) \in [b_1(t,x), b_2(t,x)]$, we have

$$x^{\delta_n}(t, x_0^1; a, b_1)(w) \le x^{\delta_n}(t, x_0; a, b)(w) \le x^{\delta_n}(t, x_0^2; a, b_2)(w)$$

for all $n \in \mathbb{N}$ and $w \in \widetilde{W}_T$ by a comparison theorem for ordinary differential equations. Hence, from (3), we have

$$x(t, x_0^1; a, b_1)(w) \le x(t, x_0; a, b)(w) \le x(t, x_0^2; a, b_2)(w),$$

which proves the boundedness of the solution $X(t)$. Since $X(t)$ is closed, it follows $X(t) \in \mathscr{K}_k$.

The latter assertion is shown by a strong comparison theorem for SDE's (see [14]). □

3 Stochastic Differential Equations with Fuzzy Set Coefficients

Due to Theorem 1, in case (1), we can not expect to extend the results to SDE's with fuzzy set coefficients. However, in case (2), it is possible and would be interesting to extend those to SDE's with fuzzy set coefficients, which is the subject of this section.

Let $I = [0,1]$. A function $u : D_u \to I$ is a *fuzzy set* if D_u is a closed subset of $\mathbb{R}$ and, for every $\alpha \in I$, its α-level set $[u]_\alpha = \{x \in D_u : u(x) \ge \alpha\}$ is closed. Let $\mathscr{F}_k = \mathscr{F}_k(\mathbb{R})$ be the space of fuzzy sets such that D_u is compact and its 1-level set $[u]_1$ is non-empty. An element u in $\mathscr{F}_k$ is considered as an element of the space $D(I;\mathscr{K}_k) = D(I;\mathscr{K}_k(\mathbb{R}))$ of functions $u : I \to \mathscr{K}_k$ which is left continuous in $(0,1]$ and has right limit in $[0,1)$. We denote by $u(\alpha)$ the α-level set $[u]_\alpha$ of u for $\alpha \in I$. Then $\mathscr{F}_k$ is identified with the subspace $D_d(I;\mathscr{K}_k)$ of $D(I;\mathscr{K}_k)$ which consists of all decreasing $u \in D(I;\mathscr{K}_k)$, that is $u(\alpha) \supset u(\beta)$ whenever $0 \le \alpha < \beta \le 1$.

Take $a \in \mathscr{A}_T$, $G(t,x) : [0,T] \times \mathbb{R} \to \mathscr{F}_k$ and $Y_0 \in \mathscr{F}_k$. We then consider the SDE's with fuzzy set coefficients

$$dY(t) = a(t, Y(t))dB_t + G(t, Y(t))dt,\ Y(0) = Y_0. \tag{6}$$

Definition 2. *Let $\mathscr{C}_T \subset \mathscr{A}_T \times \mathscr{A}_T$ satisfy conditions* (i) *and* (ii). *Then a solution $Y(t) = Y(t;\mathscr{C}_T)$ of the SDE with set coefficients* (6) *is defined by the fuzzy sets whose level sets $[Y(t)]_\alpha$ satisfy*

$$[Y(t)]_\alpha = \begin{cases} \lim_{\gamma \to \alpha, \gamma < \alpha} \widetilde{Y}(t)_\gamma, & \alpha \in (0,1], \\ \widetilde{Y}(t)_0, & \alpha = 0, \end{cases} \tag{7}$$

where

$$\widetilde{Y}(t)_\gamma = \mathrm{cl}\{x(t, x_0; a, b) : (a,b) \in \mathscr{C}_T, \\ b(s,x) \in [G(s,x)]_\gamma, \textit{ for all } (s,x) \in [0,T] \times \mathbb{R},\ x_0 \in [Y_0]_\gamma\}.$$

We note that a family of left continuous decreasing closed sets u_α, $\alpha \in I$ uniquely determines a fuzzy set on D_u (see e.g. [9, Theorem 5.1.4] with awareness of the small change of definitions).

The main problem is if the solution to equation (6) is bounded almost surely. Further, each $G(t,x)$ is included in fuzzy intervals. Thus, to make the statement simpler, we restrict ourselves to the case where $G(t,x)$ and Y_0 are fuzzy set intervals. Thus, we assume that the level sets $[G(t,x)]_\alpha$ of $G(t,x)$ are bounded closed intervals $[\beta_1(t,x)_\alpha, \beta_2(t,x)_\alpha]$ where $\beta_i(t,x)_\alpha$, $i=1,2$ belong to $\mathscr{A}_T$ as functions in $(t,x)\in[0,T]\times\mathbb{R}$, and $\beta_1(t,x)_\alpha$ and $\beta_2(t,x)_\alpha$ are non-decreasing and non-increasing left continuous functions in $\alpha\in I$ respectively with $\beta_1(t,x)_1\le\beta_2(t,x)_1$. We sometimes denote $G(t,x)=G_{\beta_1,\beta_2}(t,x)=G_{\beta_1,\beta_2}$ to designate β_i explicitly. Assume also $[Y_0]_\alpha$ are bounded closed intevals $[y^1_{0,\alpha}, y^2_{0,\alpha}]$.

Then equation (6) is reduced to

$$dY(t)=a(t,Y(t))dB_t+G_{\beta_1,\beta_2}(t,Y(t))dt,\ Y(0)=Y_0. \tag{8}$$

Theorem 2. *For each $\mathscr{C}_T$ satisfying conditions* (i) *and* (ii)*, there exists a solution Y to* (8) *such that $Y(t,Y_0)\in\mathscr{F}_k$, for all $t\in[0,T]$ a.s. Further, if $\beta_1(t,x)_0<\beta_2(t,x)_0$ for $(t,x)\in[0,T]\times\mathbb{R}$, and $y^1_{0,0}<y^2_{0,0}$, then $Y(t,Y_0)$ is non-degenerate for all $t\in[0,T]$.*

Proof. Take a $w\in\widetilde{W}_T$ and denote as $\widetilde{Y}(t)_\gamma(w)$ to designate w explicitly. By the proof of Theorem 1 (2), we have $\widetilde{Y}(t)_\gamma(w)\in\mathscr{K}_k$, for all $\gamma\in I$. Further, since

$$\begin{aligned}&\{(a,b)\in\mathscr{C}_T,\ b(t,x)\in[G(t,x)]_{\gamma_1},\ x_0\in[X_0]_{\gamma_1}\}\\ &\supset\{(a,b)\in\mathscr{C}_T,\ b(t,x)\in[G(t,x)]_{\gamma_2},\ x_0\in[X_0]_{\gamma_2}\},\end{aligned}$$

whenever $0\le\gamma_1<\gamma_2\le1$, we have $\widetilde{Y}(t)_{\gamma_1}(w)\supset\widetilde{Y}(t)_{\gamma_2}(w)$ whenever $0\le\gamma_1<\gamma_2\le1$. Hence there exists a $Y(t)(w)\in\mathscr{F}_k$ such that (7) holds (see e.g. [9, Theorem 5.1.4]). Since $P(\widetilde{W}_T)=1$, we have the first asserion of Theorem. The latter assertion is a direct consequence of Theorem 1 (2). □

Example 1 (Black-Scholes market model). The Black-Sholes market model is introduced by [2] and [10], and studied from the point of view of martingale theory by [6], and [12]. We extend it to SDE's with fuzzy coefficients to allow investors only to presume the expected return ratio as a fuzzy set. For our extended Black-Scholes market model, we assume $a(t,x)=\alpha(t)x$ and $b(t,x)=\beta(t)x$, where $\alpha(t)$ and $\beta(t)$ are continuous functions in $t\in I$. Thus the level sets $[G(t,x)_{\beta_1,\beta_2}]_\gamma$ of the fuzzy interval $G(t,x)$ are given by

$$[G(t,x)_{\beta_1,\beta_2}]_\gamma=\{\beta(t)x:\beta(t)\in[\beta_1(t)_\gamma,\beta_2(t)_\gamma]\},$$

where $\beta_i(t)_\gamma$, $i=1,2$ are continuous functions in $t\in[0,T]$ such that $\beta_1(t)_1\le\beta_2(t)_1$ and $\beta_1(t)_\gamma$ (resp. $\beta_{2,\gamma}(t)$) are non-decreasing (resp. non-increasing) in $\gamma\in I$. From Theorem 2, we can then find a solution Y to (8) such that $Y(t,Y_0)\in\mathscr{F}_k$, for all $t\in[0,T]$ a.s. Further, if $\beta_1(t,x)_0<\beta_2(t,x)_0$, $(t,x)\in[0,T]\times\mathbb{R}$ and $y^1_{0,0}<y^2_{0,0}$, then $Y(t,Y_0)$ is non-degenerate for all $t\in[0,T]$.

In this case, we can also obtain the conclusions directly. Indeed, (1) is reduced to

$$dx(t)=\alpha(t)x(t)dB_t+\beta(t)x(t)dt,\ x(0)=x_0,$$

whose solution is given by

$$x(t) = x_0 \exp\left[\int_0^t \alpha(s)dB_s + \int_0^t (\beta(s) - \frac{1}{2}\alpha(s)^2)ds\right].$$

We thus have

$$\widetilde{Y}(t)_\gamma = \mathrm{cl}\Big\{x_0 \exp\Big[\int_0^t \alpha(s)dB_s + \int_0^t (\beta(s) - \frac{1}{2}\alpha(s)^2)ds\Big] : (a,b) \in \mathscr{C}_T, \\ \beta(s) \in [\beta_1(s)_\gamma, \beta_2(s)_\gamma] \text{for all } s \in [0,T]\Big\}.$$

This gives the bounds of $\widetilde{Y}(t)_\gamma$; $\widetilde{Y}(t)_\gamma \subset [\widetilde{y}^1_\gamma(t), \widetilde{y}^2_\gamma(t)]$, where

$$\widetilde{y}(t)^i_\gamma = y^i_{0,\gamma} \exp\left[\int_0^t \alpha(s)dB_s + \int_0^t (\beta_i(s)_\gamma - \frac{1}{2}\alpha(s)^2)ds\right], \quad i = 1,2.$$

4 Conclusion

For the solutions to SDE's with set coefficients, we conclude the following. (i) If the coefficients of stochastic differential term is a set, then the solutions can be unbounded almost surely in most cases. (ii) If the coefficients of stochastic differential term is a usual function, then the solutions are bounded almost surely.

This yields that, for SDE's with fuzzy set coefficients, only the coefficients drift terms could be fuzzy sets to obtain almost surely bounded fuzzy set solutions. Further, we actually obtained almost surely bounded fuzzy set solutions in the case where the coefficients of stochastic differential term are usual functions and the coefficients drift terms are bounded fuzzy set intervals. The Black-Scholes market model in mathematical finance is extended, so that the investors could presume the expected return ratio as a fuzzy set.

References

1. Ahmed, N.U.: Nonlinear stochastic differential inclusions on Banach space. Stochastic Anal. Appl. 12, 1–10 (1994)
2. Black, F., Scholes, M.: The pricing of options and corporate liabilities. Scottish J. Political Econ. 81, 637–654 (1973)
3. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes, 2nd edn. North-Holland Mathematical Library, vol. 24. North-Holland, Amsterdam (1989)
4. Itô, K.: On a stochastic integral equation. Proc. Japan Acad. 22, 32–35 (1946)
5. Jung, E.J., Kim, J.H.: On set-valued stochastic integrals. Stochastic Anal. Appl. 21, 401–418 (2003)
6. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics, vol. 113. Springer, New York (1988)
7. Kisielewicz, M.: Properties of solution set of stochastic inclusions. J. Appl. Math. Stochastic Anal. 6, 217–236 (1993)
8. Kisielewiz, M., Michta, M., Motyl, J.: Set valued approach to Stochastic control, Part 1: Existence and regularity properties. Dynamic Systems Appl. 12, 405–432 (2003); Part 2: Viability and semimartingale issues 12, 433–466

9. Li, S., Ogura, Y., Kreinovich, V.: Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables. Theory and Decision Library. Series B: Mathematical and Statistical Methods, vol. 43. Kluwer Academic Publishers Group, Dordrecht (2002)
10. Merton, R.C.: Optimum consumption and portfolio rules in a continuous time model. J. Econom. Theory 3, 373–413 (1971)
11. Ogura, Y.: On stochastic differential equations with set coefficients and the Black-Scholes model. In: Proceedings of the 8th International Conference on Intelligent Technologies (InTech 2007, Sydney, Australia), pp. 300–304 (2007)
12. Øksendal, B.: Stochastic Differential Equations. Universitext. Springer, Berlin (1985)
13. Stroock, D.W., Varadhan, S.R.S.: On the support of diffusion processes with application to the strong maximum principle. In: Proceedings of Sixth Berkeley Symp Math Statist Prob III, pp. 333–359. Univ. California Press, Berkeley (1972)
14. Yamada, T., Ogura, Y.: On the strong comparison theorem for solutions of stochastic differential equations. Z Wharsch verw Gebiete 56, 3–19 (1981)

Strong Solution of Set-Valued Stochastic Differential Equation

Jungang Li and Shoumei Li*

Department of Applied Mathematics, Beijing University of Technology, Beijing, P.R. China

Abstract. In this paper, we shall firstly illustrate why we should introduce the Itô type set-valued stochastic differential equation. Then we shall recall the Lebesgue integral of a set-valued stochastic process with respect to the time t and discuss its some properties. We shall also obtain the theorem of existence and uniqueness of solution of Itô type set-valued stochastic differential equation.

1 Introduction

It is well-known that the classical Itô type stochastic differential equation (cf. [12]) is

$$\begin{cases} dx_t = b(t,x_t)dt + \sigma(t,x_t)dB_t \\ x_0 = \eta \end{cases}$$

or the stochastic integral form

$$x_t = \eta + \int_0^t b(s,x_s)ds + \int_0^t \sigma(s,x_s)dB_s,$$

where $(B_t)_{0\le t\le T}$ is m-dimensional Brownian motion, $b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T]\times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are Borel measurable functions with some conditions. It describes the movement law of an object with random disturbing given the initial condition. It has been widely used in the stochastic control (e.g. [29]) and financial mathematics (e.g. [5], [7], [14]). In complex systems, however, we can not decide the exactly x_t at time t but we can know it takes values, for instance, in some interval $[x_1,x_2]$, $x_1 < x_2$. This becomes a set-valued stochastic differential equation as follows:

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \tag{1}$$

where $b(t,X_t)$ and $\sigma(t,X_t)$ take the closed subsets of $\mathbb{R}^d$ and $\mathbb{R}^{d\times m}$ respectively. Equation (1) can be written as the following set-valued stochastic integral form

$$X_t = X_0 + \int_0^t b(t,X_t)dt + \int_0^t \sigma(t,X_t)dB_t.$$

* Research partially supported by NSFC(No.10771010), Research Fund of Beijing Educational Committee, PHR(IHLB) and 111 Talent Project Fund of BJUT, P.R. China.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 271–277, 2008.
springerlink.com

To solve the above set-valued differential equation, we have to consider two kinds of integrals of a set-valued stochastic process: one is the integral of a set-valued stochastic process with respect to time t, and the other is the Itô integral of a set-valued stochastic process with respect to a Brownian motion B_t. In [15], Kim used the definition of stochastic integral of set-valued stochastic process with respect to the Brownian motion introduced by Kisielewicz in [16] and discussed its properties. We called it Aumann type integral since the idea came from Aumann integral for a set-valued random variable [2]. In [13], Jung and Kim gave a new definition with basic space being $\mathbb{R}$ by taking fixed time t. It may be more suitable to treat a set-valued stochastic process as a whole. In [23], Li and Ren introduced a new way to define the Itô integral of set-valued stochastic processes and discussed their properties. Concerning the integral with respect to the time t, Kisielewicz also introduced Aumann type integral in [17]. There are many good related former works such as [6], [8], [17]–[20], [22], [27], [28], [30] and so on.

We also can extend above set-valued case to fuzzy set-valued stochastic differential equation without difficulty by using level set method. There is another way to discuss fuzzy set-valued stochastic differential equation. In [10], Hu et al. used Hukuhara difference to define the differentiability and to discuss the Itô type fuzzy stochastic differential equations in the special case $\sigma(t,X_t) \in \mathbb{R}^{d\times m}$. But since it is well-known that the space of all closed subsets of some metric space is not linear with resect to the addition and scalar multiplication, it leds to a big problem: under what conditions does the Hukuhara difference exist? It is a difficult problem so that they simply assumed that the Hukuhara difference of a stochastic process at any two times always exists. In this paper, we use selection method to consider the same type problem as in [10] without using the Hukuhara difference. We shall consider the Itô type set-valued stochastic integral equation, discuss the existence and uniqueness of its solution. We would like to mention that.

We organize our paper as follows: in Section 2, we shall introduce some necessary notations, definitions and results about set-valued stochastic processes and set-valued stochastic Lebesgue integral. In Section 3, we shall give Itô type set-valued stochastic differential equation, and then prove the theorem of existence and uniqueness of the solution to the equation.

2 Lebesgue Integral of a Set-Valued Stochastic Process and Its Properties

Throughout this paper, assume that $(\Omega,\mathscr{A},\mu)$ is a complete atomless probability space, the σ-field filtration $\{\mathscr{A}_t : t \in I\}$ satisfies the usual conditions (i.e. containing all null sets, non-decreasing and right continuous), $I=[0,T]$ with $T>0$, $\mathbb{R}$ is the set of all real numbers, $\mathbb{N}$ is the set of all natural numbers, $\mathbb{R}^d$ is the d-dimensional Euclidean space with usual norm $\|\cdot\|$, $\mathscr{B}(E)$ is the Borel field of the space E. Let $f=\{f(t),\mathscr{A}_t : t\in I\}$ be a $\mathbb{R}^d$-valued adapted stochastic process.

It is said that f is progressively measurable if for any $t\in I$, the mapping $(s,\omega)\mapsto f(s,\omega)$ from $[0,t]\times\Omega$ to $\mathbb{R}^d$ is $\mathscr{B}([0,t])\times\mathscr{A}_t$-measurable. Each right continuous (left continuous) adapted process is progressively measurable.

Assume that $\mathscr{L}^p(\mathbb{R}^d)$ denotes the set of $\mathbb{R}^d$-valued stochastic processes $f = \{f(t), \mathscr{A}_t : t \in I\}$ such that f satisfying (a) f is progressive, and (b)

$$|||f|||_p = \left[E\left(\int_0^T \|f(t,\omega)\|^p ds\right)\right]^{1/p} < \infty.$$

Let $f, f' \in \mathscr{L}^p(\mathbb{R}^d)$, $f = f'$ if and only if $|||f - f'|||_p = 0$. Then $(\mathscr{L}^p(\mathbb{R}^d), |||\cdot|||_p)$ is complete.

Now we review notation and concepts of set-valued stochastic processes. Assume that $\mathbf{K}(\mathbb{R}^d)$ is the family of all nonempty, closed subsets of $\mathbb{R}^d$, and $\mathbf{K}_c(\mathbb{R}^d)$ (*resp.* $\mathbf{K}_k(\mathbb{R}^d)$, $\mathbf{K}_{kc}(\mathbb{R}^d)$) is the family of all nonempty closed convex (*resp.* compact, compact convex) subsets of $\mathbb{R}^d$. The Hausdorff metric between two closed sets A, B may take infinite when they are unbounded. But it is known (cf. [25, Theorem 1.1.2]) that $\mathbf{K}_k(\mathbb{R}^d)$ is a complete space with respect to d_H, and $\mathbf{K}_{kc}(\mathbb{R}^d)$ is its closed subset. For $B \in \mathbf{K}(\mathbb{R}^d)$, define $\|B\|_{\mathbf{K}} = d_H(\{0\}, B) = \sup_{a\in B}\|a\|$. For more topological results and analysis results on closed subset spaces, readers may refer to Beer's book [3], Aubin and Frankowsk's book [1] respectively.

For a set-valued random variable F (cf. [4], [9], [25]), define the set

$$S_F^p = \{f \in L^p[\Omega, \mathbb{R}^d] : f(\omega) \in F(\omega)\ a.e.(\mu)\},$$

where $L^p[\Omega;\mathbb{R}^d]$ is the set of all $\mathbb{R}^d$-valued random variables f such that $\|f\|_p = [E(\|f\|^p)]^{1/p} < \infty$, and constant $p \geq 1$. The Aumann integral of F is defined as $E[F] = \{E[f] : f \in S_F^1\}$. It was introduced by Aumann in 1965 (cf. [2]). A set-valued random variable $F : \Omega \to \mathbf{K}(\mathbb{R}^d)$ is called *integrable* if S_F^1 is non-empty. F is called *integrable bounded* if $\int_\Omega \|F(\omega)\|_{\mathbf{K}} d\mu < \infty$. Let $L^p[\Omega;\mathbf{K}(\mathbb{R}^d)]$ (resp. $L^p[\Omega;\mathbf{K}_c(\mathbb{R}^d)]$, $L^p[\Omega;\mathbf{K}_{kc}(\mathbb{R}^d)]$) denote the family of $\mathbf{K}(\mathbb{R}^d)$-valued (resp. $\mathbf{K}_c(\mathbb{R}^d)$, $\mathbf{K}_{kc}(\mathbb{R}^d)$-valued) L^p-bounded random variables F such that $\|F(\omega)\|_{\mathbf{K}} \in L^p[\Omega;\mathbb{R}]$. Concerning more notation, definitions and results of set-valued random variables, readers could refer to the excellent paper [9] or the book [25].

$F = \{F(t) : t \in I\}$ is called *a set-valued process* if $F : I \times \Omega \to \mathbf{K}(\mathbb{R}^d)$ is a set-valued function such that for any fixed $t \in I$, $F(t,\cdot)$ is a set-valued random variable. A set-valued process $F = \{F(t) : t \in I\}$ is called *adapted with respect to the filtration* $\{\mathscr{A}_t : t \in I\}$, if $F(t)$ is measurable with respect to $\mathscr{A}_t$ for each $t \in I$, and denoted by $\{F(t), \mathscr{A}_t : t \in I\}$.

Definition 1. *A set-valued stochastic process $F = \{F(t), \mathscr{A}_t : t \in I\}$ is called to be* progressively measurable*, if for any $t \in I$, the mapping $(s,\omega) \mapsto F(s,\omega)$ from $[0,t] \times \Omega$ to $\mathbf{K}(\mathbb{R}^d)$ is $\mathscr{B}([0,t]) \times \mathscr{A}_t$-measurable, i.e. for any $A \in \mathscr{B}(\mathbb{R}^d)$, $\{(s,\omega) \in [0,t] \times \Omega : F(s,\omega) \cap A \neq \emptyset\} \in \mathscr{B}([0,t]) \times \mathscr{A}_t$.*

If F is progressively measurable then F is adapted.

Definition 2. *A progressively measurable set-valued stochastic process $F = \{F(t), \mathscr{A}_t : t \in I\}$ is called $\mathscr{L}^p$-bounded, if the real stochastic process $\{\|F(t)\|_{\mathbf{K}}, \mathscr{A}_t : t \in I\} \in \mathscr{L}^p(\mathbb{R})$.*

Let $\mathscr{L}^p(\mathbf{K}(\mathbb{R}^d))$ denote the set of all $\mathscr{L}^p$-bounded progressively measurable $\mathbf{K}(\mathbb{R}^d)$-valued stochastic process. Similarly, we have notations $\mathscr{L}^p(\mathbf{K}_c(\mathbb{R}^d))$, $\mathscr{L}^p(\mathbf{K}_k(\mathbb{R}^d))$ and $\mathscr{L}^p(\mathbf{K}_{kc}(\mathbb{R}^d))$.

Definition 3. *A $\mathbb{R}^d$-valued process $\{f(t), \mathscr{A}_t : t \in I\} \in \mathscr{L}^p(\mathbb{R}^d)$ is called an $\mathscr{L}^p$-selection of $F = \{F(t), \mathscr{A}_t : t \in I\}$ if $f(t,\omega) \in F(t,\omega)$ a.e.$(t,\omega) \in I \times \Omega$.*

Let $S^p(\{F(\cdot)\})$ or $S^p(F)$ denote the family of all $\mathscr{L}^p(\mathbb{R}^d)$-selections of $F = \{F(t), \mathscr{A}_t : t \in I\}$, i.e.

$$S^p(F) = \left\{\{f(t)\} \in \mathscr{L}^p(\mathbb{R}^d) : f(t,\omega) \in F(t,\omega), a.e.\ (t,\omega) \in I \times \Omega\right\}.$$

Concerning more properties of stochastic processes F and $S^p(F)$, readers may refer to [30].

Now we consider the integral of set-valued stochastic process with respect to the time t. Kisielewicz gave the following definition in [17].

Definition 4. *Suppose $F = \{F(t) : t \in I\} \in \mathscr{L}^p(\mathbf{K}_k(\mathbb{R}^d))$ $(1 \le p < +\infty)$ is a set-valued stochastic process. For any $\omega \in \Omega$, $t \ge 0$, define*

$$\int_0^t F(s,\omega)ds := \left\{\int_0^t f(s,\omega)ds : f \in S^p(F)\right\},$$

where $\int_0^t f(s,\omega)ds$ is the Lebesgue integral, $\int_0^t F(s,\omega)ds$ is called the Aumann type Lebesgue integral of set-valued stochastic process F with respect to time t. For any $0 \le u < t < T$,

$$\int_u^t F(s,\omega)ds := \int_0^t I_{[u,t]}(s)F(s,\omega)ds. \tag{2}$$

Remark 1

(1) We would like to thank professor Yukio Ogura who mentioned us that there is a delicate problem in above definition to deal with the almost everywhere problem in general case. As a matter of fact, take an $f \in \mathscr{L}^p(\mathbb{R}^d)$, then, by Fubini Theorem, for all $\omega \in \Omega$, the mapping $f(\cdot, \omega)$ is $\mathscr{B}([0,t])$-measurable and for almost every $\omega \in \Omega$ (NOT every!), we have that

$$I_t(f)(\omega) = \int_0^t f(s, \omega)ds < \infty.$$

Since the set $S^p(F)$ is uncountable in general so that (2) can not be well-defined for almost every $\omega \in \Omega$! How to deal with this problem in general, readers may refer to our recent paper [24].

If $\mathscr{A}$ is μ-separable, then $L^p[I \times \Omega, \mathscr{B}([0,1]) \times \mathscr{A}, \lambda \times \mu; \mathbb{R}^d]$ is separable. Since $S^p(F)$ is a closed subset of $L^p[I \times \Omega, \mathscr{B}([0,1]) \times \mathscr{A}, \lambda \times \mu; \mathbb{R}^d]$, $S^p(F)$ is separable so that (2) can be well-defined for almost every $\omega \in \Omega$. For simplification, we ignore this null set and assume that (2) is well-defined for each $\omega \in \Omega$ in the rest of paper.

(2) Here we choose that the set of selections is $S^p(F)$. As a matter of fact, if we only consider the Lebesgue integral, we can use $S^1(F)$. But we often consider the sum of integral of a set-valued stochastic process with respect to time t and integral of a set-valued stochastic process with respect to Brownian motion, where we have to use $S^2(F)$. Thus we here use $S^p(F)$ for more general case.

Theorem 1. *Let set-valued stochastic process $F \in \mathscr{L}^p(\mathbf{K}_k(\mathbb{R}^d))$. Then for any $t \in I$, a set-valued mapping L_t defined by*

$$L_t(\omega) = \int_0^t F(s,\omega)ds \quad (\omega \in \Omega)$$

is a compact convex set-valued random variable and

$$L_t(\omega) = \int_0^t \mathrm{co}F(s,\omega)ds$$

Theorem 2. *Let $F = \{F_t : t \in I\} \in \mathscr{L}^p(\mathbf{K}_k(\mathbb{R}^d))$, then there exists a sequence of $\mathbb{R}^d$-valued stochastic processes $\{f^i = \{f^i(t) : t \in I\} : i \geq 1\} \subset S^p(F)$ such that*

$$F(t,\omega) = \mathrm{cl}\{f^i(t,\omega) : i \geq 1\}, \quad a.e.\ (t,\omega) \in I \times \Omega,$$

and

$$L_t(\omega) = \mathrm{cl}\Big\{\int_0^t f^i(s,\omega)ds : i \geq 1\Big\} \quad a.e.\ (t,\omega) \in I \times \Omega.$$

Theorem 3. *If a set-valued stochastic process $\{F_t, \mathscr{A}_t : t \in I\} \in \mathscr{L}^2(\mathbf{K}_k(\mathbb{R}^d))$, then for $0 \leq t_1 < t \leq T$, we have*

$$L_t(\omega) = \mathrm{cl}\{L_{t_1}(\omega) + \int_{t_1}^t F(s,\omega)ds\}, a.s.$$

where the closure is taken in $\mathbb{R}^d$.

3 The Itô Type Set-Valued Stochastic Differential Equation

The Itô type set-valued stochastic differential equation

$$dF(t) = f(t,F(t))dt + g(t,F(t))dB_t \tag{3}$$

where $F(t)$ is a set-valued stochastic process and $F \in \mathscr{L}^2(\mathbf{K}(\mathbb{R}^d))$, $f : I \times \mathbf{K}_k(\mathbb{R}^d) \to \mathbf{K}_k(\mathbb{R}^d)$ is measurable, $g : I \times \mathbf{K}_k(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ is measurable. If f is integrable and g is square integrable, then equation (3) is equivalent to the integrable form:

$$F(t) = F(0) + \int_0^t f(s,F(s))ds + \int_0^t g(s,F(s))dB_s. \tag{4}$$

Theorem 4. ***(existence and uniqueness)*** *Assume that $f(t,F), g(t,F)(t \in I, F \in \mathbf{K}_k(\mathbb{R}^d))$ satisfy the following conditions:*

(i) Linear condition: there exists a positive constant K_1 such that for any $t \in I, F \in \mathbf{K}_k(\mathbb{R}^d)$,

$$\|f(t,F)\|^2_{\mathbf{K}} + |g(t,F)|^2 \leq K_1^2(1 + \|F\|^2_{\mathbf{K}});$$

(ii) Lipschitz continuous condition: there exists a positive constant K_2 such that for any $t \in I, F_1, F_2 \in \mathbf{K}(\mathbb{R}^d)$

$$d_H(f(t,F_1),f(t,F_2)) + |g(t,F_1) - g(t,F_2)| \leq K_2 d_H(F_1,F_2),$$

Then there is a strong solution to the equation (3), and the solution is unique.

To prove above theorem, we need the following important inequality formula.

Theorem 5. *Suppose set-valued stochastic process $F = \{F(t) : t \in I\}$, $G = \{G(t) : t \in I\} \in \mathscr{L}^2(\mathbf{K}_k(\mathbb{R}^d))$, then a.e. for any $t \in I$*

$$Ed_H^2\Big(\int_0^t F(s)ds, \int_0^t G(s)ds\Big) \leq tE\int_0^t d_H^2(F_s,G_s)ds,$$

especially, we have

$$E\Big[\Big\|\int_0^t F(s)ds\Big\|_{\mathbf{K}}^2\Big] \leq tE\Big[\int_0^t \|F(s)\|_{\mathbf{K}}^2 ds\Big].$$

Since page limitation, we omit all the proofs of the theorems in this paper.

4 Conclusions

In this paper, we firstly stated the definition of the Lebesgue integral of a set-valued stochastic process with respect to time t based on the former works such as Kisielewicz [16], Kisielewicz, Michta and Motyl [19]–[20]. And then we discussed some properties of set-valued Lebesgue integral, especially we proved the presentation theorem of set-valued stochastic integral. We obtained the inequality about the the Lebesgue integrals of set-valued stochastic processes. By using this inequality with other properties of integrals, we proved the theorem of existence and uniqueness of solution of Itô type set-valued stochastic differential equation.

References

1. Aubin, J.P., Frankowska, H.: Set-Valued Analysis. Systems & Control: Foundations & Applications, vol. 2. Birkhäuser Boston, Inc., Boston (1990)
2. Aumann, R.: Integrals of set valued functions. J. Math. Anal. Appl. 12, 1–12 (1965)
3. Beer, G.: Topologies on Closed and Closed Convex Sets. Mathematics and Its Applications, vol. 268. Kluwer Academic Publishers Group, Dordrecht (1993)
4. Castaing, C., Valadier, M.: Convex Analysis and Measurable Multifunctions. Lect Notes in Math, vol. 580. Springer, Berlin, New York (1977)
5. Chen, Z., Kulperger, R.: Minimax pricing and Choquet pricing. Insurance Math. Econom. 38(3), 518–528 (2006)
6. Da Prato, G., Frankowska, H.: A stochastic Filippov theorem. Stochastic Anal. Appl. 12, 409–426 (1994)
7. Etheridge, A.: A Course in Financial Calculus. Cambridge University Press, Cambridge (2002)

8. Hiai, F.: Multivalued stochastic integrals and stochastic differential inclusions (unpublished manuscript) (2008)
9. Hiai, F., Umegaki, H.: Integrals, conditional expectations and martingales of multivalued functions. J. Multivariate Anal. 7, 149–182 (1977)
10. Hu, L., Zhao, W., Feng, Y.: Fuzzy stochastic differential equations of the Itô-type. Chinese J. Eng. Math. 1, 52–62 (2006)
11. Hu, S., Papageorgiou, N.S.: Handbook of Multivalued Analysis. Kluwer Academic Publishers, Dordrecht (1997)
12. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. Noth-Holland/Kodansha, Amsterdam-Tokyo (1981)
13. Jung, E.J., Kim, J.H.: On set-valued stochastic integrals. Stochastic Anal. Appl. 21, 401–418 (2003)
14. Karatzas, I.: Lectures on the Mathematics of Finance. CRM Monograph Series, vol. 8. American Mathematical Society, Providence (1997)
15. Kim, B.K., Kim, J.H.: Stochastic integrals of set-valued processes and fuzzy processes. J. Math. Anal. Appl. 236, 480–502 (1999)
16. Kisielewicz, M.: Properties of solution set of stochastic inclusions. J. Appl. Math. Stochastic Anal. 6, 217–236 (1993a)
17. Kisielewicz, M.: Set valued stochastic integrals and stochastic inclusions. Discuss Math. 13, 119–126 (1993b)
18. Kisielewicz, M.: Set-valued stochastic integrals and stochastic inclusions. Stochastic Anal. Appl. 15, 783–800 (1997)
19. Kisielewicz, M., Michta, M., Motyl, J.: Set valued approach to stochastic control, Part I: Existence and regularity properties. Dynam. Systems Appl. 12, 405–432 (2003)
20. Kisielewicz, M., Michta, M., Motyl, J.: Set valued approach to stochastic control, Part II: Viability and semimartingale issues. Dynam. Systems Appl. 12, 433–466 (2003)
21. Klein, E., Thompson, A.C.: Theory of Correspondences Including Applications to Mathematical Economics. John Wiley & Sons, New York (1984)
22. Li, J., Li, S.: Set-valued stochastic Lebesgue integral and representation theorems. Int. J. Comput. Intell. Syst.(to appear, 2008)
23. Li, S., Ren, A.: Representation theorems, set-valued and fuzzy set-valued Itô intergal. Fuzzy Sets Syst. 158, 949–962 (2007)
24. Li, J., Li, S., Ogura, Y.: Strong solution of Itô type set-valued stochastic differential equation (unpublished manuscript) (2008)
25. Li, S., Ogura, Y., Kreinovich, V.: Limit Theorems and Applications of Set-Valuded and Fuzzy Sets-Valued Random Variables. Kluwer Academic Publishers, Dordrecht (2002)
26. Li, S., Ogura, Y., Proske, F.N., Puri, M.L.: Central limit theorems for generalized set-valued random variables. J. Math. Anal. Appl. 285, 250–263 (2003)
27. Motyl, J.: Existence of Solutions of set-valued Itô equation. Bull. Acad. Pol. Sci. 46, 419–430 (1998)
28. Ogura, Y.: On stochastic differential equations with set coefficients and the Black-Scholes model. In: Proceedings of the 8th International Conference on Intelligent Technologies (InTech 2007, Sydney, Australia), pp. 300–304 (2007)
29. Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications, 4th edn. Springer, Berlin (1995)
30. Zhang, W., Li, S., Wang, Z., Gao, Y.: Set-valued stochastic differential integrals and differential inclusions. In: An Introduction of Set-Valued Stochastic Processes, ch. 8, Sinence Press, Beijing/New York (2007)

Convergence Results

Convergences of Random Variables with Respect to Coherent Upper Probabilities Defined by Hausdorff Outer Measures

Serena Doria

Department of Geotechnologies, University G. d'Annunzio, Chieti, Italy

Abstract. Given a non-empty set Ω and a partition $\boldsymbol{B}$ of Ω let $\boldsymbol{L}$ be the class of all subsets of Ω. Upper conditional probabilities $\overline{P}(A|B)$ are defined on $\boldsymbol{L} \times \boldsymbol{B}$ by a class of Hausdorff outer measures when the conditioning event B has positive and finite Hausdorff measure in its dimension; otherwise they are defined by a 0-1 valued finitely additive (but not countably additive) probability. The unconditional upper probability is obtained as a particular case when the conditioning event is Ω. Relations among different types of convergence of sequences of random variables are investigated with respect to this upper probability. If Ω has finite and positive Hausdorff outer measure in its dimension the given upper probability is continuous from above on the Borel σ-field. In this case we obtain that the pointwise convergence implies the μ-stochastic convergence. Moreover, since the outer measure is subadditive then stochastic convergence with respect to the given upper probability implies convergence in μ-distribution.

1 Introduction

One of the main topics in probability theory and stochastic processes theory is the convergence of sequences of random variables, which plays a key role in asymptotic inference. Different kinds of convergence and their relations are considered in literature ([1]) when all the random variables are defined on the same probability space: convergence with probability 1 or strong convergence and convergence in probability or weak convergence. It is well known that convergence with probability 1 implies convergence in probability but the converse in not true (see for example Billingsley [1] p. 274 and p. 340). These convergences are used respectively for the strong law and the weak law of large numbers. In statistics, if a sequence of statistics converges in probability to the population value as the sample size goes to infinity according to the weak law of large numbers, the statistic is called *consistent*. Convergence in probability implies convergence in distribution, another type of convergence used in the central limit theorem. Moreover the convergence theorems are important because they yield sufficient conditions for the integral to be interchanged with pointwise convergence of random variables.

In Denneberg [3] these different types of convergence of sequences of random variables are considered with respect to a monotone set function instead of a probability measure. To obtain the same relations among different types of convergence some other properties are required for the monotone set function. In particular it has been proven that pointwise convergence (that is a particular case of convergence with probability 1)

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 281–288, 2008.
springerlink.com

of a sequence of random variables to a random variable X implies the μ-stochastic convergence (convergence in probability if μ is a probability measure) if either μ is continuous from above or the convergence is uniform. If μ is a subadditive monotone set function then μ-stochastic convergence implies convergence in μ-distribution. In [2] μ-stochastic convergence plus the uniform integrability have been proven to imply the convergence in mean for monotone, subadditive, normalized continuous from above set fnctions; moreover the convergence in mean of a sequence of Borel measurable random variables has been proven to imply the μ-stochastic convergence. In this paper relations among different types of convergences with respect to upper probability defined by Hausdorff outer measures are investigated. The necessity to introduce Hausdorff (outer) measures as new tool to assess (upper) conditional probability is due to some problems related to the axiomatic definition of regular conditional probability ([6]). In fact every time that the σ-field of the conditioning events is not countably generated, conditional probability, defined by the Radon-Nikodym derivative is not separately coherent as required in Walley's approach ([7]).

2 Convergences of Random Variables with Respect to a Monotone Set Function

In [3] different kinds of convergence of a sequence of random variables with respect to a monotone set function are introduced and their relations have been proven.

Given a non-empty set Ω and denoted by S a set system, containing the empty set and properly contained in $\mathrm{P}(\Omega)$, the family of all subsets of Ω, a monotone set function $\mu : S \to \overline{\mathbb{R}}_+$ is such that $\mu(\oslash) = 0$ and if $A, B \in S$ with $A \subset B$ then $\mu(A) \leq \mu(B)$. A monotone set function on S is *continuous from below* if for each non decreasing sequence of sets A_n of S such that $A = \bigcup_{n=1}^{\infty} A_n$ belongs to S we have $\lim_{n\to\infty} \mu(A_n) = \mu(A)$. A monotone set function on S is *continuous from above* if for each non increasing sequence of sets A_n of S such that $A = \bigcap_{n=1}^{\infty} A_n$ belongs to S which is closed under set difference then the following properties are equivalent

i) μ is continuous from below;
ii) μ is continuous from above;
iii) μ is continuous from above at the empty set.

If S is a σ-field then μ is σ-additive if and only if it is additive and continuous from below.

Given a monotone set function μ on S the *outer set function* of μ is the set function defined on the whole power set $\mathrm{P}(\Omega)$ by

$$\mu^*(A) = \inf\{\mu(B) \,|\, A \subset B \in S\}, \quad A \in \mathrm{P}(\Omega).$$

The inner set function of μ is the set function defined on the whole power set $\mathrm{P}(\Omega)$ by

$$\mu_*(A) = \sup\{\mu(B) \,|\, B \subset A; B \in S\}, \quad A \in \mathrm{P}(\Omega).$$

Examples of outer set functions or outer measures are the Hausdorff outer measures.

Let (Ω,d) be the Euclidean metric space with Ω= [0,1]. The diameter of a non-empty set U of Ω is defined as $|U| = \sup\{|x-y| : x,y \in U\}$ and if a subset A of Ω is such that $A \subset \bigcup_i U_i$ and $0 < |U_i| < \delta$ for each i, the class $\{U_i\}$ is called a δ-cover of A. Let s be a non-negative number. For $\delta > 0$ we define $h_{s,\delta}(A) = \inf \sum_{i=1}^{\infty} |U_i|^s$, where the infimum is over all δ-covers $\{U_i\}$.

The Hausdorff s-dimensional outer measure of A, denoted by $h^s(A)$, is defined as $h^s(A) = \lim_{\delta \to 0} h_{s,\delta}(A)$. This limit exists, but may be infinite, since $h_{s,\delta}(A)$ increases as δ decreases. The *Hausdorff dimension* of a set A, $\dim_H(A)$, is defined as the unique value, such that

$$h^s(A) = \infty \text{ if } 0 \le s < \dim_H(A), \qquad h^s(A) = 0 \text{ if } \dim_H(A) < s < \infty.$$

We can observe that if $0 < h^s(A) < \infty$ then $\dim_H(A) = s$, but the converse is not true. We assume that the Hausdorff dimension of the empty set is equal to -1 so no event has Hausdorff dimension equal to the empty set. If an event A is such that $\dim_H(A) = s < 1$ than the Hausdorff dimension of the complementary set is equal to 1 since the following relation holds: $\dim_H(A \cup B) = \max\{\dim_H(A), \dim_H(B)\}$.

A subset A of Ω is called *measurable* with respect to the outer measure h^s if it decomposes every subset of Ω additively, that is if $h^s(E) = h^s(AE) + h^s(E - A)$ for all sets $E \subseteq \Omega$.

An important property of Hausdorff outer measures is that they are *regular* ([5, Theorem 1.6]), that is for every set A there is an h^s-measurable set E containing A with $h^s(E) = h^s(A)$ and so they are continuous from below ([5, Lemma 1.3]). The restriction of h^s to the σ-field of h^s-measurable sets, containing the σ-field of the Borel sets ([5, Theorem 1.5]), is called Hausdorff s-dimensional measure. In particular the Hausdorff 0-dimensional measure is the counting measure and the Hausdorff 1-dimensional measure is the Lebesgue measure.

We recall the definitions of different types of convergence with respect to a monotone set function and their implications given in [3]. Let μ be a monotone set function defined on S properly contained in P(Ω) and $X : \Omega \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ an arbitrary function on Ω then the set function $G_{\mu,X}(x) = \mu\{\omega \in \Omega : X(\omega) > x\}$ is decreasing and it is called *decreasing distribution function* of X with respect to μ. If μ is continuous from below then $G_{\mu,X}(x)$ is right continuous. In particular the decreasing distribution function of X with respect to the Hausdorff outer measures is right continuous since these outer measures are continuous from below.

A function $X : \Omega \to \overline{\mathbb{R}}$ is called upper μ-measurable if $G_{\mu^*,X}(x) = G_{\mu_*,X}(x)$.

Given a monotone set function μ defined on a field S and a sequence $X_n : \Omega \to \overline{\mathbb{R}}$ of upper μ-measurable functions we say that X_n *converges in μ-distribution* to an upper μ measurable function $X : \Omega \to \overline{\mathbb{R}}$ if $\lim_{n\to\infty} G_{\mu,X_n} = G_{\mu,X}$ except at on at most countable set. Since a monotone set function is continuous except on an at most countable set the previous condition is equivalent to $\lim_{n\to\infty} G_{\mu,X_n}(x) = G_{\mu,X}(x)$ for all continuity points x of $G_{\mu,X}$. A sequence of random variables X_n *converges μ-stochastically* (or converges in probability if μ is a probability) to a random variable X if $|X_n - X|$ converges in μ^*-distribution to the null function $G_{\mu,0}(x)$, where $G_{\mu,0}(x) = \mu(\Omega)$ if $x \le 0$ and $G_{\mu,0}(x) = 0$ if $x > 0$. If μ is monotone and subadditive and X_n converges μ-stochastically to X then X_n converges in μ-distribution to X (Proposition 8.5 of [3]). If μ is monotone

and continuous from above and X_n converges pointwise to X then X_n converges μ-stochastically to X (Proposition 8.8 of [3]).

Given an upper μ-measurable function $X : \Omega \to \overline{\mathbb{R}}$ with decreasing distribution function $G_{\mu,X}(x)$, the Choquet integral of X with respect to μ is defined if $\mu(\Omega) < \infty$ through

$$\int X d\mu = \int_{-\infty}^{0} (G_{\mu,X}(x) - \mu(\Omega))dx + \int_{0}^{\infty} G_{\mu,X}(x)dx$$

The integral is in $\mathbb{R}$ or can assume the values $-\infty$, ∞ and 'non-existing'. If $X \geq 0$ or $X \leq 0$ the integral always exists.

Let μ be a monotone set function and let X_n be a sequence of random variables such that $Y \leq X_n \leq Z$ for every $n \in \mathbb{N}$ and Y and Z have finite Choquet integral with respect to μ; if X_n converges in μ-distribution to X then $\lim_{n\to\infty} \int X_n d\mu = \int X d\mu$ (General Dominated Convergence Theorem, Proposition 8.9 of [3]).

3 Upper Conditional Previsions Defined with Respect to Hausdorff Outer Measures

Given a non-empty set Ω, a gamble X is a bounded function from Ω to R (the set of real numbers) and let $\boldsymbol{L}$ be the set of all gambles on Ω. When $\boldsymbol{K}$ is a linear space of gambles a coherent upper prevision is a real function $\overline{P}$ defined on $\boldsymbol{K}$, such that the following conditions hold for every X and Y in $\boldsymbol{K}$:

1) $\overline{P}(X) \leq \sup(X)$;
2) $\overline{P}(\lambda X) = \lambda \overline{P}$ for each positive constant λ;
3) $\overline{P}(X+Y) \leq \overline{P}(X) + \overline{P}(Y)$.

Suppose that $\overline{P}$ is an upper prevision defined on a linear space $\boldsymbol{K}$, its conjugate lower prevision $\underline{P}$ is defined on the same domain $\boldsymbol{K}$ by $\underline{P}(-X) = -\overline{P}(X)$.

When $\boldsymbol{K}$ is an arbitrary class of events, that can be regarded as a class of 0-1 valued gambles then $\overline{P}(X)$ is called an upper coherent probability and $\underline{P}(X)$ is a lower coherent probability. If for every X belonging to $\boldsymbol{K}$ we have $P(X) = \overline{P}(X) = \underline{P}(X)$, then P is called a linear prevision.

We recall the notion of coherent conditional upper prevision.

Let $\boldsymbol{B}$ denote a partition of Ω, which is a non-empty, pair wise-disjoint subsets whose union is Ω. For B in $\boldsymbol{B}$ let $\boldsymbol{H}(B)$ be the set of all random variables defined on B. An upper conditional prevision $\overline{P}(X|B)$ is a real function defined on $\boldsymbol{H}(B)$. Upper conditional previsions $\overline{P}(X|B)$, defined for B in $\boldsymbol{B}$ and X in $\boldsymbol{H}(B)$ are required ([7]) to be *separately coherent*, that is for every conditioning event B $\overline{P}(\cdot|B)$ is a coherent upper prevision on the domain $\boldsymbol{H}(B)$ and $\overline{P}(B|B) = 1$. Moreover the upper conditional previsions $\overline{P}(X|B)$ for B in $\boldsymbol{B}$ are required to be coherent with the unconditional previsions $\overline{P}$. Upper and lower conditional probabilities are particular kinds of upper and lower conditional previsions obtained when $\boldsymbol{H}(B)$ is a class of events.

In [4] upper conditional previsions for bounded random variables are defined by the Choquet integral with respect to the Hausdorff outer measure if the conditioning event has positive and finite Hausdorff outer measure in its dimension; otherwise, they are defined by a 0-1 valued finitely, but not countably, additive probability.

Definition 1. *Let Ω be a non-empty set and let $\boldsymbol{B}$ be a partition of Ω. Denote by s the Hausdorff dimension of the conditioning event B and by h^s the Hausdorff s-dimensional outer measure. Let $\boldsymbol{L}$ be the class of all bounded random variables on Ω and let m be a 0-1 valued finitely additive, but not countably additive, probability. Then upper conditional previsions are defined on $\boldsymbol{L} \times \boldsymbol{B}$ by the functions*

$$\overline{P}(X|B) = \frac{1}{h^s(B)} \int_B X dh^s \quad \text{if } 0 < h^s(B) < \infty$$

and by

$$\overline{P}(X|B) = m(XB) \quad \text{if } h^s(B) = 0, \infty$$

Upper conditional probabilities can be obtained from the previous definition in the case where $\boldsymbol{L}$ is the class of all subsets of Ω.

Theorem 1. *Let Ω be an non-empty set with finite and positive Hausdorff measure in its dimension, let $\boldsymbol{B}$ be a partition of Ω. Denote by s the Hausdorff dimension of the conditioning event B and by h^s the Hausdorff s-dimensional measure. Let $\boldsymbol{F}$ be the class of all subsets of Ω. Moreover, let m be a 0-1 valued finitely additive, but not countably additive, probability. Then the function defined on $\boldsymbol{F} \times \boldsymbol{B}$ by*

$$\overline{P}(A|B) = \frac{h^s(AB)}{h^s(B)} \quad \text{if } 0 < h^s(B) < \infty$$

and by

$$\overline{P}(A|B) = m(AB) \quad \text{if } h^s(B) = 0, \infty$$

is an upper conditional probability.

In the next theorem we prove that when Ω has finite and positive Hausdorff outer measure in its dimension and upper probability is defined as in Theorem 1, then μ-stochastic convergence implies convergence in μ-distribution.

Theorem 2. *Let Ω be an non-empty set with finite and positive Hausdorff measure in its dimension and let $\mu = \overline{P}(A|\Omega)$ the upper probability defined as in Theorem 1. Let X_n be a sequence of random variables on Ω; if X_n converges to a random variable X μ-stochastically then X_n converges in μ-distribution to X.*

Proof. Since Ω has finite and positive Hausdorff measure in its dimension we have $\mu = \overline{P}(A|\Omega) = \frac{h^s(A)}{h^s(\Omega)}$. Moreover every outer Hausdorff measure is subadditive so we obtain that μ-stochastic convergence implies convergence in μ-distribution. □

In probability theory convergence in distribution has been proven to be equivalent to the pointwise convergence of the expectations functionals on all bounded continuous functions (see for example Theorem 29.1 of [1]). We have that this equivalence remains valid when upper probabilities are defined with respect to Hausdorff outer measures.

Definition 2. *Let Ω be an non-empty set with finite and positive Hausdorff outer measure in its dimension and let $\mu = \overline{P}(A|\Omega)$ the upper probability defined as in Theorem 1. Given a random variable X on Ω then the upper probability μ_X induced by μ on $(\mathbb{R}, \boldsymbol{F})$, where $\boldsymbol{F}$ is the Borel σ-field, is defined by $\mu_X(B) = \overline{P}(\omega \in \Omega : \omega \in X^{-1}(B)) = \frac{h^s(X^{-1}(B))}{h^s(\Omega)}$ for B belonging to $\boldsymbol{F}$.*

If X is a continuous random variable and B is a Borelian set then the set $X^{-1}(B)$ is also a Borelian set; moreover since every Hausdorff s-dimensional outer measure is countably additive on the Borel σ-field then the probability induced by the upper probability μ on $(\mathbb{R},\boldsymbol{F})$ is a probability measure. A probability measure on $(\mathbb{R},\boldsymbol{F})$ corresponds to every decreasing distribution function $G_{\mu,X}(x)$. If μ_n and μ are the probability measures on $(\mathbb{R},\boldsymbol{F})$ corresponding to $G_{\mu,X_n}(x)$ and $G_{\mu,X}(x)$ then X_n converges in μ-distribution to X if and only if $\lim_{n\to\infty}\mu_n(A)=\mu(A)$ for every $A=(x,\infty)$.

This last condition is equivalent to the pointwise convergence of expectation functionals on all bounded and continuous function f (Theorem 29.1 of [1]), that is $\lim_{n\to\infty}\int f d\mu_n=\int f d\mu$.

An important consequence of Theorem 2 is that upper probabilities defined as in Theorem 1 satisfy the General Dominated Convergence Theorem.

We prove that when Ω is a non-empty set with positive and finite Hausdorff outer measure in its dimension, the upper prevision defined by Definition 1 satisfies the following Monotone Convergence Theorem for monotone set functions ([3, Theorem 8.1]).

Theorem 3. *Let μ be a monotone set function on a σ-algebra $\boldsymbol{F}$ properly contained in P(Ω), which is continuous from below. For an increasing sequence of non negative, $\boldsymbol{F}$-measurable random variables X_n the limit function $X=\lim_{n\to\infty}X_n$ is $\boldsymbol{F}$-measurable and $\lim_{n\to\infty}\int X_n d\mu=\int X d\mu$.*

Theorem 4. *Let Ω be a non-empty set with positive and finite Hausdorff outer measure h^s in its dimension. Let $\boldsymbol{L}$ be the class of all bounded random variables on Ω. Then the coherent upper prevision defined by Definition 1 is continuous, that is given an increasing sequence of non negative random variables X_n converging to the random variable X we have that $\lim_{n\to\infty}\overline{P}(X_n|\Omega)=\overline{P}(X|\Omega)$.*

Proof. Each s-dimensional Hausdorff outer measure is regular and continuous from below. Then by the Monotone Convergence Theorem it follows that the given upper prevision is continuous. □

When Ω is a non-empty set with positive and finite Hausdorff outer measure in its dimension, the upper probabilities μ defined as in Theorem 1 are monotone, subadditive and continuous from below so they satisfy Proposition 3.7 of [2], that is a sequence of random variables μ-uniformly integrable and μ-stochastically converging to a random variable X, converges in μ-mean to X. Moreover from Proposition 3.8 of [2] we obtain that if μ is equal to the upper probability defined by Definition 1 then a sequence of Borel measurable random variables converging in μ-mean to X is μ-stochastically converging. In the next theorem we prove that when Ω has finite and positive Hausdorff outer measure in its dimension and we consider the restriction to the Borel σ-field of the upper probability defined as in Theorem 1, then pointwise convergence implies μ-stochastic convergence.

Theorem 5. *Let Ω be an non-empty set with finite and positive Hausdorff measure in its dimension, let $\boldsymbol{B}$ be a partition of Ω. Denote by s the Hausdorff dimension of the conditioning event B and by h^s the Hausdorff s-dimensional measure. Let $\boldsymbol{F}$ be the σ-field of all Borel subsets of Ω. Moreover, let m be a 0-1 valued finitely additive, but*

not countably additive, probability. Let μ be the upper probability defined on $\boldsymbol{F} \times \boldsymbol{B}$ as in Theorem 1, that is $\mu = \overline{P}(A|\Omega)$. Let X_n be a sequence of Borel measurable random variables on Ω converging pointwise to a random variable X. Then X_n converges to X μ-stochastically.

Proof. Since Ω has finite and positive Hausdorff measure in its dimension we have $\mu = \overline{P}(A|\Omega) = \frac{h^s(A)}{h^s(\Omega)}$. Moreover every outer Hausdorff measure is continuous from below and countably additive on the Borel σ-field. So every (outer) Hausdorff measure is continuous from above on the Borel σ-field. Then the pointwise convergence of a sequence of random variables X_n to X implies the μ-stochastic convergence of X_n to X. □

Remark 1. In general a coherent upper probability is not continuous from below and continuous from above; for example if we consider a coherent upper probability defined as natural extension of a merely finitely additive probability on a σ-field, then it is not continuous from above and continuous from below since an additive measure on a σ-field is continuous from above and continuous from below if and only if it is σ-additive. As a consequence we have that the pointwise convergence does not imply stochastic convergence with respect to this upper probability and the Monotone Convergence Theorem cannot always be applied. Hausdorff outer measures satisfy Theorem 4 and Theorem 5 because they are Borel regular outer measures.

4 Conclusions

In this paper it is proven that the relations among different types of convergences of random variables defined with respect to upper probability defined by Hausdorff outer measures are the same that hold if convergences are defined with respect to a probability measure. When Ω has finite Hausdorff outer measure in its dimension these results are obtained because Hausdorff outer measures are Borel regular outer measures and so continuous from below and continuous from above on the Borel σ-field. In general if upper probability is defined as natural extension of a coherent merely finitely additive probability defined on a σ-field we have that μ-stochastically convergence does not imply convergence in μ-distribution since in this case the upper probability is not continuous from above.

Acknowledgement. I wish to thank two anonymous referees for their useful comments.

References

1. Billingsley, P.: Probability and Measure. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1986)
2. Couso, I., Montes, S., Gil, P.: Stochastic convergence, uniform integrability and convergence in mean on fuzzy measure spases. Fuzzy Sets Syst 129, 95–104 (2002)
3. Denneberg, D.: Non-Additive Measure and Integral. Theory and Decision Library B: Mathematical and Statistical Methods, vol. 27. Kluwer Academic Publishers Group, Dordrecht (1994)

4. Doria, S.: Coherence and fuzzy reasoning. In: Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, pp. 165–174 (2007)
5. Falconer, K.J.: The Geometry of Fractals Sets. Cambridge Tracts in Mathematics, vol. 85. Cambridge University Press, Cambridge (1986)
6. Seidenfeld, T., Schervish, M., Kadane, J.B.: Improper regular conditional distributions. Ann. Probab. 29(4), 1612–1624 (2001)
7. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Monographs on Statistics and Applied Probability, vol. 42. Chapman and Hall, Ltd., London (1991)

On Convergence in Necessity and Its Laws of Large Numbers

Pedro Terán

Grupo Decisión Multicriterio Zaragoza, Facultad de Ciencias Económicas y Empresariales, Universidad de Zaragoza, Zaragoza, Spain

Abstract. We aim at clarifying the relationship between laws of large numbers for fuzzy sets or possibility distributions and laws of large numbers for fuzzy or possibilistic variables. We contend that these two frameworks are different and present the relationships between them that explain why this fact was unrecognized so far.

1 The Problem

A classical result in Probability Theory is the **law of large numbers**. If ξ is a variable defined on a probability space $(\Omega, \mathscr{A}, P)$, with mean μ, and $\{\xi_n\}_n$ is a sequence of independent, identically distributed variables, then

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \to \mu.$$

If convergence is almost sure, i.e. $\xi_n \to \xi$ if $\xi_n(\omega) \to \xi(\omega)$ except for at most those ω in a set of probability 0, it is called the Strong LLN. If convergence is in probability, i.e. $P(|\xi_n - \xi| > \varepsilon) \to 0$ for every $\varepsilon > 0$, then it is called the Weak LLN. Almost sure convergence is stronger than convergence in probability, hence the names.

If we replace the probability measure P by a possibility measure Π, the variable is usually called a *fuzzy* or *possibilistic variable*. One can then ask whether similar theorems hold. The notion of almost sure convergence applies verbatim and convergence in probability has an immediate analog: $\xi_n \to \xi$ in necessity if $\Pi(|\xi_n - \xi| > \varepsilon) \to 0$ for every $\varepsilon > 0$. The name comes from the dual expression $Nec(|\xi_n - \xi| \leq \varepsilon) \to 1$, with Nec being the dual measure to Π (a necessity measure).

The notion of independence is, however, more involved, as discussed elsewhere [7]. Here we will replace the notion of independence by $\top$-relatedness, without inquiring whether $\top$-related variables are 'independent' in the intuitive sense or not. For any triangular norm $\top$, two variables ξ and η are called $\top$-*related* if

$$\Pi(\xi \in A, \eta \in B) = \Pi(\xi \in A) \top \Pi(\xi \in B)$$

for any sets A, B.

It must be observed that a possibility measure Π is characterized by the function $\pi : \omega \mapsto \Pi(\{\omega\})$, since $\Pi(A) = \sup_{\omega \in A} \pi(\omega)$. This function is called the *possibility*

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 289–296, 2008.
springerlink.com

distribution of Π (this terminology is standard though in conflict with 'probability distribution'='probability measure induced by a variable') and can be identified with a fuzzy set. The possibility distribution of the possibility measure induced by the bivariate (ξ, η) can be retrieved from those of ξ and η as follows:

$$\pi_{(\xi,\eta)}(x,y) = \pi_\xi(x) \top \pi_\eta(y).$$

Notice that this implies some restrictions, related to left continuity, on the t-norm $\top$.

As a consequence of that,

$$\pi_{\xi+\eta}(x) = \sup_{x_1+x_2=x} \pi_{(\xi,\eta)}(x_1,x_2) = (\pi_\xi \oplus \pi_\eta)(x);$$

namely, the possibility distribution of a sum of $\top$-related possibilistic variables is the t-normed sum of the possibility distributions of the summands. It is sometimes thought, incorrectly, that the formula above is a *definition*.

That shows that, at least to some extent, possibility measures of sums of variables can be studied via t-normed sums of fuzzy sets. Curiously, early papers like [1, 6, 11] dealt explicitly with fuzzy variables, but Robert Fullér's paper [3] seems to have caused a shift in the perception of this problem and further papers have spoken of the 'law of large numbers for fuzzy sets' and framed it within fuzzy arithmetics, omitting any underlying variables.

Our contention is that, rather than being a mere naming issue, those two problems are not the same. They are not perceived like that because, in the formulation of the LLNs for fuzzy sets there remain two vestiges of the LLN for possibilistic variables:

(a) Convergence in the LLN for fuzzy sets 'is' still convergence in necessity.
(b) The limit is assumed to be a point.

When the limit is a point, convergence in necessity can be written without reference to the sequence of variables but only to their possibility distributions identified with fuzzy sets, as follows:

$$\xi_n \to x \text{ in necessity } \Leftrightarrow \sup_{y \notin [x-\varepsilon, x+\varepsilon]} \pi_n(y) \to 0 \quad \text{for each } \varepsilon > 0.$$

(Here and in the sequel, π_n, π always denote the possibility distributions of the variables ξ_n, ξ.) This notion of convergence of a sequence of fuzzy sets to a point cannot be extended analogously to the case of a fuzzy limit, because then convergence in necessity cannot be recast as a convergence of possibility distributions. But, in order to ensure that the limit is a point, one must require that fuzzy sets in the sequence have a single *modal value* (i.e. with possibility 1).

Arguably, those vestiges have not been left behind because a solution to these two problems was not found:

(a) It is not obvious how convergence in necessity to a point should be extended to the general case.
(b) Unnatural assumptions must be made in order to ensure that the limit is a point.

As regards vestige (b) and problem (b), it must be observed that, contrary to the situation in Probability Theory, **the limit in the LLN need not be a point.**

The aim of this contribution is to proceed to a clarification of these issues.

2 Convergence in Necessity and Levelwise Convergence

Let $\mathscr{U}$ be the family of all fuzzy subsets of $\mathbb{R}^d$ having non-empty compact α-cuts for $\alpha \in (0,1]$. In that space we consider the topology τ generated by the mappings $L_\alpha : U \mapsto U_\alpha$, namely

$$U_n \to_\tau U \Leftrightarrow (U_n)_\alpha \to U_\alpha \text{ for all } \alpha \in (0,1].$$

Convergence of the α-cuts is in the Hausdorff metric. The topology τ is called the *topology of levelwise convergence* or the *cylindrical topology*.

Our starting point is the following result, which is a corollary to [9, Theorem 13].

Proposition 1. *Let $\top$ be a continuous t-norm, and $\oplus$ its associated t-normed sum. For any $U \in \mathscr{U}$,*

$$\bigoplus_{i=1}^{n} U \to_\tau K\mathrm{co}U,$$

where $K\mathrm{co}U(x) = \sup\{a \in [0, \mathrm{co}U(x)] \mid a \top a = a\}$.

For further discussion of this and related results, the reader is referred to [9, Proposition 4], [10, Example 4] and [8].

For clarity of presentation, we consider only the Archimedean case.

Corollary 1. *Let $\top$ be a continuous Archimedean t-norm, and $\oplus$ its associated t-normed sum. For any $U \in \mathscr{U}$,*

$$\bigoplus_{i=1}^{n} U \to_\tau I_{\mathrm{co}U_1}.$$

These results tell us that, for a sequence $\{\xi_n\}_n$ of $\top$-related identically distributed possibility measures, the possibility distribution of $n^{-1}\sum_{i=1}^n \xi_i$ converges, in some sense, to another possibility distribution. That distribution represents a crisp convex set if the t-norm is Archimedean, but it can be a genuinely fuzzy set in general.

We begin with an interesting example.

Example 1. Identically distributed variables whose partial sums are identically distributed but do not converge almost surely. Let $\Omega = [0,2]$ with the ignorance possibility distribution such that $\Pi(\varnothing) = 0$ and $\Pi(A) = 1$ for any other $A \subset \Omega$. Define the functions $f_m : \Omega \to \mathbb{R}$ such that $f_m(x) = (1-x)^{-1}$ for $x \in [0,1)$, $f_m(x) = 0$ for $x \in [1,2)$ and $f_m(2) = m$. Define $\xi_n = f_{2n-1}$ and notice that all ξ_n are identically distributed, with possibility distribution $I_{[0,\infty)}$. Besides,

$$n^{-1}\sum_{i=1}^{n} \xi_i = f_{n^{-1}\sum_{i=1}^n (2i-1)} = f_n,$$

so the variables $n^{-1}\sum_{i=1}^n \xi_i$ are identically distributed as well. All f_m are identical except at point 2, but since $\Pi(\{2\}) = 1$, the sequence cannot converge almost surely.

Observe that a sequence which does not converge almost surely, cannot converge in necessity either. This is exactly the opposite of what happens in Probability Theory.

Proposition 2. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables. If $\xi_n \to \xi$ in necessity, then $\xi_n \to \xi$ almost surely.*

We turn now to the relationship between the levelwise convergence in Proposition 1 and Corollary 1 and the convergence in necessity to a point in the literature.

Proposition 3. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables. If $\pi_n \to I_{\{x\}}$ in τ, then $\xi_n \to x$ in necessity.*

In view of this result, τ-convergence provides a way to make rigorous the notion of convergence in necessity in laws of large numbers for fuzzy sets. It also shows that assumptions on the form of the limit are not needed.

One also has to notice that the converse holds in full generality.

Proposition 4. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables. If $\xi_n \to \xi$ in necessity, then $\pi_n \to \pi$ in τ.*

We denote by B the closed unit ball, and by $B(x,\varepsilon)$ the closed ball of center x and radius ε.

Proof. By hypothesis, for each $\varepsilon > 0$ and $\alpha \in (0,1]$, there exists $n_0 \in \mathbb{N}$ such that

$$\Pi(|\xi_n - \xi| > \varepsilon) < \alpha$$

for all $n \geq n_0$. Denote the event $\{|\xi_n - \xi| > \varepsilon\}$ by A.

Let $x \in \pi_\alpha$ (equivalently, $\Pi(\xi = x) \geq \alpha$). We have

$$\alpha \leq \Pi(\xi = x) = \max\{\Pi(\{\xi = x\} \cap A), \Pi(\{\xi = x\} \cap A^c)\},$$

but

$$\Pi(\{\xi = x\} \cap A) \leq \Pi(A) < \alpha,$$

so we deduce

$$\pi(x) = \Pi(\xi = x) = \Pi(\{\xi = x\} \cap A^c).$$

Now, if the event in the right-hand side occurs, $\xi = x$ and $|\xi_n - \xi| \leq \varepsilon$ imply $\xi_n \in B(x,\varepsilon)$. Therefore,

$$\pi(x) \leq \Pi(\xi_n \in B(x,\varepsilon)) = \sup_{y \in B(x,\varepsilon)} \Pi(\xi_n = y) = \sup_{y \in B(x,\varepsilon)} \pi_n(y).$$

In summary,

$$\alpha \leq \pi(x) \leq \sup_{y \in B(x,\varepsilon)} \pi_n(y),$$

whence

$$\begin{aligned}
\pi_\alpha &\subset \{x \mid \forall \beta \in (0,\alpha), \sup_{y \in B(x,\varepsilon)} \pi_n(y) > \beta\} \\
&= \bigcap_{\beta \in (0,\alpha)} \{x \mid \exists y \in B(x,\varepsilon) \mid \pi_n(y) > \beta\} = \bigcap_{\beta \in (0,\alpha)} ((\pi_n)_{\beta^+} + \varepsilon B) \\
&= \bigcap_{\beta \in (0,\alpha)} (\pi_n + \varepsilon I_B)_{\beta^+} = (\pi_n + \varepsilon I_B)_\alpha = (\pi_n)_\alpha + \varepsilon B.
\end{aligned}$$

Analogously, we prove

$$(\pi_n)_\alpha \subset \pi_\alpha + \varepsilon B.$$

Accordingly,

$$d_H((\pi_n)_\alpha, \pi_\alpha) \leq \varepsilon.$$

Thus, for each fixed $\alpha \in (0,1]$ we have

$$d_H((\pi_n)_\alpha, \pi_\alpha) \to 0,$$

and the proof is complete. □

It is essential to realize that it is not possible to provide a mode of convergence for fuzzy sets such that $\xi_n \to \xi$ in necessity if and only if $\pi_n \to \pi$. The reason is that convergence of variables and convergence of distributions or fuzzy sets are conceptually different. Indeed, let $\{\xi_n\}_n$ be an arbitrary sequence of identically distributed variables. Their distributions are identical (in particular, converge) but it cannot be expected that ξ_n converges in necessity or almost surely.

3 Levelwise Convergence and Weak Convergence

As we have shown, one has to distinguish carefully between convergences of variables and convergences of (merely) their distributions. In Probability Theory, we have almost sure convergence and convergence in probability on one side, and convergence in distribution on the other side. A sequence of variables is said to converge in distribution when their distributions converge weakly. Weak converge is metrized by e.g. the Prokhorov metric

$$\rho(P_1, P_2) = \inf\{\varepsilon > 0 \mid P_1(F) \leq P_2(F + \varepsilon B) + \varepsilon \text{ for each closed set } F\}.$$

A notion of weak convergence of capacities has been developed [5, 2] which includes, in particular, possibility measures. A sequence $\{\nu_n\}_n$ of capacities converges weakly to ν if

(i) For each closed set $F \subset \mathbb{R}^d$, $\limsup_n \nu_n(F) \leq \nu(F)$.
(ii) For each open set $G \subset \mathbb{R}^d$, $\liminf_n \nu_n(G) \geq \nu(G)$.

One can also consider the Prokhorov metric

$$\rho(\nu_1, \nu_2) = \inf\{\varepsilon > 0 \mid \nu_1(F) \leq \nu_2(F + \varepsilon B), \\ \nu_2(F) \leq \nu_1(F + \varepsilon B) \text{ for each closed set } F\}.$$

In the case of possibility measures, this convergence reduces to epiconvergence of their distributions.

Our first observation is as follows.

Proposition 5. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables. If $\xi_n \to \xi$ almost surely, then the induced possibility measures $\Pi_{\xi_n} \to \Pi_\xi$ converge weakly.*

We observe now that τ-convergence is stronger than weak convergence.

Proposition 6. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables. If $\pi_n \to \pi$ in τ, then the induced possibility measures $\Pi_{\xi_n} \to \Pi_\xi$ converge weakly.*

But the converse holds if the limit is non-fuzzy.

Proposition 7. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables, and let ξ have induced distribution $\pi = I_A$ for some set A. If $\Pi_{\xi_n} \to \Pi_\xi$ weakly, then $\pi_n \to \pi$ in τ.*

Corollary 2. *Let $\{\xi_n\}_n$ be a sequence of possibilistic variables and $x \in \mathbb{R}^d$. Then, the following are equivalent:*

(i) $\xi_n \to x$ *in necessity,*
(ii) $\xi_n \to x$ *almost surely,*
(iii) $\pi_n \to I_{\{x\}}$ *in* τ,
(iv) $\Pi_{\xi_n} \to \delta_x$ *weakly,*

where δ_x is the Dirac distribution at x.

4 Tightness and the Law of Large Numbers for Possibility Distributions

A common technique to prove LLNs in the literature, e.g. [4], is as follows: (1) fix an Archimedean t-norm, (2) consider its decreasing generator f, (3) consider the compositions $f \circ \pi_n$, (4) make assumptions on their shapes, (5) calculations provide the LLN. The use of this technique implies that the set of sequences for which the LLN is proven differs for each t-norm. We want to show that other techniques are possible which do not depend on the chosen t-norm.

We start by recalling the notion of tight families of capacities, which generalizes tight families of probabilities. The definition was given in [5], but in a wrong way which we correct here. A set $\mathscr{N}$ of capacities is called *tight* if, for every $\varepsilon > 0$, there is a compact set K such that

$$\sup_{\nu \in \mathscr{N}} \nu(K^c) < \varepsilon.$$

This notion extends to families of variables in the obvious way.

We can characterize tight families of possibility measures in terms of their possibility distributions.

Proposition 8. *A family $\mathscr{N}$ of possibility measures on $\mathbb{R}^d$ is tight if and only if there exists a fuzzy set $U \in \mathscr{U}$ dominating the possibility distribution of each element of $\mathscr{N}$.*

Then, we obtain the following law of large numbers.

Theorem 1. *Let $\top$ be any continuous Archimedean t-norm. Let $\{\xi_n\}_n$ be a tight sequence of $\top$-related possibilistic variables in $\mathbb{R}$. Then,*

$$d_H((n^{-1}\bigoplus_{i=1}^{n}\pi_i)_\alpha, n^{-1}\sum_{i=1}^{n}(\pi_i)_1) \to 0$$

for each $\alpha \in (0,1]$. In particular, if the ξ_n have a unique modal value m_n and $n^{-1}\sum_{i=1}^n m_i \to m$ then $n^{-1}\sum_{i=1}^n \xi_i \to m$ in necessity.

Proof (Scheme)

Step 1. Since any interval can be subtracted from a larger interval, we can write

$$\pi_n = I_{(\pi_n)_1} + \tilde{\pi}_n,$$

where $\tilde{\pi}_n$ are fuzzy sets with unique modal value 0.

Step 2. By [8, Lemma 3], we also have

$$\pi_n = I_{(\pi_n)_1} \oplus \tilde{\pi}_n,$$

whence

$$d_H((n^{-1}\bigoplus_{i=1}^{n}\pi_i)_\alpha, n^{-1}\sum_{i=1}^{n}(\pi_i)_1) = d_H(n^{-1}\sum_{i=1}^{n}((\pi_i)_1 + (\tilde{\pi}_i)_\alpha), n^{-1}\sum_{i=1}^{n}(\pi_i)_1))$$

$$\leq d_H(n^{-1}\sum_{i=1}^{n}(\tilde{\pi}_i)_\alpha, \{0\}).$$

Therefore it suffices to prove $n^{-1}\sum_{i=1}^{n}(\tilde{\pi}_i) \to I_{\{0\}}$ in τ.

Step 3. Since $\{\pi_n\}_n$ is tight, $\{\tilde{\pi}_n\}_n$ is so as well.

Step 4. Define $U \in \mathscr{U}$ to be the upper semicontinuous envelope of $\sup_n \tilde{\pi}_n$. Then $U_1 = \{0\}$.

Step 5. We have

$$I_{\{0\}} \subset \tilde{\pi}_i \subset U.$$

We apply Corollary 1 to U, and a sandwich argument yields $n^{-1}\sum_{i=1}^{n}(\tilde{\pi}_i) \to I_{\{0\}}$ in τ, as wished. □

Note. The author maintains a weblog on Statistics, Probability and Fuzzy Sets at the following URL: http://spfs.blogspot.com

Acknowledgement. Research partially funded by the Spanish Ministerio de Educación y Ciencia (MTM2005–02254 and TSI2005–02511, and the Gobierno de Aragón (PM2007–034).

References

1. Badard, R.: The law of large numbers for fuzzy processes and the estimation problem. Inform. Sci. 28, 161–178 (1982)
2. Baddeley, A.: Hausdorff Metric for Capacities. Technical Report, CWI Report BS-R9127 (1991)
3. Fullér, R.: A law of large numbers for fuzzy numbers. Fuzzy Sets Syst. 45, 293–303 (1992)
4. Hong, D.H., Ahn, C.H.: Equivalent conditions for laws of large numbers for $\top$-related *L-R* fuzzy numbers. Fuzzy Sets Syst. 136, 387–395 (2003)
5. Norberg, T.: Random capacities and their distributions. Probab. Theory Related Fields 73, 281–297 (1986)
6. Rao, M.B., Rashed, A.: Some comments on fuzzy variables. Fuzzy Sets Syst. 6, 285–292 (1981)

7. Terán, P.: Paradoxes in the law of large numbers without probability. Seminar talk at the Department of Mathematics, Universidad de Extremadura, Spain (paper in preparation) (February 2007)
8. Terán, P.: Probabilistic foundations for measurement modelling using fuzzy random variables. Fuzzy Sets Syst 158, 973–986 (2007)
9. Terán, P.: A strong law of large numbers for t-normed arithmetics. Fuzzy Sets Syst. 159, 343–360 (2008)
10. Terán, P., Molchanov, I.: The law of large numbers in a metric space with a convex combination operation. J. Theoret. Probab. 19, 875–898 (2006)
11. Williamson, R.C.: The law of large numbers for fuzzy variables under a general triangular norm extension principle. Fuzzy Sets Syst. 41, 55–81 (1991)

Combinatorics of Imprecise Probabilities

The Omnipresence of Cycle-Transitivity in the Comparison of Random Variables

Bernard De Baets[1], Hans De Meyer[2], and Karel De Loof[2]

[1] Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Gent, Belgium
[2] Department of Applied Mathematics and Computer Science, Ghent University, Gent, Belgium

Abstract. In this paper, the transitivity properties of reciprocal relations, also called probabilistic relations, are investigated within the framework of cycle-transitivity. Interesting types of transitivity are highlighted and shown to be realizable in applications. For example, given a collection of random variables $(X_k)_{k\in I}$, pairwisely coupled by means of a same copula $C \in \{T_{\mathbf{M}}, T_{\mathbf{P}}, T_{\mathbf{L}}\}$, the transitivity of the reciprocal relation Q defined by $Q(X_i,X_j) = \text{Prob}\{X_iX_j\} + 1/2\ \text{Prob}\{X_i = X_j\}$ can be characterized within the cycle- transitivity framework. Similarly, given a poset $(P,\leq)$ with $P = \{x_1,\ldots,x_n\}$, the transitivity of the mutual rank probability relation Q_P, where $Q_P(X_i,X_j)$ denotes the probability that x_i precedes x_j in a random linear extension of P, is characterized as a type of cycle-transitivity for which no realization had been found so far.

Keywords: Copula, Cycle-transitivity, Poset, Random variables comparison, Reciprocal relation, Transitivity, t-Norm, T-transitivity.

1 Introduction

Reciprocal ($[0,1]$-valued binary relations Q satisfying $Q(a,b)+Q(b,a) = 1$) provide a convenient tool for expressing the result of the pairwise comparison of a set of alternatives [2]. They are particularly popular in fuzzy set theory where they are used for representing intensities of preference [1, 14]. They are often called probabilistic relations as they are used for expressing the result of the pairwise comparison of random variables in various fields such as game theory [11], voting theory [16], mathematical psychology [12] and order theory [8, 13, 15].

Recently, we have presented a general framework for studying the transitivity of reciprocal relations, encompassing various types of T-transitivity and stochastic transitivity [3, 7]. It turns out that many types of transitivity encountered in practical applications, especially when the comparison of random variables is at stake, when casted into the form compatible with cycle-transitivity, become very simple. The outline of the paper is as follows. First, the standard types of transitivity are briefly reviewed. Then the cycle-transitivity framework is introduced and it is indicated how in this framework transitivity is characterized by a single function, called upper bound function, acting on three (ordered) arguments. A scheme is established that encompasses the standard and some other important types of cycle- transitivity. The second part of the paper deals

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 299–305, 2008.
springerlink.com

with two applications: in each a reciprocal relation is established through a particular way of comparing random variables. the scheme admit a transitivity.

2 Cycle-Transitivity

A fuzzy relation R on A is an $A^2 \to [0,1]$ mapping that expresses the degree of relationship between elements of A: $R(a,b) = 0$ means a and b are not related at all, $R(a,b) = 1$ expresses full relationship, while $R(a,b) \in]0,1[$ indicates a partial degree of relationship only. For such relations, the concept of T-transitivity is very natural.

Definition 1. *Let T be a t-norm. A fuzzy relation R on A is called T- transitive if for any $(a,b,c) \in A^3$ it holds that $T(R(a,b),R(b,c)) \leq R(a,c)$.*

The three basic t-norms are $T_{\mathbf{M}}$, the minimum operator, $T_{\mathbf{P}}$, the ordinary product, and $T_{\mathbf{L}}$, the Łukasiewicz t-norm.

Another class of $A^2 \to [0,1]$ mappings are the reciprocal relations Q satisfying $Q(a,b) + Q(b,a) = 1$, for any $a,b \in A$. They arise in the context of pairwise comparison. Though the semantics of reciprocal relations and fuzzy relations are different, the concept of T-transitivity is sometimes formally applied to reciprocal relations as well. However, more often the transitivity properties of reciprocal relations can be characterized as of one of the various kinds of stochastic transitivity. The following general formulation of stochastic transitivity has been proposed [3].

Definition 2. *Let g be a commutative increasing $[1/2,1]^2 \to [1/2,1]$ mapping. A reciprocal relation Q on A is called stochastic transitive w.r.t. g if for any $(a,b,c) \in A^3$ it holds that $(Q(a,b) \geq 1/2 \wedge Q(b,c) \geq 1/2) \Rightarrow Q(a,c) \geq g(Q(a,b),Q(b,c))$.*

This definition includes strong stochastic transitivity when $g = \max$, moderate stochastic transitivity when $g = \min$, weak stochastic transitivity when $g = 1/2$, and λ-transitivity, with $\lambda \in [0,1]$, when $g = \lambda \max + (1-\lambda)\min$. Clearly, strong stochastic transitivity implies λ- transitivity, which implies moderate stochastic transitivity, which, in turn, implies weak stochastic transitivity.

In the cycle-transitivity framework [7], for a reciprocal relation Q on A, the quantities

$$\alpha_{abc} = \min(Q(a,b),Q(b,c),Q(c,a))\,,\ \beta_{abc} = \mathrm{med}(Q(a,b),Q(b,c),Q(c,a))\,,$$

$$\gamma_{abc} = \max(Q(a,b),Q(b,c),Q(c,a))\,,$$

are defined for all $(a,b,c) \in A^3$. Obviously, $\alpha_{abc} \leq \beta_{abc} \leq \gamma_{abc}$. Also, the notation $\Delta = \{(x,y,z) \in [0,1]^3 \,|\, x \leq y \leq z\}$ will be used.

Definition 3. *A function $U : \Delta \to \mathbb{R}$ is called an upper bound function if it satisfies:*

(i) $U(0,0,1) \geq 0$ and $U(0,1,1) \geq 1$;
(ii) for any $(\alpha,\beta,\gamma) \in \Delta$:

$$U(\alpha,\beta,\gamma) + U(1-\gamma,1-\beta,1-\alpha) \geq 1\,. \tag{1}$$

The function $L : \Delta \to \mathbb{R}$ defined by

$$L(\alpha,\beta,\gamma) = 1 - U(1-\gamma, 1-\beta, 1-\alpha) \tag{2}$$

is called the dual lower bound function of a given upper bound function U. Inequality (1) simply expresses that $L \leq U$.

Definition 4. *A reciprocal relation Q on A is called cycle-transitive w.r.t. an upper bound function U if for any $(a,b,c) \in A^3$ it holds that*

$$L(\alpha_{abc},\beta_{abc},\gamma_{abc}) \leq \alpha_{abc}+\beta_{abc}+\gamma_{abc}-1 \leq U(\alpha_{abc},\beta_{abc},\gamma_{abc}), \tag{3}$$

where L is the dual lower bound function of U.

Due to the built-in duality, it holds that if (3) is true for some (a,b,c), then this is also the case for any permutation of (a,b,c). In practice, it is therefore sufficient to check (3) for a single permutation of any $(a,b,c) \in A^3$. Alternatively, due to the same duality, it is also sufficient to verify the right-hand inequality (or equivalently, the left-hand inequality) for two permutations of any $(a,b,c) \in A^3$ (not being cyclic permutations of one another), e.g. (a,b,c) and (c,b,a). Hence, (3) can be replaced by

$$\alpha_{abc}+\beta_{abc}+\gamma_{abc}-1 \leq U(\alpha_{abc},\beta_{abc},\gamma_{abc}). \tag{4}$$

Note that a value of $U(\alpha,\beta,\gamma)$ equal to 2 will often be used to express that for the given values there is no restriction at all (indeed, $\alpha+\beta+\gamma-1$ is always bounded by 2).

For two upper bound functions such that $U_1 \leq U_2$, it clearly holds that cycle-transitivity w.r.t. U_1 implies cycle-transitivity w.r.t. U_2. It is clear that $U_1 \leq U_2$ is not a necessary condition for the latter implication to hold. Two upper bound functions U_1 and U_2 will be called *equivalent* if for any $(\alpha,\beta,\gamma) \in \Delta$ it holds that $\alpha+\beta+\gamma-1 \leq U_1(\alpha,\beta,\gamma)$ is equivalent to $\alpha+\beta+\gamma-1 \leq U_2(\alpha,\beta,\gamma)$.

For instance, suppose that the inequality $\alpha+\beta+\gamma-1 \leq U_1(\alpha,\beta,\gamma)$ can be rewritten as $\alpha \leq h(\beta,\gamma)$, then an equivalent upper bound function U_2 is given by $U_2(\alpha,\beta,\gamma) = \beta+\gamma-1+h(\beta,\gamma)$. In this way, it is often possible to find an equivalent upper bound function in only two of the variables α, β and γ.

The different types of fuzzy and stochastic transitivity can be reformulated in the cycle-transitivity framework and are then characterized by an upper bound function $U(\alpha,\beta,\gamma)$. In Figure 1, upper bound functions are shown for $T_{\mathbf{M}}$-, $T_{\mathbf{P}}$- and $T_{\mathbf{L}}$-transitivity as well as for strong, moderate and stochastic transitivity. Three more types of transitivity are depicted: two consecutive weakenings of $T_{\mathbf{P}}$-transitivity, respectively called dice-transitivity and mutual-rank-transitivity, and a type of transitivity which is situated between moderate stochastic transitivity and mutual-rank-transitivity and is known as partial stochastic transitivity.

In fact, many examples of reciprocal relations we have encountered in our research on the comparison of random variables distributions are neither fuzzy nor stochastic transitive but have a type of transitivity that can be nicely expressed as an instance of cycle-transitivity.

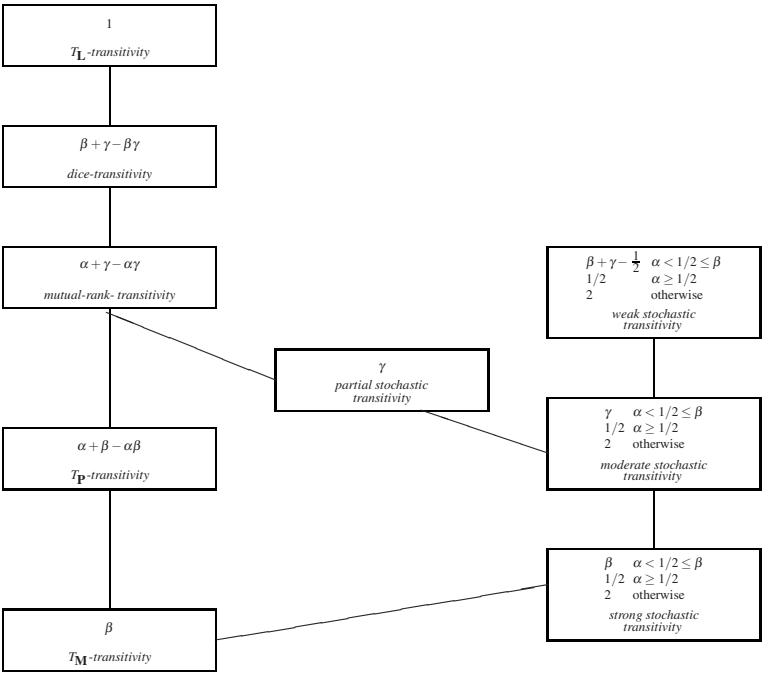


Fig. 1. Hasse-diagram with different types of cycle-transitivity characterized by their upper bound function $U(\alpha,\beta,\gamma)$. Bottom up (partial) ordering of transitivity types is in agreement with the 'stronger than' relation.

3 The Pairwise Comparison of Artificially Coupled Random Variables

An immediate way of comparing two random variables is to consider the probability that the first one takes a greater value than the second one. Proceeding along this line of thought, a random vector $(X_1,X_2,\ldots,X_m)$ generates a reciprocal relation.

Definition 5. *Given a random vector $(X_1,X_2,\ldots,X_m)$, consider the reciprocal relation Q defined by*

$$Q(X_i,X_j)=\mathrm{Prob}\{X_i>X_j\}+\frac{1}{2}\mathrm{Prob}\{X_i=X_j\}\,. \tag{5}$$

The above comparison method takes into account the bivariate marginal c.d.f. and hence the pairwise dependence of the components of the random vector. The information contained in the reciprocal relation is therefore much richer than if, for instance, we would have based the comparison of X_i and X_j solely on their expected values. Despite the fact that the dependence structure is entirely captured by the multivariate c.d.f., the pairwise comparison is only apt to take into account pairwise dependence, as only bivariate c.d.f.'s are involved. The random variables may even be pairwise independent while not mutually independent.

Since the copulas C_{ij} that couple the univariate marginal c.d.f. into the bivariate marginal c.d.f. can be different from another, the analysis of the reciprocal relation and in particular the identification of its transitivity properties appear rather cumbersome. It is nonetheless possible to state in general, without making any assumptions on the bivariate c.d.f., that the reciprocal relation Q generated by an arbitrary random vector always shows some minimal form of transitivity.

Proposition 1. *The reciprocal relation generated by a random vector is $T_{\mathbf{L}}$- transitive.*

In [4], we have considered the situation where abstraction is made that the random variables are components of a random vector, and all bivariate c.d.f. are enforced to depend in the same way upon the univariate c.d.f., in other words, we consider the situation of all copulas being the same, well knowing that this might not be possible at all. In fact, this simplification is equivalent to considering instead of a random vector, a collection of random variables and to artificially compare them, all in the same manner and based upon a same copula.

First, we have unravelled the case of the product copula $T_{\mathbf{P}}$, and the cases of the two extreme copulas, the minimum operator $T_{\mathbf{M}}$ and the Łukasiewicz t- norm $T_{\mathbf{L}}$, respectively related to a presumed but not-necessarily existing comonotonic and countermonotonic pairwise dependence of the random variables [17]. The following results have been reported.

Proposition 2 ([9, 10]). *The reciprocal relation Q generated by a collection of random variables pairwisely coupled by $T_{\mathbf{P}}$ is dice-transitive, i.e. it is cycle- transitive w.r.t. to the upper bound function U given by $U(\alpha,\beta,\gamma) = \beta + \gamma - \beta\gamma$. In particular, the reciprocal relation generated by a collection of independent random variables is dice-transitive.*

Proposition 3 ([5, 6])

(i) The reciprocal Q generated by a collection of random variables pairwisely coupled by $T_{\mathbf{M}}$ is $T_{\mathbf{L}}$-transitive, i.e. it is cycle- transitive w.r.t. to the upper bound function U given by $U(\alpha,\beta,\gamma) = 1$.

(ii) The reciprocal relation Q generated by a collection of random variables pairwisely coupled by $T_{\mathbf{L}}$ is partially stochastic transitive, i.e. it is cycle-transitive w.r.t. to the upper bound function U given by $U(\alpha,\beta,\gamma) = \gamma$.

Note that partial stochastic transitivity is a slightly weakened version of moderate stochastic transitivity (note the strict inequalities instead of weak inequalities) as it can be expressed equivalently as

$$(Q(a,b)1/2 \wedge Q(b,c)1/2) \Rightarrow Q(a,c) \geq \min(Q(a,b),Q(b,c)),$$

for any $(a,b,c) \in A^3$.

4 Mutual Rank Probabilities in Posets

Consider a finite poset $(P,\leq)$ with $P = \{x_1,\ldots,x_n\}$. A linear extension of P is an order-preserving permutation of its elements (hence, also a ranking of the elements

compatible with the partial order). We denote by $p(x_i < x_j)$ the fraction of linear extensions of P in which x_i precedes x_j. If the space of all linear extensions of P is equipped with the uniform measure, the position of x in a linear extension can be regarded as a discrete random variable X with values in $\{1,\ldots,n\}$. Since $p(x_i < x_j) = \mathrm{Prob}\{X_i < X_j\}$, it is called a mutual rank probability. Note that P uniquely determines a random vector $X = (X_1,\ldots,X_n)$ with multivariate distribution function $F_{X_1,\ldots,X_n}$, whereas the rank probabilities $p(x_i < x_j)$ are then computed from the bivariate marginal distributions F_{X_i,X_j}.

Definition 6. *Given a poset $P = \{x_1,\ldots,x_n\}$, consider the reciprocal relation Q_P defined by*

$$Q_P(X_i,X_j) = \mathrm{Prob}\{X_i < X_j\} = p(x_i < x_j)\,. \tag{6}$$

The problem of probabilistic transitivity in a finite poset P was raised by Fishburn [13]. For any $u,v \in [0,1]$, define $\delta(u,v)$ as

$$\delta(u,v) = \inf\{p(x_i < x_k) \mid p(x_i < x_j) \geq u, p(x_j < x_k) \geq v\}\,,$$

where the infimum is taken over all choices of P and distinct x_i,x_j,x_k. Fishburn proved:

$$\begin{aligned}
&\delta(u,v) = 0 \text{ if } u+v<1\,,\\
&u+v-1 \leq \delta(u,v) \leq \min(u,v)\,,\\
&\delta(u,1-u) \leq 1/e\,,\\
&\delta(u,v) \leq 1-(1-u)(1-v)(1-\ln[(1-u)(1-v)])\,.
\end{aligned} \tag{7}$$

A non-trivial lower bound on δ was proved in [15] via geometric arguments. Define

$$\gamma(u,v) = \inf\{\mathrm{Prob}\{Y_i < Y_k\} \mid \mathrm{Prob}\{Y_i < Y_j\} \geq u, \mathrm{Prob}\{Y_j < Y_k\} \geq v\}\,,$$

where the infimum is taken over $Y = (Y_1,\ldots,Y_n)$ chosen uniformly from some n-dimensional compact convex subset of $\mathbb{R}^n$. Since $\delta(u,v) \geq \gamma(u,v)$, the function γ provides a lower bound for δ. Kahn and Yu [15] proved that

$$\gamma(u,v) = \begin{cases} 0 & \text{, if } u+v<1\,,\\ \min(u,v) & \text{, if } u+v-1 \geq \min(u^2,v^2)\,,\\ \dfrac{(1-u)(1-v)}{u+v-2\sqrt{u+v-1}} & \text{, otherwise.} \end{cases} \tag{8}$$

If we translate the bounds (7) and (8) into the cycle-transitivity framework, we obtain that (8) provides an upper bound $U(\alpha,\beta,\gamma)$ on $\alpha+\beta+\gamma-1$, whereas (7) provides a lower bound, which is, however, less stringent than the lower bound function $L(\alpha,\beta,\gamma)$ associated to $U(\alpha,\beta,\gamma)$ by (2). Surprisingly, the upper bound function $U(\alpha,\beta,\gamma)$ which is the equivalent of (8), is very simple.

Proposition 4. *The reciprocal relation Q_P generated by the mutual rank probabilities in a poset P, is mutual-rank-transitive, i.e. it is cycle- transitive w.r.t. to the upper bound function U given by $U(\alpha,\beta,\gamma) = \alpha+\gamma-\alpha\gamma$.*

Mutual-rank-transitivity is stronger than dice-transitivity but weaker than $T_{\mathbf{P}}$-transitivity (see Figure 1). Note that for general P, despite the fact that the multivariate distribution function $F_{X_1,\ldots,X_n}$, or equivalently, the n-dimensional copula, can be very complex, certain pairwise couplings are trivial. Indeed, if in P it holds that $x_i < y_j$, then x_i precedes y_j in all linear extensions and X_i and X_j are comonotone, which means that X_i and X_j are coupled by (a discretization of) $T_{\mathbf{M}}$. For pairs of elements in P that are incomparable, the bivariate couplings can vary from pair to pair. The copulas are not all equal to $T_{\mathbf{L}}$, as can be seen already from the example where P is an antichain with 3 elements.

References

1. Chiclana, F., Herrera, F., Herrera-Viedma, E., Martínez, L.: A note on the reciprocity in the aggregation of fuzzy preference relations using OWA operators. Fuzzy Sets Syst. 137, 71–83 (2003)
2. David, H.A.: The Method of Paired Comparisons. Griffin's Statistical Monographs & Courses, vol. 12. Charles Griffin & Co. Ltd., London (1963)
3. De Baets, B., De Meyer, H.: Transitivity frameworks for reciprocal relations: cycle-transitivity versus FG-transitivity. Fuzzy Sets Syst. 152, 249–270 (2005)
4. De Baets, B., De Meyer, H.: On the cycle-transitive comparison of artificially coupled random variables. Internat. J. Approx. Reason 47, 306–322 (2008)
5. De Baets, B., De Meyer, H., De Schuymer, B.: Extreme copulas and the comparison of ordered lists. Theory and Decision 62, 195–217 (2007)
6. De Baets, B., De Meyer, H., De Schuymer, B.: On the transitivity of comonotonic and countermonotonic comparison of random variables. J. Multivariate Anal. 98, 177–193 (2007)
7. De Baets, B., De Meyer, H., De Schuymer, B., Jenei, S.: Cyclic evaluation of transitivity of reciprocal relations. Soc. Choice Welf. 26, 217–238 (2006)
8. De Loof, K., De Baets, B., De Meyer, H., Brüggeman, R.: A hitchhiker's guide to poset ranking. Comb Chem High Throughput Screen (in press, 2008)
9. De Schuymer, B., De Meyer, H., De Baets, B.: Cycle-transitive comparison of independent random variables. J. Multivariate Anal. 96, 352–373 (2005)
10. De Schuymer, B., De Meyer, H., De Baets, B., Jenei, S.: On the cycle-transitivity of the dice model. Theory and Decision 54, 261–285 (2003)
11. Dutta, B., Laslier, J.-F.: Comparison functions and choice correspondences. Soc. Choice Welf. 16, 513–532 (1999)
12. Fishburn, P.: Binary choice probabilities: On the varieties of stochastic transitivity. J. Mathematical Psychology 41, 48–60 (1986)
13. Fishburn, P.: Proportional transitivity in linear extensions of ordered sets. J. Combin. Theory Ser. A 10, 327–352 (1973)
14. Herrera-Viedma, E., Herrera, F., Chiclana, F., Luque, M.: Some issues on consistency of fuzzy preference relations. European J. Oper. Res. 154, 98–109 (2004)
15. Kahn, J., Yu, Y.: Log-concave functions and poset probabilities. Combinatorica 18, 85–99 (1998)
16. Laslier, J.-F.: Tournament Solutions and Majority Voting, vol. 7. Springer, Berlin (1997)
17. Nelsen, R.: An Introduction to Copulas, 2nd edn. Lecture Notes in Statistics, vol. 139. Springer, New York (2005)

Geometry of Cores of Submodular Coherent Upper Probabilities and Possibility Measures

Tomáš Kroupa[1,2]

[1] Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic
[2] Department of Mathematics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

Abstract. We study and review geometrical properties of the set of the probabilities dominated by a submodular coherent upper probability (a possibility measure, in particular) on a finite set. We mention that there exists a polynomial algorithm for vertex enumeration. A new upper bound for the number of vertices in case of possibility measures is derived.

Keywords: Coherent upper probability, Possibility measure, Core.

1 Introduction

In this contribution we study and review geometrical properties of the set of the probabilities dominated by a submodular coherent upper probability on a finite set. The aim is to identify those geometrical and algebraical properties that lead to the existence of efficient algorithms for processing the imprecise probability. One of the tasks of eminent importance is that of recovering the extreme points of the dominated set of probabilities or, at least, estimating their number. The class of imprecise probabilities for which both of these tasks are solvable are submodular coherent upper probabilities. We make an effort to single out some of its subclasses (such as possibility measures) to show that they exhibit additional properties from the geometrical viewpoint.

Basic concepts of imprecise probabilities [14] and possibility theory [4] are recalled in Sect. 2. The structure of cores is investigated in Sect. 3 and conclusions are in Sect. 4. In the paper we use definitions and results concerning polytopes (see [18], for example).

2 Basic Notions

Let $N = \{1,\dots,n\}$ be a finite set with $n \geq 2$ and let 2^N denotes the set of all subsets of N. A *set function* on 2^N is a mapping $\mu : 2^N \to \mathbb{R}$ with $\mu(\emptyset) = 0$.

We say that a set function is *monotone* when $\mu(A) \leq \mu(B)$ for every $A,B \in 2^N$ such that $A \subseteq B$; it is called *submodular* (or 2-*alternating capacity*) if the inequality $\mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B)$ holds for every $A,B \in 2^N$. When P is a probability measure on 2^N, then p denotes the corresponding *probability distribution* on N, that is, the n-dimensional vector whose i-th component p_i is $P(\{i\})$, for every $i \in N$. For any set function μ with $\mu(N) = 1$, the set $\mathscr{M}(\mu)$ of probability distributions p on N with

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 306–312, 2008.
springerlink.com

$P(A) \leq \mu(A)$, for every $A \subseteq N$, is called the *core* (or *credal set*) of μ. Hence the core $\mathscr{M}(\mu)$ is the set of n-dimensional vectors $p \in \mathbb{R}^n$ satisfying the conditions

$$\sum_{i \in A} p_i \leq \mu(A), \quad \text{for every } A \in 2^N, \tag{1a}$$

$$p_i \geq 0, \quad i = 1, \ldots, n, \tag{1b}$$

$$\sum_{i=1}^{n} p_i = 1. \tag{1c}$$

Hence $\mathscr{M}(\mu)$ is a (possibly empty) convex polytope in $\mathbb{R}^n$ of dimension at most $n-1$. By $\operatorname{ext}\mathscr{M}(\mu)$ we denote the set of all extreme points (vertices) of $\mathscr{M}(\mu)$, which is always finite because of the finite number of the affine constraints (1).

A *coherent upper probability* is a set function $\mu : 2^N \to [0,1]$ such that $\mu(A) = \sup\{\sum_{i \in A} p_i \mid p \in \mathscr{M}(\mu)\}$, for every $A \in 2^N$, and $\mathscr{M}(\mu) \neq \emptyset$. It can be deduced from a result of Walley [15, p. 14] that every submodular monotone set function μ with $\mu(N) = 1$ is a coherent upper probability. On the contrary not every coherent upper probability is submodular—see [7, Remark 5.3].

A *possibility measure* is a set function $\Pi : 2^N \to [0,1]$ such that $\Pi(N) = 1$, and where for every $A, B \in 2^N$, it holds that $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$. A *possibility distribution* on N is the n-dimensional vector π whose i-th coordinate π_i is $\Pi(\{i\})$, for every $i \in N$. Every possibility measure is submodular monotone and thus a coherent upper probability.

3 Geometry of Cores

The polytope $\mathscr{M}(\mu)$ is completely determined as the convex hull of the set of its vertices $\operatorname{ext}\mathscr{M}(\mu)$. *Vertex enumeration* is the problem of generating all the vertices of a polytope given as an intersection of finitely many halfspaces. This is a fundamental problem in computational geometry, which is in general algorithmically highly nontrivial. Nevertheless there exist efficient techniques for certain classes of polytopes, which we mention in this section. It is therefore desirable to study the geometrical structure of the cores for particular classes of coherent upper probabilities in order to exploit their special properties. Since the time and the space complexity of the enumeration algorithms can also be judged a priori by estimating the number of vertices of the core, a goal in itself is to find some upper bounds for the number of vertices.

3.1 Cores of Submodular Coherent Upper Probabilities

The next theorem, which is considered to be well-known, gives a characterization of the core of submodular coherent upper probabilities. As far as the knowledge of the author goes, it can be traced back to Edmond's result [5] concerning the so-called base polyhedra in polymatroid[1] theory [6]. Similar results appeared in coalition game theory

[1] A *polymatroid* is a pair $(\mu, 2^N)$, where μ is a submodular monotone set function on 2^N. When $\mu(N) = 1$, then a so-called *base polyhedron* of the polymatroid $(\mu, 2^N)$ is precisely the core of μ in the sense of (1).

and imprecise probabilities too, cf. [12] and [2]. The formulation below is based on the polymatroid version of the theorem from [6, Section 3.3].

Theorem 1. *Let μ be a submodular coherent upper probability on 2^N. Then $\mathscr{M}(\mu)$ coincides with the set of vectors $p \in \mathbb{R}^n$ satisfying the conditions (1a) and (1c).*

A vector $p \in \mathbb{R}^n$ is an element of $\mathrm{ext}\mathscr{M}(\mu)$ if and only if there exists an $(n+1)$-tuple of sets $A_0, \ldots, A_n$ belonging to 2^N such that

$$\emptyset = A_0 \subset A_1 \subset \cdots \subset A_n = N,$$

where $A_i \setminus A_{i-1} = \{a_i\}$, for each $i = 1, \ldots, n$, and

$$p_{a_i} = \mu(A_i) - \mu(A_{i-1}), \quad \text{for each } i \in N. \tag{2}$$

The complete description of the facial structure of $\mathscr{M}(\mu)$ can be found in [6, Theorem 3.30] (see also [13]). Note that the above theorem enables to dispense completely with the nonnegativity conditions (1b) defining the core $\mathscr{M}(\mu)$. This result justifies the game-theoretic terminology employed when calling the set $\mathscr{M}(\mu)$ "core": for every submodular coherent upper probability μ, the set $\mathscr{M}(\mu)$ coincides precisely with the core of the game μ as studied in coalition game theory. This fact however depends on the economic interpretation of the game μ since the game-theoretic core is usually defined with the reversed inequality in (1a) provided that the values of μ are profits resulting from the cooperation.

The present inequality in (1a) thus implies that the value $\mu(A)$ should be thought of as a loss inflicted to a coalition A rather than the profit generated by the coalition A. While the latter interpretation of coalition games is more common (cf. [12]), the first one also appears in the literature (for example, in the foundational Aubin's paper [1] about games with fuzzy coalitions).

Formula (2) also leads to a very inefficient algorithm for enumerating the vertices of $\mathscr{M}(\mu)$ that is based on generating all permutations of the elements of the set N. There exists, however, a vertex enumeration technique by Zhan [17], which is well-tailored to cores of submodular coherent upper probabilities. Zhan's algorithm is polynomial and enumerates all the vertices of $\mathscr{M}(\mu)$ in $O(n^3|\mathrm{ext}\mathscr{M}(\mu)|)$ time and in $O(n^2)$ space. We refrain from describing even the basic ideas of this sophisticated algorithm, which generalizes several enumeration methods. An interested reader is referred to [17] for the comprehensive details.

Submodularity is one of the properties enhancing the performance of the enumeration algorithms. Another property of a coherent upper probability leading to tractable computations is rather an intrinsic geometric property of its core: a d-dimensional polytope is called *simple* when each vertex is contained in precisely d facets.

The d-dimensional cube or simplex are examples of simple polytopes; the pyramid with a non-triangular base is not a simple polytope. Simplicity of the core enables us to recover the vertices efficiently since there exist enumerating algorithms running in polynomial time per vertex for the class of all simple polytopes (see [8] for details and references therein). The author of this paper proved recently in [8] that the core of every possibility measure is a simple polytope (see also Theorem 2 in this paper).

A very special core geometry arises from the example of Wallner in [16, p.347, Fig.2]: put $\mu(A) = 1 - f\left(\frac{n-|A|}{n}\right)$, where $f : [0,1] \to [0,1]$ is a strictly convex function

with $f(0)=0, f(1)=1$. Namely, it is not difficult to prove that the core $\mathscr{M}(\mu)$ investigated in that example is combinatorially equivalent[2] to the so-called permutahedron. Let S_n be the set of all permutations of N. An $(n-1)$-*dimensional permutahedron* is the convex hull of the set $\{(\sigma(1),\ldots,\sigma(n)) \mid \sigma \in S_n\}$. The permutahedra are very rare among all polytopes: every $(n-1)$-dimensional permutahedron is a simple polytope that is an affine projection of the $\binom{n}{2}$-dimensional cube [18, p.17].

It follows directly from Theorem 1 there are at most $n!$ vertices in the core of every submodular coherent upper probability. More generally, the result of Wallner [16, Theorem 5.13] even shows that the submodularity condition can be relaxed so the upper bound is $n!$ for every coherent upper probability.

The next section is devoted to possibility measures for which some upper bounds for the number of vertices of their cores will be derived.

3.2 Cores of Possibility Measures

When π is a possibility distribution such that $\pi_i = 0$ for every $i \in I$ with $I \subset N$, then, by the simple projection argument, the core of the possibility measure Π can be identified with the core of the possibility measure Π' such that π' is defined as the restriction of π to $N \setminus I$. Without loss of generality, we start with the following convention.

Convention. From now on we assume that $0 < \pi_1 \leq \cdots \leq \pi_n = 1$.

Moral proved in [10] that $p \in \mathscr{M}(\Pi)$ if and only if p is a probability distribution such that $\sum_{j=1}^{k} p_j \leq \pi_k$, for each $k = 1,\ldots,n$. The cores of possibility measures were characterized in [8], where the proof of Theorem 2 can be found. Put

$$S = \{i \in \{1,\ldots,n-2\} \mid \pi_{i+1} > \pi_i\} \cup \{n-1\}.$$

Theorem 2. *The core $\mathscr{M}(\Pi)$ of a possibility measure Π is a simple $(n-1)$-dimensional polytope such that $p \in \mathscr{M}(\Pi)$ if and only if*

$$\sum_{j=1}^{k} p_j \leq \pi_k, \quad k \in S, \tag{3a}$$

$$p_i \geq 0, \quad i = 1,\ldots,n-1, \tag{3b}$$

$$p_n = 1 - \sum_{i=1}^{n-1} p_i. \tag{3c}$$

Moreover, the polytope $\mathscr{M}(\Pi)$ has $n-1+|S|$ facets given by $\mathscr{K} \cap \mathscr{M}(\Pi)$, where $\mathscr{K}$ is either $\{p \in \mathbb{R}^n \mid p_i = 0\}$ or $\{p \in \mathbb{R}^n \mid \sum_{j=1}^{k} p_j = \pi_k\}$, for each $i = 1,\ldots,n-1$ and for each $k \in S$, respectively.

The representation of $\mathscr{M}(\Pi)$ by the system (3) is *irreducible*, that is, removing any inequality or equation from (3) changes the set $\mathscr{M}(\Pi)$. While the irreducible

[2] We say that two polytopes are *combinatorially equivalent* if there exists an order-preserving bijection between their face lattices.

representation (3) is easily constructed[3] for every possibility measure, it can be shown that this is not the only irreducible representation of $\mathscr{M}(\Pi)$. The representation by system (3) is however useful for obtaining some upper bounds for the number of vertices. Miranda et al. [9] derived the exponential upper bound 2^{n-1} for the number $|\mathrm{ext}\mathscr{M}(\Pi)|$ of vertices of any possibility measure Π. Using only the simplicity of $\mathscr{M}(\Pi)$, the following lower bound and the upper bound for $|\mathrm{ext}\mathscr{M}(\Pi)|$ were obtained in [8][4]:

$$|S|(n-2)+2 \leq |\mathrm{ext}\mathscr{M}(\Pi)| \leq \binom{n-2+|S|-r_1}{r_2} + \binom{n-2+|S|-r_2}{r_1}, \tag{4}$$

where r_1 is the greatest integer such that $r_1 \leq \frac{n-2}{2}$, and r_2 is the greatest integer such that $r_2 \leq \frac{n-1}{2}$. It was shown in [8] that the upper bound from (4) is not uniformly better than the exponential bound. In the sequel we will show that there exists an easily computable upper bound that is always lower than the exponential bound. To this end, put $i_0 = 0$ and let $i_1, \ldots, i_{|S|}$ denote the elements of S such that $i_j < i_{j+1}$, for each $j = 1, \ldots, |S|-1$.

Theorem 3. *If $\mathscr{M}(\Pi)$ is the core of a possibility measure Π, then*

$$|\mathrm{ext}\mathscr{M}(\Pi)| \leq 2^{|S|} \prod_{j=1}^{|S|} (i_j - i_{j-1}). \tag{5}$$

The expression on the right-hand side of (5) is always smaller or equal to 2^{n-1} with the equality holding when Π is such that $\pi_i < \pi_{i+1}$, for each $i = 1, \ldots, n-2$.

Proof. It follows from Theorem 2 and the definition of a vertex that there is a one-to-one correspondence between the vertices of $\mathscr{M}(\Pi)$ and uniquely solvable systems of $n-1$ linear equations selected from

$$\sum_{j=1}^{k} p_j = \pi_k, \quad k \in S, \tag{6a}$$

$$p_i = 0, \quad i = 1, \ldots, n-1. \tag{6b}$$

Hence it suffices to bound from above the total number of such uniquely solvable systems. Note that (6a) can be equivalently written as

$$\begin{aligned} p_1 + \ldots + p_{i_1} &= \pi_{i_1} \\ p_{i_1+1} + \ldots + p_{i_2} &= \pi_{i_2} - \pi_{i_1} \\ &\ddots \\ p_{i_{|S|-1}+1} + \ldots + p_{i_{|S|}} &= \pi_{i_{|S|}} - \pi_{i_{|S|-1}} \end{aligned} \tag{7}$$

A uniquely solvable linear system arising from (7) by appending arbitrarily chosen equations from (6b) is called an *initial linear system*. First, we will count the total

[3] Detecting whether an inequality is redundant in a description of a polytope is a nontrivial problem, cf. [18, p. 48].

[4] Unfortunately, the formula for the upper bound is misprinted in [8, Theorem 2].

number of initial linear systems. Since the right-hand sides of (7) are positive due to our Convention, it follows that precisely one variable in each linear equation of (7) must be nonzero. This means that there are

$$i_1(i_2 - i_1)\dots(i_{|S|} - i_{|S|-1}) = \prod_{j=1}^{|S|}(i_j - i_{j-1}) \tag{8}$$

initial linear systems.

To finish the proof of (5), observe that every uniquely solvable system of $n-1$ linear equations chosen from (6a)-(6b) corresponds to precisely one initial linear system in which some of the $|S|$ equations from (7) are possibly interchanged with those from (6b). Since there are precisely $2^{|S|}$ of all interchanges, there can be at most $2^{|S|}$ of the uniquely solvable systems resulting in this way from an initial linear system. Combining this with (8), the inequality (5) follows.

In order to prove the second assertion of the theorem, observe that $\sum_{j=1}^{|S|}(i_j - i_{j-1}) = i_{|S|} = n-1$. We obtain

$$2^{|S|}\prod_{j=1}^{|S|}(i_j - i_{j-1}) = \prod_{j=1}^{|S|}2(i_j - i_{j-1}) \le \prod_{j=1}^{|S|}2^{i_j - i_{j-1}} = 2^{\sum_{j=1}^{|S|}(i_j - i_{j-1})} = 2^{n-1},$$

since $2p \le 2^p$ for every positive integer p. □

In general, the upper bound from Theorem 3 is not better than the upper bound from (4) and vice versa. Note that the first part of the proof of Theorem 3 is based on a particular vertex enumeration algorithm, which can be used to recover the vertices when n and $|S|$ are rather small. However, this technique is of a very limited use as it recovers the vertices in time exponential in $|S|$.

4 Conclusions

In the paper we presented the algebraical and the geometrical properties of imprecise probabilities that facilitate computations with their cores. Namely, there exist polynomial algorithms for vertex enumeration of both submodular coherent upper probabilities and coherent upper probabilities whose core is a simple polytope. The upper bounds from Sect. 3.2 provide the preliminary judgment on the performance of such algorithms. There are many theoretical and practical facets of imprecise probabilities in which the computational properties of the core play a crucial role. For example, a concept of (conditional) independence can be defined with the help of extreme points of the core [3, 11].

An interesting open question is to describe the class of coherent upper probabilities whose cores are simple polytopes.

Acknowledgement. The author owes his gratitude to his colleague Fero Matúš who attracted his attention to the connection between the cores of imprecise probabilities and the base polyhedra in polymatroid theory. The work of the author was supported by the grant No.1M0572 of the Ministry of Education, Youth and Sports of the Czech Republic.

References

1. Aubin, J.P.: Coeur et valeur des jeux flous à paiements latéraux. C. R. Acad. Sci. Paris Sér. A. 279, 891–894 (1974)
2. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. Math. Social Sci. 17(3), 263–283 (1989)
3. Couso, I., Moral, S., Walley, P.: Examples of independence for imprecise probabilities. In: De Cooman, G., Cozman, F.G., Moral, S., Walley, P. (eds.) Proceedings of the 1st International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 1999, Ghent, Belgium), pp. 121–130 (1999)
4. Dubois, D., Prade, H.: Possibility theory. Plenum Press, New York (1988)
5. Edmonds, J.: Submodular functions, matroids, and certain polyhedra. In: Guy, R.K., Hanani, H., Sauer, N., Schönheim, J. (eds.) Combinatorial Structures and Their Applications. Proceedings of the Calgary International Conference on Combinatorial Structures and their Applications (Calgary, Alberta, Canada,1969), pp. 69–87. Gordon and Breach, New York (1970)
6. Fujishige, S.: Submodular Functions and Optimization. Annals of Discrete Mathematics, vol. 47. North-Holland Publishing Co, Amsterdam (1991)
7. Krätschmer, V.: Coherent lower previsions and Choquet integrals. Fuzzy Sets Syst. 138(3), 469–484 (2003)
8. Kroupa, T.: Geometry of possibility measures on finite sets. Internat. J. Approx. Reason 48(1), 237–245 (2008)
9. Miranda, E., Couso, I., Gil, P.: Extreme points of credal sets generated by 2-alternating capacities. Internat. J. Approx. Reason 33(1), 95–115 (2003)
10. Moral, S.: Información Difusa. Relaciones Entre Probabilidad y Posibilidad. PhD Thesis. Universidad de Granada, Spain (1985)
11. Moral, S., Cano, A.: Strong conditional independence for credal sets. Ann. Math. Artif. Intell. 35(1-4), 295–321 (2002)
12. Shapley, L.S.: Cores of convex games. Internat. J. Game Theory 1, 11–26 (1972)
13. Verdegay-López, J.F., Moral, S.: Network of probabilities associated with a capacity of order-2. Inform. Sci. 125(1-4), 187–206 (2000)
14. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Monographs on Statistics and Applied Probability, vol. 42. Chapman and Hall Ltd., London (1991)
15. Walley, P.: Measures of uncertainty in expert systems. Artificial Intelligence 83(1), 1–58 (1996)
16. Wallner, A.: Extreme points of coherent probabilities in finite spaces. Internat. J. Approx. Reason 44(3), 339–357 (2007)
17. Zhan, P.: A polynomial algorithm for enumerating all vertices of a base polyhedron. J. Oper. Res. Soc. Japan 40(3), 329–340 (1997)
18. Ziegler, G.: Lectures on polytopes. Graduate Texts in Mathematics, vol. 152. Springer, New York (1995)

On Transformations between Belief Spaces

Ralph L. Wojtowicz

Metron Inc., Reston, VA, USA

Abstract. Commutative monoids of belief states have been defined by imposing one or more of the usual axioms and assigning a combination rule. Familiar operations such as normalization and the Voorbraak map are surjective homomorphisms. The latter map takes values in a monoid of Bayesian states. The pignistic map is not a monoid homomorphism. This can impact robust decision making for frames of cardinality at least 3. We adapt the concept of measure zero reflecting functions between probability spaces to define a category **P0R** having belief states as objects and plausibility zero reflecting functions as morphisms. This definition encapsulates a generalization of the notion of absolute continuity to the context of belief spaces. We show that the Voorbraak map induces a functor valued in **P0R** that is right adjoint to the embedding of Bayesian states.

Keywords: Belief state, category, adjoint, Voorbraak map, pignistic transformation, absolute continuity.

1 Preliminaries

This article describes results from an effort to improve the performance of a large information system having components implemented by different companies and employing distinct uncertainty models. We seek concepts and theorems that help precisely characterize properties of and relationships between these models.

For a set U, $\mathscr{P}(U)$ is its powerset and $|U|$ is its cardinality. If $V \subset U$, then $U \backslash V$ is the complement of V in U. $\mathbb{R}$ is the set of real numbers and ϕ is the empty set. For a function $f : A \to B$ with source A and target B, $f^* : \mathscr{P}(B) \to \mathscr{P}(A)$ is its inverse image function and $f_* : \mathscr{P}(A) \to \mathscr{P}(B)$ is its direct image function. If $U \subset A$, for example, then $f_*(U) = \{b \in B \mid \exists a \in U.\ f(a) = b\}$. If f is a bijection then $f^{-1} : B \to A$ is its inverse. As in [5], we make frequent, implicit use of the following corollary of the Binomial Theorem.

Lemma 1. *If X is a finite set and $U \subset W \subset X$, then*

$$\sum_{U \subset V \subset W} (-1)^{|V|} = \begin{cases} (-1)^{|W|} & \text{if } U = W \\ 0 & \text{if } U \neq W. \end{cases}$$

Throughout this article, X will denote a finite, non-empty set to be called a *frame*.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 313–320, 2008.
springerlink.com

2 Belief Representations

A function $m : \mathscr{P}(X) \to \mathbb{R}$ is a *belief state*. The set of all such is $\mathscr{X}$. For $U \subset X$, $\nu_U \in \mathscr{X}$ is defined by $\nu_U(U) = 1$ and $\nu_U(V) = 0$ if $V \neq U$. ν_X is *vacuous* and ν_ϕ is *conflicted*. The *zero* belief state assigns $0 \in \mathbb{R}$ to each $U \subset X$ and is denoted $0 \in \mathscr{X}$. $u \in \mathscr{X}$ defined by $u(U) = 1/|X|$ if $|U| = 1$ and 0 otherwise is *uniform*.

$m \in \mathscr{X}$ is *non-negative* if $0 \leq m(U)$ for all $U \subset X$. m is *unitary* if $1 = \sum_{U \subset X} m(U)$. Let $\mathscr{X}_+$ and $\mathscr{X}_\Sigma$ respectively be the sets of non-negative and unitary belief states. Let $\mathscr{X}_0 = \{m \in \mathscr{X} \,|\, m(\phi) = 0\}$. Define $\mathscr{X}_\mathscr{D} = \mathscr{X}_\Sigma \cap \mathscr{X}_+$, $\mathscr{X}_\mu = \mathscr{X}_\mathscr{D} \cap \mathscr{X}_0$, $\mathscr{X}_0' = \mathscr{X}_0 \cup \{\nu_\phi\}$, and $\mathscr{X}_\mu' = \mathscr{X}_\mu \cup \{\nu_\phi\}$. m is *Bayesian* if $m \in \mathscr{X}_\mu$ and $m(U) \neq 0$ implies $|U| = 1$. The uniform belief state u, for example, is Bayesian. Let $\mathscr{X}_\mathscr{B}$ be the set of Bayesian belief states and $\mathscr{X}_\mathscr{B}' = \mathscr{X}_\mathscr{B} \cup \{0\}$.

Alternative belief state representations are useful for proving theorems, gaining insight, and supporting interpretations. Define $\beta : \mathscr{X} \to \mathscr{X}$ and $\beta^{-1} : \mathscr{X} \to \mathscr{X}$ by

$$\beta(m)(U) = \sum_{V \subset U} m(V) \tag{1}$$

$$\beta^{-1}(b)(U) = \sum_{V \subset U} (-1)^{|U \setminus V|} b(V)\,. \tag{2}$$

$\beta(m)$ is the *implicability* of m. If $m \in \mathscr{X}_\mu$, then $\beta(m)$ is a *belief function*. Proof that β and β^{-1} are inverses employs Lemma 1. Other inverse function pairs described below are similarly established. Define $\kappa : \mathscr{X} \to \mathscr{X}$ and $\kappa^{-1} : \mathscr{X} \to \mathscr{X}$ by

$$\kappa(m)(U) = \sum_{U \subset V} m(V) \tag{3}$$

$$\kappa^{-1}(q)(U) = \sum_{U \subset V} (-1)^{|V \setminus U|} q(V)\,. \tag{4}$$

$\kappa(m)$ is the *commonality* of m. Define $\lambda : \mathscr{X} \to \mathscr{X}$ and $\lambda^{-1} : \mathscr{X} \to \mathscr{X}$ by

$$\lambda(m)(U) = \begin{cases} \displaystyle\sum_{V \cap U \neq \phi} m(V) & \text{if } U \neq \phi \\ \displaystyle 1 - \sum_{V \subset X} m(V) & \text{if } U = \phi \end{cases} \tag{5}$$

$$\lambda^{-1}(\ell)(U) = \begin{cases} \displaystyle\sum_{V \subset U} (-1)^{|U \setminus V|} \big(1 - \ell(X \setminus V)\big) & \text{if } U \neq \phi \text{ and } U \neq X \\ \displaystyle \ell(\phi) + \sum_{V \subset X} (-1)^{|X \setminus V|} \big(1 - \ell(X \setminus V)\big) & \text{if } U = X \\ 1 - \ell(X) - \ell(\phi) & \text{if } U = \phi\,. \end{cases} \tag{6}$$

$\lambda(m)$ is the *plausibility* of m. For example, $\lambda(\nu_\phi) = 0$ and $\lambda^{-1}(0) = \nu_\phi$. For each $x \in X$, $\lambda(m)(\{x\}) = \kappa(m)(\{x\})$. Define $\Lambda(m) = \sum_{x \in X} \lambda(m)(\{x\})$. The *Bayesian constant* of m is

$$\mathscr{B}(m) = \begin{cases} 1/\Lambda(m) & \text{if } \Lambda(m) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

If $m \in \mathscr{X}_{\mathscr{D}}$, then $0 = \Lambda(m)$ iff $m = \nu_\phi$. If $m \in \mathscr{X}_\mu$, then $1 \leq \Lambda(m)$ with equality holding iff $m \in \mathscr{X}_{\mathscr{B}}$. Composites of the formulas above yield translations between belief representations. Maps between implicability b and commonality q are

$$(\kappa \circ \beta^{-1})(b)(U) = \sum_{V \subset U} (-1)^{|V|}\, b(X \backslash V) \tag{8}$$

$$(\beta \circ \kappa^{-1})(q)(U) = \sum_{V \subset X \backslash U} (-1)^{|V|}\, q(V). \tag{9}$$

Transformations between implicability b and plausibility ℓ are

$$(\beta \circ \lambda^{-1})(\ell)(U) = \begin{cases} 1 - \ell(X \backslash U) - \ell(\phi) & \text{if } U \neq \phi \\ 1 - \ell(\phi) & \text{otherwise} \end{cases} \tag{10}$$

$$(\lambda \circ \beta^{-1})(b)(U) = \begin{cases} b(X) - b(X \backslash U) & \text{if } U \neq X \\ 1 - b(X) & \text{otherwise.} \end{cases} \tag{11}$$

Between commonality q and plausibility ℓ the following hold

$$(\lambda \circ \kappa^{-1})(q)(U) = \begin{cases} \sum\limits_{\phi \neq V \subset U} (-1)^{|V|+1}\, q(V) & \text{if } U \neq \phi \\ 1 - q(\phi) & \text{otherwise} \end{cases} \tag{12}$$

$$(\kappa \circ \lambda^{-1})(\ell)(U) = \begin{cases} \sum\limits_{\phi \neq V \subset U} (-1)^{|V|+1}\, \ell(V) & \text{if } U \neq \phi \\ 1 - \ell(\phi) & \text{otherwise.} \end{cases} \tag{13}$$

The following conditions on $m \in \mathscr{X}$ are equivalent: m is unitary; $1 = \sum_{U \subset X} m(U)$; $\beta(m)(X) = 1$; $\kappa(m)(\phi) = 1$; $\lambda(m)(\phi) = 0$. The following are also equivalent: $m(\phi) = 0$; $\beta(m)(\phi) = 0$; $\lambda(m)(X) = \beta(m)(X)$; $0 = \sum_{V \subset X} (-1)^{|V|}\, \kappa(m)(V)$.

3 Combination Operators

The *unnormalized combination* operator $\odot : \mathscr{X} \times \mathscr{X} \to \mathscr{X}$ defined by

$$(m \odot n)(U) = \sum_{A \cap B = U} m(A)\, n(B) \tag{14}$$

satisfies the following for all m, n, and $p \in \mathscr{X}$.

1. $m \odot n = n \odot m$
2. $m \odot \nu_X = m$
3. $(m \odot \nu_\phi)(U) = \nu_\phi(U) \sum_{A \subset X} m(A)$
4. $\kappa(m \odot n)(U) = \kappa(m)(U) \cdot \kappa(n)(U)$ for all $U \subset X$
5. $m \odot (n \odot p) = (m \odot n) \odot p$
6. $\lambda(m \odot n)(\{x\}) = \lambda(m)(\{x\}) \cdot \lambda(n)(\{x\})$ for all $x \in X$

Proofs of 1–3 are direct calculations. 4 follows from equality of $\sum\limits_{U \subset V} (m \odot n)(V)$ to

$$\sum_{U\subset V}\sum_{A\cap B=V} m(A)\,n(B) = \sum_{U\subset A}\sum_{U\subset B} m(A)\,n(B) = \left(\sum_{U\subset A} m(A)\right)\left(\sum_{U\subset B} n(B)\right).$$

Property 5 follows from 4 and the fact that κ and κ^{-1} are inverse functions:

$$(m\odot n)\odot p = \kappa^{-1}\big(\kappa((m\odot n)\odot p\big) = \kappa^{-1}(\kappa(m)\,\kappa(n)\,\kappa(p)) = \kappa^{-1}\big(\kappa(m\odot(n\odot p))\big).$$

Property 6 follows from 4 and the fact that $\kappa(m)$ and $\lambda(m)$ agree on singletons.

The *combination* operator $\oplus : \mathscr{X}\times\mathscr{X}\to\mathscr{X}$ is defined by

$$(m\oplus n)(U) = \begin{cases} \frac{(m\odot n)(U)}{1-(m\odot n)(\phi)} & \text{if } (m\odot n)(\phi)\neq 1 \text{ and } U\neq\phi \\ 0 & \text{if } (m\odot n)(\phi)\neq 1 \text{ and } U=\phi \\ \nu_\phi(U) & \text{otherwise} \end{cases} \tag{15}$$

and satisfies the following for all m, n, and $p\in\mathscr{X}$.

1. $m\oplus n\in\mathscr{X}_0'$
2. $m\oplus n = n\oplus m$
3. $m\oplus\nu_X = m$ iff $m\in\mathscr{X}_0'$
4. $m\oplus\nu_\phi = \nu_\phi$ iff $m\in\mathscr{X}_\Sigma$
5. if $m,n\in\mathscr{X}_\mathscr{D}$, then
$$\kappa(m\oplus n)(U) = \begin{cases} \kappa(m)(U)\,\kappa(n)(U)/\big(1-(m\odot n)(\phi)\big) & \text{if } m\odot n\neq\nu_\phi,\ U\neq\phi \\ \nu_\phi(U) & \text{otherwise} \end{cases}$$
6. if $m,n\in\mathscr{X}_\mathscr{D}$, then $m\oplus n\in\mathscr{X}_\mathscr{D}$
7. if $m,n,p\in\mathscr{X}_\mu'$, then $m\oplus(n\oplus p) = (m\oplus n)\oplus p$
8. if $m,n\in\mathscr{X}_\mathscr{B}'$, then $m\oplus n\in\mathscr{X}_\mathscr{B}'$
9. if $m\in\mathscr{X}_\mathscr{B}'$, then $m\oplus u = m$

Proofs are left to the reader [2, 5, 6]. For $m\in\mathscr{X}$, define $\odot_0 m = m$ and, for $n\geq 1$, $\odot_n m = m\odot(\odot_{n-1} m)$. Similar definitions apply to iteration of $\oplus$.

4 Belief State Monoids and Homomorphisms

We review the commutative monoids $(\mathscr{X}_+,\odot,\nu_X)$, $(\mathscr{X}_\mathscr{D},\odot,\nu_X)$, and $(\mathscr{X}_\mu',\oplus,\nu_X)$ and the monoid homomorphisms Φ_Σ and Φ_0 introduced in [2] and the commutative monoid $(\mathscr{X}_\mathscr{B}',\oplus,u)$ and the homomorphism $\mathfrak{V}$ of [6].

Restriction of $\odot$ gives an operator $\mathscr{X}_+\times\mathscr{X}_+\to\mathscr{X}_+$. Properties 1, 2, and 5 of $\odot$ from Section 3 establish that $(\mathscr{X}_+,\odot,\nu_X)$ is a commutative monoid. Another restriction of $\odot$ produces $\mathscr{X}_\Sigma\times\mathscr{X}_\Sigma\to\mathscr{X}_\Sigma$. Since $\mathscr{X}_\mathscr{D} = \mathscr{X}_+\cap\mathscr{X}_\Sigma$, $(\mathscr{X}_\mathscr{D},\odot,\nu_X)$ is also a commutative monoid. $m\odot\nu_\phi = \nu_\phi$ for $m\in\mathscr{X}_\mathscr{D}$ by property 3 of $\odot$.

The *normalization operator* $\Phi_\Sigma : \mathscr{X}_+\to\mathscr{X}_\mathscr{D}$ defined by

$$\Phi_\Sigma(m)(U) = \begin{cases} m(U)\Big/\displaystyle\sum_{V\subset X} m(V) & \text{if } m\neq 0 \\ \nu_\phi(U) & \text{otherwise} \end{cases} \tag{16}$$

is a homomorphism $(\mathscr{X}_+, \odot, \nu_X) \to (\mathscr{X}_\mathscr{D}, \odot, \nu_X)$ (see [2]). That is, $\Phi_\Sigma(m \odot n) = \Phi_\Sigma(m) \odot \Phi_\Sigma(n)$ and $\Phi_\Sigma(\nu_X) = \nu_X$. Moreover, if $m \in \mathscr{X}_\mathscr{D}$, then $\Phi_\Sigma(m) = m$.

Restriction of $\oplus$ gives an operator $\mathscr{X}'_\mu \times \mathscr{X}'_\mu \to \mathscr{X}'_\mu$ by properties 1 and 6 of $\oplus$ from Section 3. Properties 2, 3, and 7 establish that $(\mathscr{X}'_\mu, \oplus, \nu_X)$ is a commutative monoid. $m \oplus \nu_\phi = \nu_\phi$ for $m \in \mathscr{X}'_\mu$ by property 4. Another restriction of $\oplus$ gives an operator $\mathscr{X}'_\mathscr{B} \times \mathscr{X}'_\mathscr{B} \to \mathscr{X}'_\mathscr{B}$ by property 8. These results together with property 9 establish that $(\mathscr{X}'_\mathscr{B}, \oplus, u)$ is a commutative monoid. Moreover, $m \oplus 0 = 0$ in $\mathscr{X}'_\mathscr{B}$.

The operator $\Phi_0 : \mathscr{X}_\mathscr{D} \to \mathscr{X}'_\mu$ defined below is a homomorphism $(\mathscr{X}_\mathscr{D}, \odot, \nu_X) \to (\mathscr{X}'_\mu, \oplus, \nu_X)$ satisfying $\Phi_0(m) = m$ for $m \in \mathscr{X}'_\mu$ (see [2]).

$$\Phi_0(m)(U) = \begin{cases} m(U)/(1 - m(\phi)) & \text{if } m(\phi) \neq 1 \text{ and } U \neq \phi \\ 0 & \text{if } m(\phi) \neq 1 \text{ and } U = \phi \\ \nu_\phi(U) & \text{otherwise} \end{cases} \tag{17}$$

The Voorbraak map $\mathfrak{V} : \mathscr{X}'_\mu \to \mathscr{X}'_\mathscr{B}$ defined by

$$\mathfrak{V}(m)(U) = \begin{cases} \lambda(m)(\{x\})/\Lambda(m) & \text{if } \Lambda(m) \neq 0 \text{ and } U = \{x\} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

gives a monoid homomorphism $(\mathscr{X}_\mu, \oplus, \nu_X) \to (\mathscr{X}_\mathscr{B}, \oplus, u)$ such that $\mathfrak{V}(m) = m$ for $m \in \mathscr{X}'_\mathscr{B}$ (see [6]). Note that $m \in \mathscr{X}_\mathscr{D} \subset \mathscr{X}'_\mu$ implies $\Lambda(m) = 0$ iff $m = \nu_\phi$.

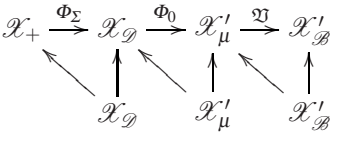
Conversion of $m \in \mathscr{X}_\mu$ to a Bayesian state is an approach to decision making with belief models. The pignistic map $\mathfrak{P} : \mathscr{X}_\mu \to \mathscr{X}_\mathscr{B}$ is not a monoid homomorphism. In [1], Cobb and Shenoy observed that $\mathfrak{P}$ may display a decision change when applied to iterates $\oplus_k m$ of a fixed m. The reader may compute $\mathfrak{V}(\oplus_n m)$ and $\mathfrak{P}(\oplus_n m)$ with m defined by the following for sufficiently small ε and δ to observe this phenomenon.

U	ϕ	$\{x\}$	$\{y\}$	$\{z\}$	$\{x,y\}$	$\{x,z\}$	$\{y,z\}$	X
$m(U)$	0	0	δ	$\delta+\varepsilon$	$\delta+2\varepsilon$	$\delta+\varepsilon$	0	$1-4(\delta+\varepsilon)$

The diagrams below summarize the results of this section. The first illustrates the commutative monoids and homomorphisms.

$$(\mathscr{X}_+, \odot, \nu_X) \xrightarrow{\Phi_\Sigma} (\mathscr{X}_\mathscr{D}, \odot, \nu_X) \xrightarrow{\Phi_0} (\mathscr{X}'_\mu, \oplus, \nu_X) \xrightarrow{\mathfrak{V}} (\mathscr{X}'_\mathscr{B}, \oplus, u)$$

The underlying functions of these homomorphisms are the horizontal arrows of the commutative diagram below. Vertical arrows are identity functions and diagonals are inclusions.



Surjectivity of Φ_Σ, Φ_0, and $\mathfrak{V}$ are consequences of commutativity (in any category, $g \circ f$ an epimorphism implies g an epimorphism).

5 Categories of Belief States

If $\mathscr{C}$ is a (locally-small) category, then $|\mathscr{C}|$ is its class of objects and $\mathscr{C}(A,B)$ is the set of morphisms with source A and target B. Categorical definitions are from [4].

In [7], Wendt defined and investigated two categories having probability spaces as objects. The first has measure zero reflecting functions as morphisms while the latter has categorical disintegrations. These constructions led Jackson to implicitly establish a semantics of intuitionistic higher-order predicate calculus in terms of probability spaces [3]. Such connections between rules-based systems and uncertainty models are of interest for their potential applications to information fusion systems and human/machine interfaces. Below we take first steps in adapting Wendt's constructions to belief models.

A *probability density* on a finite, non-empty set X is a function $p : X \to \mathbb{R}$ satisfying $0 \leq p(x)$ and $1 = \sum_{x\in X} p(x)$. A *probability space* is a pair (X,p) with p probability density on X. Assume such spaces are equipped with powerset σ-algebras. Let $\widehat{p}$ be the induced measure defined by $\widehat{p}(U) = \sum_{x\in U} p(x)$. If (X,p) and (Y,q) are probability spaces, $f : X \to Y$ is *measure zero reflecting* if, given $V \subset Y$, $\widehat{q}(V) = 0$ implies $\widehat{p}(f^*(V)) = 0$ (i.e., $\widehat{p} \circ f^*$ on Y is absolutely continuous with respect to $\widehat{q}$). Let **M0RP** be the category having probability spaces as objects and measure zero reflecting functions as morphisms. In **M0RP**, identity functions are identity morphisms and composition is function composition. **M0RP** a proper subcategory of **M0R** defined in [7].

A *belief space* is a pair (X,m) with X finite, non-empty set and $m \in \mathscr{X}_\mu$. If (X,m) and (Y,n) are belief spaces, a function $f : X \to Y$ is *plausibility zero reflecting* if $\lambda(n)(V) = 0$ implies $\lambda(m)(f^*(V)) = 0$. for any $V \subset Y$. Let **P0R** be the category having belief spaces as objects and plausibility zero reflecting functions as morphisms. Identity morphisms and composition in **P0R** are defined as in **M0RP**. Composition is well defined: if $f \in \mathbf{P0R}\big((X,m),(Y,n)\big)$, $g \in \mathbf{P0R}\big((Y,n),(Z,r)\big)$, and $W \subset Z$, then $\lambda(r)(W) = 0$ implies $\lambda(n)\big(g^*(W)\big) = 0$ implies $\lambda(m)\big(f^*(g^*(W))\big) = \lambda(m)\big((g\circ f)^*(W)\big) = 0$.

Theorem 1. *There is a faithful functor* $D : \mathbf{M0RP} \to \mathbf{P0R}$ *defined on objects by*

$$D(X,p) = (X, d(p)) \tag{19}$$

where $d(p)(\{x\}) = p(x)$ *and* $d(p)(U) = 0$ *for* $|U| \neq 1$. *On morphisms,* $D(f) = f$.

Proof. $D(X,p) \in |\mathbf{P0R}|$ since $d(p) \in \mathscr{X}_{\mathscr{B}} \subset \mathscr{X}_\mu$. Observe that

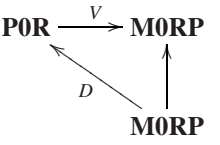
$$\lambda(d(p))(A) = \sum_{U\cap A\neq\phi} d(p)(U) = \sum_{x\in A} d(p)(\{x\}) = \sum_{x\in A} p(x) = \widehat{p}(A)$$

for $\phi \neq A \subset X$. If $f \in \mathbf{M0RP}\big((X,p),(Y,q)\big)$, then $D(f)$ is plausibility zero reflecting since, $0 = \lambda(d(q))(B) = \widehat{q}(B)$ implies $0 = \widehat{p}\big(f^*(B)\big) = \lambda(d(p))\big(f^*(B)\big)$. □

Theorem 2. *The Voorbraak map induces a faithful functor* $V : \mathbf{P0R} \to \mathbf{M0RP}$ *defined on objects by*

$$V(X,m) = (X, v(m)) \tag{20}$$

where $v(m)(x) = \mathfrak{V}(m)(\{x\})$. *On morphisms* $V(f) = f$. *Moreover, the diagram below of categories and functors is commutative*

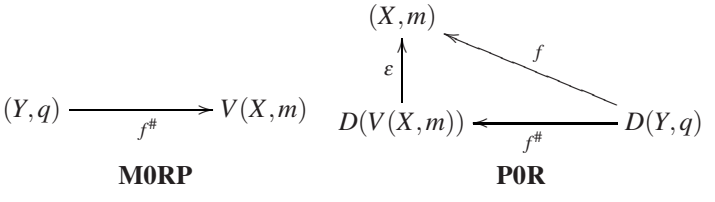


where the vertical arrow is the identity functor.

Proof. $V(X,m) \in |\mathbf{M0RP}|$ since $m \in \mathscr{X}_\mu$ implies $0 < \Lambda(m)$ which implies $\mathfrak{V}(m) \in \mathscr{X}_{\mathscr{B}}$. To verify that $V(f)$ is measure zero reflecting if $f \in \mathbf{P0R}\big((X,m),(Y,n)\big)$, note that $0 = \widehat{v(n)}(B) = \sum_{y\in B} v(n)(y) = \sum_{y\in U} \lambda(n)(\{y\})/\Lambda(n)$ implies $0 = \lambda(n)(\{y\})$ for each $y \in B$. This implies $0 = \lambda(m)(f^*(\{y\}))$ for each $y \in B$, hence, $0 = \widehat{v(m)}(f^*(\{y\}))$. Disjointness of the sets $f^*(\{y\})$ and additivity of $\widehat{v(m)}$ yields $0 = \widehat{v(m)}(f^*(U))$. Commutativity of the diagram implies that D is injective on objects, hence, **M0RP** can be identified with a subcategory of **P0R** using D as embedding. □

Theorem 3. *: The Voorbraak functor* V : **P0R** $\to$ **M0RP** *is right adjoint to the functor* D : **M0RP** $\to$ **P0R**.

Proof. For a belief space (X,m) we must show that the function $\varepsilon : X \to X$ defined by $\varepsilon(x) = x$ is a plausibility zero reflecting map $\varepsilon : D(V(X,m)) \to (X,m)$ and that, given any probability space (Y,q) and plausibility zero reflecting map $f : D(Y,q) \to (X,m)$, there is a unique $f^{\#} \in \mathbf{M0RP}\big((Y,q),V(X,m)\big)$ for which the diagram below right is commutative.



To establish the condition on ε, assume $\lambda(m)(U) = 0$. Note that

$$\lambda\big(d(v(m))\big)(U) = \widehat{v(m)}(U) = \sum_{x\in U} \lambda(m)(\{x\})/\Lambda(m).$$

If $x \in U$ and $0 < \lambda(m)(\{x\})$, then there exists $A \subset X$ with $x \in A$ and $0 < m(A)$. $x \in A$ implies $A \cap U \neq \phi$, hence $0 < m(A) \leq \lambda(m)(U)$ contradicting the assumption on U. This implies $\lambda\big(d(v(m))\big)(U) = 0$. is plausibility zero reflecting. The definition ε and the requirement of commutativity impose $f^{\#} = f$. $f^{\#}$ is measure zero reflecting and plausibility zero reflecting between the appropriate spaces. □

Dempster's Combination Rule induces a symmetric monoidal structure on **P0R** just as product measures induce such a structure on **M0RP**. The tensor product on **P0R** is defined as follows. Let $X \xleftarrow{\pi} X \times Y \xrightarrow{\pi'} Y$ be the cartesian product and associated projection functions given belief spaces, (X,m) and (Y,n). Define $(X,m) \otimes (Y,n) = (X \times Y, m \otimes n)$ where

$$m \otimes n = \Phi_0\big(\Phi_\Sigma(m \circ \pi_*)\big) \oplus \Phi_0\big(\Phi_\Sigma(n \circ \pi'_*)\big).$$

This monoidal structure and related functors will be topics of a later article.

Acknowledgement. The author would like to thank Dr. Roger Sullivan and Dr. Steve Bravy for their support of this and related work under SBIR contract HQ0006-06-C-7322.

References

1. Cobb, B.R., Shenoy, P.: On Transforming Belief Function Models to Probability Models. Working Paper No. 23, University of Kansas School of Business (2004)
2. Hummel, R., Landy, M.: Statistical view on the theory of evidence. IEEE Trans. Pattern Anal. Mach. Intell. 10(2), 235–247 (1988)
3. Jackson, M.: A Sheaf-Theoretic Approach to Measure Theory. PhD. Thesis. University of Pittsburgh (2006)
4. Mac Lane, S.: Categories for the Working Mathematician, 2nd edn. Springer, New York (1998)
5. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
6. Voorbraak, F.: A computationally efficient approximation of Dempster-Shafer theory. Int. J. Man-Mach. Stud. 30, 525–536 (1989)
7. Wendt, M.: The Category of Disintegration. Cahiers Topologie Géom 35(4), 291–308 (1994)

Variance in Generalized Probability Representations

Lower and Upper Covariance

Erik Quaeghebeur

SYSTeMS Research Group, Ghent University, Zwijnaarde, Belgium

Abstract. We give a definition for lower and upper covariance in Walley's theory of imprecise probabilities (or coherent lower previsions) that is direct, i.e., does not refer to credal sets. It generalizes Walley's definition for lower and upper variance. Just like Walley's definition of lower and upper variance, our definition for lower and upper covariance is compatible with the credal set approach; i.e., we also provide a covariance envelope theorem. Our approach mirrors the one taken by Walley: we first reformulate the calculation of a covariance as an optimization problem and then generalize this optimization problem to lower and upper previsions. We also briefly discuss the still unclear meaning of lower and upper (co)variances and mention some ideas about generalizations to other central moments.

Keywords: Variance, Covariance, Central moment, Theory of imprecise probabilities, Envelope theorem.

1 Introduction

In the statistical and probabilistic literature, the variance and covariance of a random variable are important quantities. The generalization of these and other concepts to Walley's [5] theory of imprecise probabilities – which encompasses, in some sense, classical probability theory, Dempster–Shafer theory, and possibility theory – is usually done by taking lower and upper envelopes (see, e.g., [3, 4]). Walley ([5], §G) gives a more direct, equivalent definition of lower and upper variance. In this paper, we similarly give a direct definition of the lower and upper covariance concepts.

In the rest of this introduction, we go over some necessary concepts from the theory of classical and imprecise probabilities, respectively. Then follows Section 2, treating Walley's definition of lower and upper variance. This forms a basis for Section 3, which culminates in our definition for lower and upper covariance. The concluding Section 4 briefly touches on further generalizations to other central moments and the meaning of the concepts we have been discussing.

Throughout the paper, we give a running example, recognizable by its small print.

Concepts from classical probability theory. In the classical setting, a subject is uncertain about the value of some random variable, that, e.g., represents the outcome of an experiment. We here only consider real-valued, bounded random variables, which – because they can be seen as uncertain gains or losses – we call gambles.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 323–330, 2008.
springerlink.com

Even though he may be uncertain, the subject can express what incomplete knowledge he has by giving an estimate for a number of gambles. In classical probability theory, these estimates are expected values, or fair prices for the gambles ([2]). These expected values are collected in the subject's expectation functional or prevision P, which maps gambles to fair prices (real numbers). We assume that the subject has specified prices for all gambles defined on some possibility space.

If the subject wants his prices to be reasonable and consistent, such a prevision must satisfy some rationality criteria called coherence conditions ([5], §2.8). Let f and g be any two gambles and let λ be a real number, then these criteria are:[1]

$$\text{Boundedness:} \quad \inf f \leq Pf \leq \sup f, \tag{1}$$

$$\text{Linearity:} \quad P(\lambda \cdot f + g) = \lambda \cdot Pf + Pg. \tag{2}$$

Assume that P satisfies them. Call $\mathscr{P}$ the set of previsions satisfying these criteria.[2]

Let us introduce our running example: Take the unit interval $[0,1]$ as the possibility space and let U be the prevision corresponding to the uniform distribution on this possibility space. We shall be using two gambles, the identity gamble id and $1-\mathrm{id}$. We then find $U\mathrm{id} = U(1-\mathrm{id}) = \frac{1}{2}$.

Variance and covariance. When focusing on the uncertainty governing the outcome of some gambles, there are – apart from their expected value – also other so-called statistics that summarize what we know about the gambles. One class of these are the central moments corresponding to a given prevision P: let $\mathscr{H}$ be a finite multiset of gambles,[3] then the corresponding central moment $M_P\mathscr{H}$ is defined by

$$M_P\mathscr{H} := P\big(\textstyle\prod_{h\in\mathscr{H}}(h - Ph)\big). \tag{3}$$

In this paper, we are mainly interested in the case where $\mathscr{H}$ contains two gambles, so to the so-called second order central moments. The first of these is the variance of f, defined by taking $\mathscr{H} = \{f, f\}$ above: (even in f)

$$V_P f := P(f - Pf)^2 = Pf^2 - (Pf)^2. \tag{4}$$

The second, the covariance of f and g, is defined by taking $\mathscr{H} = \{f, g\}$: (odd in f and g, even in (f, g))

$$C_P\{f,g\} := P\big((f - Pf)\cdot(g - Pg)\big) = P(f\cdot g) - Pf\cdot Pg. \tag{5}$$

For both Equation (4) and (5) the last expression follows from the boundedness (1) and linearity (2) of P. They show that both $V_P f$ and $C_P\{f,g\}$ are continuous functions of P, a fact that will be of use later on.

[1] Unless noted otherwise, all introduced symbols retain their meaning in the rest of the paper. For clarity, we do not use parentheses for function application, but only for grouping.

[2] We choose the weak* topology (i.e., the topology of pointwise convergence) on the set $\mathscr{P}$ ([5], §D3). This allows us to later talk about the compactness of its subsets (e.g., credal sets) and about the continuity of functions defined on it (i.e., that have previsions as an argument).

[3] A multiset can contain the same object multiple times, but – like a set – is unordered.

For our running example, we find that $V_U\,\mathrm{id} = V_U(1-\mathrm{id}) = \frac{1}{12}$ and $C_U\{\mathrm{id}, 1-\mathrm{id}\} = -\frac{1}{12}$.

Concepts from the theory of imprecise probabilities. The theory of imprecise probabilities and other generalizations of classical probability theory were all conceived to model aspects of uncertainty that cannot be captured classically. The major difference is that now the subject's prevision – which specifies unique fair (buying *and* selling) prices for gambles – is replaced by a lower prevision $\underline{P}$ and an upper prevision $\overline{P}$, which respectively specify acceptable supremum buying prices and acceptable infimum selling prices (both real numbers) that need not coincide.

In the context of this paper, where lower and upper previsions are defined for all gambles, it is sufficient to give either one of them, because they are related by conjugacy: $\overline{P}f = -\underline{P}(-f)$. We shall here take a lower prevision to be the fundamental concept and an upper prevision the derived one.

The conditions (1) and (2) now get replaced by a weaker set of coherence conditions ([5], §2.3), which we assume $\underline{P}$ and $\overline{P}$ to satisfy: (take λ positive)

$$\text{Boundedness:} \quad \inf f \leq \underline{P}f \leq \overline{P}f \leq \sup f, \tag{6}$$

(Mixed) super and sublinearity:

$$\lambda \cdot \underline{P}f + \underline{P}g \leq \underline{P}(\lambda \cdot f + g) \leq \lambda \cdot \underline{P}f + \overline{P}g \leq \overline{P}(\lambda \cdot f + g) \leq \lambda \cdot \overline{P}f + \overline{P}g. \tag{7}$$

For our example, the subject's lower (upper) prevision is a so-called linear-vacuous mixture: $\underline{U} = \delta \cdot U + (1-\delta) \cdot \inf$, $\overline{U} = \delta \cdot U + (1-\delta) \cdot \sup$, with $\delta \in [0,1]$. Then $\underline{U}\,\mathrm{id} = \underline{U}(1-\mathrm{id}) = \frac{\delta}{2}$ and $\overline{U}\,\mathrm{id} = \overline{U}(1-\mathrm{id}) = 1 - \frac{\delta}{2}$. Further on, we also encounter the two gambles $\frac{1}{2} \cdot \big(\mathrm{id} + (1-\mathrm{id})\big) = \frac{1}{2}$ and $\frac{1}{2} \cdot \big(\mathrm{id} - (1-\mathrm{id})\big) = \mathrm{id} - \frac{1}{2}$; for these, we find $\underline{U}\frac{1}{2} = \overline{U}\frac{1}{2} = \frac{1}{2}$, $\underline{U}(\mathrm{id} - \frac{1}{2}) = -\frac{1-\delta}{2}$, and $\overline{U}(\mathrm{id} - \frac{1}{2}) = \frac{1-\delta}{2}$. Taking δ to be 0 or 1 gives us the vacuous[4] or uniform case, respectively.

The information encoded in the lower prevision $\underline{P}$ can also be represented by the compact convex set $\mathscr{M}\underline{P}$ of previsions that dominate it ([5], §3.3). This so-called credal set is defined by $\mathscr{M}\underline{P} := \{P \in \mathscr{P} \mid P \geq \underline{P}\}$, where the inequality is a universally quantified pointwise one. The lower or upper prevision for any gamble can be calculated as a lower or upper envelope of the previsions for this gamble present in $\mathscr{M}\underline{P}$:

$$\underline{P}f = \min_{P \in \mathscr{M}\underline{P}} Pf \quad \text{and} \quad \overline{P}f = \max_{P \in \mathscr{M}\underline{P}} Pf. \tag{8}$$

The minimum and maximum are always attained in an extreme point of $\mathscr{M}\underline{P}$.

2 Lower and Upper Variance

In this section, we state the results found in Appendix G of Walley's book [5].

Variance as an optimization problem. What we need to do first, is take another look at the definition (4) of the variance of a gamble. We rewrite and reformulate it as follows: For all real numbers μ, it holds that $V_P f = P\big((f-\mu) + (\mu - Pf)\big)^2 = P(f-\mu)^2 - (Pf-\mu)^2$, so $V_P f + (Pf-\mu)^2 = P(f-\mu)^2$. The second left-hand side term is the expression for a parabola with a minimum 0 for μ in Pf; the right-hand side is therefore also an expression

[4] With a vacuous lower (or upper) prevision, the subject expresses a total lack of knowledge.

for a parabola, with the variance as a minimum, also attained in Pf. So an alternative definition for the variance is

$$V_P f = \min_{\mu\in\mathbb{R}} P(f-\mu)^2. \tag{9}$$

Lower and upper variance. Walley ([5], §G1) takes this last definition (9) as inspiration to define the lower and upper variance of a gamble: (even in f)

$$\underline{V}_{\underline{P}} f := \min_{\mu\in\mathbb{R}} \underline{P}(f-\mu)^2 \quad \text{and} \quad \overline{V}_{\underline{P}} f := \min_{\mu\in\mathbb{R}} \overline{P}(f-\mu)^2. \tag{10}$$

In this definition, minima and not infima are used, even though $\mathbb{R}$, the set of reals, is open. Let us justify this. First, let $\overline{P}$ stand for both $\underline{P}$ and $\overline{P}$, so we can do a parallel derivation for both cases. Furthermore, take $\varepsilon = \mu - \underline{P}f$, then

$$\begin{aligned}\overline{P}(f-\mu)^2 = \overline{P}(f-\underline{P}f-\varepsilon)^2 &\geq \overline{P}(f-\underline{P}f)^2 + \varepsilon^2 + \underline{P}\big(-2\cdot\varepsilon\cdot(f-\underline{P}f)\big)\\ &\geq \overline{P}(f-\underline{P}f)^2 + \varepsilon^2 \quad \text{if } \varepsilon \leq 0\\ &> \overline{P}(f-\underline{P}f)^2 \quad \text{if } \varepsilon < 0, \text{ i.e., if } \mu < \underline{P}f. \end{aligned} \tag{11}$$

The first inequality follows from (mixed) superadditivity (7). An entirely similar derivation, now with $\varepsilon = \overline{P}f - \mu$, gives us

$$\overline{P}(f-\mu)^2 > \overline{P}(f-\overline{P}f)^2 \quad \text{if } \varepsilon < 0, \text{ i.e., if } \mu > \overline{P}f. \tag{12}$$

Together with the fact that the interval $[\underline{P}f, \overline{P}f]$ is compact and that $\overline{P}(f-\mu)^2$ is a continuous[5] function of μ, inequalities (11) and (12) show that, for both the case of the lower and upper variance, a minimum is attained in a μ that belongs to $[\underline{P}f, \overline{P}f]$.

For our running example, we can use this optimization domain restriction to write $\underline{V}_{\underline{U}}\mathrm{id} = \min\{\underline{U}(\mathrm{id}-\mu)^2 \mid \mu \in [\frac{\delta}{2}, 1-\frac{\delta}{2}]\}$ and $\overline{V}_{\underline{U}}\mathrm{id} = \min\{\overline{U}(\mathrm{id}-\mu)^2 \mid \mu \in [\frac{\delta}{2}, 1-\frac{\delta}{2}]\}$. Working this out, we find that for both the minimum is attained for $\mu = \frac{1}{2}$; this gives $\underline{V}_{\underline{U}}\mathrm{id} = \frac{\delta}{12}$ and $\overline{V}_{\underline{U}}\mathrm{id} = \frac{1}{4} - \frac{\delta}{6}$. Similarly, $\underline{V}_{\underline{U}}\frac{1}{2} = \overline{V}_{\underline{U}}\frac{1}{2} = 0$, $\underline{V}_{\underline{U}}(\mathrm{id}-\frac{1}{2}) = \frac{\delta}{12}$, and $\overline{V}_{\underline{U}}(\mathrm{id}-\frac{1}{2}) = \frac{1}{4} - \frac{\delta}{6}$.

The variance envelope theorem. Walley ([5], §G2) also proves the variance envelope theorem; it allows calculation of lower and upper variances via the credal set:

$$\underline{V}_{\underline{P}} f = \min_{P\in\mathscr{M}\underline{P}} V_P f \quad \text{and} \quad \overline{V}_{\underline{P}} f = \max_{P\in\mathscr{M}\underline{P}} V_P f. \tag{13}$$

In Walley's proof of this theorem, the minimax theorem is used, whose application requires the continuity of $V_P f$ as a function of P, which we mentioned below (5). Note that the maximum above is not necessarily attained in an extreme point of $\mathscr{M}\underline{P}$ ([5], §G3), in contrast to the situation for upper previsions (8).

3 Lower and Upper Covariance

We are now ready to attack this paper's central topic: to find a *direct* definition for the lower and upper covariance of a pair of gambles, i.e., one that does not involve credal

[5] A proof consists of the ε–δ-technique together with (7).

sets. The approach we are going to take is analogous to the one Walley [5] has taken for the lower and upper variance of a gamble: (i) We first define the covariance of a gamble as an optimization problem and (ii) Then use this to find expressions for the lower and upper covariance of a pair of gambles. However, for reasons that will become clear at that point, we are going to posit a covariance envelope theorem that mirrors the variance envelope theorem (13) right in between these two steps.

Covariance as an optimization problem. As we did in to obtain (9), we first reformulate the definition (5) of covariance as an optimization problem: For all real μ and ν, it holds that $C_P\{f,g\} = P\big((f-\mu+\mu-Pf)\cdot(g-\nu+\nu-Pg)\big) = P\big((f-\mu)\cdot(g-\nu)\big) - (Pf-\mu)\cdot(Pg-\nu)$, so then $C_P\{f,g\} + (Pf-\mu)\cdot(Pg-\nu) = P\big((f-\mu)\cdot(g-\nu)\big)$. The second left-hand side term is the expression in (μ,ν) for a hyperbolic paraboloid (or saddle surface); the same therefore again holds for the right-hand side. Its saddle point $(\underline{P}f, \underline{P}g)$ can be reached using a minimax (or maximin) operator. This is clearer after a substitution and some rewriting: let α and β be real numbers such that $\mu = \alpha+\beta$ and $\nu = \alpha-\beta$, then it holds for all real α and β that $P\big((\frac{f+g}{2}-\alpha)^2 - (\frac{f-g}{2}-\beta)^2\big) = C_P\{f,g\} + (P\frac{f+g}{2}-\alpha)^2 - (P\frac{f-g}{2}-\beta)^2$, which gives rise to the following defining expression:

$$C_P\{f,g\} = V_P\tfrac{f+g}{2} - V_P\tfrac{f-g}{2} = \mathrm{opt}_{\alpha,\beta\in\mathbb{R}} P\big((\tfrac{f+g}{2}-\alpha)^2 - (\tfrac{f-g}{2}-\beta)^2\big), \qquad (14)$$

where $\mathrm{opt}_{\alpha,\beta\in\mathbb{R}}$ can be either $\min_{\alpha\in\mathbb{R}}\max_{\beta\in\mathbb{R}}$ or $\max_{\beta\in\mathbb{R}}\min_{\alpha\in\mathbb{R}}$. Because P is linear and its argument is the sum of terms in *either* α *or* β, it is the same whether we use a maximin or minimax operator.

Proposing a definition for lower and upper covariance would ideally have consisted of just replacing the linear prevision P with the lower prevision $\underline{P}$ and its conjugate upper prevision $\overline{P}$. However, the fact that we have two operators to choose from leaves us with a dilemma: as neither $\underline{P}$ or $\overline{P}$ is linear, does it matter which operator to use for the definition of lower and upper covariance? Perhaps working with credal sets can shed some light on this issue and clarify which of the two choices should be taken (if they do not turn out to be equivalent). This is the next topic.

However, there is one thing we can already say: Independently of the operator, using the same reasoning that led to (11) and (12), it follows that the minimizing α belongs to $[\underline{P}\frac{f+g}{2}, \overline{P}\frac{f+g}{2}]$ and – invoking conjugacy – that the maximizing β belongs to $[\underline{P}\frac{f-g}{2}, \overline{P}\frac{f-g}{2}]$.

For our running example, where we let $f = \mathrm{id}$ and $g = 1-\mathrm{id}$, these intervals respectively become the singleton $\{\frac{1}{2}\}$ and $[-\frac{1-\delta}{2}, \frac{1-\delta}{2}]$.

The covariance envelope theorem. Due to the confusing double options we have – with expression (14) – for generalizing the definition of covariance to imprecise probabilities, we need something that can help us choose between them (or show they are both good enough). This something is the covariance envelope theorem, which now rather functions as a definition, and not a theorem to be proven:

$$\underline{C}_{\underline{P}}\{f,g\} := \min_{P\in\mathscr{M}_{\underline{P}}} C_P\{f,g\} \quad \text{and} \quad \overline{C}_{\underline{P}}\{f,g\} := \max_{P\in\mathscr{M}_{\underline{P}}} C_P\{f,g\}. \qquad (15)$$

It states that the lower and upper covariance corresponding to $\underline{P}$ can be seen as lower and upper envelopes over the credal set $\mathscr{M}\underline{P}$ of classical covariances. Due to the compactness of $\mathscr{M}\underline{P}$ and the continuity of the covariance as a function of P, the minimum and maximum are attained, so this theorem is sensible as an indirect definition. Note that the minimum and maximum above are not necessarily attained in an extreme point of $\mathscr{M}\underline{P}$, in contrast to the situation for lower and upper previsions ([8]). This theorem is an immediate analog of the variance envelope theorem ([13]) and it expresses a desirable if not conceptually necessary property.

Looking back at the two equivalent definitions of covariance given by ([14]), it becomes clear we must investigate whether the maximin and minimax operator encountered there can be interchanged with the maximum or minimum over P encountered in ([15]). Let us write this out more explicitly. First define the convex functions u and v by $u\alpha = (\frac{f+g}{2} - \alpha)^2$ and $v\beta = (\frac{f-g}{2} - \beta)^2$ then the question is: Which, if any, of the following statements can we ascertain to be true:

$$\begin{aligned} \min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} &\stackrel{?}{=} \operatorname{opt}_{\alpha,\beta\in\mathbb{R}} \min_{P\in\mathscr{M}\underline{P}} P(u\alpha - v\beta), \\ \max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} &\stackrel{?}{=} \operatorname{opt}_{\alpha,\beta\in\mathbb{R}} \max_{P\in\mathscr{M}\underline{P}} P(u\alpha - v\beta), \end{aligned} \tag{16}$$

where, as before, $\operatorname{opt}_{\alpha,\beta\in\mathbb{R}}$ can be either $\min_{\alpha\in\mathbb{R}}\max_{\beta\in\mathbb{R}}$ or $\max_{\beta\in\mathbb{R}}\min_{\alpha\in\mathbb{R}}$.

First of all, note that consecutive minimum operators or consecutive maximum operators can always be interchanged.

Whether the interchange of a minimum and a maximum operator is allowed, can only be checked after a more thorough study of the functions involved: As function application is a linear operation, $P(u\alpha - v\beta)$ is linear in P (and thus both convex and concave); as P is linear, $P(u\alpha - v\beta)$ is convex in α and concave in β. Furthermore, $P(u\alpha - v\beta)$ is continuous in α, β, and P. Together with the fact that the maximum or minimum is always attained in some convex compact set, this is enough to do a first operator interchange, i.e., we can apply the minimax theorem ([5], §E6) if needed. We find the following modified statements:

$$\min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} \begin{cases} \stackrel{?}{=} \min_{\alpha\in\mathbb{R}} \min_{P\in\mathscr{M}\underline{P}} \max_{\beta\in\mathbb{R}} P(u\alpha - v\beta), \\ \stackrel{?}{=} \max_{\beta\in\mathbb{R}} \min_{P\in\mathscr{M}\underline{P}} \min_{\alpha\in\mathbb{R}} P(u\alpha - v\beta), \end{cases}$$

$$\max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} \begin{cases} \stackrel{?}{=} \min_{\alpha\in\mathbb{R}} \max_{P\in\mathscr{M}\underline{P}} \max_{\beta\in\mathbb{R}} P(u\alpha - v\beta), \\ \stackrel{?}{=} \max_{\beta\in\mathbb{R}} \max_{P\in\mathscr{M}\underline{P}} \min_{\alpha\in\mathbb{R}} P(u\alpha - v\beta). \end{cases}$$

As maximizing is a convex operation and minimizing is a concave operation, $\max_{\beta\in\mathbb{R}} P(u\alpha - v\beta)$ is convex as a function of P and α and $\min_{\alpha\in\mathbb{R}} P(u\alpha - v\beta)$ is concave as a function of P and β. This means that a second application of the maximin theorem is not possible, and a second interchange is not generally possible for all cases. We can therefore only be sure about two of the four initial statements ([16]):

$$\begin{aligned} \min_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} &= \min_{\alpha\in\mathbb{R}} \max_{\beta\in\mathbb{R}} \underline{P}(u\alpha - v\beta), \\ \max_{P\in\mathscr{M}\underline{P}} C_P\{f,g\} &= \max_{\beta\in\mathbb{R}} \min_{\alpha\in\mathbb{R}} \overline{P}(u\alpha - v\beta). \end{aligned}$$

Thus, the covariance envelope theorem implies a direct definition of lower and upper covariance, the starting point of what follows just below.

Lower and upper covariance. By combining the covariance envelope theorem (15) with the last expressions we encountered, we find definitions for the lower and upper covariance of a pair of gambles that use a lower or upper prevision, but not the corresponding credal set: (odd in f and g, even in (f,g))

$$\begin{aligned}\underline{C}_{\underline{P}}\{f,g\} &= \min_{\alpha\in\mathbb{R}}\max_{\beta\in\mathbb{R}}\underline{P}\big((\tfrac{f+g}{2}-\alpha)^2-(\tfrac{f-g}{2}-\beta)^2\big),\\ \overline{C}_{\underline{P}}\{f,g\} &= \max_{\beta\in\mathbb{R}}\min_{\alpha\in\mathbb{R}}\overline{P}\big((\tfrac{f+g}{2}-\alpha)^2-(\tfrac{f-g}{2}-\beta)^2\big).\end{aligned} \tag{17}$$

For our running example, we know we must take $\alpha=\frac{1}{2}$, so then

$$\begin{aligned}\underline{C}_{\underline{U}}\{\mathrm{id},1-\mathrm{id}\} &= \max_{\beta\in\mathbb{R}}\underline{U}\big(-(\mathrm{id}-\tfrac{1}{2}-\beta)^2\big) = -\min_{\mu\in\mathbb{R}}\overline{U}(\mathrm{id}-\mu)^2 = -\overline{V}_{\underline{U}}\mathrm{id} = -\tfrac{1}{4}+\tfrac{\delta}{6},\\ \overline{C}_{\underline{U}}\{\mathrm{id},1-\mathrm{id}\} &= \max_{\beta\in\mathbb{R}}\overline{U}\big(-(\mathrm{id}-\tfrac{1}{2}-\beta)^2\big) = -\min_{\mu\in\mathbb{R}}\underline{U}(\mathrm{id}-\mu)^2 = -\underline{V}_{\underline{U}}\mathrm{id} = -\tfrac{\delta}{12}.\end{aligned}$$

An interesting property of classical covariance is that it can be written as a difference of two variances (see equation (14)). For our generalized definition, this identity becomes a string of inequalities:

$$\begin{aligned}\underline{V}_{\underline{P}}(\tfrac{f+g}{2})-\overline{V}_{\underline{P}}(\tfrac{f-g}{2}) &\le \underline{C}_{\underline{P}}\{f,g\}\\ &\le \min\big\{\underline{V}_{\underline{P}}(\tfrac{f+g}{2})-\underline{V}_{\underline{P}}(\tfrac{f-g}{2}),\overline{V}_{\underline{P}}(\tfrac{f+g}{2})-\overline{V}_{\underline{P}}(\tfrac{f-g}{2})\big\}\\ &\le \max\big\{\underline{V}_{\underline{P}}(\tfrac{f+g}{2})-\underline{V}_{\underline{P}}(\tfrac{f-g}{2}),\overline{V}_{\underline{P}}(\tfrac{f+g}{2})-\overline{V}_{\underline{P}}(\tfrac{f-g}{2})\big\}\\ &\le \overline{C}_{\underline{P}}\{f,g\} \le \overline{V}_{\underline{P}}(\tfrac{f+g}{2})-\underline{V}_{\underline{P}}(\tfrac{f-g}{2}).\end{aligned} \tag{18}$$

These inequalities are obtained starting from the definitions of lower and upper covariance (17). They are related to lower and upper variance (10) by using (mixed) super and sublinearity (7).

For our running example, all but the third of the inequalities in (18) become equalities.

4 Conclusions

Musing on other lower and upper central moments. We started with the definition (3) of a central moment and then restricted ourselves to the second order ones. An obvious (still open) question would now be: Can we generalize the ideas of this paper to higher order central moments?

Independently of whether it is possible or not to give definitions for arbitrary lower and upper central moments using a lower or upper prevision only, it seems desirable that these definitions satisfy a central moment envelope theorem:

$$\underline{M}_{\underline{P}}\mathscr{H} := \min_{P\in\mathscr{M}_{\underline{P}}} M_P\mathscr{H} \quad \text{and} \quad \overline{M}_{\underline{P}}\mathscr{H} := \max_{P\in\mathscr{M}_{\underline{P}}} M_P\mathscr{H}. \tag{19}$$

As before, we could try to write the definition (3) of a central moment as an optimization problem by replacing each $h-Ph$ by $(h-\mu_h)+(\mu_h-Ph)$ (where $h\in\mathscr{H}$ and the

μ_h are real numbers) and separating the term $\underline{P}\big(\prod_{h\in\mathscr{H}}(h-\mu_h)\big)$, as we did to obtain (9) and (14). This could lead to central moments defined as optimization problems; the difficulty with this is that we are going to be dealing with much more complex multilinear expressions (in the μ_h) than parabola or saddle surfaces.

But what does it mean? Another thing that is still an open question – to me personally – is: What is the meaning of a lower and upper variance and covariance?

An intuitive interpretation is the one typically given to their precise counterparts:

(i) variance is a statistic describing how much a gamble is believed to vary,
(ii) covariance is a statistic describing how and how much a pair of gambles is believed to vary together.

I have found no satisfactory behavioral interpretation; they could be seen as prices – as we do for previsions –, but trying to say for what leads to all too convoluted explanations, I think. Perhaps variance and covariance should just be seen as useful for the description of probability density or mass functions, and any 'generalization' as mathematically interesting at most.

On the other hand, the fact that no appealing, or intuitively simple behavioral interpretation is known to me, does not mean that it could not be found, e.g., in the economic literature or in other non-classical theories for uncertainty and indeterminacy. Couso et al. [1], for example, give a nice overview of definitions for the variance of a *fuzzy* random variable and their interpretation.

Acknowledgement. I wish to thank Gert de Cooman, Filip Hermans, and two anonymous referees for their useful comments and suggestions. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author.

References

1. Couso, I., Dubois, D., Montes, S., Sánchez, L.: On various definitions of the variance of a fuzzy random variable. In: De Cooman, G., Vejnarová, J., Zaffalon, M. (eds.) Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2007, Prague, Czech Republic), pp. 135–144. Action M Agency, SIPTA, Prague (2007)
2. De Finetti, B.: Theory of Probability. Wiley, Chichester (1974–1975)
3. Hall, J.W.: Uncertainty-based sensitivity indices for imprecise probability distributions. Reliab. Eng. Syst. Saf. 91(10-11), 1443–1451 (2006)
4. Kreinovich, V., Xiang, G., Ferson, S.: Computing mean and variance under Dempster-Shafer uncertainty: Towards faster algorithms. Internat. J. Approx. Reason 42(3), 212–227 (2006)
5. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Some Properties of the d_K-Variance for Interval-Valued Random Sets

Ángela Blanco[1], Norberto Corral[1], Gil González-Rodríguez[2], and M. Asunción Lubiano[1]

[1] Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Oviedo, Spain
[2] European Centre for Soft Computing, Mieres, Spain

Abstract. The suitability of the family of d_K-distances for intervals to quantify and estimate the variability of a random interval with respect to its Aumann expectation is discussed. To be precise, we will review some properties of the metrics and the associated variance. Then, we will show that the sample variance is a consistent estimator of the population one, which make it adequate for inferential problems. The use of the d_K-variance will be illustrated by means of a real-life example.

Keywords: Random interval, d_K-Metric, Variance, Estimation.

1 Introduction

The random intervals are useful to model many random experiments in which the characteristic observed on each outcome cannot be described by means of a single real number, but rather by a range or interval. For instance, whenever fluctuations of economical or medical magnitudes over different days are observed, or whenever physical measures recorded by a machine with a known maximum error are obtained.

The measurement of the variability about a central point is one of the most important issues in Statistics. If the expected value is considered and the aim is to quantify the dispersion (or, in other words, the error of predicting the values of a variable X by the expected value $E(X)$), it is common to consider the expected value of the square-distances to the mean (i.e., $\mathrm{Var}_d(X) = E(d^2(X, E(X)))$). The utility of this concept regarding, for instance, the least-squares criterion is strongly connected with the Fréchet property, which holds whenever

$$E(X) = \arg\min_U E(d^2(X,U))$$

and, in this case, $\mathrm{Var}_d(X) = \min_U E(d^2(X,U))$. However, the fulfilment of this property depends on the tandem metric/expected value being considered.

When the usual interval arithmetic (based on the Minkowski addition and the product by a scalar in $\mathbb{R}$) is considered, the coherent expected value in terms of the SLLN is the Aumann's one (see [1]). Specifically, let $\mathscr{K}_c(\mathbb{R})$ be the space of nonempty closed and bounded intervals of $\mathbb{R}$, $(\Omega, \mathscr{A}, P)$ be a probability space, and $\mathscr{X} : \Omega \to \mathscr{K}_c(\mathbb{R})$ be an interval-valued random set (i.e., a Borel-measurable mapping w.r.t. the topology

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 331–337, 2008.
springerlink.com

generated by d_H). The particularization of the Aumann expectation can be expressed as the compact interval $E^A(X) = [E(\inf X), E(\sup X)]$, provided that this interval exists.

On the other hand, one of the best known distances in $\mathscr{K}_c(\mathbb{R})$ is the particularization of the Hausdorff metric, that is,

$$d_H(A,B) = \max\{|\inf A - \inf B|, |\sup A - \sup B|\}.$$

One drawback of this metric from a statistical point of view is that it is not easy-to-use in applying the least squares criterion when the usual definition of mean square error and the Aumann expected value are considered, since it is well-known that the Fréchet property is not satisfied (see [8]).

Example 1. ([8]). Let X be an interval-valued random set so that

$$\mathrm{mid}X = \begin{cases} 2 \text{ , with probability } & p_1 = 2/3 \\ 3 \text{ , with probability } & p_2 = 1/3 \end{cases}$$

and

$$\mathrm{spr}X = \begin{cases} 1 \text{ , with probability } & q_1 = 2/3 \\ 2 \text{ , with probability } & q_2 = 1/3 \end{cases}$$

where 'mid' stands for the centre of the interval and 'spread' stands for the radius.

Thus, it is possible to check that $E^A(X) = [1, 11/3]$, whereas the real interval which minimizes $E\left(d_H^2(X,U)\right)$ corresponds to $U = [1, 3.4]$.

To avoid this inconvenience, the particularization of the ρ_2 distance, which is defined in terms of the support functions, can be considered (see [3]). Let $s_A : \mathbb{R} \to \mathbb{R}$ be the support function of the interval A, which is defined so that $s_A(u) = \sup_{a \in A} \langle a, u \rangle$ for every $u \in \mathbb{R}$, $\langle \cdot, \cdot \rangle$ being the usual inner product on $\mathbb{R}$. The support function restricted to the unit sphere $\mathbb{S}^0 = \{-1, 1\}$ characterizes the intervals; specifically, $s_A(1) = \sup A$ and $s_A(-1) = -\inf A$. If μ denotes the Lebesgue measure on the unit sphere, the ρ_2 distance is expressed as:

$$\rho_2(A,B) = \int_{\mathbb{S}^0} (s_A(u) - s_B(u))^2 d\mu(u) = \frac{1}{2}(\sup A - \sup B)^2 + \frac{1}{2}(\inf A - \inf B)^2$$

This metric has interesting properties in connection with the statistical analyses. However, from an intuitive point of view, it presents some inconveniences, as also happens with the Hausdorff metric.

Example 2. ([2]) It is straightforward to check that the Hausdorff metric assigns the same distance to the two pairs of intervals $A_1 = [0,5]$, $B_1 = [6,7]$ and $A_2 = [0,5]$, $B_2 = [6,10]$. However, it seems more intuitive for a distance measure to assign a greater value to the second pair.

In an analogous way, it is possible to show that the ρ_2 distance between the intervals $C_1 = [-2,2]$ and $D_1 = [-1,1]$ is the same that the one between $C_2 = [-2,1]$ and $D_2 = [-1,2]$, whereas it seems more natural for the second one to be greater.

Bertoluzza et al.'s metric ([2]) is an L^2-type metric that has a better intuitive behaviour in situations like that in Example 2. It is defined in terms of a non-degenerate probability measure W on the measurable space $([0,1],\beta_{[0,1]})$ as follows:

$$d_W(A,B)=\left(\int_{[0,1]}(f_A(\lambda)-f_B(\lambda))^2 dW(\lambda)\right)^{1/2},$$

where $f_A(\lambda)=\lambda\sup A+(1-\lambda)\inf A$, for all $\lambda\in\mathbb{R}$. The distance d_W can be alternatively expressed in terms of a discrete measure $\overrightarrow{\varphi}^W=(\varphi_1^W,\varphi_2^W,\varphi_3^W)$ as

$$d_W(A,B)=\left(\varphi_1^W(\sup A-\sup B)^2+\varphi_2^W(\inf A-\inf B)^2+\varphi_3^W(\mathrm{mid}\,A-\mathrm{mid}\,B)^2\right)^{1/2}. \tag{1}$$

In addition, if $\int_{[0,1]} t dW=1/2$, d_W can be expressed in an intuitive way in terms of the centre and the spread of the intervals as

$$d_W(A,B)=\left((\mathrm{mid}\,A-\mathrm{mid}\,B)^2+\sigma_W^2(\mathrm{spr}\,A-\mathrm{spr}\,B)^2\right)^{1/2}, \tag{2}$$

where $\sigma_W^2=\int_{[0,1]}(2\lambda-1)^2 dW(\lambda)$. Thus, d_W takes into account the 'location' of the intervals, through the distance between the mid-points, and the 'degree of imprecision', through the weighted distance between the spreads. Note that since $\sigma_W^2\leq 1$, then the relative importance of the spreads is always lower than or equal to that of the centres (see [9]).

Bertoluzza et al.'s metric and ρ_2 can be generalized by considering a wide family of L^2-distances w.r.t. certain kernels (see [5]). To be precise, the d_K-distance is defined as

$$d_K(A,B)=\int_{\mathbb{S}^0}(s_A(u)-s_B(u))(s_A(v)-s_B(v))dK(u,v)$$

where $K:\mathbb{S}^0\times\mathbb{S}^0\to\mathbb{R}$ is a symmetric and definite positive kernel. The d_K distance can be also expressed in a way similar to (1) as follows:

$$\begin{aligned} d_K(A,B)=&\big([K(1,1)+K(1,-1)](\sup A-\sup B)^2\\ &+[K(-1,-1)+K(1,-1)](\inf A-\inf B)^2-4K(1,-1)(\mathrm{mid}\,A-\mathrm{mid}\,B)^2\big)^{1/2} \end{aligned} \tag{3}$$

It is easy to show that d_W can be written as a distance of the family d_K by considering the kernel

$$K(u,v)=\begin{cases}\int_0^1\lambda^2 dW(\lambda) & \text{if } u=v=1\\ \int_0^1(1-\lambda)^2 dW(\lambda) & \text{if } u=v=-1\\ -\int_0^1\lambda(1-\lambda)dW(\lambda) & \text{if } u=-v\end{cases} \tag{4}$$

However, there are metrics in the d_K family that cannot be expressed as a d_W metric. Thus,

Example 3. Consider

$$K(u,v)=\begin{cases}1/4 & \text{if } u=v=1\\ 1/4 & \text{if } u=v=-1\\ 1/8 & \text{if } u=-v\end{cases}$$

By comparing (1) and (3) it is easy to show that there is no discrete measure ϕ^W that would allow us write this d_K distance as a d_W distance.

One of the advantages of the metrics in the d_K-family is that they can be expressed in terms of an inner product in the space of continuous functions defined on $\mathbb{R}$ by considering the restriction of the support function to the unit sphere (see [5]), that is,

$$d_K(A,B) = (\langle s_A - s_B, s_A - s_B \rangle_K)^{1/2},$$

where $\langle \cdot, \cdot \rangle_K$ is the inner product associated with the L^2-distance in the space of the continuous functions w.r.t. the kernel K.

2 The d_K-Variance for Random Intervals

The d_K-variance is a particularization to the interval case of the one considered in [5] (see also, [4], [6] and [7]). Its expression corresponds to that of the above-introduced variance w.r.t. a general metric d, that is, the d_K-variance of an interval-valued random set X is given by

$$\mathrm{Var}_{d_K}(X) = E(d_K^2(X, E^A(X)),$$

whenever this exists.

This variance fulfils most of the suitable properties of a classical variance. However, in order to obtain that $\mathrm{Var}_{d_K}(X) = \mathrm{Var}_{d_K}(-X)$ (where the usual product of an interval by a scalar is considered) it is necessary to assume that the kernel satisfies that $K(1,1) = K(-1,-1)$; otherwise this suitable property fails.

Example 4. Consider a probabilistic space $(\Omega, \mathscr{A}, P)$ with $\Omega = \{\omega_1, \omega_2\}$, $\mathscr{A} = \mathscr{P}(\Omega)$ and $P(\omega_i) = 1/2$ for $i = 1,2$, and define the interval-valued random set X associated with $(\Omega, \mathscr{A}, P)$ as $X(\omega_1) = [0,2]$ and $X(\omega_2) = [0,4]$. Then, it is easy to show that $\mathrm{Var}_{d_K}(X) = K(1,1)$ and $\mathrm{Var}_{d_K}(-X) = K(-1,-1)$.

For this reason, from now we will choose a kernel $K : \{-1,1\} \times \{-1,1\} \to \mathbb{R}$ so that the associated matrix is doubly symmetric and positive definite.

3 Estimation of the d_K-Variance

Let $X : \Omega \to \mathscr{K}_c(\mathbb{R})$ be an interval-valued random set with finite variance $\sigma_X^2 = \mathrm{Var}_{d_K}(X)$ and $\{X_i\}_{i=1}^n$ a random sample obtained from X (i.e., a set of independent random elements distributed as X). The variance σ_X^2 can be estimated, as usual, by its analogue sample version, that is,

$$\widehat{\sigma_n^2(X)} = \frac{1}{n}\left[d_K^2(X_1, \overline{X_n}) + \ldots + d_K^2(X_n, \overline{X_n})\right]$$

where the sample mean $\overline{X_n} = (1/n)(X_1 + \ldots + X_n)$ is defined in terms of the Minkowski sum and the product of an interval by a scalar.

The linearity of the inner product and the properties of the support function allow us to express $\widehat{\sigma_n^2(X)}$ as

$$\widehat{\sigma_n^2(X)} = \frac{1}{n}\sum_{i=1}^{n}\langle s_{X_i}, s_{X_i}\rangle_K - \langle s_{\overline{X_n}}, s_{\overline{X_n}}\rangle_K.$$

From this expression, and by applying the SLLN for real- and Hilbert-valued random elements, it is possible to verify that $\widehat{\sigma_n^2(X)}$ converges almost surely to

$$E\left(\langle s_X, s_X\rangle_K\right) - \langle s_{E^A(X)}, s_{E^A(X)}\rangle_K = E\left(\langle s_X - s_{E^A(X)}, s_X - s_{E^A(X)}\rangle_K\right)$$

$$= E\left(d_K^2(X, E^A(X))\right) = \sigma_X^2,$$

i.e., the sample d_K-variance is a strongly consistent estimator of the population d_K-variance, which supports its suitability.

4 Simulation Studies

In order to illustrate empirically the consistency result for the sample d_K-variance, some simulations have been carried out.

In this section we have chosen as d_K-distance that of Bertoluzza et al.'s corresponding to the Lebesgue measure, that is, K is that in (4) with $W =$ Lebesgue measure on $[0,1]$. Let X be an interval-valued random set with finite variance σ_X^2 and let $\{X_i\}_{i=1}^n$ be the simple random sample obtained to analyze the estimator $\widehat{\sigma_n^2(X)}$. Different distributions for $\mathrm{mid}X$ and $\mathrm{spr}X$, as well as different sample sizes n have been considered.

Table 1. Empirical estimate of d_K-variance

Theoretical situation	n	$E(\widehat{\sigma_n^2(X)})$	MSE
$\mathrm{spr}X \sim \chi_1^2$	10	1.4988	0.6980
$\mathrm{mid}X \sim N(0,1)$	30	1.6072	0.2606
independent	50	1.6313	0.1582
	100	1.6487	0.0800
$\sigma_X^2 = 5/3$	500	1.6635	0.0167
$\mathrm{spr}X \sim \chi_1^2$ $Y \sim N(0,1)$	10	3.2565	9.2879
independent	30	3.5713	3.6265
$\mathrm{mid}X = \mathrm{spr}X + Y$	50	3.5936	2.0418
	100	3.6209	1.0829
$\sigma_X^2 = 11/3$	500	3.6612	0.2137

The distribution of the sample d_K-variance $\widehat{\sigma_n^2(X)}$ has been approximated by Monte Carlo method on the basis of 10,000 iterations, and the corresponding mean value and mean square error(MSE) have been computed.

In Table 1 the obtained results are gathered. We can appreciate that in both theoretical situations (one in which mids and spreads are dependent and another one in which they are independent) the considered estimator $\widehat{\sigma_n^2(X)}$ for the variance of X seems to be asymptotically unbiased, and the mean square error is closer to 0 as n increases.

5 Example

The consistency of the sample d_K-variance and the suitable theoretical properties that it fulfils for the kernels considered here make it suited for inferences concerning the variability about the Aumann expected value. In this section a point estimation will be illustrated, although further statistical analyses, as confidence intervals or hypothesis testing can be developed.

In Table 2 a sample data set corresponding to the systolic blood pressure ranges over a day of 59 patients of Hospital Valle del Nalón of Asturias (Spain) is shown. Since only the minimum and maximum value over a day were recorded, it is suitable to represent the sample values as realizations of an interval-valued random set X. The aim is to estimate the degree of variability of the systolic blood pressure range of the patients.

Table 2. Systolic blood pressure ranges during a day

11.8-17.3	11.9-21.2	9.8-16.0	10.4-16.1	12.2-17.8	9.7-15.4	13.1-18.6	12.7-18.9
8.7-15.0	10.5-15.7	11.3-21.3	14.1-25.6	12.0-17.9	14.1-20.5	10.8-14.7	10.1-19.4
9.9-16.9	11.5-19.6	10.9-17.4	12.6-19.7	9.9-17.2	12.8-21.0	9.9-20.1	11.3-17.6
9.4-14.5	8.8-22.1	11.4-18.6	14.8-20.1	11.3-18.3	14.5-21.0	11.1-19.2	9.4-17.6
12.0-18.0	11.6-20.1	10.2-15.6	10.0-16.1	10.2-16.7	10.3-15.9	15.9-21.4	10.4-16.1
10.2-18.5	13.8-22.1	10.6-16.7	11.1-19.9	8.7-15.2	11.2-16.2	13.0-18.0	12.0-18.8
13.6-20.1	10.3-16.1	9.5-16.6	9.0-17.7	12.5-19.2	9.2-17.3	11.6-16.8	9.7-18.2
8.3-14.0	9.8-15.7	12.7-22.6					

The sample mean of X is $[11.1881, 18.1678]$, and the sample variance is $\widehat{\sigma_n^2(X)} = 3.582$, so $\widehat{\sigma_n(X)} = 1.8929$. Thus, we can conclude that the estimated expected value for the systolic blood pressure range is $[11.1881, 18.1678]$, with an approximated dispersion of 1.8929 about this value.

Acknowledgement. The research in this paper has been partially supported by the Spanish Ministry of Education and Science Grant MTM2006-07501, and the PhD Grant BP06-010 of the Govern of Asturias, though the Innovation, Technology and Science Plan (PCTI) 2006-2009. This financial support is gratefully acknowledged.

References

1. Arstein, Z., Vitale, R.A.: A strong law of large numbers for random compact sets. Ann. Probab. 5, 879–882 (1975)
2. Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. Mathware Soft. Comput. 2, 71–84 (1995)
3. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc., River Edge (1994)
4. Körner, R.: On the variance of fuzzy random variables. Fuzzy Sets Syst. 92, 83–93 (1997)
5. Körner, R., Näther, W.: On the variance of random fuzzy variables. In: Bertoluzza, C., Gil, M., Ralescu, D. (eds.) Statistical Modeling, Analysis and Management of Fuzzy Data, pp. 22–39. Physica-Verlag, Heidelberg (2002)
6. Lubiano, A.: Generalized measure of the absolute variation of a random set. Definition and statistical applications. In: Proceedings of 11th European Young Statisticians Meeting (EYSM 1999, Marly-le-Roi, Francia), pp. 122–127 (1997)
7. Lubiano, M.A., Gil, M.A., López-Díaz, M., López, M.T.: The $\overrightarrow{\lambda}$-mean squared dispersion associated with a fuzzy random variable. Fuzzy Sets Syst. 111(3), 307–317 (2000)
8. Näther, W.: Linear statistical inference for random fuzzy data. Statistics 29, 221–240 (1997)
9. Trutschnig, W., González-Rodríguez, G., Colubi, A., Gil, M.A.: A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread (submitted for publication) (2008)

Aggregation Functions

Triangular Conorms on the Space of Non-decreasing Lists of Non-negative Real Numbers

Isabel Aguiló, Javier Martín, Gaspar Mayor, and Jaume Suñer

Dept. de Matemàtiques i Inf., Universitat de les Illes Balears, Palma de Mallorca, Spain

Abstract. Motivated by the study of $\triangle_n^+$-valued distances, where $\triangle_n^+$ is the set of distance distribution functions with range in $\{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$, we deal with triangular conorms defined on the bounded lattice Σ_n of non-decreasing lists $(a_1, \ldots, a_n) \in [0, +\infty]^n$ equipped with the natural (product) order. Using triangular conorms on $[0, +\infty]$ and triangular norms on $\{0, 1, \ldots, n\}$ we describe different classes of appropriate triangular conorms on $[0, +\infty]^n$.

Keywords: Lattice, Generalized distance, Probabilistic metric space, t-Norm, t-Conorm, Discrete t-norm, Triangle inequality.

1 Introduction

The history of triangular norms started with the paper "Statistical Metrics" (K. Menger, 1942). The main idea of Karl Menger was to introduce the notion of a space in which distances between points are determined by probability distribution functions rather than by real numbers. To extend the triangle inequality to this setting, he employed a function T from $[0,1]^2$ to $[0,1]$, which he called a triangular norm (t-norm). The original set of axioms for t-norms was considerably weaker, including among others also the functions which are known today as triangular conorms. In 1961, in the paper "Associative functions and statistical triangle inequality", Abe Sklar and Berthold Schweizer provided the axioms of t-norms as they are used today and, in 1962, Sherstnev gave a redefinition of statistical metric space by defining the concept of triangle function. Many results concerning t-norms and triangle functions were obtained in the course of this development, most of which are summarized in [6]. An excellent book on t-norms and applications is [3]. Recently, relevant publications on t-norms are [5] and [1]. Motivated by the fact that in most practical situations it is sufficient to reduce the range $[0,1]$ of probabilities to some finite subset of this interval, we deal with distance distribution functions with range in $\{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$ and we pay attention to the problem of finding appropriate triangular functions to be used in the triangle inequality condition.

In this paper we call triangular norm any binary operation defined on a bounded partially ordered set satisfying the Sklar and Schweizer axioms.

2 Preliminaries

Let $(L; \leq)$ be a bounded lattice. We denote by 0 and 1 the minimum and maximum elements respectively $(0 < 1)$, and $x \wedge y = \inf\{x, y\}$, $x \vee y = \sup\{x, y\}$ for all x, y in L.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 341–348, 2008.
springerlink.com

Definition 1. *A triangular norm (briefly t-norm) on L is a binary operation $T : L \times L \to L$ such that for all $x, y, z \in L$ the following axioms are satisfied:*

(1) $T(x,y) = T(y,x)$ *(commutativity)*
(2) $T(T(x,y),z) = T(x,T(y,z))$ *(associativity)*
(3) $T(x,y) \leq T(x',y')$ *whenever* $x \leq x'$, $y \leq y'$ *(monotonicity)*
(4) $T(x,1) = x$ *(boundary condition)*

A triangular conorm (t-conorm for short) is a binary operation $S : L \times L \to L$ which, for all $x, y, z \in L$, satisfies (1)-(3) and $S(x,0) = x$.

Example 1. Basic examples of t-norms and t-conorms on a lattice L are the well-known drastic T_D and S_D and the infimum $T_\wedge$ and the supremum $S_\vee$ given by: $T_\wedge(x,y) = x \wedge y$ and $S_\vee(x,y) = x \vee y$.

A simple but important particular case is $L = \{0,1,2,\ldots,n\}$ equipped with the usual ordering. In addition to T_D and T_M (now $T_M(x,y) = \min(x,y) = T_\wedge(x,y)$), another basic t-norm on L can be considered: $T_{\text{Ł}}(x,y) = \max(0, x+y-n)$, that we call the Łukasiewicz t-norm. Also we can consider the bounded sum or Łukasiewicz t-conorm $S_{\text{Ł}}(x,y) = \min(x+y, n)$. Note that (T_D, S_D), (T_M, S_M) and $(T_{\text{Ł}}, S_{\text{Ł}})$ are dual pairs with respect to the unique strong negation on L: $N(x) = n - x$.

Let L_1 and L_2 be bounded lattices. We denote by $\mathscr{T}_i$, $\mathscr{S}_i$, $i = 1,2$, the family of t-norms and t-conorms on L_i respectively. Let $\varphi : L_1 \to L_2$ be an isomorphism (φ is a bijection such that φ and φ^{-1} are order-preserving) then we can associate to each t-norm (t-conorm) T_1 (S_1) on L_1 a t-norm (t-conorm) T_2 (S_2) on L_2 defined by $T_2(x,y) = \varphi(T_1(\varphi^{-1}(x), \varphi^{-1}(y)))$ ($S_2(x,y) = \varphi(S_1(\varphi^{-1}(x), \varphi^{-1}(y)))$) for all $x, y \in L_2$. Thus we have a correspondence one-to-one from $\mathscr{T}_1$ to $\mathscr{T}_2$ (from $\mathscr{S}_1$ to $\mathscr{S}_2$). In this sense we can say that the sets $\mathscr{T}_1$ and $\mathscr{T}_2$ ($\mathscr{S}_1$ and $\mathscr{S}_2$) are equal (up to the isomorphism φ).

Proofs and more details on t-norms defined on bounded posets can be found in [2], [4] and [5].

In 1967, E.Trillas ([7, 8]) introduced the notion of generalized metric space, when considering abstract metrics valued in ordered semigroups, unifying with this approach the real metric structures of M. Fréchet and the probabilistics metric spaces of K. Menger. We adopt here as ordered semigroup a bounded lattice $(L; \leq)$ with a t-conorm S defined on it.

3 $\triangle_n^+$-Valued S-Distances

Definition 2 (Generalized Metric Space). *Given a bounded lattice $(L; \leq)$ and a t-conorm S on L and a non-empty set X, we say that a function $d : X \times X \to L$ is an L-valued S-distance on X if the following conditions hold for all $p, q, r \in X$:*

(i) $d(p,q) = 0 \Leftrightarrow p = q$, *where* 0 *is the minimum of* L.
(ii) $d(p,q) = d(q,p)$
(iii) *S-triangle inequality:*

$$d(p,r) \leq S(d(p,q), d(q,r)) \tag{1}$$

In this case we say that X is an (L,S,d)-metric space.

Example 2 (Ordinary Metric Spaces). The ordinary metric spaces are the $([0,+\infty],S,d)$-metric spaces where $[0,+\infty]$ is equipped with the usual order and $S(a,b)=a+b$.

Example 3 (Probabilistic Metric Spaces)
Let $\triangle^+$ be the set of functions $F:[0,+\infty]\to[0,1]$ which are non-decreasing and left-continuous on $(0,+\infty)$ with $F(0)=0$ and $F(+\infty)=1$.

A function $F\in\triangle^+$ is usually called a distance distribution function. Basic elements of $\triangle^+$ are the so called Dirac distributions:

$$\varepsilon_a(x)=\begin{cases}0 & \text{if } x\le a\\ 1 & \text{if } x>a\end{cases} \text{ where } a<+\infty\,;\quad \varepsilon_{+\infty}(x)=\begin{cases}0 & \text{if } x<+\infty\\ 1 & \text{if } x=+\infty\end{cases}$$

The set $\triangle^+$ is a complete lattice with respect to the usual pointwise ordering: $F\le G$ if and only if $F(x)\le G(x)\ \forall x\in[0,+\infty]$. The functions $\varepsilon_{+\infty}$ and ε_0 are the minimum and maximum of $\triangle^+$, respectively.

Consider $(\triangle^+;\le^{op})$ the set $\triangle^+$ equipped with the opposite order of the usual one ($F\le^{op}G\Leftrightarrow G\le F$) and S a t-conorm on $(\triangle^+;\le^{op})$. Then, a non-empty set X is a $(\triangle^+,S,d)$-metric space if and only if $d:X\times X\to\triangle^+$ satisfies

- $d(p,q)=\varepsilon_0\Leftrightarrow p=q$
- $d(p,q)=d(q,p)$
- $d(p,r)\le^{op}S(d(p,q),d(q,r))$

It is worth to observe that the triangle inequality can be rewritten in the form $d(p,r)\ge T(d(p,q),d(q,r))$, where T is a t-norm on $(\triangle^+;\le^{op})$. Thus we recover the probabilistic metric spaces ([6]) as examples of $(\triangle^+,S,d)$-metric spaces.

Example 4 (Discrete Probabilistic Metric Spaces)
Given a positive integer n, we denote by $\triangle_n^+$ the set $\{F\in\triangle^+;\mathrm{Ran}F\subset\{0,\frac{1}{n},\ldots,\frac{n-1}{n},1\}\}$. It is clear that $\triangle_n^+$ is a sublattice of $\triangle^+$ containing ε_a for all $a\in[0,+\infty]$. In this paper we mainly deal with $\triangle_n^+$-valued S-distances.

Note that a natural way to construct t-conorms on $\triangle_n^+$ is from a t-conorm on $\{0,1,2,\ldots,n\}$: given a t-conorm S on $\{0,1,2,\ldots,n\}$ we define $\sigma(F,G)(x)=\frac{1}{n}S(nF(x),nG(x))$ for all F, G in $\triangle_n^+$ and all x in $[0,+\infty]$. Then it is easily verified that σ is a t-conorm on $(\triangle^+;\le^{op})$ for any t-conorm S on $\{0,1,2,\ldots,n\}$.

On the other hand, if S is a t-conorm on $(\triangle^+;\le^{op})$ and $F,G\in\triangle_n^+$ then in general $S(F,G)$ is not an element of $\triangle_n^+$, thus the restriction of S to $\triangle_n^+$ is not a t-conorm on $\triangle_n^+$.

Next proposition proves that the lattice $(\triangle_n^+;\le^{op})$ is isomorphic to $(\Sigma_n;\le)$ where Σ_n is the set of n-ordered lists $(a_1,\ldots,a_n)\in[0,+\infty]^n$ satisfying $a_1\le\ldots\le a_n$ equipped with the usual product ordering.

Proposition 1. *The function $\varphi:\triangle_n^+\to\Sigma_n$ defined by $\varphi(F)=(a_1,\ldots,a_n)$ where $a_i=\max\{x;F(x)<\frac{i}{n}\}$ is an isomorphism. The inverse of φ is given by $\varphi^{-1}(a_1,\ldots,a_n)=F$ where $F(x)=\frac{1}{n}|\{i;a_i<x\}|$ for all x in $[0,+\infty]$.*

Remark 1. Observe that this application φ inverts the usual order, in the sense that, given $F, G \in \Delta_n^+$, $F \leq G \Longleftrightarrow \varphi(F) \geq \varphi(G)$.

Corollary 1

(i) The family of t-conorms on $(\triangle_n^+; \leq^{op})$ *"is" the family of t-conorms on* $(\Sigma_n; \leq)$.

(ii) A $\triangle_n^+$*-metric space "is" a* Σ_n*-metric space, and reciprocally.*

(iii)The ordinary metric spaces are just the $(\triangle_1^+ = \Sigma_1 = [0, +\infty], S, d)$*-metric spaces where* $S(a,b) = a + b$.

Example 5. Consider the set $\mathbb{Z} \times \mathbb{Z}$ of ordered pairs of integers equipped with the distance between (a,b), (c,d) measured along axes at right angles: $|a-c| + |b-d|$ (Manhattan distance). Let us suppose that each pair (a,b) moves at random (with the same probability) to any one of the four surrounding points $(a \pm 1, b \pm 1)$, and consider the set $X = \{\{(a \pm 1, b \pm 1)\}; (a,b) \in \mathbb{Z} \times \mathbb{Z}\}$. Now we define $d : X \times X \to \triangle_{16}^+$ as follows: $d(p,q) = F$, where $F(x) =$ the probability of "the distance between p and q is less than x".

If we denote $p = \{(a \pm 1, b \pm 1)\}$, $q = \{(c \pm 1, d \pm 1)\}$, $h = |a-c|$, $v = |b-d|$ and $\delta = h + v$ then using the above representation of distance distribution functions by means of 16-ordered lists, we have that $d(p,q)$ is given by

$$\begin{cases} (d-2, d-2, d-2, d-2, d, d, d, d, d, d, d, d, d+2, d+2, d+2, d+2) \\ \qquad \text{if } v, h \geq 2 \\ (d-2, d-2, d-2, d, d, d, d, d, d, d, d, d, d+2, d+2, d+2, d+2) \\ \qquad \text{if } v = 1 \text{ and } h \geq 2 \text{ or vice versa} \\ (d-2, d, d, d, d, d, d, d, d, d+2, d+2, d+2, d+2, d+2, d+2, d+2) \\ \qquad \text{if } v = 0 \text{ and } h \geq 2 \text{ or vice versa} \\ (0,0,2,2,2,2,2,2,2,2,2,2,4,4,4,4) \qquad \text{if } v = h = 1 \\ (1,1,1,1,1,1,1,1,3,3,3,3,3,3,3,3) \qquad \text{if } v = 0 \text{ and } h = 1 \text{ or vice versa} \end{cases}$$

In order to get an $\triangle_{16}^+$-valued distance on X, we have to introduce appropriate t-conorms S on Σ_{16} such that the triangle inequality holds: $d(p,r) \leq S(d(p,q), d(q,r))$. We deal with this matter in the next section.

4 Triangular Conorms on Σ_n

According to (i) in Corollary 1, to study t-conorms on $(\triangle_n^+; \leq^{op})$ is equivalent to study t-conorms on the very simple structure $(\Sigma_n; \leq)$. We analyze in this section different methods to construct t-conorms on this ordered set of ordered lists.

Proposition 2. *The t-conorms on* $(\Sigma_n; \leq)$ *are the restrictions to* Σ_n *of the t-conorms on* $([0, +\infty]^n; \leq)$ *such that* Σ_n *is closed under them.*

Proof. Let S be a t-conorm on $[0, +\infty]^n$ such that $S(a,b) \in \Sigma_n$ whenever $a, b \in \Sigma_n$. It is obvious that the restriction of S to Σ_n is a t-conorm on Σ_n. Reciprocally, let us consider a t-conorm S' on Σ_n and define for all $a, b \in [0, +\infty]^n$:

$$S(a,b) = \begin{cases} a & \text{if } b = 0 \\ b & \text{if } a = 0 \\ S'(\sigma(a),\sigma(b)) & \text{otherwise} \end{cases}$$

where $\sigma(a)$ is the non-decreasing reordering of the n-list a. We can easily prove that S is a t-conorm on $[0,+\infty]^n$ and Σ_n is closed under S. □

Example 6. These are some t-conorms on $[0,+\infty]^n$ such that Σ_n is closed under them:

(i) S_D and $S_\vee$
(ii) $S((a_1,\ldots,a_n),(b_1,\ldots,b_n)) = (a_1+b_1,\ldots,a_n+b_n)$
(iii) $S((a_1,\ldots,a_n),(b_1,\ldots,b_n)) = (a_1+b_1,a_2+b_2+a_1b_1\ldots,a_n+b_n+a_1b_1)$

Note that $S((a_1,a_2),(b_1,b_2)) = (a_1+b_1,\max(a_2,b_2))$ is a t-conorm on $[0,+\infty]^2$ but Σ_2 is not closed under it: $S((2,3),(3,4)) = (5,4) \notin \Sigma_2$.

4.1 Direct Products from t-Conorms on $[0,+\infty]$

Proposition 3. *Let S be a t-conorm on $[0,+\infty]^n$ which is the direct product of $S_1,\ldots,S_n$, where $S_1,\ldots,S_n$ are t-conorms on $[0,+\infty]$. Then, Σ_n is closed under S if and only if $S_1 \leq \ldots \leq S_n$.*

In particular, $S((a_1,\ldots,a_n),(b_1,\ldots,b_n)) = (S_1(a_1,b_1),\ldots,S_1(a_n,b_n))$ where S_1 is a t-conorm on $[0,+\infty]$ is a t-conorm on $[0,+\infty]^n$ such that Σ_n is closed under it. In (i) and (ii) of Example 6 we show t-conorms of this type, whereas the t-conorm given in (iii) is not a direct product. Finally, $S((a_1,a_2),(b_1,b_2)) = (a_1+b_1,\max(a_2,b_2))$ is the direct product of $S_1(a_1,b_1) = a_1+b_1$ and $S_2(a_2,b_2) = \max(a_2,b_2)$ with $S_2 \leq S_1$.

Remark 2. Note that direct products are not appropriate t-conorms to be used in Example 5. Thus, consider the points p,q,r in X defined by (a,b), $(a+2,b)$, $(a+4,b)$, and suppose that $S = (S_1,\ldots,S_{16})$ is the direct product of the t-conorms $S_1,\ldots,S_{16}$. Let us show that the S-triangle inequality $d(p,r) \leq S(d(p,q),d(q,r))$ fails: $d(p,r) = (2,4,4,4,4,4,4,4,4,6,6,6,6,6,6,6,6)$, $d(p,q) = d(q,r) = (0,2,2,2,2,2,2,2,2,4,4,4,4,4,4,4,4)$, thus $d(p,r)$ is not less than or equal to $(S_1(0,0),S_2(2,2,),\ldots,S_{16}(4,4))$ due to $S_1(0,0) = 0$ for any t-conorm S_1.

Let us introduce now a family of t-conorms on $[0,+\infty]^n$ that could be appropriate in particular in Example 5.

Proposition 4. *Given $1 \leq k,r \leq n$, consider $S_{(k,r)}(a,b) = (a_1+b_1,a_{r-1}+b_{r-1}\ldots,a_r+b_r+a_kb_k,\ldots,a_n+b_n+a_kb_k)$ where $a = (a_1,\ldots,a_n)$, $b = (b_1,\ldots,b_n)$). Then, $S_{(k,r)}$ is a t-conorm on $[0,+\infty]^n$ such that Σ_n is closed under it.*

Remark 3

- Note that the t-conorm given in (iii) of Example 6 is just $S_{(1,n)}$.
- The t-conorms $S_{(k,r)}$ with $k \neq r$ are not direct products of t-conorms on $[0,+\infty]$.
- The t-conorms $S_{(k,1)}$ with $k \geq 5$ can be used in Example 5.

4.2 Sherstnev's Construction

By observing that the left side of the Menger's inequality $F_{pr}(x+y) \geq T(F_{pq}(x), F_{qr}(y))$ only depends on the sum $x+y$, then the inequality holds for all $x, y \geq 0$ if and only if $F_{pr}(z) \geq \sup_{x+y=z} T(F_{pq}(x), F_{qr}(y))$. From this fact Sherstnev showed that, for any left-continuous t-norm T, the function τ_T defined via $\tau_T(F,G)(z) = \sup_{x+y=z} T(F(x), G(y))$ for all $z \geq 0$ is a t-norm on $(\triangle^+; \geq)$. In this subsection we adapt this construction to our discrete setting and we define $\sigma_T(F,G)(z) = \sup_{x+y=z} \frac{1}{n} T(nF(x), nG(y))$, where T is a t-norm on $L_n = \{0, 1, \ldots, n\}$, for all $z \geq 0$ and $F, G \in \triangle_n^+$. In the next proposition we show that σ_T is a t-conorm on $(\triangle_n^+; \leq^{op})$ by translating first the expression of σ_T in terms of lists in Σ_n. It is worth to observe that, in contrast to the continuous case, we obtain a t-conorm from any t-norm T, with no restrictions on it.

Proposition 5. *Given a t-norm T on L_n, let us define a function $\sigma_T : \Delta_n^+ \times \Delta_n^+ \longrightarrow \Delta_n^+$ by*

$$\sigma_T(F,G)(x) = \sup_{y+z=x} \tfrac{1}{n} T(nF(y), nG(z)) \ \ \forall\, x \in [0, +\infty].$$

Then σ_T is a t-conorm on Δ_n^+.

Proof. According to the Corollary of Proposition 1, we will prove that the corresponding function $\hat{\sigma}_T(a,b) = \varphi(\sigma_T(\varphi^{-1}(a), \varphi^{-1}(b)))$ is a t-conorm on Σ_n. For this reason, let us obtain the expression of $\hat{\sigma}_T$ as a function of the lists. Let us consider $a, b, c \in \Sigma_n$ and suppose that $\hat{\sigma}_T(a,b) = c$ and let $F = \varphi^{-1}(a)$ and $G = \varphi^{-1}(b)$. Thus, for all $i \in \{1, \ldots, n\}$, $c_i = \max\{x : \sigma_T(F,G)(x) < \frac{i}{n}\} = \max\{x : \sup_{y+z=x} T(F(y), G(z)) < \frac{i}{n}\}$, and this means that for all y, z such that $y + z = x$, $T(F(y), G(z)) < \frac{i}{n}$, that is, $T\left(\frac{|\{j : a_j < y\}|}{n}, \frac{|\{j : b_j < z\}|}{n}\right) < \frac{i}{n}$. Finally, we obtain $c_i = \min\left\{a_j + b_k : T\left(\frac{j}{n}, \frac{k}{n}\right) \geq \frac{i}{n}\right\}$.

From this expression, it is a straightforward calculation to prove that $\hat{\sigma}_T$ is a t-conorm on Σ_n. □

Example 7. The related t-conorms on $\triangle_n^+$ to the basic t-norms minimum, drastic and Łukasiewicz are the following:

- If $T = \min$, then $\hat{\sigma}_{\min}((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = (a_1 + b_1, \ldots, a_n + b_n)$.
- If $T = T_{\text{Ł}}$, then $\hat{\sigma}_{\text{Ł}}((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = (c_1, \ldots, c_n)$, where $c_i = \min\{a_j + b_{n+i-j}; j = i, i+1, \ldots, n\}$.
- For the drastic t-norm T_D, $\hat{\sigma}_D((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = (c_1, \ldots, c_n)$, with $c_i = \min\{a_i + b_n, a_n + b_i\}$.

Example 8. There are six t-norms on $\{0,1,2,3\}$. They are listed in Table 1 and the corresponding t-conorms on Σ_3 are listed in Table 2.

Example 9. Let us consider $n = 4$ and two lists $a = (2,3,5,5)$ and $b = (1,4,6,7)$. Let T be the t-norm given in Table 3.

Then $c_1 = \min\left\{a_j + b_k : T\left(\frac{j}{n}, \frac{k}{n}\right) \geq \frac{1}{n}\right\}$ and, from the table of T we observe that $T(\frac{j}{n}, \frac{k}{n}) \geq \frac{1}{n}$ for $j \geq 3, k \geq 1$, for $j, k \geq 2$ and for $j \geq 1, k \geq 3$. The minimum sum $a_j + b_k$ for all of these pairs (j,k) is taken by $j = 3, k = 1$ and it is equal to $a_3 + b_1 = 6$. Thus $c_1 = 6$. Similar calculations allow to obtain the other components of the list c and $c = \hat{\sigma}(a,b) = (6,9,11,12)$.

Table 1. The six t-norms on $\{0,1,2,3\}$

T_1	T_2	T_3	T_4	T_5	T_6
0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
0 0 1 2	0 0 2 2	0 0 0 2	0 1 1 2	0 1 2 2	0 1 2 2
0 0 0 1	0 0 0 1	0 0 0 1	0 1 1 1	0 0 1 1	0 1 1 1
0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0

Table 2. The six induced t-conorms on Σ_3

σ_1 $(\min\{a_1+b_3, a_2+b_2, a_3+b_1\}, \min\{a_2+b_3, a_3+b_2\}, a_3+b_3)$
σ_2 $(\min\{a_1+b_3, a_2+b_2, a_3+b_1\}, a_2+b_2, a_3+b_3)$
σ_3 $(\min\{a_1+b_3, a_3+b_1\}, \min\{a_2+b_3, a_3+b_2\}, a_3+b_3)$
σ_4 $(a_1+b_1, \min\{a_2+b_3, a_3+b_2\}, a_3+b_3)$
σ_5 $(\min\{a_1+b_2, a_2+b_1\}, a_2+b_2, a_3+b_3)$
σ_6 $(a_1+b_1, a_2+b_2, a_3+b_3)$

Table 3. The t-norm T

4	0	1	2	3	4
3	0	1	2	3	3
2	0	0	1	2	2
1	0	0	0	1	1
0	0	0	0	0	0
	0	1	2	3	4

Acknowledgement. The authors acknowledge the support of the Govern Balear grant PCTIB2005GC1-07 and the Spanish DGI grant MTM2006-08322.

References

1. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific Publishing Co., Singapore (2006)
2. De Baets, B., Mesiar, R.: Triangular norms on product lattices. Fuzzy Sets Syst. 104, 61–75 (1999)
3. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Trends in Logic, vol. 8. Kluwer Academic Publishers, Boston (2000)
4. Mayor, G., Torrens, J.: On a class of operators for expert systems. Int. J. Intelligent Systems 8, 771–778 (1993)

5. Mayor, G., Torrens, J.: Triangular norms on discrete settings. In: Klement, E.P., Mesiar, R. (eds.) Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms, pp. 189–230. Elsevier, Dordrecht (2005)
6. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. North Holland, Amsterdam (1983)
7. Trillas, E.: Sobre Distancias Estadísticas. Tesis doctoral. Pub. Univ. Barcelona (1972)
8. Trillas, E., Alsina, C.: Introducción a los Espacios Métricos Generalizados. Serie Universitaria, 49. Fundación Juan March, Madrid (1978)

On Patchwork Techniques for 2-Increasing Aggregation Functions and Copulas

Fabrizio Durante, Susanne Saminger-Platz, and Peter Sarkoci

Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, Linz, Austria

Abstract. In recent years, there has been a raise of interest in the determination of copulas with given values at some fixed points, or with given horizontal, vertical, affine, diagonal, or sub-diagonal sections and combinations thereof. Closely related to these investigations are the determination and characterization of increasing and 2-increasing functions with given margins whose domain is a subset of the unit square as well as necessary and sufficient conditions providing that the combination (patchwork) of such functions on sub-domains yields a (new) 2-increasing aggregation function on $[0,1]^2$, in particular a copula. In the present contribution we provide a full characterization of increasing, 2-increasing functions with prescribed margins acting on a sub-rectangle of the unit square. The characterization allows to determine easily the greatest and smallest such functions and to look at the results on copulas with given horizontal and/or vertical sections and its boundaries from a more general and unified viewpoint. We further discuss necessary and sufficient conditions for a patchwork based on triangular sub-domains.

1 Introduction

In this contribution we focus on 2-increasing binary aggregation functions [9, 12], the most prominent and most studied examples thereof being copulas [22]. Various methods for constructing copulas are already known (see [22] for a comprehensive overview and further references). However, particularly in recent years, there has been a raise of interest in the determination of copulas with given values at some fixed points, or with given horizontal, vertical, affine, diagonal, or sub-diagonal sections and combinations thereof [4, 6, 7, 8, 11, 14, 15, 16, 17, 24, 25]. In some cases also (partial) results for the boundaries of these classes have been established [7, 9, 16, 24, 25]. It is interesting to see that, for the different cases, the results have been obtained more or less independently from each other, although many of these constructions refer to the determination of a 2-increasing and increasing function on some sub-domain of the unit square with predescribed margins (not necessarily coinciding with the margins of the unit square). Two different problems are directly related to this viewpoint and can be distinguished: first, the determination and characterization of increasing and 2-increasing functions on sub-domains of the unit square with given margins. Second, necessary and sufficient conditions providing that the combination (patchwork) of such functions on sub-domains yields a (new) 2-increasing aggregation function, in particular a copula.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 349–356, 2008.
springerlink.com

In Section 3 we discuss the second question for a patchwork on triangular subdomains. We provide a full characterization of increasing and 2-increasing functions acting on a sub-rectangle of the unit square with prescribed margins in Section 4. The characterization allows to determine immediately the greatest and smallest such functions and allows, e.g., to look at the results on copulas with given horizontal and/or vertical sections and its boundaries from a more general and unified viewpoint. We illustrate this by some selected examples in Section 5.

2 Preliminaries

Definition 1. *A binary function* $A\colon [0,1]^2 \to [0,1]$ *which is increasing in each place and fulfills* $A(0,0) = 0$, $A(1,1) = 1$ *is called a* (binary) aggregation function. *If, in addition, A fulfills, for all* $x_1, x_2 \in [0,1]$ *and for all* $y_1, y_2 \in [0,1]$ *with* $x_1 \leq x_2$ *and* $y_1 \leq y_2$,

$$V_A([x_1,x_2] \times [y_1,y_2]) := A(x_1,y_1) + A(x_2,y_2) - A(x_1,y_2) - A(x_2,y_1) \geq 0 \qquad (1)$$

then A is 2-increasing. *The class of all (2-increasing) aggregation functions will be denoted by* $\mathscr{A}$ *(*$\mathscr{A}_2$*).*

Note that the value $V_A([x_1,x_2] \times [y_1,y_2])$ is often referred to as the *A-volume* of the rectangle $[x_1,x_2] \times [y_1,y_2]$.

Definition 2. *A* 2*-increasing aggregation function* $C\colon [0,1]^2 \to [0,1]$ *is called a* copula *if it has neutral element* 1*, i.e.* $C(x,1) = C(1,x) = x$ *for every* $x \in [0,1]$. *We will denote the set of all copulas by* $\mathscr{C}$.

Among the most prominent copulas let us mention the monotone dependence copula M and its counterpart, the countermonotone dependence copula W given by $M(x,y) = \min(x,y)$ and $W(x,y) = \max(x+y-1,0)$. For every copula C it holds that $W \leq C \leq M$.

Definition 3. *Given* $A \in \mathscr{A}$*, the* diagonal *of A is the mapping* $\delta_A\colon [0,1] \to [0,1]$ *given by* $\delta_A(t) = A(t,t)$. *The* opposite diagonal *of A is, instead, the mapping* $\eta\colon [0,1] \to [0,1]$ *defined by* $\eta(t) = A(t,1-t)$.

3 The Diagonal Patchwork

Let Δ_+, Δ_-, Γ_+, and Γ_- be the following subsets of the unit square:

$$\Delta_+ := \{(x,y) \in [0,1]^2 : x \geq y\}; \qquad \Gamma_+ := \{(x,y) \in [0,1]^2 : x+y \leq 1\};$$
$$\Delta_- := \{(x,y) \in [0,1]^2 : x < y\}; \qquad \Gamma_- := \{(x,y) \in [0,1]^2 : x+y > 1\}.$$

For every subset S of $[0,1]^2$, $\mathbf{1}_S$ denotes the *indicator function* of S, i.e., $\mathbf{1}_S(x,y)$ is equal to 1, if $(x,y) \in S$, and 0, otherwise.

Given the aggregation functions A and B, we introduce the functions $F_{A,B}, F^{A,B} : [0,1]^2 \to [0,1]$ given, for all x,y in $[0,1]$, by

$$F_{A,B}(x,y) := A(x,y)\,\mathbf{1}_{\Delta_+}(x,y) + B(x,y)\,\mathbf{1}_{\Delta_-}(x,y);$$
$$F^{A,B}(x,y) := A(x,y)\,\mathbf{1}_{\Gamma_+}(x,y) + B(x,y)\,\mathbf{1}_{\Gamma_-}(x,y).$$

Note that the diagonal patchwork restricted to copulas has also been discussed in [10, 11] and, under the name *diagonal splice*, in [24]. It is easy to show (see also [10]) that if A and B are aggregation functions with the same (opposite) diagonal section, then $F_{A,B}$ ($F^{A,B}$) is an aggregation function. Moreover, if A and B have neutral element 1, so has $F_{A,B}$. If B has neutral element 1, then so has $F^{A,B}$. Further the following sufficient resp. necessary conditions hold:

Proposition 1. *Let A,B be in $\mathscr{A}_2$ such that $\delta_A = \delta_B = \delta$.*

(i) $F_{A,B} \in \mathscr{A}_2$ if, and only if, for all $x_1, x_2 \in [0,1]$ with $x_1 \le x_2$,

$$\delta_A(x_1) + \delta_A(x_2) \ge A(x_2,x_1) + B(x_1,x_2).$$

(ii) If A and B are symmetric, then $F_{A,B} \in \mathscr{A}_2$.
(iii) If $A \ge B$ or $B \ge A$, then $F_{A,B} \in \mathscr{A}_2$.

Moreover, $\delta_{F_{A,B}} = \delta$.

Proposition 2. *Let A,B be in $\mathscr{A}_2$ such that $\eta_A = \eta_B = \eta$.*

(i) Then $F^{A,B} \in \mathscr{A}_2$ if, and only if, for all $x_1, x_2 \in [0,1]$ with $x_1 \le x_2$

$$A(x_1, 1-x_2) + B(x_2, 1-x_1) \ge \eta(x_1) + \eta(x_2).$$

Moreover, $F^{A,B}$ is a copula if, and only if, B is a copula.
(ii) If $B \ge A$, then $F^{A,B} \in \mathscr{A}_2$.

Moreover, $\eta_{F_{A,B}} = \eta$.

4 The Rectangular Patchwork

Now, let us consider

- A rectangle $R = [a_1,a_2] \times [b_1,b_2] \subseteq [0,1]^2$ with $a_1 < a_2$, $b_1 < b_2$;
- A 2-increasing aggregation function $A\colon [0,1]^2 \to [0,1]$;
- A binary function $B\colon R \to [0,1]$, increasing in each place, such that $B = A$ on the boundaries of R, i.e., $B\restriction_{\partial R} = A\restriction_{\partial R}$ with $\restriction$ indicating the restriction of the function to ∂R.

Then the function $\big(B \hookrightarrow_R A\big)\colon [0,1]^2 \to [0,1]$ defined by

$$\big(B \hookrightarrow_R A\big)(x,y) = \mathbf{1}_R(x,y)B(x,y) + \big(1 - \mathbf{1}_R(x,y)\big)A(x,y)$$

is an aggregation function, called the *rectangular patchwork* of B in A on R. Notice that in [6] a similar approach, called *orthogonal grid construction*, for conjunctors, quasi-copulas, and copulas based on two operations both acting on the unit square has been discussed. However, for the rectangular patchwork $B \hookrightarrow_R A$ as discussed here, it

is sufficient to have some 2-increasing function $B\colon R \to [A(a_1,b_1),A(a_2,b_2)]$ at hand which is increasing in each argument and coincides with A on ∂R. We call such functions as *being admissible for the rectangular patchwork*. Clearly, the restriction of any 2-increasing aggregation function A' to R with $A' \restriction_{\partial R} = A \restriction_{\partial R}$ is admissible for a rectangular patchwork. The relationship between B and $B \hookleftarrow_R A$ and its properties is easily shown (compare also [6]):

Proposition 3. *Under the above assumptions, if B is 2-increasing, then so is $B \hookleftarrow_R A$. Moreover, if A is a copula, then $B \hookleftarrow_R A$ is also a copula.*

Moreover, we can provide the following characterization of functions B being admissible for a rectangular patchwork (compare also [13]) showing that, in fact, only the values of B at ∂R are of relevance.

Theorem 1. *Let A be in $\mathscr{A}_2$ and consider some rectangle $R = [a_1,a_2] \times [b_1,b_2] \subseteq [0,1]^2$ with $a_1 < a_2$, $b_1 < b_2$. Define $\lambda_{A,R} := V_A(R)$.*

(i) If $\lambda_{A,R} = 0$, then the only function B_R admissible for a rectangular patchwork is $B_R\colon R \to [A(a_1,b_1),A(a_2,b_2)]$ given by

$$B_R(x,y) = A(x,b_1) + A(a_1,y) - A(a_1,b_1) = A(x,b_2) + A(a_2,y) - A(a_2,b_2)$$

(ii) If $\lambda_{A,R} > 0$, then for every copula C, the function $B_R^C\colon R \to [A(a_1,b_1),A(a_2,b_2)]$ defined by

$$B_R^C(x,y) := \lambda_{A,R} C\left(\frac{V_A([a_1,x]\times[b_1,b_2])}{\lambda_{A,R}}, \frac{V_A([a_1,a_2]\times[b_1,y])}{\lambda_{A,R}}\right) + A(x,b_1) + A(a_1,y) - A(a_1,b_1) \tag{2}$$

is admissible for the rectangular patchwork.

(iii) If $\lambda_{A,R} > 0$, then for every function B being admissible for the rectangular patchwork, there exists a copula C such that $B = B_R^C$.

Remark 1. Notice that for some copula C, $\lambda_{C,R}$ denotes the value $V_C(R)$. Moreover, for every copula C and every rectangle R, $C \restriction_R$ is clearly a 2-increasing increasing function on R. Because of Theorem 1 it follows that there exists a copula C' such that $C \restriction_R = B_R^{C'}$. However, in general, C' might be different from C. Consider, e.g., the copula C given as a non-trivial convex combination of M and the product copula Π. Then C is a copula with a singular component just along the main diagonal. For a rectangle $R = [1-a,1] \times [0,a]$, $a \in \left]0,\frac{1}{2}\right[$, the rectangular patchwork $(B_R^C \hookleftarrow_R C)$ has an additional singular component along the line segment joining the points $(1-a,0)$ and $(1,a)$ such that $(B_R^C \hookleftarrow_R C)$ differs from C.

The following upper and lower boundaries for functions being admissible for the rectangular patchwork can be given:

Proposition 4. *Let A be in $\mathscr{A}_2$ and consider some rectangle $R = [a_1,a_2] \times [b_1,b_2] \subseteq [0,1]^2$ with $a_1 < a_2$, $b_1 < b_2$. Then $B_*, B^*\colon R \to [A(a_1,b_1),A(a_2,b_2)]$ defined by*

$$B_*(x,y) = \max\Big(A(x,b_1)+A(a_1,y)-A(a_1,b_1), A(x,b_2)+A(a_2,y)-A(a_2,b_2)\Big)$$
$$B^*(x,y) = \min\Big(A(x,b_2)+A(a_1,y)-A(a_1,b_2), A(x,b_1)+A(a_2,y)-A(a_2,b_1)\Big)$$

are admissible for the rectangular patchwork. Moreover, it holds that all functions B admissible for the rectangular patchwork fulfill

$$B_*(x,y) \le B(x,y) \le B^*(x,y)$$

for all $(x,y) \in R$.

So far, we have introduced the rectangular patchwork for a single rectangle only. However, by consecutive repetition of the patchwork technique we can extend the construction to being applicable for several rectangles: consider an aggregation function A and some index set $\mathscr{I}$ which is at most countable. Let $(R_i)_{i\in\mathscr{I}}$ be a family of rectangular subsets of the unit square with $R_i \cap R_j \subseteq \partial R_i \cap \partial R_j$ for any $i,j \in \mathscr{I}$ with $i \neq j$, expressing that any two members of this family overlap at most in their boundaries. In addition consider a family $(C_i)_{i\in\mathscr{I}}$ of copulas, indexed by the same index set, for which it holds that, if $R_i \cap R_j \neq \emptyset$, then $C_i \restriction_{R_i\cap R_j} = C_j \restriction_{R_i\cap R_j}$. Define the rectangular patchwork $(\langle R_i, C_i\rangle)^A_{i\in\mathscr{I}} \colon [0,1]^2 \to [0,1]$ by

$$(\langle R_i, C_i\rangle)^A_{i\in\mathscr{I}}(x,y) := \begin{cases} B^{C_i}_{R_i}(x,y), & \text{if } (x,y) \in R_i, \\ A(x,y), & \text{otherwise.} \end{cases} \tag{3}$$

Based on the results presented in this chapter we briefly discuss selected and already established constructions of copulas and its boundaries from the perspective of rectangular patchwork.

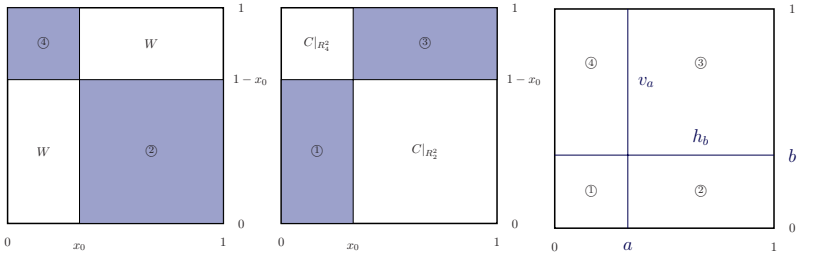
5 Examples

5.1 W-Ordinal Sum

Let us first consider the so called W-ordinal sum [4, 5, 21] of two copulas C', C'' defined in the following way

$$C(u,v) = \begin{cases} x_0\, C'\big(\frac{u}{x_0}, \frac{x_0+v-1}{x_0}\big), & \text{if } (u,v) \in [0,x_0]\times[1-x_0,1], \\ (1-x_0)\, C''\big(\frac{u-x_0}{1-x_0}, \frac{v}{1-x_0}\big), & \text{if } (u,v) \in [x_0,1]\times[0,1-x_0], \\ W(u,v), & \text{otherwise.} \end{cases} \tag{4}$$

It is known that C is again a copula, moreover, that any copula C with $C(x_0, 1-x_0) = 0$ for some $x_0 \in]0,1[$ can be represented as an W-ordinal sum [4, 5, 6, 21]. Indeed, C coincides with the rectangular patchwork $(\langle 0, x_0, C'\rangle, \langle x_0, 1, C''\rangle)^W$ (see Fig. 1(a)). Upper resp. lower bounds are obtained by choosing $C' = C'' = W$ resp. $C' = C'' = M$. On the other hand, for some copula C and some $x_0 \in]0,1[$ such that $C(x_0, 1-x_0) = 0$, it holds that $\lambda_{C,R_1} = 0$, with $R_1 = [0,x_0]\times[0,1-x_0]$, and therefore uniquely $B_{R_1}(x,y) = 0 = W(x,y)$. Further, clearly for $R_3 = [x_0,1]\times[1-x_0,1]$, $\lambda_{C,R_3} = 1-x_0-(1-x_0)+0 = 0$ such that and $B_{R_3}(x,y) = C(x,1)+C(1,y)-1 = W(x,y)$ (see also Fig. 1(b)).



W-ordinal sum (a) W-ordinal sum (b) Cross copula (c)

Fig. 1. Examples of rectangular patchwork operations

5.2 Cross Copulas

Now consider a copula C and $a,b \in [0,1]$ with $C(a,b) = c > 0$ and define $h_b(x) := C(x,b)$ and $v_a(y) := C(a,y)$. The function h_b is the *horizontal b-section* and v_a is the *vertical a-section* of C. Copulas with given horizontal and vertical section are known as *cross copulas* [7]. As shown in Figure 1 (c), the horizontal and vertical section divide the unit square into four rectangles R_i, $i = 1,\ldots 4$. Moreover, if we assume that $C(a,b) = c$ with $0 < c < \min(a,b)$ it follows that for each rectangle R_i, it holds that $\lambda_{C,R_i} > 0$, $i = 1,...,4$. As a consequence, for all copulas C_1, C_2, C_3, C_4, the function $C' \colon [0,1]^2 \to [0,1]$, defined by

$$C'(x,y) = \begin{cases} B_{R_1}^{C_1}(x,y), & \text{if } (x,y) \in R_1, \\ B_{R_2}^{C_2}(x,y), & \text{if } (x,y) \in R_2, \\ B_{R_3}^{C_3}(x,y), & \text{if } (x,y) \in R_3, \\ B_{R_4}^{C_4}(x,y), & \text{if } (x,y) \in R_4, \end{cases}$$

is a cross-copula with the prescribed horizontal and vertical section [7]. It is immediate from Theorem 1 that choosing $C_i = W$ (resp. M), $i = 1,\ldots 4$, leads to the smallest (resp. greatest) such crosscopula. On the other hand, for any cross-copula C as described above there exist copulas C_i, $i = 1,\ldots 4$ such that

$$C' = (\langle R_1, C_1\rangle, \langle R_2, C_2\rangle, \langle R_3, C_3\rangle, \langle R_4, C_4\rangle)^C.$$

which again demonstrates how the general notion of rectangular patchwork reduces to more particular crosscopula construction.

6 Conclusion

We have presented necessary and sufficient conditions for 2-increasing aggregation functions obtained via diagonal or rectangular patchwork techniques. We have illustrated by two examples how the characterization obtained in Theorem 1 allows to revisit known constructions for copulas and its boundaries.

References

1. Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Studies in Fuzziness and Soft Computing, vol. 221. Physica-Verlag, Heidelberg (2007)
2. Bouchon-Meunier, B. (ed.): Aggregation and Fusion of Imperfect Information. Studies in Fuzziness and Soft Computing, vol. 12. Physica-Verlag, Heidelberg (1998)
3. Calvo, T., Mayor, G., Mesiar, R. (eds.): Aggregation operators: New trends and applications. Studies in Fuzziness and Soft Computing, vol. 97. Physica-Verlag, Heidelberg (2002)
4. De Baets, B., De Meyer, H.: From the comparison of ordered lists to the construction of copulas. In: López-Díaz, M., Gil, M.A., Grzegorzewski, P., Hyrniewicz, O., Lawry, L. (eds.) Soft Methodology and Random Information Systems, pp. 165–172. Springer, Heidelberg (2004)
5. De Baets, B., De Meyer, H.: Copulas and pairwise probabilistic comparison of ordered lists. In: Proceedings of Tenth International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2004, Perugia, Italy), pp. 1091–1098 (2004)
6. De Baets, B., De Meyer, H.: Orthogonal grid constructions of copulas. IEEE Trans. Fuzzy Syst. 15, 1053–1062 (2007)
7. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Copulas with given values on a horizontal and a vertical section. Kybernetika 43, 209–220 (2007)
8. Durante, F., Kolesárová, A., Mesiar, R., Sempi, C.: Copulas with given diagonal sections: novel constructions and applications. Int. J. Uncertain. Fuzziness Knowl-Based Syst. 15(4), 397–410 (2007)
9. Durante, F., Mesiar, R., Papini, P.L., Sempi, C.: 2-increasing binary aggregation operators. Inform. Sci. 177(1), 111–129 (2007)
10. Durante, F., Mesiar, R., Sempi, C.: Copulas with given diagonal section: some new results. In: Montseny, E., Sobrevilla, P. (eds.) Proceedings of EUSFLAT–LFA Conference, Universitat Politecnica de Catalunya, Barcelona, pp. 932–936 (2005)
11. Durante, F., Mesiar, R., Sempi, C.: On a family of copulas constructed from the diagonal section. Soft Computing 10(6), 490–494 (2006)
12. Durante, F., Saminger-Platz, S., Sarkoci, P.: On representations of 2-increasing binary aggregation functions (submitted for publication) (2008)
13. Durante, F., Saminger-Platz, S., Sarkoci, P.: Rectangular patchwork for bivariate copulas and tail dependence (submitted for publication) (2008)
14. Fredricks, G.A., Nelsen, R.B.: Diagonal copulas. In: Beneš, V., Štepán, J. (eds.) Distributions with Given Marginals and Moment Problems, pp. 121–128. Kluwer Academic Publishers, Dordrecht (1997)
15. Fredricks, G.A., Nelsen, R.B.: Copulas constructed from diagonal sections. In: Beneš, V., Štepán, J. (eds.) Distributions with Given Marginals and Moment Problems, pp. 129–136. Kluwer Academic Publishers, Dordrecht (1997)
16. Klement, E.P., Kolésarová, A.: Extensions to copulas and quasi-copulas as special 1-Lipschitz aggregation operators. Kybernetika 41, 329–348 (2005)
17. Klement, E.P., Kolésarová, A.: Intervals of 1-Lipschitz aggregation operators, quasi-copulas, and copulas with given affine section. Monatsh Math. 152, 151–167 (2007)
18. Klement, E.P., Kolésarová, A., Mesiar, R., Sempi, C.: Copulas constructed from horizontal sections. Comm. Statist. Theory Methods 36(16), 2901–2911 (2008)
19. Klement, E.P., Mesiar, R. (eds.): Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms. Elsevier, Amsterdam (2005)
20. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms, Trends in Logic. Studia Logica Library, vol. 8. Kluwer Academic Publishers, Dordrecht (2000)

21. Mesiar, R., Szolgay, J.: W-ordinals sum of copulas and quasi-copulas. In: Proceedings of MAGIA 2004 Conference (Kočovce, Slovak Republic), pp. 78–83 (2004)
22. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer Series in Statistics. Springer, New York (2006)
23. Nelsen, R.B., Quesada Molina, J.J., Rodríguez Lallena, J.A., Úbeda Flores, M.: Best-possible bounds on sets of bivariate distribution functions. J. Multivariate Anal. 90, 348–358 (2004)
24. Nelsen, R.B., Quesada Molina, J.J., Rodríguez Lallena, J.A., Úbeda Flores, M.: On the construction of copulas and quasi–copulas with given diagonal sections. Insur. Math. Econ. 42, 473–483 (2008)
25. Quesada-Molina, J.J., Saminger-Platz, S., Sempi, C.: Quasi-copulas with a given sub-diagonal section. Nonlinear Anal Theory Methods Appl. (in press, 2008) doi: 10.1016/j.na.2007.11.021
26. Siburg, K.F., Stoimenov, P.A.: Gluing copulas (2008), http://www.sfb475.uni-dortmund.de/berichte/tr31-07.pdf
27. Sklar, A.: Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231 (1959)
28. Torra, V., Narukawa, Y.: Modeling Decisions: Information Fusion and Aggregation Operators. Cognitive Technologies Series. Springer, Berlin (2007)

Part V

Engineering

Computational Methods

A Hierarchical Fusion of Expert Opinion in the TBM

Minh Ha-Duong

CIRED, CNRS, Nogent-sur-Marne, France

Abstract. We define a hierarchical method for expert opinion aggregation that combines consonant beliefs in the Transferable Belief Model. Experts are grouped into schools of thought, then opinions are aggregated using the cautious conjunction operator within groups and the non-interactive disjunction across. This is illustrated with a real-world dataset including 16 experts.

Keywords: Evidence theory, Information fusion, Expert opinion.

1 Introduction

Aggregating the opinion of experts is challenging for several reasons. Scientists interact and share evidence, so assuming independence leads to over-precise results. Conjunctive aggregation methods do not work well when there is contradiction among experts. In a dissent minority situation, aggregation methods like averaging that weight views proportionally to the number of proponents are arguably unbalanced scientifically. Finally, it is difficult to assume that some experts are less reliable than others.

To address these challenges, this paper proposes a hierarchical method: experts are grouped into schools of thought. Within groups, beliefs are combined using a cautious conjunction rule, and across with the non-interactive disjunction.

Section 2 recalls elements of the Transferable Belief Model, based on [3]. Section 3 presents the data, the implementation and discusses the proposed hierarchical approach. Section 4 compares it theoretically and numerically to other fusion procedures. Section 5 concludes.

2 The Transferable Belief Model

The Transferable Belief Model represents and combines uncertain beliefs elaborating upon Dempster-Shafer evidence theory [3]. Uncertainty is represented by allocating the unit mass of belief among *subsets* of a frame of reference Ω. Let 2^Ω denote the power set of Ω. Its elements will be denoted with upper case letters such as $A \subseteq \Omega$ or $X \subseteq \Omega$. The empty subset will be denoted $\emptyset$. A basic belief assignment (BBA) is a function $m : 2^\Omega \to [0,1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 361–368, 2008.
springerlink.com

Belief that the state of the world is in $A \subseteq \Omega$, with certainty, is represented by the BBA $\mathbf{1}_A : 2^\Omega \to [0,1]$ defined as $\mathbf{1}_A(A) = 1$, and $\mathbf{1}_A(X) = 0$ if $X \neq A$. Discounting the BBA m means replacing it by $\mathrm{dis}(m,r) = rm + (1-r)\mathbf{1}_\Omega$.

Belief that the state of the world is in $A \subsetneq \Omega$, with a weight of belief s, is represented by the function $A^s = \mathrm{dis}(\mathbf{1}_A, 1-e^{-s})$.

The Transferable Belief Model allows non-zero belief mass to the empty set. The number $m(\emptyset)$, called weight of conflict, represents the internal contradiction arising when beliefs result from information sources pointing in different directions. The extreme case $\mathbf{1}_\emptyset$ represents being completely confused by contradictory information sources. Renormalizing a BBA m means replacing it by the BBA m^* defined as:

$$\begin{cases} m^*(\emptyset) = 0 \\ m^*(A) = \frac{m(A)}{1-m(\emptyset)} \quad \text{if } A \neq \emptyset \end{cases}$$

Several functions can be defined from a BBA m. These functions are defined by $bel(\emptyset) = pl(\emptyset) = 0$, $q(\emptyset) = 1$, and for any $X \neq \emptyset$ as:

$$bel(X) = \sum_{\emptyset \neq A \subseteq X} m(A) \qquad pl(X) = \sum_{A \cap X \neq \emptyset} m(A)$$

$$q(X) = \sum_{A \supseteq X} m(A) \qquad s(X) = \sum_{A \supseteq X} (-1)^{|X|-|A|} \ln\big(q(A)\big)$$

These are called the belief function bel, the plausibility function pl, the commonality function q and the weights of belief function s, only defined here when $m(\Omega) > 0$. An intuitive interpretation of the theory of evidence sees $m(X)$ as a mass of belief that can flow to any subset of X. In this view, $bel(X)$ represents the minimal amount of mass that is constrained to stay within X (the belief that X must happen), while $pl(X)$ the maximal amount of belief that could flow into X (the plausibility that X could happen). Commonality $q(X)$ represents the amount of belief that can flow to every point of X. An interpretation of the weights of belief s will be discussed with the factorization Eq. (2) below.

The two basic combination rules of the Transferable Belief Model are denoted $\textcircled{\wedge}$ and $\textcircled{\vee}$. The non-interactive conjunction rule $\textcircled{\wedge}$ should be used when assuming that both information sources are reliable, and the non-interactive disjunction rule $\textcircled{\vee}$ when assuming that at least one is reliable. These rules reduce to classical set intersection and set union when BBAs are reduced to a single set receiving all the mass, i.e. $\mathbf{1}_A \textcircled{\vee} \mathbf{1}_B = \mathbf{1}_{A \cup B}$. These operators are commutative, associative and give a BBA if μ_1 and μ_2 are BBAs.

$$(\mu_1 \textcircled{\wedge} \mu_2)(X) = \sum_{A \cap B = X} \mu_1(A) \times \mu_2(B), \qquad (\mu_1 \textcircled{\vee} \mu_2)(X) = \sum_{A \cup B = X} \mu_1(A) \times \mu_2(B)$$

Dempster's combination rule $\oplus$ is normalized non-interactive conjunction:

$$m_1 \oplus m_2 = (m_1 \textcircled{\wedge} m_2)^*.$$

Weights of belief allow to express any BBA m such that $m(\Omega) > 0$ as the conjunction of elementary pieces of evidence [8]:

$$m = \bigcirc\!\!\!\!\!\!\bigcap_{A \subset \Omega} A^{s(A)} \tag{2}$$

Weights s can be negative. If $s < 0$, then A^s is not a BBA but a generalized BBA: a real-valued subset function which verifies Eq. (1), but may take values outside of [0,1]. We suggest to interpret A^s as the change in one's beliefs realized by giving confidence s to a new piece of evidence stating that the state of the world is in A. Positive infinity for s represents the limit case of a perfectly convincing proof. Negative weight $s < 0$ have an algebraic justification comparable to negative numbers: considering A with weight s exactly counterbalances considering A with weight $-s$.

For any two BBA m_1 and m_2 having weight functions s_1 and s_2:

$$m_1 \textcircled{\cap} m_2 = \bigcap_{\substack{A \subset \Omega \\ A \neq \Omega}} A^{s_1(A)+s_2(A)}.$$

The non-interactive conjunction of two pieces of evidence pointing in the same direction adds up to stronger beliefs, i.e. $A^s \textcircled{\wedge} A^s = A^{2s}$. Using this rule is adequate when information sources can be assumed to be independent. But scientific experts share data, models and theories, so we have to consider also an alternative to $\textcircled{\wedge}$. The cautious conjunction operator, denoted $\textcircled{\cap}$, was recently proposed [3]. It combines any two BBA such that $m_1(\Omega) > 0$ and $m_2(\Omega) > 0$ by:

$$m_1 \textcircled{\wedge} m_2 = \bigcap_{A \subset \Omega} A^{\max(s_1(A), s_2(A))}.$$

If m_1 and m_2 are BBAs, then $m_1 \textcircled{\cap} m_2$ is also a BBA. The rule $\textcircled{\cap}$ is commutative and associative, idempotent ($m \textcircled{\cap} m = m$), and distributes over the noninteractive rule $(m_1 \textcircled{\wedge} m_2) \textcircled{\cap} (m_1 \textcircled{\wedge} m_3) = m_1 \textcircled{\wedge} (m_2 \textcircled{\cap} m_3)$. Distributivity has an interesting interpretation related to the fusion of beliefs. Consider two experts in the following scenario. Expert 1's belief results from the noninteractive conjunction of two pieces of evidence, $m_1 = A^s \textcircled{\wedge} B^t$. Expert 2 shares one piece of evidence with expert 1, and has an independent piece, so that $m_2 = A^s \textcircled{\wedge} C^u$. Then distributivity implies that in the fusion, the shared evidence A^s is not counted twice $m_1 \textcircled{\cap} m_2 = A^s \textcircled{\wedge} (B^t \textcircled{\cap} C^u)$.

3 An Expert Aggregation Situation

Climate sensitivity is a proxy for the severity of the climate change problem. It is denoted $\Delta T_{2\times}$, and defined as the equilibrium global mean surface temperature change following a doubling of atmospheric CO_2 concentration, compared to pre-industrial levels. The value of this parameter, critical for climate policy, is not known precisely: for a long time, the [1.5°C, 4.5°C] interval has been regarded as the canonical uncertainty range of $\Delta T_{2\times}$.

Ref. [5] conducted structured interviews using expert elicitation methods drawn from decision analysis with 16 leading U.S. climate scientists. The authors obtained judgments about a number of key climate variables, including subjective PDFs for the climate sensitivity parameter. In the dataset, no probability is allocated to climate sensitivity lower than −6°C, or larger than 12°C. For the sake of numerical tractability, this range was subdivided in seven ranges at the subdivision points {-6, 0, 1.5, 2.5, 3.5, 4.5, 6, 12}, and PDFs were discretized to obtain, for each expert, a probability distribution p_i on $\Omega = \{\omega_1, \ldots, \omega_7\}$.

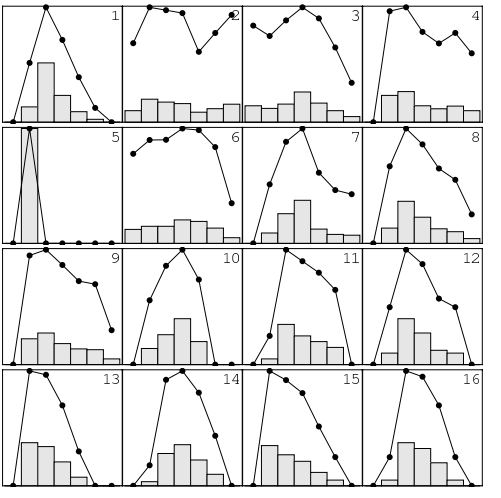


Fig. 1. Beliefs of the 16 experts. Vertical axis goes from 0 to 1. The horizontal axis discretizes the [-6°C, 12°C] climate sensitivity range into seven intervals using a non-uniform subdivision at -6, 0, 1.5, 2.5, 3.5, 4.5, 6 and 12°C. On each graph, the grey histogram represents the actual elicited probability distribution, and the dotted lines represents the possibility distribution corresponding to the implicit consonant belief function.

Then we transformed each p_i into an implicit consonant belief function m_i. Given a probability distribution p, a consonant belief function m is defined as follows [9]. Order the states of the world from most to least probable, that is $p(\omega_{n_1}) > \cdots > p(\omega_{n_{|\Omega|}})$. Consider the sets $A_k = \{\omega_{n_1}, \ldots, \omega_{n_k}\}$ and assign to A_k the belief mass $m(A_k) = |A| \times \big(p(\omega_{n_k}) - p(\omega_{n_{k+1}})\big)$ with the convention that $p_{n_{|\Omega|+1}} = 0$. Figure 1 represents the results, the beliefs of the 16 experts.

Most experts BBA m_i verify $m_i(\Omega) = 0$, and cannot be factorized as described above. But no information source is 100% reliable; scientific knowledge is necessarily based on a possibly large but finite number of human observation; many philosophers consider that scientific knowledge should always be open to revision in front of new experimental evidence; and the elicitation of opinions was necessarily coarse so experts who allocated no significant probability weight to extreme outcomes might have agreed that there was a very small possibility. These reasons justify discounting the opinion of experts using a high reliability factor such as $r = 0.999$.

Simple ways to combine expert opinions pool all beliefs together symmetrically using a fusion operator like $\textcircled{\wedge}$, $\textcircled{\circ}$, $\textcircled{\vee}$, or $\oplus$. These simple methods have problems when aggregating conflicting beliefs, which have led researchers to suggest *adaptative* fusion rules instead [6, 1, 4]. The general idea is to merge conjunctively subgroups of coherent sources, before disjunctively merging the different results. We propose a hierarchical fusion procedure based on this idea. It aims to be relevant when science is not yet stabilized, and the notion of 'competing theories' can be used. Sociology of science suggests that at some moments in the progress of science, in front of a big

unexplained problem, scientists tend to group into schools of thought, which correspond to alternative candidate theories.

Within each group, contradiction should be low because experts share an explanation of the way the world works. Experts can all be considered reliable information sources, but the independence assumption does not hold. Beliefs will be combined with a cautious conjunction operator.

To combine across groups, we assume that only one theory will be adopted in the end and use the non-interactive disjunction operator. This deals with the challenge of representing equally minority views because all theories are treated equally, regardless of the number of experts in the group. Using the $\copyright$ operator also assumes that schools of though are non-interactive. The bold disjunctive combination rule could be used to relax this assumption [3].

This hierarchical approach is a mathematical aggregation method based on a qualitative behavioral analysis: a partition of experts into a small numbers of schools of thought. Representing the diversity of viewpoints by a small numbers of schools of though is admittedly a strong simplification of complex social reality, but treating all experts symmetrically is even simpler. If it were clear from the start what the different schools of though are, one could select a single expert to represent each position, and then pool the opinions symmetrically. Otherwise, it is only after a formal sociological study of the experts community that the population of experts can be organized around a small number of archetypes.

Adaptative fusion methods determine the groups using the beliefs themselves, but for experts the tools of sociology could be used. For example, the network of experts can be analyzed with catalogues of publications since experts who have published together have seen the same data, they are more likely to share evidence. Another classical method to determine how a group of people is organized is to analyze the content of the semi-structured face-to-face interviews conducted in the expert elicitation. Finally, the experts themselves know their community, they can help to discover how it is organized, and they can validate the results of the sociological analysis.

4 Comparing Fusion Methods

This section critically assesses the different ways to combine opinions defined above, theoretically and numerically. It discusses the results presented on Figure 2, which allows to compare 8 alternative ways to perform the fusion of expert opinion. On each plot, the vertical axis goes from 0 to 1, and horizontally the numbers (from 1 to 7) denote the states of the world ω_1 to ω_7. There are three series of points on each plot. The top one is labelled pl, while the middle one is labelled p and the bottom bel. They display respectively the plausibility of singletons $pl(\{\omega_i\})$, the pignistic probability $p^m(\omega_i) = \frac{1}{1-m(\emptyset)} \sum_{\omega \in X} \frac{m(X)}{|X|}$, and the belief $bel(\{\omega_i\})$.

The top left plot presents the result obtained with Dempster's rule, $\bigoplus_{i=1\ldots n} \text{dis}(m_i, 0.8)$. This rule has many drawbacks for pooling expert opinion. It requires discounting otherwise it is known to give counter-intuitive results when there is contradiction [10]. Having no evidence to determine the reliability factor, we used $r = 0.8$. Dempster's rule also assumes independence, reducing the plausibility of states of the

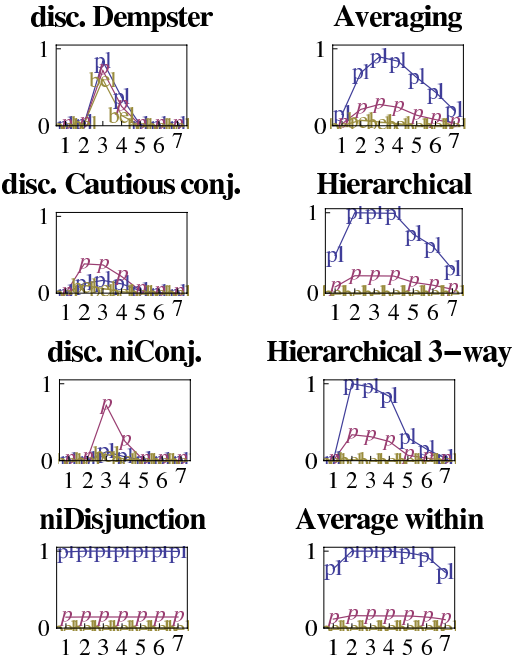


Fig. 2. Comparing 8 alternative procedures to fusion expert opinion

world that are outside the central range: on the figure, beliefs are very focused around ω_3. Much of this precision is unwarranted as experts are not independent.

Consider now the second and third cases in the left column. They show respectively the results obtained with cautious conjunction $\bigcap_{i=1\ldots n}\mathrm{dis}(m_i, 0.8)$ and with non-interactive conjunction $\bigwedge_{i=1\ldots n}\mathrm{dis}(m_i, 0.8)$. Since these rules produce the trivial result $\mathbf{1}_\emptyset$ when the information sources conflict completely, discounting is needed to recover informative results. This is hard to justify, when the whole point of the Transferable Belief Model is to accept $\mathbf{1}_\emptyset$ as a theoretically correct result. The precision of the non-interactive conjunction is also highly questionable given that experts interact.

The left bottom plot shows the result of the non interactive disjunction $\bigcup_{i=1\ldots n}$ $\mathrm{dis}(m_i, 0.999)$. This operator produces uninformative beliefs close to $\mathbf{1}_\Omega$. Discounting could only make the result even less informative. This operator has potential to combine Bayesian, but not consonant, beliefs.

Turn now to the right column and averaging, also called the linear opinion pool: $\frac{1}{n}\sum_{i=1\ldots n}\mathrm{dis}(m_i, 0.999)$. The theoretical criticism of averaging is that the weight of an opinion increases with the number of experts holding it. Yet scientific arguments should be evaluated on their own merits, not by *argumentum ad populum* (Latin: "appeal to the people"). It is only at the social decision-making stage that the quality and number of people behind each view should matter. Groupthink and bandwagon effects are known

dangers when pooling opinions. Thus, a fusion method that gives equal attention to the minority and the majority views would be preferable.

Figure 2, the second plot on the right shows our central result: the hierarchical fusion: $\bigcirc_{k=1\ldots N} \bigcirc_{i\in G_k} \mathrm{dis}(m_i, 0.999)$. The different schools of thought are as follows (see Figure 1): Experts in group $G_1 = \{2,3,6\}$ allow cooling. Those in group $G_2 = \{4,7,8,9\}$ allow high outcomes but no cooling, $G_3 = \{1,10\ldots16\}$ disallow extreme cases, and the opinion of the single expert outlying in $G_4 = \{5\}$ is concentrated on [0°C,1°C].

According to the hierarchical aggregation of opinions, the uncertainty range $\{\omega_2, \omega_3, \omega_4\}$, i.e. 0°C to 3.5°C is completely plausible, and higher values are very plausible. Experts were interviewed in 1995 and these results should be compared to the more recent scientific literature from a policy perspective. Our understanding is that the plausibility of the $\Delta T_{2\times} < 1.5$°C case has decreased a lot, while the plausibility that $\Delta T_{2\times} > 3.5$°C has remained high.

We conducted two sensitivity analysis. We merged G_1 and G_2 together for a 3-way hierarchical fusion (third plot). Results appear significantly sensitive to the clustering of experts: the plausibility of the 'above 4.5°C' case, drops from 0.61 to 0.15. Lastly, we examined a hierarchical fusion where the first step is averaging, rather than the cautious conjunction. The plausibility levels increase: the result is more ambiguous.

5 Summary and Conclusion

Symmetrical fusion procedures have problems to aggregate expert opinion when there is a range of competing scientific theories. Conjunctions only say 'Experts contradict each other', while disjunctions say 'Everything is possible'. Dempster's rule lead to overconfidence. Averaging assumes that finding scientific truth is like majority voting.

The proposed hierarchical fusion procedure is built around a simple model of experts' social relations: they are divided into schools of thought. Beliefs are aggregated using the cautious conjunction operator within, and combined using the non-interactive disjunction rule across groups. This solves several theoretical problems with opinion aggregation. It does not use discounting, thus avoiding the issue of expert calibration. Within groups, cautious conjunction does pool together distinct streams of evidence to make beliefs firmer. But it is not assumed that opinions are independent: this would overestimate the precision of actual information. Disjunction allows to deal with complete contradiction among opinions without falling into degenerate results or paradoxes. When several scientific theories compete to explain the same observations, it should not be assumed that both are true at the same time (conjunction), but that at least one will remain (disjunction). Pooling opinions across schools of thoughts, rather than across individual experts, is arguably a more balanced procedure. Unlike averaging, minority views are equally taken into account.

Even with purely mathematical expert aggregation methods, one has to make sure that no major point of view is omitted when selecting the experts. Therefore sociological considerations on the population of experts cannot really be avoided. The hierarchical approach brings forward transparently that qualitative analysis. Finding out the detailed

structure of epistemic communities to explain the differences between theories may be as policy-relevant as computing aggregate beliefs.

Acknowledgement. We gratefully thank David Keith for the data and acknowledge the precious comments of T. Denœux, CIRED colleagues and two anonymous referees.

References

1. Ayoun, A., Smets, P.: Data association in multi-target detection using the Transferable Belief Model. Int. J. Intell. Syst. 16(10), 1167–1182 (2001)
2. Cooke, R.M.: Experts in Uncertainty. In: Opinion and Subjective Probability in Science. Environmental Ethics and Science Policy Series, The Clarendon Press, Oxford University Press, New York (1991)
3. Denœux, T.: Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. Artificial Intelligence 172(2–3), 234–264 (2007)
4. Destercke, S., Dubois, D., Chojnacki, E.: Fusion d'opinions d'experts et théories de l'incertain. In: Proceedings of Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2006, Toulouse, France) Cépaduès edn., pp. 295–302 (2006)
5. Granger Morgan, M., Keith, D.W.: Subjective judgments by climate experts. Environ. Sci. Technol. 29(10), 468A–476A (1995)
6. Schubert, J.: Cluster-based specification techniques in Dempster-Shafer theory. In: Symbolic and Quantitative Approaches to Reasoning and Uncertainty, pp. 395–404 (1995)
7. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
8. Smets, P.: The canonical decomposition of a weighted belief. In: International Joint Conference on AI (IJCAI 1995, Montréal, Québec, Canada), pp. 1896–1901. Morgan Kaufmann, San Francisco (1995)
9. Dubois, D., Prade, H., Smets, P.: A definition of subjective possibility. Internat. J. Approx. Reason (in Press, 2008) doi:10.1016/j.ijar.2007.01.005
10. Zadeh, L.A.: On the Validity of Dempster's Rule of Combination of Evidence. UCB/ERL Memo M79/24, University of California, Berkeley (1979)

Application of Non-convex Fuzzy Variables to Fuzzy Structural Analysis

Uwe Reuter

Department of Civil Engineering, Technische Universität Dresden, Dresden, Germany

Abstract. Data situation in engineering is usually characterized by uncertainty. Fuzzy set theory provides adequate modeling of specific uncertainty phenomena. In this paper a new approach for modeling and processing of imprecise data by means of non-convex fuzzy variables is presented. For numerical processing an enhanced discretization of non-convex fuzzy variables is introduced. Furthermore fuzzy structural analysis is enhanced in order to be applicable to non-convex fuzzy variables. Fuzzy structural analysis under consideration of non-convex fuzzy variables is demonstrated by way of a two bay steel structure.

Keywords: Imprecise data, Non-convex fuzzy variables, Discretization, Fuzzy structural analysis.

1 Introduction

Data situation in engineering is usually characterized by uncertainty. The term uncertainty summarizes phenomena like imprecise, vague, fluctuating, fragmentary, linguistic, or subjective data. Realistic numerical modeling of engineering structures requires an adequate modeling of those uncertainty phenomena with respect to their sources. In general uncertainty may thus be subdivided into aleatoric and epistemic uncertainty [1]. Aleatoric uncertainty takes variability into account and is described by means of stochastic models. Epistemic uncertainty takes data into account, which overextend pure stochastic models. Non-stochastic models are amongst others provided by the fuzzy set theory (e. g. [6, 7]). Imprecise data are modeled as fuzzy variables (e. g. [4]) and are taken into account within the fuzzy structural analysis (e. g. [2, 3, 5]).

In this paper a new approach for modeling and processing of imprecise data by means of non-convex fuzzy variables is presented. For numerical processing an enhanced discretization of non-convex fuzzy variables is introduced. Furthermore fuzzy structural analysis is enhanced in order to be applicable to non-convex fuzzy variables. Non-convex fuzzy variables occur amongst others within the modeling of smooth transitions between two complementary states 1 and 2, e. g., measurement results of interfaces are only available as grey-tones. Such measurement results may be characterized by continuous, monotonic increasing functions $B(x)$. The function $B(x)$ maps the measured values onto the interval $[0;1]$. All measured values assignable to state 1 are rated with $B(x) = 0$ and all measured values assignable to state 2 are rated with $B(x) = 1$. The shape of the function inbetween has to be defined by the observer. By derivating

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 369–375, 2008.
springerlink.com

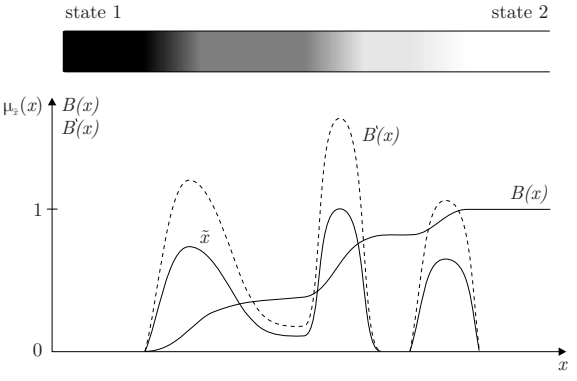


Fig. 1. Modeling of a smooth transition between two complementary states 1 and 2 as non-convex fuzzy variable $\tilde{x}$

function $B(x)$ and standardization the membership function $\mu_{\tilde{x}}(x)$ is received. In dependency of the function $B(x)$ non-convex fuzzy variables may be the result (see Fig. 1).

2 Discretization of Non-convex Fuzzy Variables

A fuzzy variable $\tilde{x}$ is defined as an uncertain subset of the fundamental set $\mathbf{X}$.

$$\tilde{x} = \{x, \mu_{\tilde{x}}(x) \,|\, x \in \mathbf{X}\} \tag{1}$$

The uncertainty is assessed by the membership function $\mu_{\tilde{x}}(x)$.

A normalized membership function $\mu_{\tilde{x}}(x)$ is defined as follows:

$$0 \le \mu_{\tilde{x}}(x) \le 1 \quad \forall\, x \in \mathbb{R} \tag{2}$$

$$\exists\, x_l,\ x_r \ \text{ with } \ \mu_{\tilde{x}}(x) = 1 \quad \forall\, x \in [x_l; x_r]\,. \tag{3}$$

A fuzzy variable $\tilde{x}$ is referred to as convex if its membership function $\mu_{\tilde{x}}(x)$ monotonically decreases on each side of the maximum value, i. e. if

$$\mu_{\tilde{x}}(x_2) \ge \min\left[\mu_{\tilde{x}}(x_1);\ \mu_{\tilde{x}}(x_3)\right] \quad \forall x_1, x_2, x_3 \in \mathbb{R} \ \text{ with } \ x_1 \le x_2 \le x_3 \tag{4}$$

applies.

For all $\alpha \in (0,1]$ a set of e_α closed finite intervals $[x_{\alpha l}^{j^*}; x_{\alpha r}^{j^*}]$, $j^* = 1, 2, ..., e_\alpha$, may be extracted from a non-convex fuzzy variable $\tilde{x}$. It holds:

$$x_{\alpha l}^1 \le\ x_{\alpha r}^1 < x_{\alpha l}^2 \le\ x_{\alpha r}^2 < \ldots < x_{\alpha l}^{e_\alpha} \le\ x_{\alpha r}^{e_\alpha} \tag{5}$$

Thus a non-convex fuzzy variable $\tilde{x}$ may be characterized by a family of α-level sets X_α according to Eq. (6).

$$\tilde{x} = \left(X_\alpha = \left\{[x_{\alpha l}^1; x_{\alpha r}^1],\, [x_{\alpha l}^2; x_{\alpha r}^2],\, ...,\, [x_{\alpha l}^{e_\alpha}; x_{\alpha r}^{e_\alpha}]\right\} \,|\, \alpha \in [0,1]\right) \tag{6}$$

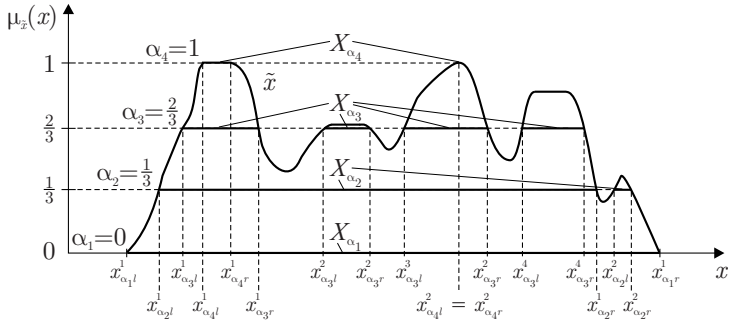


Fig. 2. Non-convex fuzzy variable $\tilde{x}$ discretized by $n = 4$ α-level sets X_α

If the number of α-level sets is denoted by n, then for $i = 1, 2, ..., n-1$ the following holds provided that $n \geq 2$:

$$0 \leq \alpha_i \leq \alpha_{i+1} \leq 1 \tag{7}$$

$$\alpha_1 = 0 \quad \text{and} \quad \alpha_n = 1 \tag{8}$$

$$X_{\alpha_{i+1}} \subseteq X_{\alpha_i} \tag{9}$$

In Fig. 2 a non-convex fuzzy variable $\tilde{x}$ disretized by $n = 4$ α-level sets X_α is shown.

3 Fuzzy Structural Analysis under Consideration of Non-convex Fuzzy Variables

By means of fuzzy structural analysis it is possible to map the fuzzy input variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_l$ onto the fuzzy result variables $\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_m$.

$$\underline{\tilde{z}} = (\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_m) = \tilde{f}(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_l) \tag{10}$$

The solution of Eq. (10) may be found by applying the extension principle. Under the condition that the fuzzy variables $\tilde{x}_k$, $k = 1, 2, ..., l$, are convex, however, α-level optimization is numerically more efficient [3]. In case of non-convex fuzzy variables $\tilde{x}_k$ the α-level optimization has to be generalized. The generalized α-level optimization approach is based on multiple discretization. All fuzzy variables $\tilde{x}_k$ and $\tilde{z}_j$, $j = 1, 2, ..., m$, are discretized using the same number of α-levels α_i, $i = 1, 2, ..., n$. The α-level sets X_{k,α_i} of the fuzzy input variables $\tilde{x}_k$, $k = 1, 2, ..., l$, are sets of $e_{\tilde{x}_k,\alpha_i}$ intervals $[x^{j^k}_{k,\alpha_i l}; x^{j^k}_{k,\alpha_i r}]$, $j^k = 1, 2, ..., e_{\tilde{x}_k,\alpha_i}$ and form thus a non-continuous l-dimensional crisp subspace $\underline{X}_{\alpha_i}$.

If no interaction exists between the fuzzy input variables, the subspace $\underline{X}_{\alpha_i}$ forms $\prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$ l-dimensional hypercuboids. The crisp subspace $\underline{Z}_{\alpha_i}$ is assigned to the crisp subspace $\underline{X}_{\alpha_i}$ on the same α-level. These are constructed from the α-level sets Z_{j,α_i}, $j = 1, 2, ..., m$, of the fuzzy result variables $\tilde{z}_j$. The α-level sets Z_{j,α_i} of the fuzzy result variables $\tilde{z}_j$ are sets of $e_{\tilde{z}_j,\alpha_i}$ intervals $[z^{s^j}_{j,\alpha_i l}; z^{s^j}_{j,\alpha_i r}]$, $s^j = 1, 2, ..., e_{\tilde{z}_j,\alpha_i}$. Each subspace $\underline{Z}_{\alpha_i}$ forms $\prod_{j=1}^{m} e_{\tilde{z}_j,\alpha_i}$ m-dimensional hypercuboids.

Each point of the subspace $\underline{X}_{\alpha_i}$ is uniquely described by the coordinates $x_1, x_2, \ldots, x_l$. Each point in the subspace $\underline{Z}_{\alpha_i}$ may be computed by means of

$$\underline{z} = (z_1, z_2, ..., z_m) = f(x_1, x_2, ..., x_l)\,. \tag{11}$$

The mapping $f(\cdot)$ is referred to as the deterministic fundamental solution. This represents an arbitrary computational model, e. g. a finite element model.

Under the condition that the fuzzy variables are convex, it is sufficient to compute the largest element $z_{j,\alpha_i r}$ and the smallest element $z_{j,\alpha_i l}$ of Z_{j,α_i}. If these two elements are known for a sufficient number of α-levels, the membership function $\mu_{\tilde{z}_j}(z_j)$ may be stated in discretized form.

In the case of non-convex fuzzy input variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_l$ no convex fuzzy result variables $\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_m$ can be assumed. The largest element $z_{j,\alpha_i r}$ and the smallest element $z_{j,\alpha_i l}$ of Z_{j,α_i} only yields the envelope of the membership function $\mu_{\tilde{z}_j}(z_j)$. The complete description of the membership function $\mu_{\tilde{z}_j}(z_j)$ requires the computation of the interval boundaries $z^{s^j}_{j,\alpha_i l}$ and $z^{s^j}_{j,\alpha_i r}$, $s^j = 1, 2, ..., e_{\tilde{z}_j,\alpha_i}$, of the α-level sets Z_{j,α_i}.

The computation of the interval boundaries $z^{s^j}_{j,\alpha_i l}$ and $z^{s^j}_{j,\alpha_i r}$, $s^j = 1, 2, ..., e_{\tilde{z}_j,\alpha_i}$, succeeds in two steps. In the first step the optimization problems, each with the objective functions

$$z_j = f(x_1, x_2, ..., x_l) \Rightarrow \max \tag{12}$$

$$z_j = f(x_1, x_2, ..., x_l) \Rightarrow \min \tag{13}$$

and the constraints

$$x_1 \in [x^{j^1}_{1,\alpha_i l}; x^{j^1}_{1,\alpha_i r}] \tag{14}$$

$$x_2 \in [x^{j^2}_{2,\alpha_i l}; x^{j^2}_{2,\alpha_i r}] \tag{15}$$

$$\vdots$$

$$x_l \in [x^{j^l}_{l,\alpha_i l}; x^{j^l}_{l,\alpha_i r}] \tag{16}$$

has to be solved for all possible combinations $j^1, j^2, ..., j^l$ with $j^k = 1, 2, ..., e_{\tilde{x}_k,\alpha_i}$ and $k = 1, 2, ..., l$. That is, maximum and minimum for each of the $\prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$ l-dimensional hypercuboids has to be located.

For each of the $\prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$ objective functions according Eq. (12) a local maximum $z_j = h_{j,p}$ ($p = 1, 2, ..., \prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$) and for each of the $\prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$ objective functions according Eq. (13) a local minimum $z_j = t_{j,p}$ is given by the constraints according Eqs. (14) till (16) in the subspace $\underline{X}_{\alpha_i}$. The major local maximum correspond to the global maximum within the subspace $\underline{X}_{\alpha_i}$ and the minor local minimum correspond to the global minimum.

By means of the local maxima $h_{j,p}$ and the local minima $t_{j,p}$ the computation of the interval boundaries $z^{s^j}_{j,\alpha_i l}$ and $z^{s^j}_{j,\alpha_i r}$, $s^j = 1, 2, ..., e_{\tilde{z}_j,\alpha_i}$, succeeds in the second step. The local maxima $h_{j,p}$ and minima $t_{j,p}$ are sorted and indexed from the smallest to the largest value according to the following inequalities.

$$h_{j,1} \leq h_{j,2} \leq ... \leq h_{j,num} \quad \text{with} \quad num = \prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i} \tag{17}$$

$$t_{j,1} \leq t_{j,2} \leq ... \leq t_{j,num} \quad \text{with} \quad num = \prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i} \tag{18}$$

For each α-level set Z_{j,α_i} of the fuzzy result variable $\tilde{z}_j$ the number $e_{\tilde{z}_j,\alpha_i}$ of intervals $[z^{s^j}_{j,\alpha_i l}; z^{s^j}_{j,\alpha_i r}]$ is then given by Eq. (19).

$$e_{\tilde{z}_j,\alpha_i} = \#\left\{s \,|\, h_{j,s} < t_{j,s+1},\, s = 1, 2, ..., num\right\} + 1 \tag{19}$$

At this $\#\{\cdot\}$ is the number of values s, for which the condition $h_{j,s} < t_{j,s+1}$, $s = 1, 2, ..., num$, is fulfilled.

The local maxima and local minima located according Eqs. (12) and (13) do not yield directly discrete points of the membership function $\mu_{\tilde{z}_j}(z_j)$. Each of the located $\prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$ maxima and each corresponding minimum define an interval, in which the fuzzy result variable $\tilde{z}_j$ may take a value on the α-level α_i. For this reason it is to check, if the located $\prod_{k=1}^{l} e_{\tilde{x}_k,\alpha_i}$ intervals overlap each other. In the case that an interval does not overlap any other interval, the corresponding local maximum and minimum are discrete points of the membership function $\mu_{\tilde{z}_j}(z_j)$. That is, they are directly interval boundaries of the α-level set Z_{j,α_i}. In the case of overlapping intervals, the intervals are combined to one enveloping interval and only the minor minimum and the major maximum of the overlapping intervals are discrete points of the membership function $\mu_{\tilde{z}_j}(z_j)$. The computation rule for the interval boundaries $z^{s^j}_{j,\alpha_i l}$ and $z^{s^j}_{j,\alpha_i r}$ of the fuzzy result variables $\tilde{z}_j$ is thus given recursively by Eqs. (20) till (24).

$$z^1_{j,\alpha_i l} = t_{j,1} \tag{20}$$

$$z^1_{j,\alpha_i r} = \min\left[h_{j,s} \,|\, h_{j,s} < t_{j,s+1},\, s = 1, 2, ..., num\right] \tag{21}$$

$$z^2_{j,\alpha_i l} = \min\left[t_{j,s} \,|\, t_{j,s} > z^1_{j,\alpha_i r},\, s = 1, 2, ..., num\right] \tag{22}$$

$$z^2_{j,\alpha_i r} = \min\left[h_{j,s} \,|\, z^2_{j,\alpha_i l} < h_{j,s} < t_{j,s+1},\, s = 1, 2, ..., num\right] \tag{23}$$

$$\vdots$$

$$z^{e_{\tilde{z}_j,\alpha_i}}_{j,\alpha_i r} = h_{j,num} \tag{24}$$

By means of the interval boundaries $z^{s^j}_{j,\alpha_i l}$ and $z^{s^j}_{j,\alpha_i r}$, $s^j = 1, 2, ..., e_{\tilde{z}_j,\alpha_i}$, the membership functions $\mu_{\tilde{z}_j}(z_j)$ of the fuzzy result variables $\tilde{z}_j$ are given discretely.

4 Example

Fuzzy structural analysis under consideration of non-convex fuzzy variables is demonstrated by the example of a two bay steel structure with a crane runway. The crane runway is used by two independent cranes with a lifting force of 320 kN. The steel structure without crane runway is shown in Fig. 3.

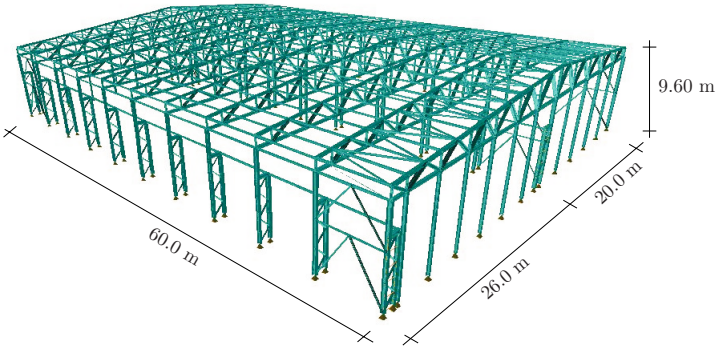


Fig. 3. Two bay steel structure

In the context of the example the sideways horizontal impact forces of the crane runway on the girder are regarded. The horizontal forces complying with the standard base on a series of simplifications and are thus not a comprehensive reflection of reality. On this account the horizontal forces $H_{S,1}$ and $H_{S,2}$ of the two cranes are modeled as fuzzy variables. The forces complying with the standard are chosen as the 'best possible crisped' impact values ($\mu = 1$). Zero impact forces are valuated by membership value zero. Further possible forces are taken into account by membership values less than one. Fig. 4 shows the membership function of the resulting non-convex fuzzy variables $\tilde{H}_{S,1}$ and $\tilde{H}_{S,2}$.

The deterministic fundamental solution within the fuzzy structural analysis is the program system SSt-micro. The modeling is carried out using 1494 bars and 638 nodes. In the following the normal force at the upper edge of bar 738 is analyzed. As worst loading case a position of the cranes in the middle of the girder with a distance in-between of 6 m is chosen.

The non-convex fuzzy input variables $\tilde{H}_{S,1}$ and $\tilde{H}_{S,2}$ are discretized by means of $n = 4$ α-levels $\alpha_1 = 0$, $\alpha_2 = \frac{1}{3}$, $\alpha_3 = \frac{2}{3}$ and $\alpha_4 = 1$ according to Sect. 2. As a result of the fuzzy structural analysis the normal force is obtained as a non-convex fuzzy variable

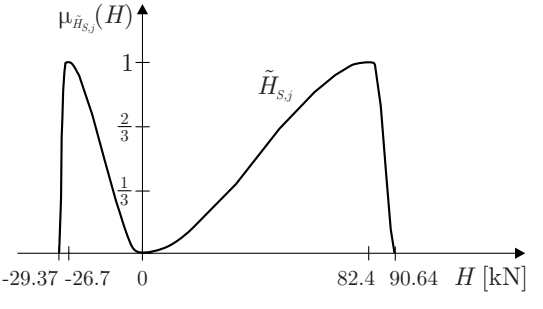


Fig. 4. Non-convex fuzzy impact forces $\tilde{H}_{S,j}$ $(j = 1, 2)$

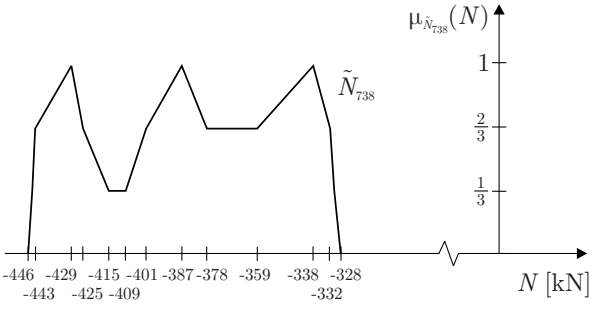


Fig. 5. Non-convex fuzzy normal force $\tilde{N}_{738}$

$\tilde{N}_{738}$ (see Fig. 5). The three peak points of the membership function represent the normal force corresponding to the combinations of the impact forces complying with the standard. Because of the symmetry of the system there are only three and not four peak points as expected. Analysis of the α-level $\alpha_1 = 0$ yields approximately four percent difference in comparison to the results complying with the standard. Furthermore the effect of overlapping intervals occurs on α-level $\alpha_2 = \frac{1}{3}$. In other words, non-continuous hypercuboids in the input space yield overlapping intervals in the result space.

5 Conclusion

In this paper a new approach for fuzzy structural analysis under consideration of non-convex fuzzy variables is presented. Therefor an enhanced discretization of non-convex fuzzy variables is introduced. Non-convex fuzzy variables are characterized by a family of α-level sets which are sets of closed finite intervals. Within the framework of fuzzy structural analysis the α-level optimization is generalized. Fuzzy structural analysis under consideration of non-convex fuzzy variables is demonstrated by way of a two bay steel structure.

References

1. Helton, J.C., Burmaster, D.E.: Special issue on the treatment of aleatory and epistemic uncertainty. Reliab. Eng. System Saf. 54(2–3), 91–262 (1996)
2. Möller, B., Reuter, U.: Uncertainty Forecasting in Engineering. Springer, Heidelberg (2007)
3. Möller, B., Graf, W., Beer, M.: Fuzzy structural analysis using α-level optimization. Comput. Mech. 26(6), 547–565 (2000)
4. Viertl, R.: Statistical Methods for Non-Precise Data. CRC Press, Boca Raton (1996)
5. Wood, K.L., Otto, K.N., Antonsson, E.K.: Engineering design calculations with fuzzy parameters. Fuzzy Sets Syst. 52, 1–20 (1992)
6. Zadeh, L.A.: Fuzzy sets. Information & Control 8, 338–353 (1965)
7. Zimmermann, H.J.: Fuzzy Set Theory and Its Applications. Kluwer Academic Publishers, Boston (1992)

Handling Uncertainty in Higher Dimensions with Potential Clouds towards Robust Design Optimization

Martin Fuchs and Arnold Neumaier

Faculty of Mathematics, University of Vienna, Wien, Austria

Abstract. Robust design optimization methods applied to real life problems face some major difficulties: how to deal with the estimation of probability densities when data are sparse, how to cope with high dimensional problems and how to use valuable information provided in the form of unformalized expert knowledge. We introduce the *clouds* formalism as means to process available uncertainty information reliably, even if limited in amount and possibly lacking a formal description. We provide a worst-case analysis with confidence regions of relevant scenarios which can be involved in an optimization problem formulation for robust design.

Keywords: Clouds, Potential clouds, Robust design, Design optimization, Confidence regions, Uncertainty modeling.

1 Background

Robust design is the art of safeguarding reliably against uncertain perturbations while seeking an optimal design point. In every design process an engineer faces the task to qualify the object he has designed to be robust. That means the design should not only satisfy given requirements on functionalities, but should also work under uncertain, adverse conditions that may show up during employment of the designed object.

Hence the process of robust design demands both the search of an optimal design with respect to a given design objective, and an appropriate method of handling uncertainties. In particular for early design phases, it is frequent engineering practice to assign and refine intervals or safety margins to the uncertain variables. These intervals or safety margins are propagated within the whole optimization process. Thus the design arising from this process is supposed to include robustness intrinsically. Note that the assessment of robustness is exclusively based on expert knowledge of the engineers who assign and refine the intervals. There is no quantification of reliability, no rigorous worst-case analysis involved.

Several methods exist to approach reliability quantification from a rigorous mathematical background, originating from classical probability theory, statistics, or fuzzy theory. However, real life applications of many methods disclose various problems. One of the most prominent is probably the fact that the dimension of many uncertain real life scenarios is very high. This can cause severe computational effort, even given the complete knowledge about the multivariate probability distributions of the uncertainties, also famous as the curse of dimensionality [8]. Often standard simulation techniques

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 376–382, 2008.
springerlink.com

are used to tackle the dimensionality issue, as the computational effort they require seems to be independent of the dimension. Advancements have been made based on sensitivity analysis [15], or on α-level optimization, cf. [11]. Moreover, if the amount of available uncertainty information is very limited, well-known current methods like Monte Carlo either do not apply at all, or are endangered to critically underestimate error probabilities, [2]. A simplification of the uncertainty model, e.g., a reduction of the problem to an interval analysis after assigning intervals to the uncertainties as described before (e.g., so called 3 σ boxes), entails a loss of valuable information which would actually be available, maybe only unformalized, but not at all considered in the uncertainty model. Incomplete information supplemented by expert statements can be handled with different methods, e.g.: p-boxes [3], fuzzy sets [1], random sets [10]. A combination of uncertainty methods and design optimization is addressed in approaches to reliability based design optimization: based on reliability methods [7]; based on possibility theory in [12]; based on evidence theory in [13].

This paper is organized as follows. We will start introducing our approach based on the clouds formalism and lead to the the special case of interest in this study, cf. Section 2.1, the concept of potential clouds. Several remarks on suitable potential function choices are given in Section 2.2. A short introduction how clouds can be involved in an optimization problem formulation for robust design can be studied in Section 2.3. Section 3 concludes our studies.

2 Introducing the New Approach

Our work deals with a new approach based on the *clouds* formalism, cf. [14]. Clouds can process limited amounts of stochastic information in an understandable and computationally attractive way, even in higher dimensions, in order to perform a reliable worst-case analysis, reasonably safeguarded against perturbations that result from unmodeled or unavailable information. Since the strength of our new methodology lies especially in the application to real life problems with a very limited amount of uncertainty information available, we focus in particular on problem statements arising in early design phases where today's methods handling limited information are very immature. On the one hand, the information is usually available as bounds or marginal probability distributions on the uncertain variables, without any formal correlation information. On the other hand, unformalized expert knowledge will be captured to improve the uncertainty model adaptively by adding dependency constraints to exclude scenarios deemed irrelevant. The information can also be provided as real sample data, if available.

If we have a look at Figure 1, we see confidence levels on some two dimensional random variable ε. The curves displayed can be considered to be level sets of a function $V(\varepsilon) : \mathbb{R}^2 \to \mathbb{R}$, called the potential. The potential characterizes confidence regions $C_\alpha := \{\varepsilon \in \mathbb{R}^2 \mid V(\varepsilon) \leq V_\alpha\}$, where V_α is determined by the condition $\Pr(\varepsilon \in C_\alpha) = \alpha$. If the probability information is not precisely known nested regions are generated

$$\underline{C}_\alpha := \{\varepsilon \in \mathbb{R}^2 \mid V(\varepsilon) \leq \underline{V}_\alpha\},$$

where $\underline{V}_\alpha$ is largest such that $\Pr(\varepsilon \in \underline{C}_\alpha) \leq \alpha$, and

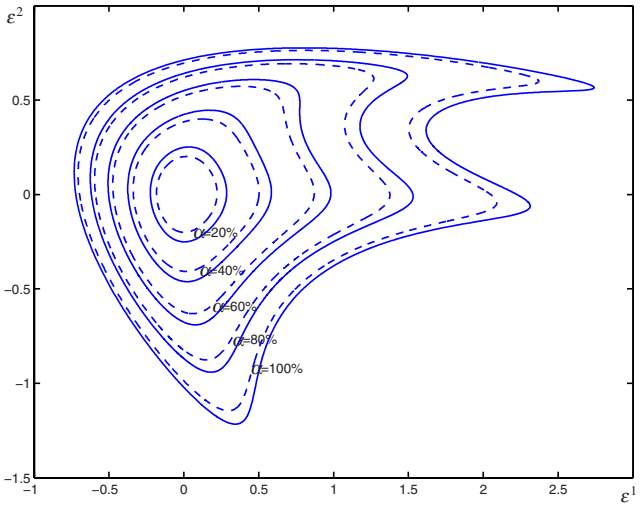


Fig. 1. Nested confidence regions in two dimensions for confidence levels $\alpha = 0.2, 0.4, 0.6, 0.8, 1$. The lower confidence regions $\underline{C}_\alpha$ plotted with dashed lines, the upper confidence regions $\overline{C}_\alpha$ with solid lines.

$$\overline{C}_\alpha := \{\varepsilon \in \mathbb{R}^2 \mid V(\varepsilon) \leq \overline{V}_\alpha\},$$

where $\overline{V}_\alpha$ is smallest such that $\Pr(\varepsilon \in \overline{C}_\alpha) \geq \alpha$. The information in $\underline{C}_\alpha$ and $\overline{C}_\alpha$ is called a *potential cloud*. The values $V_\alpha, \underline{V}_\alpha, \overline{V}_\alpha$ can be found from the cumulative distribution function (CDF) of $V(\varepsilon)$ and lower and upper bounds of it. These bounds in turn can be determined empirically using the Kolmogoroff-Smirnov (KS) distribution [9].

2.1 Potential Cloud Generation

We assume that the uncertainty information consists of given samples, boxes, non-formalized dependency constraints or continuous marginal CDFs $F_i, i \in I \subseteq \{1, 2, \ldots, n\}$, on the n-dimensional vector of uncertainties ε, without any formal knowledge about correlations or joint distributions. In case there is no sample provided or the given sample is very small, a sample S has to be generated. For these cases we first use a Latin Hypercube Sampling (LHS) inspired method to generate S, i.e., the sample points $x_1, x_2, \ldots, x_{N_S}$ are chosen from a grid satisfying $x_i^j \neq x_k^j \; \forall i, k \in \{1, 2, \ldots, N_S\}, k \neq i, \forall j \in \{1, 2, \ldots, n\}$, where x_i^j is the projection of x_i to the j^{th} coordinate. If only boxes for ε are given, then the grid is equidistant, if marginal distributions are given the grid is transformed with respect to them to ensure that each grid interval has the same marginal probability. LHS introduces some preference for a simple structure. The effect of this preference will be diminished by weighting of the sample points.

The generated sample represents the marginal distributions. However after a modification of S, e.g., by cutting off sample points as we will do later, an assignment of weights to the sample points is necessary to preserve the marginal CDFs. In

order to do so the weights $\omega_1, \omega_2, \ldots, \omega_{N_S} \in [0,1]$, corresponding to the sample points $x_1, x_2, \ldots, x_{N_S}$, are required to satisfy the following conditions (11)

$$\sum_{j=1}^{k} \omega_{\pi_i(j)} \in [F_i(x^i_{\pi_i(k)}) - d, F_i(x^i_{\pi_i(k)}) + d], \qquad \sum_{k=1}^{N_S} \omega_k = 1. \tag{1}$$

for all $i \in I$, $k = 1, \ldots, N_S$, where π_j is a sorting permutation of $\{1, \ldots, N_S\}$, such that $x^j_{\pi_k(1)} \leq x^j_{\pi_k(2)} \leq \cdots \leq x^j_{\pi_k(N_S)}$, and I the index set of those entries of the uncertainty vector ε where a marginal CDF $F_i, i \in I$ is given. The constraints (11) require the weights to represent the marginal CDFs with some reasonable margin d. In practice, one chooses d with KS statistics.

We determine bounds on the CDF of $V(\varepsilon)$ by $\overline{F} := \min(\widetilde{F} + D, 1)$ and $\underline{F} := \max(\widetilde{F} - D, 0)$, where $\widetilde{F}(\xi) := \sum_{\{j | V(x_j) \leq \xi\}} \omega_j$ the weighted empirical distribution for $V(\varepsilon)$, and D is again chosen with KS statistics. Finally we fit the two step functions $\underline{F}$, $\overline{F}$ to smooth, monotone lower bounds $\underline{\alpha}$ and upper bounds $\overline{\alpha}$. From these bounds the regions $\underline{C}_\alpha, \overline{C}_\alpha$ can be computed straightforward.

Lower and upper bounds of empirical CDFs remind of p-boxes. In fact a potential cloud can be considered as a p-box on the potential of a random vector. Clouds extend the p-box concept to the multivariate case without the exponential growth of work in the conventional p-box approach. Furthermore, clouds can be considered as fuzzy sets with interval valued membership function or as a special case of random sets.

2.2 Choice of the Potential

We see that given a potential the corresponding potential cloud is easy to estimate, even for high dimensional data. The choice of the potential is dictated by the shape of the points set defined by the sample of available ε. We are looking for a way to find a good choice of V that gives the possibility to improve the potential iteratively and allows for a simple computational realization of the confidence regions, e.g., by linear constraints. This leads us to the investigation of polyhedron-shaped potentials. A polyhedron potential centered at $m \in \mathbb{R}^n$ can be defined as:

$$V_p(\varepsilon) := \max_k \frac{(A(\varepsilon - m))^k}{b^k}, \tag{2}$$

where $\varepsilon, b \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, $(A(\varepsilon - m))^k, b^k$ the k^{th} component of the vectors $A(\varepsilon - m)$ and b, respectively.

But how to achieve a polyhedron that reflects the given uncertainty information in the best way? As mentioned we assume the uncertainty information to consist of given samples, boxes or marginal distributions, and unformalized dependency constraints. After providing a sample S as described in Section 2.1 we define a box b_0 containing 100% of the sample points, and we define our potential $V_0(\varepsilon)$ box-shaped taking the value 1 on the margin of b_0. Based on expert knowledge, a user-defined variation of V_0 can be performed afterwards by cutting off sample points deemed irrelevant for the worst-case, cf. Figure 2. The optimization phase (cf. Section 2.3) provides a worst-case scenario which is highlighted in the graphical user interface. The expert can decide to exclude, e.g., the

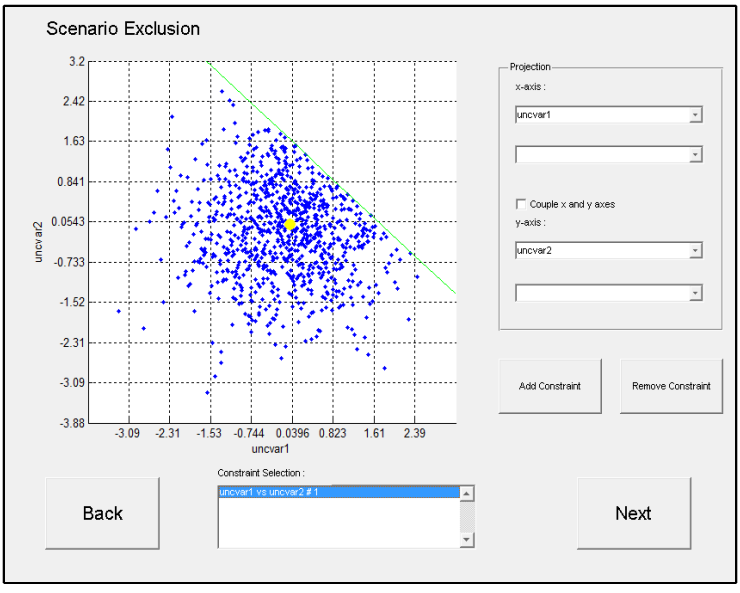


Fig. 2. Graphical user interface for an interactive scenario exclusion. The exclusion is performed in 1 and 2 dimensional projections.

worst-case or different scenarios, based on his technical knowledge. Thus an expert can specify the uncertainty information in the form of dependency constraints adaptively, even if the expert knowledge is only little formalized, resulting in a polyhedron shaped potential.

Assume the linear constraints $A(\varepsilon - \mu) \leq b$ represent the exclusion of sample points and the box constraint from b_0, we define our polyhedron shaped potential as in (2) with $m = \mu$.

Further details on the construction of potential clouds and the choice of the potential function can be studied in [5].

2.3 Robust Design Optimization Problem

In this section we give a short introduction how potential clouds can be involved in an optimization problem formulation for robust design. Provided an underlying model of a given structure to be designed, with an objective function $g(z)$ and input vector z. Let θ be a design point, i.e., it fully defines the design. Let T be the set of all allowed designs. The input variables z consist of design variables which depend on the design θ, e.g., the thrust of a thruster, and external inputs with a nominal value that cannot be controlled for the underlying model, e.g., a specific temperature. Let $Z(\theta)$ be the input vector given θ and given the external inputs at their nominal values.

The input variables are affected by uncertainties. Let ε denote the related vector of uncertain errors. One can formulate the optimization problem as a mixed-integer, bi-level problem of the following form:

$$\begin{aligned} &\min_{\theta} \max_{\varepsilon} g(z) && && \text{(objective function)} \\ &\text{s.t.} && z = Z(\theta) + \varepsilon && \text{(input constraints)} \\ &&& \theta \in T && \text{(selection constraints)} \\ &&& V_p(\varepsilon) \leq \underline{V}_\alpha && \text{(cloud constraint)} \end{aligned} \tag{3}$$

The cloud constraint involves confidence regions as level sets of the potential function $V = V_p(\varepsilon)$ as described previously. The confidence level α should be chosen to reflect the seriousness of consequences of the worst case event. In our applications from spacecraft system design we used $\alpha = 0.95$, cf. [6]. The cloud constraint models the embedding of the uncertainty methods into the optimization phase.

A detailed view on the optimization techniques used to solve (3) is given in [4].

3 Summary

We present a new methodology based on clouds to provide confidence regions for safety constraints in robust design optimization. We can process the uncertainty information from expert knowledge towards a reliable worst-case analysis, even if the information is limited in amount and high dimensional.

We can summarize the basic concept of our methodology in three essential steps within an iterative framework. First, the expert provides the underlying system model, given as a black-box model, and all a priori available information on the input variables of the model. Second, the information is processed to generate a potential cloud thus producing safety constraints for the optimization. Third, optimization methods minimize a certain objective function subject to the functional constraints which are represented by the system model, and subject to the safety constraints from the cloud. The results of the optimization are returned to the expert, who is given an interactive possibility to provide additional information a posteriori and to rerun the procedure, adaptively improving the uncertainty model.

The adaptive nature of our uncertainty model, i.e., the possibility of manually adding dependency constraints, is one of the key features. The iteration steps significantly improve the uncertainty information and we are able to process the new information to an improved uncertainty model.

All in all, the presented approach offers an attractive novel point of view on high dimensional uncertainty handling and its involvement to robust design.

References

1. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1986)
2. Ferson, S.: What monte carlo methods cannot do. Hum. Ecol. Risk Assess 2, 990–1007 (1996)

3. Ferson, S.: Ramas Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers. Lewis Publishers, Boca Raton, Florida (2002)
4. Fuchs, M., Neumaier, A.: Autonomous robust design optimization with potential clouds (submitted for publication 2008), http://www.martin-fuchs.net/publications.php
5. Fuchs, M., Neumaier, A.: Potential based clouds in robust design optimization. Special Issue on Imprecision of J. Stat. Theory Appl. (in press, 2008), http://www.martin-fuchs.net/publications.php
6. Fuchs, M., Girimonte, D., Izzo, D., Neumaier, A.: Robust and automated space system design. In: Robust Intelligent Systems (in press, 2008), http://www.martin-fuchs.net/publications.php
7. Kaymaz, I., Marti, K.: Reliability-based design optimization for elastoplastic mechanical structures. Comput. & Structures 85(10), 615–625 (2007)
8. Koch, P.N., Simpson, T.W., Allen, J.K., Mistree, F.: Statistical approximations for multidisciplinary optimization: The problem of size. Special Issue on Multidisciplinary Design Optimization of J. Aircr. 36(1), 275–286 (1999)
9. Kolmogoroff, A.: Confidence limits for an unknown distribution function. Ann. Math. Statist. 12(4), 461–463 (1941)
10. Molchanov, I.: Theory of Random Sets. Springer, London (2005)
11. Möller, B., Beer, M.: Fuzzy Randomness: Uncertainty in Civil Engineering and Computational Mechanics. Springer, Berlin (2004)
12. Mourelatos, Z., Zhou, J.: Reliability estimation and design with insufficient data based on possibility theory. AIAA J. 43(8), 1696–1705 (2005)
13. Mourelatos, Z., Zhou, J.: A design optimization method using evidence theory. J. Mech. Des. 128(4), 901–908 (2006)
14. Neumaier, A.: Clouds, fuzzy sets and probability intervals. Reliab. Comput. 10(4), 249–272 (2004), http://www.mat.univie.ac.at/~neum/ms/cloud.pdf
15. Oberguggenberger, M., King, J., Schmelzer, B.: Imprecise probability methods for sensitivity analysis in engineering. In: De Cooman, G., Vejnarova, J., Zaffalon, M. (eds.) Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2007, Prague, Czech Republic), pp. 317–326. SIPTA, Prague (2007)

Reliability of Structures under Consideration of Uncertain Time-Dependent Material Behaviour

Steffen Freitag, Wolfgang Graf, Stephan Pannier, and Jan-Uwe Sickert

Institute for Structural Analysis, Technische Universität Dresden, Dresden, Germany

Abstract. In this paper a concept for time-dependent reliability assessment of civil engineering structures is presented. This concept bases on the uncertainty model fuzzy randomness. The time-dependent behaviour of materials with fading memory is modelled with the aid of rheological elements using uncertain fractional time derivatives of strain. The presented method is applied to the reliability assessment of a pavement construction.

1 Introduction

Generally, material behaviour, geometry, and stresses of civil engineering structures are time-dependent, which results in a time-dependent reliability. Here we focus on the effect of uncertain time-dependent material behaviour. Realistic computation of structural reliability requires mapping models specifying the nonlinear time-dependent structural behaviour appropriately. The structural response during a load period may consist of instant elastic, instant plastic, viscoelastic and viscoplastic deformations. These phenomena have to be incorporated in a time-dependent structural analysis. Rheological models [3, 10], which are based mathematically on time derivatives, enable the description of the time-dependent deformation behaviour of particular engineering materials, e.g. asphalt, elastomer or textile reinforced concrete. They are characterized by their susceptibility of loading rate and loading history, which is modelled with the aid of fractional time derivatives in the paper.

Because of material tests yields uncertain results, a conventional rheological formulation is enhanced by uncertain time derivatives of real order and uncertain parameters, see Sect. 2. A respective assessment of structural reliability requires appropriate uncertainty models. Currently, stochastic models are applied in engineering practice for uncertainty specification and reliability assessment, see e.g. [11]. Here, we introduce a method for reliability assessment considering structural long-term behaviour in conjunction with uncertain parameters. The uncertain parameters are assessed with the aid of imprecise probabilities [8, 16]. This model is appropriate to consider the uncertainty, vagueness, and ambiguity of the stochastic model specification [7, 9]. The time-dependent input parameters are modelled as fuzzy random processes, random processes or fuzzy processes. The establishment of fuzzy random processes within the fuzzy safety concept is enabled with the aid of the uncertainty measure fuzzy failure probability, outlined in Sect. 4. The fuzzy failure probability is computed by means of the numerical algorithm Fuzzy Monte-Carlo Simulation (FMCS) [9, 13].

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 383–390, 2008.
springerlink.com

In Sect. 5 the algorithms are applied for the assessment of the time-dependent reliability of a pavement construction. The long-term material behaviour of the asphalt layer is modelled by an uncertain fractional rheological element.

2 Fractional Rheological Models and Uncertainty

Conventional rheological formulations contain time derivatives of integer order [3], whereas fractional rheological formulations base on time derivatives of real order [18]. The extension of conventional rheological models to fractional rheological models facilitates an improved fitting of material parameters to experimental data. Moreover, the theory of fractional derivatives is well-founded in the physics of material behaviour [18]. Conventional rheological formulations can be interpreted as a subclass of fractional formulations.

As an example, the fuzzy stochastic fractional NEWTON body is introduced to model the long-term behaviour of materials with fading memory, see e.g. [10]. The differential equation of the fractional NEWTON body is

$$\sigma_v(\varepsilon, \tau, \alpha, p) = p(\sigma_v) \frac{\mathrm{d}^\alpha}{\mathrm{d}\tau^\alpha} \varepsilon(\tau) \,. \tag{1}$$

In terms of rheology $\sigma_v(\varepsilon, \tau, \alpha, p)$ could be interpreted as a stress, which is a function of the stress dependent parameter $p(\sigma_v)$ and the fractional derivative of the strain $\varepsilon(\tau)$ with respect to the time τ. The operator α represents the order of the strain derivative, which can adopt real values between 0 and 1. For $\alpha = 1$ an ideal dashpot is described with the viscosity $p(\sigma_v)$. Against this, the fractional NEWTON body represents an elastic spring for $\alpha = 0$ with the elasticity $p(\sigma_v)$. The operator α is assumed to be constant in time for further remarks.

The differential equation (1) may be solved by a LAPLACE transform [1]. Regarding to the creep and relaxation behaviour of the material the LAPLACE transform of Eq. (1) requires the definition of stress or strain boundary conditions. Here, the stress boundary condition $\sigma_v(\varepsilon, \tau, \alpha, p) = \sigma_v^*$ is selected. Under consideration of this boundary condition the creep strain $\varepsilon(\tau)$ is obtained for a time-constant stress σ_v^*.

$$\varepsilon(\tau) = \frac{\sigma_v^* \, \tau^\alpha}{p(\sigma_v^*) \, \Gamma(\alpha + 1)} \tag{2}$$

Thereby, the Gamma function $\Gamma(\alpha + 1)$ is an extension of the factorial function to real numbers.

However, in many engineering applications the stress is time-variant. This results from time-dependent structural loads as well as load rearrangements in layered or composite constructions caused by different time-dependent material behaviour. In order to account for arbitrary loading regimes a convolution of the creep function (2) can be conducted, see [10].

The determination of the strain at time τ requires knowledge of the entire history of loading (i.e. stresses). Thereby, changes of stress, occurred in the past, are less influential on the actual strain state than stress changes occurred quite recently. This material

property is usually denoted as fading memory. In [10] this property is considered by means of the convolution of approximated creep functions. Therewith, an efficient solution with internal variables is provided.

In general, the long-term material behaviour of structural members is affected by uncertainty because of the fact that quasi identical experiments yield different results. This uncertainty may be reflected in the parameters of the fractional rheological model. Thereby, the uncertainty results from both aleatory and epistemic sources. Thus, the parameter $p(\sigma_v^*)$ of Eq. (2) is modelled as fuzzy random function $\tilde{p}(\sigma_v^*)$. That means, for a given stress σ_v^* the functional value $\tilde{p}(\sigma_v^*)$ is a fuzzy random variable according to the definition given in Sect. 3. Furthermore, the operator α of Eq. (2) is described by means of a fuzzy quantity $\tilde{\alpha}$.

3 Numerical Analysis Considering Fuzzy Random Functions

The uncertainty of the long-term material behaviour may be described appropriately by means of the generalized uncertainty model fuzzy randomness [7], which bases on the uncertain measure fuzzy probability as a special form of the imprecise probability [8]. Therewith, fractional rheological models may be extended modelling the parameters by means of fuzzy random variables. Different definitions of fuzzy random variables, e.g. given in [6] and [12], are summarized in [5]. The theory of fuzzy stochastic processes and fuzzy random functions respectively is contained in [17].

Based on these definitions fuzzy probability density functions $\tilde{f}(x)$, introduced in [7], represents a fuzzy set of probability density function $f(x)$. Therefore, it is referred to as assessed bunch of functions $f(x)$. The bunch is described by means of fuzzy bunch parameters $\tilde{s}$. If the bunch depends on more then one $\tilde{s}$, all n bunch parameters are joined in the vector $\underline{\tilde{s}}$, which represents a vector of fuzzy quantities. The fuzzy probability density function $\tilde{f}(x)$ results therewith in a function $f(\tilde{s}, x)$.

Typical fuzzy bunch parameters in engineering applications are the functional parameters of $\tilde{f}(x)$. For instance, a GUMBEL distributed fuzzy random variable, that means each $f(x) \in \tilde{f}(x)$ is GUMBEL distributed, may depend on the fuzzy bunch parameters $\tilde{s}_1 = \tilde{a}$ and $\tilde{s}_2 = \tilde{b}$. Then the fuzzy probability density function is

$$f(\underline{\tilde{s}}, x) = \tilde{s}_1 \exp\left(-\tilde{s}_1 (x - \tilde{s}_2) - \exp\left(-\tilde{s}_1 (x - \tilde{s}_2)\right)\right) . \tag{3}$$

This approach is different to the approaches introduced e.g. in [2, 4] where only the expected value can be fuzzified. However, regarding the engineering application the bunch parameter representation introduced in [7] is appropriate. Further, the approach conforms to the definitions given in [5] as shown in [7]. On the basis of the fuzzy bunch parameter representation the numerical solution of a fuzzy stochastic analysis may be formulated.

A fuzzy stochastic analysis aims at the mapping of fuzzy random variables $\tilde{X}$ on fuzzy random variables $\tilde{Z}$ according to

$$M_{FSA} : \underline{\tilde{X}} \rightarrow \underline{\tilde{Z}} . \tag{4}$$

With the aid of the bunch parameter representation, Eq. (4) is transformed into the mapping

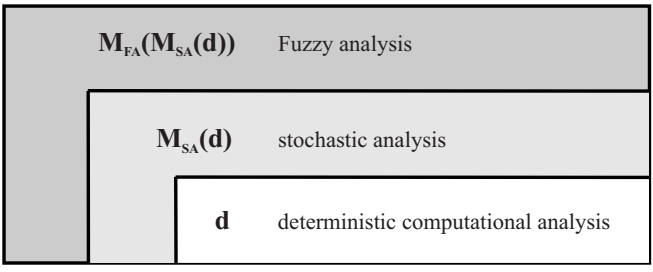


Fig. 1. Fuzzy stochastic analysis M_{FSA}

$$\underline{\tilde{\sigma}} = (\tilde{\sigma}_1, \ldots, \tilde{\sigma}_j, \ldots, \tilde{\sigma}_{m_1}) = m(\tilde{s}_1, \ldots, \tilde{s}_k, \ldots, \tilde{s}_{n_1}) \tag{5}$$

Applying α-discretization to the fuzzy bunch parameter, an optimization problem is solved in order to determine the α-level sets of the fuzzy bunch parameters $(\tilde{\sigma}_1, \tilde{\sigma}_2, ..., \tilde{\sigma}_{m_1})$. This algorithm is referred to as fuzzy analysis and described, e.g., in [7]. It replaces the solution of the extension principle. Each element of the input α-level sets yields a stochastic analysis. Within the stochastic analysis, e.g., applying the Monte Carlo simulation, a deterministic fundamental solution $d(.)$ is processed repeatedly. Therewith a three-loop computational algorithm is constituted, see Fig. 1, also referred to as Fuzzy Monte-Carlo Simulation (FMCS).

4 Time-Dependent Reliability under Consideration of Fuzzy Randomness

The reliability of structures alters during the lifetime. Time-dependent statical and dynamic loads, environmental conditions, and long-term material behaviour (e.g. according to Sect. 2) result in time-dependent structural responses. The consideration of these time-dependencies within the framework of reliability analysis requires a computational model with time-dependent uncertain parameters. Here, these parameters are described as fuzzy random processes, utilizing the generalized uncertainty model fuzzy randomness. A fuzzy random process $\tilde{X}(\tau)$ is a family of fuzzy random variables according to [17].

If input parameters are modelled as fuzzy random processes $\tilde{X}(\tau)$, the reliability assessment yields uncertain and time-dependent results. In [9, 13] the time-dependent uncertainty measure fuzzy failure probability $\tilde{P}_f(\tau)$ is introduced to process the influence of subjective evaluation in the safety assessment. Thereby, the fuzzy failure probability $\tilde{P}_f(\tau)$ represents a fuzzy set of failure probabilities $P_f(\tau)$ which are real-valued functions.

Generally, the structural reliability is influenced by nonlinear structural behaviour. Thus, the structural resistance depends on the stress process. Then, the fuzzy failure probability $P_f(\tau)$ at time points τ has to be computed in the space $\underline{\mathbf{X}}$ of the fuzzy random input parameters, such as material, geometry, and load parameters. In view of the numerical solution, the space is constituted by means of the one-dimensional

fuzzy random variables $\tilde{X}$ obtained by the discretization of all fuzzy random processes $\tilde{X}(\tau)$. Additionally, real random variables X, e.g., as result of discretized real random processes $X(\tau)$, may be accounted for as special case of fuzzy random variables.

The space of fuzzy random variables is subdivided into a fuzzy survival domain $\underline{\tilde{X}}_S$ and a fuzzy failure domain $\underline{\tilde{X}}_F$ by the fuzzy set of limit state equations $g(\underline{x}) = 0$. Thereby, $g(\underline{x})$ represents the performance function of a structure, e.g., $g(\underline{x}) = R(\underline{x}) - S(\underline{x})$ with $R(\underline{x})$ structural resistance and $S(\underline{x})$ stresses. The fuzzy set is modelled by means of fuzzy bunch parameters $\tilde{s}_g$ joined in $\underline{\tilde{s}}_g$. For each $\underline{s}_g \in \underline{\tilde{s}}_g$, the inequality $g_s(s_g, \underline{x}) \leq 0$ represents the failure domain $\underline{X}_F(\underline{s}_g)$, where the limit state is exceeded.

For each fuzzy random variable $\tilde{X}$, which results from discretization in time, the assigned fuzzy probability density function $f(\tilde{s}, x)$ has to be known. The joint behaviour may be described with the aid of the multivariate probability density function $f(\underline{\tilde{s}}, \underline{x})$ in $\underline{\mathbf{X}}$ considering fuzzy correlation. For each certain vector $\underline{s} \in \underline{\tilde{s}}$ the failure probability is computed by

$$P_f = \int\limits_{\underline{x}|g(\underline{s}_g, \underline{x}) \leq 0} f(\underline{s}, \underline{x})\, d\underline{x}. \tag{6}$$

The fuzzy failure probability is computed with the fuzzy stochastic analysis, see Fig. 1. Thereby, Eq. (6) represents the stochastic analysis and $P_f = \sigma$ an element of the fuzzy result $\tilde{\sigma} = \tilde{P}_f$.

5 Reliability Assessment of a Pavement Construction

The FMCS, see Sect. 3, is applied for reliability assessment of a layered pavement construction with a width of 6.6 m. The pavement construction consists of a 0.30 m asphalt layer, a 0.15 m unbonded sub-base layer, and a 0.45 m frost protection layer. Figure 2 displays the discretized structure of the pavement for the Finite Element Method with 20-node displacement elements taking into account the system symmetry.

The aim is to investigate the long-term deformation behaviour of the pavement construction. Specially, the permanent deformations of the pavement surface are of interest to assess the serviceability. For that purpose, the load of the pavement construction is modelled by a periodic loading process. The footprints of the load model are marked in Fig. 2. In contrast to a pure application of linear elastic 3D material models for the sublayers, the viscoplastic material behaviour of the asphalt layer is modelled additionally

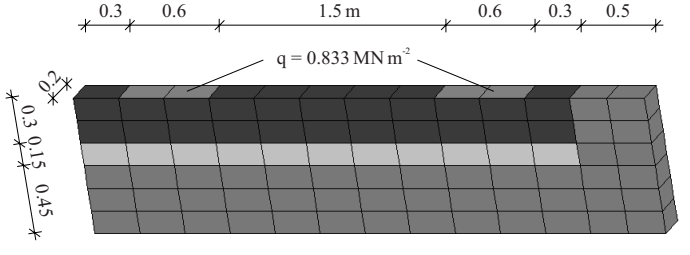


Fig. 2. Discretized pavement construction

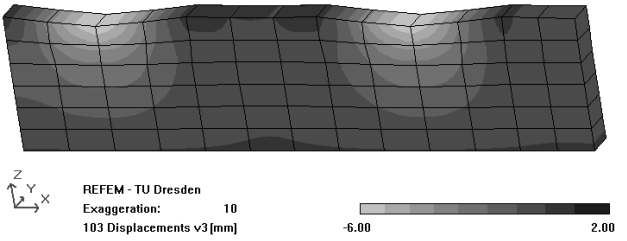


Fig. 3. Realisation of the vertical pavement displacement after 10^6 load cycles

with the aid of the fuzzy stochastic fractional NEWTON body. Due to a short duration of a load cycle viscoelastic strain components are neglected. The long-term deformation behaviour of the pavement construction is investigated in the load cycle domain. The time τ in the fractional differential equation (1) is substituted by the number of the load cycle lc. The parameters of the fuzzy stochastic fractional NEWTON body are identified on experimental data of cyclic loading tests. The uncertain operator $\tilde{\alpha}$ is modelled by a fuzzy number $\tilde{\alpha} =< 0.055; 0.0775; 0.1 >$. The stress dependent parameter $\tilde{p}(\sigma_v)$ is identified as

$$\tilde{p}(\sigma_v) = -227.778\,\sigma_v^2 + \tilde{p}_{lin}\,|\sigma_v| + 3.066 \tag{7}$$

The uncertainty of the factor $\tilde{p}_{lin}$ is modelled by means of a normal distributed fuzzy random variable $\tilde{X}$ with mean value $\mu_x = 265.95$ and fuzzy standard deviation $\tilde{\sigma}_x = \tilde{s}_1 =< 6.0; 8.62; 10.0 >$.

The deterministic fundamental solution $d(.)$ is represented by a 3D computational finite element model based on the numerical solution strategy introduced in [10]. In Fig. 3 a realisation of the vertical displacement after 10^6 load cycles is shown. The load cycle dependent fuzzy failure probability $\tilde{P}_f(lc)$ is computed with the aid of

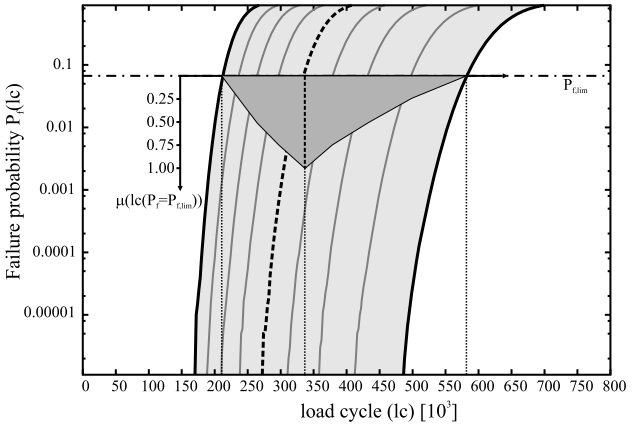


Fig. 4. Load cycle dependent fuzzy safety level $\tilde{P}_f(lc)$

the FMCS [9, 13] utilizing response surface approximation based on artificial neural networks, see [14]. The serviceability limit state is achieved, if the deformation is 5.0 mm. As a result of the investigation the fuzzy function $\tilde{P}_f(lc)$ of the pavement construction is shown in Fig. 4. Comparing the fuzzy failure probability $\tilde{P}_f(lc)$ with the permissible failure probability, e.g., for the serviceability limit state predefined in codes like $\beta = 1.5$ ($P_{f,lim} = 6.6807\,10^{-2}$), the number of sustainable load cycles is determined.

6 Conclusions

In the paper a method to account for uncertain long-term material behaviour utilizing a fuzzy stochastic fractional rheological model is introduced. Furthermore, a concept of fuzzy stochastic reliability assessment is outlined. The reliability is quantified by means of the fuzzy failure probability. This enables the safety assessment under consideration of subjective evaluations, which are frequently present in the engineering practice.

Acknowledgement. The authors gratefully acknowledge the support of the German Research Foundation (DFG).

References

1. Bronstein, I.N., Semendjajew, K.A., Musiol, G., Mühlig, H.: Taschenbuch der Mathematik (Pocket book of mathematics). Harry Deutsch, Frankfurt (1997)
2. Feng, Y.: The variance and covariance of fuzzy random variables and their applications. Fuzzy Sets Syst. 120, 487–497 (2001)
3. Gross, D., Hauger, W., Schnell, W., Wriggers, P.: Technische Mechanik, Bd. 4. Springer, Berlin (1995)
4. Körner, R.: Linear Models with Random Fuzzy Variables. Dissertation, Bergakademie Freiberg (1997)
5. Krätschmer, V.: A unified approach to fuzzy random variables. Fuzzy Sets Syst. 123, 1–9 (2001)
6. Kwakernaak, H.: Fuzzy random variables – I. Definitions and theorems. Inf. Sci. 15, 1–29 (1978)
7. Möller, B., Beer, M.: Fuzzy Randomness: Uncertainty in Civil Engineering and Computational Mechanics. Springer, Heidelberg (2004)
8. Möller, B., Beer, M.: Engineering computation under uncertainty – Capabilities of non-traditional models. Comput. Struct. 86, 1024–1041 (2008)
9. Möller, B., Beer, M., Graf, W., Sickert, J.-U.: Time-dependent reliability of textile strengthened RC structures under consideration of fuzzy randomness. Comput. Struct. 84, 585–903 (2006)
10. Oeser, M., Freitag, S., Möller, B., Wellner, F., Werkmeister, S.: 3-D constitutive model for asphalt on the basis of fractional creep functions. In: Proceedings on the 10th International Conference on Asphalt Pavements (ISAP, Québec, Canada), CD-ROM, pp. 755–765 (2006)
11. Petryna, Y.S., Krätzig, W.B.: Computational framework for long-term reliability analysis of RC structures. Comput. Methods Appl. Mech. Engrg. 194, 1619–1639 (2005)
12. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)

13. Sickert, J.-U.: Fuzzy Random Functions and thier Application in Structural Analysis and Safety Assessment (in german). Dissertation, TU Dresden, Veröffentlichungen Institut für Statik und Dynamik der Tragwerke, Heft 9, Dresden (2005)
14. Simpson, T.W., Peplinski, J.D., Koch, P.N., Allen, J.K.: Metamodels for computer-based engineering design: Survey and recommendations. Eng. Comput. 17, 129–150 (2001)
15. Viertl, R.: Statistical Methods for Non-Precise Data. CRC Press, Boca Raton (1996)
16. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, New York (1991)
17. Wang, G., Zhang, Y.: The theory of fuzzy stochastic processes. Fuzzy Sets Syst. 51, 161–178 (1992)
18. West, B., Bologna, M., Grigolini, P.: Physics of Fractal Operators. Springer, New York (2003)

Data Analysis

A New Insight into the Linguistic Summarization of Time Series Via a Degree of Support: Elimination of Infrequent Patterns

Janusz Kacprzyk and Anna Wilbik*

Systems Research Institute, Polish Academy of Sciences, Warszawa, Poland

Abstract. We extend our previous works on using a fuzzy logic based calculus of linguistically quantified propositions for linguistic summarization of time series (cf. Kacprzyk, Wilbik and Zadrożny [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. That approach, using the classic degree of truth (validity) to be maximized, is here extended by adding a degree of support. On the one hand, this can reflect in natural language the essence of traditional statistical approaches, and on the other hand, can help discard linguistic summaries with a high degree of truth but a low degree of support so that they concern infrequently occurring patterns and may be uninteresting. We show an application to the absolute performance analysis of an investment (mutual) fund.

1 Introduction

A linguistic summary of a data (base) is meant as a concise, human-consistent description of a (numerical) data set expressed in a (quasi)natural language. It was introduced by Yager [21] and then further developed and implemented by Kacprzyk and Yager [14], and Kacprzyk, Yager and Zadrożny [15]. The contents of a numeric set of data (relational database) is summarized via linguistically quantified propositions using Zadeh's calculus of linguistically quantified propositions [22].

We are concerned with time series. Traditionally, for their analyses, statistical methods are applied. Though they reflect world (commonsense) knowledge because they try to grasp what *usually* happens, they do not try to take into account an imprecise meaning of this usuality. As precise and powerful as they are, thay may be viewed not fully human consistent.

In this paper we adopt this perspective. Our intention is to provide an additional tool based on an explicit use of natural language. It is an extension of our previous works (cf. Kacprzyk, Wilbik and Zadrożny [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]), mainly towards a more complex evaluation of the results (linguistic summaries) obtained.

The analysis of time series is a complex task involving many aspects (cf. Batyrshin and Sheremetov [2, 3]). First, we need to identify the consecutive parts of time series, (partial)trends, within which the data exhibit some uniformity as to their variability. The (linguistic) summaries of time series refer to the (linguistic) summaries of (partial) trends. We have to aggregate the (characteristic features of) consecutive trends over an entire time span (horizon) assumed. We follow the idea initiated by Yager [21] and then

* Supported by the Ministry of Science and Higher Education under Grant No. NN516 4309 33.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 393–400, 2008.
springerlink.com

shown made to be in Kacprzyk and Yager [14], and Kacprzyk, Yager and Zadrożny [15], that the most comprehensive and meaningful will be a linguistic quantifier driven aggregation resulting in linguistic summaries exemplified by *"Most trends are short"* or *"Most long trends are increasing"* which are easily derived and interpreted using Zadeh's fuzzy logic based calculus of linguistically quantified propositions. A new quality, and an increased generality was obtained by using Zadeh's [23] protoforms as proposed by Kacprzyk and Zadrożny [16].

Here we employ the classic Zadeh's fuzzy logic based calculus of linguistically quantified propositions as in our source papers but in addition to the degree of truth (validity), we use a degree of support as the second criterion to ease the ranking of the summaries, indicating if an observed behavior concerns a frequent or infrequent pattern. Notice that since the concept of support has a clear probabilistic flavor because it basically specifies the chance (probability, in the frequentistic sense) that the object exhibits a specified property (or some properties). The use of a fuzzy quantifier base aggregation can again be viewed as an attempt to expand the traditional line of reasoning to a widespread use of human perception related to the use of natural language.

As an example, we will show a linguistic summarization of daily quotations over an eight year period of an investment (mutual) fund. We will present in detail the characteristic features of trends derived under some reasonable granulations, variability, trend duration, etc.

The approach is in line with some other modern approaches to linguistic summarization of time series, for instance with `SumTime` (cf. `www.csd.abdn. ac.uk/research/sumtime/`).

Notice that our purpose is not to forecast the future behavior of time series but to summarize its past behavior. This may be very useful for various purposes, for instance for the comparison of the past performance of various investment funds.

2 Data Preprocessing

There are many algorithms for the piecewise linear segmentation of time series data, including e.g. on-line (sliding window) algorithms, bottom-up or top-down strategies (cf. Keogh [18, 19]). In our works [4, 5, 7, 8, 9, 10, 11, 12, 13] we used a simple on-line algorithm, a modification of the Sklansky and Gonzalez one [20].

We consider the following three features of (global) trends in time series: (1) dynamics of change, (2) duration, and (3) variability.

By *dynamics of change* we understand the speed of change of the consecutive values of time series. It may be described by the slope of a line representing the trend. *Duration* is the length of a single trend, and is also represented by a linguistic variable. *Variability* describes how "spread out" a group of data is. We compute it as a weighted average of values taken by some measures used in statistics: (1) the range, (2) the interquartile range (IQR), (3) the variance, (4) the standard deviation, and (5) the mean absolute deviation (MAD). This is also treated as a linguistic variable. For practical reasons for all we use a fuzzy granulation (cf. Bathyrshin at al. [1, 2]) to represent the values by a small set of linguistic labels as, e.g.: quickly increasing, increasing, slowly increasing,

constant, slowly decreasing, decreasing, quickly decreasing. These values are equated with fuzzy sets.

3 Linguistic Data Summaries and Their Protoforms

A linguistic summary is a (usually short) natural language like sentence(s) that subsume the very essence of numeric and large data set which is hard to be comprehended by the human – cf. Kacprzyk and Zadrożny [16, 17]. A linguistic summary of a data set comprises: (1) a summarizer P (e.g. "low" for attribute "salary"), (2) a quantity in agreement Q (a linguistic quantifier, e.g. most), (3) truth (validity) $\mathcal{T} \in [0,1]$, and possibly (4) a qualifier R (e.g. "young" for attribute "age"). It may be exemplified by $\mathcal{T}$(most of employees earn low salary) $= 0.7$ or, in an extended form including a qualifier (e.g. young): $\mathcal{T}$(most of young employees earn low salary) $= 0.9$.

Then, $\mathcal{T} \in [0,1]$, i.e. the truth (validity) of a linguistic summary, directly corresponds to the truth value of the above linguistically quantified statements. The formulas employed will be given later.

We employ Zadeh's [23] protoforms for dealing with linguistic summaries [16]. Here we use different protoforms of time series summaries (cf. Kacprzyk, Wilbik and Zadrożny [4] – [13]):

- For a short form: Among all segments, Q are P
- For an extended form: Among all R segments, Q are P

3.1 Quality Measures of Linguistic Summaries

In our previous works (cf. Kacprzyk, Wilbik, Zadrożny [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]) we employed the degree of truth (validity) as advocated by Yager in his source paper [21]. Now, following Kacprzyk and Yager [14] and Kacprzyk, Yager and Zadrożny [15] – where other quality measures were given: degrees of imprecision, covering, appropriateness, and the length of a summary – we use the degrees of truth and support, a modification of the degree of covering.

The truth values for the simple and extended summaries are ($\wedge$ is the minimum or a t-norm, and Q is a nondecreasing fuzzy linguistic quantifier quantifier):

$$\mathcal{T}(\text{Among all } Y,\ Q \text{ are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{1}$$

$$\mathcal{T}(\text{Among all } RY,\ Q \text{ are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_R(y_i)\wedge\mu_P(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{2}$$

In Kacprzyk and Yager [14] and Kacprzyk, Yager and Zadrożny [15] the degree of covering of "QRy's are P" is

$$d_c(QRy's \text{ are } P) = \frac{\text{card}\{y : \mu_P(y) > 0 \wedge \mu_R(y) > 0\}}{\text{card}\{y : \mu_R(y) > 0\}} \tag{3}$$

The $d_c(.)$ says how many objects corresponding to the query are "covered" by the summary. It yields the proportion of elements exhibiting both P and R to those exhibiting

R only. This is similar to the confidence measure in association rules where it is accompanied usually by the support measure. Basically, if the degree of support is low, such a summary describes a (local) pattern seldomly occurring. This is the main motivation for using this measure in our context since we wish to avoid summaries that seldom happen. Hence, the degrees of support for the simple and extended protoforms, is calculated, respectively, as:

$$d_s(\text{Among all } Y,\ Q \text{ are } P) = \frac{1}{n}|\{y : \mu_P(y) > 0\}| \tag{4}$$

$$d_s(\text{Among all } RY,\ Q \text{ are } P) = \frac{1}{n}|\{y : \mu_P(y) > 0 \wedge \mu_R(y) > 0\}| \tag{5}$$

and, using a t-norm (e.g. 'min") and the Σ-count, we obtain the fuzzified degrees of support:

$$d_{sf}(\text{Among all } Y,\ Q \text{ are } P) = \frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i) \tag{6}$$

$$d_{sf}(\text{Among all } RY,\ Q \text{ are } P) = \frac{1}{n}\sum_{i=1}^{n}\mu_R(y_i) \wedge \mu_P(y_i) \tag{7}$$

4 Numerical Experiments

We tested our approach on daily quotations of an investment (mutual) fund that invests at most 50% of assets in shares listed at the Warsaw Stock Exchange. Quotations from April, 1998 – July, 2007, beginning with PLN 10.00 per share, and ending with PLN 55.27, were considered. The minimal share value was PLN 6.88, the maximal was PLN 57.85, the highest daily increase was e PLN 1.27, while the highest daily decrease was PLN 2.41.

Using the modified Sklansky and Gonzalez algorithm and $\varepsilon = 0.25$ we obtained 326 extracted (partial) trends, from 2 days to 71 days.

For lack of space, we will only give some examples of the summaries obtained under various granulations (number of linguistic values), with the $\mathcal{T}$, and d_s and d_{sf}:

- For 3 labels for the dynamics of change (*decreasing, constant, increasing*), the duration (*short, medium length, long*) and the variability (*low, medium, high*) in Table 1.

Notice that the last 4 summaries have a high degree of truth but a low degree of support so that they do not refer to frequently occurring situations and may be not relevant enough for the user.

- For 5 labels for the dynamics of change (*quickly decreasing, decreasing, constant, increasing, quickly increasing*), the duration (*very short, short, medium length, long, very long*) and the variability (*very low, low, medium, high, very high*) are in Table 2.

Notice that, again, the last 4 summaries have a high degree of truth but a low degree of support so that they do not refer to frequently occurring situations and may be not relevant enough.

Table 1.

linguistic summary	$\mathscr{T}$	d_s	d_{sf}
Among all trends, most are short	0.7129	0.6871	0.6564
Among all trends, most are constant	0.6318	0.6748	0.6159
Among all trends, most are of a low variability	0.7028	0.7730	0.6514
Among all short trends, most are of a low variability	0.8784	0.5399	0.4852
Among all trends of a low variability, most are short	0.8898	0.5399	0.4852
Among all increasing trends, most are of a low variability	0.8867	0.1871	0.1615
Among all medium trends, at least around a half is of medium variability	1.0000	0.2270	0.1257
Among all trends of a high variability, at least around a half is increasing	0.9161	0.0184	0.0137
Among all decreasing trends, almost all are short	1.0000	0.1933	0.1668

Table 2.

linguistic summary	$\mathscr{T}$	d_s	d_{sf}
Among all trends, most are constant	0.6318	0.6748	0.6159
Among all trends, at least around a half is very short	1.0000	0.5460	0.5077
Among all trends, at least around a half is of a very small variability	1.0000	0.5225	0.4626
Among all very short trends, most are of a very small variability	0.9350	0.4202	0.3896
Among all trends of a very small variability, most are very short	1.0000	0.4202	0.3896
Among all constant trends, at least around a half are short	0.7476	0.3006	0.2538
Among all trends of a very high variability, at least around a half are quickly increasing	0.8299	0.0031	0.0020
Among all trends of medium length, almost all are constant	1.0000	0.1472	0.1383
Among all quickly decreasing trends, almost all are very short	1.0000	0.0675	0.0635
Among all long trends, much more than a half is of a low variability	0.5519	0.0521	0.0375

- For 7 labels for the dynamics of change (*quickly decreasing, decreasing, slowly decreasing, constant, slowly increasing, increasing, quickly increasing*), the duration (*very short, short, rather short, medium length, rather long, long, very long*) and the variability (*very low, low, rather low, medium, rather high, high, very high*) are in Table 3.

Here, we show first the simple summaries and then the extended ones. Notice that the first summary is totally true but its degree of support is very low which is clearly caused by a very restrictive linguistic quantifiers (almost none), and such a summary may not refer to a commonly appearing situation, hence is of a limited use. The same is with the last 3 summaries.

Table 3.

linguistic summary	$\mathcal{T}$	d_s	d_{sf}
Among all trends, almost none are of rather high variability	1.0000	0.0460	0.0245
Among all trends, at least around a half are very short	1.0000	0.5460	0.5077
Among all trends, at least around a half are constant	1.0000	0.5614	0.4633
Among all trends, at least around a half are of very low variability	0.9138	0.4570	0.4370
Among all trends, at most around one third is slowly increasing	1.0000	0.2147	0.1348
Among all very short trends, most are of a very low variability	0.9045	0.3957	0.3819
Among all trends of a very low variability, most are very short	1.0000	0.3957	0.3819
Among all slowly decreasing trends, most are very short	0.9709	0.1411	0.0914
Among all constant trends, at least around a half is short	0.6827	0.2577	0.1864
Among all trends of medium length, almost all are constant	1.0000	0.0798	0.0631
Among all quickly decreasing trends, almost all are very short	1.0000	0.0675	0.0635
Among all long trends, much more than a half is of rather low variability	0.7344	0.0123	0.0069

It can be noticed that the use of the degree of support in addition to the degree of truth provides much help because it helps eliminate summaries which are very true but concern situations whose chance of appearance is low. A further extension would be a full fledged bi-crietria analysis with the degree of validity and support.

5 Concluding Remarks

We extended our works (cf. Kacprzyk, Wilbik, Zadrożny [4,5,6,7,8,9,10,11,12,13]) by adding to the basic degree of truth (validity), a degree of support. This can reflect in natural language the very essence of statistical approaches. Moreover, the use of the degree of support can help discard linguistic summaries which concern infrequently occurring patterns. We show an application to the absolute performance type analysis of daily quotations of an investment fund. The results are promising. Since a final assessment of the results should involve a human evaluation, a questionnaire type technique will be shown later.

References

1. Batyrshin, I.: On granular derivatives and the solution of a granular initial value problem. Int. J. Appl. Math. Comput. Sci. 12(3), 403–410 (2002)
2. Batyrshin, I., Sheremetov, L.: Perception based functions in qualitative forecasting. In: Batyrshin, I., Kacprzyk, J., Sheremetov, L., Zadeh, L.A. (eds.) Perception-based Data Mining and Decision Making in Economics and Finance, pp. 119–134. Springer, Heidelberg (2006)
3. Batyrshin, I., Sheremetov, L.: Towards perception based time series data mining. In: Nikravesh, M., Kacprzyk, J., Zadeh, L.A. (eds.) Forging New Frontiers. Fuzzy Pioneers I, pp. 217–230. Springer, Heidelberg (2007)

4. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summarization of trends: a fuzzy logic based approach. In: Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 2006, Paris, France), pp. 2166–2172 (2006)
5. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summaries of time series via a quantifier based aggregation using the Sugeno integral. In: Proceedings of 2006 IEEE World Congress on Computational Intelligence (WCCI 2006, Vancouver, BC, Canada), pp. 3610–3616. IEEE Press, Los Alamitos (2006)
6. Kacprzyk, J., Wilbik, A., Zadrożny, S.: On some types of linguistic summaries of time series. In: Proceedings of the 3rd International IEEE Conference on Intelligent Systems (IS 2006, London, UK), pp. 373–378. IEEE Press, Los Alamitos (2006)
7. Kacprzyk, J., Wilbik, A., Zadrożny, S.: A linguistic quantifier based aggregation for a human consistent summarization of time series. In: Lawry, J., Miranda, E., Bugarin, A., Li, S., Gil, M.A., Grzegorzewski, P., Hryniewicz, O. (eds.) Soft Methods for Integrated Uncertainty Modelling, pp. 186–190. Springer, Heidelberg (2006)
8. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Capturing the essence of a dynamic behavior of sequences of numerical data using elements of a quasi-natural language. In: Proceedings of the 2006 IEEE International Conference on Systems, Man, and Cybernetics (IEEE–SMC 2006, Taipei, Taiwan), pp. 3365–3370. IEEE Press, Los Alamitos (2006)
9. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summarization of time series by using the Choquet integral. In: Melin, P., Castillo, O., Aguilar, L.T., Kacprzyk, J., Pedrycz, W. (eds.) IFSA 2007. LNCS (LNAI), vol. 4529, pp. 284–294. Springer, Heidelberg (2007)
10. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summarization of time series under different granulation of describing features. In: Kryszkiewicz, M., Peters, J.F., Rybinski, H., Skowron, A. (eds.) RSEISP 2007. LNCS (LNAI), vol. 4585, pp. 230–240. Springer, Heidelberg (2007)
11. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Linguistic summaries of time series via an OWA operator based aggregation of partial trends. In: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007, London, UK), pp. 467–472. IEEE Press, Los Alamitos (2007)
12. Kacprzyk, J., Wilbik, A., Zadrożny, S.: Mining time series data via linguistic summaries of trends by using a modified Sugeno integral based aggregation. In: Proceedings of IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007, Honolulu, Hawaii), pp. 742–749. IEEE Press, Los Alamitos (2007)
13. Kacprzyk, J., Wilbik, A., Zadrożny, S.: On linguistic summaries of time series via a quantifier based aggregation using the Sugeno integral. In: Castillo, O., Melin, P., Kacprzyk, J., Pedrycz, W. (eds.) Hybrid Intelligent Systems Analysis and Design, pp. 421–439. Springer, Heidelberg (2007)
14. Kacprzyk, J., Yager, R.R.: Linguistic summaries of data using fuzzy logic. Int. J. Gen. Syst. 30, 33–154 (2001)
15. Kacprzyk, J., Yager, R.R., Zadrożny, S.: A fuzzy logic based approach to linguistic summaries of databases. Int. J. Appl. Math. Comput. Sci. 10, 813–834 (2000)
16. Kacprzyk, J., Zadrożny, S.: Linguistic database summaries and their protoforms: toward natural language based knowledge discovery tools. Inform. Sci. 173, 281–304 (2005)
17. Kacprzyk, J., Zadrożny, S.: Fuzzy linguistic data summaries as a human consistent, user adaptable solution to data mining. In: Gabrys, B., Leiviska, K., Strackeljan, J. (eds.) Do Smart Adaptive Systems Exist?, pp. 321–339. Springer, Heidelberg (2005)
18. Keogh, E., Chu, S., Hart, D., Pazzani, M.: An online algorithm for segmenting time series. In: Proceedings of IEEE International Conference on Data Mining (ICDM 2001, San Jose, California, USA), pp. 289–296 (2001)

19. Keogh, E., Chu, S., Hart, D., Pazzani, M.: Segmenting time series: A survey and novel approach. In: Last, M., Kandel, A., Bunke, H. (eds.) Data Mining in Time Series Databases, World Scientific Publishing, Singapore (2004)
20. Sklansky, J., Gonzalez, V.: Fast polygonal approximation of digitized curves. Pattern Recognition 12(5), 327–331 (1980)
21. Yager, R.R.: A new approach to the summarization of data. Inform. Sci. 28, 69–86 (1982)
22. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Comput. Math. Appl. 9, 149–184 (1983)
23. Zadeh, L.A.: A prototype-centered approach to adding deduction capabilities tosearch engines – the concept of a protoform. In: Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2002, New Orleans, USA), pp. 523–525 (2002)

An Autoregressive Model with Fuzzy Random Variables

Dabuxilatu Wang

School of Mathematics and Information Sciences, Guangzhou University,
Guangzhou, P.R. China

Abstract. An autoregressive model is defined for fuzzy random variables under the concept of Fréchet variance and covariance as well as Gaussian fuzzy random variable. In some special case, by using the Hukuhara difference between fuzzy sets, the conditions for stationary solution of a p-order autoregressive process ($AR(p)$) are extended to the case of fuzzy data in the manner of conventional stochastic setting.

1 Introduction

In time series analysis (cf. [1]) the p-order autoregressive ($AR(p)$) process is a fundamental aspect for time series modelling. It has been a basic model for deriving out other important time series model such as ARCH model and ARIMA model,etc.. Recently, some hybrid modelling methods (e.g. neural network [8], fuzzy coefficients [13, 14], etc.) have been applied for time series modelling. Among them the time series with linguistic values proposed by [12] is a typical example of fuzzy data valued time series and it had been used for forecasting the enrollments. In this model the fuzzy data are only taken on values of fuzzy sets defined on some finite discrete domain. Such fuzzy data valued time series approach had been developed by [2], [7] extensively. However, their approaches are mainly based on the fuzzy relational equations and approximate reasonings. There have been signs of some advantages in these models over the traditional one such as the conveniance on computation. It should be pointed out that in which the stochastic characteristics of the time series are completely ignored. Such an ignorance would led to some biased results of forecasting and lack of a reasonable testing procedure. In fact, concerning the both randomness and fuzziness implied in such time series, it seems optimal to integrate the conventional statistical tools, concepts, proposals into such time series analysis, so as to carry out more flexible and reliable models for the fuzzy data valued time series. Motivated by the works on the stochastic regression models for fuzzy data (cf. [6, 10, 16]), and the consideration on complicated mutual relationship among the fuzzy data appeared in time series, we may consider the fuzzy data valued time series under stochastic point of view, and allow the fuzzy data may taking on values of fuzzy sets defined on some continuous domain like real line R or n-dimensional vector space R^n. Since the non-linear structure of the set of all fuzzy random variables (cf. [11, 5]) as well as the complicated formulas for the expectation and variance and covariance (cf. [4, 10]), it is usually hard to define an $AR(p)$ model for fuzzy data time series and apply it in the real life forecasting process. In this paper,

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 401–408, 2008.
springerlink.com

from theoretical point of view, we try to give a simple $AR(p)$ model for fuzzy data valued time series by using support function of set and Gaussian fuzzy error as well as the Fréchet variance and covariance.

2 Preliminaries

A fuzzy set $\tilde{u}$ of $\mathbb{R}^n$ equivalents to its membership function $\tilde{u}:\mathbb{R}^n\to[0,1]$, where the number $\tilde{u}(x)$ represents the degree of membership that x belongs to $\tilde{u}$. By $F(\mathbb{R}^n)$ we denote the collection of all normal, convex and compact fuzzy sets on $\mathbb{R}^n$, i.e. for $\tilde{u}\in F(\mathbb{R}^n)$, (1) There exists $x_0\in\mathbb{R}^n$ such that $\tilde{u}(x)=1$; (2) The $\alpha-$cut of $\tilde{u}$, $\tilde{u}_\alpha:=\{x\in\mathbb{R}^n:\tilde{u}(x)\geq\alpha\}$, $\alpha\in(0,1]$, is a convex and compact set of $\mathbb{R}^n$; (3) $\tilde{u}_0:=cl\{x\in\mathbb{R}^n:\tilde{u}(x)>0\}$, the support of $\tilde{u}$, is compact.

Zadeh's extension principle (cf. [3, 17]) allows us to proceed addition and scale multiplication on $F(\mathbb{R}^n)$:

$$(\tilde{u}\oplus\tilde{v})(x)=\sup_{s+t=x}\min(\tilde{u}(s),\tilde{v}(t)),\quad x\in\mathbb{R}^n.$$

$$(a\odot\tilde{u})(x)=\begin{cases}\tilde{u}(\frac{x}{a}),a\neq0\\ 0,a=0\end{cases}\qquad a\in\mathbb{R}.$$

and (cf. [6]) for any $a,b\in\mathbb{R}$, it holds

$$(ab)\odot\tilde{u}=a\odot(b\odot\tilde{u}),\qquad a\odot(\tilde{u}\oplus\tilde{v})=(a\odot\tilde{u})\oplus(a\odot\tilde{v}).$$

But it holds only for $ab\geq0$, $a,b\in\mathbb{R}$,

$$(a+b)\odot\tilde{u}=(a\odot\tilde{u})\oplus(b\odot\tilde{u}).$$

It indicates that $(F(\mathbb{R}^n),\oplus,\odot)$ is not a linear space. With Minkowvski's sets operation it holds

$$(\tilde{u}\oplus\tilde{v})_\alpha=\tilde{u}_\alpha\oplus\tilde{v}_\alpha,\quad \alpha\in(0,1].$$

$$(a\odot\tilde{u})_\alpha=a\odot\tilde{u}_\alpha,\quad \alpha\in(0,1].$$

A support function of $\tilde{u}\in F(\mathbb{R}^n)$ is defined as

$$S_{\tilde{u}_\alpha}(x)=\begin{cases}\sup_{t\in\tilde{u}_\alpha}\{x\cdot t\}, & \alpha\in(0,1],\\ 0, & \alpha=0.\end{cases}\qquad x\in S^{n-1}=\{x:\|x\|=1\}.$$

where $\cdot$ denotes the inner product in the Euclidean space $\mathbb{R}^n$. It holds that for $\tilde{u},\tilde{v}\in F(\mathbb{R}^n)$ and $a\in\mathbb{R}$,

$$S_{\tilde{u}\oplus\tilde{v}}=S_{\tilde{u}}+S_{\tilde{v}}.$$

$$S_{a\odot\tilde{u}}(x)=aS_{\tilde{u}}(x),\ a>0;\qquad S_{a\odot\tilde{u}}(x)=-aS_{\tilde{u}}(-x),\ a<0.$$

thus, it holds that

$$S_{((a\odot\tilde{u})\oplus(b\odot\tilde{v}))_\alpha}(x)=\begin{cases}(aS_{\tilde{u}_\alpha}+bS_{\tilde{v}_\alpha})(x), & a,b>0\\ -(aS_{\tilde{u}_\alpha}+bS_{\tilde{v}_\alpha})(-x), & a,b<0.\end{cases}$$

where $\alpha\in[0,1]$. We define the distance between $\tilde{u},\tilde{v}$ by

$$\delta_2(\tilde{u},\tilde{v}) := \left(n\int_0^1\int_{S^{n-1}} |S_{\tilde{u}_\alpha}(x) - S_{\tilde{v}_\alpha}(x)|^2 \mu(dx)d\alpha\right)^{1/2},$$

and let

$$\langle \tilde{u},\tilde{v}\rangle := n\int_0^1\int_{S^{n-1}} S_{\tilde{u}_\alpha}(x)S_{\tilde{v}_\alpha}(x)\mu(dx)d\alpha.$$

where μ is a normalized Lebesgue measure.

Let $(\Omega,\mathscr{B},P)$ be a complete probability space. The mapping $\tilde{X}:\Omega\to F(\mathbb{R}^n)$ is said to be a fuzzy random variable (f.r.v.) if $\tilde{X}$ is $\mathscr{B}-\mathscr{A}$ measurable, where $\mathscr{A}$ is a σ-algebra induced by $\tilde{X}$ associated with δ_2. Let $\tilde{X}$ be a f.r.v., then $S_{\tilde{X}_\alpha}$ is a random element and $E(S_{\tilde{X}_\alpha}) = S_{E(\tilde{X}_\alpha)}$ (cf. [10, 16]) if the expectation $E(\tilde{X}_\alpha)$ exists, where $E(\tilde{X}_\alpha)$ is an Aumann expectation of $(\tilde{X}_\alpha),\alpha\in[0,1]$ (cf. [11, 5]).

In the sequel, we assume that f.r.v. $\tilde{X}$ is with second order, i.e.

$$E(\|\tilde{X}\|) := E(\delta_2^2(\tilde{X},\{0\})) < +\infty,$$

The Fréchet variance of $\tilde{X}$ is defined by

$$Var(\tilde{X}) := E(\delta_2^2(\tilde{X},E(\tilde{X}))) = n\int_0^1\int_{S^{n-1}} Var(S_{\tilde{X}_\alpha}(x))\mu(dx)d\alpha.$$

and the covariance of f.r.v.'s $\tilde{X},\tilde{Y}$ is defined by

$$Cov(\tilde{X},\tilde{Y}) := n\int_0^1\int_{S^{n-1}} Cov(S_{\tilde{X}_\alpha}(x),S_{\tilde{Y}_\alpha}(x))\mu(dx)d\alpha.$$

and the usual classical form

$$Cov(\tilde{X},\tilde{Y}) = E\langle \tilde{X},\tilde{Y}\rangle - \langle E\tilde{X},E\tilde{Y}\rangle$$

holds. Note that,

$$Cov((a\odot\tilde{X})\oplus(b\odot\tilde{Y}),c\odot\tilde{Z}) = acCov(\tilde{X},\tilde{Z}) + bcCov(\tilde{Y},\tilde{Z})$$

holds only for $ac\geq 0$, $bc\geq 0$, $a,b,c\in\mathbb{R}$. In the case of $n=1$,

$$Var(\tilde{X}) = \int_0^1 (Var(\inf\tilde{X}_\alpha) + Var(\sup\tilde{X}_\alpha))d\alpha.$$

$$Cov(\tilde{X},\tilde{Y}) = \int_0^1 (Cov(\inf\tilde{X}_\alpha,\inf\tilde{Y}_\alpha) + Cov(\sup\tilde{X}_\alpha,\sup\tilde{Y}_\alpha))d\alpha.$$

The independence of f.r.v.'s can be followed by the independence of the random elements which is already defined (cf. [9]). Thus, obviously, if f.r.v. $\tilde{X}$ and $\tilde{Y}$ are independent , then $Cov(\tilde{X},\tilde{Y}) = 0$. However, if $Cov(\tilde{X},\tilde{Y})\neq 0$, then they will be dependent, there is some degree of dependence between them. We assume that the degree of dependence is represented by the value of the so called "correlation" $R(\tilde{X},\tilde{Y}) = Cov(\tilde{X},\tilde{Y})/\sqrt{Var(\tilde{X})Var(\tilde{Y})}$, and it holds $P(\delta_2(\tilde{Y}\oplus(\lambda\odot E\tilde{X}),E\tilde{Y}\oplus(\lambda\odot\tilde{X}))=0)=1$ when $R(\tilde{X},\tilde{Y})=1$; $P(\delta_2(\tilde{Y}\oplus(\lambda\odot\tilde{X}),E\tilde{Y}\oplus(\lambda\odot E\tilde{X}))=0)=1$ when $R(\tilde{X},\tilde{Y})=-1$ (cf. [15]). We say that $\tilde{X}$ and $\tilde{Y}$ are uncorrelated if $R(\tilde{X},\tilde{Y})=0$.

Denoting $\tilde{u}\oplus((-1)\odot\tilde{v}))$ by $\tilde{u}\ominus\tilde{v}$ for $\tilde{u},\tilde{v}\in F(\mathbb{R}^n)$, then by the linearity of expectation for f.r.v. with respect to the non-linear structure $(F(\mathbb{R}^n),\oplus,\odot)$ (cf. [5]), it holds that $E[\tilde{X}\ominus E(\tilde{X})]=\tilde{0}$, where $\tilde{0}$ (we call fuzzy zero) denotes a normal compact convex fuzzy set on $\mathbb{R}^n$ with $\tilde{0}(0)=1$ and it is not always true that $\tilde{0}=\{0\}$. This formula allows us to proceed a kind of centralization for a given f.r.v..

3 A Sort of Autoregressive Series of Fuzzy Random Variables

Definition 1. *Let $\{\tilde{X}_t\}(t \in \mathbb{Z})$ be a series of f.r.v.s valued in $F(\mathbb{R}^n)$ with second order, $\{\tilde{X}_t\},(t \in \mathbb{Z})$ is said to be a weak stationary process if it holds that (i) $E(\tilde{X}_t) = \tilde{c}$ and $Var(\tilde{X}_t) = \sigma^2$ for all t; (ii) $Cov(\tilde{X}_{t+h}, \tilde{X}_t) = Cov(\tilde{X}_h, \tilde{X}_0)$. Where $\mathbb{Z}$ is the set of all integers.*

Definition 2. *Let $\tilde{\varepsilon}$ be a fuzzy random variable valued in $F(\mathbb{R}^n)$. $\tilde{\varepsilon}$ is said to be a Gaussian error f.r.v. if $\tilde{\varepsilon}$=$E(\tilde{\varepsilon}) \oplus \varepsilon$ and $E(\tilde{\varepsilon}) = \tilde{0}$, random vector $\varepsilon \sim N_n(0, \Sigma)$.*

Definition 3. *A time series of fuzzy random variables $\{\tilde{X}_t\}$ is said to be a p-order autoregressive ($AR(p)$) process if $\{\tilde{X}_t\}$ is a weak stationary process and for any $t \in \mathbb{Z}$ it holds that*

$$\tilde{X}_t = (\theta_1 \odot \tilde{X}_{t-1}) \oplus (\theta_2 \odot \tilde{X}_{t-2}) \oplus \ldots \oplus (\theta_p \odot \tilde{X}_{t-p}) \oplus \tilde{\varepsilon}_t,$$

where θ_i is a real-valued parameter, $\{\tilde{\varepsilon}_t\}$ is a series of Gaussian error f.r.v.'s satisfying that $\tilde{\varepsilon}_t = E(\tilde{\varepsilon}_t) \oplus \varepsilon_t$ and $\{\varepsilon_t\}$ is a series of i.i.d. normal random vectors with $\varepsilon_t \sim N_n(0, \Sigma)$

Obviously, $E(\tilde{\varepsilon}_t) = \tilde{0}$, $Var(\tilde{\varepsilon}_t) = n \int_{S^{n-1}} x' \Sigma x \mu(dx)$. (cf. [16])

A series of f.r.v.'s $\{\tilde{X}_t\}$ is said to be a causal $AR(p)$ process if it has a stationary solution almost everywhere, i.e., there exists a number series $\{b_j\}$ such that $\tilde{X}_t$=$\oplus_{j=0}^{\infty}(b_j \odot \tilde{\varepsilon}_{t-j}), a.e.$.

Proposition 1. *Let $\{\tilde{\xi}_t\}$ be a series of Gaussian fuzzy random variables: $\tilde{\xi}_t = E(\tilde{\xi}_t) \oplus \xi_t$, and $\xi_t \sim N_n(0, \Sigma)$. If $\{\xi_t\}$ is an i.i.d. series of random vectors, then $\{\tilde{\xi}_t\}$ is uncorrelated, i.e. $Cov(\tilde{\xi}_t, \tilde{\xi}_j) = 0, t, j \in \mathbb{Z}$.*

Proof. By the definition of Gaussian f.r.v., it holds for $t, j \in \mathbb{Z}$

$$S_{(\tilde{\xi}_t)_\alpha}(x) = S_{E(\tilde{\xi}_t)_\alpha}(x) + S_{\{\xi_t\}}(x); S_{(\tilde{\xi}_j)_\alpha}(x) = S_{E(\tilde{\xi}_j)_\alpha}(x) + S_{\{\xi_j\}}(x),$$

by the i.i.d. of $\{\xi_t\}$, we get that $S_{\{\xi_t\}}(x)$ and $S_{\{\xi_j\}}(x)$ are independent, thus,

$$\begin{aligned} E\left(S_{(\tilde{\xi}_t)_\alpha}(x) S_{(\tilde{\xi}_j)_\alpha}(x)\right) &= E\left(S_{(\tilde{\xi}_t)_\alpha}(x)\right) E\left(S_{(\tilde{\xi}_j)_\alpha}(x)\right) + E\left(S_{(\tilde{\xi}_j)_\alpha}(x)\right) S_{\{E\xi_t\}}(x) \\ &\quad + E\left(S_{(\tilde{\xi}_t)_\alpha}(x)\right) S_{\{E\xi_j\}}(x) + S_{\{E\xi_t\}}(x) S_{\{E\xi_j\}}(x) \\ &= E\left(S_{(\tilde{\xi}_t)_\alpha}(x)\right) E\left(S_{(\tilde{\xi}_j)_\alpha}(x)\right), \end{aligned}$$

which means that

$$\begin{aligned} Cov(\tilde{\xi}_t, \tilde{\xi}_j) = n \int_0^1 \int_{S^{n-1}} \Big[& E\left(S_{(\tilde{\xi}_t)_\alpha}(x) S_{(\tilde{\xi}_j)_\alpha}(x)\right) \\ & - E\left(S_{(\tilde{\xi}_t)_\alpha}(x)\right) E\left(S_{(\tilde{\xi}_j)_\alpha}(x)\right) \Big] \mu(dx) d\alpha = 0. \end{aligned}$$

□

For a weak stationary process we define $r(h) := Cov(\tilde{X}_{t+h}, \tilde{X}_t)$ as an auto-covariance function. The Hukuhara deference $\ominus_H$ between two fuzzy sets refers to [16].

Proposition 2. *Let* $\{\tilde{X}_t\}(t \in \mathbb{Z})$ *be an* $AR(1)$ *process:* $\tilde{X}_t = (\theta \odot \tilde{X}_{t-1}) \oplus \tilde{\varepsilon}_t$, *then* $\{\tilde{X}_t\}$ *possesses a stationary solution almost everywhere if and only if* $0 < |\theta| < 1$.

Proof. By the non-linear structure of $(F(\mathbb{R}^n), \oplus, \odot)$, for arbitrary $\theta \in \mathbb{R}$ and each $\omega \in \Omega$, it holds for $t \in \mathbb{Z}$

$$\begin{aligned}\tilde{X}_t &= \theta \odot ((\theta \odot \tilde{X}_{t-1}) \oplus \tilde{\varepsilon}_{t-1}) \oplus \tilde{\varepsilon}_t \\ &= (\theta^2 \odot \tilde{X}_{t-2}) \oplus (\theta \odot \tilde{\varepsilon}_{t-1}) \oplus \tilde{\varepsilon}_t \\ &= \ldots\ldots \\ &= (\theta^{k+1} \odot \tilde{X}_{t-k-1}) \oplus (\theta^k \odot \tilde{\varepsilon}_{t-k}) \oplus \ldots \oplus (\theta \odot \tilde{\varepsilon}_{t-1}) \oplus \tilde{\varepsilon}_t,\end{aligned}$$

formally we have

$$\theta^{k+1} \odot \tilde{X}_{t-k-1} = \tilde{X}_t \ominus_H (\oplus_{j=0}^{k}(\theta^j \odot \tilde{\varepsilon}_{t-j})),$$

then, if $0 < |\theta| < 1$, it holds

$$\begin{aligned}&E\|\tilde{X}_t \ominus_H (\oplus_{j=0}^{k}(\theta^j \odot \tilde{\varepsilon}_{t-j}))\| \\ &= E\|\theta^{k+1} \odot \tilde{X}_{t-k-1}\| \\ &= n\int_0^1 \int_{s^{n-1}} E\left|(S_{(\theta^{k+1} \odot \tilde{X}_{t-k-1})_\alpha}(x))\right|^2 \mu(dx)d\alpha \\ &= \theta^{2(k+1)}\Big(Var(\tilde{X}_{t-k-1}) + n\int_0^1 \int_{s^{n-1}} \left|S_{E(\tilde{X}_{t-k-1})_\alpha}(x))\right|^2 \mu(dx)d\alpha\Big) \\ &\to 0 (k \to \infty).\end{aligned}$$

which means $E\|\tilde{X}_t \ominus_H (\oplus_{j=0}^{\infty}(\theta^j \odot \tilde{\varepsilon}_{t-j}))\| = 0$, i.e.,

$$E\Big[n\int_0^1 \int_{s^{n-1}} (S_{(\tilde{X}_t)_\alpha}(x) - S_{(\oplus_{j=0}^{\infty}\theta^j \odot \tilde{\varepsilon}_{t-j})_\alpha}(x))^2 \mu(dx)d\alpha\Big] = 0,$$

by the property of integral w.r.t. the measure mentioned above, it holds

$$n\int_0^1 \int_{s^{n-1}} (S_{(\tilde{X}_t)_\alpha}(x) - S_{(\oplus_{j=0}^{\infty}\theta^j \odot \tilde{\varepsilon}_{t-j})_\alpha}(x))^2 \mu(dx)d\alpha = 0, a.e.$$

then $(S_{(\tilde{X}_t)_\alpha}(x) - S_{(\oplus_{j=0}^{\infty}\theta^j \odot \tilde{\varepsilon}_{t-j})_\alpha}(x))^2 = 0, a.e. \Leftrightarrow S_{(\tilde{X}_t)_\alpha}(x) = S_{(\oplus_{j=0}^{\infty}(\theta^j \odot \tilde{\varepsilon}_{t-j}))_\alpha}(x)$, $a.e.$ which leads to $\tilde{X}_t = \oplus_{j=0}^{\infty}(\theta^j \odot \tilde{\varepsilon}_{t-j}), a.e.$. This is a stationary solution for above $AR(1)$ process. Obviously, the converse conclusion holds. □

In the classical time series theory, the concept of lag operator is already defined. Here we formally define a lag operator B for series of f.r.v.'s $\{\tilde{X}_t\}$ by $B^j \odot \tilde{X}_t = \tilde{X}_{t-j}$.

Let $\{\tilde{Y}_t\}$ be a weak stationary series of f.r.v.'s with second order, set $\tilde{X}_t = \tilde{Y}_t \ominus E(\tilde{Y}_t)$, then $\{\tilde{X}_t\}$ is a weak stationary series of f.r.v.'s with fuzzy zero expectation.

Proposition 3. *Let* $\{\tilde{X}_t\}(t \in Z)$ *be weak stationary p-order autoregressive process, then it possesses a stationary solution almost everywhere if and only if the polynomial* $a(z)=\theta_1 + \theta_2 z + \ldots + \theta_p z^{p-1}$ *satisfies* $0 < |a(z)| < 1$ *and all coefficients* θ_i $(i = 1, 2, \ldots, p.)$ *are positive (or negative)and* $\theta_i z^{i-1} (i = 2, 3, \ldots, p)$ *are positive.*

Proof. Since

$$\begin{aligned}\tilde{X}_t &= (\theta_1 \odot \tilde{X}_{t-1}) \oplus (\theta_2 \odot \tilde{X}_{t-2}) \oplus \ldots \oplus (\theta_p \odot \tilde{X}_{t-p}) \oplus \tilde{\varepsilon}_t \\ &= (a(B) \odot \tilde{X}_{t-1}) \oplus \tilde{\varepsilon}_t \\ &= (a(B) \odot (a(B) \odot \tilde{X}_{t-2} \oplus \tilde{\varepsilon}_{t-1})) \oplus \tilde{\varepsilon}_t \\ &= \ldots\ldots \\ &= (a(B)^{k+1} \odot \tilde{X}_{t-k-1}) \oplus (a(B)^k \odot \tilde{\varepsilon}_{t-k}) \oplus \ldots \oplus (a(B) \odot \tilde{\varepsilon}_{t-1}) \oplus \tilde{\varepsilon}_t \\ &= \ldots\ldots\end{aligned}$$

by Proposition 2 we obtain the conclusion immediately. □

Proposition 4. *Let $\{\tilde{X}_t\}$ be an AR(p) process defined in Definition 3. If $\{\tilde{X}_t\}$ is positive (negative) causal autoregressive , i.e. there exists a positive (or negative) number series $\{b_j\}$ such that $\tilde{X}_t = \oplus_{j=0}^{\infty}(b_j \odot \tilde{\varepsilon}_{t-j}), a.e.$, and all coefficients $\theta_i (i = 1, \ldots, p)$ are positive (or negative), then the Yule-Walker equation can be carried out almost everywhere for the process $\{\tilde{X}_t\}$.*

Proof. For $j = 0, 1, 2, \ldots, p$, it holds that

$$\begin{aligned}r(j) &= Cov(\tilde{X}_t, \tilde{X}_{t-j}) \\ &= Cov((\theta_1 \odot \tilde{X}_{t-1}) \oplus (\theta_2 \odot \tilde{X}_{t-2}) \oplus \ldots \oplus (\theta_p \odot \tilde{X}_{t-p}) \oplus \tilde{\varepsilon}_t, \tilde{X}_{t-j}) \\ &= \theta_1 Cov(\tilde{X}_{t-1}, \tilde{X}_{t-j}) + \theta_2 Cov(\tilde{X}_{t-2}, \tilde{X}_{t-j}) + \ldots + Cov(\tilde{\varepsilon}_t, \tilde{X}_{t-j}) \\ &= \theta_1 r(j-1) + \theta_2 r(j-2) + \ldots + \theta_p r(j-p) + b_0 Cov(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t-j}), a.e.,\end{aligned}$$

i.e.

$$r(0) = \theta_1 r(1) + \theta_2 r(2) + \ldots + \theta_p r(p) + b_0 Var(\tilde{\varepsilon}_t), a.e.,$$

and

$$\begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix} = \begin{pmatrix} r(0) & r(-1) & \ldots & r(1-p) \\ r(1) & r(0) & \ldots & r(2-p) \\ \vdots & \vdots & \ldots & \vdots \\ r(p) & r(p-1) & \ldots & r(0) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}, a.e..$$

□

Now we consider the problem of estimation with respect to the $AR(p)$ model based on sampling observations $\tilde{x}_1, \ldots, \tilde{x}_m$. Same like the classical case, we can instead of auto-covariance function $r(j)$ by the sample auto- covariance function $\hat{r}(h)$ in the above Yule-Walker equation to estimate the unknown coefficients parameters θ_i. Here

$$\hat{r}(h) := \frac{1}{m} \sum_{t=1}^{m-h} n \int_0^1 \int_{S^{n-1}} (S_{(\tilde{x}_{t+h})_\alpha}(x) - S_{\overline{\tilde{x}}_\alpha}(x))(S_{(\tilde{x}_t)_\alpha}(x) - S_{\overline{\tilde{x}}_\alpha}(x)) \mu(dx) d\alpha,$$

and $\hat{r}(h) = \hat{r}(-h)$, $0 \le h < m, -m < -h \le 0$. where $\overline{\tilde{x}} = \frac{1}{n} \oplus_{i=1}^{m} \tilde{x}_i$. If the matrix of self covariance functions is inversive , then we could obtain the estimated $AR(p)$ process as

$$\tilde{X}_t = (\hat{\theta}_1 \odot \tilde{X}_{t-1}) \oplus (\hat{\theta}_2 \odot \tilde{X}_{t-2}) \oplus \ldots \oplus (\hat{\theta}_p \odot \tilde{X}_{t-p}) \oplus \tilde{\varepsilon}_t.$$

where

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_p \end{pmatrix} = \begin{pmatrix} \hat{r}(0) & \hat{r}(-1) & \dots & \hat{r}(1-p) \\ \hat{r}(1) & \hat{r}(0) & \dots & \hat{r}(2-p) \\ \vdots & \vdots & \dots & \vdots \\ \hat{r}(p) & \hat{r}(p-1) & \dots & \hat{r}(0) \end{pmatrix}^{-1} \begin{pmatrix} \hat{r}(1) \\ \hat{r}(2) \\ \vdots \\ \hat{r}(p) \end{pmatrix}, a.e..$$

The estimated $AR(p)$ model gives a way to express the series of residual as $\tilde{\varepsilon}_t = \tilde{X}_t \ominus_H \hat{\tilde{X}}_t$, where $\hat{\tilde{X}}_t = (\hat{\theta}_1 \odot \tilde{X}_{t-1}) \oplus (\hat{\theta}_2 \odot \tilde{X}_{t-2}) \oplus \dots \oplus (\hat{\theta}_p \odot \tilde{X}_{t-p})$. By this consideration, an autoregressive (or autocorrelated) series of f.r.v.'s $\{\tilde{X}_t\}$ could be modelled through an uncorrelated series of Gaussian f.r.v.'s $\{\tilde{X}_t \ominus_H \hat{\tilde{X}}_t\}$, for which, following classical approaches for modelling uncorrelated time series, various statistical quality control charts as well as other useful stochastic models depending on uncorrelated time series could be easily established, therefore, we may partially solved the modelling problems for autoregressive (or autocorrelated) series of f.r.v.'s.

Conclusions. Based on the introduced hybrid notion Gaussian fuzzy error and the already established notions like Fréchet variance and covariance as well as the Aumann expectation for f.r.v.'s, we have proposed a novel special $AR(p)$ model for time series of f.r.v.'s, which is different from all of previous obtained models of fuzzy time series. We are aware of that such a model is hard to be applied in real life because of the operation caused by the complicated Hukuhara difference, and also the correlation between f.r.v.'s is much more complicated than the ordinary random variables case, even if $R(\tilde{X}, \tilde{Y}) = 1$, it is still uncertain on the linear dependence between the two f.r.v.'s $\tilde{X}$ and $\tilde{Y}$. It is desirable to reconsider the notion of correlation for f.r.v.'s, and propose more suitable error term instead of the simple Gaussian fuzzy error for the time series of f.r.v.'s modelling.

Acknowledgement. The author would like to thank reviewer for their helpful suggestions on the manuscript. The research in this paper is supported by the funds No.2005–546 of Education Ministry of P.R.China and the funds No. 62033 of Education bureau of Guangzhou city. Their financial support is gratefully acknowledged.

References

1. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1991)
2. Chen, S.M.: Forecasting enrollments based on fuzzy time series. Fuzzy Sets Syst. 81, 311–319 (1996)
3. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific Publishing Co., Inc., River Edge (1994)
4. Feng, Y., Hu, L., Shu, H.: The variance and covariance of fuzzy random variables and their applications. Fuzzy Sets Syst. 120, 487–497 (2001)
5. Krätschmer, V.: Probability theory in fuzzy sample space. Metrika 60(2), 167–189 (2004)
6. Krätschmer, V.: Limit distributions of least squares estimators in linear regression models with vague concepts. J. Multivariate Anal. 97, 1044–1069 (2006)
7. Lee, L.W., Wang, L.H., Chen, S.M., Leu, Y.H.: Handling forecasting problem based on two-factors high-order fuzzy time series. IEEE Trans. Fuzzy Syst. 14(3), 468–477 (2006)

8. Medeiros, M.C., Teräsvirta, T., Rech, G.: Building neural network models for time series: A statistical approach. J. Forecast 25(1), 49–75 (2006)
9. Näther, W.: On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data. Metrika 51, 201–221 (2000)
10. Näther, W.: Regression with fuzzy random data. Comput. Statist. Data Anal. 51, 235–252 (2006)
11. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)
12. Song, S., Chissom, B.S.: Forecasting enrollments with fuzzy time series. Fuzzy Sets Syst. 54(1), 1–19 (1993)
13. Tseng, F.M., Tzeng, G.H., Yu, H.C., Yuan, B.J.C.: Fuzzy ARIMA model for forecasting the foreign exchenge market. Fuzzy Sets Syst. 118, 9–19 (2001)
14. Thavaneswaran, A., Thiagarajah, K., Appadoo, S.S.: Fuzzy coefficient volatility (FCV) models with applications. Math. Comput. Modelling 45, 777–786 (2007)
15. Wang, D., Yasuda, M.: Some asymptotic properties of point estimation with *n*-dimensional fuzzy data. Statistics 38(2), 167–181 (2004)
16. Wünsche, A., Näther, W.: Least-square fuzzy regression with fuzzy random variables. Fuzzy Sets Syst. 130, 43–50 (2002)
17. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Part 1. Inform. Sci. 8, 199–249 (1975); Part 2. Inform. Sci. 8, 301–353; Part 3. Inform. Sci. 9, 43–80

Tackling Multiple-Instance Problems in Safety-Related Domains by Quasilinear SVM

Christian Moewes[1], Clemens Otte[2], and Rudolf Kruse[1]

[1] Department of Knowledge and Language Engineering, University of Magdeburg, Magdeburg, Germany
[2] Siemens AG, Corporate Technology, Munich, Germany

Abstract. In this paper we introduce a preprocessing method for safety-related applications. Since we concentrate on scenarios with highly unbalanced misclassification costs, we briefly discuss a variation of multiple-instance learning (MIL) and recall soft margin hyperplane classifiers; in particular the principle of a support vector machine (SVM). According to this classifier, we present a training set selection method for learning quasilinear SVMs which guarantee both high accuracy and model complexity to a lower degree. We conclude with annotating on a real-world application and potential extensions for future research in this domain.

1 Introduction

Safety-related systems can be found in manifold fields where a failure may lead to fatalities or severe injuries to human beings, loss or very bad damage of equipment, or environmental harm [9]. The usage of machine learning methods is not that straightforward compared to other applications where learning machines have been applied very successfully.

Main differences to other classification domains are highly unbalanced classification costs and the infrequency of positive events, e.g., trigger events, and alarms. We try to compare this domain with multiple-instance (MI) learning [5] of which problems partly resemble safety-related applications. In contrast to single-instance supervised learning where one given example is represented by one feature vector (so-called instance), here an example is a set of feature vectors. Therefore, this setting of the learning problem is called *multiple-instance learning* problem. A set of multiple instances is named *bag*.

In binary pattern recognition with class labels $\{+1, -1\}$, a bag will be classified as positive if at least one of its instances is positive. It is negative if all of its instances are negative. This is also called the *MI assumption* [13]. This assumption is too general for safety-related applications where the final model must be highly interpretable. Thence we will tighten the MI assumption to have a MIL framework for the present domain. Before we will introduce our assumption and a possible approach to tackle safety-related problems, let us briefly recall MIL and safety-related applications.

1.1 Multiple-Instance Learning

In multiple-instance problems, one single training example (a positive or negative bag) is constituted by many different feature vectors, so-called *instances*. At least one is

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 409–416, 2008.
springerlink.com

responsible for the observed class of the given example. Hence the class label is attached to the bag instead of the instances themselves.

Let us denote positive bags as B_i^+ and the jth observation of this bag as $\mathbf{x}_{ij}^+ \in \mathbb{R}^n$ where n is the dimensionality of the input space $\mathscr{X}$. The bag B_i^+ consists of l_i^+ instances $\mathbf{x}_{ij}^+$ for $j = 1,\ldots,l_i^+$. Consequently, the ith negative bag is denoted by B_i^-, its jth observation by $\mathbf{x}_{ij}^-$. Likewise, l_i^- symbolizes the number of instances in this negative bag. We denote the number of positive and negative bags as N^+ and N^-. The overall number of instances is referred to $l = l^+ + l^- = \sum_{i=0}^{N^+} l_i^+ + \sum_{i=0}^{N^-} l_i^-$. Thus the sample of all instances in negative and positive bags is listed by $\mathbf{x}_1,\ldots,\mathbf{x}_l$.

Nowadays many learning problems have been treated as MI problems, i.e., drug activity prediction [5, 6], stock market prediction [7], image retrieval [14, 15], natural scene classification [7], text categorization [1], and image categorization [4]. With the application to safety-related domains, another type of problem is identified as MI formulation under certain requirements.

1.2 Safety-Related Applications

Safety-related applications can be found in many real-world problems, e.g., condition monitoring of plants, automobiles, airplanes, and trains. These systems are supervised by many sensors collecting a (nearly) continuous multidimensional signal in time, e.g., speed, temperature, pressure, global position. Every time series itself describes one certain event of multiple instances. Regarding the MIL setting, we can state that every event corresponds to one bag which is either positive (e.g., a machine breakdown, alarm) or negative (e.g., proper machine operation, no-alarm). Thus it is necessary to binary classify these multivariate time series.

No instance in time of a negative bag must be classified as positive. A false positive in such an application usually involves severe injuries or harm to humans or machines. On the other hand, all positive events or bags have to be correctly classified before a certain limiting time has passed (e.g., time to exchange a machine before breakdown). If a positive event is recognized early enough, then certain countermeasures can be performed to prevent or moderate heavy accidents. These requirements meet the MIL setting. The following ones tighten the general MI assumption.

Since tests of such complex systems are very expensive and thus quite rare, there does not exist a vast of data (especially positive events). Hence a main disadvantage in those domains is the fact that formal proofs of the correctness of the learned classifier are not feasible [9]. Therefore, the model has to be enriched by experts' knowledge to ensure security requirements. Furthermore, we find unbalanced misclassification costs in safety-related domains very often s.t. constraints have to be added to the model as well.

It is not trivial to find the best trade-off between accuracy and model complexity. There exist some classifier for instance, support vector machine (SVM), that implicitly tries to satisfy both criteria. Taking advantages of the SVM's flexibility, we can even incorporate knowledge to meet unbalanced misclassification costs. This soft computing method will be introduced as extension of a linear separating hyperplane classifier in Sect. 2.

In safety-related applications, the model complexity in terms of simple functional dependencies is frequently the most important point. Quasilinear functions[1] with a good generalization performance must be found to establish a physical interpretation of human experts. Section 3 describes requirements for simple models regarding SV machines. After that, a combination of two methods is proposed to obtain a somehow simple and still accurate classifier. We conclude and discuss potential future work in Sect. 4.

2 Support Vector Machines

Let us formally introduce the basic concepts that we are going to talk about. Suppose we are given the input space $\mathscr{X}$ (not necessarily a vector space) and the output space $\mathscr{Y}$. Since we deal with a binary classification problem, $\mathscr{Y} = \{\pm 1\}$. We observe l training patterns $(x_i, y_i) \in \mathscr{S} \subseteq \mathscr{X} \times \mathscr{Y}$ where $i = 1, \ldots, l$. They have been drawn i.i.d. from an unknown distribution. If $\mathscr{X} \subset \mathbb{R}^n$, then $x_i \mapsto \mathbf{x}_i$. Our goal is to separate the data with a linear hyperplane $\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle + b = 0\}$ where $\mathbf{w} \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are the norm vector and the bias of the hyperplane, respectively. The decision function of a hyperplane classifier which shall predict y' for any $\mathbf{x}$ corresponds to

$$f(\mathbf{x}) = \operatorname{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b). \tag{1}$$

We are looking for the hyperplane that maximizes the margin between every training pattern and the hyperplane. Such a hyperplane is called optimal since it is unique and has the best generalization performance on unseen data. If all points $(x_i, y_i) \in \mathscr{S}$ can be separated linearly by a hyperplane, we can obtain the optimal hyperplane by solving a quadratic optimization problem with linear inequality constraints. Usually not all training patterns can be separated perfectly. Therefore we introduce slack variables ξ_i with $i = 1, \ldots, l$ in order to relax the optimization problem to

$$\underset{\mathbf{w}, b, \xi}{\text{minimize}} \ \tau(\mathbf{w}, \xi) = \tfrac{1}{2} \|\mathbf{w}\| + C \textstyle\sum_{i=1}^{l} \xi_i \tag{2}$$

$$\text{subject to} \quad y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \tag{3}$$

$$\text{and} \quad \xi_i \geq 0, \ \forall i = 1, \ldots, l. \tag{4}$$

Here, $\xi = (\xi_1, \ldots, \xi_l)$ corresponds to the slack variables ξ_i and C is a global parameter that has to be determined by the user. The bigger C, the easier training patterns may violate the constraint (3). By introducing the Lagrangian of the primal problem (2), we end up solving the dual

$$\underset{\alpha_1, \ldots, \alpha_l}{\text{maximize}} \ \textstyle\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,i'=1}^{l} y_i y_{i'} \alpha_i \alpha_{i'} \langle \mathbf{x}_i, \mathbf{x}_{i'} \rangle \tag{5}$$

$$\text{subject to} \quad \textstyle\sum_{i=1}^{l} y_i \alpha_i = 0 \tag{6}$$

$$\text{and} \quad 0 \leq \alpha_i \leq C, \ \forall i = 1, \ldots, l. \tag{7}$$

In practice, only few problems can be solved by a linear classifier. Hence the problem has to be reformulated in a nonlinear way. This is done by mapping the input space $\mathscr{X}$ to

[1] Quasilinear functions should be monotonic. They can be approximated with very few linear functions.

some high-dimensional feature space $\mathscr{H}$ by $\Phi : \mathscr{X} \mapsto \mathscr{H}$ where Φ satisfies Mercer's condition [11]. We can thus solve our nonlinear optimization problem linearly in $\mathscr{H}$ by computing the scalar product $K(x,x') = \langle \Phi(x), \Phi(x') \rangle$ which is called kernel. We simply replace the occurrence of the scalar product in (5) with a chosen kernel function. Finally, the discrimination function (1) becomes $f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i K(x,x_i) + b\right)$.

For our purpose, let us have a look at the following two kernel functions[2]. First of all, we can apply the linear kernel

$$K(\mathbf{x},\mathbf{x}') = \langle \mathbf{x},\mathbf{x}' \rangle = \sum_{d=1}^{n} [\mathbf{x}]_d [\mathbf{x}']_d. \tag{8}$$

which performs the identical mapping $\Phi : \mathscr{X} \mapsto \mathscr{X}$. Second, kernel functions $K(\mathbf{x},\mathbf{x}') = K(\|\mathbf{x}-\mathbf{x}'\|)$ generate radial basis functions e.g., the Gaussian kernel

$$K(\mathbf{x},\mathbf{x}') = \exp\left(-\gamma \left\|\mathbf{x}-\mathbf{x}'\right\|^2\right). \tag{9}$$

3 Quasilinear Support Vector Machines

With respect to the SVM, the "linearity" of a SVM is expressed by the capacity of the function (so-called hypothesis) chosen by the principle of structural risk minimization (SRM) [12]. This principle of minimizing the expected risk controls the capacity s.t. the chosen hyperplane will guarantee the lowest error on unseen instances. Thus it heavily influences the complexity of the SVM.

The classification problem we deal with does not demand to correctly classify all positive instances. A positive bag will be correctly discriminated by at least one of its instances. On the contrary, all instances of negative bags have to be correctly recognized. Thence a pruning of positive examples that are hard to classify before the actual training is a way to simplify the process of model selection. The actual decision function will be selected by any suitable binary classifier e.g., SVM. This classifier might select a quasilinear hypothesis since conflicting positive instances have been removed.

Using a SVM to classify the pruned instance, we must choose an appropriate kernel function. Without having knowledge about the underlying distribution that generates the data, the Gaussian kernel (9) has shown good results in practice [11]. It is based on the Euclidean distance metric and thus intuitive. Due to the interpretability constraint in safety-related domains, $\mathscr{H}$ should geometrically correspond to $\mathscr{X}$. In [11] the authors argue that the linearity of the Gaussian kernel only depends on γ. For small γ, the SVM will determine a quasilinear discriminant function. A rather large γ causes narrow kernels which lead to complex nonlinear functions in $\mathscr{X}$.

The linearity of a SV machine is a necessity to accept and approve our model. By decreasing the number of possible hypothesis, the potential solution becomes probably more linear and thus less complex. However, a too simple model might not generalize well on unseen bags. Thus the hyperplane must have the local ability to become more complex in order to ensure a higher accuracy for crucial bags.

[2] See [11] for a collection of kernel functions and further details on SVMs.

3.1 Support Vector Pruning

A learning machine that discriminates events of security-related systems must be rather simple to be approved of security standards. Easier machines are favored instead of more complex ones. The SVM principle is theoretically well motivated, however, the feature space $\mathscr{H}$ is never expressed explicitly. If we construct $\mathscr{H}$ geometrically similar to $\mathscr{X}$ without using the linear kernel (8), it is possible to understand the resulting machine to a higher degree.

Quasilinear classifiers are preferred to complex models by pruning instances that are very hard to classify by a linear SVM. Since we deal with security-related domains, it is strictly forbidden to prune negative instances. Pruning is, however, feasible for positive bags (cf. Sect. 3). It removes candidates for critical instances from the dataset. It does not prune complete bags since every bag corresponds to a real-world event that has to be recognized. Thence at least m instances of every positive bag are kept even if they would have been linearly misclassified.

The pruning process is motivated by the search for a quasilinear classifier since linear dependencies are geometrically easy to interpret. Furthermore, misclassified points will automatically become support vectors. The farthest positive SVs on the negative side of the hyperplane have a big influence on the model selection step. It is particularly very probable that those points will become support vectors even with a more sophisticated kernel.

The pruning can be explained briefly by the following 4 procedures:

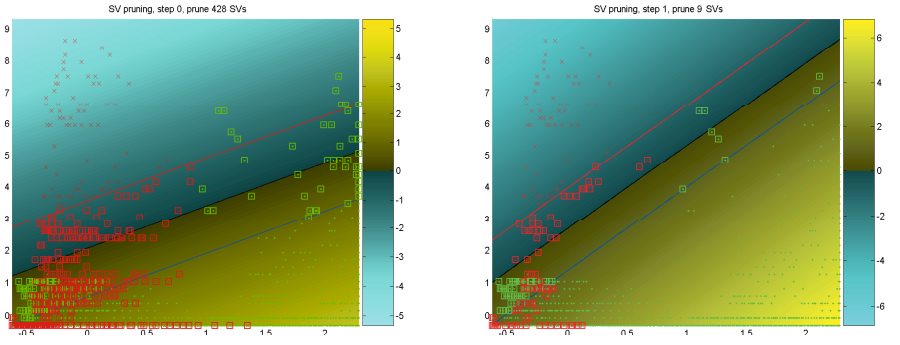
1. Train a linear SV machine with all positive and negative patterns.
2. Identify misclassified positive support vectors.
3. Create a training set without these positive samples.
4. Repeat training until a stable model is obtained.

The third procedure has to assure that none of the bags will get empty. This is done by only pruning the farthest wrong positive support vectors of every bag s.t. the number of remaining instances is at least m. After all bags have been inspected, a new linear classifier is trained. The procedure begins again until no SV has been pruned. This algorithm converges relatively fast after approximately 6 iterations.

Fig. 1 shows an artificial application of SV pruning. The training of the linear SV machines has been performed with $C = 10$. At least $m = 10$ instances of every bag had to remain after pruning. In the first step, more than 400 instances of some positive bags have been pruned. Then 9 instances have been removed and thus not less than one linear SVM would have been trained for further pruning.

3.2 Bag Weighting

Quasilinear SV machines are very nice to have. On the contrary, the model shall still deploy all positive bags and prevent deployment of every negative bag. Thus a trade-off between simplicity and complexity has to be found. This section will introduce a modification of the standard SVM i.e., we locally allow the discriminant function to become more complex.



(a) (b)

Fig. 1. Example of SV pruning's first two steps on an artificial MI problem. Positive (negative) instances are shown as red crosses (green dots). The black line represents the decision boundary. The red (blue) line symbolizes the class margin of 1. Positive (negative) SVs are distinguished by red (green) squares around their instances. The color legends on the right side of the plots clarify the distance to the hyperplane. (a) Initial step found 428 positive instances for pruning. (b) Second step with pruned dataset determined 9 further positive instances which will be removed.

Reviewing (2) we find the global parameter C that expresses misclassification costs of all patterns. In particular, there is no a priori preference or priority of any pattern. Thence solving

$$\underset{\mathbf{w},b,\xi}{\text{minimize}}\ \tau(\mathbf{w},\xi) = \tfrac{1}{2}\|\mathbf{w}\| + C\sum_{i=1}^{l} C_i\xi_i$$

subject to (3) and (4), we introduce weights C_i for $1 \leq i \leq l$. It is straightforward to assign weights to complete bags as well. The user can influence the learning step by incorporating experts' knowledge in form of bag weights. The choice of the C_i is performed heuristically since these weights differ from problem to problem.

In combination with SV pruning, bag weighting can be a powerful tool to ensure both a quite simple model and the fulfillment of customer requirements i.e., high accuracy. It might be a good procedure to first apply the pruning method and then assign weights to misclassified bags. Remaining conflicts due to global model simplicity might thus be either removed or resolved.

4 Conclusions

In this paper we presented an hybrid approach for preprocessing MI problems in safety-related domains. Whereas classifiers for standard MI datasets aim to be as accurate as possible, we focused on learning machines of which model simplicity is essential. We introduced SV pruning to favor quasilinear classifiers. Bag weighting has been suggested to enable both the input of expert's knowledge and the trade-off between model simplicity and accuracy. The presented idea has been successfully applied to a safety-related system in automobile industry [8]. Due to the nondisclosure of this project, however, its empirical evaluation cannot be presented.

There are many possible extensions and improvements to the proposed methods. We will focus our research on generating fuzzy rules based on SV learning since fuzzy classifiers have been successfully implemented in safety-related applications (see [9]). Some approaches recently came up to construct fuzzy graphs from support vectors [2, 3, 10].

Preprocessing bags by SV pruning and bag weighting can be the basis for the following approach. The SVM could directly output fuzzy rules. Therefore, one would either have to formulate a differentiated optimization problem or define a special kernel that already includes domain knowledge. Both ways might result in understandable fuzzy rules that still guarantee a good generalization. In addition, SV machines allow domain experts to comprise their knowledge to model learning. The whole concept might establish a powerful framework to find less complex classifiers not only in safety-related domains.

Acknowledgement. We would like to thank Eyke Hüllermeier who inspired us to cast safety-related problems into the framework of multiple-instance learning. Furthermore we acknowledge the critical and helpful comments of two anonymous reviewers to finalize this paper.

References

1. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Proceedings of the NIPS-2002 Conference. Advances in Neural Information Processing Systems, MA, USA, vol. 15, pp. 561–568. MIT Press, Cambridge (2003)
2. Chaves, A., Vellasco, M.B.R., Tanscheit, R.: Fuzzy rule extraction from support vector machines. In: Nedjah, N., Mourelle, L., Abraham, A., Köppen, M. (eds.) Proceedings of Fifth International Conference on Hybrid Intelligent Systems (HIS2005, Rio de Janeiro, Brazil), pp. 335–340. IEEE Computer Society Press, Los Alamitos (2005)
3. Chen, Y., Wang, J.Z.: Support vector learning for fuzzy rule-based classification systems. IEEE FS 11(6), 716–728 (2003)
4. Chen, Y., Wang, J.Z.: Image categorization by learning and reasoning with regions. J. Mach. Learn Res. 5, 913–939 (2004)
5. Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence 89(1-2), 31–71 (1997)
6. Maron, O., Lozano-Pérez, T.: A framework for multiple-instance learning. In: Jordan, M.I., Kearns, M.J., Solla, S.A. (eds.) Proceedings of the NIPS-1997 Conference. Advances in Neural Information Processing Systems, vol. 10, pp. 570–576. MIT Press, Cambridge (1998)
7. Maron, O., Ratan, A.: Multiple-instance learning for natural scene classification. In: Shavlik, J. (ed.) Proceedings of the 15th International Conference on Machine Learning (ICML-98, Madison, WI), pp. 341–349. Morgan Kaufmann, San Francisco (1998)
8. Moewes, C.: Application of Support Vector Machines to Discriminate Vehicle Crash Events. Master's Thesis. FIN, University of Magdeburg (2007)
9. Otte, C., Nusser, S., Hauptmann, W.: Machine learning methods for safety-related domains: Status and perspectives. In: Proceedings of the Symposium on Fuzzy Systems in Computer Science (FSCS 2006, Magdeburg, Germany), pp. 139–148 (2006)
10. Papadimitriou, S., Terzidis, C.: Efficient and interpretable fuzzy classifiers from data with support vector learning. Intell. Data Anal. 9(6), 527–550 (2005)

11. Schölkopf, B., Smola, A.J.: Learning with Kernels. In: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge (2002)
12. Vapnik, V., Chervonenkis, A.: Theory of Pattern Recognition [in Russian], Nauka, Moscow; Wapnik, W., Tscherwonenkis, A.: Theorie der Zeichenerkennung. Akademie–Verlag, Berlin (1979) (German Translation)
13. Xu, X.: Statistical Learning in Multiple Instance Problems. Master's Thesis. Department of Computer Science, University of Waikato (2003)
14. Yang, C., Lozano-Pérez, T.: Image database retrieval with multiple-instance learning techniques. In: Proceedings of the 16th International Conference on Data Engineering (ICDE, San Diego, CA), pp. 233–243 (2000)
15. Zhang, Q., Goldman, S.A., Yu, W., Fritts, J.E.: Content-based image retrieval using multiple-instance learning. In: Sammut, C., Hoffmann, A.G. (eds.) The 19th International Conference on Machine Learning (ICML 2002, Sydney, Australia), pp. 682–689. Morgan Kaufmann, San Francisco (2002)

Decision-Making

An Efficient Normal Form Solution to Decision Trees with Lower Previsions

Nathan Huntley and Matthias C.M. Troffaes

Department of Mathematical Sciences, Durham University, Durham, United Kingdom

Abstract. Decision trees are useful graphical representations of sequential decision problems. We consider decision trees where events are assigned imprecise probabilities, and examine their normal form decisions; that is, scenarios in which the subject initially decides all his future choices. We present a backward induction method for efficiently finding the set of optimal normal form decisions under maximality. Our algorithm is similar to traditional backward induction for solving extensive forms in that we solve smaller subtrees first, however it is different in that solutions of subtrees are only used as intermediate steps to reach the full solution more efficiently—in particular, under maximality, a decision that is optimal in a subtree can be potentially absent in any optimal policy in the full tree.

1 Introduction

This paper studies sequential decision problems: a subject must choose from a finite set of options, where each option has a finite number of mutually exclusive consequences, possibly random, that lead to either rewards or new decisions. Based on his preferences between rewards and beliefs about the consequences, the subject tries to determine his optimal policy. Such problems are often represented by *decision trees*, which are graphical representations of relationships between the choices, consequences, and rewards [2, 5, 6].

Usually, decision trees are modelled by assigning a precise probability to each event, and a precise utility to each reward, and the policy yielding maximal expected utility is elected. Finding this optimal policy can be done in two ways (equivalent for precise probabilities): the extensive form and the normal form. We present our interpretation of these forms.

Extensive form decision making is concerned with solving a decision problem by eliminating decision branches that are by some method judged non-optimal. Upon reaching a decision node, the subject is then free to choose from the remaining decision branches. Normal form decision making involves specifying all possible decision paths for every possible combination of events, and then aims to reduce that set. Such a decision path for each possible combination of events will be called a *normal form decision* or a *policy*.

In many cases of interest, a subject's knowledge is too limited to specify a probability for all events. Following [9], in such cases we can still represent a subject's knowledge by specifying a so-called coherent lower prevision, or equivalently, a set of probability distributions.

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 419–426, 2008.
springerlink.com

In such situations, it is known that normal and extensive forms are not equivalent [3, 7]. In particular, natural ways of using backward induction to solve extensive forms potentially leaves one open to taking bad decisions. We shall therefore, for this paper, focus on solving the normal form.

Traditionally, one aims to arrive at the single optimal normal form decision which maximises expected utility. Arguably, however, if for instance only a set of probability distributions can be identified, not all policies may be comparable, and, therefore, decision criteria that apply in such cases typically determine a *set* of optimal policies rather than electing a single one. How can a subject determine optimal policies given limited knowledge? Some solutions have been proposed in the literature [3, 4, 7], however, a systematic study of solving decision trees under partially ordered preferences is still lacking.

In principle, one can find the optimal policies by first listing all normal form decisions, and then for instance applying Walley's *maximality* [9, §3.9.2]. However, it is well known that the number of normal form decisions typically grows too large too quickly with the size of the tree. Therefore, we propose a recursive algorithm, which is similar to traditional backward induction for solving extensive forms in that we solve smaller subtrees first, however our recursive algorithm is different from traditional backward induction in that solutions of the subtrees are only used as intermediate steps to reach the full solution more efficiently—in particular, under maximality, a decision that is optimal in a subtree can be potentially absent in any optimal policy in the full tree. This contrasts with backward induction algorithms such as in [4], in which the solutions of the subtrees are indeed part of the final solution.

The key result which leads to our algorithm is that, for maximality, a non-optimal normal form decision in a subtree cannot be part of an optimal normal form decision in the full tree. This allows us to eliminate many decisions in the full tree early on without further computation.

Section 2 introduces coherent lower previsions and an optimality criterion, maximality. Section 3 explains decision trees. Section 4 explains how to reduce a tree to normal form. Section 5 provides the main result of the paper and a simple example. Section 6 concludes the paper.

2 Coherent Lower Previsions

Let Ω be a finite set of possible states of the world. Elements of Ω are called *outcomes* and are denoted by ω. Subsets of Ω are called *events*, and are denoted by A, B, etc. Bounded real-valued functions on Ω are called *gambles* when interpreted as an uncertain payoff expressed in utiles. So, a gamble X pays a reward of $X(\omega)$ if ω turns out to be the true outcome. The set of all possible gambles on Ω is denoted by $\mathscr{L}(\Omega)$.

It can be argued that even when a subject has only little information, his/her information about ω can be represented by means of a *coherent lower prevision* $\underline{P}$ on $\mathscr{L}(\Omega)$, through a process called *natural extension* [1, 9]. For the purpose of this paper, it is sufficient to assume a subject can model his information through a compact convex set $\mathscr{M}$ of (finitely additive) probability measures, and [9, p. 135] $\underline{P}(X) = \min_{\mu \in \mathscr{M}} \int X \, \mathrm{d}\mu$. In other words, coherent lower previsions correspond to lower envelopes of expectations.

For simplicity, assume the lower probability of all conditioning events is strictly positive, so the conditional lower prevision is [9, p. 298, §6.4.2]

$$\underline{P}(X|B) = \min_{\mu \in \mathscr{M}} \frac{\int BX \, \mathrm{d}\mu}{\mu(B)}.$$

When probabilities are precisely known, it is rational to choose gambles that maximise expectation [8]. With lower previsions, there is no expectation, so we need a different notion of optimality. Several have been proposed, and in this paper we consider one called maximality [9].

For an event A and gambles X and Y, say $X >_{\underline{P}|A} Y$ if $\underline{P}(X - Y\,|\,A) > 0$. For a set of gambles $\mathscr{X}$, a gamble $X \in \mathscr{X}$ is said to be *maximal given* A if there is no $Y \in \mathscr{X}$ such that $Y >_{\underline{P}|A} X$:

$$\mathrm{opt}(\mathscr{X}|A) = \{X \in \mathscr{X} : (\forall Y \in \mathscr{X})(Y \not>_{\underline{P}|A} X)\}.$$

Because $>_{\underline{P}|A}$ is a partial order, maximality leads to a set of optimal gambles rather than electing a single one as is usually the case when maximising expectation. Therefore maximality yields a more realistic approach to normative decision making when information is limited.

3 Decision Trees

A decision tree [2, 5] is a graphical causal representation of decisions, outcomes, payoffs (in utiles), and probabilities in a decision problem. They consist of decision nodes and chance nodes, growing from left to right, with time flowing from left to right, and with payoffs at the end of each branch.

Consider the following problem, which is both simple enough to admit easy study, and complex enough to demonstrate all concepts involved. Tomorrow, a subject is going for a walk in the lake district. Tomorrow, it may rain (ω_1), or not (ω_2). The subject can either take an umbrella (d_1), or not (d_2). But the subject may also choose to buy today's newspaper to learn about tomorrow's weather forecast (d_B), or not ($d_{\overline{B}}$), before leaving for the lake district. For the sake of simplicity, we assume that the forecast can have one of two outcomes: predicting rain (B_1), or not (B_2). The utility of each combination, if the subject does not buy the newspaper, is summarised in Table 1. If the subject buys the newspaper, then a cost c is subtracted from the utilities.

The decision tree corresponding to this example is depicted in Fig. 1. Decision nodes are depicted by squares, and chance nodes by circles. From each node, branches emerge. For decision nodes, each branch corresponds to a decision; for chance nodes, each branch corresponds to an event. In the lake district problem, the subject is first confronted with the decision to buy the newspaper or not, hence the tree starts off with

Table 1. Payoff table for the rain and umbrella problem

	rain (ω_1)	no rain (ω_2)
umbrella (d_1)	-5	0
no umbrella (d_2)	-10	5

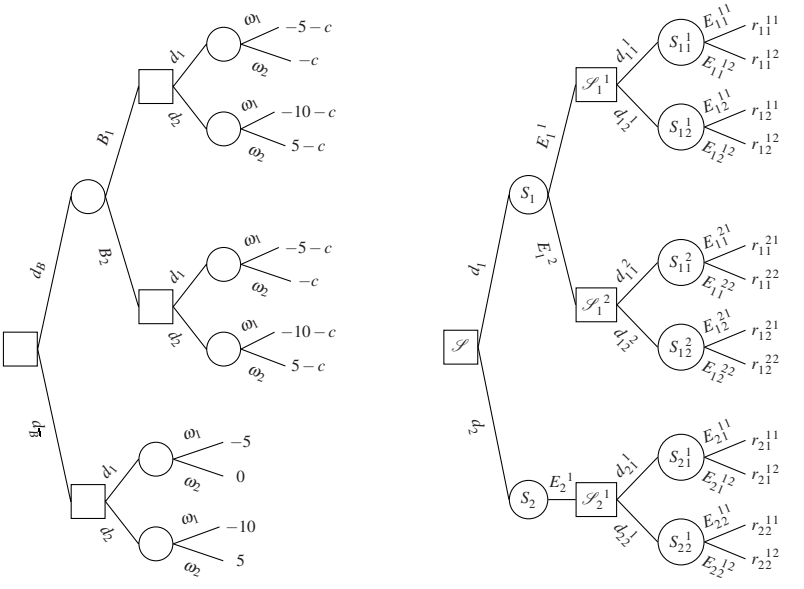


Fig. 1. The lake district decision tree (left), and its standard form (right)

a decision node. If the subject buys the newspaper (d_B), then it can inform him about tomorrow's weather forecast. Thus, the chance node following the subject's decision d_B has two branches, forecasting rain (B_1), or no rain (B_2). Next, when the subject leaves for the lake district, he can either take his umbrella with him (d_1) or not (d_2), hence the decision node following B_1. Finally, during the walk, it either rains (ω_1) or not (ω_2), as depicted by a chance node for each possible combination of events and decisions preceding the actual walk.

So, each path in a decision tree corresponds to a particular sequence of decisions and events, with resulting payoff at the end.

Before elaborating on solving decision trees, we introduce some notation. First, let us assume that the tree is of the following standard form:

- The root node of a decision tree must be a decision node. If not, add a dummy decision node before it.
- A decision node may not directly succeed another decision node. Separate successive decision nodes by a dummy chance node.
- A chance node may not directly succeed another chance node. Merge successive chance nodes.
- All paths from the root node to a reward node must pass through the same number of nodes. Add dummy decision and chance nodes to the shorter branches in a way that does not violate the above rules.

As an example, applying these transformations is shown in Fig. 1.

This transformation allows a natural way of labelling events and decisions. Decisions correspond to arcs from decision nodes to chance nodes, and events correspond to arcs from chance nodes to decision nodes.

- Label decisions at the root node by $d_1, d_2, \ldots$.
- Label events at the chance node reached by decision d_i by ${E_i}^1, {E_i}^2, \ldots$.
- Continue this method of labelling until all arcs are labelled.

We also label nodes similarly. Label a decision node with $\mathscr{S}$, and a chance node with S, using the same sub- and superscript notation as for the arcs. So, the root is $\mathscr{S}$, its children S_i, grandchildren $\mathscr{S}_i^j$, and so on.

4 Reducing Decision Trees to the Normal Form

A normal form decision specifies beforehand how the subject will act in any event. The events then obtain, and at each decision node the subject takes the prescribed decision. The reward received at the end depends on the events obtained, and therefore a normal form decision simply amounts to a gamble.

Formally, we define the normal form $\mathscr{N}$ of a decision tree to be the set of gambles induced by all normal form decisions. It is useful to consider normal forms of subtrees too, using the same sub- and superscripts as used for nodes. For instance, ${\mathscr{N}_1}^1$ is the normal form of the subtree at ${\mathscr{S}_1}^1$ in Fig. 1.

For example, at S_1, the subject has already chosen decision d_1. His normal form $\mathscr{N}_1$ at this node must specify his future policy before observing which event obtains. So he must choose between the following four options:

$$d_{11}^{\ 1} \text{ and } d_{11}^{\ 2}, \qquad d_{11}^{\ 1} \text{ and } d_{12}^{\ 2}, \qquad d_{12}^{\ 1} \text{ and } d_{11}^{\ 2}, \qquad d_{12}^{\ 1} \text{ and } d_{12}^{\ 2}.$$

To solve decision trees in the normal form, we must first find the gamble to which each normal form decision corresponds. First of all, a final chance node simply amounts to a decision yielding a gamble mapping its branches—the events—to the corresponding rewards. For instance, on Fig. 1, the chance node $S_{11}^{\ 2}$ yields (identifying events ${E_*}^*$ with their indicator functions)

$$\mathscr{N}_{11}^{\ 2} = \{X_{11}^{\ 2}\} = \{{E_{11}}^{21} r_{11}^{\ 21} + {E_{11}}^{22} r_{11}^{\ 22}\}.$$

Next, imagine we are at node ${\mathscr{S}_1}^2$ and choose $d_{11}^{\ 2}$: then we receive the gamble $X_{11}^{\ 2}$. On the other hand, we could also have chosen $d_{12}^{\ 2}$, in which case we receive the gamble $X_{12}^{\ 2}$. So, at decision node ${\mathscr{S}_1}^2$, we may choose between

$${\mathscr{N}_1}^2 = \{X_{11}^{\ 2}, X_{12}^{\ 2}\} = \mathscr{N}_{11}^{\ 2} \cup \mathscr{N}_{12}^{\ 2}.$$

In other words, the normal form set at a decision node is the union of the normal form sets of its children.

At the intermediate chance node S_1, either ${E_1}^1$ or ${E_1}^2$ may obtain, and depending on which obtains, either ${\mathscr{N}_1}^1$ or ${\mathscr{N}_1}^2$ becomes available. Hence,

$$\mathscr{N}_1 = \{{E_1}^1 {X_1}^1 + {E_1}^2 {X_1}^2 \colon {X_1}^1 \in {\mathscr{N}_1}^1, {X_1}^2 \in {\mathscr{N}_1}^2\}.$$

Finally, $\mathscr{N}$ again corresponds to the union $\mathscr{N}_1 \cup \mathscr{N}_2$. Summarising,

- $\mathscr{N}_{*i}^{*} = \{X_{*i}^{*}\}$ with $X_{*i}^{*} = \sum_j E_{*i}^{*j} r_{*i}^{*j}$ for final chance nodes,
- $\mathscr{N}_{*}^{*} = \bigcup_i \mathscr{N}_{*i}^{*}$ for decision nodes,
- $\mathscr{N}_{*i}^{*} = \left\{ \sum_j E_{*i}^{*j} X_{*i}^{*j} \colon X_{*i}^{*j} \in \mathscr{N}_{*i}^{*j} \right\}$ for intermediate chance nodes,

where $\mathscr{N}_{*}^{*}$ refers to the normal form of a subtree at a decision node, and $\mathscr{N}_{*i}^{*}$ refers to the normal form of a subtree at a chance node. The stars in $\mathscr{N}_{*}^{*}$ emphasise that these equations hold for normal forms at any subtree, and represent previous decisions and events required to reach $\mathscr{N}_{*}^{*}$.

5 Backward Induction to Solve the Normal Form

Let A_{*}^{*} be the union of all events preceding a decision node. For instance, in Fig. 1, ${A_1}^2 = {E_1}^2$ and $A = \Omega$. So A_{*}^{*} summarises the events we have observed at $\mathscr{S}_{*}^{*}$ and S_{*i}^{*}. The optimal set of policies is therefore $\operatorname{opt}(\mathscr{N}_{*}^{*}|A_{*}^{*})$ at decision nodes $\mathscr{S}_{*}^{*}$, and $\operatorname{opt}(\mathscr{N}_{*i}^{*}|A_{*}^{*})$ at chance nodes S_{*i}^{*}.

The traditional backward induction algorithm connects the optimal solution of a tree with the optimal solutions of subtrees further down the tree. The elements of $\mathscr{N}_{*}^{*}$ are gambles of the form $\sum_j E_{*i}^{*j} X_{*i}^{*j}$ for $X_{*i}^{*j} \in \mathscr{N}_{*i}^{*j}$. Under expected utility, elements of $\operatorname{opt}(\mathscr{N}_{*}^{*}|A_{*}^{*})$ are gambles of the form $\sum_j E_{*i}^{*j} X_{*i}^{*j}$ for $X_{*i}^{*j} \in \operatorname{opt}(\mathscr{N}_{*i}^{*j}|A_{*i}^{*j})$. So knowing the optimal solutions of subtrees to the right allows us to immediately eliminate many non-optimal elements of $\mathscr{N}_{*}^{*}$, and more quickly find the optimal solutions of a large tree.

It turns out that we can do something similar if our information is described by a coherent lower prevision, and using (conditional) maximality:

Theorem 1. *The following equations hold.*

$$\operatorname{opt}(\mathscr{N}_{*}^{*}|A_{*}^{*}) \subseteq \bigcup_i \operatorname{opt}(\mathscr{N}_{*i}^{*}|A_{*}^{*}) \tag{1}$$

$$\operatorname{opt}(\mathscr{N}_{*i}^{*}|A_{*}^{*}) \subseteq \left\{ \sum_j E_{*i}^{*j} X_{*i}^{*j} \colon X_{*i}^{*j} \in \operatorname{opt}(\mathscr{N}_{*i}^{*j}|A_{*i}^{*j}) \right\} \tag{2}$$

To show how the theorem leads to an efficient backward induction algorithm, we solve the decision tree in Fig. 1 for a linear-vacuous mixture

$$\underline{P}(X) = (1-\varepsilon)P_0(X) + \varepsilon \inf X$$

where $P_0(X)$ is an expectation and $\varepsilon \in (0,1)$. The conditional is [9, p. 309]

$$\underline{P}(X|E) = \frac{(1-\varepsilon)P_0(EX) + \varepsilon \inf_{\omega \in E} X(\omega)}{(1-\varepsilon)P_0(E) + \varepsilon}.$$

Let $P_0(\omega_1) = 1/2$, $P_0(\omega_1|B_1) = 7/10$, and $P_0(\omega_1|B_2) = 2/10$, so $P_0(B_1) = 3/5$. Let $\varepsilon = 1/10$, and do not specify c.

First, the tree is transformed (Fig. 1). Obviously, at the final chance nodes $\operatorname{opt}(\mathscr{N}_{i\,k}^{\,j}|A_i^{\,j}) = \mathscr{N}_{i\,k}^{\,j} = \{X_{ik}^{\,j}\}$. Now find $\operatorname{opt}(\mathscr{N}_i^{\,j}|A_i^{\,j})$. At $\mathscr{N}_2^{\,1}$, ${A_2}^1 = \Omega$ and

$$\underline{P}(X_{21}^1 - X_{22}^1) = \underline{P}(X_{22}^1 - X_{21}^1) = -5\varepsilon = -\frac{1}{2},$$

so both decisions are in $\text{opt}(\mathscr{N}_2^1|A_2^1)$. Next, $\text{opt}(\mathscr{N}_2|A) \subseteq \{X_{21}^1, X_{22}^1\}$ by (2), but there is no preference between these, so $\text{opt}(\mathscr{N}_2|A) = \{X_{21}^1, X_{22}^1\}$.

Considering $\mathscr{N}_1^1$ and $\mathscr{N}_1^2$, we have $A_1^1 = E_1^1$, $A_1^2 = E_1^2$, and

$$\underline{P}(X_{11}^1 - X_{12}^1|A_1^1) = \tfrac{6-31\varepsilon}{3+2\varepsilon} > 0, \qquad \underline{P}(X_{12}^2 - X_{11}^2|A_1^2) = \tfrac{6-31\varepsilon}{2+3\varepsilon} > 0$$

for $\varepsilon = 1/10$, and thus $\text{opt}(\mathscr{N}_1^1|A_1^1) = \{X_{11}^1\}$ and $\text{opt}(\mathscr{N}_1^2|A_1^2) = \{X_{12}^2\}$. By (2) we have that $\text{opt}(\mathscr{N}_1|A) \subseteq \{E_1^1 X_{11}^1 + E_1^2 X_{12}^2\}$, and since the right hand side is a singleton, equality holds.

Finally, by (1), $\text{opt}(\mathscr{N}|A) \subseteq \{E_1^1 X_{11}^1 + E_1^2 X_{12}^2, X_{21}^1, X_{22}^1\}$:

$$\begin{aligned}
\underline{P}\left(X_{21}^1 - (E_1^1 X_{11}^1 + E_1^2 X_{12}^2)\right) &= c - (6+19\varepsilon)/5 = c - 79/50 \\
\underline{P}\left(E_1^1 X_{11}^1 + E_1^2 X_{12}^2 - X_{21}^1\right) &= -c + (6-31\varepsilon)/5 = -c + 29/50 \\
\underline{P}\left(X_{22}^1 - (E_1^1 X_{11}^1 + E_1^2 X_{12}^2)\right) &= c - (6+19\varepsilon)/5 = c - 79/50 \\
\underline{P}\left(E_1^1 X_{11}^1 + E_1^2 X_{12}^2 - X_{22}^1\right) &= -c + (6-31\varepsilon)/5 = -c + 29/50.
\end{aligned}$$

Concluding, if the newspaper costs less than $29/50$, then we buy and follow its advice. If it costs more than $79/50$, then we do not buy, but still we have insufficient information to decide whether to take the umbrella or not. If it costs between $29/50$ and $79/50$, then we have insufficient information to choose between the three remaining options.

6 Conclusion

We presented a recursive method for finding the normal form solution of a decision tree when information is insufficient to specify a unique probability measure. In such a case, the optimal solution is usually a set of normal form decisions rather than a single one. Similar to classical backward induction, our method can often be much more efficient than comparing all normal form decisions. Besides maximality, our method also works with E-admissibility, but not with Γ-maximinity nor with interval dominance.

However, we do not argue that the normal form always yields the best solution to a sequential decision problem. Indeed, the normal form solution might have undesirable properties. For instance, a decision in a subtree can be affected by options refused some time in the past. Due to lack of space we leave a discussion of these intriguing matters to another paper.

References

1. Boole, G.: An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities. Walton and Maberly, London (1854)
2. Clemen, R.T., Reilly, T.: Making Hard Decisions. Duxbury Press, Belmont (2001)
3. Jaffray, J.: Rational decision making with imprecise probabilities. In: Proceedings of 1st International Symposium on Imprecise Probabilities and Their Applications, pp. 183–188 (1999)

4. Kikuti, D., Cozman, F., de Campos, C.P.: Partially ordered preferences in decision trees: Computing strategies with imprecision in probabilities. In: Proceedings of Multidisciplinary Workshop on Advances in Preference Handling (IJCAI-05, Edinburgh, UK), pp. 118–123 (2005)
5. Lindley, D.V.: Making Decisions, 2nd edn. Wiley, London (1985)
6. Raiffa, H., Schlaifer, R.: Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration. Harvard University, Boston (1961)
7. Seidenfeld, T.: A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus E-admissibility. Synthese 140, 69–88 (2004)
8. Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1953)
9. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)

Group Decision Making with Soft AHP Based on the Random Set View of Fuzzy Sets

Pedro Terán and José María Moreno-Jiménez

Grupo Decisión Multicriterio Zaragoza, Facultad de Ciencias Económicas y Empresariales, Universidad de Zaragoza, Zaragoza, Spain

Abstract. We present a variant of the Analytic Hierarchy Process intended to facilitate consensus search in group decision making. This soft methodology combines fuzzy sets and probabilistic information to provide judgements oriented by the actors' attitude towards negotiation. A Monte Carlo approach is taken to derive a preference structure distribution which should finally be studied to extract knowledge about the resolution process.

1 Introduction

The Analytic Hierarchy Process [15], AHP, is a multicriteria decision making technique that provides in an absolute scale the priorities corresponding to the alternatives being compared in a context with multiple scenarios, multiple actors and multiple criteria.

Its methodology consists on three stages: (i) modelling, (ii) valuation and (iii) priorization and synthesis. In the first stage, a hierarchy of the relevant aspects of the problem is constructed. In the second stage, the preferences of the actors involved in the resolution process are elicited by means of reciprocal pairwise comparison matrices using judgements based on the fundamental scale $\{1/9, 1/8, \ldots, 1, \ldots, 8, 9\}$ proposed by Saaty [15]. Finally, in the third stage, local, global and total priorities are obtained.

In its initial formulation (Conventional AHP), the valuation process is a deterministic one. However, most real applications require considering actors' uncertainty when comparing tangible and intangible aspects. There are a number of procedures to deal with the uncertainty inherent in the judgement elicitation process. Interval judgements [12], reciprocal random distributions [6] and fuzzy numbers [10, 4, 5] are some of the most extended procedures.

Using fuzzy judgements to elicit the actors' preferences, we present a new approach to include the actors' attitude in the negotiation process in AHP-group decision making. This complements, in the fuzzy setting, the recent Bayesian approach to Stochastic AHP in [2]. To incorporate their attitude, we associate a probability distribution to the α-level parameter. Jointly considered, the fuzzy judgements and the α-level distributions are built into a soft AHP allowing us to deal with the AHP-group decision making problem in a more realistic and effective way than traditional approaches [16, 14, 9].

The two traditional approaches followed in AHP-group decision making are aggregation of individual judgements (AIJ) and aggregation of individual priorities (AIP). Other approaches can be found in [7, 13]. In the AIJ procedure, a judgement matrix for

D. Dubois et al. (Eds.): Soft Methods for Hand. Var. and Imprecision, ASC 48, pp. 427–434, 2008.
springerlink.com

the group is constructed from the individual judgements and group priorities are calculated from it. The three most commonly used methods to determine the entries in the group judgement matrix are consensus, voting and aggregation of judgements. In AIP, group priorities are computed from individual priorities using an aggregation method. In both procedures, the most widely used aggregation technique is the weighted geometric mean.

The paper is structured as follows. After this brief Introduction, Section 2 describes in an intuitive way the new approach based on the well-known relationships between fuzzy sets and random sets. Section 3 explains how individual judgements for the negotiation are elicited. Section 4 shows how to obtain the preference structure distribution on the possible rankings of the alternatives. Finally, Section 5 suggests how to exploit this distribution from a learning perspective.

2 A Soft AHP Approach

The proposed approach extends AHP methods which allow imprecise judgements in the form of real intervals. That extension to group decision problems aims at incorporating the actors' attitude towards negotiation with emphasis on consensus search. The basic notion is that actors may accept enlarging their interval judgements, moving farther from their personal judgement, in an attempt to find overlap areas of larger compatibility with the others' views. The role of the analyst is to facilitate the process and extract knowledge from the problem resolution.

Briefly, the steps are as follows.

First, actors elicit pairwise comparison matrices whose elements are fuzzy intervals. These basic judgements fix the framework for the process, establishing the less imprecise position matching to the actor's ideas, a more imprecise interval with the maximal admissible concessions and a continuum of intermediate positions. The underlying fuzzy sets semantics is that of preference: the membership function denotes how well a number qualifies as an acceptable quantification of the actor's judgement of relative importance.

Second, for a specific negotiation process, the actors decide, on the basis of subjective factors and interests, the kind of position to be adopted: tougher or more open. That attitude towards the negotiation is represented by a probability distribution on the interval $[0,1]$ of membership values. The negotiation weight distribution assesses more weight to the positions more comfortable or convenient to the actor.

Within a fixed context, e.g. in decisions repeated over time, basic judgements may remain the same while the negotiation weight distribution varies with the circumstances of each negotiation.

Third, both kinds of information are fused, using the notions from random set theory and its connection to fuzzy sets, so that basic judgements are revised yielding adequate negotiation judgements. Now, the correct semantic interpretation of these 'posterior' fuzzy judgements is the possibility semantics.

Interval methods cannot capture the graduality which appears naturally in this setting. The proposed method is soft since it conjugates several approaches to imprecision and uncertainty, even with different semantic interpretations. Since it does not rely

exclusively on Fuzzy Set Theory and does not lead to fuzzy priorities, we would call it a 'Soft AHP' method, rather than a 'Fuzzy AHP' method.

The next stage concern the aggregation of the actors' views and the analysis of the possible preference structures are more supported by the imprecise judgements emitted. The preference structure distribution should be analyzed to gain understanding of the decision and negotiation process and detect patterns, something harder or not possible in methods which ultimately provide a single ranking.

3 Individual Negotiation Judgements

We start with m actors who approach the decision process willing to make a consensus decision. Each actor has a weight $\beta_k > 0$ in the decision, with $\sum_k \beta_k = 1$. For simplicity of presentation, we assume a local context, i.e. a single criterion. The input for the analysis is a matrix of fuzzy pairwise comparisons between alternatives and a probability distribution encoding each actor's potential framework and attitude towards negotiation, respectively.

Each actor provides a matrix $A^{[k]} = [a_{ij}^{[k]}]_{i,j=1}^{n}$, *whose entries are fuzzy intervals.*

The core and support of the fuzzy interval represent the most and less restrictive positions actor k might be willing to adopt. Thus we also call it a *basic judgement*. The remaining α-cuts

$$(a_{ij}^{[k]})_\alpha = \{x \mid a_{ij}^{[k]}(x) \geq \alpha\}$$

represent intermediate positions between those two extremes. As α approaches 0, consensus becomes easier since the overlap between the actors' positions increases.

In practice, the $a_{ij}^{[k]}$ can often be chosen to be trapezoidal fuzzy sets for half the entries of $A^{[k]}$, but not for all of them since that would lead to a violation of the reciprocity property of AHP. The other entries are determined by reciprocity between $a_{ij}^{[k]}$ and $a_{ji}^{[k]}$, so that, for each fixed $\alpha \in [0,1]$,

$$\min(a_{ij}^{[k]})_\alpha = (\max(a_{ji}^{[k]})_\alpha)^{-1}, \quad \max(a_{ij}^{[k]})_\alpha = (\min(a_{ji}^{[k]})_\alpha)^{-1}.$$

Each actor provides a probability distribution $P^{[k]}$ *on the real interval* $[0,1]$.

After basic judgements are elicited, for one particular negotiation each actor chooses a distribution on $[0,1]$ according to his specific attitude towards that negotiation. This *negotiation weight distribution* can be given in the form of a density function with support $[0,1]$ and determines the sort of position which will be given more preponderance in the negotiation. Distributions concentrated close to 1 represent tougher positions with little room for concessions, while distributions concentrated close to 0 represent very open positions primarily willing to ease consensus, even if reached farther from the actor's ideal position.

Intuitively, the density function should be unimodal in the sense of being a quasiconvex function. It represents weighting, rather than random behaviour.

For clarity of presentation, assume that the mode $m^{[k]}$ is unique. Then, $(a_{ij}^{[k]})_{m^{[k]}}$ represents the central position of actor k. Level sets $(a_{ij}^{[k]})_\alpha$ for $\alpha < m^{[k]}$, being longer

intervals, represent less demanding positions given a lesser weight by the negotiation weight distribution $P^{[k]}$. Actors may or may not choose $m^{[k]} = 1$. In general, they need not, since it is unclear that an actor's dominating attitude in judgement modelling will bring the decision process to a more satisfactory conclusion for him. Inversely, α-cuts for $\alpha > m^{[k]}$ represent more stringent positions which are given smaller weight too.

Once both $A^{[k]}$ and $P^{[k]}$ are fixed, we must merge those two pieces of information. We will do that by using consonant random sets.

Indeed, each fuzzy basic judgement $a_{ij}^{[k]}$, together with the negotiation weight distribution $P^{[k]}$, easily provides a random set (a random interval) which is consonant, i.e. monotonic. We just have to take the level mapping $L_{ij}^{[k]}$ defined on the interval $[0,1]$, endowed with the probability measure $P^{[k]}$, and with interval values given by

$$L_{ij}^{[k]}(\alpha) = (a_{ij}^{[k]})_\alpha, \quad \alpha \in [0,1].$$

It must be stressed that different choices of basic judgements and negotiation weight distribution may encode the same information. For any increasing bijective transformation $\phi : [0,1] \to [0,1]$, the pair $([\phi \circ a_{ij}^{[k]}]_{i,j}; P^{[k]} \circ \phi^{-1})$ represents the same information as $(A^{[k]}; P^{[k]})$. Therefore, the procedure is invariant under increasing bijective transformations of the scale interval $[0,1]$, a nice property from the measurement-theoretical point of view.

In turn, all the information of that random set is contained in its one-point coverage function $\pi_{ij}^{[k]}$ given by

$$\pi_{ij}^{[k]}(x) = P^{[k]}(x \in a_{ij}^{[k]}).$$

We call $\pi_{ij}^{[k]}$ a *negotiation judgement* or *final judgement*. Observe that $\pi_{ij}^{[k]}$ can be reinterpreted as a fuzzy set, by invoking again the connection between random sets and fuzzy sets.

Let us show how negotiation judgements combine the information in $a_{ij}^{[k]}$ and $P^{[k]}$.

Denote by $F^{[k]}$ the distribution function of $P^{[k]}$. Then, one can prove that

$$\pi_{ij}^{[k]}(x) = F^{[k]}(\max\{\alpha \in [0,1] \mid x \in (a_{ij}^{[k]})_\alpha\}) = F^{[k]}(a_{ij}^{[k]}(x)).$$

If $P^{[k]}$ is given by a density function with full support, as seems reasonable, then $F^{[k]}$ is invertible and a classical theorem of Probability Theory tells us that $P^{[k]} \circ (F^{[k]})^{-1}$ is a uniform distribution in $[0,1]$. Therefore, we have

$$\pi_{ij}^{[k]} = F^{[k]} \circ a_{ij}^{[k]},$$

$$P^{[k]} \circ (F^{[k]})^{-1} \sim \mathscr{U}[0,1].$$

Taking $\phi = F^{[k]}$ above, we deduce that the pair $([\pi_{ij}^{[k]}]_{i,j}; \mathscr{U}[0,1])$ contains the same information as the original pair $(A^{[k]}; P^{[k]})$. But since the uniform distribution gives equal weight to each α, all the information is now in the $\pi_{ij}^{[k]}$.

It is possible to compare the actors' positions via $\pi_{ij}^{[k]}$, since it recasts the information in a common scale, with uniform weighting for all actors. In this representation, a fuzzy

set very steep in the area surrounding the $F^{[k]}(m^{[k]})$-cut means that actor k strongly wishes to remain close to his central position, while a more flexible position would be characterized by a fast 'opening' towards larger intervals for $\alpha < F^{[k]}(m^{[k]})$.

In order to simplify the elicitation process, the analyst may predetermine the shape of the fuzzy intervals and the density function so that only a few parameters, easily interpretable, are left for actors to specify. One possible way is as follows.

The frame judgements $a_{ij}^{[k]}$ are taken to be trapezoidal, so that only the end-points of their core and support must be elicited. Note that both intervals need not have the same center. In some situations, it may be easier to elicit the support end-points indirectly by indicating the percentage of the corresponding core end-point the actor might eventually be willing to concede.

For the negotiation weight distribution $P^{[k]}$, the simplest choice is a triangular distribution, which is determined once the mode $m^{[k]}$ is specified. The value $m^{[k]}$ reflects intuitively the attitude toward negotiation, with $m^{[k]} = 1$ representing a tough attitude and $m^{[k]} = 0$ a fully open one. Trapezoidal distributions are possible as well.

Another possibility for the $P^{[k]}$ is the beta $\beta(p,q)$ family of distributions. Appropriate choices of p,q control not only the position of the center of the distribution but also its dispersion around the actor's central position.

4 The Preference Structure Distribution

Our final aim is to quantify how much support receives each possible ranking (preference structure) of the alternatives in view of the information collected so far. A way to overcome the difficulty to solve the problem analitically is to simulate by Monte Carlo methods many crisp judgement matrices which are compatible with the positions expressed by the actors.

We begin by choosing a random value η in [0,1] according to a uniform distribution. For each k, we select the η quantile of the negotiation weight distribution $P^{[k]}$,

$$q^{[k]} = (F^{[k]})^{-1}(\eta).$$

Then we perform simulations to select crisp values

$$\xi_{ij}^{[k]} \in (a_{ij}^{[k]})_{q^{[k]}} = (\pi_{ij}^{[k]})_{\eta}.$$

It is enough to simulate only for those entries of the matrix which were directly chosen by actor k. For the rest of the matrix, reciprocity is enforced by the relationship

$$\xi_{ji}^{[k]} = (\xi_{ij}^{[k]})^{-1}.$$

A uniform distribution or another distribution, if deemed appropriate, can be used. That overcomes some problems with reciprocity appearing in many variants of Fuzzy AHP.

The latter part is analogous to known stochastic methods to solve AHP with imprecise judgements [12]. The computational complexity is the same as for those interval methods, since obtaining a crisp judgement matrix involves $(n-1)n/2$ simulations per actor and only one additional simulation is needed to fix $q^{[k]}$.

Note that simulation is applied at a different height level $q^{[k]}$ for each actor. In the long run, the Glivenko-Cantelli Theorem ensures that, for each actor, the empirical distribution approximates the weights provided by actor k, as the number of simulations increases.

Once Monte Carlo judgement matrices are obtained, well-established methods for Group Decision Making with AHP can be used to obtain the preference structure distribution. For the sake of completeness, we describe a possible continuation of the analysis until its conclusion.

Each actor k and each simulated crisp judgement matrix $[\xi_{ij}^{[k]}]_{i,j}$ provide a vector of priority values for the alternatives. There are several methods for obtaining the priorities, and several ways to aggregate individual preferences. We suggest the methods based on the geometric mean, for their good properties in the group decision setting. Barzilai and Golany [3] proved that AIJ using the weighted geometric mean method (WGGM) followed by derivation of priorities by the rowwise geometric mean method (RGGM) yields the same result than derivation of priorities by RGGM followed by AIP by WGGM. Moreover, Escobar et al. [8] showed that AIJ has good properties with respect to consistency, in that the aggregate judgement matrix tends to decrease the inconsistency levels of the less consistent actors.

For instance, in the AIJ method we calculate the matrix Ξ^G of aggregate group judgements

$$\xi_{ij}^G = \prod_{k=1}^{m} (\xi_{ij}^{[k]})^{\beta_k}, \quad i,j = 1,\ldots,n.$$

Then, priorities for the alternatives are derived as

$$\omega_i^G = \prod_{k=1}^{m} (\xi_{ij}^G)^{1/k}, \quad i = 1,\ldots,n.$$

Alternatives are ranked according to the values ω_i^G. There are $n!$ possible rankings or preference structures, which can be identified with permutations of n elements.

After sufficiently many simulations, we end up with an empirical distribution on preference structures. For each possible preference structure $\mathscr{R}$, it gives us the proportion $\lambda_{\mathscr{R}}$ of samples leading to that ranking.

5 Exploiting the Model

From the standpoint that we should seek to extract knowledge from the resolution of the decision problem, the preference structure distribution contains rich information which should be explored in search for patterns, see [7].

Visual methods for representing the group information, e.g. [17], provide a starting point for exploring the preference structures. Individual preferences can be compared to group preferences to detect similarities and patterns. Graphical and statistical tools such as clustering, fuzzy clustering and multidimensional scaling are appropriate for this stage of the analysis, see e.g. [11]. Our research group (GDMZ) is currently working in this area with application to large e-democracy and e-cognocracy decision problems.

A reasonable approach to synthesizing the information in the preference structure distribution goes by applying voting methods well-studied in Social Choice Theory. In order to take into account the information contained in individual rankings, methods using the whole ranking seem more appropriate. An example is the Borda count method, other methods are available.

With the Borda method, the best alternative in a preference structure is given n points, the second best $n-1$ points, and so on. Each preference structure has its weight given in the preference structure distribution, resulting

$$v_i = \sum_{\mathscr{R}} \lambda_{\mathscr{R}}[(n+1) - \mathscr{R}(i)], \quad i = 1, \ldots, n.$$

Alternatives can be ranked or chosen according to the values v_i, which result from aggregation over all $n!$ preference structures.

An alternative to AIJ and AIP allowing interval judgements is the AIPS (aggregation of individual preference structures) method in [7]. In that paper, preference structures are calculated for each actor, then aggregated, allowing to compare each actor's preference structure to the group's.

Acknowledgement. Research partially supported by the Spanish *Ministerio de Educación y Ciencia* (TSI2005–02511 and MTM2005–02254), and the *Gobierno de Aragón* (PM2007–034).

References

1. Aguarón, J., Moreno-Jiménez, J.M.: The geometric consistency index: approximated thresholds. European J. Oper. Res. 147, 137–145 (2003)
2. Altuzarra, A., Moreno-Jiménez, J.M., Salvador, M.: Consensus building in AHP- group decision making: A Bayesian approach. (submitted for publication) (2008)
3. Barzilai, J., Golany, B.: AHP rank reversal normalization and aggregation rules. INFOR 32, 57–64 (1994)
4. Buckley, J.J.: Fuzzy hierarchical analysis. Fuzzy Sets Syst. 17, 223–247 (1985)
5. Csutora, R., Buckley, J.J.: Fuzzy hierarchical analysis: the Lambda-Max method. Fuzzy Sets Syst. 120, 181–195 (2001)
6. Escobar, M.T., Moreno-Jiménez, J.M.: Reciprocal distributions in the Analytic Hierarchy Process. European J. Oper. Res. 123, 154–174 (2000)
7. Escobar, M.T., Moreno-Jiménez, J.M.: Aggregation of individual preference structures in AHP-Group Decision Making. Special issue on Frontiers in GDN research of Group Decis Negot 16, 287–301 (2007)
8. Escobar, M.T., Aguarón, J., Moreno-Jiménez, J.M.: A note on AHP group consistency for the row geometric priorization procedure. European J. Oper. Res. 153, 318–322 (2004)
9. Forman, E., Peniwati, K.: Aggregating individual judgments and priorities with the Analytic Hierarchy Process. European J. Oper. Res. 108, 165–169 (1998)
10. Laarhoven, P.J.M., Pedrycz, W.: A fuzzy extension of Saaty's priority theory. Fuzzy Sets Syst. 11, 229–241 (1983)
11. Gargallo Valero, P., Moreno-Jiménez, J.M., Salvador Figueras, M.: AHP-Group Decision Making: A Bayesian approach based on mixtures for group pattern identification. Group Decis. Negot. 16, 485–506 (2007)

12. Moreno-Jiménez, J.M., Vargas, L.G.: A probabilistic study of preference structures in the Analytic Hierarchy Process with interval judgments. Math. Comput. Modelling 17, 73–81 (1993)
13. Moreno-Jiménez, J.M., Aguarón, J., Escobar, M.T.: The core of consistency in AHP-Group Decision Making. Group Decis. Negot. 17(3), 249–265 (2008)
14. Ramanathan, R., Ganesh, L.S.: Group preference aggregation methods employed in AHP: an evaluation and intrinsic process for deriving members weightages. European J. Oper. Res. 79, 249–265 (1994)
15. Saaty, T.L.: Multicriteria Decision Making: the Analytic Hierarchy Process (2nd edn. in 1990, RSW, Pittsburgh) McGraw-Hill, New York (1980)
16. Saaty, T.L.: Group decision-making and the AHP. In: Golden, B.L., Wasil, E.A., Harker, P.T. (eds.) The Analytic Hierarchy Process: Applications and Studies, pp. 59–67. Springer, Heidelberg (1989)
17. Turón, A., Moreno-Jiménez, J.M.: Graphical visualization tools in AHP-Group Decision Making. In: Proceedings of Group Dec. and Negotiation 2006 Intl. Conf. IISM, Karlsruhe, Germany, pp. 150–152 (2007)

Author Index