Metrika (2000) 51: 201–221

Metrika

On random fuzzy variables of second order and their application to linear statistical inference with fuzzy data

Wolfgang Näther

TU Bergakademie Freiberg, Fakultät für Mathematik und Informatik, 09596 Freiberg, Germany (e-mail: naether@math.tu-freiberg.de)

Received: November 1999

Abstract. This paper summarizes some results on random fuzzy variables with existing expectation and variance, called random fuzzy variables of second order. Using the Frechét-principle and – via support functions – the embedding of convex fuzzy sets into a Banach space of functions it especially presents a unified view on expectation and variance of random fuzzy variables. These notions are applied in developing linear statistical inference with fuzzy data. Detailed investigations are presented concerning best linear unbiased estimation in linear regression models with fuzzy observations.

Key words: Expectation and variance of random fuzzy variables, linear regression with fuzzy data

1 Introduction and overview

Consider the clouding y for given atmospheric pressure x. Our experience is that y cannot be predicted exactly for given x. Therefore y is modelled by a suitable random variable Y. Assume further that the observation scheme is of such kind that the clouding Y is recorded by linguistic (and more or less vague) expressions like "Cloudless", "Clear", "Fair", "Cloudy" and "Overcast". Thus, we have to consider Y as a random variable with vague outcomes. This is an example for a so-called random fuzzy variable and we may ask e.g. for a suitable regression model for the relation between Y and x (see section 3.2).

Also in other real situations uncertainty of data comes from two sources: from randomness and from fuzziness. Randomness models the stochastic variability of all possible outcomes of an experiment and fuzziness describes the vagueness of the given or just realized outcome. Randomness is more an instrument of a normative analysis which thinks about the question "What will happen in future?"; fuzziness is more an instrument of a descriptive analysis reflecting questions like "What has happened?" or "What is meant by the data?".

Vague outcomes of an experiment can be described by fuzzy sets. Following Zadeh (1965), a fuzzy set A on $\mathbb{R}^d$ is identified by its membership function $m_A \mid \mathbb{R}^d \to [0, 1]$ where $m_A(x)$ is interpreted as the degree of acceptance that $x \in \mathbb{R}^d$ is a member of A. Obviously, for sets $A \subseteq \mathbb{R}^d$ in the usual sense, called crisp sets in fuzzy set theory, the membership function m_A coincides with the indicator function of A.

The crisp set

$$A_\alpha \stackrel{\text{def}}{=} \{x \in \mathbb{R}^d : m_A(x) \geq \alpha\}, \quad 0 < \alpha \leq 1 \tag{1}$$

is called the α-cut of A. For $\alpha = 0$ define: $A_{\alpha=0} \stackrel{\text{def}}{=} \text{closure}\{x \in \mathbb{R}^d : m_A(x) > 0\}$. A fuzzy set A is called convex and compact if all α-cuts A_α have this property. A is called normal if $A_1 = \{x \in \mathbb{R}^d : m_A(x) = 1\} \neq \emptyset$.

Basic to fuzzy set theory is Zadeh's extension principle

$$m_{g(A_1,\ldots,A_k)}(z) = \sup_{\substack{x_1,\ldots,x_k: \\ g(x_1,\ldots,x_k)=z}} \inf\{m_{A_1}(x_1), \ldots, m_{A_k}(x_k)\} \tag{2}$$

which provides a general method to extend classical functions $g(x_1, \ldots, x_k)$ on $\mathbb{R}^k$ to allow for fuzzy input $A_1, \ldots, A_k$. Given the fuzzy input, (2) presents the membership function of the fuzzy image by g.

In modelling realistic situations, fuzziness is often tied to randomness since possible random outcomes have to be described by fuzzy sets, especially in the case of linguistically expressed outcomes. Consider the introductory example. "Classical" statisticians probably consider data like "Cloudless", "Clear", "Flair", "Cloudy" and "Overcast" as ordinal data and "defuzzify" them in a formal way by numbers, for example by 0, 1, 2, 3, 4. Now the basis for statistical inference is an arbitrary classical discrete random variable.

The advantage of considering random fuzzy variables is the following: At the level of experimentation it is often possible to use the experience of the experimenter for a justified modelling of the fuzziness of the outcomes. Statistical inference with random fuzzy variables transfers the fuzziness, e.g. into parameter estimators. Now, at this level, the level of decision, it may be necessary to defuzzify the vague parameter estimate, but it can be done in a more responsible way as on the level of experimentation because consequences of wrong decisions can be taken into account. Hence, the vagueness of experimental outcomes is carried with a random fuzzy variable and the associated statistical procedures up to the level of decision. Roughly speaking, the "philosophy" of random fuzzy variables is: Do not defuzzify at the level of experimentation, take care of a fair transfer of fuzziness and defuzzify – if necessary – at the level of decision.

There are several concepts of a random fuzzy variable Y.

Following Kwakernaak (1978), a random fuzzy variable Y is considered as a vague perception of a crisp but unobservable random variable X. A typical example for Kwakernaak's approach is the following (see Kruse/Meyer (1987)): The currentage X of a randomly chosen mayfly is an ordinary random variable on the positive real line. But, since mayflies have no certificate of birth, we only can perceive a random variable Y through a set of "windows" like "Young", "Middle age" and "Old", i.e. we perceive fuzzy sets as observation results since the original X is not observable.

A conceptional different approach is given by the so called probabilistic fuzzy sets introduced by Hirota (1981). There Y is considered, first of all, as a fuzzy set with random membership values, i.e. m_Y can be considered as a random process on $\mathbb{R}^d$ with values in [0,1] (which describes, for example, the grey values of X-ray photographs).

In this paper, however, we mainly will consider (randomly chosen) real vague objects. Consider e.g. the notion "nose". There is no natural crisp boundary between "nose" and "cheek" or "nose" and "forehead" and there is no possibility to reduce this vagueness by more precise observations. Consider a population of people. Then the "nose" of a randomly chosen individual is a random fuzzy variable Y but without a crisp original X behind them. The most successful approach to random fuzzy variables of this kind was presented by Puri/Ralescu (1986) where Y is considered as a fuzzification of a random set (therefore, sometimes Y is called random fuzzy set, too). An exact definition is the following:

Definition 1. *Denote by $\mathscr{F}_C^d$ the set of all normal compact convex fuzzy subsets of $\mathbb{R}^d$. Let $(\Omega, \mathscr{B}, P)$ be a probability space. Then, $Y \mid \Omega \to \mathscr{F}_C^d$ is called a* random fuzzy variable *(rfv) on $\mathbb{R}^d$ if for any $\alpha \in [0,1]$ the α-cut Y_α is a convex compact random set (e.g. in the sense of Matheron (1975)).*

One essential advantage of the Puri/Ralescu-approach is the embedding of the concept of a rfv into the well-developped concept of random sets. So, complicate measurability considerations which are necessary e.g. for Kwakernaak's approach, can be avoided. Let us remind the reader that a compact random set M is a random variable for which for any compact set K all "hit-events" $M \cap K \neq \emptyset$ and all "miss-events" $M \cap K = \emptyset$ are measurable. More exactly: M can be characterized by the so called capacity functional $c(K) \overset{\text{def}}{=} P(M \cap K \neq \emptyset)$. For an overview on random sets Stoyan (1998) is recommended.

The concept of rfv's in the sense of Def. 1 has been studied successfully, e.g. for limit theorems (starting with Klement/Puri/Ralescu (1986)) and has been applied to asymptotical statistics with vague data by Kruse/Meyer (1987).

Our attention is focussed on so-called rfv's of second order, i.e. on rfv's with existing expectation and variance. This paper summarizes some earlier results on this topic from a unified point of view and is organized as follows:

In section 2, a well justified approach for defining expectation and variance of rfv's is presented. As a guide, we use the so called Frechét-principle (see subsection 2.1). Using their support functions convex (fuzzy) sets can be embedded isomorphically in a Banach space of functions (see Radström (1952)). Therefore a convex rfv can be identified with a special Banach-space-valued random variable. From this point of view concepts for expectation and (co)-variance of rfv's can be deduced from the corresponding well-defined notions for Banach-space-valued random variables. Some special properties of expectation and variance of rfv's as well as estimators of them are presented in subsection 2.5.

In section 3 we refer to some results on rfv's of second order, e.g. to strong laws of large numbers and to asymptotical tests of hypotheses on the expectation. Some more detailed, we discuss ideas to develop some kind of linear statistical inference for rfv's of second order, especially for linear regression with random fuzzy observations. Unfortunately, a straightforward generalization of known classical results on this topic is not possible since $\mathscr{F}_C^d$ is not a linear space w.r.t. (later defined) addition and scalar multiplication of fuzzy sets.

2 Expectation and variance of random fuzzy variables

2.1 The Frechét-principle

For defining expectation and variance of a rfv Y the Frechét-principle is used as a methodological principle. Frechét (1948) has defined the expectation $\mathbb{E}^{(d)}Z$ for a random variable Z with values in a metric space (M, d) as a (not necessary unique) solution of the problem

$$\mathbb{E}d(Z, \mathbb{E}^{(d)}Z)^2 = \inf_{a \in M} \mathbb{E}d(Z, a)^2. \tag{3}$$

Note that $\mathbb{E}d(Z, a)^2$ is the usual expectation of the real-valued variable $d(Z, a)^2$. The variance of Z, denoted by $\mathbf{Var}^{(d)}Z$, is then defined by

$$\mathbf{Var}^{(d)}Z = \mathbb{E}d(Z, \mathbb{E}^{(d)}Z)^2. \tag{4}$$

This is a generalization of the known fact that for a real-valued random variable X the expectation $\mathbb{E}X$ minimizes $\mathbb{E}|X - a|^2$ and $\mathbf{Var}\,X$ equals $\mathbb{E}|X - \mathbb{E}X|^2$.

In the following, $\mathbb{E}^{(d)}Z$ satisfying (3) is called Frechét-expectation w.r.t. d. For rfv Y, the Frechét approach opens the way for defining several types of expectations and (via (4)) their associated variances, each induced by a given metric between fuzzy sets. Therefore, first of all, we have to discuss on suitable distances between fuzzy sets.

2.2 Distances between fuzzy sets

It seems to be natural to start with the Hausdorff-metric between crisp sets $A, B \subset \mathbb{R}^d$, given by

$$d_H(A, B) = \max\left\{ \sup_{b \in B} \inf_{a \in A} \|a - b\|, \sup_{a \in A} \inf_{b \in B} \|a - b\| \right\}.$$

Using the α-cuts (1), for two fuzzy sets A, B this can be generalized to

$$d_p(A, B) = \begin{cases} \left(\int_0^1 (d_H(A_\alpha, B_\alpha))^p d\alpha \right)^{1/p}, & p \in [1, \infty) \\ \sup\limits_{\alpha \in [0,1]} d_H(A_\alpha, B_\alpha), & p = \infty \end{cases}$$

where especially d_1 and d_∞ are investigated in the literature. For example, $(\mathcal{F}_C^d, d_1)$ appears as a complete and separable metric space, $(\mathcal{F}_C^d, d_\infty)$, however, is a complete but non-separable metric space (see Puri/Ralescu (1986)).

Another type of distances can be defined via support functions. For any compact convex set $A \subset \mathbb{R}^d$ the support function s_A is defined as

$$s_A(u) = \sup_{y \in A} \langle u, y \rangle; \quad u \in S^{d-1},$$

where $\langle \cdot , \cdot \rangle$ is the scalar product in $\mathbb{R}^d$ and S^{d-1} the $(d-1)$-dimensional unit sphere in $\mathbb{R}^d$. Note that for convex and compact $A \subset \mathbb{R}^d$ the support function s_A is uniquely determined. A fuzzy set $A \subset \mathcal{F}_C^d$ can be characterized α-cut-wise by its support function:

$$s_A(u, \alpha) \overset{\text{def}}{=} s_{A_\alpha}(u); \quad \alpha \in [0, 1], \quad u \in S^{d-1}. \tag{5}$$

Thus, via support functions (5), $\mathcal{F}_C^d$ can be embedded in a space of functions on $S^{d-1} \times [0, 1]$ and we can define a metric in $\mathcal{F}_C^d$ using e.g. a special L_2-metric in $L_2(S^{d-1} \times [0, 1])$, i.e.

$$\delta_2(A, B) = \sqrt{d \int_0^1 \int_{S^{d-1}} (s_A(u, \alpha) - s_B(u, \alpha))^2 v(du)\, d\alpha}, \tag{6}$$

where v is the Lebesgue measure on S^{d-1}. Note that $(\mathcal{F}_C^d, \delta_2)$ is complete and separable (see Diamond/Kloeden (1994)).

As an example, consider so-called LR-fuzzy numbers $A \overset{\text{def}}{=} (\mu, l, r)_{LR}$ with central value $\mu \in \mathbb{R}^1$, left and right spread $l \in \mathbb{R}^+$ and $r \in \mathbb{R}^+$, decreasing left and right shape functions $L|\mathbb{R}^+ \to [0, 1]$, $R|\mathbb{R}^+ \to [0, 1]$ with $L(0) = R(0) = 1$ and finite support, i.e. a fuzzy set A with

$$m_A(x) = \begin{cases} L\left(\dfrac{\mu - x}{l}\right) & \text{if } x \le \mu \\ R\left(\dfrac{x - \mu}{r}\right) & \text{if } x \ge \mu. \end{cases} \tag{7}$$

Note that the α-cuts of $A = (\mu, l, r)_{LR}$ are given by the intervals

$$A_\alpha = [\mu - L^{-1}(\alpha)l, \mu + R^{-1}(\alpha)r]; \quad \alpha \in [0, 1]$$

and that the support function of an interval is defined on $S^0 = \{-1, 1\}$ by

$$s_{[a,b]}(u) = \begin{cases} -a & \text{if } u = -1 \\ b & \text{if } u = 1. \end{cases}$$

An LR-fuzzy number $A = (\mu, l, r)_{LR}$ with $L = R$ and $l = r \overset{\text{def}}{=} \Delta$ is called symmetric and abbreviated by $A \overset{\text{def}}{=} (\mu, \Delta)_L$.

Then, we have especially for two symmetric fuzzy numbers $A = (\mu_A, \Delta_A)_L$, $B = (\mu_B, \Delta_B)$:

$$d_H(A_\alpha, B_\alpha) = |\mu_A - \mu_B| + L^{-1}(\alpha)|\Delta_A - \Delta_B|$$

$$d_1(A,B) = |\mu_A - \mu_B| + l_1|\Delta_A - \Delta_B|; \quad l_1 = \int_0^1 L^{-1}(\alpha)\,d\alpha$$

$$d_\infty(A,B) = |\mu_A - \mu_B| + l_\infty|\Delta_A - \Delta_B|; \quad l_\infty = \sup_{\alpha\in[0,1]} L^{-1}(\alpha) \tag{8}$$

$$\delta_2(A,B) = \sqrt{(\mu_A - \mu_B)^2 + l_2(\Delta_A - \Delta_B)^2}; \quad l_2 = \int_0^1 (L^{-1}(\alpha))^2\,d\alpha. \tag{9}$$

2.3 Expectation

Asking for laws of large numbers for rfv's, Puri/Ralescu (1986) have proposed to use the so-called Aumann-expectation as a suitable expectation of rfv's. The Aumann-expectation goes back to the paper Aumann (1965) on integrals of set valued functions and is defined as follows:

Definition 2. *Let Y be a rfv on $\mathbb{R}^d$ with $\mathbb{E}s_Y < \infty$. The* Aumann-expectation *of Y is defined as the fuzzy set $\mathbb{E}^{(A)}Y \in \mathcal{F}_C^d$ with*

$$\forall\alpha \in [0,1] : (\mathbb{E}^{(A)}Y)_\alpha = \mathbb{E}^{(A)}Y_\alpha$$

where $\mathbb{E}^{(A)}Y_\alpha$ is the Aumann-expectation of the random set Y_α defined by

$$\mathbb{E}^{(A)}Y_\alpha = \{\mathbb{E}\eta : \eta(\omega) \in Y_\alpha(\omega) P\text{-a.e. and } \eta \in L^1(\Omega, \mathcal{B}, P)\}$$

Note that $\mathbb{E}s_Y < \infty$ ensures the existence of $E^{(A)}Y$ and that the Aumann-expectation of the convex compact random set Y_α is the set of all (usual) expectation of random vectors η in $\mathbb{R}^d$ which P-a.e. lie in M. Often such η is called a selector of Y_α.

It should be emphasized that there are also other proposals to define the expectation of a rfv. E.g. for a rfv in form of a probabilistic fuzzy set Y an expectation $\mathbb{E}^{(m)}Y$ can be defined via the expected membership function, i.e. (see Hirota (1981))

$$m_{\mathbb{E}^{(m)}Y}(x) = \mathbb{E}m_Y(x); \quad x \in \mathbb{R}^d,$$

which in general does not coincide with Def. 2. There are further definitions of expectations for random sets (see e.g. Molchanov (1993), Stoyan/Stoyan (1994)) which can be used for further alternative α-cut-wise definitions of expectations for rfv's.

Now we will discuss the reasons why the Aumann-expectation is preferable.

Denote by $\oplus$ the addition between fuzzy sets which comes from application of extension principle (2) to $g(x_1, x_2) = x_1 + x_2$, i.e.

$$m_{A\oplus B}(z) = \sup_{x_1,x_2:x_1+x_2=z} \inf\{m_A(x_1), m_B(x_2)\}. \tag{10}$$

Note that $\oplus$ for crisp A, B coincides with the Minkowski-addition. It is simple also to define by (2) a scalar multiplication $\odot_s$ of fuzzy sets. For simplicity, we write λY instead of $\lambda \odot_s Y$.

A nice property of the Aumann expectation is its linearity w.r.t. $\oplus$, i.e.

$$\mathbb{E}^{(A)}(\lambda_1 Y_1 \oplus \lambda_2 Y_2) = \lambda_1 \mathbb{E}^{(A)} Y_1 \oplus \lambda_2 \mathbb{E}^{(A)} Y_2. \tag{11}$$

The main question, however, is: Can $\mathbb{E}^{(A)} Y$ be interpreted as a Frechét-expectation w.r.t. a certain metric? If this is true this metric can be used by (4) for a well defined variance. Examples show that $\mathbb{E}^{(A)}$ is not Frechét w.r.t. the Hausdorff metric d_p and that the Frechét-expectation w.r.t. d_p is a nonlinear operator (see Näther (1997)). But it holds:

Theorem 1. *The Aumann-expectation* $\mathbb{E}^{(A)}$ *is Frechét-expectation w.r.t.* δ_2 *from (6).*

For the proof see Näther (1997).

The result in Theorem 1 holds not only for the distance δ_2 from (6) but also for the more general L_2-distance in $L_2(S^{d-1} \times [0,1])$

$$\rho_2(A,B)^2 = \int_{[0,1]^2 \times (S^{d-1})^2} (s_A(u,\alpha) - s_B(u,\alpha)) \times (s_A(v,\beta) - s_B(v,\beta))\, dK(u,\alpha,v,\beta) \tag{12}$$

with a symmetric and positive definite kernel K.

Now, let us mention that $\mathbb{E}^{(A)}$ coincides with the so-called Pettis-expectation. Let $(U, \|\cdot\|)$ be a separable Banach space.

Definition 3. *Let Z be a random variable on U with* $\mathbb{E}\|Z\| < \infty$. *The* Pettis-expectation $\mathbb{E}^{(P)} Z$ *is that element of U with*

$$f(\mathbb{E}^{(P)} Z) = \mathbb{E} f(Z)$$

for each linear functional on U.

Using the Hahn-Banach-Theorem the uniqueness of $\mathbb{E}^{(P)}$ easily can be shown (see e.g. Gänssler/Stute (1977)). Since the Aumann-expectation $\mathbb{E}^{(A)}$ satisfies (11) the following theorem is not surprising:

Theorem 2. *Let Y be a rfv with* $\mathbb{E}\|Y\|_\rho < \infty$, *where* $\|\cdot\|_\rho = \rho_2(\cdot, \{0\})$ *is generated by* ρ_2 *from (12). Then the Aumann-expectation* $\mathbb{E}^{(A)} Y$ *is equal to the Pettis-expectation* $\mathbb{E}^{(P)} Y$.

2.4 Variance

Using the Frechét-principle the variance of a rfv Y corresponding to its Aumann-expectation $\mathbb{E}^{(A)} Y$ exists if $\mathbb{E}\|Y\|_\rho^2 < \infty$. Then it is given by (4), i.e.

$$\mathbf{Var}\ Y = \mathbb{E}\rho_2^2(Y, \mathbb{E}^{(A)} Y).$$

Since $s_{E^{(A)} Y} = \mathbb{E}s_Y$, which goes back to a standard result on random convex sets (see Stoyan/Stoyan (1994)), the variance can be written as

$$\mathbf{Var}\ Y = \int_{[0,1]^2 \times (S^{d-1})^2} \mathbf{Cov}(s_Y(u, \alpha), s_Y(v, \beta))\, dK(u, \alpha, v, \beta). \tag{13}$$

Let us mention but not stress here that (13) can be deduced also from a more general point of view using the notion of covariance operator for random variables on a Banach space $(U, \|\cdot\|)$, see e.g. Araujo/Giné (1980).

Moreover, let us mention that for two rfv's X and Y by use of the $L_2(S^{d-1} \times [0,1])$-scalar-product we can define

$$\langle X, Y \rangle \stackrel{\text{def}}{=} \int_{[0,1]^2 \times (S^{d-1})^2} s_X(u, \alpha) s_Y(v, \beta)\, dK(u, \alpha, v, \beta). \tag{14}$$

Then (13) can be written in a well known form as

$$\mathbf{Var}\ Y = \mathbb{E}\langle Y, Y \rangle - \langle \mathbb{E}Y, \mathbb{E}Y \rangle. \tag{15}$$

Consequently,

$$\mathbf{Cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}\langle X, Y \rangle - \langle \mathbb{E}X, \mathbb{E}Y \rangle$$

$$\rho(X, Y) \stackrel{\text{def}}{=} \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}\, X\ \mathbf{Var}\, Y}} \tag{16}$$

are definitions of covariance and correlation between rfv's X and Y which are suggesting themselves.

Let us point out, however, that $\langle X, Y \rangle$ is not a scalar product in $(\mathscr{F}_C^d, \oplus, \odot_s)$. The reason is that the image of $\mathscr{F}_C^d$ by the embedding via support functions is only a cone in $L_2(S^{d-1} \times [0,1])$, i.e. $\langle X, Y \rangle$ has only linearity properties w.r.t. scalar multiplication with positive constants. One consequence is, for example, that from $\rho(X, Y) = 1$ we cannot deduce a linear dependence between X and Y.

Let us show now that (13) contains several earlier approaches as special cases. Using the special kernel

$$dK(u, \alpha, v, \beta) = d \cdot \delta_u(v) \delta_\alpha(\beta) \nu(du)\, d\alpha$$

ρ_2 reduces to δ_2 from (6). This leads to the special variance

$$\mathbf{Var}\ Y = d \int_0^1 \int_{S^{d-1}} \mathbf{Var}\, s_Y(u, \alpha) \nu(du)\, d\alpha \tag{17}$$

which is used by Näther (1997).

The class of distances introduced by Bertoluzza et. al (1995) are special cases of (12). They defined a distance between two normal convex fuzzy sets A

and B of the real line $\mathbb{R}^1$ by

$$D(A,B)^2 = \int_0^1 \int_0^1 [t(\inf A_\alpha - \inf B_\alpha) \\ + (1-t)(\sup A_\alpha - \sup B_\alpha)]^2 \, dg(t) \, d\varphi(\alpha),$$

where g and φ are normalized weight measures on $[0,1]$. Straightforward calculations show that

$$D(A,B)^2 = c_2 \int_0^1 (\inf A_\alpha - \inf B_\alpha)^2 \, d\varphi(\alpha) \\ + (1 - 2c_1 + c_2) \int_0^1 (\sup A_\alpha - \sup B_\alpha)^2 \, d\varphi(\alpha) \\ + 2(c_1 - c_2) \int_0^1 (\inf A_\alpha - \inf B_\alpha)(\sup A_\alpha - \sup B_\alpha) \, d\varphi(\alpha)$$

with

$$c_1 = \int_0^1 t \, dg(t) \quad \text{and} \quad c_2 = \int_0^1 t^2 \, dg(t).$$

Because $S^{d-1} = S^0 = \{+1, -1\}$, the integral with respect to S^{d-1} is a sum of two terms for $u = +1$ and $u = -1$. Hence, with the kernel

$$dK(u,\alpha,v,\beta) = \delta_\alpha(\beta) \, d\varphi(\alpha) \begin{cases} 1 - 2c_1 + c_2 & \text{for } u = v = +1 \\ c_2 & \text{for } u = v = -1 \\ c_1 - c_2 & \text{for } u = -v \end{cases}$$

$D(A,B)^2$ is a special case of (12). Especially the $\boldsymbol{\lambda}$-mean squared dispersion defined by Lubiano et al. (2000) is based on a special D-distance and is included in the variance concept above. Lubiano defines the $\boldsymbol{\lambda}$-mean squared dispersion of a random fuzzy variable Y by

$$\Delta_{\boldsymbol{\lambda}}^2(Y) = \mathbb{E} D_{\boldsymbol{\lambda}}(Y, \mathbb{E}Y)^2$$

with

$$D_{\boldsymbol{\lambda}}(A,B)^2 = \int_0^1 [\lambda_1 (\sup A_\alpha - \sup B_\alpha)^2 + \lambda_2 (\operatorname{mid} A_\alpha - \operatorname{mid} B_\alpha)^2 \\ + \lambda_3 (\inf A_\alpha - \inf B_\alpha)^2] \, d\alpha,$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ with $\lambda_i \in [0,1)$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and mid $A_\alpha = 1/2(\sup A_\alpha + \inf A_\alpha)$ denotes the middle point of the interval. Here

$$dK(u,\alpha,v,\beta) = \delta_\alpha(\beta)\,d\alpha \begin{cases} \lambda_1 + \lambda_2/4 & \text{for } u = v = +1 \\ \lambda_3 + \lambda_2/4 & \text{for } u = v = -1 \\ -\lambda_2/4 & \text{for } u = -v. \end{cases}$$

Also the Hagaman distance (see Bardossy et al. (1992)) between two different LR-fuzzy numbers $A = (\mu_A, l_A, r_A)_{L_A R_A}$ and $B = (\mu_B, l_B, r_B)_{L_B R_B}$ (see (7))

$$D_f(A,B)^2 \stackrel{\text{def}}{=} \frac{1}{2}\int_0^1 \{(\mu_A - l_A L_A^{-1}(\alpha) - \mu_B + l_B L_B^{-1}(\alpha))^2 + (\mu_A + r_A R_A^{-1}(\alpha) - \mu_B - r_B R_B^{-1}(\alpha))^2\} f(\alpha)\,d\alpha$$

with density function f is obtained by

$$dK(u,\alpha,v,\beta) = \frac{1}{2}\delta_\alpha(\beta) f(\alpha)\,d\alpha.$$

In this way, also Diamonds distance (Diamond (1988)) in the case of triangular fuzzy numbers $A = (\mu_A, l_A, r_A)_\Delta$

$$d(A,B)^2 \stackrel{\text{def}}{=} (\mu_A - l_A - \mu_B + l_B)^2 + (\mu_A + r_A - \mu_B - r_B)^2 + (\mu_A - \mu_B)^2$$

is included. Note that a triangular fuzzy number is an LR-fuzzy number with linear left and right shape, i.e. $L(x) = R(x) = \max\{1 - x, 0\}$.

2.5 Properties and estimation

We will summarize some properties of the Aumann-expectation and the associated variance, and we will show that they both can be estimated in a similar way as in the classical case. In the following the superscript in $\mathbb{E}^{(A)}$ is omitted and we only write $\mathbb{E}$ for the Aumann-expectation.

The following theorem deals with independent rfv's. It is emphasized that it is not necessary to develop a special concept of independence for rfv's: Due to the embedding, rfv's Y can be considered as random (support) functions s_Y, and the independence of random functions is well-defined.

Theorem 3. *Let Y_1 and Y_2 be independent integrable random fuzzy variables on $\mathbb{R}^1$. Then $\mathbb{E}(Y_1 \odot Y_2) = \mathbb{E}Y_1 \odot \mathbb{E}Y_2$, where $\odot$ is the extended multiplication via* (2).

Proof. The operation $A \odot B$ is linear in both arguments, i.e. for all fuzzy sets A, B_1, B_2 and all real numbers λ

$$A \odot (\lambda B) = \lambda(A \odot B) \quad \text{and} \quad A \odot (B_1 \oplus B_2) = (A \odot B_1) \oplus (A \odot B_2)$$

(for the first argument use the symmetry $A \odot B = B \odot A$). By the independence of Y_1 and Y_2 and by the linearity of the expectation we obtain

$$\mathbb{E}(Y_1 \odot Y_2) = \mathbb{E}_{Y_1}\mathbb{E}_{Y_2}(Y_1 \odot Y_2 | Y_1) = \mathbb{E}_{Y_1}(Y_1 \odot \mathbb{E}Y_2 | Y_1) = \mathbb{E}Y_1 \odot \mathbb{E}Y_2.$$

Theorem 4. *Let Y and Y_1, Y_2 be random fuzzy variables and define the variance with respect to any L_2 distance ρ_2 in (12) and let $\|\cdot\|_\rho = \rho_2(\cdot, \{o\})$ be the induced norm. Then for any positive squared integrable random variable ξ, for any $A \in \mathscr{F}_C^d$ and for any real number λ we obtain*

1. $\mathbf{Var}(\xi) = \mathbb{E}(\xi - \mathbb{E}\xi)^2$,
2. $\mathbf{Var}(Y) = \mathbb{E}\|Y\|_\rho^2 - \|\mathbb{E}Y\|_\rho^2$,
3. $\mathbf{Var}(\lambda Y) = \lambda^2 \mathbf{Var}(Y)$,
4. $\mathbf{Var}(\xi A) = \|A\|_\rho^2 \cdot \mathbf{Var}(\xi)$,
5. $\mathbf{Var}(A \oplus Y) = \mathbf{Var}(Y)$,
6. $\mathbf{Var}(\xi Y) = \mathbb{E}\|Y\|_\rho^2 \cdot \mathbf{Var}(\xi) + \mathbb{E}\xi^2 \cdot \mathbf{Var}(Y)$ *if ξ and Y are independent.*
7. $\mathbf{Var}(Y_1 \oplus Y_2) = \mathbf{Var}(Y_1) + \mathbf{Var}(Y_2)$ *if Y_1 and Y_2 are independent.*

For the proof see Körner (1997).

Theorem 5. *Let $Y_1, Y_2, \ldots$ be independent and identically distributed random fuzzy variables with $\mathbb{E}\|Y_1\| < \infty$. Then $\overline{Y}_n = (Y_1 \oplus \cdots Y_n)/n$ is an unbiased and consistent estimator of the Aumann expectation, i.e.*

$$\mathbb{E}\,\overline{Y}_n = \mathbb{E}\,Y_1 \quad \textit{and} \quad \overline{Y}_n \xrightarrow[n\to\infty]{P\text{-a.s.}} \mathbb{E}\,Y_1.$$

Proof. Because of the linearity of the Aumann expectation we obtain that $\overline{Y}_n$ is unbiased. The consistency follows from the law of large number of Klement/Puri/Ralescu (1986).

Theorem 6. *Let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random fuzzy variables with $\mathbb{E}\|Y_1\|^2 < \infty$.*

Then $S_n^2 = (n-1)^{-1}\sum_{k=1}^n \rho_2(Y_k, \overline{Y}_n)^2$ is an unbiased and consistent estimator of $\mathbf{Var}(Y_1)$, *i.e.*

$$\mathbb{E}S_n^2 = \mathbf{Var}(Y_1) \quad \textit{and} \quad S_n^2 \xrightarrow[n\to\infty]{P\text{-a.s.}} \mathbf{Var}(Y_1).$$

Proof. Note that

$$\frac{n-1}{n} S_n^2 = \frac{1}{n}\sum_{k=1}^n \|Y_k\|_\rho^2 - \|\overline{Y}_n\|_\rho^2.$$

Now, $\overline{Y}_n \xrightarrow[n\to\infty]{P\text{-a.s.}} \mathbb{E}\,Y_1$ and $\|\cdot\|_\rho^2$ is continuous, therefore,

$$\|\overline{Y}_n\|_\rho^2 \xrightarrow[n\to\infty]{P\text{-a.s.}} \|\mathbb{E}\,Y_1\|_\rho^2. \quad (\text{Note that } \|\mathbb{E}\,Y_1\|_\rho^2 \le \mathbb{E}\|Y_1\|_\rho^2 < \infty.)$$

Furthermore, $d_k = \|Y_k\|_\rho^2$ is a sequence of independent identically distributed variables, such that

$$\bar{d}_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=1}^n d_k \xrightarrow[n\to\infty]{P\text{-a.s.}} \mathbb{E}d_1 = \mathbb{E}\|Y_1\|_\rho^2.$$

Hence,

$$S_n^2 \xrightarrow[n\to\infty]{P\text{-a.s.}} \mathbb{E}\| Y_1\|_\rho^2 - \|\mathbb{E} Y_1\|_\rho^2 = \mathbf{Var}(Y_1).$$

Furthermore S_n^2 is unbiased, because $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})$ is an unbiased estimator of $\mathbf{Cov}(X, Z)$ in the classical real case, i.e. we have

$$\begin{aligned}\mathbb{E}S_n^2 &= \int_{[0,1]^2\times(S^{d-1})^2} \mathbb{E}\frac{1}{n-1}\sum_{i=1}^{n}(s_{Y_i}(u,\alpha) - s_{\bar{Y}_n}(u,\alpha))\cdot \\ &\quad \cdot (s_{Y_i}(v,\beta) - s_{\bar{Y}_n}(v,\beta))\, dK(u,\alpha,v,\beta) \\ &= \int_{[0,1]^2\times(S^{d-1})^2} \mathbf{Cov}(s_Y(u,\alpha), s_Y(v,\beta))\, dK(u,\alpha,v,\eta) = \mathbf{Var}\, Y.\end{aligned}$$

3 Applications

3.1 Asymptotic results

Having an expectation of a rfv Y it is naturally to ask for a law of large numbers, i.e. for the convergence of

$$\frac{1}{n}(Y_1 \oplus \cdots \oplus Y_n) \stackrel{\text{def}}{=} \bar{Y}_n \tag{18}$$

to $\mathbb{E}^{(A)} Y$ for n to infinity where $Y_1, \ldots, Y_n$ are iid replications of Y. A first strong law of large numbers (SLLN) was presented by Klement/Puri/Ralescu (1986) where a.s.-convergence is proved w.r.t. the d_1-metric.

Let us point out, however, that measurability conditions of rfv's are mainly characterized by the d_∞-metric. Therefore one is more interested in SLLN with a.s.-convergence w.r.t. d_∞. Since $(\mathscr{F}_C^d, d_\infty)$ is non-seperable, the proof of such a SLLN is relatively complicated. It was achieved only recently by Colubi et al. (2000) and by Proske (1998) (see also Molchanov (1999)).

Proving SLLN for rfv's of second order, the existing variance allows the application of the well-known tools from classical theory, e.g. a Kolmogorov inequality can be proven. A typical SLLN for rfv's of second order is the following:

Theorem 7. *(see Körner (1997)) Let $Y_1, \ldots, Y_n, \ldots$ be a sequence of independent rfv's with $\mathbb{E}\| Y_n\|_\rho^2 < \infty$. If the series $\sum_{k=1}^{\infty} \mathbf{Var}\, Y_k/k^2$ converges then*

$$P\left(\lim_{n\to\infty} \rho_2\left(\bar{Y}_n, \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}^{(A)} Y_k\right) = 0\right) = 1.$$

Applying results for probability distributions in linear spaces (see Vakhania (1981)) it can be shown further that for an independent identical distributed sequence $Y_1, \ldots, Y_n, \ldots$ of rfv's with $\mathbb{E}^{(A)} Y_1 = M_o$

$$n\delta_2(\overline{Y}_n, M_o)^2 \tag{19}$$

has asymptotically a so-called ω^2-distribution, i.e. the distribution of (19) is asymptotically equivalent to the distribution of

$$\sum_{k=1}^{\infty} \lambda_k \xi_k^2 \tag{20}$$

where $\xi_1, \ldots, \xi_k, \ldots$ are iid $N(0,1)$ and $\lambda_1 \geq \lambda_2 \geq \cdots$ are the eigenvalues of the covariance operator C_{Y_1} of Y_1. At least in that case where the Y_i are random LR-fuzzy numbers (7) the distribution of (19) is relatively easy to compute since C_{Y_1} reduces to a 3×3-matrix with three eigenvalues. (19) and (20) can be used to construct an asymptotic significance test of the hypothesis $\mathbb{E}^{(A)} Y = M_o$. For more details see Körner (2000).

3.2 Linear statistical inference for regression with random fuzzy data

Let us denote a classical crisp linear regression model by a random variable y which depends on the regressor $x \in \mathbb{R}^k$ by

$$\mathbb{E} y = f_1(x)\beta_1 + \cdots + f_m(x)\beta_m = f(x)^T \beta \tag{21}$$

where $f|\mathbb{R}^k \to \mathbb{R}^m$ is a known setup-function and $\beta \in \mathbb{R}^m$ is an unknown regression parameter. Given observations $y_1, \ldots, y_n$ of y at the design points $x_1, \ldots, x_n \in \mathbb{R}^k$, the parameter β has to be estimated, e.g. by the classical least-squares estimator

$$\check{\beta} = (F^T F)^{-1} F^T \underline{y} \tag{22}$$

where $\underline{y} = (y_1, \ldots, y_n)^T$ and $F = (f(x_1), \ldots, f(x_n))^T$ is assumed to have full rank.

The problem now is that only fuzzy observations $Y_1, \ldots, Y_n$ are available, for example the clouding y for given atmospheric pressure x is reported by linguistic expressions like: Cloudless, Clear, Fair, Cloudy, Overcast (see the introductory example). Thus (21) has to be generalized by

$$\mathbb{E}^{(A)} Y = f(x)^T B, \tag{23}$$

where now Y is a random fuzzy variable and B is a fuzzy parameter vector. The question is: How to estimate B?

There are some data-analytic approaches from literature where it is not necessary to have a stochastic model like (23). For example Tanaka and his school looks for such a fuzzy parameter $\hat{B}$ that the fuzzy function $f(x)^T \hat{B}$ covers all the fuzzy data $Y_1, \ldots, Y_n$ at least to a given degree (see e.g. Tanaka

(1987)). A second approach consists in a straightforward application of the extension principle (2) to well justified classical crisp estimators, e.g. to the least squares estimator (22) (see Viertl (1996) but also Körner/Näther (1998) with some critizism of this approach). A third possibility is to find a suitable fuzzy parameter estimate $\hat{B}$ by use of a least squares approximation principle for fuzzy data (see e.g. Diamond (1988), (1992), but also Körner/Näther (1998) with a certain generalization).

Here we will emphasize the stochastic background of the data, i.e. we assume that the data $Y_1, \ldots, Y_n$ are realizations of a rfv. Y, especially of a random fuzzy number, and we are interested especially in best linear unbiased estimation (BLUE) of B.

At first let us specialice the expectation and variance formulas for a random LR-fuzzy number $Y = (\mu, l, r)_{LR}$. We restrict ourselfes to δ_2 from (6). Having in mind that in the LR-fuzzy-number-case $S^d = S^o = \{-1, 1\}$ and

$$s_Y(u, \alpha) = \begin{cases} -\mu + lL^{-1}(\alpha) & \text{if } u = -1 \\ \mu + rR^{-1}(\alpha) & \text{if } u = 1 \end{cases}$$

we have for $A = (\mu_A, l_A, r_A)_{LR}$, $B = (\mu_B, l_B, r_B)_{LR}$

$$\delta_2(A, B)^2 = |\mu_A - \mu_B|^2 + \frac{1}{2} r_2 |r_A - r_B| + \frac{1}{2} l_2 |l_A - l_B| \\ - l_1(\mu_A - \mu_B)(l_A - l_B) + r_1(\mu_A - \mu_B)(r_A - r_B)$$

with

$$l_1 = \int_0^1 L^{-1}(\alpha)\, d\alpha, \quad r_1 = \int_0^1 R^{-1}(\alpha)\, d\alpha$$

$$l_2 = \int_0^1 (L^{-1}(\alpha))^2\, d\alpha, \quad r_2 = \int_0^1 (R^{-1}(\alpha))^2\, d\alpha,$$

(compare also with (9)).

Then it is easy to compute

$$\mathbb{E}^{(A)}(\mu, l, r)_{LR} = (\mathbb{E}\mu, \mathbb{E}l, \mathbb{E}r)_{LR} \tag{24}$$

$$\mathbf{Var}(\mu, l, r)_{LR} = \mathbf{Var}\,\mu + \frac{1}{2} l_2 \mathbf{Var}\, l + \frac{1}{2} r_2 \mathbf{Var}\, r \\ - l_1 \mathbf{Cov}(\mu, l) + r_1 \mathbf{Cov}(\mu, r). \tag{25}$$

For a random symmetric fuzzy number $Y = (\mu, \Delta)_L$ (24) and (25) reduces to

$$\mathbb{E}^{(A)}(\mu, \Delta)_L = (\mathbb{E}\mu, \mathbb{E}\Delta)_L$$

$$\mathbf{Var}(\mu, \Delta)_L = \mathbf{Var}\,\mu + l_2 \mathbf{Var}\,\Delta \tag{26}$$

Looking for a BLUE, we have to carry out linear operations with the fuzzy data. The advantage of LR-fuzzy numbers is that $\oplus$ and $\odot_s$ can be expressed by simple operations w.r.t. the parameters μ, l and r (see e.g. Dubois/Prade (1980)):

$$(\mu_1, l_1, r_1)_{LR} \oplus (\mu_2, l_2, r_2)_{LR} = (\mu_1 + \mu_2, l_1 + l_2, r_1 + r_2)_{LR} \tag{27}$$

$$\lambda(\mu, l, r)_{LR} = \begin{cases} (\lambda\mu, \lambda l, \lambda r)_{LR} & \text{if } \lambda > 0 \\ (\lambda\mu, -\lambda r, -\lambda l)_{RL} & \text{if } \lambda < 0 \\ 1_{\{0\}} & \text{if } \lambda = 0. \end{cases} \tag{28}$$

Here 1_A is the indicator function of a set A.

For symmetric fuzzy numbers (27) and (28) reduces to

$$\lambda_1(\mu_1, \Delta_1)_L \oplus \lambda_2(\mu_2, \Delta_2)_L = (\lambda_1\mu_1 + \lambda_2\mu_2, |\lambda_1|\Delta_1 + |\lambda_2|\Delta_2)_L. \tag{29}$$

a) BLUE

First of all let us point out that (similar to interval arithmetics) $(\mathscr{F}_C^d, \oplus, \odot_s)$ is not a linear space since $\oplus$ is not a group operation. Especially LR-fuzzy numbers with given L and R form a convex cone w.r.t. $\oplus$ and $\odot_s$. This will be the reason (and the following discussion will show it) that there is no straightforward analogy between classical BLUE and BLUE with fuzzy data. Now the starting point is the fuzzified regression model (23), i.e. more exactly

$$\mathbb{E}^{(A)} Y(x) = f(x)^T B = f_1(x)B_1 \oplus \cdots \oplus f_m(x)B_m. \tag{30}$$

Let n fuzzy number data be given, say

$$Y_i = (y_i, \Delta_i)_L; \quad i = 1, \ldots, n. \tag{31}$$

The aim is to estimate the B_j; $j = 1, \ldots, m$; by a linear estimator

$$\hat{B}_j = \lambda_{1j} Y_1 \oplus \cdots \oplus \lambda_{nj} Y_n \stackrel{\text{def}}{=} \lambda_j^T Y \tag{32}$$

which is unbiased in the sense that

$$\mathbb{E}^{(A)} \hat{B}_j = B_j. \tag{33}$$

Taking into account (29), for unbiasedness it is necessary that B_j is modeled as a symmetric fuzzy number with the same shape function L, say

$$B_j = (\beta_j, \delta_j)_L; \quad j = 1, \ldots, m.$$

Then (30) writes

$$\begin{aligned}\mathbb{E}^{(A)} Y &= (f_1(x)\beta_1 + \cdots + f_m(x)\beta_m, |f_1(x)|\delta_1 + \cdots + |f_m(x)|\delta_m)_L \\ &\stackrel{\text{def}}{=} (f(x)^T\beta, |f(x)^T|\delta)_L\end{aligned}$$

and, since the Y_i are assumed to be a realization of $Y(x_i)$,

$$\mathbb{E}^{(\Lambda)}\hat{B}_j = (\lambda_j^T F\beta, |\lambda_j^T||F|\delta)_L$$

with $\lambda_j^T = (\lambda_{1j}, \ldots, \lambda_{nj})$, $|\lambda_j^T| = (|\lambda_{1j}|, \ldots, |\lambda_{nj}|)$; $j = 1, \ldots, m$; $F = (f(x_1), \ldots, f(x_n))^T$, $|F| = (|f(x_1)|, \ldots, |f(x_N)|)^T$. More condensed, with $\Lambda \stackrel{\text{def}}{=} (\lambda_1, \ldots, \lambda_m)^T$ and $B = (B_1, \ldots, B_m)^T \stackrel{\text{def}}{=} (\beta, \delta)_L$, an estimator $\hat{B}$ is unbiased iff

$$\mathbb{E}^{(\Lambda)}\hat{B} = (\Lambda F\beta, |\Lambda|\,|F|\delta)_L = (\beta, \delta)_L$$

which is satisfied iff simultaneously

$$\Lambda F = I_m, \quad |\Lambda|\,|F| = I_m. \tag{34}$$

The first equation ensures unbiasedness of the centre, the second unbiasedness of the spreads. Unfortunately, in general it is not possible to obtain unbiasedness of the spreads. This can be seen already in the simple linear regression case

$$\mathbb{E}Y = B_1 x \oplus B_2.$$

Here $F^T = \begin{pmatrix} x_1, \ldots, x_n \\ 1, \ldots, 1 \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda_{11} \cdots \lambda_{1n} \\ \lambda_{21} \cdots \lambda_{2n} \end{pmatrix}$ and the second equation of (34) writes

$$\sum_i |\lambda_{1i}| = 1, \quad \sum_i |\lambda_{1i}||x_i| = 0,$$

$$\sum_i |\lambda_{2i}| = 0, \quad \sum_i |\lambda_{2i}||x_i| = 1,$$

which is inconsistent, since e.g. from $\sum |\lambda_{2i}| = 0$ it follows $\lambda_{2i} = 0$ for all i. But then we cannot obtain $\sum |\lambda_{2i}||x_i| = 1$.

Therefore it holds:

Theorem 8. *For the model (30) with $m \geq 2$, there is in general no unbiased linear estimator for B of the form (32).*

For $m = 1$, i.e. $\mathbb{E}Y = f(x)B$, unbiasedness can be forced. Then (34) reduces to

$$\sum_i \lambda_i f(x_i) = 1, \quad \sum_i |\lambda_i|\,|f(x_i)| = 1 \tag{35}$$

which is equivalent to $\sum_i \lambda_i f(x_i) = 1$ and sign $\lambda_i =$ sign $f(x_i)$. E.g. $\lambda_i = 1/nf(x_i)$ automatically leads to an unbiased linear estimator for B. In this one-dimensional special case it is easy to find the BLUE: Using Theorem

4 point 3 and 7 with $\mathbf{Var}\, Y_i \overset{\text{def}}{=} \sigma^2$ we have

$$\mathbf{Var}\,\hat{B} = \sum_{i=1}^{n} \lambda_i^2 \sigma^2.$$

The BLUE coefficients

$$\lambda_i^* = f(x_i) \Big/ \sum_{j=1}^{n} (f(x_j))^2 \tag{36}$$

are the solutions of $\sum_{i=1}^{n} \lambda_i^2 = \text{Min}$ with side condition (35). As a special case of (36), clearly, the arithmetic fuzzy mean $\overline{Y}_n$ (see (18)) is BLUE for the expectation $\mathbb{E}^{(\Lambda)} Y = B$.

b) Weak BLUE

One way out of the situation described in Theorem 8 is to make setups and requirements only for the central values of the data and not for the spreads (see Näther (1997)). Given data of the form (31), i.e. $Y_i = (y_i, \Delta_i)_L$; $i = 1, \ldots, n$; instead of (30) we use a model $Y_i = Y(x_i)$ with

$$\begin{aligned}\mathbb{E}^{(\Lambda)} Y(x) &= (f_1(x)\beta_1 + \cdots + f_m(x)\beta_m, \Delta_0(x))_L \\ &= (f(x)^T \beta, \Delta_0(x))_L.\end{aligned} \tag{37}$$

For estimating the β_j we consider analogously to (32) for $j = 1, \ldots, m$

$$\hat{\beta}_j = \lambda_{1j} Y_1 \oplus \cdots \oplus \lambda_{nj} Y_n = \lambda_j^T Y = (\lambda_j^T y, |\lambda_j|^T \Delta)_L,$$

where $y = (y_1, \ldots, y_n)^T$, $\Delta = (\Delta_1, \ldots, \Delta_n)^T$. Now, with the terminology from subsection a) we have for $j = 1, \ldots, m$

$$\mathbb{E}^{(\Lambda)} \hat{\beta}_j = (\lambda_j^T F \beta, |\lambda_j|^T \Delta_0)_L; \Delta_0 = (\Delta_0(x_1), \ldots, \Delta_0(x_n))^T$$

or more condensed

$$\mathbb{E}^{(\Lambda)} \hat{\beta} = (\Lambda F \beta, |\Lambda| \Delta_0)_L.$$

Much more weaker as in (34), we only require unbiasedness of the centre, i.e.

$$\Lambda F = I_m. \tag{38}$$

An estimator $\hat{\beta}$ with (38) is called weak unbiased. Analogously to classical linear inference a *weak linear unbiased estimator* $\hat{\beta}$ exists iff F is of full rank. To find the weak BLUE for β let us consider $\mathbf{Var}\,\hat{\beta}_j$ which we can compute by use of (26) as

$$\begin{aligned}\mathbf{Var}\,\hat{\beta}_j &= \mathbf{Var}(\lambda_j^T y) + l_2 \mathbf{Var}(|\lambda_j|^T \Delta) \\ &= \lambda_j^T \textstyle\sum_y \lambda_j + l_2 |\lambda_j|^T \textstyle\sum_\Delta |\lambda_j|; j = 1, \ldots, m;\end{aligned} \tag{39}$$

where $\sum_y$, like $\sum_\Delta$ are the covariance matrices of the observed centres y and of the observed spreads Δ. Minimization of (39) w.r.t. λ_j under side condition (38) gives the coefficient vector λ_j^* of the weak BLUE $\hat{\beta}_j^*$. As an essential difference to the classical linear estimation theory, (39) can in general not be reduced to minimization of a quadratic form. However, reduction to a quadratic form is possible, if the spreads are uncorrelated, i.e. if

$$\sum_\Delta = \mathrm{diag}(\sigma_\Delta(x_i)).$$

Then (39) reduces to

$$\mathbf{Var}\,\hat{\beta}_j = \lambda_j^T \sum \lambda_j; \sum = \sum_y + l_2 \sum_\Delta. \tag{40}$$

Minimization of (40) under side condition (38) coincides with the classical BLUE-problem for linear regression with observations correlated by $\sum$. Thus, the solution is given by

Theorem 9. *If $F^T \sum^{-1} F$ is regular and $\sum_\Delta$ is diagonal, the weak BLUE for β in the model (37) is given by*

$$\hat{\beta}^* = \left(F^T \sum^{-1} F\right)^{-1} F^T \sum^{-1} Y.$$

The main disadvantage of a weak BLUE is that the spreads are uncontrolled, and indeed examples show too large spreads of the estimated regression function.

c) Componentwise BLUE

A more satisfactory approach in this connection seems to be the following: The idea is to split up the problem. For estimation of the centre of the regression parameter only the observed centres y_i are used, while for estimation of the spread only the observed spreads Δ_i are taken into account. This means that we give up the requirement that the estimator should be a linear form of the "unsplitted" fuzzy data $Y_i = (y_i, \Delta_i)_L$. Somewhat more detailed: Now the model is

$$\mathbb{E}^{(A)} Y(x) = (\mathbb{E}y, \mathbb{E}\Delta)_L = (f(x)^T\beta, g(x)^T\gamma)_L; \beta \in \mathbb{R}^m, \gamma \in \mathbb{R}^q,$$

i.e. for the centre and for the spread different setups are used. Clearly, in the interesting region, say $H \subseteq \mathbb{R}^k$, positivity of spread must be ensured, i.e.

$$\forall x \in H : g(x)^T\gamma > 0.$$

Given observations $Y_i = (y_i, \Delta_i)_L$ we consider estimators of the form

$$\hat{\beta} = \Lambda y, \quad \hat{\gamma} = \Gamma\Delta, \quad \forall x \in H : g(x)^T\hat{\gamma} \geq 0$$

where $y = (y_1, \ldots, y_n)^T$ and $\Delta = (\Delta_1, \ldots, \Delta_n)^T$. $\hat{\beta}$ is a classical linear estimator of the centre-parameter β based only on the observed central values y and $\hat{\gamma}$ is a classical linear estimator of the spread-parameter γ based on the observed spreads Δ with a side condition for positivity. Unbiasedness of $\hat{\beta}$ and $\hat{\gamma}$ is ensured if

$$\Lambda F = I_m, \Gamma G = I_q; \quad G = (g(x_1), \ldots, g(x_n))^T.$$

For estimation of the regression function $\mathbb{E}Y(x)$ we will use

$$\hat{Y}(x) = \left(f(x)^T \hat{\beta}, g(x)^T \hat{\gamma}\right)_L. \tag{41}$$

An unbiased estimator $\hat{Y}^*$ of the form (41) with minimal variance is called *componentwise BLUE*. Clearly, $\hat{Y}$ is unbiased iff $\hat{\beta}$ and $\hat{\gamma}$ are unbiased. To find the componentwise BLUE, firstly **Var** $\hat{Y}(x)$ is obtained from (26) as

$$\begin{aligned}\mathbf{Var}\ \hat{Y}(x) &= \mathbf{Var}(f(x)^T \hat{\beta}) + l_2 \mathbf{Var}(g(x)^T \hat{\gamma}) \\ &= f(x)^T \mathbf{Cov}\, \hat{\beta} f(x) + l_2 g(x)^T \mathbf{Cov}\, \hat{\gamma} g(x).\end{aligned}$$

From this the following is clear:

Theorem 10. *If $\hat{\beta}^*$ is BLUE for β in the linear model $Ey = F\beta$ and if $\hat{\gamma}^*$ is BLUE for γ in the linear model $E\Delta = G\gamma$ under the side condition $\forall x \in H : g(x)^T \hat{\gamma}^* > 0$, then*

$$\hat{Y}^*(x) = (f(x)^T \hat{\beta}^*, g(x)^T \hat{\gamma}^*)_L$$

is componentwise BLUE. Clearly, if regularity of the matrices is ensured, $\hat{\beta}^*$ is given by

$$\hat{\beta}^* = \left(F^T \sum\nolimits_y^{-1} F\right)^{-1} F^T \sum\nolimits_y^{-1} y$$

and $\hat{\gamma}^*$ by

$$\hat{\gamma}^* = \left(G^T \sum\nolimits_\Delta^{-1} G\right)^{-1} G^T \sum\nolimits_\Delta^{-1} \Delta, \tag{42}$$

if

$$\forall x \in H : g(x)^T \hat{\gamma}^* > 0 \tag{43}$$

is fulfilled.

The crucial point is that $\hat{\gamma}^*$ from (42) satisfies (43) only in special cases. For example, if $\{H_j\}_{j=1,\ldots,q}$, is a partition of H and if we model the spread on

H_j by $\gamma_j \geq 0$, i.e. if we use $g(x) = (1_{H_1}(x), \ldots, 1_{H_q}(x))^T$, then $\hat{\gamma}^*$ from (42) is given by $\hat{\gamma}^* = (\bar{\Delta}_1, \ldots, \bar{\Delta}_q)^T$ where $\bar{\Delta}_j$ is the arithmetic mean of spreads for observations from H_j. Clearly, $\bar{\Delta}_j \geq 0$ for $x \in H_j$ and the requirement (43) is automatically satisfied.

By straightforward considerations, the results of this section can be generalized to LR-type-data using the model

$$\mathbb{E}^{(A)} Y(x) = (\mathbb{E}y, \mathbb{E}l, \mathbb{E}r)_{LR} = (f(x)^T \beta, g_l(x)^T \gamma_l, g_r(x)^T \gamma_r)_{LR}.$$

Some numerical examples w.r.t. section 3.2. can be found in Körner/Näther (1998).

4 Concluding remarks

We have discussed in section 3.2 the application of rfv's in the context of BLUE for regression models with fuzzy data. This is a contribution to non-asymptotic results for statistical inference with rfv's. Note that there are many results on the asymptotic behaviour of rfv's. Some of them are quoted after Definition 1 and in section 3.1. Let us emphasize, however, that particularly non-asymptotic results are important, since fuzzy data are available often in such situations where the amount of information on an experiment is restricted, i.e. we cannot expect large sample sizes.

An interesting problem for further research on regression with fuzzy data is the statistical analysis of that case, where not only the "output" Y is fuzzy but also the "input" x.

Acknowledgement. I am indebted to the referees for their careful reading and valuable proposals which have led to on improvement of the manuscript.

References

1. Araujo A, Giné E (1980) The central limit theorem for real and Banach valued random variables. John Wiley and Sons. Chichester
2. Aumann RJ (1965) Integrals of set-valued functions. J. Math. Anal. Appl. 12:1–12
3. Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer Academic Publishers, Dordrecht-Boston-London
4. Bardossy A, Hagaman R, Duckstein L, Bogardi I (1992) Fuzzy least squares regression: Theory and application. In: Kacprzyk J and Fedrizzi M (eds.) Fuzzy regession analysis. Omnitech Press, Warsaw and Physica-Verlag, Heidelberg, 181–193
5. Bertoluzza C, Salas A, Corral N (1995) On a new class of distances between fuzzy numbers. Mathware & Soft Computing 2:71–84
6. Colubi A, Lopez-Diaz M, Dominguez-Menchero JS, Gil MA (1999) A generalized strong law of large numbers. Prob. Theor. Rel. Field 114:401–417
7. Diamond P (1988) Fuzzy least squares. Information Sciences 46:141–157
8. Diamond P (1992) Least squares and maximum likelihood regression for fuzzy linear models. In: Kacprzyk J and Fedrizzi M (eds.) Fuzzy regression analysis. Omnitech Press, Warsaw and Physica-Verlag, Heidelberg, 137–151
9. Diamond P, Kloeden P (1994) Metric spaces of fuzzy sets. World Scientific, Singapore, New Jersey, London, Hongkong

10. Dubois D, Prade H (1980) Fuzzy sets and systems: Theory and application. Academic Press, New York
11. Fréchet M (1948) Les éléments aléatoires de natures quelconque dans un éspace distancié. Ann. Inst. H. Poincaré 10:215–310
12. Gänssler P, Stute W (1977) Wahrscheinlichkeitstheorie. Springer, Berlin-Heidelberg
13. Hirota K (1981) Concepts of probabilistic sets. Fuzzy Sets and Systems 5:31–46
14. Klement EP, Puri ML, Ralescu DA (1986) Limit theorems for fuzzy random variables. Proceedings of Royal Society of London – Series A 19, 171–182
15. Körner R (1997) On the variance of fuzzy random variables. Fuzzy Sets and Systems 92:83–93
16. Körner R (2000) An asymptotic α-test for the expectation of random fuzzy variables. Journ. of Statist. Plann. and Inference 83:331–346
17. Körner R, Näther W (1998) Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. Inf. Science 109:95–118
18. Kruse R, Meyer KD (1987) Statistics with vague data. Reidel Publ. Comp., Dordrecht
19. Kwakernaak H (1978) Fuzzy random variables – I. Definitions and theorems. Inform. Sci. 15:1–29
20. Lubiano MA, Gil MA, López AT, López-Diaz M (2000) The λ-mean squared dispersion associated with a fuzzy random variable. Fuzzy sets and systems 111:307–317
21. Matheron G (1975) Random sets and integral geometry. J. Wiley & Sons, New York
22. Molchanov, IS (1993) Limit theorems for unions of random closed sets. Lecture Notes in Mathematics 1561: Springer-Verlag, Berlin-Heidelberg-New York
23. Molchanov, IS (1999) On strong laws of large numbers for random upper semicontinuous functions. J. Math. Anal. Appl. 235:349–355
24. Näther W (1997) Linear statistical inference for random fuzzy data. statistics 29:221–240
25. Proske F (1998) Grenzwertsätze für Fuzzy-Zufallsvariablen. Thesis, University of Ulm
26. Puri ML, Ralescu DA (1986) Fuzzy random variables. J. Math. Anal. Appl. 114:409–422
27. Radström H (1952) An embedding theorem for spaces of convex sets. Proc. Amer. Math. Soc. 3:165–169
28. Stoyan D, Stoyan H (1994) Fractals, Random Shapes and Point Fields, Wiley & Sons, Chichester
29. Stoyan D (1998) Random sets: models and statistics. Internat. Statist. Review 66:1–27
30. Tanaka H (1987) Fuzzy data analysis by possibilistic linear models. Fuzzy Sets and Systems 24:363–375
31. Vakhania NN (1981) Probability distribution on linear spaces. Elsevier Science Publishers B.V., North Holland. Amsterdam
32. Viertl R (1996) Statistical methods for non-precise data. CRC Press, Boca Raton, New York, London, Tokyo
33. Zadeh LA (1965) Fuzzy sets. Inf. and Control 8:338–353