Frontiers in Statistical Quality Control

Sven Knoth
Wolfgang Schmid *Editors*

Frontiers in Statistical Quality Control 11

Springer

Frontiers in Statistical Quality Control

More information about this series at
http://www.springer.com/series/10821

Sven Knoth • Wolfgang Schmid
Editors

Frontiers in Statistical Quality Control 11

Springer

Editors
Sven Knoth
Helmut Schmidt University
Hamburg
Germany

Wolfgang Schmid
Europa-Universität Viadrina
Frankfurt (Oder)
Germany

Frontiers in Statistical Quality Control
ISBN 978-3-319-12354-7 ISBN 978-3-319-12355-4 (eBook)
DOI 10.1007/978-3-319-12355-4

Library of Congress Control Number: 2015938305

Mathematics Subject Classification: 62-06, 62P30, 62L99, 62N05, 62K99

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

The XIth International Workshop on *Intelligent Statistical Quality Control* took place in Sydney, Australia from August 20 to August 23, 2013. It was hosted by Professor Ross Sparks, CSIRO Mathematics, Informatics and Statistics, North Ryde, Australia. The invitational workshop was jointly organized by Professors S. Knoth, W. Schmid, and R. Sparks. The 23 papers in this volume were carefully selected by the scientific program committee, reviewed by its members, revised by the authors and, finally, adapted by the editors for this volume.

The focus of the book lies on three major areas of statistical quality control: statistical process control (SPC), acceptance sampling and design of experiments. The majority of the papers deal with statistical process control while acceptance sampling, and design of experiments are treated to a lesser extent.

The book is divided into four parts. Subject of Part I is statistical process control. Part II is devoted to acceptance sampling. Part III covers the design of experiments, while in Part IV related fields are considered.

Part I: Statistical Process Control

Social networks are increasingly attracting the attention of academic and industry researchers. Monitoring communications between clusters of suspicious individuals is important in flagging potential planning activities for terrorism events or crime. Governments are interested in methodology that can forewarn them of future terrorist attacks or social uprisings in disenchanted groups of their populations. In the paper of **Sparks** a range of approaches is examined that could be used to monitor communication levels between suspicious individuals.

Woodall and **Driscoll** deal with the monitoring of a rare event. A review of some recent results is given and several new approaches are offered. Because some of the competing monitoring procedure have implicit headstart features, there are compelling arguments for the use of steady-state performance metrics. The choice

of an appropriate performance metrics is discussed in detail. The strong adverse effect of Phase I parameter estimation on Phase II performance of various charts is summarized. In addition, the important practical issue of the effect of aggregation of counts over time, some generalizations of standard methods, and some promising research ideas are discussed.

Big data is a popular term that is used to describe the large, diverse, complex, and/or longitudinal datasets generated from a variety of instruments, sensors, and/or computer-based transactions. In **Megahed** and **Jones-Farmer** several big data applications are discussed to highlight the opportunities and challenges for applied statisticians interested in surveillance and statistical process control. The goal of the authors is to bring the research issues into better focus and encourage methodological developments for big data analysis in these areas.

Epprecht presents a survey of the research on techniques for the statistical control of industrial multiple-stream processes. These are processes in which the same type of item is manufactured in several streams of output in parallel, or still continuous processes in which several measures are taken at a cross section of the product. This paper seems to be the first literature review on this topic. Essential differences in the underlying models are stressed and issues for further research are pointed out.

Yashchin considers a unified methodology based on the use of likelihood ratio tests to monitor processes. This approach leads to control schemes that provide good statistical performance and are easy to implement. They depend on just one design parameter and require a limited computational effort that is dynamically adjusted based on the process conditions. An example pertaining to multivariate control of the normal mean is discussed in detail.

Lazariv and **Schmid** give an overview about variance control charts for time dependent processes. In their paper, they consider charts based on the likelihood ratio approach and the generalized likelihood ratio approach, the sequential probability ratio method and the generalized sequential probability ratio procedure, the Shiryaev–Roberts procedure and a generalized Shiryaev–Roberts approach, and different types of exponentially weighted moving average (EWMA) charts. Within an extensive simulation study, these schemes are compared with each other. In order to measure the performance of the schemes, the average run length and the average delay are used.

Controlling both increases and decreases in a parameter by using a control statistic with an asymmetrical distribution, frequently leads to an ARL-biased chart. This means that some out-of-control ARL values are larger than the in-control ARL. **Knoth** and **Morais** discuss the idea of ARL-unbiased charts, provide instructive illustrations of ARL-(un)biased charts, relate ARL-unbiased Shewhart charts with the notions of unbiased and uniformly most powerful unbiased tests, and briefly discuss the design of EWMA charts not based on ARL(-unbiasedness).

The paper of **Su**, **Gan**, and **Tang** concerns with cumulative sum (CUSUM) charts. If the density of the in-control process is unknown, they propose to estimate its density by a kernel density estimator. The performance of this chart is

investigated for unimodal distributions. The obtained results reveal that this chart works well if a sufficient number of observations of the in-control process are available.

In **Yang** and **Arnold**, EWMA control charts based on the process proportions and an arcsin transformation are proposed to monitor the process mean and variance simultaneously. The sampling properties of the new monitoring statistics are analyzed. EWMA recursions are applied to these quantities. The behavior of the new schemes is analyzed by making use of its average run lengths.

Capizzi and **Masarotto** concern with the application of variable selection (VS) algorithms for monitoring multivariate data. These charts share the common idea that process faults usually affect a small fraction of the monitored quality characteristics. Hence, VS methods can be used to identify the subset of the variables for which the shift may have occurred. However, the suggested VS-based control charts differ in many aspects such as the particular VS algorithm and the type of the control statistic. In this paper, some VS-based control charts are compared with each other in a variety of out-of-control scenarios.

Göb and **Lurz** use an extension of the Camp-Meidell inequality to determine the control limits of a Shewhart chart. This procedure does not make use of any distributional assumption but only needs the existence of moments higher than 2. It is discussed how the moments in the bounds can be estimated from a Phase I sample. Appropriate estimators, their properties, and the effect of estimation on the properties of the process monitoring charts are investigated. In particular, the use of empirical Camp-Meidell bounds in quantile control charts is studied.

When simultaneous schemes are used, the quality characteristic is deemed to be out of control whenever a signal is triggered by either individual chart. **Morais**, **Ramos**, and **Pacheco** deal with the problem of misleading signals (MS), meaning that a shift in the process mean can be misinterpreted as a shift in the process variance and vice versa. Conditions are discussed to achieve values for the probability of a misleading signal smaller than or equal to 0.5 and alternative simultaneous Shewhart-type schemes are explored.

Saniga, **Davis**, **Faraz**, **McWilliams**, and **Lucas** investigate the characteristics of economic control chart designs for both Shewhart and CUSUM control charts. Authors in the past have made some suggestions regarding the design of these charts, where design is defined as finding the values of sample size, intersample interval and control limit (Shewhart), or control parameters for the CUSUM chart. In the present paper, the authors run a large number of experiments consisting of many configurations of the parameters and describe and model the results in terms of the actual economic designs.

Hryniewicz analyzes SPC procedures when the quality parameters of interest can be hardly directly monitored. Training data are used to build a model that is used for the prediction of the value of an unobservable variable of interest. In the paper, a model of a process is considered when traditionally applied assumptions are violated. In such a case, it is shown that some non-statistical prediction models proposed in the area of data-mining perform better than popular linear prediction

models. However, new problems have to be considered when shifts in the levels of process parameters may influence the performance of applied classification algorithms.

Nowadays, there are many applications of SPC outside engineering, e.g., in public health and finance. **Di Bucchianicco** and **van den Heuvel** investigate to which extent modern statistical theory like hypothesis testing, prediction intervals, and tolerance intervals may be used to extend Shewhart's ideas in the above-mentioned wider range of application domains. Alternative settings of statistical control proposed in the literature including Bayesian settings are discussed as well.

Part II: Acceptance Sampling

Wilrich considers sampling plans which have besides the specification limit an additional limit. Such extended sampling plans are, e.g., used for the evaluation of bacterial contamination in foods, the amount of active ingredient used in formulating drug products, and the strength of concrete. The operating characteristic function of these extended sampling plans for inspection by variables is derived, and the advantages/disadvantages in comparison with unextended sampling plans are discussed.

The concept of a fractional acceptance number is particularly useful for short-run food manufacturing processes involving a measurable quality characteristic such as the percentage sugar or fat content. **Govindaraju** and **Jones** propose a new fractional acceptance number sampling plan, which is a mix of attribute and variables methods. The operating characteristics of the proposed plan are evaluated using common error distributions and the incomplete beta function.

Steland extends the methodology of acceptance sampling for variables with unknown distributions when additional sampling information is available to such settings. Based on appropriate approximations of the operating characteristic, new acceptance sampling plans are derived that control the overall operating characteristic. The results cover the case of independent sampling as well as the case of dependent sampling. In particular, a modified panel sampling design and the case of spatial batch sampling are studied. The latter allows to detect and analyze local clusters of degraded or damaged modules in photovoltaics.

Part III: Design of Experiments

Vining, **Freeman**, and **Kensler** point out that the reliability of products and processes will become increasingly important in the near future. The paper begins with a review of the current practice for planning reliability experiments. It then reviews some recent work that takes into proper account the experimental protocol.

A basic issue is that most reliability engineers have little training in planning experiments while most experimental design experts have little background in reliability data.

Hassler, **Silvestrini**, and **Montgomery** analyze Bayesian D-optimal designs, which have become computationally feasible to construct for simple prior distributions. They identify several concerns for DB-optimal designs. It is shown that some parameter values give rise to models that have little utility to the practitioner for effect screening. For some generalized linear models such as the binomial, inclusion of such models can cause the optimal design to spread out toward the boundary of the design space. This can reduce the D-efficiency of the design over much of the parameter space and result in the Bayesian D-optimal criterion's divergence from the concerns of a practitioner designing a screening experiment.

The Bayesian Lasso is a variable selection method that can be applied in situations where there are more variables than observations. Thus both main effects and interaction effects can be considered in screening experiments. To apply the Bayesian framework to experiments involving the effect heredity principle, which governs the relationships between interactions and their corresponding main effects, several initial tunings of the Bayesian framework are required. **Noguchi**, **Ojima**, and **Yasui** propose models that do not require the initial tuning values to be specified in advance.

Part IV: Related Areas

The time scale need not be chronological if a failure occurs based on the cumulative damages suffered from its usage or exposure to some risks. The true time scale need not be supported by the observed field reliability data. **Yamamoto** and **Takeshita** use simulations to investigate the properties of time scale models and the sample properties of their estimates. The estimator and the time scale functions are applied to a problem of finding a suitable time scale for field reliability data.

Lenz analyzes the phenomenon that the naive Bayesian classifier may dominate the proper one as happened in clinical studies. The reason for the dominance relation lies in a mix of an a-priori not fixed dimension of the state-space (symptom space) given a disease, the feature selection procedure, and the parameter estimation. Estimating conditional probabilities in high dimensions when using a proper Bayesian model can lead to an "over fitting," a missing value problem, and, consequently, to a loss of classification accuracy.

The level of a workshop on *Intelligent Statistical Quality Control* is determined by the quality of its papers. We believe that this volume truly represents the frontiers of statistical quality control. The editors would like to express their deep gratitude to the members of the scientific program committee, who carefully invited researchers from around the world and refereed all submitted papers:

Olgierd Hryniewicz, Poland
Sven Knoth, Germany
Yoshikazu Ojima, Japan
Wolfgang Schmid, Germany
Ross Sparks, Australia
Peter-Th. Wilrich, Germany
William H. Woodall, U.S.A.
Emmanuel Yashchin, U.S.A.

We would like to cordially thank our host at Sydney, Ross Sparks. He created a perfect atmosphere for inspiring and fruitful discussions. Moreover, we thank Springer, Heidelberg, for the continuing collaboration.

Hamburg, Germany — Sven Knoth
Frankfurt (Oder), Germany — Wolfgang Schmid
July 2014

Contents

Contributors

Barry C. Arnold University of California, Riverside, CA, USA

Giovanna Capizzi Department of Statistical Sciences, University of Padua, Padua, Italy

Darwin Davis Department of Business Administration, Alfred Lerner College of Business & Economics, University of Delaware, Newark, DE, USA

Alessandro Di Bucchianico Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands

Eugenio K. Epprecht PUC-Rio, Department of Industrial Engineering, R. Marquês de S. Vicente Rio de Janeiro, Brazil

Alireza Faraz HEC-management School, University of Liège, Liège, Belgium

Laura J. Freeman Institute for Defense Analyses, Alexandria, VA, USA

Fah Fatt Gan Department of Statistics and Applied Probability, National University of Singapore, Singapore

Rainer Göb Institute for Applied Mathematics and Statistics, University of Würzburg, Würzburg, Germany

K. Govindaraju Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Edgar Hassler Arizona State University, Tempe, AZ, USA

Olgierd Hryniewicz Systems Research Institute, Warsaw, Poland

G. Jones Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

L. Allison Jones-Farmer Department of Information System and Analytics, Miami University, Oxford, OH, USA

Jennifer L.K. Kensler Shell Global Solutions (USA), Houston, TX, USA

Sven Knoth Institute of Mathematics and Statistics, Department of Economics and Social Sciences, Helmut Schmidt University, Hamburg, Germany

Taras Lazariv Department of Statistics, European University Viadrina, Frankfurt(Oder), Germany

Hans - J. Lenz Institut für Statistik und Ökonometrie, Freie Universität Berlin, Berlin, Germany

James Lucas James Lucas and Associates, Wilmington, DE, USA

Kristina Lurz Institute for Applied Mathematics and Statistics, University of Würzburg, Würzburg, Germany

Guido Masarotto Department of Statistical Sciences, University of Padua, Padua, Italy

Thomas McWilliams Decision Sciences Department, LeBow College of Business, Drexel University, Philadelphia, PA, USA

Fadel M. Megahed Department of Industrial and Systems Engineering, Auburn University, Auburn, AL, USA

Douglas C. Montgomery Arizona State University, Tempe, AZ, USA

Manuel Cabral Morais CEMAT & Department of Mathematics, Instituto Superior Técnico, Lisbon, Portugal

Hidehisa Noguchi Department of Industrial Administration, Tokyo University of Science, Chiba, Japan

Yoshikazu Ojima Department of Industrial Administration, Tokyo University of Science, Chiba, Japan

António Pacheco CEMAT & Department of Mathematics, Instituto Superior Técnico, Lisbon, Portugal

Patrícia Ferreira Ramos CEMAT, Instituto Superior Técnico, Lisbon, Portugal

Anne R. Driscoll Department of Statistics, Virginia Tech, Blacksburg, VA, USA

Erwin Saniga Department of Business Administration, Alfred Lerner College of Business & Economics, University of Delaware, Newark, DE, USA

Wolfgang Schmid Department of Statistics, European University Viadrina, Frankfurt(Oder), Germany

Rachel T. Silvestrini Naval Postgraduate School, Monterey, CA, USA

Ross Sparks CSIRO Digital Productivity, North Ryde, NSW, Australia

Ansgar Steland Institute of Statistics, RWTH Aachen University, Aachen, Germany

Jessie Y. Su Department of Statistics and Applied Probability, National University of Singapore, Singapore

Kazuki Takeshita University of Electro-Communications, Tokyo, Japan

Xu Tang Department of Statistics and Applied Probability, National University of Singapore, Singapore

Edwin R. van den Heuvel Department of Mathematics, Eindhoven University of Technology, Eindhoven, The Netherlands

G. Geoffrey Vining Department of Statistics, Virginia Tech, Blacksburg, VA, USA

Peter-Th. Wilrich Institut für Statistik und Ökonometrie, Freie Universität Berlin, Berlin, Germany

William H. Woodall Department of Statistics, Virginia Tech, Blacksburg, VA, USA

Watalu Yamamoto University of Electro-Communications, Tokyo, Japan

Su-Fen Yang National Chengchi University, Taipei, Taiwan

Emmanuel Yashchin IBM, Thomas J. Watson Research Center, Yorktown Heights, NY, USA

Seiichi Yasui Department of Industrial Administration, Tokyo University of Science, Chiba, Japan

Part I
Statistical Process Control

Social Network Monitoring: Aiming to Identify Periods of Unusually Increased Communications Between Parties of Interest

Ross Sparks

Abstract Social networks are increasingly attracting the attention of academic and industry researchers. Monitoring communications within clusters of suspicious individuals is important in flagging potential planning activities for terrorism events or crime. Governments are interested in methodology that can forewarn them of future terrorist attacks or social uprisings in disenchanted groups of their populations. This paper will examine a range of approaches that could be used to monitoring communication levels between suspicious individuals. The methodology could be scaled up to either understand changes in social structure for larger groups of people, to help manage crises such are bushfires in densely populated areas, or early detection of disease outbreaks using surveillance methods. The methodology could be extended into these other application domains that are less invasive of individuals' privacy.

Keywords Communication monitoring • Multivariate EWMA • Social network data

1 Introduction

Security organisations see value in capturing and analysing social media content for the early detection of emerging threats. These could involve either physical threats to facilities, civil unrest, threats of attacks or terrorism. Early detection of emerging risks enables security organisations to vary their surveillance efforts by focusing on areas that minimise negative impacts. Although security organisations are able to collect a massive amount of information, they are only able to process limited amounts due to the total volume of activity. The ability to monitor e-mail, phone and text conversation between two parties has been possible for some time, e.g., tracking suspects of money laundering through journaling. Journaling involves obtaining

R. Sparks (✉)
CSIRO Digital Productivity, Locked Bad 17, North Ryde, NSW 1670, Australia
e-mail: Ross.Sparks@csiro.au

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_1

everything sent and received by suspects. It is therefore possible to monitor the email conversations between two people without them knowing.

The surveillance of user activity within social networks or communication between people generates additional ethical and legal risks. Only the truly suspicious people should be the target of surveillance activities and only aggregated trends in communications should be analysed. It is important that the monitoring technology avoids the targeting of innocent parties as much as possible as well as preserving the privacy of innocent parties when alarming unusual interactions or trends.

Knowing where to focus analytical and surveillance efforts is the challenge. Large volumes of data make it difficult to know where to invest the effort. Other benefits of focusing the surveillance effort on the higher risk aspects of social media are reduced privacy breaches and improving the scalability of the surveillance task. Surveillance of social networks in some countries is a growing activity.

The Department of Homeland Security in the USA searches digital media for information on disasters, suspicious packages, street closures, risks, bomb threats, etc. The popularity of consumer cloud services, such as Face-book, YouTube and Linked-In, provides new targets for security monitoring. Agencies such as the Federal Bureau of Investigation and the National Security Agency routinely collect and monitor information about internet activity for reasons related to national security. The FBI and other government agencies have been monitoring email communications for several years. One of the FBI's earlier email monitoring programs, codenamed "Carnivore", used conventional packet-sniffing technologies to monitor email communications. Packet-sniffing technologies are legal as long as the data are filtered adequately (e.g., see Ohm et al. 2007). Carnivore was designed to harvest only the information sought, allowing all other information to pass through the system, but was criticised for several security flaws.

The Telegraph reported on Friday 24 August 2012 that Ministers in the United Kingdom are preparing to monitor the email exchanges and website visits of every person in the UK. The Sunday Times reported that internet companies will be instructed to install hardware enabling the government's security agencies to monitor any phone calls, text messages, emails and websites accessed in real time. Ministers believe it is essential that the police and security agencies have access to such communications data in order to tackle terrorism and protect the public. The technology is meant to be able to access the content of such communications without a warrant. The legislation would enable it to trace individuals or groups, recording how often and for how long, they are in communication.

Face-book, to abide by the law, cooperates with the police. Face-book's software focuses on conversations between its members. The scanning technology uses historical chat records from criminals and sexual predators to search for phrases in electronic conversations. The relationship analyses need to highlight concerns before Face-book employees examine communications and make the final decision on whether to warn the authorities.

Criminals self-promote and provide evidence on sites such as Face-book and Twitter. The following examples were reported by *The Washington Post*:

- Members of an alleged street gang were reportedly using MySpace to funnel information regarding a potential hit on a federal witness.
- One alleged drug dealer voiced his complaints on an online site regarding how an individual was "watering down the pack" of the hallucinogen PCP.
- Authorities used Face-book to arrest a man wanted in a gang-related stabbing after he posted his new hometown and contact information.

Law enforcement agencies are monitoring social network pages for any gang-related conversations that could be crime-related. Among the most useful items are the lists of friends, photos and contact information. The New York Police Department has a unit whose task is to hunt down criminals using social media networks. The public are using sites like Face-book to express concerns and complaints regarding law enforcement actions in and around where they live thus offering useful information for prevention strategies.

The many forms of communications and the increasing flow of information between groups of people add to the increasing challenge. There is the difficulty with gauging the quality of the information. Developing methods for analysing this information quickly in a way that assesses what is quality information and avoids the influence of biased/wrong information is difficult. For example, initial reports from scenes of emergencies tend to be inaccurate. Perceptions over time change and these are not always documented on social media. The analytical challenge relating to selection biases and disinformation is enormous.

Although the final aim is to come up with a unified approach to monitoring social media interaction, this paper focuses on the smaller issue of trying to identify cells with increased levels of communications indicating some activity is potentially being planned and also flagging who is involved with these plans from the target interest group. Section 2 discusses the methodology. Section 3 introduces the exponentially weighted moving average (EWMA) statistic that is useful for both smoothing the counts in time to reveal trends and for accumulating memory to gain sufficient power to detect significant increases from expected levels of communication. Section 4 outlines the algorithms proposed in this paper for identifying significant increases in levels of communication.

2 The Methodology

We will start very simply by looking at who contacts whom at time period t (taken as a day in this paper). We aggregate over all means of communication, e.g., Twitter, Face-book pages, sms, phone, e-mail, etc. In a simple illustrative example, we monitor the communication levels within a targeted group of 20 individuals. Label them as persons A to T. We monitor the number of times A contacts: B, C, . . ., T; B contacts: A, C, . . ., T; etc. Say we have daily total numbers of directional contacts (e.g., A phones B) between each person in the target group. Figure 1 represents the cells in which we investigate counts over a sequence of time periods. Figure 1 limits

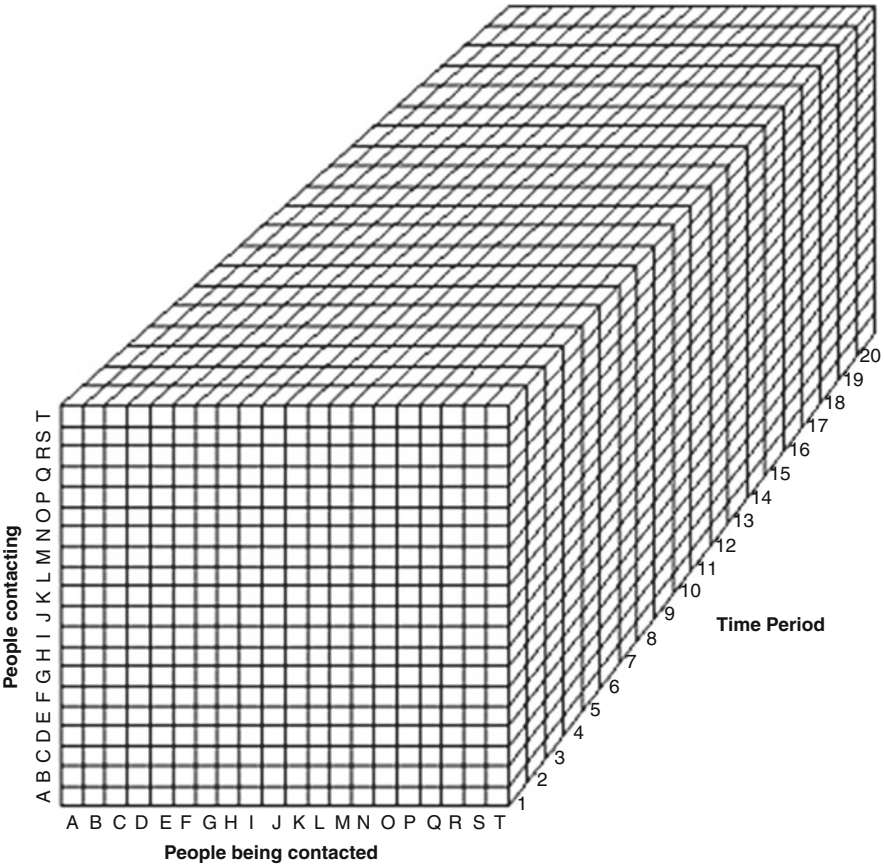


Fig. 1 The target communication network with temporal memory

itself to recording communication counts for the past twenty time periods. Looking at a fixed group of the same 20 people without changes may be limiting so we will explore ways of changing the group members by swapping people in and out based on their communication level with the targeted group.

The diagonal elements of Fig. 1 for each time period are zero because people don't contact themselves, but I prefer the diagonal elements to be the number of times that potential criminals contact close family members. This could be used as a surrogate for a family incident being the reason for an escalation of communication levels with an individual. This paper aims to develop technology that will be useful in identifying a sudden unusual increase in communications between individuals highlighting that some criminal activity or terrorism event is potentially being planned. The remaining sub-sections will propose several ways of achieving the aim of detecting unusual period of increased communication between sub-networks in the population.

Before we outline the approach we define some notation. Let $y_{i,j,t}$ be the number of times the ith person contacted the jth person during time period t, and let $y_{i,i,t}$ be the number times the ith person contacted close family members during time period t. Let $\mu_{i,j,t}$ be the expected value for count $y_{i,j,t}$ at time t. This mean count $\mu_{i,j,t}$ is considered stationary and not a function of time t, although the theory is developed for the non-stationary mean situation.

If spatio-temporal surveillance task is tackled with a lattice structure as in Glaz et al. (2001), then this task looks decidedly similar to that which is outlined in Glaz et al. (2001) (see Fig. 1). Where this task differs from that outlined in Glaz et al. (2001) is that their lattice dimensions are longitude, latitude and time. Each dimension has a natural ordering in this spatio-temporal setting (see also Sparks et al. 2012). When monitoring communication levels in the network the dimensions are people contacting others and time; while time has a natural order, there is no natural ordering of the people in the lattice. This means that the spatio-temporal monitoring technology can't be taken over into the network communication volume monitoring without developing a meaningful order for people in the network. This paper develops a plan that does not require the people to be ordered into neighbours.

2.1 Multivariate EWMA Statistic

The multivariate EWMA statistic (e.g., see Lowry et al. 1992; Sparks 1992) is used to both retain memory of past communication counts as well as the smoothing of the counts to reveal trends. We therefore examine the multivariate EWMAs of both the counts and their means, that is, by calculating

$$\text{ewma}_{i,j,t} = \alpha y_{i,j,t} + (1-\alpha)\text{ewma}_{i,j,t-1}$$

and

$$ewm\mu_{i,j,t} = \alpha\mu_{i,j,t} + (1-\alpha)ewm\mu_{i,j,t-1}$$

where $0 < \alpha < 1$ (Although the means are treated as known here, in practice, they are estimated using historical records, e.g., see Sparks and Patrick 2014). This temporal memory management differs from the scan statistic which advocates a moving time window for retaining temporal memory (e.g., see Kulldorf and Nagarwalla 1995). The temporal memory management approach is similar to the EWMA scan used in Sparks et al. (2012) where it has been shown in certain circumstances to be superior to the moving window scan of $T = 10$.

The scan plan (Tango 1995) uses an exhaustive search of different size windows thus requiring significant computational effort if the targeted group involves several thousand individuals, particularly when the order does not matter. Therefore its application is unrealistic unless it is limited to monitoring small groups of targeted people, e.g., when monitoring a 1,000 person network for crime gangs of 5 then

there are over 8 billion potential gangs of 5 to scan. Therefore technology that is less computational onerous is needed for monitoring several thousand targeted peoples' communications.

2.2 An Algorithm for Detecting Unusually High Communication Networks Based on Order Statistics

An alternative to the scan statistic is considered next. This is used to provide early warnings of significantly increased communication periods for clusters of people in the communication network. The approach tries to identify groups of people (communication networks) with elevated communication levels.

A very simple example is now used below to demonstrate the plan. This assumes that we are examining the daily communication counts between ten people labelled A, B, C, D, E, F, G, H, I and J. Consider the communication counts in Table 1 as those for a particular day. Their respective expected level of communication

Table 1 Simple example of communication counts and expected counts

Contacting \ Contacted	A	B	C	D	E	F	G	H	I	J
A	0	6	5	0	0	0	0	0	0	0
B	4	0	3	0	0	0	0	0	0	0
C	3	3	0	0	0	0	0	0	0	0
D	0	0	0	0	5	2	1	0	0	0
E	0	0	0	2	0	2	0	0	0	0
F	0	0	0	2	0	0	1	0	0	0
G	0	0	0	1	1	1	0	0	0	0
H	0	0	0	0	0	0	0	0	3	2
I	0	0	0	0	0	0	0	2	0	2
J	0	0	0	0	0	0	0	2	2	0
Contacting person	Expected daily communication levels (expected/average number of calls)									
A	0	2.1	1.7	0	0	0	0	0	0	0
B	1.9	0	1.3	0	0	0	0	0	0	0
C	0.6	0.4	0	0	0	0	0	0	0	0
D	0	0	0	0	4.5	2.1	0.6	0	0	0
E	0	0	0	2.5	0	2.40	0.4	0	0	0
F	0	0	0	1.5	0.6	0	0.9	0	0	0
G	0	0	0	0.5	1.0	0.8	0	0	0	0
H	0	0	0	0	0	0	0	0	2.5	2.4
I	0	0	0	0	0	0	0	2.1	0	2.0
J	0	0	0	0	0	0	0	1.9	1.8	0

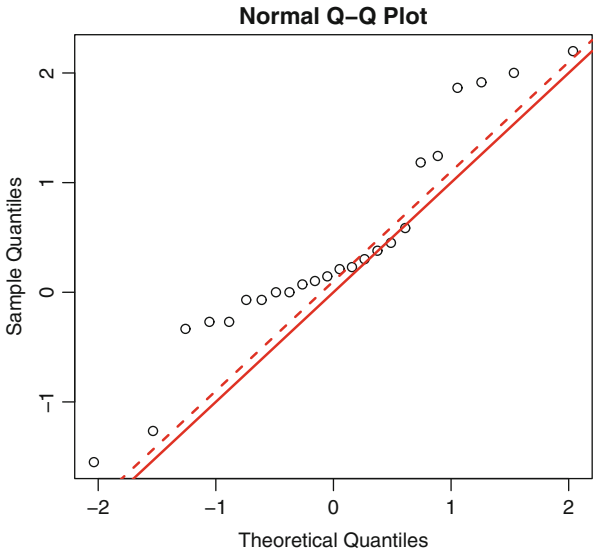


Fig. 2 The qq-plot used to determine the number of people to scan for detecting increased communication levels

is included at the bottom of the table. In other words, Table 1 reports the daily count in rows 2 to 11 of the table, and the expected daily communication levels in rows 13 to 22. We will assume that counts are Poisson distributed and that communications between individuals are independent once we have corrected for their mean. The individual communication counts are investigated in term of their departure from expected. Their signal-to-noise ratios are ordered from largest to smallest. For the example in Table 1, we calculate the individual communication signal-to-noise ratios for those with non-zero expected values using square root of cell counts minus square root of expected values all times 2. This is approximately standard normally distributed for Poisson counts. These signal-to-noise ratios are plotted against standard normal quantiles using the qq-plot (e.g., see Fig. 2 for the data in Table 1). We ignore diagonal cells (because they don't vary). The solid line in Fig. 2 is when the sample matches the theoretical values. The number of high end order statistics above the dashed line determines the number of connections that are significantly higher than expected. For those higher than expected there are six directed communications that are above the dashed line. The departure in these from expected is considered, i.e., total counts for these are $6+5+4+3+3+3 = 24$ and total expected values are $2.1 + 1.7 + 1.9 + 1.3 + 0.6 + 0.4 = 8$ giving a signal-to-noise ratio of 5.7. This is then compared to a threshold (found by simulation) to see if communication has elevated significantly. The size of the threshold determines the average time between signals when there is no change in communication levels. The approach can be scaled up to the more complex task where the average length of conversations between individuals, and text mining are included. An example is the monitoring of the average length of conversations and counts of key phrases of the conversations simultaneously with communication counts. This would be useful to

assess the nature of the increased communications as well as identifying the nature of changes in conversations. The main advantage of the group of ordered statistics approach is that the methodology is invariant of the ordering of the target group, whereas applying the scan statistic is highly dependent on the ordering of people in the lattice structure.

3 Simulated Example of Application

Assume we are monitoring 1,000 people who have been convicted of a violent crime before and are currently not in jail. We will simulate a network of criminal gangs involving 100 independent groups of ten individuals; however, this knowledge will be treated as unknown. The mean communication daily counts between individuals within a gang are taken as uniform on the interval of 0.1 to 3 during periods when no crime is being planned, and 0.0001 for individuals between gangs. A step change in communications of δ for all individuals in the gangs will be simulated at the start of their plans for the crime. This may be a little unrealistic, but we have no information on common communications trends while gangs plan a crime; besides this may vary from gang to gang. The plan used $m = 25$, 50, 75 or 100 rather than using the qq-plot to determine the level of aggregation.

The following outbreaks, which are always taken as within the social network, were considered: (the plan would detect it much sooner if gangs spanned networks that usually did not communicate)

Scenario 1: One cell of ten individuals. (Total communication mean count = 136.61).
Scenario 2: Two neighbouring cells of ten individuals. (Total communication mean count = 275.25).
Scenario 3: Two non-neighbouring cells of ten individuals. (Total mean count = 294.76).
Scenario 4: Three independent cells involving 7 of the 10 within each cell. (Total mean count = 204.2).
Scenario 5: Four independent cells involving 6 of the 10 within each cell. (Total mean count = 194.95).
Scenario 6: Four non-neighbouring cells of ten individuals (Total mean count = 385.82).

It is surprising that the best level of aggregation is for the 25 highest order statistics for all shifts in Tables 2, 3 and 4. The relative advantage for selecting $m = 25$ is less clear if the communication outbreak involves more people, e.g., for Scenarios 5 & 6 there is less of a difference between the choice of $m = 25$ and $m = 50$. For improved diagnosis the best choice is to have m nearly as large as the number of cells in the outbreak. Therefore, the choice of m that is best for early detection and diagnosis unfortunately differ.

Table 2 Average days to detection of the hidden simulated outbreaks for Scenarios 1 & 2 where δ is the increase in the mean number of calls for each individual in the network planning a crime

Scenario	1				2			
δ \ m	25	50	75	100	25	50	75	100
0.0	100.8	99.8	100.3	100.2	101.4	100.2	100.3	100.2
0.5	21.7	22.2	22.6	23.8	15.9	16.0	16.2	16.2
1.0	8.8	9.7	10.5	10.9	7.1	7.5	7.8	8.1
2.0	4.5	5.0	5.5	5.9	3.8	3.9	4.3	4.8
3.0	3.4	3.7	3.9	4.1	2.8	2.9	2.9	3.4
4.0	2.8	2.9	2.9	3.4	1.9	1.9	2.6	2.9
6.0	1.9	2.0	2.0	2.7	1.6	1.6	1.9	2.0

Table 3 Average days to detection of the hidden simulated outbreaks for Scenarios 3 & 4 where δ is the increase in the mean number of calls for each individual in the sub-network planning a crime

Scenario	1				2			
δ \ m	25	50	75	100	25	50	75	100
0.0	101.4	99.8	100.3	100.2	101.4	100.2	100.3	100.2
0.5	15.8	15.9	16.3	16.9	15.6	16.4	16.5	16.8
1.0	7.2	7.4	8.0	8.2	7.1	7.5	8.1	8.3
2.0	3.8	3.9	4.1	4.8	3.6	3.9	4.2	4.8
3.0	2.8	2.9	2.9	3.4	2.9	3.0	3.0	3.2
4.0	1.9	1.9	2.5	2.9	2.0	2.0	2.5	2.9
6.0	1.6	1.5	1.9	2.0	1.8	1.8	1.9	2.0

Table 4 Average days to detection of the hidden simulated outbreaks for Scenarios 5 & 6 where δ is the increase in the mean number of calls for each individual in the network planning a crime

Scenario	1				2			
δ \ m	25	50	75	100	25	50	75	100
0.0	101.4	99.8	100.3	100.2	101.4	100.2	100.3	100.2
0.5	14.7	14.8	15.5	15.6	12.0	12.1	12.3	12.6
1.0	6.6	7.3	7.5	8.1	5.4	5.9	6.1	6.4
2.0	3.2	3.9	3.9	4.8	2.9	2.9	3.4	3.8
3.0	2.2	2.9	2.9	3.4	2.2	2.2	2.6	2.9
4.0	1.9	2.0	2.0	2.9	1.9	1.9	2.0	2.0
6.0	1.7	1.6	1.7	2.0	1.6	1.6	1.6	1.6

As long as the increase in communication levels is equal to their normal value (i.e., $\delta \geq 1.5$), then the planning of a crime is detected within a week, which is probably sufficient to intervene before the crime is committed. Thus this monitoring tool could be useful for preventing serious criminal activities.

4 Concluding Remarks

Many systems are designed to collect social network data. Analytical methods are needed to turn these large volumes of data into information. This paper focuses on surveillance techniques that aim at detecting changes in communication volume within networks of individuals. Although the focus of the paper is on applications relating to the risk of crime or terrorism, it can easily be used in other applications, e.g., for monitoring the number of people from different groups of a population that visit several different blog sites per week.

Social media information is also useful for business intelligence reasons. According to Gartner (an information technology company focusing on social media) by 2015 more than half of companies are likely to be eavesdropping on the publics' e-conversations on social media trying to assess security risks to their organisations. Many public relations firms provide social media monitoring as a standard client service. For example, if companies do not directly invest in social media, they will miss information when someone, somewhere, starts talking about either their brand, staff, products, price, or value. Even knowing what people are saying about competitors is a rich source of information. Large companies that operate globally are interested in how innovation links to network communications within their organisations and across organisations (e.g., see Rothwell 1992). Therefore open access data inside and outside organisations can uncover new sources of value and help drive performance improvements and innovation. Social media monitoring is evolving towards real-time data-driven business improvement based on "real-time" customer insights.

Most organisations do not want to set up an environment where they have employees looking at private communications, so it is important that they use technologies that preserve privacy and have low false-positive rates. Focusing on mass trends rather than individual behaviour is a good start.

Social media monitoring is vital not only for security reasons, but also for improving business intelligence or improving an individual's health management with targetted advice. This area of research is not without its technical, ethical and legal challenges, however.

References

Glaz, J., Naus, J. & Wallenstein, S. (2001). *SCAN statistics*. New York: Springer.

Kulldorff, M. & Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. *Statistics in Medicine, 14(8)*, 799–810.

Lowry, C. A., Woodall, W. H., Champ, C. W. & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics, 34*(1), 46–53.

Ohm, P., Sicker, D. & Grunwald, D. (2007). Legal issues surrounding monitoring during network research. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC'07*. San Diego, California, October 24–26. (2007).

Rothwell, R. (1992). Successful industrial innovation: critical factors for the 1990s. *R & D Management, 22*, 221–240.

Sparks, R. S. (1992). Quality control with multivariate data. *Australian Journal of Statistics, 34*, 375–390.

Sparks, R. S., Okugami, C. & Bolt, S. (2012). Outbreak detection of spatio-temporally smoothed crashes. *Open Journal of Safety Science and Technology, 2*(3), 98–107.

Sparks, R., & Patrick, E. (2014). Detection of multiple outbreaks using spatio-temporal EWMA-ordered statistics. *Communications in Statistics-Simulation and Computation, 43*(10), 2678–2701.

Tango T. (1995). A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine, 14*(21–22), 2323–2334.

Some Recent Results on Monitoring the Rate of a Rare Event

William H. Woodall and Anne R. Driscoll

Abstract A growing number of applications involve monitoring with rare event data. The event of interest could be, for example, a nonconforming manufactured item, a congenital malformation, or an industrial accident. The most common approaches for monitoring such processes involve using an exponential distribution to model the time between the events or using a Bernoulli distribution to model whether or not each opportunity for the event results in its occurrence. The use of a sequence of independent Bernoulli random variables leads to a geometric distribution for the number of non-occurrences between the occurrences of the rare events. One surveillance method is to use a power transformation on the exponential or geometric observations to achieve approximate normality of the in-control distribution and then use a standard individuals control chart. We add to the argument that use of this approach is very counterproductive and cover some alternative approaches. We discuss the choice of appropriate performance metrics. The strong adverse effect of Phase I parameter estimation on Phase II performance of various charts is then summarized. In addition, the important practical issue of the effect of aggregation of counts over time, some generalizations of standard methods, and some promising research ideas are discussed.

Keywords Impact of data aggregation • Monitoring geometric distribution • Phase I parameter estimation • Power transformation

1 Introduction

In an increasing number of applications interest is in the monitoring of a relatively rare event. It is often assumed that the practitioner has the results of a sequence of independent Bernoulli random variables, where a value of one indicates the event occurred and a value of zero indicates nonoccurrence of the event. Thus the number of trials between events has a geometric distribution for a stable process. Under

W.H. Woodall (✉) • A.R. Driscoll
Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0439, USA
e-mail: bwoodall@vt.edu; agryan@vt.edu

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_2

this scenario one wants to monitor the event probability. In other applications it is commonly assumed that the time between events has an exponential distribution and one monitors the average time between events. Most often the focus is on detecting process deterioration, i.e., an increase in the probability of the adverse event or a decrease in the average time between events.

Szarka and Woodall (2011) reviewed the extensive number of methods that have been proposed for monitoring processes using Bernoulli data. Generally, it is difficult to better the performance of the Bernoulli cumulative sum (CUSUM) chart of Reynolds and Stoumbos (1999). The Bernoulli and geometric CUSUM charts can be designed to be equivalent, as discussed by Szarka and Woodall (2012). With respect to monitoring the mean of an exponential distribution it is difficult to outperform the exponential CUSUM chart studied by Lucas (1985) and Gan (1994). Levinson (2011) argued that control charts should not be used with healthcare rare event data because in many situations there is an assignable cause for each error, e.g., each hospital-acquired infection or serious prescription error, and each incident should be investigated. We agree that serious adverse events should be investigated whether or not they result in a control chart signal. The investigation of rare adverse events, however, and the implementation of process improvements to prevent future such errors, does not preclude using a control chart to determine if the rate of such events has increased or decreased over time. In fact, a control chart can be used to evaluate the success of any process improvement initiative.

2 Performance Metrics

The choice of appropriate performance metrics for comparing surveillance schemes for monitoring Bernoulli and exponential data is quite important. The usual Average Run Length (ARL) metric refers to the average number of points plotted on the chart until a signal is given. This metric is most clearly appropriate when the time between the plotted points is constant. For exponential and geometric random variables each plotted point corresponds to the occurrence of the event of interest. Thus the ARL is the expected number of events until the chart signals. If the event is quite adverse such as a serious accident or medical error, then this interpretation of the ARL is very useful even though the time between events varies. If the process does deteriorate, then we would like to detect it with as few adverse events as possible.

In some cases, such as in monitoring the number of near-miss accidents, it may be informative to use a metric that reflects the actual time required to obtain an out-of-control signal. Thus one can consider the number of Bernoulli trials until an out-of-control signal is given for Bernoulli data, leading to its average, the ANOS. The ANOS will be proportional to the average time before a signal if the rate at which the Bernoulli trials are observed is constant over time. For exponentially distributed data one could consider the average time to signal, the ATS. If the process is stable, then ANOS = ARL/ p and ATS = ARL* θ, where p and θ are the Bernoulli probability and the exponential mean, respectively.

To assess out-of-control performance we believe it is most realistic to consider steady-state performance where the shift in the parameter occurs at some time after monitoring has begun. This approach was discussed by Zhang et al. (2007) and Szarka and Woodall (2011), among others. Under this scenario one cannot easily convert the ARL metric to the ANOS and ATS metrics. Consideration of steady-state performance of competing methods is important because some methods have an implicit headstart feature that results in good zero-state performance, but poor steady-state performance. An example is the sets method of Chen (1978) studied by Capizzi (1994) and Sego et al. (2008). The basic sets method based on Bernoulli data signals a rate increase if the most recent k geometric waiting times are all below a specified constant. The sets method and its variations have been proposed for monitoring the rate of congenital malformations.

In addition, we note, as an example, that Liu et al. (2006) used zero-state performance comparisons instead of steady-state comparisons. With zero-state performance comparisons the shift in the parameter is assumed to occur at the start of monitoring or when the chart statistic is at its initial value. They considered charts based on the times between r events, where $r > 1$, that will seem to be more effective than they would be for the delayed shifts assumed for steady-state analysis. It is important to note that with these methods one waits until r events have occurred to obtain one count, waits until r additional events occur to obtain the next count, and so forth. When methods are based on waiting times which are aggregated over time like this or when comparing methods based on counts taken at differing levels of aggregation over time, then the ARL metric is not meaningful and either the ATS or ANOS metric should be used.

3 The Use of a Power Transformation

We first consider the time-between-occurrence control chart as discussed by Montgomery (2013, pp. 332–333). We assume that when the process is in control that the successive times are independent exponential random variables with a mean of θ_0. Montgomery suggested the transformation method proposed by Nelson (1994) under which the exponential random variables, $X_1, X_2, X_3, \ldots, X_m$ in Phase I are transformed using a power transformation $Y = X^{0.2777}$ in order to achieve approximate normality. After the transformation is made then one uses the standard individuals control chart with 3-sigma control limits based on the average moving range. This chart is frequently referred to as a "t-chart" in the literature, where t refers to "time."

Even though this seems like a reasonable approach, the performance of the resulting t-chart is quite poor. McCool and Joyner-Motley (1998) studied Nelson's method and the use of a logarithmic transformation. More recently, Santiago and Smith (2013) showed that if the in-control parameter is assumed to be known then with 3-sigma control limits the out-of-control average run length (ARL) values when there are decreases in the mean are far higher than the in-control ARL. This

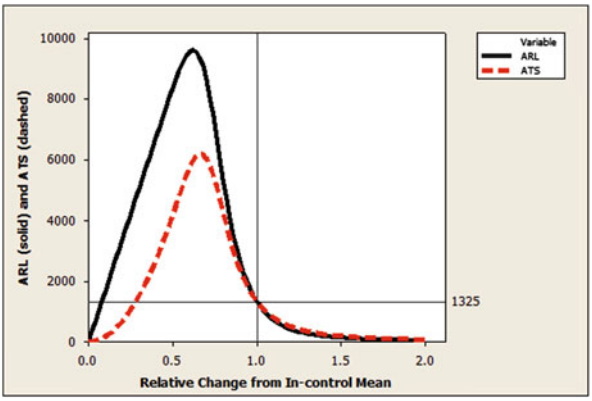


Fig. 1 Exact ARL and ATS values for the t-Chart based on Nelson (1994) Transformation, ($\theta_0 = 1$ and $ARL_0 = 1325$)

striking behavior is illustrated by the solid line in Fig. 1. The average time between events must decrease to a very small fraction of the in-control value θ_0 in order for a signal to be given on average with fewer plotted points.

Even though there are more points on average plotted on the chart before a signal is given when the mean decreases, one should note that there is a shorter average time period between the plotted points. Thus, in some cases it may be informative to consider the ATS metric, which is ARL*θ. This metric gives a better indication of how quickly a signal is given. One can see from the dashed line we added in Fig. 1 that the t-chart does not do well in detecting decreases in the exponential mean quickly, but the performance based on the ATS metric is not as poor as for the ARL metric. One could adjust the control limits of the chart to reduce the amount of ARL-bias, but this would take away the simplicity of the approach. Szarka and Woodall (2011) discussed this issue for similar charts.

We next consider the case in which the value of the in-control parameter p_0 is unknown for a Bernoulli process. Table 1 shows the percentage of the time one will have a useful lower control limit (LCL) if one uses Nelson (1994) power transformation with geometrically distributed data and the standard individuals chart with limits based on the average moving range. Each geometric observation is the number of trials required for the event to occur. Because interest is most likely to be in detecting increases in the probability of the event probability, and thus decreases in the mean of the geometric waiting time, there is a very good chance that the control chart resulting from the use of Nelson (1994) transformation method will not have an LCL greater than one and not be able to detect such process deterioration. Here p_0 is the in-control probability of the event of interest occurring. We based each percentage in the table on 100,000 simulations of Phase I data containing m geometric observations. This table shows that as events become rarer, the probability of LCL > 1 increases, but still remains below 70 % even for $p_0 = 0.000001$ for $m = 100$. Note that the probability of a useful LCL decreases as m increases for all values of p_0 except $p_0 = 0.000001$. We do not know the reason for this phenomenon.

Table 1 Proportion of geometric Shewhart charts based on a Phase I sample of size m and Nelson's transformation which yield a useful LCL

p_0	$m = 25$	$m = 50$	$m = 100$
0.05	0.0095	0.0004	0.0000
0.01	0.0930	0.0277	0.0032
0.005	0.1545	0.0698	0.0182
0.001	0.3090	0.2314	0.1437
0.0001	0.4755	0.4535	0.4287
0.00005	0.5079	0.5018	0.4964
0.000001	0.6128	0.6466	0.6973

Table 2 Proportion of exponential Shewhart charts based on a Phase I sample size of m and Nelson's transformation which yield a useful LCL

$m = 25$	$m = 50$	$m = 100$
0.66029	0.71463	0.78248

Table 2 shows the corresponding percentages for exponentially distributed data, leading to the same conclusion. For the exponential distribution, we can assume without loss of generality that $\theta_0 = 1$ since for any other in-control value we can rescale the observations by dividing by θ_0.

Since the transformation method does not work well, other approaches must be used. A wide variety of methods have been proposed in the literature, including Shewhart, CUSUM, and EWMA charts based on sequences of exponential or Bernoulli data. Most papers in the literature are on Phase II methods with the in-control parameter value assumed to be known. In practice the practitioner must estimate the in-control parameters, so the effect of parameter estimation is discussed in the next section.

4 Performance Metrics

Steiner and MacKay (2004) pointed out that extremely large Phase I sample sizes are needed in order to establish control limits for high quality Bernoulli processes. If one uses Nelson's (1994) recommendation of basing the estimator of the in-control probability p_0 on 24 observed events in Phase I, then if $p_0 = 0.000005$ (i.e., 5 ppm) the expected number of items in Phase I would be 4.8 million units. Steiner and MacKay (2004) also pointed out that the out-of-control expected number of items to signal can also be impractically large.

In some respects the problem is worse than Steiner and MacKay (2004) portray it. Zhang et al. (2013) studied the effect of estimation error on the Shewhart-type geometric chart. For practitioners to have confidence in their control chart design in Phase II, they must have Phase II charts with the mean in-control ARL (or other metric) near the desired value with sufficiently small variation about that value. Figure 2 shows the average in-control ARL and Fig. 3 the standard deviation of the

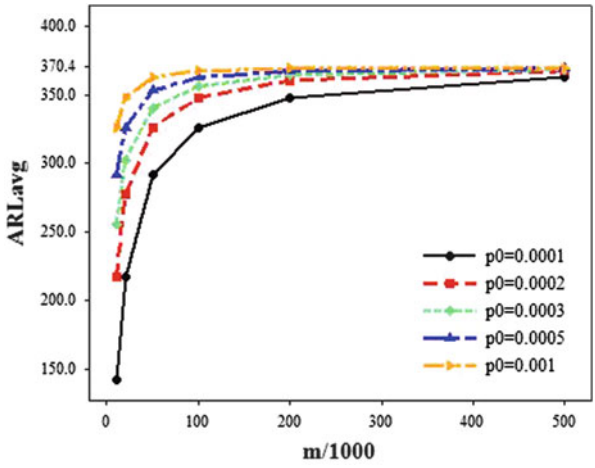


Fig. 2 The expected in-control ARL for a given in-control proportion p_0 and Phase I sample size m

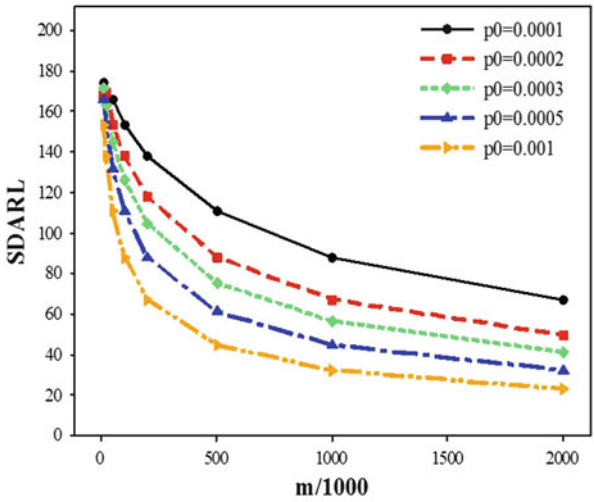


Fig. 3 The standard deviation of the in-control ARL for a given in-control proportion p_0 and Phase I sample size m

in-control ARL when the desired in-control ARL is 370.4. Note that the horizontal scales of Figs. 2 and 3 are not the same. Also, if needed, the in-control ANOS = in-control ARL/ p_0. The number of Phase I Bernoulli observations is denoted by m.

Even though the average in-control ARL converges relatively quickly to the desired value as m increases in Fig. 2, the variation in the in-control ARL converges very slowly to zero in Fig. 3. This means that it is very difficult to design the geometric control chart to have a specified in-control performance. The situation becomes worse for lower values of p_0. Using Nelson's (1994) recommendation of using a sample size required to observe 24 events in Phase I is inadequate. Lee et al. (2013) showed an even larger adverse effect of estimation error in designing the

Table 3 The average and standard deviation (in parentheses) of the in-control ARL for one-sided exponential CUSUM charts based on a Phase I sample size m and a shift size of interest δ

m	$\delta = 0.8$	$\delta = 0.5$	$\delta = 0.2$
50	10740.7	1157.3	642.6
	(213171.2)	(2561.7)	(474.4)
500	611.0	538.8	512.3
	(402.5)	(197.4)	(101.9)
10,000	504.8	501.8	500.6
	(60.8)	(38.9)	(21.9)
50,000	500.9	500.4	500.1
	(26.8)	(17.3)	(9.8)
100,000	500.5	500.2	500.1
	(18.9)	(12.2)	(6.9)

Bernoulli CUSUM chart. The effect of estimation error on the Bernoulli CUSUM chart increases as the targeted shift size in the underlying proportion decreases. The overall conclusion here is that practitioners should not expect charts based on estimated in-control parameters to perform like they would if these parameters were known, even if the Phase I sample is large.

Zhang et al. (2014) studied the effect of estimation error on the one-sided exponential CUSUM chart designed to detect a change in the mean from θ_0 to $\delta\theta_0$, where $0 < \delta < 1$. Some of their results are shown in Table 3, where one can see that the effect of estimation error is greater if one wishes to detect a smaller shift in the mean time between events. A large number of in-control observations are needed to have the expected in-control ARL near the targeted value of 500. In order for the variation of the in-control ARL to be reasonably low, however, an inordinate number of Phase I observations is required. Note that, if needed, the in-control ATS = in-control ARL* θ.

5 The Effect of Aggregating Data

Data aggregation is frequently done when monitoring rare events and for count data generally. For example, one might monitor the number of accidents per month in a plant or the number of patient falls per week in a hospital. Montgomery (2013, pp. 332–333) indicated that when there are many samples (i.e., time periods) with no events, then a c-chart is not useful and one should use a time-between-occurrence control chart. Schuh et al. (2013) showed, however, that there can be significantly long expected delays in detecting process deterioration when data are aggregated over time even when there are few samples with zero events. One can always aggregate data over long enough time periods to avoid zero counts, but the consequence is slower detection of increases in the rate of the adverse event.

We assume a homogeneous Poisson process with an average time between events of θ time units. Under this model the interarrival times are independently distributed exponential random variables with mean θ. For an aggregation period of length d time units, counts are independent and Poisson distributed with a mean of $\mu = d/\theta$. The time unit considered can vary depending on the application. We assume that the in-control value of the parameter is $\theta = \theta_0 = 1$ and that we wish to detect only decreases in the average time between events. Our results, however, are generalizable. For example, if events occur at an average rate of 4 per 28-day period, then this is equivalent to one event per week. If we consider an aggregation period d of seven-time units, then this would correspond to aggregating the data over a 7 week (or 49 day) period. Thus, it is only necessary to consider the $\theta_0 = 1$ case. Basically, the expected count is always one for some time period.

Instead of monitoring the exponentially distributed time-between-event data, it is very common to monitor instead the Poisson distributed data obtained by aggregating the event data over time intervals of a specified length. Schuh et al. (2013) showed that there is a price to be paid, however, for the data aggregation. As an example, Table 4 shows some of their results when monitoring a Poisson process with an average of one time unit between events of interest. Table 4 shows the additional number of adverse events that would be expected to occur before a signal is given for a decrease in the average time between events to given values θ, where $\theta < \theta_0 = 1$. The comparison is to the chart with the lowest steady-state

Table 4 Number of additional adverse events expected for each less-effective chart when detecting various decreases in the average time between events.

θ	POIS30	POIS4	POIS7	POIS1	EXP	Performance for best chart
0.95	-	105.8	178.2	57.7	54.0	2239.37
0.9	-	127.2	201.3	95.9	103.3	1084.78
0.85	-	87.4	130.8	72.9	70.1	551.06
0.8	-	44.4	68.6	33.1	36.9	298.87
0.75	-	19.9	28.3	8.9	10.7	178.13
0.7	8.0	9.1	10.9	2.6	-	113.00
0.65	7.8	8.9	8.0	0.5	-	73.54
0.6	18.8	9.5	6.7	0.5	-	51.67
0.55	23.6	10.5	6.4	0.5	-	38.73
0.5	28.8	11.8	6.6	0.6	-	30.60
0.45	34.0	13.6	6.9	0.9	-	25.11
0.4	39.8	16.0	7.8	1.3	-	21.25
0.35	46.0	18.9	8.9	1.1	-	18.57
0.3	53.7	23.0	10.7	1.3	-	16.33
0.25	63.6	28.0	12.8	1.2	-	14.80
0.2	79.5	35.5	17.0	2.0	-	13.00

(The — symbol appears when a chart has the lowest steady-state ATS). For reference the average number of adverse events until detection is given in the far right column for the best performing chart

ATS for the given shift. The exponential CUSUM chart is compared here to Poisson CUSUM charts based on four different levels of aggregation. In the table POIS14, for example, refers to the Poisson CUSUM chart based on counts aggregated over 14 time periods. All charts are designed to optimally detect a sustained shift to $\theta = 0.5$ under zero-state conditions. The in-control Average Time to Signal (ATS) values are near 4,500 for all five charts. The out-of-control performance is based on the steady-state ATS values with a shift to the out-of-control parameter value at some point in time after monitoring has begun. For more details, the reader is referred to Schuh et al. (2013).

Some other work has been done on the effect of aggregating count data. Reynolds and Stoumbos (2000) compared the performance of Bernoulli CUSUM charts to that of binomial CUSUM charts for aggregated samples of a specified size, finding that Bernoulli CUSUM charts showed better overall performance, especially for detecting large increases in the rates of nonconforming items. Szarka and Woodall (2011) further discussed this topic in their review of Bernoulli-based charts. Another type of aggregation is to wait until one has observed a given number of events before updating a control chart based on a proportion or waiting time. See, for example, Zhang et al. (2007) and Dzik et al. (2008). This type of aggregation, however, does not appear to delay the detection of process changes nearly as much as aggregating data over fixed time periods.

6 Some Generalizations

There have been some generalizations of the Bernoulli and exponential distribution-based methods we have discussed. Ryan et al. (2011), for example, extended the monitoring of Bernoulli data to monitoring with more than two categories through use of the multinomial distribution. Having more than two categories provides more information about the process.

Steiner et al. (2000) proposed a widely used method for monitoring Bernoulli data when the in-control probability varies over time. Their risk-adjusted CUSUM method is used to monitor surgical and other health-related outcomes while adjusting for patient risk factors. In these applications one typically works with serious events, such as death within 30 days of surgery, but the overall rate is frequently too high, e.g., around 0.01, for the adverse event to be considered rare. See Woodall et al. (2015) for a review of risk-adjusted monitoring.

Mousavi and Reynolds (2009) considered the monitoring of autocorrelated Bernoulli data where adverse events are more likely to follow other adverse events than to follow trials where the event does not occur. Finally, as generalizations of methods designed for exponentially distributed time-between-event data, methods have been proposed for Weibull and gamma distributed time-between-event data. See, for example, Xie et al. (2002) and Zhang et al. (2007).

7 Research Ideas

Given the disappointing performance of many of the methods for monitoring the rate of a rare event, it is important to identify improved methods if at all possible. We recommend the alternative methods proposed by Saleh et al. (2015) and others. The focus of much of the research on monitoring rare events has been on detecting sustained step shifts corresponding to process deterioration. Additional research on other forms of process deterioration, such as drifts, and on detecting process improvement is needed.

We believe that the adverse effect of aggregating data over time has not been fully appreciated in practice and more research work is needed on this topic. Only a couple of the most basic scenarios for count data have been studied. Some interesting and important topics include the effect of data aggregation on the performance of charts based on seasonal or autocorrelated count data, risk-adjusted data, multinomial and multivariate data, and Weibull-distributed time-between-event data.

Virtually all of the work on monitoring the rate of rare events is based on the assumption that there is a sustained shift in the rate. In some applications the rate change may be transient. In this scenario other performance metrics would be needed, such as the probability of detecting the process shift during the transient period. The effect of data aggregation over time might be larger if shifts in the parameter are not sustained.

As reviewed by Szarka and Woodall (2011), there are several dozen papers on Phase II methods for monitoring a Bernoulli probability. Even though recent research has shown that very large Phase I samples are needed, relatively little work has been done on Phase I analysis to check the adequacy of the Bernoulli model and the stability of the process. Some related references are the papers by Pettitt (1980), Worsley (1983), Wallenstein et al. (1994), Bell et al. (1994), Borror and Champ (2001), Balakrishnan et al. (2001), Krauth (2003), and Tikhomirova and Christyakov (2010). It is not clear which approach, or combinations of approaches, is the best. Similarly, work on the Phase I analysis of exponential data is also needed. The Phase I methods of Jones and Champ (2002) and Dovoedo and Chakraborti (2012) have quite low power to detect shifts in the exponential mean.

In many applications the event of interest may vary in severity. The event of interest may be an industrial accident, for example, but the impact and consequences of accidents vary. How should both rate and severity be monitored? A related paper is that of Wu et al. (2010).

Although we and Xie et al. (2010) have reviewed some of the work on monitoring continuous time-between event data, we believe a more thorough review is needed of the rather extensive literature on this subject. This review would be similar to that done by Szarka and Woodall (2011) for monitoring with Bernoulli data.

Toubia-Stucky et al. (2012) proposed a Bayesian approach to monitoring the proportion associated with Bernoulli-trials. Although a Bayesian approach may seem appealing, their method does not seem to be a viable alternative to the frequentist methods due to the required assumptions and its ad hoc construction.

Finally,Wheeler (2011) was very critical of using either the Bernoulli or exponential models and preferred the use of an individuals control chart with empirically determined 3-sigma control limits based on the median moving range. He advocated plotting either the time-between-event data or converting the time-between event data to an "instantaneous rate." For example, if an adverse event occurred after 110 days, this would be equivalent to an instantaneous rate of (1/110)*365 = 3.32 events per year. Study of these methods is needed since Wheeler (2011) gave only a case study illustration of his proposed methods. Our preliminary investigation shows, however, that his methods have an unacceptably large false alarm rate. His recommendation of using a minimum of five events to determine the control limits will likely prove to lead to highly unpredictable chart performance.

8 Conclusions

We have provided a review of some recent results on the monitoring of rare events along with our perspective. This recent work has demonstrated that the effect of estimation error in Phase I is more severe than for other charts studied in the literature (Jensen et al. 2006). There are compelling arguments for the use of steady-state performance metrics. In addition, recent work has demonstrated that aggregating event data over fixed time intervals, as frequently done in practice, can result in significant delays in detecting increases in the rate of adverse events. We believe that the monitoring of the rate of rare events is an important and challenging area and have offered some research ideas.

We agree with Steiner and MacKay (2004) that the monitoring of the rate of rare events is indeed a very difficult problem. Steiner and MacKay (2004) proposed solution to many of these difficulties is to use logistic regression or some other approach to try to identify a continuous underlying variable, if possible, to monitor instead of tracking simply the occurrence or non-occurrence of events. There could be considerably more information in such an underlying continuous variable, which could lead to more effective monitoring.

Acknowledgements The authors appreciate the helpful comments of a referee, Joel Smith of Minitab, Inc. and the following Virginia Tech graduate students: Rebecca Dickinson, Gregory Purdy, Sarah Richards, Mohammed S. Shafae, Hongyue Sun, Wenmeng Tian, and Xiang Zhang.

References

Balakrishnan, N., Brito, M. R., & Quiroz, A. J. (2001). Using large order statistics of runs for tracking a changing Bernoulli probability. *Communications in Statistics – Theory and Methods, 31*(5), 719–732.

Bell, C., Gordon, C., & Pollak, M. (1994). An efficient nonparametric detection scheme and its application to the surveillance of a Bernoulli process with unknown baseline. In E. Carlstein, H. -G. Müller, & D. Siegmund (Eds.), *Change-point problems*. IMS Lecture Notes – Monograph Series (Vol. 23, pp. 7–27). Hayward, CA: Institute of Mathematical Statics.

Borror, C. M., & Champ, C. W. (2001). Phase I control charts for independent Bernoulli data. *Quality and Reliability Engineering International, 17*(5), 391–396.

Capizzi, G. (1994). On stopping rules for the sets scheme: An application of the success runs theory. In E. Carlstein, H. -G. Müller, & D. Siegmund (Eds.), *Change-point problems*. IMS Lecture Notes – Monograph Series (Vol. 23, pp. 66–77). Hayward, CA: Institute of Mathematical Statics.

Chen, R. A. (1978). Surveillance system for congenital malformations. *Journal of the American Statistical Association, 73*, 323–327.

Dovoedo, Y. H., & Chakraborti, S. (2012). Boxplot-based Phase I control charts for time between events. *Quality and Reliability Engineering International, 28*, 123–130.

Dzik, W. S., Beckman, N., Selleng, K., Heddle, N., Szczepiorkowski, Z., Wendel, S., & Murphy, M. (2008). Errors in patient specimen collection: Application of statistical process control. *Transfusion, 48*, 2143–2151.

Gan, F. F. (1994). Design of optimal exponential CUSUM control charts. *Journal of Quality Technology, 26*(2), 109–124.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology, 38*(4), 349–364.

Jones, L. A., & Champ, C. W. (2002). Phase I control charts for times between events. *Quality and Reliability Engineering International, 18*, 479–488.

Krauth, J. (2003). Change-points in Bernoulli trials with dependence. In W. Gaul, & M. Schader (Eds.), *Between data science and everyday web practice*. Heidelberg, Germany: Springer.

Lee, J., Wang, N., Xu, L., Schuh, A., & Woodall, W. H. (2013). The effect of parameter estimation on upper-sided Bernoulli CUSUM charts. *Quality and Reliability Engineering International, 29*, 639–651.

Levinson, W. A. (2011). Are control charts suitable for health care applications? Quality Digest. Available at http://www.qualitydigest.com/inside/health-care-article/are-control-charts-suitable-health-care-applications.html.

Liu, J. Y., Xie, M., Goh, T. N., & Sharma, P. R. (2006). A comparative study of exponential time between events charts. *Quality Technology & Quantitative Management, 3*(3), 347–359.

Lucas, J. M. (1985). Counted data CUSUM's. *Technometrics, 27*(2), 129–144.

McCool, J.I., & Joyner-Motley, T. (1998). Control charts applicable when the fraction nonconforming is small. *Journal of Quality Technology, 30*(3), 240–247.

Montgomery, D. C. (2013). *Introduction to statistical quality control* (7th ed.). Hoboken, NJ: John Wiley.

Mousavi, S., & Reynolds, M. R., Jr. (2009). A CUSUM chart for monitoring a proportion with autocorrelated binary observations. *Journal of Quality Technology, 41*(4), 401–414.

Nelson, L. S. (1994). A control chart for parts-per-million nonconforming items. *Journal of Quality Technology, 26*, 239–240.

Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika, 67*(1), 79–84.

Reynolds, M. R., Jr., & Stoumbos, Z. G. (1999). A CUSUM chart for monitoring a proportion when inspecting continuously. *Journal of Quality Technology, 31*(1), 87–108.

Reynolds, M. R., Jr., & Stoumbos, Z. G. (2000). A general approach to modeling CUSUM charts for a proportion. *IIE Transactions, 32*, 515–535.

Ryan, A. G., Wells, L. J., & Woodall, W. H. (2011). Methods for monitoring multiple proportions when inspecting continuously. *Journal of Quality Technology, 43*(3), 237–248.

Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I., & Woodall, W. H. (2015). Another look at the EWMA control chart with estimated parameters. To appear in the Journal of Quality Technology.

Santiago, E., & Smith, J. (2013). Control charts based on the exponential distribution: Adapting runs rules for the *t* chart. *Quality Engineering, 25*(2), 85–96.

Schuh, A., Woodall, W. H., & Camelio, J. A. (2013). The effect of aggregating data when monitoring a Poisson process. *Journal of Quality Technology, 45*(3), 260–272.

Sego, L. H., Woodall, W. H., & Reynolds, M. R., Jr. (2008). A comparison of surveillance methods for small incidence rates. *Statistics in Medicine, 27*(8), 1225–1247.

Steiner, S. H., Cook, R. J., Farewell, V. T., & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics, 1*, 441–452.

Steiner, S. H., & MacKay, R. J. (2004). Effective monitoring of processes with parts per million defective – A hard problem! In H. J. Lenz, & P. -Th. Wilrich (Eds.), *Frontiers in statistical quality control 7*. Heidelberg, Germany: Springer.

Szarka, J. L., III., & Woodall, W. H. (2011). A review and perspective on surveillance of Bernoulli processes. *Quality and Reliability Engineering International, 27*, 735–752.

Szarka, J. L., III, & Woodall, W. H. (2012). On the equivalence of the Bernoulli and geometric CUSUM charts. *Journal of Quality Technology, 44*(1), 54–62.

Tikhomirova, M. I., & Christyakov, V. P. (2010). On properties of the likelihood function in the problem of estimation of the instants of changes of the success probability in nonhomogeneous Bernoulli trials. *Discrete Mathematics and Applications, 20*(2), 213–230.

Toubia-Stucky, G., Liao, H., & Twomey, J. (2012). A sequential Bayesian cumulative conformance count approach to deterioration in high yield processes. *Quality and Reliability Engineering International, 28*, 203–214.

Wallenstein, S., Naus, J., & Glaz, J. (1994). Power of the scan statistic in detecting a changed segment in a Bernoulli sequence. *Biometrika, 81*(3), 595–601.

Wheeler, D. J. (2011). Working with rare events – What happens when the average count gets very small? Quality Digest. Available at http://www.qualitydigest.com/inside/quality-insider-article/\working-rare-events.html.

Woodall, W. H., Fogel, S. L., & Steiner, S. H. (2015). The monitoring and impovement of surgical outcome quality. To appear in Journal of Quality Technology.

Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika, 70*(2), 455–464.

Wu, Z., Liu, Y., He, Z., & Khoo, M. (2010). A cumulative sum scheme for monitoring frequency and size of an event. *Quality and Reliability Engineering International, 26*, 541–554.

Xie, M., Goh, T. N., & Ranjan, P. (2002). Some effective control chart procedures for reliability monitoring. *Reliability Engineering & System Safety, 77*, 143–150.

Xie, Y. J., Tsui, K. -L., Xie, M., & Goh, T. N. (2010). Monitoring time-between-events for health management. In *Prognostics & System Health Management Conference*, Macau.

Zhang, C.W., Xie, M., Liu, J.Y., & Goh, T.N. (2007). A control chart for the gamma distribution as a model of time between events. *International Journal of Production Research, 45*(23), 5649–5666.

Zhang, M., Peng, Y., Schuh, A., Megahed, F. M., & Woodall, W. H. (2013). Geometric charts with estimated parameters. *Quality and Reliability Engineering International, 29*, 209–223.

Zhang, M., Megahed, F. M., & Woodall, W. H. (2014). Exponential CUSUM charts with estimated control limits. *Quality and Reliability Engineering International, 30*, 275–286.

Statistical Perspectives on "Big Data"

Fadel M. Megahed and L. Allison Jones-Farmer

Abstract As our information infrastructure evolves, our ability to store, extract, and analyze data is rapidly changing. *Big data* is a popular term that is used to describe the large, diverse, complex and/or longitudinal datasets generated from a variety of instruments, sensors and/or computer-based transactions. The term *big data* refers not only to the size or volume of data, but also to the variety of data and the velocity or speed of data accrual. As the volume, variety, and velocity of data increase, our existing analytical methodologies are stretched to new limits. These changes pose new opportunities for researchers in statistical methodology, including those interested in surveillance and statistical process control methods. Although it is well documented that harnessing *big data* to make better decisions can serve as a basis for innovative solutions in industry, healthcare, and science, these solutions can be found more easily with sound statistical methodologies. In this paper, we discuss several *big data* applications to highlight the opportunities and challenges for applied statisticians interested in surveillance and statistical process control. Our goal is to bring the research issues into better focus and encourage methodological developments for *big data* analysis in these areas.

Keywords Analytics • Control charts • Data mining • High-dimensional data • Image-monitoring • Surveillance • Text mining

1 Introduction

The volume of data produced from purchase transactions, social media sites, production sensors, healthcare information systems, cellphone GPS signals, etc. continues to expand at an estimated annual rate of 60–80 % (The Economist 2010a). To put these numbers into perspective, an 80 % annual rate—if compounded daily—

F.M. Megahed (✉)
Department of Industrial and Systems Engineering, Auburn University, Auburn, AL 36849, USA
e-mail: fmegahed@auburn.edu

L.A. Jones-Farmer
Department of Information Systems and Analytics, Miami University, Oxford, OH 45056, USA
e-mail: farmerl2@miamioh.edu

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_3

means that $\sim 90\,\%$ of that data existing in the world today has been created since 2011. The acquisition of massive datasets is expected to transform the approach to major problems such as: combating crime, preventing diseases, and predicting the occurrence of natural disasters. Additionally, businesses are using their collected data to spot business trends and customize their services according to customer profiles. For example, Wal-Mart, the leading U.S. discount retailer, processes more than 1 million customer transactions per hour, resulting in databases estimated to be in the magnitude of 2,500 terabytes (The Economist 2010b). Wal-Mart is not the only corporation that handles and stores such massive amounts of data. Others include social media corporations, gaming companies, airlines, insurance companies, healthcare providers, electric companies, and others. Indeed, most industries are experiencing data inflation and combating the growing need to make sense of *big data*.

The acquisition of data does not automatically transfer to new knowledge about the system under study. This notion is not new to the statistical community. Deming (2000, p. 106) said that "information, no matter how complete and speedy, is not knowledge. Knowledge has temporal spread. Knowledge comes from theory. Without theory, there is no way to use the information that comes to us on the instant." In the current age of *big data* there seem to be two competing camps: those who place problems and data into theoretical frameworks (e.g., statisticians, operations researchers, etc.), and those who use inductive analysis tools such as data mining, artificial intelligence, and knowledge discovery methods.

Although theoretical framing of data has proven successful, sometimes the theoretical frameworks are unrealistic for *real data*, and analyses based on the frameworks fail. Similarly, inductive, often ad hoc, methods have worked well, but often the focus on empirical results to the exclusion of domain expertise and statistical or mathematical theory has given wrong answers. The need to balance theory and empiricism is not new, but is growing in attention with the evolution of the *big data* attention. A recent survey in *The Economist* shows that 62 % of workers have indicated that the quality of their work is hampered because they cannot make sense of the data that they already have (The Economist 2011). With large and diverse data sets, we need to consider new and creative solutions to the problems that are posed. MacGregor (2013) states "as we continue to collect more and more information through more advanced sensors, our traditional reliance on fundamental or structured empirical models becomes less and less viable and hybrid approaches are needed."

Unfortunately, for many organizations in the private and public sector, processing and making sense of *big data* is still somewhat of a science project (with exceptions in chemometrics where multivariate statistics are used in the monitoring and control of processes governed by large amounts of sensor data). In 2010, the Library of Congress (LOC) signed an agreement with Twitter to archive all public tweets since 2006. The LOC receives and processes nearly 500,000,000 tweets per day, and as of December, 2012, had amassed 170 billion tweets, or 133.2 terabytes of compressed data. A single search of the archive can take up to 24 h. As scholars anxiously await access to the tweet archive, the LOC released a public statement stating that, "It is

clear that technology to allow for scholarship access to large data sets is not nearly as advanced as the technology for creating and distributing that data." (Library of Congress 2013)

In this article we give an overview of *big data* and how the researchers in statistics and statistical process control (SPC) can aid in making sense of *big data*. We begin in Sect. 2 by trying to answer the question "What is *Big Data*?", with the understanding that there is no clear-cut answer to this question. In Sect. 3 we discuss the challenges with data quality. We provide in Sect. 4 some examples on how SPC and statistics are being used in several application domains of *big data*, highlighting a few of the challenges. In Sect. 5, we discuss some differences between traditional and *big data* applications of statistical process control (SPC). We provide concluding remarks in Sect. 6.

2 What is Big Data?

Big data is a popular term that is used to describe the large, diverse, complex and/or longitudinal datasets generated from a variety of instruments, sensors and/or computer-based transactions. Although there is no clearly agreed upon definition of what constitutes *big data* most agree that the term refers to data that is so large it requires some form of distributed computing to process and analysis (SAS 2013; Manyika et al. 2011; Zikopoulos et al. 2013). To be able to gain knowledge from *big data*, it is imperative to understand both the scale and scope of *big data*. The challenges with processing and analyzing *big data* are not only limited to the size of the data. These challenges include the size, or *volume*, as well as the *variety* and *velocity* of the data (Zikopoulos et al. 2012). Known as the 3*V's*, the volume, variety, and/or velocity of the data are the three main characteristics that distinguish *big data* from the data we have had in the past.

2.1 Volume

There are two major factors that contribute to the increase in *data volume* (Carter 2011). First, the continued advancements in sensing and measurement technologies have made it possible to collect large amounts of data in manufacturing, healthcare, service, telecommunications, and military applications. Second, decreasing storage costs have made it economically feasible to store large amounts of data relatively inexpensively. With increasing technology, storage can take different forms from traditional databases to cloud storage (e.g., Amazon Web Services and Microsoft Azure). It is interesting to note that advancements in sensing/measurement technology and inexpensive storage are two of the driving forces behind advancements in statistical methodology, including SPC methodology. With specific regard to SPC, the 1920s were characterized with low levels of automation and high levels

of manual inspection. Therefore, control charts focused on small samples and univariate quality characteristics. With automation and increased computational capabilities, there has been increased emphasis on multivariate control charts, time-series methods and profile monitoring (Montgomery 2013; Noorossana et al. 2011). It should be noted that much of this research has originated in chemometrics (e.g., see Duchesne et al. 2012; MacGregor and Cinar 2012; Nomikos and MacGregor 1995).

2.2 Variety

With increased sensing technology, the explosion in social media and networking, and the willingness of companies to store everything, we have new challenges in terms of data variety. For example, manufacturers are no longer restricted by traditional dimensional measurement since they accumulate sensor data on their equipment, images for inspecting product aesthetics, radio-frequency identification (RFID) technology in supply chain management, and social network data to capture their customers' feedback. In theory, this variety of data should allow for a 360° view of manufacturing quality. In reality, even if we had the infrastructure to match the customer feedback to the exact product identification, and tracking through the supply chain, our methods to analyze and gain knowledge from this data are limited. Most statistical methods (including control charts) are classified according to the type of data with which they should be used. For example, few methods are available for monitoring multivariate processes with a mixture of categorical and continuous data (Jones-Farmer et al. 2014; Ning and Tsung 2010). Monitoring and analyzing mixed data types becomes more challenging when we include non-numeric, unstructured data.

Some reports estimate that nearly 80 % of an organization's data is not numeric (SAS 2013). Although there are many approaches to text analytics, our experience with text analytics suggests there remain many limitations. Most text analytics approaches convert unstructured text into numerical measures (e.g., word counts and/or measures of sentiment captured in a phrase). The numerical measures are then analyzed using traditional data analytic approaches (e.g., data mining). While a thorough discussion of text analytics methods is well beyond the scope of this article, it is important to note that most approaches have been developed in a specific context (e.g., dictionaries that assign sentiment to word combinations from literature or from interpreting political manifestos). The applicability of these methods to the cryptic and evolutionary nature of the language of social media is questionable. Thus, the analyses based on the numerical scores resulting from these methods may not always be meaningful. A major challenge of *big data* analysis is how to automatically process and translate such data into new knowledge.

2.3 *Velocity*

Zikopoulos et al. (2012) state that the conventional notion of data velocity is the rate at which the data arrives, is stored, and retrieved for processing. These authors note, however, that this definition of velocity should be expanded to include, not only the growth rate of data, but also the speed at which data are flowing. The rate of flow can be quite massive in certain domains. Considering microblog data, data consumers have the option to purchase near real-time access to the entire Twitter data or a randomly selected portion of the data over a fixed prospective time range. Recalling that Twitter processes around 500,000,000 tweets daily, the entire data set is appropriately referred to as the "fire hose." Social media is not the only domain with high velocity data. Recall the estimated 1,000,000 transactions per hour processed by Wal-Mart.

One of the main challenges with high velocity data is how fast any organization can transform the high velocity data into knowledge. Near real-time responsiveness to high velocity data may allow for competitive advantage in terms of managing stock-outs, product introductions, customer service, and public relations interventions. Similar advantages may be found with faster responsiveness to high velocity data in areas including public health, network security, safety, traffic, and other surveillance domains. Essentially, the opportunity cost clock starts once the data point has been generated. Zikopoulos et al. (2012) state "velocity is perhaps one of the most overlooked areas in the *Big Data* craze." A similar sentiment is echoed by *SAS* in a recent white paper that states, "reacting quickly enough to deal with velocity is a challenge to most organizations"(SAS 2013).

Another challenge related to data velocity is the fact that social media and transactional data can produce highly variable flow rates, with enormously large peaks, and near zero valleys. In the context of social media, the phrase "gone viral"is often used to describe a video, blog, or microblog that has been shared on a large scale very rapidly. With internet and social media data, transactions or activity on a particular site/topic can go from a manageable baseline velocity up to a level that is beyond the capacity of the infrastructure to handle in a matter of seconds. For example, in September, 2011, Target's website was flooded with visitors and transactions within hours of making a limited-edition line of clothing and home products by Missoni available to customers. The influx of transactions crashed the entire website, rendering it unavailable for the better part of a day.

In an attempt to manage unpredictable data velocity and volume, there has been an increasing market for computer Infrastructure as a Service (IaaS) providers, where customers pay as they go for scalable computing, storage and data access over the internet (for a detailed introduction on IaaS, see Mell and Grance 2011; Prodan and Ostermann 2009; U.S. General Services Administration 2013). While the emergence of IaaS providers has allowed companies a failsafe for some of the computing and storage needs that arise from the variable nature of *big data*, analyzing and monitoring data with extreme fluctuations in velocity remains a difficult problem. In addition to potential data overload problems, extreme variations

in data velocity makes it very difficult to develop a baseline to account for daily, seasonal, and event-triggered loads, and to detect performance changes and trends (SAS 2013).

3 Beyond the 3V's

The *volume*, *variety*, and *velocity* of data are three fundamental characteristics that distinguish *big data* from any other type of data (e.g., medium or small). Many have suggested that there are more *V's* that are important to the *big data* problem such as *veracity* and *value* (IEEE BigData 2013) . *Veracity* refers to the trustworthiness of the data, and value refers to the value that the data adds to creating knowledge about a topic or situation. While we agree that these are important data characteristics, we do not see these as key features that distinguish *big data* from regular data. It is important to evaluate the veracity and value of all data, both big and small.

Both *veracity* and *value* are related to the concept of data quality, an important research area in the Information Systems (IS) literature for more than 50 years. The research literature discussing the aspects and measures of data quality is extensive in the IS field, but seems to have reached a general agreement that the multiple aspects of data quality can be grouped into several broad categories (Wang and Strong 1996). Two of the categories relevant here are *contextual* and *intrinsic* dimensions of data quality.

Contextual aspects of data quality are context specific measures that are subjective in nature, including concepts like value-added, believability, and relevance. Many of the measures for these contextual dimensions are based on surveys of the end-users of the data. Batini et al. (2009) provide an overview of the previously used measures regarding contextual data quality measures. *Intrinsic* aspects of data quality are more concrete in nature, and include four main dimensions: *accuracy*, *timeliness*, *consistency*, and *completeness* (Batini et al. 2009; Huang et al. 1999; Parssian 2006; Scannapieco and Catarci 2002). The term *accuracy* implies data which are as free from error as possible. *Timeliness* implies that data are as up-to-date as possible. *Consistency* measures how closely the data's content and occurrences are structured in the same way in every instance. *Completeness* refers to data which are full and complete in content, with no missing data. Jones-Farmer et al. (2014) gave an overview of data quality and discuss opportunities for research in statistical methods for evaluating and monitoring the quality of data.

From our perspective, many of the contextual and intrinsic aspects of data quality are related to the *veracity* and *value* of the data. That said, *big data* presents new challenges in conceptualizing, evaluating, and monitoring data quality. An excellent example of the unique difficulties with *big data* quality, including value and veracity, occurs with social media data. On April 23, 2013, at 1:07 p.m., Eastern Time, a tweet from the Associated Press (AP) account stated "Breaking: Two Explosions in the White House and Barack Obama is injured" (Strauss et al. 2013). The fraudulent tweet, originated from the hacked AP Twitter account led to an immediate

drop in the Dow Jones industrial average. Although the Dow quickly recovered following a retraction from the AP and a press release from the White House, this example illustrates the immediate and dramatic effects of poor quality data. There are other examples of problems with veracity and value of *big data* in the social media realm. For example, political analysts covering the 2012 Mexican presidential election struggled with tweets due to the creation of a large number of fake Twitter accounts that polluted the political discussion, and introduced derogatory hash tags (Zikopoulos et al. 2013, pp. 14–15). It became very difficult to sift through the spam tweets from the tweets of voters, and even understand the impact of spam tweets on the voting population.

4 Applications of SPC to Big Data Problems

In this section, we discuss several examples on how SPC is currently being used or could potentially be used with *big data*, highlighting some differences between traditional and *big data* applications of SPC. As a disclaimer, not all of the data that might be used with these applications will fit some definitions of "*big data*"; however, they all fall into the general class of high volume, and/or, high velocity, and/or, high variety data. The types of data we discuss include (a) labeled data, (b) unlabeled data, (c) functional data, (d) graph data, and (e) data from multiple streams.

4.1 Labeled Data

Machine learning algorithms are among the most frequently used methods to analyze large and diverse data sets. Machine learning algorithms are typically classified as supervised or unsupervised learning algorithms. The focus of supervised algorithms is to classify data based on training set that includes known classifications. In the case of supervised learning, the classifications or dependent variables are considered *labeled*. Specifically, the training set consists of pairs $(\boldsymbol{x}, y)$ where $\boldsymbol{x}$ is a vector of independent variables (or features) and $\boldsymbol{y}$ is the classification value for $\boldsymbol{x}$ (the label). The objective of the analysis is to discover a function $y = f(\boldsymbol{x})$ that best predicts the value of y that is associated with unseen values of $\boldsymbol{x}$ (Rajaraman et al. 2012). It is important to note that y is not necessarily a real number, but might be a dichotomous or polytomous variable. In some cases, such as text analytics, y can represent a near-infinite set of classes.

Recently, machine learning algorithms have been proposed for use within SPC (see, e.g., Chinnam 2002; Cook and Chiu 1998; Deng et al. 2012; Hwang et al. 2007; Sun and Tsung 2003). Although all of these methods are not presented in a *big data* setting, we discuss them since they may be scalable to high volume, high velocity data. Several of these methods can be used with mixed data types and

data measured using different measurement scales. Usually, an artificial data set is generated to represent out-of-control data, which is used to train a classifier. This approach converts the monitoring problem to a supervised classification problem which classifies future observations as either in- or out-of-control. Supervised learning methods have also been used in Phase I analysis, in an attempt to identify an in-control baseline sample (e.g., Liu et al. 2008; Deng et al. 2012)

4.2 Unlabeled Data

Unsupervised learning algorithms are commonly used when data are unlabeled, often because it is expensive, difficult, or impossible to assign a label identifying the classification or outcome. The goal of unsupervised learning methods is to assign each point to one of a finite set of classifications. There are several approaches to unsupervised learning, the most common of which include cluster analysis and mixture modeling. Several authors have applied unsupervised learning methods to unlabeled data in Phase I. Sullivan (2002) introduced a clustering method to detect multiple outliers in a univariate continuous process. Thissen et al. (2005) used Gaussian mixture models to establish an in-control reference sample in a Phase I analysis. Jobe and Pokojovy (2009) introduced a computer intensive multi-step clustering method for retrospective outlier detection in multivariate processes. More recently, Zhang et al. (2010) introduced a univariate clustering-based method for finding an in-control reference sample from a long historical stream of univariate continuous data. There remain many opportunities to investigate the strengths and limitations of unsupervised learning methods for Phase I analysis.

We see opportunities for using machine learning in both Phase I and Phase II applications in SPC. Much work and process understanding is often required in the transition between a Phase I and Phase II analysis Woodall (2000). With *big data* applications, machine learning algorithms can be used to develop insights and understand the root-causes for underlying problems. For example, Wenke et al. (1999) included a classification step as a part of a general framework for intrusion detection in computer systems. Cruz and Wishart (2006) provided a review on using machine learning algorithms in cancer prediction and prognosis. In their findings, they stated that "machine learning is also helping to improve our basic understanding of cancer development and progression."

4.3 Functional Data

Functional data originated within the fields of chemometrics and climatology in the 1960s. The topic has received a great deal of attention from the statistics community ever since, as it covers a wide range of important statistical topics, including:

classification, inference, factor-based analysis, regression, resampling, time-series, and random processes. Additionally, it allows for infinitesimal calculations due to the continuous nature of the data (Ferraty and Romain 2011). The reader is referred to Ramsay and Silverman (2002, 2005) for an in-depth coverage of the topics involved in functional data analysis (FDA). Ferraty and Romain (2011) provide an excellent discussion on various benchmarking methodologies and the fundamental mathematical aspects that are related to the *big data* aspect of FDA.

Woodall et al. (2004) related the application of profile monitoring to FDA. Profile monitoring is a control charting method used when the quality of a process or product can be characterized by a functional relationship between a response variable and explanatory variable(s). In such applications features are often extracted from the monitored curves, such as regression coefficients, and then, used as input to standard control charting methods. This transformation represents an initial phase of preprocessing that has not been explicitly identified in the traditional SPC literature. Extensions of the existent profile monitoring techniques can be seen in image analysis and other related areas. We provide an overview on these applications in the paragraphs below.

We view the use of images for process monitoring as a very promising area of statistical research with a wide range of applications. In manufacturing settings, image-based monitoring adds the capability of monitoring a wide variety of quality characteristics, such as dimensional data, product geometry, surface defect patterns, and surface finish, in real-time. This is somewhat different from traditional SPC applications that focus on dimensional and discrete process data. Additionally, image monitoring plays an important role in many medical, military, and scientific applications. Megahed et al. (2011) provided a review on the use of control charts with 2D image data. In their review, they noted that the use of control charting with image data is not yet well developed (with the exception of applications in chemometrics).

A common theme in the literature discussing FDA is the lack in statistical and mathematical tools that enables these data-rich sources to achieve their full potential. For example, Ferraty and Romain (2011, pp. viii–ix) identified several challenges for the future in the context of FDA including the need for semiparametric methods, the nonlinear structure in high-dimensional spaces, and the increasing sophistication of the datasets. These challenges are important research areas to consider, especially since surveillance requires Phase I analyses where the baseline is established and the parameters are estimated. In addition to these challenges, there exist several control charting opportunities that are specific to 2D images and data from 3D scanners. For example, there is no discussion on the effect of estimation error on image-based control charts, and therefore it is not known what Phase I sample sizes are needed and whether they would be practical. Generally speaking, there is a need to evaluate the assumptions inherent in many of the papers involving control charting for image and point cloud data. It may be beneficial to create a large repository of functional data and share it with researchers to accelerate scientific discovery in this data-driven area. Such repositories exist for non-surveillance

applications involving functional data (see, e.g., http://archive.ics.uci.edu/ml/, http://www.ncbi.nlm.nih.gov/geo/, and http://aws.amazon.com/publicdatasets/).

4.4 Graph Data

In several *big data* applications, there is an interest to extract relationships between the different entities in the data. This relational knowledge is best described through a graph representation (Cook and Holder 2006). A graph allows the entities to be represented by nodes. Two nodes are connected by an edge if there is a relationship between them. The graph representation encompasses a large number of applications that include: (a) social networks (e.g., Facebook, LinkedIn); (b) cyber-security and computer networks (e.g., detecting network intrusion); (c) influence propagation (e.g., propagation of disease, or product adoption); (d) e-commerce (e.g., consumer ratings); and (e) other application domains Chakrabarti and Faloutsos (2012). Some examples of opportunities to apply SPC techniques to these types of graph data are given below.

McCulloh and Carley (2008) proposed an SPC-based approach to detect change in social networking based on the notion that the data collected on such networks is similar in structure to continuous variables that are typically monitored by control charts in manufacturing applications. In their approach, they calculated the average graph measures for density, closeness, and betweenness centrality for several consecutive time-periods of the social network. They estimated the in-control mean and variance for each of these three measures by taking a sample average and sample variance of the stabilized measures. Then, they used a CUSUM chart (with a reference value $k = 0.5$ and control limit, $h = 4$) to monitor each of these three measures. Their control charts were deployed to detect structural changes on two publically available social network datasets: *Tactical Officer Education Program e-mail Network* and *open source Al-Qaeda communications network*. Using the CUSUM chart, they were able to identify significant changes in both networks, with Al-Qaeda network structurally changing prior to the attacks of September 11th. In addition, the authors discussed in detail the limitations of their method and some future research extensions that should be examined. We agree with their recommendations and suggest further investigation in that area.

Wu et al. (2007) presented a simple approach using a Shewhart control chart to detect distributed denial of service (DDoS) attacks on a computer network. Underbrink et al. (2012) suggested using Shewhart control charts in the context of testing cyber-attack penetration to software applications. Hale and Rowe (2012) provided an introduction on statistical process control, Shewhart control chart and Shewhart chart with runs rules, and suggested its use to detect and eliminate problems in developing software for military applications. From this discussion, one can see that the use of control charting for intrusion detection in cyber-security applications has been introductory. This can be seen by the absence of the phrase "control chart" in the *Guide to Intrusion Detection and Prevention Systems*

developed by the U.S. National Institute of Standards and Technology (Scarfone and Mell 2012). We see the cyber-security (CS) problem to be far more complex than what is presented as the architecture of the computer network is continuously changing and the state of the network can be represented by data of different types (e.g., counts, continuous variables, and probabilities/ratios). At this time, we know of only a limited class of multivariate control charting methods that can be used to successfully monitor multivariate data of mixed data types. These methods are generally in the class of supervised machine learning techniques (e.g., neural network based control charts), and the properties of these methods are still being discussed in the academic literature. Thus, further work in this area warrants a detailed investigation of these methods.

4.5 Multistream Data

Increased availability of data and data tracking sources makes it possible to simultaneously monitor data from many, often disparate sources. This presents opportunities and challenges in terms of managing reasonable false alarm rates while maintaining the ability to quickly detect unusual events. A control chart for monitoring multiple streams collected at regular time intervals was introduced by Boyd (1950). This approach considered plotting the maximum and minimum subgroup mean from the group of streams on a control chart to detect location changes. Mortell and Runger (1995) proposed alternative control charts based on the pooled subgroup means for detecting overall mean changes and the range of the subgroup means to detect location changes in an individual stream. Liu et al. (2008) compared the method of Mortell and Runger (1995) to several approaches in the context of monitoring using multiple gauges in a truck assembly process. Liu et al. (2008) proposed methods based on F-tests comparing the between gauge to the within gauge variability and likelihood ratio tests to detect specific changes in only one stream. Meneces et al. (2008) considered the effect of correlation across the multiple streams on the use of individual Shewhart X-bar charts to monitor each stream. Using simulation, Meneces et al. (2008) suggested alternative control chart constants to maintain a desired false alarm probability when monitoring multiple correlated streams. Lanning et al. (2002) used an adaptive fractional-sampling approach to monitor a large number of independent streams when it is only possible to monitor a fraction of the streams. Recently, Jirasettapong and Rojanarowan (2011) provided a framework for selecting a control chart method for monitoring multiple streams based on the stream characteristics including the degree of correlation among the streams, the number of streams, and the shift size to be detected.

While some methods to monitor multiple streams have been addressed in the SPC literature, most focus on a manufacturing context. The need to monitor multiple, often correlated, data streams occurs outside of manufacturing in several *big data* contexts. These include surveillance in public health, bioterrorism, social

media, and others. Many authors have considered the statistical challenges present in monitoring multiple data streams, particularly in the biosurveillance context. Shmueli and Burkom (2010) noted that the context of biosurveillance, including the data types, the mechanisms underlying the time series, and the applications all deviate from the traditional industrial setting for which control chart methods were developed. Rolka et al. (2007) gave an overview of the methodological issues in public health and bioterrorism surveillance in the context of multiple stream data. In addition to traditional methods including the multivariate cumulative sum and exponentially weighted moving average approaches, Rolka et al. (2007) considered the spatiotemporal aspect of biosurveillance data and the use of spatial scan statistics as well as Bayesian network models to combine multiple data streams. Tsui et al. (2008) also considered public health and disease surveillance, giving an overview of the most popular methods, and discussing research opportunities in spatiotemporal surveillance. Fraker et al. (2008) and Megahed et al. (2012) discussed the advantages and disadvantages of several performance metrics applied to health surveillance, recommending metrics based on time-to-signal properties.

The advances in the use of the Web as an information sharing tool, the simplicity with which publicly available data can be merged and filtered, and the willingness of individuals to openly broadcast through social media is making the already challenging area of multiple stream surveillance even richer and more widely communicated. For example, Brownstein et al. (2009) listed a sample of Web-based digital resources for disease detection and discusses the use of search-term surveillance. They gave an example using Google Insights to investigate the search volume for several terms related to a salmonella outbreak attributed to contaminated peanut butter in January, 2009. Chunara et al. (2012) illustrated the effectiveness of the use of informal data sources such as microblogs volume and query search volumes to gain early insight into the 2010 cholera outbreak in Haiti. In both the case of the salmonella outbreak and cholera in Haiti, researchers were able to look retrospectively at a known event, asking the question "could we have detected this event using alternate sources?" In addition to the statistical challenges with monitoring multiple data streams, the real challenge may be non-statistical, and relate to knowing what to listen and look for in these emerging data sources. The reader is referred to (Hay et al. 2013) for a discussion on the opportunities of *big data* in global infectious disease surveillance.

5 Some Challenges in Big Data SPC

The application of SPC methods to *big data* is similar in many ways to the application of SPC methods to regular data. However, many of the challenges inherent to properly studying and framing a problem can be more difficult in the presence of massive amounts of data. There exist several frameworks for solving

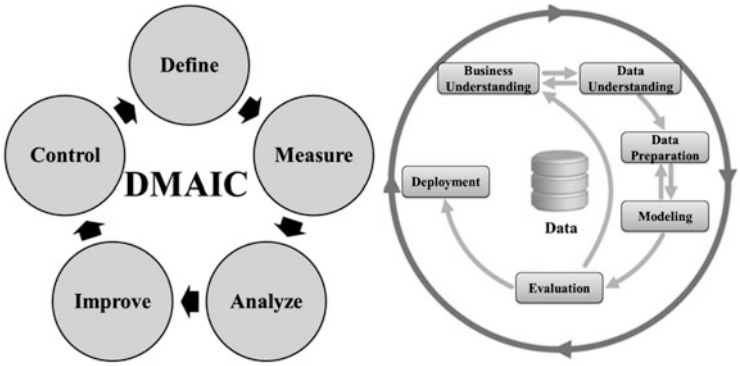


Fig. 1 The DMAIC and CRISP-DM cycles

problems in the Total Quality Management (TQM), SPC, or Six Sigma area. For example, there is the Plan Do Check Act (PDCA) cycle, or the Define, Measure, Analyze, Improve, Control (DMAIC) cycle. In the data mining and knowledge discovery areas, common problem solving frameworks include the Cross Industry Standard Process for Data Mining (CRISP-DM), the knowledge discovery in data mining (KDD) (Azevedo and Santos 2008). Figure 1 shows the DMAIC cycle and the CRISP-DM cycle side-by-side.

It is clear from Fig. 1 that there is some overlap between the traditional SPC framework of DMAIC and that of the CRISP-DM process. However, in the CRISP-DM cycle, there is specific emphasis on business understanding, data understanding, and data preparation. These stages are implicit within the DMAIC cycle, in the *define* and *measure* stages, but the importance of these stages is magnified when working with large and diverse data sets. For example, business and process understanding is emphasized heavily within Phase I of process improvement. In this Phase, the goal is to understand the process and process variability, define the in-control state of the process, and establish a baseline for estimation of the process model and parameters. Completing Phase I is critical to successful Phase II monitoring. However, Phase I cannot be successfully completed until there is clear business understanding of the process and the goal of the analysis. Additionally, data understanding is critical, and in a *big data* scenario, often with thousands or more variables, identifying the critical key outcome measures can be challenging. Finally, the importance of data quality cannot be overemphasized. Jones-Farmer et al. (2014) gave a review of the data quality literature and how this relates to SPC.

A key element in the CRISP-DM cycle is the importance of model deployment. Model deployment may be as simple as presenting the analysis in a PowerPoint or a white paper, or may be as complex as building an adaptive, online model into the information system. In any case, it is important to note that building the model is not the end-game. The actual use of the analysis in practice is the goal. Thus, some consideration needs to be given to the actual implementation of the statistical surveillance applications. This brings us to another important challenge, that of the

complexity of many *big data* applications. SPC applications have a tradition of *back of the napkin* methods. The custom within SPC practice is the use of simple methods that are easy to explain like the Shewhart control chart. These are often the best methods to use to gain credibility because they are easy to understand and easy to explain to a non-statistical audience. However, *big data* often does not lend itself to easy-to-compute or easy-to-explain methods. While a control chart based on a neural net may work well, it may be so difficult to understand and explain that it may be abandoned for inferior, yet simpler methods. Thus, it is important to consider the dissemination and deployment of advanced analytical methods in order for them to be effectively used in practice.

Aside from the general framework of problem solving, there are other important challenges with applying SPC to *big data* problems. For example, the data processing speed is a major area of focus in *big data* applications, especially with 100 % sampling. Therefore, methods compete based on the overall running time. The focus on data processing speed can also reflect on how the surveillance problem is approached. This effect can be seen in the monitoring of massive amounts of labeled data, where historical data is not stored and the analysis is done on the fly (Rajaraman et al. 2012). Accordingly, the analysis is often carried out using a representative set of points from the data, and a selective set of summary statistics that are useful for moment calculations. See Zhang et al. (2010); Bradley et al. (1998) and Guha et al. (1998) for three highly cited examples in this area. It is important for researchers in statistical surveillance to consider processing speed when developing and refining methodologies.

Another challenge in monitoring high dimensional data sets is the fact that not all of the monitored variables are likely to shift at the same time; thus, some method is necessary to identify the process variables that have changed. In high dimensional data sets, the decomposition methods used with multivariate control charts can become very computationally expensive. Several authors have considered variable selection methods combined with control charts to quickly detect process changes in a variety of practical scenarios including fault detection, multistage processes, and profile monitoring. Zou and Qiu (2009) considered the least absolute shrinkage and selection operator (LASSO) test statistic of Tibshirani (1996) combined with a multivariate exponentially weighted moving average (EWMA) control chart to quickly identify process changes and to identify the shifted mean components. Zou et al. (2012) considered the use of the LASSO statistic applied to profile monitoring. Zou et al. (2011) developed a LASSO based diagnostic framework for statistical process control that applies to high dimensional data sets. Capizzi and Masarotto (2011) developed a multivariate (EWMA) control chart using the Least Angle Regression (LAR) algorithm to detect changes in both the mean and variance of a high dimensional process. All of these methods based on variable selection techniques are based on the idea of monitoring subsets of potentially faulty variables. Capizzi and Masarotto (2011) listed several important research areas in this domain including the need to adapt these methods for nonparametric process monitoring, better monitoring of process dispersion, and adapting the methods for variable sampling interval (VSI) methods to improve detection performance. Some

variable reduction methods are needed to better identify shifts. We believe that further work in the areas combining variable selection methods and surveillance are important for quickly and efficiently diagnosing changes in high-dimensional data.

Another important challenge when using SPC methods with *big data* applications is that, traditionally, SPC methods were developed for numeric data. While there are some attributes control charts, these tend to be a distant choice to using methods designed for quantitative variables. However, one of the great challenges of *big data* is the ability to process and analyze unstructured data. Most of *big data* applications are concerned with non-numeric data obtained from several databases. This introduces several challenges. First, data quality becomes a focal point. Pipino et al. (2002) defined data quality to have several dimensions that include: completeness, consistency, free-of-error, security, timeliness, and understandability. These metrics are essential to evaluate prior to performing any analysis since the reliability of any produced model is associated with the quality of the input data. Moreover, this evaluation is not straightforward since the data is often non-numeric, unstructured, and imputed by a large number of users. Second, the modeling of the data is often based on disciplines that are not studied by statisticians and quality engineers. With text data, the models draw from linguistic sciences, psychology, and computer science, and often integrate data from different languages. A third challenge is that the arrival rate of the data fluctuates based on factors that are often not understood prior to analyzing the data. This phenomenon is referred to as trending/viral for online content.

6 Concluding Remarks

In this paper, we provided an overview of *big data* and the role of statisticians/quality engineers in understanding and advancing *big data*. It is clear that *big data* analytics is an evolving field with numerous applications, some of which can present solutions to global challenges pertaining to public-health and science. SPC and statistics are currently being deployed in some of these applications; however, we encourage further research in these areas. In particular, work is needed to examine how we can best preprocess large datasets and evaluate this preprocessing phase. This includes a better understanding of how to evaluate the quality of the data, select important features from the dataset, and perform Gage R&R. Additionally, establishing the baseline is much more complex than standard SPC applications since it requires domain knowledge, the data is heavily autocorrelated and is very complex. The development of visualization techniques may play a role in better understanding the problem. In Phase 1, we believe that data mining methods such as classification and clustering approaches will become more commonly used to gain process understanding. Additionally, we expect that the monitoring of functional, graph, and multiple stream data to become increasingly important.

References

Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: A parallel overview. In *Paper Presented at the IADIS European Conference on Data Mining*. Amsterdam, The Netherlands.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys, 41*(3), 16–16.52. doi:10.1145/1541880.1541883.

Boyd, D. F. (1950). Applying the group chart for *X* and *R*. *Industrial Quality Control, 7* (3), 22–25.

Bradley, P. S. , Fayyad, U., & Reina, C. (1998). Scaling clustering algorithms to large databases. In *Proceedings of the 4th International Conference on Knowledge Discovery & Data Mining Knowledge Discovery and Data Mining* (pp. 9–15). http://www.aaai.org/Papers/KDD/1998/KDD98-002.pdf.

Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection — harnessing the web for public health surveillance. *New England Journal of Medicine, 360* (21), 2153–2157. doi:10.1056/NEJMp0900702.

Capizzi, G., & Masarotto, G. (2011). A least angle regression control chart for multidimensional data. *Technometrics, 53*(3), 285–296. doi:10.1198/Tech.2011.10027.

Carter, P. (2011). *Big data analytics: Future architectures, skills and roadmaps for the CIO.* International Data Corporation (IDC).http://www.sas.com/resources/asset/BigDataAnalytics-FutureArchitectures-Skills-RoadmapsfortheCIO.pdf.

Chakrabarti, D., & Faloutsos, C. (2012). Graph mining: laws, tools, and case studies. *Synthesis Lectures on Data Mining and Knowledge Discovery, 3*(3), 1–207. doi:10.2200/S00449ED1V01Y201209DMK006.

Chinnam, R. B. (2002). Support vector machines for recognizing shifts in correlated and other manufacturing processes. *International Journal of Production Research, 40* (17), 4449–4466.

Chunara, R., Andrews, J. R., & Brownstein, J. S. (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene, 86* (1), 39–45. doi:10.4269/ajtmh.2012.11-0597.

Cook, D. F., & Chiu, C. C. (1998) Using radial basis function neural networks to recognize shifts in correlated manufacturing process parameters. *IIE Transactions, 30*(3), 227–234.

Cook, D. J., & Holder, L. B. (2006). *Mining graph data.* Hoboken, NJ: Wiley.

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Inform, 2*, 59–77. http://www.ncbi.nlm.nih.gov/pubmed/19458758.

Deming, W. E. (2000). *The new economics: for industry, government, education* (2nd ed.). Cambridge: The MIT Press.

Deng, H., Runger, G. C., & Tuv, E. (2012). Systems monitoring with real time contrasts. *Journal of Quality Technology, 44*(1), 9–27.

Duchesne, C., Liu, J. J., & MacGregor, J. F. (2012). Multivariate image analysis in the process industries: A review. *Chemometrics and Intelligent Laboratory Systems, 117*, 116–128. http://dx.doi.org/10.1016/j.chemolab.2012.04.003.

Ferraty, F., & Romain, Y. (2011). *The Oxford handbook of functional data analysis.* Oxford Handbooks. Oxford: Oxford University Press.

Fraker, S. E., Woodall, W. H., & Mousavi, S. (2008). Performance metrics for surveillance schemes. *Quality Engineering, 20*(4), 451–464. doi:10.1080/08982110701810444.

Guha, S., Rastogi, R., & Shim, K. (1998). CURE: an efficient clustering algorithm for large databases. In *Paper presented at the Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, Seattle, Washington, USA.

Hale, C., & Rowe, M. (2012). Do not get out of control: Achieving real-time quality and performance. *CrossTalk: The Journal of Defense Software Engineering, 25*(1), 4–8. http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA554677.

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine, 10* (4), e1001413.

Huang, K, -T., Lee, Y. W., & Wang, R. Y. (1999). *Quality information and knowledge.* Upper Saddle River: Prentice Hall.

Hwang, W., Runger, G., & Tuv, E. (2007). Multivariate statistical process control with artificial contrasts. *IIE Transactions, 39*(6), 659–669.

IEEE BigData (2013). http://www.ischool.drexel.edu/bigdata/bigdata2013/ (Accessed 05/19/2013).

Jirasettapong, P., & Rojanarowan, N. (2011). A guideline to select control charts for multiple stream processes control. *Engineering Journal, 15*(3), 1–14. doi:10.4186/ej.2011.15.3.1.

Jobe, J. M., & Pokojovy, M. (2009). A multistep, cluster-based multivariate chart for retrospective monitoring of individuals. *Journal of Quality Technology, 41*(4), 323–339.

Jones-Farmer, L. A., Ezell, J. D., & Hazen, B. T. (2014). Applying control chart methods to enhance data quality. *Technometrics, 56*(1), 29–41.

Lanning, J. W., Montgomery, D. C., & Runger, G. C. (2002). Monitoring a multiple stream filling operation using fractional samples. *Quality Engineering, 15*(2), 183–195. doi:10.1081/QEN-120015851.

Library of Congress (2013). Update on the twitter archive at the Library of Congress. http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf (Accessed 05/19/2013).

Liu, X., MacKay, R. J., & Steiner, S. H. (2008). Monitoring multiple stream processes. *Quality Engineering, 20*(3), 296–308. doi:10.1080/08982110802035404.

MacGregor, J. F. (2013). Some perspectives on the impact of big data on process systems engineering. Big data: The next frontier for innovation, competition, and productivity. In *2013 AIChE Annual Meeting.* http://www3.aiche.org/proceedings/Abstract.aspx?PaperID=342936.

MacGregor, J., & Cinar, A. (2012). Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods. *Computers & Chemical Engineering, 47*, 111–120. http://dx.doi.org/10.1016/j.compchemeng.2012.06.017.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute.* http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (Accessed 5/8/2013).

McCulloh, I. A, & Carley, K. M. (2008). *Social network change cetection.* Center for the Computational Analysis of Social and Organizational Systems Technical Report. Carnegie Mellon University, Pittsburg, PA. http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA488427.

Megahed, F. M., Fraker, S. E., & Woodall, W. H. (2012). A note on two performance metrics for public-health surveillance schemes. *Journal of Applied Probability and Statistics, 7*(1), 35–41.

Megahed, F. M., Woodall, W. H., & Camelio, J. A. (2011). A review and perspective on control charting with image data. *Journal of Quality Technology, 43*(2), 83–98.

Mell P, Grance T. (2011). *The NIST definition of cloud computing.* National Institute of Standards and Technology. http://docs.lib.noaa.gov/noaa_documents/NOAA_related_docs/NIST/special_publication/sp_800-145.pdf.

Meneces, N. S., Olivera, S. A., Saccone, C. D., & Tessore, J. (2008). Statistical control of multiple-stream processes: a Shewhart control chart for each stream. *Quality Engineering, 20*(2), 185–194. doi:10.1080/08982110701241608.

Montgomery, D. C. (2013). *Introduction to statistical quality control* 7th Ed. Wiley, Hoboken, NJ.

Mortell, R. R., & Runger, G. C. (1995). Statistical process control of multiple stream processes. *Journal of Quality Technology, 27*(1), 1–12.

Ning, X., & Tsung, F. (2010). Monitoring a process with mixed-type and high-dimensional data. In *2010 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1430–1432), 7–10 Dec 2010. doi:10.1109/IEEM.2010.5674333.

Nomikos, P., & MacGregor, J. F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics, 37*(1), 41–59. doi:10.1080/00401706.1995.10485888.http://www.tandfonline.com/doi/abs/10.1080/00401706.1995.10485888.

Noorossana, R., Saghaei, A., & Amiri, A. (2011). *Statistical analysis of profile monitoring. Wiley series in probability and statistics*. Hoboken: Wiley.

Parssian, A. (2006). Managerial decision support with knowledge of accuracy and completeness of the relational aggregate functions. *Decision Support Systems, 42*(3), 1494–1502. doi:10.1016/j.dss.2005.12.005.

Pipino, L. L, Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218. doi:10.1145/505248.506010.

Prodan, R., & Ostermann, S. (2009). A survey and taxonomy of infrastructure as a service and web hosting cloud providers. In *10th IEEE/ACM International Conference on Grid Computing* (pp. 17–25), 13–15 Oct 2009. doi:10.1109/GRID.2009.5353074.

Rajaraman, A., Leskovec, J., & Ullman, J. D. (2012). *Mining of massive datasets.* ISBN:978-1-107-01535-7.

Ramsay, J. O., & Silverman BW. (2002). *Applied functional data analysis: methods and case studies.* Springer Series in Statistics. New York: Springer.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* 2nd ed. Springer Series in Statistics. New York: Springer.

Rolka, H., Burkom, H., Cooper, G. F., Kulldorff, M., Madigan, D., & Wong, W. K. (2007). Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: Research needs. *Statistics in Medicine, 26*(8), 1834–1856. doi:10.1002/Sim.2793

SAS (2013). *Big data - What is it?* http://www.sas.com/big-data/ (Accessed 5/8/2013).

Scannapieco, M., & Catarci, T. (2002). Data quality under a computer science perspective. *Archivi & Computer, 2*, 1–15.

Scarfone, K., & Mell, P. (2012). Guide to intrusion detection and prevention systems (IDPS) (Draft): *Recommendations of the National Institute of Standards and Technology.* ttp://csrc.nist.gov/publications/drafts/800-94-rev1/draft_sp800-94-rev1.pdf.

Shmueli, G., & Burkom, H. (2010). Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics, 52*(1), 39–51. doi:10.1198/Tech.2010.06134.

Strauss, G., Shell, A., Yu, R., & Acohido, B. (2013). Hoax and ensuing crash on Wall Street show the new dangers of our light-speed media world. http://www.usatoday.com/story/news/nation/2013/04/23/hack-attack-on-associated-press-shows-vulnerable-media/2106985/ (Accessed 05/16/2013).

Sullivan, J. H. (2002). Detection of multiple change points from clustering individual observations. *Journal of Quality Technology, 34*(4), 371–383.

Sun, R., & Tsung, F. (2003). A kernel-distance-based multivariate control chart using support vector methods. *International Journal of Production Research, 41*(13), 2975–2989.

The Economist (2010a). All too much. http://www.economist.com/node/15557421 (Accessed 5/8/2013).

The Economist (2010b). Data, data everywhere. http://www.economist.com/node/15557443 (Accessed 5/8/2013).

The Economist (2011). Schumpeter: Too much buzz. http://www.economist.com/node/21542154 (Accessed 5/8/2013).

Thissen, U., Swierenga, H., de Weijer, A. P., Wehrens, R., Melssen, W. J., & Buydens, L. M. C. (2005). Multivariate statistical process control using mixture modelling. *Journal of Chemometrics, 19*(1), 23–31. doi:10.1002/Cem.903.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B, 58*(1), 267–288.

Tsui, K. -L., Wenchi, C., Gierlich, P., Goldsman, D., Xuyuan, L., & Maschek, T. (2008). A review of healthcare, public health, and syndromic surveillance. *Quality Engineering, 20*(4), 435–450. doi:10.1080/08982110802334138.

Underbrink, A., Potter, A., & Jaenisch, H., & Reifer, D. J. (2012). Application stress testing Achieving cyber security by testing cyber attacks. In: *2012 IEEE Conference on Technologies for Homeland Security (HST)* (pp. 556–561), 13–15 Nov. 2012. doi:10.1109/THS.2012.6459909.

U.S. General Services Administration (2013). *Infrastructure as a Service (IaaS).* http://www.gsa.gov/portal/content/112063 (Accessed 5/8/2013).

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.

Wenke, L., Stolfo, S. J., & Mok, K. W. (1999). A data mining framework for building intrusion detection models. In *Proceedings of the IEEE Symposium on Security and Privacy 1999* (pp. 120–132). doi:10.1109/SECPRI.1999.766909.

Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology, 32*(4), 341–350.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology, 36*(3), 309–320.

Wu, Q., Zhang, H., & Pu, J. (2007). Mitigating distributed denial-of-service attacks using network connection control charts. In *Proceedings of the 2nd International Conference on Scalable Information Systems*, Suzhou, China.

Zhang, H., Albin, S. L., Wagner, S. R., Nolet, D. A., & Gupta, S. (2010). Determining statistical process control baseline periods in long historical data streams. *Journal of Quality Technology, 42*(1), 21–35.

Zikopoulos, P., deRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D., & Giles, J. (2013). *Harness the Power of Big Data: The IBM Big Data Platform.* ISBN:978-0-07180818-7.

Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T., & Lapis, G. (2012). *Understanding big data: Analytics for enterprise class hadoop and streaming data.* New York: McGraw-Hill.

Zou, C., Ning, X., & Tsung, F. (2012). LASSO-based multivariate linear profile monitoring. *Annals of Operations Research, 192*(1), 3–19. doi:10.1007/s10479-010-0797-8.

Zou, C. L., Jiang,W., & Tsung, F. (2011). A LASSO-based diagnostic framework for multivariate statistical process control. *Technometrics, 53*(3), 297–309. doi:10.1198/Tech.2011.10034.

Zou, C. L., & Qiu, P. H. (2009). Multivariate statistical process control using LASSO. *Journal of the American Statistical Association, 104*(488), 1586–1596. doi:10.1198/jasa.2009.tm08128.

Statistical Control of Multiple-Stream Processes: A Literature Review

Eugenio K. Epprecht

Abstract This paper presents a survey of the research on techniques for the statistical control of industrial multiple-stream processes—processes in which the same type of item is manufactured in several streams of output in parallel, or still continuous processes in which several measures are taken at a cross section of the product. The literature on this topic is scarce, with few advances since 1950, and experiencing a resurgence from the mid-1990s. Essential differences in the underlying models of works before and after 1995 are stressed, and issues for further research are pointed out.

Keywords Group control charts • Parallel monitoring • Variance components

1 Introduction

A multiple stream process (MSP) is a process that generates several streams of output. From the statistical process control standpoint, the quality variable and its specifications are the same in all streams. A classical example is a filling process such as the ones found in beverage, cosmetics, pharmaceutical and chemical industries, where a filler machine may have many heads. Another one would be a mould with several cavities. Other processes may still produce only one stream of output but the quality variable is measured at several points at the same time. Consider, for instance, the fabrication of paper, sheets of steel, or the production of rubber hoses by extrusion, where at every sampling time, the thickness of the outcoming material is measured in different locations of its cross-section. For the purposes of modelling and monitoring, these can also be seen as multiple-stream processes.

Although multiple-stream processes are found very frequently in industry, the literature on schemes for the statistical control of such kind of processes is far from abundant. This paper presents a survey of the research on this topic. The focus

E.K. Epprecht (✉)
PUC-Rio, Department of Industrial Engineering, R. Marquês de S. Vicente 225, 22451-900, Rio de Janeiro, Brazil
e-mail: eke@puc-rio.br

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_4

is on industrial multiple-stream processes, but the last section (which indicates a number of issues for further research) will briefly comment on the similarities (and differences) between the problem of monitoring such processes and the problem of monitoring multiple streams of data in other contexts of application. There is a potential for the development of methods inspired on the techniques for industrial MSPs, even when these techniques cannot be directly applied due to the greater complexity of such other application contexts.

2 The Beginnings: The Group Control Chart and Nelson's Run Scheme

The first specific techniques for the statistical control of MSPs are the *group control charts* (GCCs), described by Boyd (1950) and also in Burr (1976), Pyzdek (1992, Chapter 21) and Montgomery (2012, Section 10.3.2). Boyd (1950) reported that "the group chart for $\bar{X}$ and R was developed by the British during World War II and was described in 'A First Guide to Quality Control for Engineers', a British Ministry of Supply publication compiled by Dr. E. H. Sealy and issued in 1943". Clearly the chief motivation for these charts was to avoid the proliferation of control charts that would arise if every stream were controlled with a separate pair of charts (one for location and other for spread). Assuming the in-control distribution of the quality variable to be the same in all streams (an assumption which is sometimes too restrictive), the control limits should be the same for every stream. So, the basic idea is to build only one chart (or a pair of charts) with the information from all streams. Specifically: at each sampling time t, every stream i is sampled and the corresponding $\bar{x}_i$ and R_i are calculated; the largest and the smallest $\bar{x}$ are plotted in the $\bar{x}$ GCC, and the largest R is plotted in the R GCC. If these points lie within the control limits, the other points (not plotted) would necessarily be within the limits too. Of course an S chart could be used instead of the R chart, but in the 50s the R chart was the usual one.

In the case the means of the streams differ but their variabilities are similar, the group chart can still be used by subtracting the mean of each stream from the values observed on that stream.

The question naturally arises of the overall false-alarm probability of a group chart. Indeed, with a great number of streams, an adjustment of the control limits should be made to ensure that this probability does not exceed the acceptable level. Two possibilities are the Bonferroni (Johnson and Wichern 2007, p. 232) and the Dunn-Sidak (Dunn 1958; Sidak 1967) corrections. Curiously, we found no reference to the need of such adjustments in any of the previously mentioned references. Colbeck (1999) and Grimshaw et al. (1999) are the first works, to my knowledge, that explicitly point out the inappropriateness of using 3-sigma limits with the GCC. The latter prescribed adjusting the limits with a formula that corresponds to the Dunn-Sidak correction and gave tables of constants for computing the limits.

With MSPs, special causes can be of two types: causes that affect all streams and causes that affect only one or some of the streams. For instance, in a filling process, a problem with a common pump may affect all spindles, while a clog in one spindle will affect only that spindle (if it is due to impurities in the liquid, other spindles will gradually become affected too). Nelson (1986) proposed a runs test (to be used with the group charts, in addition to the "one point beyond the 3-sigma limits" criterion) that would be sensitive to causes affecting only one stream. Namely, if the same stream gives an extreme reading (the highest or the lowest $\bar{x}$, or the highest R among all streams) r times in a row, this is a signal that the mean of that stream has shifted (or that its dispersion has increased). The critical value of r will depend on the number of streams and on the in-control ARL desired. Quoting him, "if the outputs of all the spindles have the same (not necessarily normal) distribution then the mean recurrence time or average run length (ARL) for a run of length r is given by $(k^r - 1)/(k-1)$, where k = number of spindles or, more generally, the number of streams". Using the above expression, in-control ARLs for different values of r and k can be obtained and tabulated.

Mortell and Runger (1995) evaluated a two-sided version of this runs scheme, in which a signal is the event that the mean of a stream is either a maximum or a minimum in r consecutive samples.

Of course, the control limits of the GCC already provide sensitivity against causes that affect only one stream; but the runs test increases this sensitivity. Evidently, with both criteria for an alarm, the overall false-alarm risk will also increase.

The runs test has two drawbacks, which have been pointed out by Mortell and Runger (1995) and reported by Montgomery (2012). First, if two or more streams shift, they are likely to alternate the extreme reading, so it may take quite some time to observe a run of r consecutive observations of a same stream. In addition, due to the discreteness of r, for some numbers of streams in the process there is no r value that corresponds to a good tradeoff between false-alarm risk and power: with any value of r, either the false-alarm risk will be too high or the sensitivity to shifts will be too low.

Wise and Fair (1998) recommended using the GCC only with Nelson's runs test, and no control limit. The reason, while not explicitly stated (their book is a kind of manual or cookbook; performance evaluation or theoretical arguments are not provided), may either be the fact that using both the run test and the control limits for detection would increase the false-alarm risk or, more probably, be the already high false-alarm risk of a group chart with 3-sigma limits. I cannot concur with their recommendation, however, at least because of the two limitations just mentioned of the runs test. It would be preferable to keep the control limits and adjust them for the acceptable false-alarm risk. They still proposed extending the use of group charts to the case of individuals and moving range statistics (considering that when the number of streams is large, samples with more than one observation per stream may become impractical or even infeasible). They also proposed using group charts for controlling multiple (different) quality characteristics, which can be achieved by standardizing the different variables. In my opinion, however, this proposal is

questionable. Their implicit motivation is that sometimes the number of different quality characteristics to consider is huge; a counterargument is that, nowadays, statistical process control (SPC) does not require plotting different charts on paper, so different processes and characteristics can be controlled individually. (This is supposing, of course, independent characteristics, otherwise the issue would belong to the field of multivariate SPC).

Wise and Fair's (1998) book deserves mentioning, nevertheless, for completeness of this review and for being motivated by the practical experience of the authors in the application of the techniques at Boeing.

Not only the runs scheme has its limitations, but also the "one point beyond the control limits" criterion provides the group chart with limited sensitivity, for at least two reasons: the widening of control limits required to control the false-alarm rate, and the fact that, with many streams, often it is infeasible to take samples of more than one or two measurements per stream. Schemes that are more sensitive both to causes that affect one stream and to causes that affect all streams will be presented in the next section.

From the standpoint of process analysis and improvement, Ott and Snee (1973) presented a detailed off-line analysis of a multiple-head machine, with the purpose of comparing "averages and variabilities of individual heads and changes with time". They make use of semi-residuals, complete residuals, plots of raw data arranged in different ways and ANOVA to separate the time effects and head effects. This work is remarkable as the only one, to the best of my knowledge, about the analysis of MSPs; all other works are about monitoring them.

3 Two Sources of Variation

The GCC will work well if the values of the quality variable in the different streams are independent and identically distributed, that is, if there is no cross-correlation between streams. However, such an assumption is often unrealistic. In many real multiple-stream processes, the value of the observed quality variable is typically better described as the sum of two components: a common component (let's refer to it as "mean level"), exhibiting variation that affects all streams in the same way, and the individual component of each stream, which corresponds to the difference between the stream observation and the common mean level. In formal notation: supposing that at time t a subgroup of measures is taken at each stream, the value of j-th observation of the quality variable in stream i in time t is given by:

$$x_{tij} = b_t + e_{tij} \quad , \quad t = 0, 1, 2, \ldots \quad , \quad i = 1, 2, \ldots, s \quad , \quad j = 1, 2, \ldots n \qquad (1)$$

where b_t is the value of the "mean level" in time t and e_{tij} is the value of the individual component of the j-th observation of the i-th stream at time t. The values of b_t may be i.i.d. or present some dynamics over time, depending on each particular

process, and e_{tij} is assumed to be i.i.d. $\sim \mathcal{N}(0, \sigma^2)$ over t, i and j, and independent from b.

As a result, the variance of $\bar{X}_i$ (the sample average of the observations in any stream i) is

$$V(\bar{X}_i) = \sigma_b^2 + \sigma^2/n \tag{2}$$

where σ_b^2 is the variance of the mean level.

The presence of the mean level component can be identified in a given MSP by examining the correlations between different streams (that is, between the $\bar{X}_{ti}$'s: the averages (over j) of the n values observed in each stream i in a same time t), because the mean level introduces cross correlations between the streams. Indeed, the correlation between any two streams is

$$\sigma_b^2/(\sigma_b^2 + \sigma^2/n)\,.$$

Since the control limits of the $\bar{X}$ GCC should be based on the total variance of $\bar{X}_i$ (and, as stated before, should still be "widened" to avoid inflating the total false-alarm rate), the presence of the mean level component leads to reduced sensitivity of Boyd's GCC to shifts in the individual component of a stream if the variance σ_b^2 of the mean level is large with respect to the variance σ^2 of the individual stream components. Moreover, the GCC is a Shewhart-type chart; if the data exhibit autocorrelation, the traditional form of estimating the process standard deviation (for establishing the control limits) based on the average range or average standard deviation of individual samples (even with the Bonferroni or Dunn-Sidak correction) will result in too frequent false alarms, due to the underestimation of the process total variance.

Mortell and Runger (1995) nicely illustrated this issue through an example with simulated data of a process with 2 streams, in which $\sigma_b = 4\sigma$ and a sustained shift of magnitude 1σ was applied to the mean of one of the streams. The time-series plot of the 2 streams (against control limits) shows clearly that the GCC is virtually unable to signal the shift.

Another effect is that, in the converse situation when σ is large with respect to σ_b, the GCC will have little sensitivity to causes that affect all streams—at least, less sensitivity than would have a chart on the average of the measurements across all streams, since this one would have tighter limits than the GCC. Mortell and Runger (1995) stated as generally true that the latter "can have (...) better performance (than the GCC) when a shift common to all streams occurs".

Therefore, to monitor MSPs with the two components described, Mortell and Runger (1995) proposed using two control charts: First, a chart for the grand average between streams, to monitor the mean level. The type of chart to be used should be chosen according to the dynamics of the mean level: if it has a constant mean, without any serial correlation, a classical $\bar{X}$ chart, or an EWMA or CUSUM chart would be appropriate; if it exhibits autocorrelation, some procedure for monitoring

an autocorrelated process should be used. In any case, it would be a known procedure in the previous literature for univariate processes, so Mortell and Runger (1995) did not focus on it. For monitoring the individual stream components, they proposed using a special range chart (R_t chart), whose statistic is the range *between* streams, that is, the difference between the largest stream average and the smallest stream average (at any time t, the j values x_{tij} in each stream i are averaged; R_t is then the difference between the maximum and the minimum of these averages). If the process is in control and all individual stream components have mean equal to zero, the R_t statistic has mean $d_2\sigma$ and standard deviation $d_3\sigma$, where the constants d_2 and d_3 are based on a "sample size" of s, the number of streams. If a stream undergoes a shift in the mean, R_t will increase. Performance analysis of this scheme has shown its efficiency in detecting causes that affect one stream. In addition, they analysed enhanced versions of the idea, namely, EWMA and CUSUM versions of the R_t chart. They compared the performance of the different schemes considering also the two-sided runs scheme and a pair of CUSUMs, in the maximum and in the minimum of all stream averages. In most cases, either the CUSUM or the runs scheme performed the best, depending on the size of the shift and the number of streams (depending on the number of streams, one scheme is better for small shifts and the other for larger shifts, or the other way around). The EWMA scheme was never the best, although only a few values of the smoothing constant were used, and not optimized. Finally, the authors commented that both the chart on the average of all streams and the R_t chart can be used even when at each sampling time only a subset of the streams are sampled (provided that the number of streams sampled remains constant). The subset can be varied periodically or even chosen at random. They pointed out that it is common in practice to measure only a subset of streams at each sampling time, especially when the number of streams is large. (Lanning et al. 2002, analysed a fractional sampling scheme for an MSP, as we will see in the next section).

Although almost the totality of Mortell and Runger's paper is about the monitoring of the individual streams, the importance of the chart on the average of all streams for monitoring the mean level of the process cannot be overemphasized. Not only is it in general more sensitive than the traditional group chart but also the latter is not appropriate for MSPs of the two-component nature described above.

Still about univariate charts for monitoring the mean level of MSPs, Nelson (1986) mentioned that P. C. Clifford stated to him in a personal correspondence that he used, "instead of an $\bar{x}$ chart, (...) a median chart for overall control (...). Such chart is quite unaffected by the behavior of extreme spindles".

Runger et al. (1996) modeled the MSP with two sources of variation as a particular instance of a multivariate process and analysed their decomposition into principal components. The result is that the major principal component turns out to be the average of all streams, while the remaining components are any set of $s-1$ orthonormal vectors that are orthogonal to the s-dimensional vector $[1\,1\,\ldots\,1]'$, s being the number of streams. They proposed, then, to use a control chart for the main component (average of all streams), as proposed in Mortell and Runger (1995), and to use a univariate chart on the squared norm of the orthogonal projection of the

multivariate vector on the subspace of the last $s - 1$ components, which turns out to be proportional to the between-streams sample variance).

So, their proposal corresponds to replacing Mortell and Runger's (1995) R_t chart by an S^2 chart (whose statistic is in fact the numerator of the sample variance of the stream averages). Both are measures of spread between streams. Analogously to the R_t chart, if there is more than one observation per stream, these should be averaged (in the stream) to yield only one value per stream.

The ARL performance of this S^2 chart is similar to the one of the R_t chart when there are changes in one stream. However, the ARLs decrease when the number of streams that have shifted increases, since the overall variance increases in this case, while the ARLs of the R_t chart are not expected to significantly decrease with the number of streams that shift, since the R_t statistic takes into account only the extreme values. The ARLs of the S^2 chart reach a minimum when half of the streams have shifted. This happens because, with more than half streams shifted, the values tend to concentrate again, just around another location, which makes the expected value of the S^2 statistic start to decrease, reaching its minimum when all streams have shifted—a change to which this chart is insensitive—but which should be detected by the chart on the average of all streams. Finally, they also analysed a MEWMA version of the S^2 chart—which, as expected, shows substantially smaller out-of-control ARLs.

A third scheme that considers explicitly the two sources of variation described above is Epprecht et al. (2011). For controlling the mean level, they proposed (exactly as Mortell and Runger 1995, and Runger et al. 1996) using a chart on the average of all streams; and, for controlling the individual streams, they proposed a GCC on the differences between each stream and the average of all streams. It is worth noting that Mortell and Runger (1995) had pointed out this possibility, even if they have not pursued it. (They reported in the paper, however, that they had evaluated the performance of a CUSUM chart on the extreme differences, but it had been outperformed by the R_t CUSUM chart).

For controlling MSPs well represented by the model described in the beginning of this section, this GCC has not the drawbacks mentioned above of the traditional GCC on the values of the original variables in each stream, because the differences between the streams and their average have zero mean when the process is in control and are, each one, i.i.d. over time. Operationally, the scheme is a bit more cumbersome than working with a single statistic (that is, R_t or S^2), but, if it is implemented in software (even in a spreadsheet), and if all values should be input in any case, there is no practical difference—with the advantage over the other schemes that a signal indicates the stream that has shifted. Conditional formatting in the spreadsheet can also be used to indicate every stream that exceeds a control limit. There are some complicating aspects, however: first, the fact that the differences are taken with respect to an average of themselves introduces a cross-correlation between the differences (even if there is no cross-correlation between the individual components e_{ij}—not directly observable—of the different streams). This cross-correlation is, however, negligible for more than 3 streams. Simulation analysis of the scheme for the case of shifts in only one stream has shown that it outperforms

the R_t and S^2 charts for shifts in the mean larger than 1σ. Since σ is only part of the process variance, it may be the case that shifts smaller than this be not relevant to detect.

Simões (2010) analysed an EWMA version of this scheme, and also schemes for monitoring the variance of the individual streams—an issue that had not been investigated before under the context of the model with two sources of variation. For this last problem, group control charts on the variances of the differences between each stream and the average of streams, in Shewart-like and EWMA versions, have been developed, and their performance has been analysed. Moving-range group charts and their EWMA versions have also been analysed, to cater for the case in which it is not feasible to take more than one observation per stream at each sampling time. (The work is written is Portuguese.)

Meneces et al. (2008) advocated using one separate chart for each stream, due to the enhanced diagnostic features of such scheme, which is now eased by current computer resources. They used real data from an industrial process to compare the in-control performance of this scheme with the performance of Nelson's (1986) runs scheme (which they called "the group method") and of Mortell and Runger's (1995) R_t chart. The comparison was limited to one example, though: a series of 53 samples. The performance analysis of their proposed scheme (one chart for each stream) was made by simulation and was more extensive, considering different numbers of streams and different degrees of correlation between them; also, they considered the cases of shifts in one stream, and in all streams. They also provided values for the control limits coefficients of the Shewhart charts that yield the desired in-control ARL, as a function of the correlation between streams. Although they acknowledged the presence of such correlation, they did not consider separating the two components of variability (common variability and intra-stream variability), as the individual charts they recommended use directly, as monitoring statistics, the values observed in each stream—the same monitoring statistics used by Boyd's group charts with control limits. This may lead to the same drawback of these charts, namely, reduced sensitivity to special causes that affect just one or a few streams. They also ignored the limitations of the runs scheme. In the Conclusions section, they recommended using one chart for each stream, for two reasons. One is its diagnostic feature (for example, when the R_t chart signals, an investigation is needed to determine which stream(s) is (are) affected by special causes). The other is that this scheme is more robust to differences in centring between different streams because it admits adjusting for this case, through simply replacing the observation in one stream by its difference to the in-control mean of the stream, while the other methods would exhibit too many false alarms in this case. (I cannot fully agree with this point though, because a similar adjustment could be applied to the other methods, simply by considering the range of these adjusted observations instead of the range of the raw observations).

4 Other Approaches and Particular Problems

Amin et al. (1999) proposed an EWMA chart for the largest and smallest observations in each sample (the MaxMin EWMA chart). This chart can serve to simultaneously monitor the mean and variability of univariate processes, but serves as well to monitor multiple-stream processes, being sensitive to changes in the overall process mean, in the inter-stream dispersion (as when individual streams shift) or both. It can be seen as an EWMA extension of the $\bar{x}$ (or individuals) group chart. Besides the objective detection criterion provided by the control limits, the visual patterns of the two lines (Max EWMA and Min EWMA) give information about the type of change. With a change in the overall mean, both lines would move up or down together, and with an increase (decrease) in the process variance (or inter-stream variance, in the case of an MSP), the lines will diverge (converge); and so on. Optionally an $\bar{x}$-EWMA line can be added to the chart. They reported that the chart has been successfully applied in the nylon fibers industry to monitor 100 similar 8-stream processes.

Amin et al. (1999) analysed the performance of the chart and compared it with a pair of EWMA charts, a two-sided EWMA chart for $\bar{x}$ and a one-sided EWMA chart for $\ln(S^2)$. For shifts only in the mean, the joint charts perform better; with a slight increase in the standard deviation, the difference in performance reduces and for larger increases in the standard deviation the MaxMin EWMA chart outperforms the joint scheme. They gave guidance for designing the chart, and proposed a variant that is also sensitive to reductions in the process standard deviation.

Their scheme assumes a classical in-control process. In the case of MSPs, this corresponds to the case of all streams being independent and identically distributed, with no dynamics over time. An interesting issue that remains open is, then, how the scheme would perform in the case of MSPs with two sources of variation such as described in Sect. 3, and whether and how it should be adapted (and designed) in order to deal with such processes.

In a very comprehensive paper, Wludyka and Jacobs (2002a) extended the group chart and runs scheme to the case of multi-stream binomial processes, both for the case of homogeneous streams and of streams with different nonconforming rates. They made an extensive analysis of the use of the group chart, of the runs scheme and also of a p chart for monitoring the overall process nonconforming rate, and gave several tables of ARLs for shifts in a single stream and in the process (all streams). They also discussed the possible impact of the estimation error in Phase I over the Phase II performance of the control schemes, as well as the design of a useful control scheme. In a couple of conference papers (Wludyka and Jacobs 2002b; and Wludyka 2002), the same authors and the first one alone proposed chi-squared control charts for controlling homogeneous and non-homogeneous multistream binomial processes, respectively. To the best of my knowledge, these are the only authors to have investigated MSP control by attributes,

Lasi et al. (2004) presented an application of group control charts. The process exhibited an autocorrelated common component and the first attempt at using joint $\bar{X}$ and R group charts (with 3-sigma control limits established by the traditional expressions) led to too many false alarms of the $\bar{X}$ group chart. This was due to the underestimation of the total process variance (since the $\bar{R}/d_2$ estimator does not capture the variation of the common component over time). Recall this is one of the limitations mentioned earlier of the group control charts in their traditional form. They overcame the problem by changing the estimator of the variance to capture the total variance. However, they still used 3-sigma control limits; that is, even if they used for "sigma" the estimate of the total standard deviation, they made no correction for the number of streams. By the way, the correction to apply would not be a straightforward one. The Bonferroni or the Dunn-Sidak formulae are not appropriate because they assume independence between streams (no common component). Anyway, ignoring the need for a correction would increase the false-alarm rate. Their chart still worked because the number of streams of the particular process was small with only 4 streams.

Lanning et al. (2002) considered an MSP "where it is possible to monitor only a fraction of the total streams at a given time", which is "of interest in those processes where the spread of production is great and includes a large number of streams, but the ability to monitor the process is not fully automated". The benefit is, obviously, less sampling effort. In addition, they applied the variable sample size and sampling interval (VSSI) technique to the sampling, increasing the efficiency of the monitoring. (For descriptions of VSSI control charts, see Prabhu et al. 1994, and Costa 1997.) In their proposed scheme, the measurements from a subset of streams constituted a sample, and their average was plotted into a VSSI $\bar{X}$ chart. The purpose was to control the common component, and the chart should signal when all streams shift or a significant number of them do. In that particular process shifts in a single stream (or in a very few streams) were rare and/or not relevant. Even so, the authors proposed the use of an R_t chart (as proposed by Mortell and Runger 1995) together with the VSSI $\bar{X}$ chart to detect possible shifts in individual streams. The performance analysis of the scheme (with the ATS — average time to signal — as performance measure) revealed its effectiveness in monitoring the process overall mean (as opposed to the means of individual streams) when the number of streams is large, a problem that had not received much attention before.

Bothe (2008) introduced a capability index he named *average* C_{pk} as a measure of capability of the combined output of all streams. The index is derived from the overall nonconforming fraction of the process, and has the advantage of being applicable to both variables and attribute type data. This seems to be the only work thus far on the capability of MSPs.

Liu et al. (2008) examined a somewhat different application of multiple stream monitoring, the one of a (single stream) truck assembly process alignment where product characteristics are measured by multiple gauges in parallel, and the goal is to detect biases in one (or some) of the gauges. The task of detecting shifts in the production process (which could be left to standard Shewhart charts) was out of the scope of their work. Gauges were assigned haphazardly to units of

the product, each part being measured by one gauge, and there was no "common component" of variation introducing cross-correlation between streams (gauges). They considered both the situation in which there are retrospective data to estimate the unknown parameters (especially the variance) of the measurement system under stable conditions and the situation in which such data are not available. For the former case (which they called the known variance case) they proposed a likelihood ratio method and analysed its performance as well as the performance of an F test for simultaneous comparison of the means of the different gauges. For the latter case (which they called the unknown variance case), they analysed the performances of another version of likelihood ratio test, of Mortell and Runger's R_t chart and of an S chart on the standard deviation of the output averages of gauges in each subgroup (to measure the between-gauges dispersion). With shifts (biases) in only one gauge, the likelihood ratio methods consistently outperformed their competitors in each case. The larger the number of gauges, the larger the differences in performance between methods in a same group (known variance or unknown variance); with only two gauges, the methods perform identically. However, if the number of gauges is small and two gauges shift simultaneously in the same direction, the performance of the likelihood methods is not as good.

Naturally, with increases in the sample size the power of all methods increases and the differences between their performances decrease.

Liu et al. (2008) mentioned that extensions of their work include "adapting any of the proposed test statistics to a sequential control chart such as a (...) CUSUM or (...) EWMA".

An open issue is to verify whether the likelihood test statistics proposed can be applied (with some adaptation if necessary) to the context of a model with a common component exhibiting autocorrelation.

Xiang and Tsung (2008) used a multiple-stream representation to monitor a multi-stage process. They used a state-space model to represent and predict the vector of variables in the different stages (which constitutes the output of a stage and the input to the next one), and converted "the complex multi-stage monitoring problem (...) to a simple multi-stream monitoring problem by applying group exponentially weighted moving average charts to the one-step ahead forecast errors of the model". They illustrated the application and effectiveness of the method with data from automobile hood manufacturing and workpiece assembly.

Mei (2010) proposed and analysed a scheme based on the sum (over all streams) of CUSUM statistics of the individual streams. The individual CUSUM statistics are based on the logarithm of a likelihood ratio statistic. The procedure is designed to detect changes in a moderately large number of data streams, and is robust and omnibus in the sense that it is sensitive to changes in different combinations of affected streams, even with different changepoints (changes occurring at different times), since the CUSUM statistics accumulate information over time. In the performance analysis, the author compared this scheme with another one, based on the maximum CUSUM statistic among the streams (see Woodall and Ncube 1985, and Tartakovsky et al. 2006), which performs better when one or just a few streams change (simulations were conducted for the case of 100 streams

and the max CUSUM outperformed the sum of CUSUMs when up to 5 streams changed). The development of the scheme was motivated by the general (not necessarily industrial) problem of monitoring multiple streams of data, of which the author mentioned security and biosurveillance systems as examples. The post-alarm problem of identifying the stream(s) that changed was left to one side. The validity conditions for the method are based on the strong assumption that the observations are independent over time as well as among data streams. This assumption will be reasonable, however, if the scheme is applied to "the residuals of a spatio-temporal model rather than the original observations".

Epprecht and Barros (2013) studied a filling process application where the stream variances were similar, but the stream means differed, wandered, changed from day to day, were very difficult to adjust, and the production runs were too short to enable good estimation of the parameters of the individual streams. The solution adopted to control the process was to adjust the target above the nominal level to compensate for the variation between streams, as a function of the lower specification limit, of the desired false-alarm rate and of a point (shift, power) arbitrarily selected. This would be a MSP version of "acceptance control charts" (Montgomery 2012, Sect. 10.2) if taking samples with more than one observation per stream were feasible. Since this was not the case, the target, the upper specification limit (less solid a restriction than the lower one) and the upper control limit of the chart were redefined as a function of the desired false-alarm rate, the point (shift, power) specified, the lower specification limit and the variability between stream means.

Jirasettapong and Rojanarowan (2011) were concerned with the selection of the appropriate control charts for monitoring an MSP, considering the degree of correlation among streams, the number of streams, the feasibility of using one chart per stream, whether the streams can or cannot be centred on the same target, and the relevant shift size. They provided a taxonomy, in the form of a tree diagram, to classify the MSPs according to these factors, where the leaves are the recommended control charts. The work has the merit of making explicit some aspects that may invalidate the use of some methods, and of illustrating the proposal with a real case. However, they limited the options considered to just a few schemes and have not considered all relevant aspects for the characterization of MSPs. This is a beginning, but more research is needed on this issue of selection of the appropriate control scheme for MSPs.

5 Perspectives: Other Applications, Open Issues, Challenges and Opportunities for Research

This review is constrained to the industrial context of multiple-stream processes. The need to monitor multiple streams of data, however, arises in other applications. As said before, Mei (2010) mentioned security and biosurveillance systems as examples. Woodall et al. (2010) pointed out the similarity between MSPs and the

"large number of multiple streams of data" used in health-related surveillance, such as "data available over time on a number of different subregions, hospitals or physicians", but added that the underlying assumptions of the models and methods for industrial MSPs "may be too restrictive for these methods to be applied directly to health-related monitoring". He commented that "when many patient subgroupings or several healthcare indicators are simultaneously considered, it can be a challenge to control the rate of false alarms and yet retain power to detect meaningful outbreaks". On this specific issue, see Marshall et al. (2004).

Still in this context, Burkom et al. (2005) analysed alternative tools for monitoring multiple data streams in public health monitoring systems. They underlined the distinction between *consensus monitoring*, which is "the testing of a single hypothesis using multiple sources of evidence" (which in turn may be constituted by data of different natures) and *parallel monitoring*, which is "the monitoring of time series representing different physical locations … that are possibly stratified by other covariates". This last problem is more similar to, although usually more complex than, the one of monitoring industrial MSPs. The "streams" in this case are no longer identically distributed, the process dynamics is more complex and most often is a function of factors (the "other covariates" mentioned) that differ from one location to another; seasonality and weekly patterns may be present (not to mention autocorrelation and cross-correlation between different locations) and, as a consequence, such monitoring requires specific methods or adaptation of existing ones.

Burkom et al. (2005) analysed the performance of some univariate and multivariate procedures, focusing mainly on consensus monitoring with procedures based on the combination of p-values from different sources, but concluded by "the need to blend the parallel and consensus monitoring tools to achieve a system with distributed sensitivity and controlled alert rates", since consensus monitoring tools may not detect or identify individual outputs.

I will not linger on biosurveillance, which is, as said, out of the scope of this review. The idea is to show that the development of specific multiple-stream methods or extension of the existing ones for monitoring multiple streams of data in biosurveillance systems is a need and an opportunity for challenging research.

Keeping confined to the industrial context, some open issues, which, of course, are far from constituting an exhaustive list, are the following:

- Most research works consider a small to moderate number of streams. Some processes may have hundreds of streams, and in this case the issue of how to control the false-alarm rate while keeping enough detection power (as mentioned above in the context of healthcare monitoring) becomes a real problem. Textile industries, where, for instance, the number of spools may be of about one thousand, obviated this problem by having an automatic controller in each spool. When this is not the case, FDR methods (Benjamini and Hochberg 1995; and, again, see Marshall et al. 2004) may be considered but it is not clear how to employ them in the design of monitoring schemes for industrial processes.

- Some of the methods mentioned in this review that were developed for specific problems or types of process deserve analysis to verify the possibility of being applied (with adaptation if needed) to other problems/processes. For example, how would the methods in Liu et al. (2008) extend to the case of processes with a common component exhibiting autocorrelation? How would the MaxMin EWMA chart by Amin et al. (1999) perform in the case of an MSP with two sources of variation? Or in which conditions (for which values of the ratio between variance components) does it perform well? Can it be modified to be adapted for this kind of process?
- Real multiple-stream processes can be very ill-behaved. The author of this paper has seen a plant with six 20-stream filling processes in which the stream levels had different means and variances and could not be adjusted separately (one single pump and 20 hoses). For many real cases with particular twists like this one, it happens that no previous solution in the literature is applicable. Developing methods for such specific applications is a need.
- Different monitoring methods were developed assuming processes with different characteristics. The appropriateness and efficiency of each of them depends on the dynamic behaviour of the process over time, on the degree of cross-correlation between streams, on the ratio between the variabilities of the individual streams and of the common component (note that these three factors are interrelated), on the type and size of shifts that are likely and/or relevant to detect, on the ease or difficulty to adjust all streams in the same target, on the process capability, on the number of streams, on the feasibility of taking samples of more than one observation per stream at each sampling time (or even the feasibility of taking one observation of *every* stream at each sampling time!), on the length of the production runs, and so on. So, the first problem in a practical application is to characterize the process and select the appropriate monitoring scheme (or to adapt one, or to develop a new one). This analysis may not be trivial for the average practitioner in industry. A methodological guide for such an analysis can be a very useful contribution. Jirasettapong and Rojanarowan (2011) is the only work I have found on the issue of selecting the most suitable monitoring scheme for an MSP. It considers only a limited number of alternative schemes and a few aspects of the problem. More comprehensive analyses are needed.
- An issue that has hardly been tackled is the one of Phase I. There is a huge body of research on Phase I analysis of control charts (for a review, see Jensen et al. 2006); we found no such a study in the context of MSPs, with the only exception of Wludyka and Jacobs (2002a,b) for the case of MSP control by attributes.
- Similarly to multivariate process control (of which multiple-stream process control is a particular case), after a signal there may be the need to identify the stream(s) that have shifted. Many methods do not provide this identification automatically, so there is opportunity to investigate procedures for accomplishing it.
- There is place for a more extensive comparative analysis of performance of existing methods, since the performance evaluations in different papers do not consider the same cases, shifts, and number of streams affected.

Acknowledgements This work has been supported by the CNPq (the Brazilian Council for Scientific and Technological Development), project number 307453/2011-1. I thank Bill Woodall for his comments on the first version of this paper, and for giving me the idea of writing it, some time ago.

References

Amin, R., Wolff, H., Besenfelder, W., & Baxley, R., Jr. (1999). EWMA control charts for the smallest and largest observations. *Journal of Quality Technology*, *31*(2), 189–206.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289–300.

Bothe, D. R (2008). Process Capability Indices for Multiple Stream Processes. In *Encyclopedia of statistics in quality and reliability*. New York: Wiley

Boyd, D. F. (1950). Applying the group chart for $\bar{X}$ and R. *Industrial Quality Control*, *7*, 22–25.

Burkom, H. S., Murphy, S., Coberly, J., & Hurt-Mullen, K. (2005). Public health monitoring tools for multiple data streams. In *Syndromic Surveillance: Reports from a National Conference, 2004* (54 (Suppl.), pp. 55–62). Morbidity and Mortality Weekly Report 2005.

Burr, I. W. (1976). *Statistical quality control methods*. New York: Marcel Dekker

Colbeck, J. P. (1999). *Some alternative methods for monitoring multiple-stream processes*. (Electronic thesis and dissertations). Department of Statistics, University of Manitoba, Winnipeg. (Found in http://hdl.handle.net/1993/1941).

Costa, A. F. B (1997). $\bar{X}$ Charts with variable sample size and sampling intervals. *Journal of Quality Technology*, *29*, 197–204.

Dunn, O. J. (1958). Estimation of the means of dependent variables. *Annals of Mathematical Statistics*, *29*, 2775–279.

Epprecht, E. K., & Barros, I. P. (2013). *Monitoring a Multiple Stream Process with Varying Means. Technical Memorandum 01/2013*, Department of Industrial Engineering, PUC-Rio, Rio de Janeiro.

Epprecht, E. K., Barbosa, L. F. M., & Simões, B. F. T. (2011). SPC of multiple stream processes—a chart for enhanced detection of shifts in one stream. *Produção*, *21*(2), 242–253. doi:10.1590/S0103-65132011005000022.

Grimshaw, S. D., Bryce, G. R., & Meade, D. J. (1999). Control limits for group charts. *Quality Engineering*, *12*(2), 177–184.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, *38*(4), 349–364.

Jirasettapong, P., & Rojanarowan, P. (2011). A guideline to select control charts for multiple stream processes control. *Engineering Journal*, *15*(3), 1–14.

Johnson, R. A., & Wichern, D. W. (2007) *Applied multivariate statistical analysis*, (6th ed.). New York: Pearson Prentice Hall.

Lanning, J. W., Montgomery, D. C., & Runger, G. C. (2002). Monitoring a multiple stream filling operation using fractional samples. *Quality Engineering*, *15*(2), 183–195.

Lasi, G., Mongiello, C., & Scagliarini, M. (2004). Il controllo statistico per processi a flussi multipli: problemi e soluzioni di un caso aziendale. *Statistica*, anno *LXIV*(4), 707–719 (In Italian).

Liu, X., MacKay, R. J., & Steiner, S. H. (2008). Monitoring Multiple Stream Processes. *Quality Engineering*, *20*, 296–308.

Marshall, C., Best, N., Bottle, A., & Aylin, P. (2004). Statistical Issues in the Prospective Monitoring of Health Outcomes Across Multiple Units. *Journal of the Royal Statistical Society, Series A*, *167*(3), 541–559.

Mei, Y. (2010). Efficient scalable systems for monitoring a large number of data streams. *Biometrika*, *97*(2), 419–433.

Meneces, N. S., Olivera, S. A., Saccone, C. D., & Tessore, J. (2008). Statistical Control of Multiple-Stream Processes: A Shewhart Control Chart for Each Stream. *Quality Engineering*, *20*, 185–194.

Montgomery, D. C. (2012). *Introduction to statistical quality control* 7th edn. New york: Wiley

Mortell, R. R., & Runger, G. C. (1995). Statistical Process control for multiple stream processes. *Journal of Quality Technology*, *27*(1), 1–12.

Nelson, L. S. (1986). Control Chart for multiple stream processes. *Journal of Quality Technology*, *18*(4), 225–226.

Ott, E. R., & Snee, R. D. (1973). Identifying useful differences in a multiple-head machine. *Journal of Quality Technology*, *5*(2), 47–57.

Prabhu, S. S., Montgomery, D. C., & Runger, G. C. (1994). A combined adaptive sample size and sampling interval control scheme. *Journal of Quality Technology*, *26*, 164–176.

Pyzdek, T. (1992). *Pyzdek's Guide to SPC, Vol. Two: Applications and Special Topics*. ASQ-Quality Press: Milwaukee, WI, and Quality Publishing, Inc.: Tucson, AZ.

Runger, G. C., Alt, F. B., & Montgomery, D. C. (1996). Controlling multiple stream processes with principal components. *International Journal of Production Research*, *34*(11), 2991–2999.

Sidak, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distribution. *Journal of the American Statistical Association*, *62*, 626–633.

Simões, B. F. T. (2010). *Controle Estatístico de Processos Multicanal*. (Doctoral thesis). Departamento de Engenharia Industrial, PUC-Rio, Rio de Janeiro (In Portuguese).

Tartakovsky, A. G., Rozovskha, B. L., Blazeka, R. B., & Kim, H. (2006). Detection of intrusions in information systems by sequential change-point methods (with Discussion). *Statistical Methodology*, *3*, 252–340.

Wise, S. A., & Fair, D. C. (1998). *Innovative control charting-practical spc solutions for today's manufacturing environment*. ASQ-Quality Press: Milwaukee, WI.

Wludyka, P. S. (2002). Controlling Non-Homogeneous Multistream Binomial Processes with a Chi-Squared Control Chart. In *Proceedings of the ASA Meeting 2002, Quality and Productivity Research Conference*, Tempe, Arizona [CD-ROM].

Wludyka, P. S., Jacobs, S. L. (2002a). Runs rules and p-charts for binomial multistream processes. *Communications in Statistics: Simulation and Computation*, *31*(1), 97–142.

Wludyka, P. S., Jacobs, S. L. (2002b). Controlling homogeneous multistream binomial processes with a chi-squared control chart. In *Proceedings of the 33rd Annual Meeting of the Decision Sciences Institute* (pp.2254–2263). San Diego, California, [CD-ROM].

Woodall, W. H., Ncube, M. M. (1985). Multivariate CUSUM Quality Control Procedures. *Tecnometrics*, *27*(3), 285–292.

Woodall, W. H., Grigg, O. A., Burkom, H. S. (2010). Research issues and ideas on health-related surveillance. In H.-J. Lenz, P.-T. Wilrich & W. Schmid (Eds.), *Frontiers in Statistical Quality Control 9* (pp. 145–155). Heidelberg: Physica-Verlag

Xiang, L., Tsung, F. (2008). Statistical monitoring of multi-stage processes based on engineering models. *IIE Transactions*, *40*(10), 957–970.

Regenerative Likelihood Ratio Control Schemes

Emmanuel Yashchin

Abstract We discuss the problem of monitoring where models driving the data are undergoing abrupt changes in time, such as shifts or drifts. We introduce a unified methodology based on the use of likelihood ratio tests that enables one to obtain control schemes that provide both good (and, under some conditions, optimal) statistical performance and are relatively easy to implement. These schemes depend on just one design parameter and require a limited computational effort that is dynamically adjusted based on process conditions. An example pertaining to multivariate control of the normal mean is discussed in detail.

Keywords Average Run Length • Detection • False Alarms • Monitoring

1 Introduction

In many applications involving monitoring of data streams one can justify an assumption that the model describing the data undergoes abrupt changes of unknown magnitude that occur at unknown moments of time. Depending on the objectives, such models can be used, among other things, to

- detect as quickly as possible onset of unfavorable process conditions while maintaining a low rate of false alarms (the detection problem)
- estimate the current process parameters (filtering); this problem is of importance in situations where on-line corrective actions are possible that can bring the situation under control without directly addressing the cause of the unfavorable change
- test retrospectively the assumption of data homogeneity; this problem is of importance in process capability analysis
- locate retrospectively points of change and estimate magnitudes of change in process parameters; this segmentation problem is used in the problem of process diagnosis.

E. Yashchin (✉)
IBM, Thomas J. Watson Research Center, Box 218, Yorktown Heights, NY 10598, USA
e-mail: yashchi@us.ibm.com

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_5

The above problems have been subject of intensive research (e.g., see Telksnys 1986; Basseville and Nikiforov 1994; Tsui and Woodall 1993; Carlstein et al. 1995; Lai 1995; Sullivan and Woodall 1996; Hawkins and Olwell 1998; Stoumbos et al. 2000; Sullivan 2002; Garthoff et al. 2013; Kenett et al. 2014). They can typically be addressed in either fixed sample or sequential settings. Recently, new procedures based on the generalized likelihood ratio (GLR) have been proposed by several authors, e.g., see Reynolds and Lou (2010); Wang and Reynolds (2013).

In this article we focus on the problem of sequential detection of an unfavorable change that is of high relevance in statistical process control (SPC) applications. We address the problem in a general setting by using groups of likelihood ratio tests performed sequentially at the current point in time. The depth of history considered relevant for detection purposes will vary depending on process conditions. At some points all the prior history will be declared irrelevant and discarded. The key feature of our approach is selection of such points in a manner that does not discard useful information.

In Sect. 2 we introduce the problem setup and formulate the likelihood strategy. In Sect. 3 we present the basic approach, based on selection of regeneration points, that modifies this strategy in a way where it becomes practical. Note that this strategy contains some important enhancements compared to that considered in earlier literature, eg. see Yashchin (1995, 1997). In Sects. 4 and 5 we apply this strategy to the well known problem of monitoring the mean of a multivariate normal population. As the reader will see, one appealing feature of the proposed scheme is that it cannot be beaten by a Shewhart scheme, even for large changes in the process mean, and yet, like a Shewhart scheme, it only depends on a single design parameter. In Sect. 6 we discuss generalizations involving weighting information based on time and other criteria. Finally, Sect. 7 contains a discussion.

2 The Problem Setup

Let $\{\boldsymbol{X}_i\}$, $i = 1, 2, \ldots$ be a sequence of (generally multivariate) observations that we intend to monitor. In practical applications, they may represent, for example, component-wise deviations between observed and projected features in a lithography step of a chip manufacturing process, goodness-of-fit characteristics of an assumed model, reliability characteristics of manufactured components, and so forth.

The stochastic behavior of the sequence is determined by the vector of parameters $\boldsymbol{\theta}$. We shall initially assume that all the components of this vector are of primary interest. In the last section we discuss the case where behavior of $\{\boldsymbol{X}_i\}$ also depends on nuisance parameters. In general, the vectors $\{\boldsymbol{X}_i\}$ can form a serially correlated sequence; however, to simplify the presentation, in what follows we will assume independence except where stated otherwise. Denote the most recent moment of time by T and the corresponding most recently observed observation by $\boldsymbol{X}_T$. Denote the joint distribution of m most recent observations by

$f_{\boldsymbol{\theta}}(\boldsymbol{X}_{T-m+1}, \ldots, \boldsymbol{X}_{T-1}, \boldsymbol{X}_T)$, and denote its natural logarithm (the log-likelihood) by $L_{m,T}(\boldsymbol{\theta})$.

To set up the problem of detection, we first specify the acceptable region Ω_0 in which $\boldsymbol{\theta}$ should reside under normal operating conditions and the unacceptable region, Ω_1. Note that the union of Ω_0 and Ω_0 does not need, in general, to cover the whole parameter space: there will generally exist a "grey" area in between. This three-zone approach to design of control schemes (e. g., see Woodall 1985; Yashchin 1985) is motivated by practical convenience: in many industrial applications, an engineer will have no difficulty specifying areas that are distinctly "good" or "bad"; however, dividing the parameter space into two regions to separate "good" values from "bad" ones could prove to be a challenge.

Performance of detection schemes is typically measured in terms of the run length (RL), a random variable representing the number of observations taken until a signal is triggered. In general, one would like this variable to be large when $\boldsymbol{\theta} \in \Omega_0$ (i.e., a low false alarm rate) and small when $\boldsymbol{\theta} \in \Omega_1$ (i.e., good sensitivity with respect to out-of-control conditions). The most popular measure is the average run length (ARL).

Now let us denote, for a given set of last m observations,

$$L^*_{m0,T} = \max_{\boldsymbol{\theta} \in \Omega_0} L_{m,T}(\boldsymbol{\theta}), \quad L^*_{m1,T} = \max_{\boldsymbol{\theta} \in \Omega_1} L_{m,T}(\boldsymbol{\theta}), \quad D^*_{m,T} = L^*_{m1,T} - L^*_{m0,T}. \tag{1}$$

Denote by m_0 the minimal depth of data m for which $\boldsymbol{\theta}$ is estimable. Then one can define a general strategy that leads to powerful control schemes as follows.

Likelihood Ratio (LR) Strategy: Trigger an out of control signal at time T if $D^*_{m,T} > h$ for some $m \geq m_0$ and pre-specified threshold $h \geq 0$.

The above strategy leads to powerful procedures for a wide class of situations involving control of univariate and multivariate processes with or without serial correlation, e.g., see Basseville and Nikiforov (1994); Hawkins and Olwell (1998). Unfortunately, its practical usefulness is limited since it requires one to examine the whole data set to reach a decision as to whether a signal is to be triggered at time T. Therefore, any practical application of the LR strategy involves choosing a window of size M and triggering a signal only if $D^*_{m,T} > h$ for values $m_0 \leq m \leq M$. In effect, this amounts to running a truncated SPRT backwards in time. As shown in Lai (1995), in the univariate case one can achieve asymptotic efficiency of the LR test by examining only a subset of values m. However, this approach could still require a search going deep into history to establish whether a signal is to be triggered. Another approach proposed by Lai and Shan (1999) involves using a mixture of GLR window-limited tests. Nikiforov (2001) proposes an approach based on separation of parameter space into several subsets and running a GLR window-limited scheme for various subsets in parallel.

In the next section we introduce an alternative procedure, the regenerative likelihood ratio (RLR). Its primary strengths are statistical power and simplicity. It can also be used in conjunction with other monitoring schemes, for example those

by Lai and Shan (1999) and Nikiforov (2001), to dynamically establish the size of a window in a window-limited GLR scheme.

3 The Regenerative Likelihood Ratio Approach

Based on pre-specified threshold $h \geq 0$ and minimal window size m_0, this procedure calls for determining the depth M_T dynamically, based on the previous history:

Regenerative Likelihood Ratio (RLR) Scheme: Given that at time T the last regeneration point was registered M_T units of time ago,

- trigger a signal if $D^*_{m,T} > h$ for **some** $m_0 \leq m \leq M_T$.
- if $D^*_{m,T} \leq 0$ for **every** $m_0 \leq m \leq M_T$, declare T the new regeneration point.
- otherwise, denote by m_T the maximal value of m in $[m_0, M_T]$ for which $D^*_{m,T} > 0$, and declare $T - m_T$ a new regeneration point.

We note that this formulation is different from that appearing in the earlier literature (see Yashchin 1995, 1997). The new method for selecting the regeneration point leads to substantial reduction in computing effort without a significant loss of statistical power. Within the framework of the RLR scheme selection of window size is completely automated, and thus it does not play the role of a design parameter, in effect making the design problem one-dimensional.

The process of selection of a new regeneration point is shown in Fig. 1. The intuitive explanation of the reason for moving the regeneration point from $T - M_T$ to $T - m_T$ is as follows:

- if at the present point in time the process is in control then there is no reason to have the history window at all, and so the move makes sense;
- if the process is at some "bad" level and will continue at this level in the future, the points $T - M_T + 1, \ldots T - m_T$ that are not contributing to detection now are not likely to contribute to detection later, and so can safely be discarded.

To implement the above procedure as a chart, it is convenient to define a process s_T at time T as follows:

$$s_T = \max_{m_0 \leq m \leq M_T} D^*_{m,T}. \tag{2}$$

Now we can plot s_T on a control chart with domain $(0, h)$ and trigger a signal at time T if $s_T > h$. If $s_T \leq 0$ declare T the new regeneration point. Otherwise, we find m_T as described above and declare $T - m_T$ a new regeneration point. This type of chart is liked by end users because it is relatively easy to interpret: the value of s_T represents evidence accumulated at time T against the hypothesis that the process is in control.

It is not difficult to see that in the "simple vs. simple" case where one seeks to detect a change from a fixed "good" value θ_0 of a parameter to a *known* "bad"

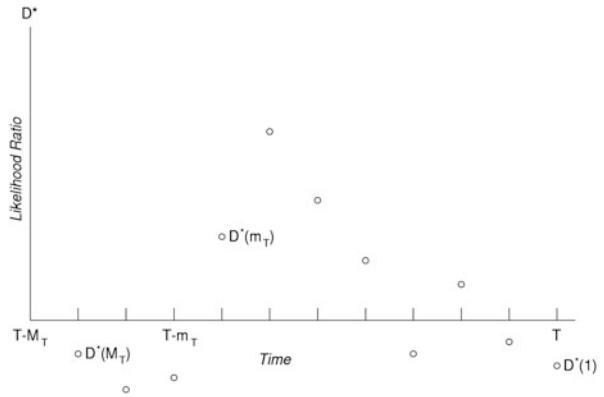


Fig. 1 The process of establishing a new regeneration point at the current point of time T. The points shown on the plot are computed at time T: $D^*(m)$ is the score in favor of the hypothesis that the parameter is in the "bad" region computed based on the data window of size m. $T - M_T$ is the current regeneration point. The largest window size for which evidence in favor of the "bad" hypothesis is positive is m_T—therefore, $T - m_T$ is declared a new regeneration point

value θ_1, the value of s_T coincides with that of the Page's Cusum scheme (also known as the decision interval scheme). Therefore, the RLR can be viewed as the natural extension of the Cusum technique to cover more complex data models, e.g. involving serial correlation, cross-correlation, and multivariate data.

In practical applications one may prefer more simple schemes that sacrifice some statistical efficiency to reduce complexity even further. For example, under "mild" out-of-control conditions the regeneration point could remain intact for a prolonged period in time, resulting in increasing computing costs related to necessity of exploring progressively deeper data windows - one may want impose a limit on such costs. We define three possible simplifications:

- *Window-limited RLR*: Regular RLR scheme but the depth of search is limited to M_0. In other words, the above rules defining the RLR call for triggering a signal $D^*_{m,T} > h$ for **some** $m_0 \le m \le min(M_0, M_T)$; T is declared a new regeneration point if $D^*_m \le 0$ for **every** $m_0 \le m \le min(M_0, M_T)$. If this does not hold, then we denote by m_T the maximal value of m in $[m_0, min(M_0, M_T)]$ for which $D^*_{m,T} > 0$, and declare $T - m_T$ a new regeneration point.
- RLR(k): Regular RLR scheme, but explores only values of m on a k - spaced grid. In addition, always explore $m = M_T$. Note that in the special case $k = \infty$ this scheme reduces to the sequence of generalized Sequential Probability Ratio Tests:

  1. Define scheme s_T at time T by $s_T = D^*_{M_T,T}$. Signal if $s_T > h$.
  2. If $s_T \le 0$ declare T the new regeneration point.

- Same as RLR(k), but also explore $m = m_0$. When $k = \infty$ we obtain a generalized Cusum-Shewhart scheme.

As k increases, these simplified schemes develop an undesirable property as a price for reduced complexity: their worst case performance deteriorates. The Lorden (1971) criterion demands, roughly speaking, that for a given rate of false alarms the average detection capability will be guaranteed to be at a certain prescribed level *independently of the data history preceding the change*. The simplified RLR procedures can appear to have good steady state or initial state statistical performance, but they run into situations where previously recorded favorable process history impedes detection capabilities with respect to subsequent changes. An example is given in Sect. 5.

The window-limited approach can also be used in conjunction with RLR(k). In general, one needs to exercise caution when selecting a suitable value of M_0 for RLR. Choosing too small a value for M_0 could lead to low detection capability with respect to unacceptable changes of relatively low or moderate magnitude. Choosing an excessively large value for M_0 could result in higher than necessary computational costs when the process level is acceptable.

4 Example: Monitoring of the Multivariate Normal Mean

To illustrate the Regenerative Likelihood Ratio approach, consider the problem of monitoring the mean of the multivariate normal population. Let us assume that

$$\boldsymbol{X}_i \sim N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma}), \qquad i = 1, 2, \ldots, \tag{3}$$

where $\boldsymbol{\mu}$ is the p-dimensional process mean and $\boldsymbol{\Sigma}_{(p\times p)}$ is assumed to be known. Let us assume that the monitored parameter is the distance from the target centroid, $\boldsymbol{\mu}_0$:

$$\lambda = \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|_\Sigma = \sqrt{(\boldsymbol{\mu} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)}, \tag{4}$$

and that the acceptable and unacceptable regions for the monitored parameter are $\lambda \le \lambda_0$ and $\lambda \ge \lambda_1$, respectively, where $\lambda_0 < \lambda_1$. For a two-dimensional case the graphical representation of the problem setup is given in Fig. 2.

To implement the RLR scheme we have to optimize, for every window m, the log-likelihood, in both acceptable and unacceptable regions, and then construct the likelihood ratio test. The log-likelihood at time T for window size m is:

$$L_{m,T}(\boldsymbol{\mu}) = -0.5m \times ln\,[(2\pi)^p \|\boldsymbol{\Sigma}\|] - 0.5 \sum_{i=1}^{m} \|\boldsymbol{X}_{T-i+1} - \boldsymbol{\mu}\|_\Sigma^2 \tag{5}$$

Denote the MLE based on window of size m and its distance from the centroid by

$$\boldsymbol{\mu}^*(m) = \bar{\boldsymbol{X}}(m) = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{X}_{T-i+1}; \quad \lambda^*(m) = \|\boldsymbol{\mu}^*(m) - \boldsymbol{\mu}_0\|_\Sigma. \tag{6}$$

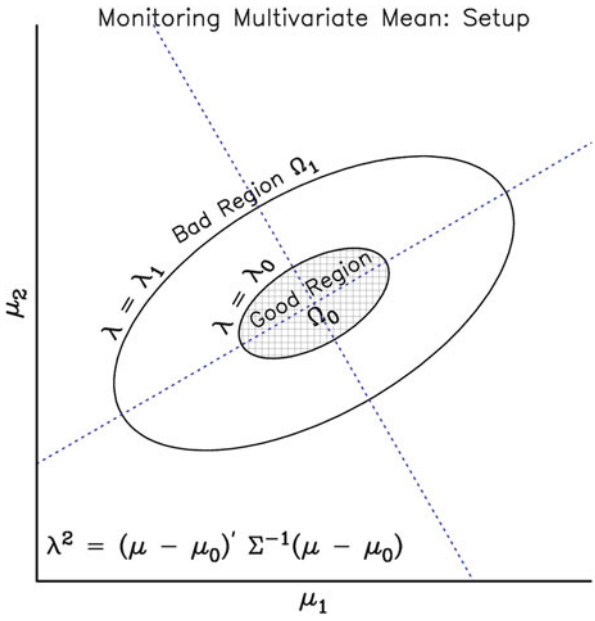


Fig. 2 The setup of the problem for monitoring the multivariate normal mean

Note that $\boldsymbol{\mu}^*(m)$ and $\lambda^*(m)$ defined above depend implicitly on T. Here and in what follows, this dependence will not be explicitly reflected in the formulas in order to keep the notation simple. We will also note that λ is estimable based on the data window of size $m = 1$; therefore, the minimal search window in RLR schemes can be set to $m_0 = 1$.

After some algebra involving constrained optimization with a Lagrange multiplier, one can show that $\boldsymbol{\mu}^*(m, \lambda)$, the MLE computed under the constraint $\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|_\Sigma = \lambda$, is given by a linear combination of the target centroid $\boldsymbol{\mu}_0$ and the unconstrained MLE. In particular,

$$\boldsymbol{\mu}^*(m; \lambda) = \boldsymbol{\mu}_0 + \frac{(\boldsymbol{\mu}^*(m) - \boldsymbol{\mu}_0) \times \lambda}{\lambda^*(m)}. \tag{7}$$

The likelihood optimization process with mean constrained to the acceptable region is illustrated in Fig. 3. By using the property (7) one can derive the RLR scheme:

- Trigger a signal at time T if for some $1 \le m \le M_T$

$$\begin{array}{lll} k_\lambda \le \lambda^*(m) \le \lambda_1 & \text{and} & m(\lambda_1 - \lambda_0)\,[\lambda^*(m) - k_\lambda] > h_\lambda \quad \text{or} \\ \lambda^*(m) \ge \lambda_1 & \text{and} & 0.5m\,[\lambda^*(m) - \lambda_0]^2 > h_\lambda, \end{array}$$

where h_λ = *signal level* and k_λ = *reference value* = $(\lambda_0 + \lambda_1)/2$.

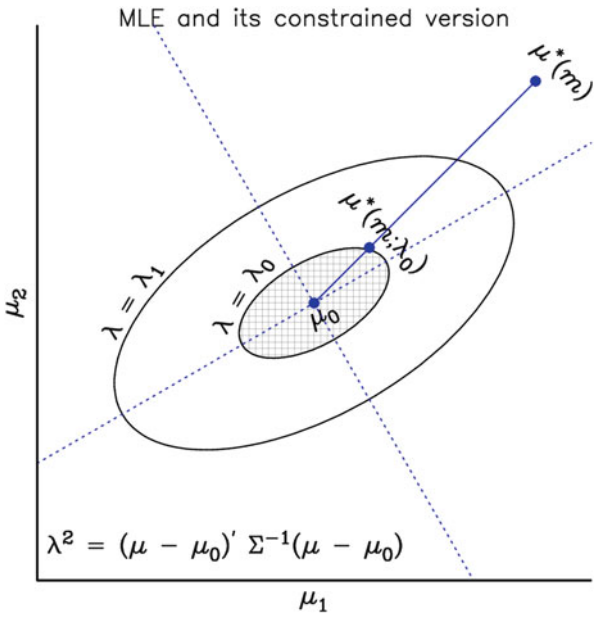


Fig. 3 The MLE $\boldsymbol{\mu}^*(m)$ of the mean $\boldsymbol{\mu}$ based on the window of depth m. The MLE $\boldsymbol{\mu}^*(m, \lambda_0)$ under the constraint $\lambda \leq \lambda_0$ is given by the linear combination of $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}^*(m)$

- If $\lambda^*(m) \leq k_\lambda$ for all $1 \leq m \leq M_T$, then declare T a new regeneration point. Otherwise, find m_T using the process described in Sect. 3, and declare $T - m_T$ a new regeneration point.

5 Selected ARL Comparisons

To illustrate the method, let us consider the bivariate normal case with $\boldsymbol{\Sigma} = I$, acceptable region consisting of a single point, $\lambda = 0$ and unacceptable region $\lambda \geq 1$. We consider several RLR schemes along with two other popular schemes, namely, the MC1 scheme proposed by Pignatiello and Runger (1990) and the Shewhart T^2 scheme which signals when $\|\boldsymbol{x}_i - \boldsymbol{\mu}_0\|^2_\Sigma$ exceeds a pre-specified threshold. The ARLs in the table below are computed under the assumption that no data is available prior to change-point.

Distance λ	RLR	RLR(∞)	MC1	T^2
0.0	200	200	203	200
1.0	10.5	9.43	9.28	42.0
1.5	5.61	4.91	5.23	15.8
2.0	3.55	3.07	3.69	6.9
2.5	2.52	2.20	2.91	3.5
3.0	1.91	1.72	2.40	2.2

As the table above illustrates, the RLR scheme is *uniformly* more powerful than the Shewhart-type T^2 scheme that is prevalent in today's industrial environment. This popularity is explained not only by its relative simplicity, but also by the fact that it is very difficult to beat in detection of large changes. Such changes are of great importance to scheme designers, even though the most typical change they are interested in detecting may be of smaller magnitude. For example, the MC1 scheme cannot beat the T^2 scheme for larger changes ($\lambda \geq 3$). Compared to MC1, the RLR scheme is more powerful for changes above $\lambda = 1.5$, while remaining competitive in the domain $\lambda < 1.5$. Finally, note that while RLR(∞) appears to be more powerful in this comparison, its "worst case" performance suffers from the fact that it allows previously recorded "good" history of the process to interfere with detection capability.

6 Weighted Likelihood Schemes

In this section we discuss a generalization of the RLR approach that involves weighting the observations. We refer to this approach as the weighted RLR technique. In practical situations weights can be introduced when:

- there is a reason to believe that some of the observations should carry more weight in decision making, because of reasons not reflected in the model. For example, a more stable data gathering environment may give more credibility to corresponding observations—and this is reflected in higher weights (type 1 weighting).
- it is not known a-priori what type of change the monitored parameter will undergo. If the only possibilities are shifts or drifts, then one could use two RLR schemes: one to detect the presence of unfavorable shifts, and another to detect a drift. However, one can opt for a less complex route and apply a single RLR scheme with higher weights given to more recent observations. Such a scheme will have much better performance than the usual RLR with respect to *drifts* in $\boldsymbol{\theta}$, while maintaining good statistical power with respect to *shifts* (type 2 weighting). In many practical applications it is convenient to choose weights of type $w_i = \gamma^i, \ \ i = 0, 1, \ldots$ which lead to *Geometric* RLR schemes.

We will now illustrate construction of a weighted RLR scheme of type 2. Let $w_0, w_1, \ldots, w_{m-1}$ be weights associated with $\boldsymbol{X}_T, \boldsymbol{X}_{T-1}, \ldots, \boldsymbol{X}_{T-m+1}$, respectively, for any fixed m (in other words, w_0 is associated with the last observation, w_1—with the previous one, etc.). As noted above, the weights will typically form a decreasing sequence to provide emphasize on the most recent information. The weighted log-likelihood function corresponding to the last m observations is given by:

$$L_{m,T}^{(w)}(\boldsymbol{\theta}) = \sum_{i=T-m+1}^{T} w_{T-i} \log f_{\boldsymbol{\theta}}(\boldsymbol{X}_i | \boldsymbol{X}_{i-1}, \boldsymbol{X}_{i-2}, \ldots). \tag{8}$$

In light of the RLR approach described in Sect. 3, one can now construct a weighted RLR control scheme by defining

$$L_{m0,T}^{*(w)} = \max_{\boldsymbol{\theta} \in \Omega_0} L_{m,T}^{(w)}(\boldsymbol{\theta}), \qquad L_{m1,T}^{*(w)} = \max_{\boldsymbol{\theta} \in \Omega_1} L_{m,T}^{(w)}(\boldsymbol{\theta}), \qquad D_{m,T}^{*(w)} = L_{m1,T}^{*(w)} - L_{m0,T}^{*(w)} \tag{9}$$

and triggering an out of control signal at time T if $D_{m,T}^{*(w)} > h$ for some $m \geq m_0$ and threshold $h \geq 0$, while adhering to rules for determining regeneration points given in Sect. 3.

7 Discussion

In general, when considering a simplified RLR scheme, one should take into consideration not only sacrifices this will require in terms of the ARL curve (they often turn out to be quite tolerable), but also the way in which nuisance parameters are handled in the particular application. Let us denote the vector of nuisance parameters by $\boldsymbol{\eta}$. Then we can define

$$L_{m0,T}^{*} = \max_{\boldsymbol{\theta} \in \Omega_0, \boldsymbol{\eta}} L_{m,T}, \qquad L_{m1,T}^{*} = \max_{\boldsymbol{\theta} \in \Omega_1, \boldsymbol{\eta}} L_{m,T}, \qquad D_{m,T}^{*} = L_{m1,T}^{*} - L_{m0,T}^{*}, \tag{10}$$

and apply an RLR scheme in the way described above (the value m_0 will have to be large enough to enable one to estimate not only $\boldsymbol{\theta}$ but also $\boldsymbol{\eta}$). In this mode we will assure that even an abrupt change in $\boldsymbol{\eta}$ will not prevent us from detecting unfavorable changes in $\boldsymbol{\theta}$ reasonably fast; simplified RLR, however, could be much slower in detecting changes in $\boldsymbol{\theta}$ under such circumstances.

In situations where changes in $\boldsymbol{\eta}$ tend to be infrequent, one can chose to obtain an estimate of its current value at time T and treat it as a known quantity. In the process of estimating the current value of $\boldsymbol{\eta}$ one will typically use (explicitly or implicitly) data that extend beyond the window M_T; for example, one can use exponentially weighted averages or other filtering techniques. In situations of this kind simplified RLR schemes tend to be less vulnerable to instability in $\boldsymbol{\eta}$.

The above summary only relates to the frequentist approach to the problem of detection. In the literature one can find a number of techniques stemming from the Bayesian approach to this problem that cannot be discussed here because of the limited scope of this article (e. g., see Kenett et al. 2014; Tartakovsky and Moustakides 2010, for information and references on this topic). Situations in which data can be viewed as being generated by models with changepoints are very common in industry, especially in areas related to Quality Control. In this article we presented several methods for detection of changes in $\boldsymbol{\theta}$. These methods are based on the concept of Likelihood and Likelihood Ratio and they do not require

assumptions about the process of changes. The RLR detection schemes introduced here appear to be statistically powerful and easily designable, as the number of tuning parameters is relatively small. Their modest computational requirements enable efficient implementation on a massive scale. These schemes can be viewed as a natural generalization of the conventional Cusum technique for relatively complex data models.

A large number of issues arise in relation to any given practical case where use of such techniques is considered. How to determine "good" and "bad" process windows? How to handle the nuisance parameters? Should any transformations be applied to the data? What are the relevant sources of variability? Is there any serial correlation present and, if so, what is its origin and nature? What modifications are needed if presence of outliers cannot be ruled out? How to obtain good performance estimates? What actions to take when we are quite confident that we are into a new regime but there is not enough data to estimate its characteristics? Under what conditions should we consider Bayesian methods more suitable? In any more or less complex situation designing a robust monitoring system involves not only solid science but also a great deal of art.

Acknowledgements I would like to thank the Editor, the Referee and Dr. Sara Basson (IBM Research) for valuable help on bringing this work to a form suitable for publication.

References

Basseville, M., & Nikiforov, I. (1994). *Detection of abrupt changes—theory and application.* Englewood Cliffs, New Jersey: Prentice Hall,

Carlstein, E., Muller, H., & Siegmund, D. (Eds.) (1995). *Change-point problems*. Lecture Notes—Monograph Series, Vol. 23. Institute of Mathematical Statistics.

Garthoff, R., Golosnoy, V., & Schmid, W. (2014). Monitoring the Mean of Multivariate Financial Time Series. *Applied stochastic models in business and industry*, *30*(3), 328–340.

Hawkins, D., & Olwell, H. (1998). *Cumulative sum charts and charting for quality improvement.* New York: Springer

Kenett, R., Zacks, S., & Amberti, D. (2014). *Modern industrial statistics: with applications in R, MINITAB and JMP* 2nd edn. New York: Wiley

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamic systems. *Journal of the Royal Statistical Society*, *57*, 613–658.

Lai, T. L., & Shan, J.Z. (1999). Efficient Recursive algorithms for detection of abrupt changes in signals and control systems. *IEEE Transactions on Automatic Control*, *44*(5), 952–966.

Lorden, G. (1971). Procedures for reacting to a change in distribution. Annals of Mathematical Statistics, *42*, 1897–1908.

Nikiforov, I. V. (2001). A simple change detection scheme. *Signal Processing*, *81*, 149–172

Pignatiello, J.J., & Runger, G.C. (1990). Comparison of multivariate CUSUM charts. *Journal Quality Technology*, *22*(3), 173–186.

Reynolds, M.R., Lou, J. (2010). An Evaluation of a glr control chart for monitoring the process mean. *Journal of Quality Technology*, *42*(3), 287–310.

Stoumbos, Z.; Reynolds, M. R., JR.; Ryan, T. P., & Woodall, W. H. (2000). The state of statistical process control as we proceed into the 21st century. *Journal of American Statistical Association*, *95*, 992–998.

Sullivan, J.H. (2002). Detection of multiple change points from clustering individual observations. *Journal of Quality Technology*, *34*(4), 371–383.

Sullivan, J. H., & Woodall, W. H. (1996). A comparison of multivariate quality control charts for individual observations. *Journal of Quality Technology*, 28, 398–408.

Tartakovsky, A. G., & Moustakides, G. V. (2010). State-of-the-art in bayesian changepoint detection. *Sequential Analysis*, *29*(2), 125–145.

Telksnys, L. (Ed.). (1986). *Detection of changes in random processes.* New York: Optimization Software

Tsui, K.-L., & Woodall, W. (1993). Multivariate control charts based on loss functions. *Sequential Analysis*, *12*(1), 79–92.

Wang, S., & Reynolds, M. R. (2013). A GLR control chart for monitoring the mean vector of a multivariate normal process. *Journal of Quality Technology*, 45(1), 18–33.

Woodall, W. H. (1985). The statistical design of quality control charts. *The Statistician*, *34*(2), 155–160.

Yashchin, E. (1985). On analysis and design of CUSUM-Shewhart control schemes. *IBM Journal of Research and Development*, *29*(4), 377–391.

Yashchin, E. (1995). Likelihood ratio methods for monitoring parameters of a nested random effect model. *Journal American Statistical Association*, *90*, 729–738.

Yashchin, E. (1997). Statistical change-point models in industrial applications. *Proceeding of 2nd World Congress of Nonlinear Analysts*, *30*(7), 3997–4006.

Variance Charts for Time Series: A Comparison Study

Taras Lazariv and Wolfgang Schmid

Abstract Most of the literature on control charts is focused on the surveillance of the mean behavior of the observed process. In our contribution we are dealing with monitoring the variance of a time series, in particular monitoring the increase of the variance. The underlying process is assumed to be a time series.
In this paper we give an overview about the existing literature. In an extensive simulation study several control charts are compared with each other. The target process is assumed to be an autoregressive process of order one. In order to measure the performance of the schemes the average run length and the average delay are used. We consider charts based on the likelihood ratio approach and the generalized likelihood ratio approach, the sequential probability ratio method and the generalized sequential probability ratio procedure, the Shiryaev-Roberts procedure and a generalized Shiryaev-Roberts approach and different types of exponentially weighted moving average charts.

Keywords Control charts • CUSUM charts • EWMA charts • Generalized likelihood ratio • Sequential detection • Shiryaev-Roberts procedures • SPRT • Statistical process control • Time series • Variance charts

1 Introduction

In many applications we are interested to detect a change in a process characteristic as soon as possible. For instance, in engineering it is necessary to control a production process in order to reduce the amount of defective parts, in economics an investor wants to earn money and thus he has to react on changes on the market, in public health the outbreak of a disease should be detected fast. In most of these cases the underlying process has a complicate structure and has to be modeled by a time series approach.

T. Lazariv (✉) • W. Schmid
Department of Statistics, European University Viadrina, PO Box 1786, 15207 Frankfurt(Oder), Germany
e-mail: lazariv@europa-uni.de; schmid@europa-uni.de

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_6

Control charts for time series have been intensively discussed in the last 25 years. However, nearly all publications deal with the detection of a change in the mean of the underlying target process. In this paper we consider control charts for the variance of a time series. Since in economics the variance is used as a measure for the risk and in engineering it reflects the performance of a production process such schemes are of wide interest in practice. This motivates the need for monitoring tools, that can detect a shift in the variance of process as quickly as possible.

Exponentially Weighted Moving Average (EWMA) control charts for the variance of a time series were introduced by MacGregor and Harris (1993). Schipper and Schmid (2001) introduced several one-sided variance charts for stationary processes, however, their main focus was in the area of nonlinear time series. An overview about variance charts for univariate and multivariate time series is given in Okhrin and Schmid (2008). A lot of new control schemes obtained by using the (generalized) likelihood ratio approach, the (generalized) sequential probability ratio test and the (generalized) Shiryaev-Roberts approach were proposed by Lazariv et al. (2013). The aim of this paper is to discuss the existing control charts from a more practical point of view. To do this we focus on the important special case of an autoregressive process of order one. All charts of Sect. 3 are compared with each other by using the Average Run Length (ARL) and the (worst, maximum) average delay.

In Sect. 2 the underlying model of our paper is introduced. It is explained how the target process and the observed process are related with each other. In Sect. 3 the control schemes considered in the paper are briefly described. The results of our comparison study are given in Sect. 4. Our simulations show that for small variance changes the generalized Shiryeav-Roberts procedure proposed by Lazariv et al. (2013) provides the best results while for moderate and large changes the EWMA residual chart dominates.

2 Modelling

In the following it is assumed that at a given time point exactly one observation is available. The target process is denoted by $\{Y_t\}$. Let $\{Y_t\}$ be a stationary Gaussian process with mean μ and autocovariance function $Cov(Y_t, Y_{t+h}) = \gamma(h)$. The relationship between the target process $\{Y_t\}$ and the observed process $\{X_t\}$ is given by

$$X_t = \begin{cases} Y_t & \text{for} \quad 1 \leq t < \tau \\ \mu + \Delta(Y_t - \mu) & \text{for} \quad t \geq \tau \end{cases} \tag{1}$$

for $t \in \mathbb{Z}$ with $\Delta > 1$ and $\tau \in \mathbb{N} \cup \{\infty\}$. Both processes coincide up to time point $\tau - 1$. Then there is a change in the scale. The observed process is said to be out of control if $\tau < \infty$. Else, if $\tau = \infty$, $\{X_t\}$ is called to be in control. Note that the change in the scale does not influence the mean structure. It holds that $E(X_t) = \mu$ as well in the in-control state as in the out-of-control state. Moreover, we get that

$$Var(X_t) = \begin{cases} Var(Y_t) & \text{for} \quad 1 \leq t < \tau \\ \Delta^2 Var(Y_t) & \text{for} \quad t \geq \tau \end{cases},$$

$$Cov(X_t, X_{t+h}) = \begin{cases} \gamma_h & \text{for} \quad t < \min\{\tau, \tau - h\} \\ \Delta\gamma_h & \text{for} \quad \min\{\tau, \tau - h\} \leq t < \max\{\tau, \tau - h\} \\ \Delta^2\gamma_h & \text{for} \quad t \geq \max\{\tau, \tau - h\} \end{cases}.$$

Note that the observed process $\{X_t\}$ is not stationary in the out-of-control case.

A situation as described in (1) can be frequently found in applications. For instance, in finance the variance is the most applied measure of risk (McNeil et al. 2005). For an analyst it is of interest to rapidly detect an increase in the variance since the probability of losing money increases. A locally constant volatility model has been investigated in various studies, e.g., Hsu et al. (1974) and Chen et al. (2010).

3 Variance Charts for Time Series

In the following it is assumed without restriction that $\mu = 0$ since else Y_t can be replaced by $Y_t - \mu$. Let $\hat{Y}_t$ denote the best linear predictor of Y_t based on $Y_{t-1}, .., Y_1$. This quantity can be recursively calculated using the innovations algorithm (cf. Brockwell and Davis 1991, p.172). It holds that $\hat{Y}_t = \sum_{j=1}^{t-1} a_{tj} Y_j$ with some coefficients a_{tj}. For a Gaussian process the best linear predictor is equal to the predictor obtained by minimizing the mean-square distance. Moreover, let $v_j = E(Y_{j+1} - \hat{Y}_{j+1})^2$ denote the mean-square error and let $\hat{X}_t = \sum_{j=1}^{t-1} a_{tj} X_j$.

3.1 CUSUM-Type Charts

Lazariv et al. (2013) introduced several new CUSUM control charts for the variance of a time series. Next we present their main results.

3.1.1 A Chart Based on the Likelihood Ratio

The idea behind this chart is to apply the likelihood ratio approach to a Gaussian process with possible changes in the variance. The run length of this chart is given by

$$N_{LR}(c;\Delta) = \inf\{n \in \mathbb{N} : \max\{0, \max_{1\le i\le n}\Big(-(n-i+1)\log(\Delta^2) + \sum_{j=i}^{n}\frac{1}{v_{j-1}}\Big[2(1-\frac{1}{\Delta})(X_j-\hat{X}_j)(X_j-T_{j,i}) - (1-\frac{1}{\Delta})^2(X_j-T_{j,i})^2\Big]\Big)\} > c\}$$

where $c > 0$ and $T_{j,\tau} = \sum_{v=\tau}^{j-1} a_{jv}X_v$. Note that $T_{j,\tau} = 0$ for $j \le \tau$.

Now we want to consider the special case of a stationary AR(1) process. Thus $Y_t = \alpha Y_{t-1} + \varepsilon_t$. It is assumed that the white noise process $\{\varepsilon_t\}$ is independent distributed with mean zero and variance σ^2. Then it holds that $\hat{X}_t = \alpha X_{t-1}$ for $t \ge 2$, $\hat{X}_1 = 0$, $v_0 = \sigma^2/(1-\alpha^2)$, and $v_n = \sigma^2$ for $n \ge 1$. Using

$$K(\Delta) = \frac{\log(\Delta^2)}{1-1/\Delta^2},$$

$$A_n^+(\Delta) = \frac{(X_n-\hat{X}_n)^2}{v_{n-1}} - K(\Delta) + \max\left\{-\frac{\hat{X}_n^2}{v_{n-1}} + \frac{2/\Delta}{1+1/\Delta}\frac{X_n\hat{X}_n}{v_{n-1}}, A_{n-1}^+(\Delta)\right\}$$

for $n \ge 1$ and $A_0^+ = 0$ we obtain that

$$N_{LR}(c;\Delta) = \inf\{n \in \mathbb{N} : \max\{0, (1-1/\Delta^2)A_n^+(\Delta)\} > c\}. \tag{2}$$

Putting $\alpha = 0$ we get the well-known variance chart for independent variables (cf. Hawkins and Olwell 1998).

3.1.2 A Chart Based on the Sequential Probability Ratio Test

The chart is obtained by applying the sequential probability ratio test of Wald. Using

$$S_n(\Delta) = \sum_{j=1}^{n}\frac{(X_j-\hat{X}_j)^2}{v_{j-1}} - nK(\Delta)$$

the run length of this scheme is equal to

$$N_{SPRT}(c;\Delta) = \inf\{n \in \mathbb{N} : \max_{0\le i\le n}\{S_n(\Delta) - S_i(\Delta)\} > c\}. \tag{3}$$

Let $S_n^+(\Delta) = \max_{0 \le i \le n} \{S_n(\Delta) - S_i(\Delta)\}$. Then

$$S_n^+(\Delta) = \max\{S_{n-1}^+(\Delta) + \frac{(X_n - \hat{X}_n)^2}{v_{n-1}} - K(\Delta), 0\}$$

and thus the control statistic can be recursively calculated.

It is important to note that contrary to the i.i.d. case the LR approach and the SPRT approach lead to different control charts.

Note that this scheme is equal to the variance chart for independent variables but instead of the observations it is applied to the normalized residuals. Thus it coincides with the CUSUM residual chart for the variance.

3.1.3 A Chart Based on Shiryaev-Roberts Procedure

The Shiryaev-Roberts (SR) procedure is based on papers of Shiryaev (1963) and Roberts (1966). Contrary to the likelihood ratio approach not the maximum of the likelihood ratio over all possible positions of the change points is taken, but the maximum is replaced by the sum. This leads to the following statistic.

Lets consider

$$R_n(\Delta) = \sum_{i=1}^{n} \frac{1}{\Delta^{n-i+1}} \exp\left\{ \sum_{j=i}^{n} \frac{1}{v_{j-1}} \left((1 - \frac{1}{\Delta})(X_j - \hat{X}_j)(X_j - T_{j,i}) \right.\right.$$
$$\left.\left. -\frac{1}{2}(1 - \frac{1}{\Delta})^2 (X_j - T_{j,i})^2 \right) \right\}.$$

The run length of the SR chart is given by

$$N_{SR}(c; \Delta) = \inf\{n \in \mathbb{N} : R_n(\Delta) > c\}. \tag{4}$$

3.2 *Generalized Control Charts*

A great disadvantage of the charts presented in the last section consists in the fact that in the derivation of the charts it is assumed that the size of the change Δ is known. In practice this quantity is replaced by a value against which the practitioner wants to protect himself. The following charts do not use this prior information. Generalized control charts have been mostly discussed in literature on change point analysis (e.g., Lai 2001). Recently it has received more attention in SPC literature (Capizzi (2001), Capizzi and Masarotto (2008), and Reynolds and Lou (2012)). The following charts were introduced in Lazariv et al. (2013).

3.2.1 Generalized Likelihood Ratio Chart

In this approach the size of the change is considered as an unknown parameter and the maximum of the likelihood function is taken over Δ as well.

Using

$$\Delta_{\tau,n} = \max\{1, \frac{\dot{S}_{n,\tau} - \ddot{S}_{n,\tau} + \sqrt{(\dot{S}_{n,\tau} - \ddot{S}_{n,\tau})^2 + 4(n-\tau+1)\ddot{S}_{n,\tau}}}{2(n-\tau+1)}\},$$

$$\dot{S}_{n,\tau} = \sum_{j=\tau}^{n} \frac{(X_j - \hat{X}_j)(X_j - T_{j,\tau})}{v_{j-1}}, \ddot{S}_{n,\tau} = \sum_{j=\tau}^{n} \frac{(X_j - T_{j,\tau})^2}{v_{j-1}}$$

the run length of the GLR chart is given by

$$N_{GLR}(c) = \inf\Big\{n \in \mathbb{N} : \max_{1\le i\le n}\Big\{-(n-i+1)\log(\Delta_{i,n}) \tag{5}$$

$$-\frac{1}{2}(\frac{1}{\Delta_{i,n}} - 1)(2\dot{S}_{n,i} + (\frac{1}{\Delta_{i,n}} - 1)\ddot{S}_{n,i})\Big\} > c\Big\}. \tag{6}$$

3.2.2 Generalized SPRT Chart

In the same way as for the GLR approach the SPRT is maximized over Δ as well. Using the notation

$$h_n(x) = nh(x) = n(x - 1 - \log(x))/2 \text{ and } T_n = \sum_{i=1}^{n}(X_i - \hat{X}_i)^2/v_{i-1}.$$

the run length of this scheme is

$$N_{GSPRT}(c) = \inf\{n \in \mathbb{N} : \max_{0\le i\le n}(h_n(T_n/n)) - h_i(T_i/i)) > c\}. \tag{7}$$

3.2.3 Generalized Modified SR Chart

Extending the Shiryaev-Roberts procedure in this way we have to take the maximum over a quantity which is difficult to handle from the analytical point of view. For that reason Lazariv et al. (2013) proposed to replace the sum of the likelihood ratios by the sum of the logarithm of the likelihood ratios. In principle this means that instead of the arithmetic mean the geometric mean is maximized. This modification can be treated much easier.

Let $\dot{U}_n = \sum_{k=1}^{n} \dot{S}_{n,k}, \quad \ddot{U}_n = \sum_{k=1}^{n} \ddot{S}_{n,k}$ and $\dot{S}_{n,k}$ and $\ddot{S}_{n,k}$ as above. Let

$$g_n(\dot{U}_n, \ddot{U}_n) = -\frac{n(n+1)}{2} \log(\hat{\Delta}_n^2) + 2(1 - \frac{1}{\hat{\Delta}_n})\dot{U}_n - \frac{1}{2}(1 - \frac{1}{\hat{\Delta}_n})^2 \ddot{U}_n$$

and $\hat{\Delta}_n = \max\{1, \frac{\dot{U}_n - \ddot{U}_n + \sqrt{(\dot{U}_n - \ddot{U}_n)^2 + 2n(n+1)\ddot{U}_n}}{n(n+1)}\}$. Then the run length of the generalized modified Shiryaev-Roberts approach is

$$N_{GMSR}(c) = \inf\{n \in \mathbb{N} : g_n(\dot{U}_n, \ddot{U}_n) > c\}. \tag{8}$$

3.3 EWMA-Type Charts

EWMA charts for the variance of a time series have been discussed for a longer time. The first papers dealing with this topic seem to be MacGregor and Harris (1993) and Schipper and Schmid (2001).

3.3.1 A Chart Based on Squared Observations

If we apply the EWMA recursion to the squared observations, we obtain

$$Z_t = (1 - \lambda) Z_{t-1} + \lambda \frac{X_t^2}{\gamma_0},$$

where $t \geq 1, \lambda \in (0, 1]$. As a starting value for the EWMA recursion Z_0 we choose the in-control value of $Var(X_t)$ which is equal to γ_0. This chart was discussed by MacGregor and Harris (1993) for an $ARMA(1, 1)$ process and by Schipper and Schmid (2001) for ARMA and GARCH processes.

Let $E_1(Z_t)$ and $Var_1(Z_t)$ denote the in-control values of $E(Z_t)$ and $Var(Z_t)$, respectively. Then $E_1(Z_t) = 1$ and

$$\lim_{t \to \infty} Var_1(Z_t) = \frac{1}{2 - \lambda}(2 + \lambda + 4\lambda \sum_{i=1}^{\infty} \rho_i^2 (1 - \lambda)^i)$$

with $\rho_i = \gamma_i / \gamma_0$ as shown in Schipper and Schmid (2001). For an AR(1) process we get that

$$\lim_{t \to \infty} Var_1(Z_t) = \frac{1}{2 - \lambda}(2 + \lambda + 4\lambda\alpha^2 \frac{1 - \lambda}{1 - (1 - \lambda)\alpha^2}).$$

Thus the stopping rule of the chart is

$$N_S(c;\lambda) = \inf\{t \in \mathbb{N} : \frac{Z_t - E_1(Z_t)}{\sqrt{\lim\limits_{t\to\infty} Var_1(Z_t)}} > c\}. \tag{9}$$

3.3.2 A Chart Based on the Logarithm of Squared Observations

In this case the EWMA recursion looks like

$$Z_t = (1-\lambda)Z_{t-1} + \lambda \ln(X_t^2/\gamma_0)$$

for $t \geq 1$ and $Z_0 = E(\ln(Y_t^2/\gamma_0)) = \gamma_0^* = -ln(2) - \gamma$ where γ stands for the Euler-Mascheroni constant, i.e. $\gamma \approx 0.57721$. Crowder and Hamilton (1992) discussed this chart for independent variables.

The run length is given by

$$N_{LS}(c;\lambda) = \inf\{t \in \mathbb{N} : \frac{Z_t - \gamma_0^*}{\sqrt{\lim\limits_{t\to\infty} Var_1(Z_t)}} > c\}. \tag{10}$$

3.3.3 The EWMA Residual Chart

The EWMA statistics for the variance is as follows:

$$Z_t = (1-\lambda)Z_{t-1} + \lambda \frac{(X_t - \hat{X}_t)^2}{v_{t-1}}$$

for $t \geq 1, \lambda \in (0,1]$. The starting value is $Z_0 = 1$. The process is concluded to be out-of-control at time t if $Z_t > c$

$$N_R(c;\lambda) = \inf\{t \in \mathbb{N} : Z_t > c\}. \tag{11}$$

Note that for an AR(1) process it holds that for $t \geq 2$

$$\frac{X_t - \hat{X}_t}{\sqrt{v_{t-1}}} = \begin{cases} \varepsilon_t/\sigma & \text{for } t < \tau \\ (\varepsilon_\tau + (\Delta - 1)Y_\tau)/\sigma & \text{for } t = \tau \\ \Delta\varepsilon_t/\sigma & \text{for } t > \tau \end{cases}$$

and

$$\frac{X_1 - \hat{X}_1}{\sqrt{v_0}} = \begin{cases} \Delta Y_1/\sqrt{\gamma_0} & \text{for } \tau = 1 \\ Y_1/\sqrt{\gamma_0} & \text{for } \tau \geq 2 \end{cases}.$$

4 Comparison Study

In this section we want to give recommendations which of the control charts of Sect. 3 should be used in a specific situation. Because no explicit formulae for the performance criteria are known we have estimated these quantities within a simulation study. All charts were calibrated such that the in-control ARL is always the same. We fixed the in-control ARL equal to 500. The out-of-control behavior of all schemes is analyzed using the obtained control limits, here c. In each case the ARL was determined within a simulation study based on 10^6 repetitions. The only exception is the GLR chart. Since no recursive presentation of the control statistic is known its calculation turns out to be more complicate. For this scheme we made use of 10^5 repetitions. To evaluate the performance of the charts the ARL $E_{\tau=1,\Delta}(N(c))$ and the average delay $E_{\tau,\Delta}(N(c) - \tau + 1 | N(c) \geq \tau)$ are taken. The reference value Δ^* is chosen within the set $\{1.10, 1.20, 1.30, 1.40, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00\}$. For EWMA control charts the parameter λ is taking values within $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$.

In our comparison study the target process is assumed to be an AR(1) process with standard normally distributed white noise. The coefficient of the process α takes values from $\{-0.9, -0.8, -0.7, \ldots, 0.7, 0.8, 0.9\}$. Because the performance criteria of the charts behave symmetric with respect to α we only considered nonnegative values. Note that some charts depend on the unknown value Δ. In practice this quantity is replaced by a fixed value which the practitioner chooses a priori. This can be a problem, however, sometimes some information is available about the size of the expected change.

In Tables 1, 2 and 3 the out-of-control ARLs of the considered charts are given. In each row and each column the ARLs of nine control charts are given, above the LR chart (Sect. 3.1.1, cf. (2)), followed by the SPRT chart (Sect. 3.1.2, cf. (3)), the Shiryaev-Roberts chart (Sect. 3.1.3, cf. (4)), the GLR chart (Sect. 3.2.1, cf. (5)), the GSPRT chart (Sect. 3.2.2, cf. (7)), and the GMSR chart (Sect. 3.2.3, cf. (8)), and three EWMA chart, namely the EWMA chart for squared observations (Sect. 3.3.1, cf. (9)), the EWMA chart for the logarithm of squared observations (Sect. 3.3.2, cf. (10)), and the EWMA residual chart (Sect. 3.3.3, cf. (11)). The first three charts and the EWMA charts depend on a reference value. For these charts the smallest out-of-control ARL over all Δ and λ is listed, respectively. In parenthesis the value of the parameter is given where the minimum is attained. The other three charts are generalized schemes and do not depend on a reference value. The results for the best charts are given in bold. Because our results are based on simulations they are subject to some random error which, however, due to the large amount of repetitions, is small. For that reason we have printed the ARLs of the charts that do not deviate from the smallest ARL by more than 2 % in bold as well. This is also the reason why in the tables there seem to be sometimes jumps for the best smoothing parameters and reference values but this is due to the fact that sometimes the out-of-control

Table 1 Out-of-control ARLs of several variance charts (LR chart, SPRT chart, Shiryaev-Roberts chart, GLR chart, GSPRT chart and Generalized Modified Shiryaev-Roberts chart, EWMA chart for squared observations, log-EWMA chart and EWMA residual chart)

Δ/α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.10	116.76(1.10)	116.84(1.10)	116.87(1.10)	117.07(1.10)	117.16(1.10)	117.11(1.10)	117.49(1.10)	117.75(1.10)	118.10(1.10)	120.28(1.10)
	116.89(1.10)	116.88(1.10)	116.96(1.10)	117.06(1.10)	117.08(1.10)	117.04(1.10)	117.22(1.10)	117.34(1.10)	117.46(1.10)	117.66(1.10)
	127.61(1.20)	127.67(1.20)	127.55(1.20)	127.42(1.20)	127.59(1.20)	127.51(1.20)	127.64(1.20)	127.71(1.20)	127.51(1.20)	127.60(1.20)
	121.65	121.66	121.75	121.83	121.97	122.13	122.47	122.54	122.95	123.76
	351.98	353.76	353.46	354.41	355.07	355.05	358.02	356.63	358.36	359.05
	79.06	**79.04**	**79.08**	**79.12**	**79.16**	**79.24**	**79.29**	**79.44**	**79.80**	**80.51**
	133.91(0.1)	136.11(0.1)	142.59(0.1)	152.84(0.1)	165.73(0.1)	181.27(0.1)	198.91(0.1)	218.61(0.1)	232.56(1.0)	257.91(1.0)
	168.84(0.1)	170.67(0.1)	175.09(0.1)	182.86(0.1)	195.19(0.1)	209.21(1.0)	213.99(1.0)	221.06(1.0)	233.33(1.0)	257.74(1.0)
	134.02(0.1)	134.02(0.1)	134.11(0.1)	134.14(0.1)	134.39(0.1)	134.29(0.1)	134.09(0.1)	134.33(0.1)	134.49(0.1)	134.58(0.1)
1.20	54.20(1.20)	54.09(1.20)	54.18(1.20)	54.35(1.20)	54.40(1.20)	54.62(1.20)	54.94(1.20)	55.11(1.20)	55.54(1.20)	56.80(1.20)
	54.15(1.20)	54.15(1.20)	54.23(1.20)	54.32(1.20)	54.41(1.20)	54.53(1.20)	54.68(1.20)	54.83(1.20)	54.95(1.20)	55.06(1.20)
	58.44(1.40)	58.43(1.40)	58.57(1.40)	58.60(1.40)	58.50(1.40)	58.57(1.40)	58.68(1.40)	58.79(1.40)	58.80(1.40)	58.84(1.40)
	61.43	61.44	61.49	61.59	61.72	61.89	61.90	62.38	62.73	63.51
	99.68	99.49	99.33	99.43	99.24	99.69	100.77	100.34	101.91	102.92
	45.70	**45.69**	**45.71**	**45.75**	**45.82**	**45.88**	**45.99**	**46.19**	**46.62**	**47.22**
	57.14(0.1)	58.31(0.1)	61.74(0.1)	67.66(0.1)	75.79(0.1)	86.52(0.1)	100.07(0.1)	116.51(0.1)	130.41(1.0)	155.24(1.0)
	82.32(0.1)	83.21(0.1)	86.29(0.1)	91.56(0.1)	100.01(0.1)	107.29(1.0)	112.01(1.0)	118.59(1.0)	130.42(1.0)	155.23(1.0)
	57.18(0.1)	57.16(0.1)	57.21(0.1)	57.24(0.1)	57.37(0.1)	57.32(0.1)	57.43(0.1)	57.61(0.1)	57.69(0.1)	57.79(0.1)

1.30	**32.32**(1.30)	**32.30**(1.30)	**32.39**(1.30)	**32.50**(1.30)	**32.66**(1.30)	**32.82**(1.30)	**33.02**(1.30)	33.24(1.30)	33.70(1.30)	34.58(1.30)
	32.32(1.30)	**32.32**(1.30)	**32.39**(1.30)	**32.47**(1.30)	**32.57**(1.30)	**32.78**(1.30)	**32.94**(1.30)	**33.11**(1.30)	**33.24**(1.30)	**33.27**(1.30)
	34.98(1.60)	34.98(1.60)	35.00(1.60)	35.04(1.60)	35.08(1.60)	35.16(1.60)	35.26(1.60)	35.30(1.60)	35.39(1.60)	35.55(1.60)
	39.11	39.11	39.18	39.28	39.42	39.59	39.81	40.04	40.39	41.11
	48.97	49.02	49.08	49.19	49.14	49.24	49.87	49.95	50.69	51.08
	32.54	**32.55**	**32.57**	**32.61**	**32.66**	**32.74**	**32.83**	**33.01**	**33.39**	33.96
	32.09(0.1)	32.77(0.1)	34.71(0.1)	38.01(0.1)	43.01(0.1)	49.88(0.1)	59.09(0.1)	71.15(0.1)	82.56(1.0)	103.86(1.0)
	49.91(0.1)	50.63(0.1)	52.48(0.1)	55.97(0.1)	60.99(1.0)	63.49(1.0)	67.27(1.0)	72.78(1.0)	82.81(1.0)	103.76(1.0)
	32.07(0.1)	**32.05**(0.1)	**32.18**(0.1)	**32.19**(0.1)	**32.26**(0.1)	**32.32**(0.1)	**32.46**(0.1)	**32.55**(0.1)	**32.72**(0.1)	**32.82**(0.1)

Note: ARLs/AD of all charts are written in bold which for a fixed value of Δ deviate from the smallest out-of-control ARL/AD by only 2 %

Table 2 Out-of-control ARLs of several variance charts (LR chart, SPRT chart, Shiryaev-Roberts chart, GLR chart, GSPRT chart and Generalized Modified Shiryaev-Roberts chart, EWMA chart for squared observations, log-EWMA chart and EWMA residual chart)

Δ/α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.40	22.04(1.40)	22.04(1.40)	22.08(1.40)	22.21(1.40)	22.35(1.40)	22.51(1.40)	22.72(1.40)	22.99(1.40)	23.27(1.40)	23.94(1.40)
	22.03(1.40)	22.03(1.40)	22.10(1.40)	22.18(1.40)	22.29(1.40)	22.43(1.40)	22.60(1.40)	22.75(1.40)	22.90(1.40)	22.99(1.40)
	23.79(1.80)	23.79(1.80)	23.80(1.80)	23.87(1.80)	23.92(1.80)	24.01(1.80)	24.16(1.80)	24.24(1.80)	24.36(1.80)	24.59(1.80)
	27.97	27.98	28.05	28.14	28.26	28.43	28.60	28.89	29.23	29.84
	30.44	30.48	30.60	30.60	30.64	30.76	31.03	31.04	31.54	31.88
	25.42	25.43	25.45	25.49	25.54	25.62	25.74	25.90	26.20	26.72
	21.19(0.1)	**21.58**(0.1)	22.79(0.1)	24.92(0.1)	28.15(0.1)	32.75(0.1)	39.21(0.1)	48.03(0.1)	57.17(1.0)	74.84(1.0)
	32.97(0.6)	33.91(0.5)	36.31(0.2)	38.11(1.0)	39.69(1.0)	41.66(1.0)	44.67(1.0)	49.28(1.0)	57.24(1.0)	74.76(1.0)
	21.21(0.1)	**21.21**(0.1)	**21.26**(0.1)	**21.32**(0.1)	**21.41**(0.1)	**21.44**(0.1)	**21.57**(0.1)	**21.67**(0.1)	**21.82**(0.1)	**21.96**(0.1)
1.50	16.30(1.50)	16.32(1.50)	16.40(1.50)	16.51(1.50)	16.61(1.50)	16.78(1.50)	16.97(1.50)	17.20(1.50)	17.48(1.50)	17.95(1.50)
	16.31(1.50)	16.33(1.50)	16.37(1.50)	16.45(1.50)	16.57(1.50)	16.72(1.50)	16.88(1.50)	17.04(1.50)	17.18(1.50)	17.29(1.50)
	17.52(2.00)	17.53(2.00)	17.56(2.00)	17.65(2.00)	17.74(2.00)	17.83(2.00)	17.94(1.80)	18.06(1.80)	18.18(1.80)	18.31(1.80)
	21.39	21.39	21.46	21.55	21.67	21.83	21.99	22.25	22.57	23.13
	21.52	21.54	21.49	21.46	21.60	21.59	21.89	21.77	22.31	22.58
	20.92	20.92	20.95	20.99	21.04	21.12	21.22	21.38	21.67	22.14
	15.52(0.1)	**15.76**(0.1)	16.59(0.1)	18.03(0.1)	20.23(0.1)	23.47(0.1)	28.11(0.1)	34.81(0.1)	42.17(1.0)	56.99(1.0)
	23.17(0.7)	23.82(0.8)	25.26(0.9)	26.68(1.0)	27.84(1.0)	29.54(1.0)	31.93(1.0)	35.68(1.0)	42.33(1.0)	56.98(1.0)
	15.52(0.1)	**15.51**(0.1)	**15.54**(0.1)	**15.58**(0.1)	**15.67**(0.1)	**15.76**(0.1)	**15.87**(0.1)	**15.98**(0.1)	**16.16**(0.1)	**16.29**(0.1)

1.75	9.50(1.75)	9.53(1.75)	9.57(1.75)	9.65(1.75)	9.77(1.75)	9.91(1.75)	10.09(1.75)	10.28(1.75)	10.53(1.75)	10.79(1.75)
	9.51(1.75)	9.52(1.75)	9.56(1.75)	9.64(1.75)	9.74(1.75)	9.87(1.75)	10.02(1.75)	10.20(1.75)	10.35(1.75)	10.47(1.75)
	10.08(2.40)	10.09(2.40)	10.14(2.40)	10.20(2.40)	10.29(2.40)	10.43(2.40)	10.55(2.40)	10.70(2.20)	10.85(2.20)	11.03(2.20)
	13.05	13.07	13.12	13.19	13.31	13.45	13.67	13.85	14.12	14.55
	11.98	11.94	11.97	11.99	12.02	12.09	12.22	12.16	12.39	12.57
	14.61	14.61	14.63	14.67	14.73	14.80	14.91	15.06	15.31	15.69
	9.07(0.1)	**9.21**(0.1)	9.59(0.1)	10.31(0.1)	11.39(0.1)	13.05(0.1)	15.45(0.1)	19.08(0.1)	23.82(1.0)	33.74(1.0)
	12.09(0.8)	12.33(0.9)	12.95(0.9)	13.73(1.0)	14.47(1.0)	15.53(1.0)	17.08(1.0)	19.56(1.0)	23.88(1.0)	33.73(1.0)
	9.07(0.1)	**9.08**(0.1)	**9.12**(0.1)	**9.17**(0.1)	**9.24**(0.1)	**9.32**(0.1)	**9.43**(0.1)	**9.58**(0.1)	**9.72**(0.1)	**9.89**(0.1)
2.00	6.60(2.00)	6.61(2.00)	6.64(2.00)	6.72(2.00)	6.82(2.00)	6.95(2.00)	7.10(2.00)	7.30(2.00)	7.51(2.00)	7.71(2.00)
	6.59(2.00)	6.61(2.00)	6.65(2.00)	6.71(2.00)	6.80(2.00)	6.92(2.00)	7.07(2.00)	7.24(2.00)	7.42(2.00)	7.55(2.00)
	6.92(3.00)	6.93(3.00)	6.97(2.80)	7.04(2.80)	7.13(2.80)	7.25(2.80)	7.39(2.80)	7.57(2.80)	7.72(2.60)	7.92(2.40)
	9.19	9.21	9.25	9.33	9.44	9.58	9.75	9.94	10.19	10.54
	8.38	8.37	8.31	8.35	8.37	8.37	8.45	8.46	8.58	8.73
	11.29	11.30	11.32	11.36	11.42	11.49	11.60	11.74	11.96	12.28
	6.41(0.1)	**6.48**(0.1)	6.73(0.1)	7.15(0.1)	7.82(0.1)	8.83(0.1)	10.31(0.1)	12.54(0.1)	15.78(0.1)	22.99(1.0)
	7.76(0.9)	7.88(0.9)	8.29(0.9)	8.73(1.0)	9.26(1.0)	9.99(1.0)	11.08(1.0)	12.79(1.0)	15.85(1.0)	22.97(1.0)
	6.39(0.1)	**6.41**(0.1)	**6.44**(0.1)	**6.49**(0.1)	**6.56**(0.1)	**6.64**(0.1)	**6.75**(0.1)	**6.89**(0.1)	**7.03**(0.1)	**7.20**(0.1)

Note: ARLs/AD of all charts are written in bold which for a fixed value of Δ deviate from the smallest out-of-control ARL/AD by only 2 %

Table 3 Out-of-control ARLs of several variance charts (LR chart, SPRT chart, Shiryaev-Roberts chart, GLR-chart, GSPRT chart and Generalized Modified Shiryaev-Roberts chart, EWMA chart for squared observations, log-EWMA chart and EWMA residual chart)

Δ/α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2.25	5.05(2.25)	5.06(2.25)	5.09(2.25)	5.16(2.25)	5.24(2.25)	5.37(2.25)	5.51(2.25)	5.70(2.25)	5.90(2.25)	6.09(2.25)
	5.04(2.25)	5.06(2.25)	5.09(2.25)	5.15(2.25)	5.23(2.25)	5.34(2.25)	5.47(2.25)	5.64(2.25)	5.83(2.25)	6.00(2.25)
	5.28(3.00)	5.29(3.00)	5.31(3.00)	5.37(3.00)	5.46(3.00)	5.57(3.00)	5.71(3.00)	5.88(2.80)	6.07(2.80)	6.26(2.60)
	7.07	7.08	7.12	7.18	7.28	7.41	7.56	7.76	7.99	8.31
	6.52	6.52	6.52	6.54	6.55	6.53	6.56	6.57	6.67	6.76
	9.24	9.25	9.27	9.32	9.37	9.45	9.55	9.69	9.88	10.15
	4.94(0.2)	**5.01**(0.2)	5.21(0.1)	5.51(0.1)	5.96(0.1)	6.65(0.1)	7.66(0.1)	9.17(0.1)	11.24(0.1)	17.09(1.0)
	5.66(0.9)	5.75(0.9)	6.01(0.9)	6.31(1.0)	6.68(1.0)	7.22(1.0)	8.03(1.0)	9.31(1.0)	11.61(1.0)	17.06(1.0)
	4.94(0.2)	**4.94**(0.2)	**4.96**(0.2)	**5.02**(0.2)	**5.08**(0.2)	**5.17**(0.2)	**5.27**(0.2)	**5.41**(0.2)	**5.57**(0.2)	**5.76**(0.2)
2.50	**4.12**(2.50)	**4.13**(2.50)	4.17(2.50)	4.21(2.50)	4.29(2.50)	4.40(2.50)	4.53(2.50)	4.71(2.50)	4.92(2.50)	5.12(2.50)
	4.12(2.50)	**4.13**(2.50)	**4.16**(2.50)	4.21(2.50)	4.27(2.50)	4.38(2.50)	4.51(2.50)	4.67(2.50)	4.87(2.50)	5.06(2.50)
	4.31(3.00)	4.32(3.00)	4.34(3.00)	4.39(3.00)	4.47(3.00)	4.57(3.00)	4.70(3.00)	4.86(3.00)	5.06(3.00)	5.26(2.80)
	5.72	5.74	5.77	5.84	5.92	6.04	6.20	6.38	6.60	6.88
	5.44	5.45	5.44	5.45	5.47	5.46	5.49	5.49	5.55	5.63
	7.84	7.85	7.87	7.91	7.97	8.05	8.15	8.29	8.48	8.70
	4.04(0.2)	4.09(0.2)	4.25(0.2)	4.51(0.1)	4.85(0.1)	5.35(0.1)	6.11(0.1)	7.18(0.1)	8.57(0.1)	13.38(1.0)
	4.48(0.9)	4.55(0.9)	4.73(0.9)	4.93(1.0)	5.21(1.0)	5.62(1.0)	6.26(1.0)	7.25(1.0)	9.07(1.0)	13.38(1.0)
	4.04(0.2)	**4.05**(0.2)	**4.08**(0.2)	**4.12**(0.2)	**4.17**(0.2)	**4.26**(0.2)	**4.36**(0.2)	**4.49**(0.2)	**4.65**(0.2)	**4.85**(0.2)

2.75	**3.52**(2.75)	**3.52**(2.75)	**3.54**(2.75)	**3.59**(2.75)	3.66(2.75)	3.75(2.75)	3.88(2.75)	4.05(2.75)	4.25(2.75)	4.47(2.75)
	3.51(2.75)	**3.52**(2.75)	**3.55**(2.75)	**3.59**(2.75)	3.66(2.75)	3.74(2.75)	3.86(2.75)	4.02(2.75)	4.22(2.75)	4.43(2.75)
	3.68(3.00)	3.68(3.00)	3.71(3.00)	3.75(3.00)	3.82(3.00)	3.91(3.00)	4.03(3.00)	4.19(3.00)	4.38(3.00)	4.60(3.00)
	4.82	4.83	4.87	4.93	5.01	5.12	5.26	5.45	5.66	5.93
	4.76	4.76	4.76	4.76	4.75	4.77	4.79	4.79	4.83	4.89
	6.83	6.83	6.86	6.90	6.96	7.04	7.14	7.27	7.44	7.65
	3.46(0.2)	3.51(0.2)	3.63(0.2)	3.84(0.2)	4.12(0.1)	4.51(0.1)	5.08(0.1)	5.89(0.1)	6.85(0.1)	10.94(1.0)
	3.74(0.9)	3.78(0.9)	3.94(0.9)	4.06(1.0)	4.29(1.0)	4.62(1.0)	5.13(1.0)	5.93(1.0)	7.39(1.0)	10.94(1.0)
	3.46(0.2)	**3.47**(0.2)	**3.49**(0.2)	**3.53**(0.2)	**3.58**(0.2)	**3.65**(0.2)	**3.75**(0.2)	**3.89**(0.2)	**4.05**(0.2)	**4.25**(0.2)
3.00	**3.09**(3.00)	**3.10**(3.00)	**3.12**(3.00)	**3.16**(3.00)	**3.22**(3.00)	3.30(3.00)	3.42(3.00)	3.57(3.00)	3.77(3.00)	4.01(3.00)
	3.09(3.00)	**3.09**(3.00)	**3.12**(3.00)	**3.15**(3.00)	**3.22**(3.00)	3.30(3.00)	3.41(3.00)	3.56(3.00)	3.76(3.00)	3.99(3.00)
	3.23(3.00)	3.24(3.00)	3.26(3.00)	3.30(3.00)	3.37(3.00)	3.45(3.00)	3.56(3.00)	3.72(3.00)	3.91(3.00)	4.13(3.00)
	4.19	4.21	4.23	4.28	4.36	4.46	4.60	4.78	4.99	5.25
	4.29	4.29	4.29	4.29	4.29	4.30	4.30	4.30	4.34	4.39
	6.06	6.07	6.10	6.14	6.20	6.28	6.38	6.49	6.66	6.85
	3.05(0.3)	**3.09**(0.3)	3.18(0.2)	3.36(0.2)	3.61(0.1)	3.92(0.1)	4.34(0.1)	5.01(0.1)	5.69(0.1)	9.18(1.0)
	3.25(0.9)	3.28(0.9)	3.38(1.0)	3.49(1.0)	3.67(1.0)	3.94(1.0)	4.34(1.0)	5.01(1.0)	6.22(1.0)	9.17(1.0)
	3.05(0.3)	**3.06**(0.3)	**3.09**(0.3)	**3.12**(0.3)	**3.17**(0.3)	**3.23**(0.3)	**3.33**(0.3)	**3.46**(0.3)	**3.62**(0.3)	**3.83**(0.3)

Note: ARLs/AD of all charts are written in bold which for a fixed value of Δ deviate from the smallest out-of-control ARL/AD by only 2 %

ARLs does not change a lot with the parameter and that additionally we still have a small random error in the simulations.

The tables show that the smallest ARLs for the LR and the SPRT charts are obtained if the reference value is chosen equal to the true value of the change. For the SR chart, however, the best choice is greater or equal to the true value. For the EWMA charts the optimal value of λ is increasing with Δ if α is fixed. For the EWMA residual chart the best choice does not depend on α and it is very slowly increasing with Δ. For the EWMA chart for squared observations we observe the same behavior provided that α is smaller or equal 0.7. For larger values of α the best choice is $\lambda = 1.0$. This shows that these charts react very sensible on the choice of λ if the coefficient of the AR(1) process is large. A similar behavior can be observed for the EWMA chart for the logarithm of the squared observations. For this chart, however, the dependence on α is much stronger and therefore it is difficult to make general recommendations about the best choice of the smoothing parameter.

For a fixed value of Δ the out-of-control ARLs of the charts only slightly change with α. The exceptions are the EWMA charts for squared observations and the EWMA charts for the logarithm of the squared observations. For these two schemes the out-of-control ARLs increase with α and the out-of-control ARL for $\alpha = 0.9$ (strong correlation) may be more than three times larger than the value for $\alpha = 0.0$ (no correlation).

Next we want to compare the charts with each other. For small changes in the variance, i.e. $\Delta \leq 1.3$, the GMSR chart provides the smallest out-of-control ARL. Note that it is even better than all charts with a reference parameter. For changes of medium and large size, i.e. $\Delta > 1.3$ the EWMA residual chart turns out to be the best scheme. For $1.3 < \Delta \leq 2.0$ the smallest out-of-control ARL is obtained for $\lambda = 0.1$ while for larger changes the smoothing parameter should be chosen larger. It has even turned out that the EWMA residual chart behaves quite robust with respect to the choice of the smoothing parameter. Choosing $\Delta = 2.0$ and $\Delta = 3.0$ even for the worst choice of the smoothing parameter the EWMA residual chart provides a smaller out-of-control ARL than the best generalized scheme.

For the calculation of the ARL it is assumed that the change already happened at the beginning, i.e. that $\tau = 1$. This is a great disadvantage. For that reason we analyzed the schemes with respect to another performance criteria, the average delay. In Table 4 it is assumed that the change arises up to 50th observation ($1 \leq \tau \leq 50$). We focus on the changes $\Delta = 1.3$ and $\Delta = 2.0$. The values of Δ and λ refer to the optimal choice of the reference value as given in Tables 1, 2 and 3. In the table we give the ARL, the worst average delay for $1 \leq \tau \leq 50$, and the average delay for $\tau = 50$. The coefficient of the AR(1) process is chosen to be equal to $\alpha = 0.4$. These results are again based on 10^6 repetitions.

Except the GSPRT chart the worst average delay of all other charts is always equal to the average delay. If τ increases the average delay is decreasing and it does not change a lot for $\tau \geq 20$. The situation is different for the GSPRT chart where the worst average delay is attained at $\tau = 50$ and its minimum ARL is observed for a small value of τ. From this table we can draw similar conclusions as from Tables 1, 2 and 3. The GMSR chart should be used for the detection of

Table 4 ARL (above), worst average delay for $1 \leq \tau \leq 50$ (middle), and the value of the average delay at position $\tau = 50$ (below) for the LR, the SPRT, the SR, and the EWMA residual chart for optimal reference parameter, and the GLR, the GSPRT chart, and the GMSR chart ($\alpha = 0.4$)

	LR	SPRT	SR	ERC	GLR	GSPRT	GMSR
$\Delta = 1.3$	**32.52**	**32.59**	35.09	**32.25**	39.40	49.18	**32.68**
	32.52	**32.59**	35.09	**32.25**	39.40	73.46	**32.68**
	29.85	29.85	30.70	31.97	33.43	73.46	**21.91**
$\Delta = 2.0$	6.78	6.79	7.14	**6.56**	9.42	8.38	11.41
	6.78	6.79	7.14	**6.56**	9.42	14.22	11.41
	6.42	6.41	6.48	6.43	8.16	14.22	**6. 18**

Note: ARLs/AD of all charts are written in bold which for a fixed value of Δ deviate from the smallest out-of-control ARL/AD by only 2 %

small changes while the EWMA residual chart dominates for moderate and larger changes. If the change is expected not to arise at the beginning, then the best of all considered schemes is the GMSR chart.

5 Summary

In this paper we compare several variance charts for Gaussian processes with each other. Our aim is to detect an increase in the variance. In order to measure the performance of the charts the ARL and the (worst, maximum) average delay are used. Besides EWMA and CUSUM type charts we consider generalized schemes as well. Such type of variance charts were proposed by Lazariv et al. (2013). They have the great advantage that they do not depend on an additional parameter whose choice is frequently unclear in applications.

For all schemes the performance criteria are estimated within an extensive simulation study. The target process is assumed to be an AR(1) process. Our results show that for small changes ($\Delta \leq 1.3$) the generalized Shiryaev-Roberts scheme of Lazariv et al. (2013) has the smallest out-of-control ARL while for moderate and larger changes the EWMA residual scheme behaves quite well. This scheme also behaves quite robust with respect to the choice of the smoothing parameter. Using the worst average delay we get the same ranking. For most schemes the worst average delay is obtained for $\tau = 1$ which means that the change happened at the first position. If we know that the change does not arise at the beginning, then the maximum average delay can be applied. In the present case its smallest value is attained for the generalized Shiryaev-Roberts scheme introduced in Lazariv et al. (2013).

References

Brockwell, P., & Davis, R. (1991). *Time Series: Theory and Methods.* New York: Springer.

Capizzi, G. (2001). Design of change detection algorithms based on the generalized likelihood ratio test. *Environmetrics*, *12*, 749–756.

Capizzi, G., & Masarotto, G. (2008). Practical design of generalized likelihood ratio control charts for autocorrelated data. *Technometrics*, *50*(3), 357–370.

Chen, Y., Härdle, W. K., & Pigorsch, U. (2010). Localized realized volatility modelling. *Journal of American Statistical Association*, *105*, 1376–1393.

Crowder, S. V., & Hamilton, M. D. (1992). EWMA charts for monitoring a process standard deviation. *Journal of Quality Technology*, *24*, 12–21.

Hawkins, D., & Olwell, D. (1998). *Cumulative Sum Charts and Charting for Quality Improvement.* New York: Springer.

Hsu, D., Miller, R., & Wichern, D. (1974). On the stable Paretian behaviour of stock market prices. *Journal of American Statistical Association*, *69*, 108–113.

Lai, T. L. (2001). Sequential analysis: some classical problems and new challenges (with discussion). *Statistica Sinica*, *11*, 303–408.

Lazariv, T., Schmid, W., & Zabolotska, S. (2013). On control charts for monitoring the variance of a time series. *Journal of Statistical Planning and Inference*, *143*, 1512–1526.

MacGregor, J. F., & Harris, T. J. (1993). The exponentially weighted moving variance. *Journal of Quality Technology*, *25*, 106–118.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management.* Princeton: Princeton University Press.

Okhrin, Y., & Schmid, W. (2008). Surveillance of univariate and multivariate linear time series. In M. Frisén (Ed.), *Financial Surveillance* (pp. 115–152). Wiley.

Reynolds, M. R., & Lou, J. (2012). A GLR control chart for monitoring the process variance. In H.J. Lenz, W. Schmid, P.-T. Wilrich (Eds.), *Frontiers in Statistical Process Control*, Physica, Vol 10, pp. 3–18.

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics*, *8*, 411–430.

Schipper, S., & Schmid, W. (2001). Sequential methods for detecting changes in the variance of economic time series. *Sequential Analysis*, *20*, 235–262.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, *8*, 22–46.

On ARL-Unbiased Control Charts

Sven Knoth and Manuel Cabral Morais

Abstract Manufacturing processes are usually monitored by making use of control charts for variables or attributes. Controlling both increases and decreases in a parameter, by using a control statistic with an asymmetrical distribution, frequently leads to an *ARL-biased* chart, in the sense that some out-of-control average run length (ARL) values are larger than the in-control ARL, i.e., it takes longer to detect some shifts in the parameter than to trigger a false alarm.
In this paper, we are going to:

- explore what Pignatiello et al. (4th Industrial Engineering Research Conference, 1995) and Acosta-Mejía et al. (J Qual Technol 32:89–102, 2000) aptly called an *ARL-unbiased* chart;
- provide instructive illustrations of *ARL-(un)biased* charts of the Shewhart-, exponentially weighted moving average (EWMA)-, and cumulative sum (CUSUM)-type;
- relate *ARL-unbiased* Shewhart charts with the notions of unbiased and uniformly most powerful unbiased (UMPU) tests;
- briefly discuss the design of EWMA charts not based on *ARL(-unbiasedness)*.

Keywords Power function • Run length • Statistical process control

1 Introduction

In 1924 Walter A. Shewhart (1891–1967) prepared a memorandum only about a page in length; in a third of that page there was a simple diagram which we would all recognize today as a control chart. It is essentially a graphical device used to

S. Knoth (✉)
Institute of Mathematics and Statistics, Department of Economics and Social Sciences, Helmut Schmidt University Hamburg, Postfach 700822, 22008 Hamburg, Germany
e-mail: Sven.Knoth@hsu-hh.de

M.C. Morais
CEMAT & Department of Mathematics, Instituto Superior Técnico, Av. Rovisco Pais, Lisbon, Portugal
e-mail: maj@math.ist.utl.pt

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_7

monitor a measurable characteristic X of a process with the purpose of establishing whether the process is operating within its limits of expected variation (see Nelson 1982, p. 176).

To detect increases or decreases in the process mean μ, Shewhart suggested the use of the sample mean and the control limits $\mu_0 \pm 3\sigma_0/\sqrt{n}$, where μ_0 and σ_0^2 are the target values of the process mean μ and variance σ^2, and n the sample size. The resulting SPC tool is usually termed a $\bar{X}$-chart with 3σ limits and has very interesting properties under certain assumptions, as shown in the next paragraphs, inspired by Morais (2002, pp. 19–20).

Assume that: the quality characteristic X is normally distributed; the process output is i.i.d.; it is possible to specify the target values μ_0 and σ_0^2 of the process mean and variance so that an analysis of past data is not required; the process mean relates to these target values as follows, $\mu = \mu_0 + \delta\sigma_0/\sqrt{n}$, $\delta \in (-\infty, +\infty)$; the process variance remains constant at the target level σ_0^2.

Now, consider a $\bar{X}$-chart with control limits $\mu_0 \pm \gamma\sigma_0/\sqrt{n}$, where γ is a positive constant chosen in such way that the in-control average run length (ARL) of the $\bar{X}$-chart, say $ARL_\mu(0)$, takes a fixed and large value. Recall that the run length (RL) of this control chart, say $RL_\mu(\delta)$, has a geometric distribution with parameter

$$\xi_\mu(\delta) = 1 - [\Phi(\gamma - \delta) - \Phi(-\gamma - \delta)], \tag{1}$$

which is a continuous and an even function of δ. The first derivative of $\xi_\mu(\delta)$ is an odd function equal to

$$\sqrt{2/\pi}\, e^{-(\gamma^2+\delta^2)/2} \times \sinh(\gamma\delta), \tag{2}$$

which is nonpositive for $\delta \in (-\infty, 0]$ and nonnegative for $\delta \in [0, +\infty)$. Therefore we can assert that $\xi_\mu(\delta)$ takes its minimum value at the origin and increases with $|\delta|$, i.e., $\xi_\mu(\delta)$ monotonically increases, as μ tends away from its target value. In other words, the associated hypotheses test possesses what Ramachandran (1958) called the *monotonicity property*.

Furthermore, taking into account that $ARL(\delta) = 1/\xi_\mu(\delta)$, we can add the in-control ARL is never smaller than any out-of-control ARL. Such a behavior of the ARL function means that: the chart satisfies what Ramalhoto and Morais (1995) and Ramalhoto and Morais (1999) called the *primordial criterion*; and we are dealing with what Pignatiello et al. (1995) and Acosta-Mejía and Pignatiello (2000) expertly termed an *ARL-unbiased* chart. Moreover,

$$P[RL_\mu(\delta) > x] = \left[1 - \xi_\mu(\delta)\right]^x \geq \left[1 - \xi_\mu(\delta')\right]^x = P[RL_\mu(\delta') > x], \tag{3}$$

for any $x \in \mathbb{N}$ and any δ, δ' such that $|\delta| \leq |\delta'|$, i.e., the number of collected samples required to detect a change in μ from its target value μ_0 to $\mu_0 + \delta\sigma_0/\sqrt{n}$ stochastically decreases (in the usual sense) with $|\delta|$—in short

$$RL_\mu(\delta) \downarrow_{st} \text{ with } |\delta|. \tag{4}$$

For more details on stochastic ordering, please refer to Shaked and Shanthikumar (1994).

As put by Pignatiello et al. (1995), most of the SPC literature on monitoring process dispersion focuses on the detection of increases in the variance, regardless of the fact that the detection of decreases in dispersion plays a crucial role in the effective implementation of any quality improvement program.

Curiously enough, the S^2-chart—recommended by, we daresay, most Statistical Quality Control (SQC) textbooks[1] to effectively signal the occurrence of both decreases and increases in the standard deviation—is an *ARL-biased* chart, in the sense that some out-of-control ARL values are larger than the in-control ARL, as put by Pignatiello et al. (1995). Moreover, most users are unaware of how to use the S^2-chart (or the R- and S-charts for that matter) to effectively monitor both increases and decreases in the process dispersion (Pignatiello et al. 1995).

Furthermore, SQC textbooks and QC practitioners tend to adopt S- or R-charts, using the famous quality control constants c_4, d_2 and d_3 ($E(S) = c_4\sigma$, $E(R) = d_2\sigma$, $\mathrm{Var}(R) = d_3^2\sigma^2$ under normality) and the following symmetric limits (in-control $\sigma_0 = 1$):

$$S\text{-chart:} \quad c_4 \pm 3\sqrt{1 - c_4^2} \qquad \text{and} \qquad R\text{-chart:} \quad d_2 \pm 3d_3\,. \tag{5}$$

The trouble is twofold—for small sample sizes n ($\leq$5 and $\leq$6 for S- and R-charts, respectively) the lower control limit is negative and the actual in-control ARL differs considerably from the nominal value 370.4, as illustrated by Fig. 1. Using

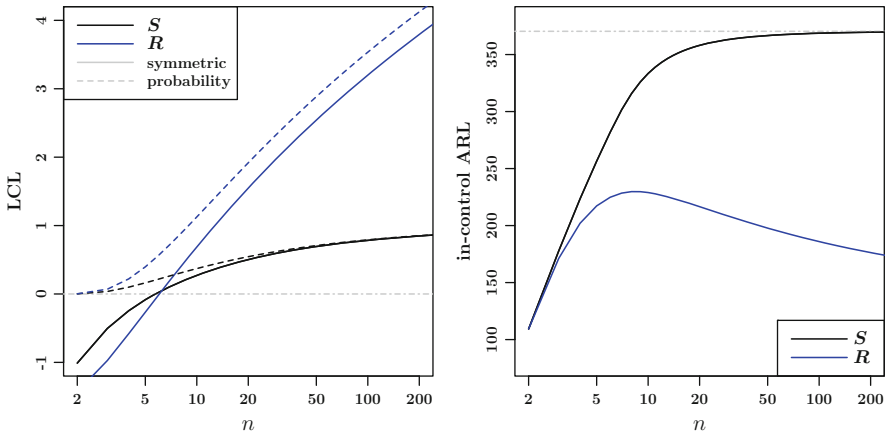


Fig. 1 Lower control limits (*left*)—for symmetric following (5) and probability limits *solid* and *dashed lines*, respectively—and actual in-control ARL (*right*) of R- and S-charts using symmetric limits (the probability limits ensure for every n the target in-control ARL of 370.4)

[1] As far as we have investigated, there is an honorable exception: Uhlmann (1982, pp. 212–215).

the symmetric limits for R-charts is certainly misleading—see also Barbosa et al. (2012) for a more detailed discussion of this phenomenon. S-charts with symmetric limits are fairly reasonable with sample sizes $n \geq 10$, whereas those limits are not recommended for smaller sample sizes.

2 A Closer Look at the ARL-Biased S^2-Chart

The use of a standard S^2-chart is recommended to practitioners to detect changes in the variance of normally distributed output, from its target value σ_0^2 to $\sigma^2 = \theta^2 \times \sigma_0^2$, where $\theta \in (0, +\infty)$.

Let $\alpha = 1/ARL_\sigma(1) \in (0, 1)$, where $ARL_\sigma(1)$ is a prespecified large value of in-control ARL of the S^2-chart. Then the lower and upper control limits found in most SQC textbooks depend on the *equal tail* quantiles of the χ^2_{n-1} distribution,

$$a(\alpha, n) = F^{-1}_{\chi^2_{n-1}}(\alpha/2) \quad \text{and} \quad b(\alpha, n) = F^{-1}_{\chi^2_{n-1}}(1 - \alpha/2), \tag{6}$$

and are determined by

$$\frac{\sigma_0^2}{n-1} \times a(\alpha, n) \quad \text{and} \quad \frac{\sigma_0^2}{n-1} \times b(\alpha, n). \tag{7}$$

Furthermore, the RL of this chart has a geometric distribution with parameter

$$\xi_\sigma(\theta) = 1 - \left\{ F_{\chi^2_{n-1}}\left[b(\alpha, n)/\theta^2\right] - F_{\chi^2_{n-1}}\left[a(\alpha, n)/\theta^2\right] \right\}; \tag{8}$$

as a consequence, $ARL_\sigma(\theta) = 1/\xi_\sigma(\theta)$.

Two graphs of $ARL_\sigma(\theta)$, for $\sigma_0^2 = 1$, $\alpha = 0.002$ and $n = 5, 10$, can be found in Fig. 2. It shows that the in-control ARL is smaller than the out-of-control ARL values for some $\theta \in (0, 1)$, i.e., it takes longer to detect some decreases in the variance than to trigger a false alarm. For instance, when $n = 5$ and $\alpha = 0.002$, the ARL associated with a 5 % decrease in σ, $ARL_\sigma(0.95) \simeq 614.1$ is 23 % larger than $ARL_\sigma(1) = 1/\alpha = 500$. We are indeed dealing with an *ARL-biased* chart, as previously noted by Pignatiello et al. (1995). The next proposition provides an expression for the argmax of $ARL_\sigma(\theta)$; this result and its proof were adapted from Morais (2002, Example 2.6, pp. 20–22) and can also be found in Zhang et al. (2005) and Uhlmann (1982, p. 213).

Proposition 1 *Consider the standard S^2-chart with control limits defined by (7). Then $ARL_\sigma(\theta)$ takes its maximum value at*

$$\theta^*(\alpha, n) = \sqrt{\frac{b(\alpha, n) - a(\alpha, n)}{(n-1)\left\{\ln[b(\alpha, n)] - \ln[a(\alpha, n)]\right\}}}. \tag{9}$$

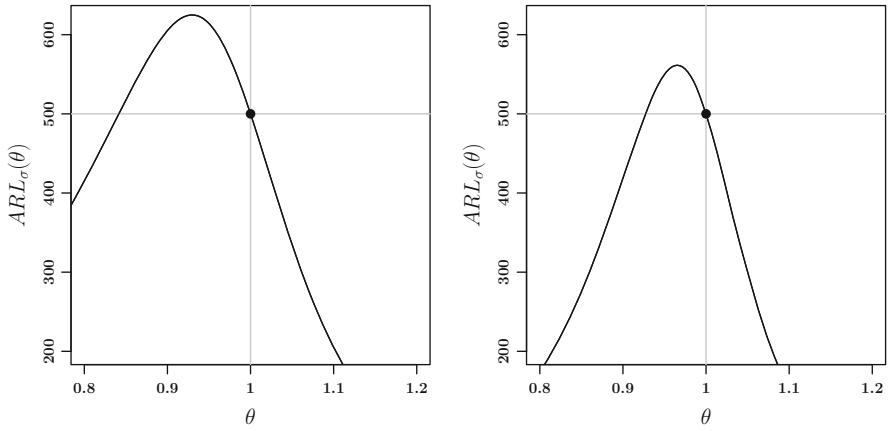


Fig. 2 $ARL_\sigma(\theta)$ of standard S^2-charts with $\sigma_0^2 = 1$ and $\alpha = 0.002$, for $n = 5$ (*left*) and $n = 10$ (*right*)

Proof In order to determine $\theta^*(\alpha, n)$ we have to study the behavior of $\xi_\sigma(\theta)$. Its first derivative has the same sign as

$$b(\alpha, n) \times f_{\chi^2_{n-1}}\left[b(\alpha, n)/\theta^2\right] - a(\alpha, n) \times f_{\chi^2_{n-1}}\left[a(\alpha, n)/\theta^2\right], \tag{10}$$

which in turn has the same sign as

$$k(\theta) = \left[\frac{b(\alpha, n)}{a(\alpha, n)}\right]^{\frac{n-1}{2}} \times \exp\left[-\frac{b(\alpha, n) - a(\alpha, n)}{2\theta^2}\right] - 1. \tag{11}$$

Since $k(\theta)$ is continuous and strictly increasing function of θ in $(0, +\infty)$, such that

$$\lim_{\theta \to 0^+} k(\theta) = -1 \quad \text{and} \quad \lim_{\theta \to +\infty} k(\theta) = \left[\frac{b(\alpha, n)}{a(\alpha, n)}\right]^{\frac{n-1}{2}} - 1 > 0, \tag{12}$$

the first derivative of $\xi_\sigma(\theta)$ changes sign (from negative to positive) only once in $(0, +\infty)$. As a consequence, $\xi_\sigma(\theta)$ takes its minimum value at the unique root of equation $k(\theta) = 0$ in $(0, +\infty)$, which is indeed $\theta^*(\alpha, n)$. Hence, $ARL_\sigma(\theta)$ increases (resp. decreases) with θ for $0 < \theta \leq \theta^*(\alpha, n)$ (resp. $\theta \geq \theta^*(\alpha, n)$). □

By capitalizing on the fact that $\xi_\sigma(\theta)$ decreases (resp. increases) with θ for $0 < \theta \leq \theta^*(\alpha, n)$ (resp. $\theta \geq \theta^*(\alpha, n)$), we can add that

$$RL_\sigma(\theta) \uparrow_{st} \text{ with } \theta \in \left(0, \theta^*(\alpha, n)\right] \tag{13}$$

$$RL_\sigma(\theta) \downarrow_{st} \text{ with } \theta \in \left[\theta^*(\alpha, n), +\infty\right). \tag{14}$$

Table 1 Values of $\theta^*(\alpha, n)$ for standard S^2-charts, with $\sigma_0^2 = 1$, $\alpha = 0.001, 0.002, 1/370.4, 0.003, 0.004, 0.005, 0.010, 0.050$ and $n = 2, 3, 4, 5, 7, 10, 15, 100$

	n							
α	2	3	4	5	7	10	15	100
0.001	0.838196	0.888443	0.914767	0.931301	0.950817	0.965705	0.977286	0.996653
0.002	0.829239	0.883921	0.912191	0.929700	0.950062	0.965367	0.977149	0.996651
1/370.4	0.825003	0.881846	0.911036	0.928994	0.949734	0.965221	0.977089	0.996650
0.003	0.823457	0.881099	0.910624	0.928744	0.949618	0.965170	0.977069	0.996649
0.004	0.819079	0.879013	0.909488	0.928057	0.949302	0.965031	0.977012	0.996648
0.005	0.815510	0.877345	0.908591	0.927519	0.949057	0.964923	0.976968	0.996647
0.010	0.803343	0.871885	0.905733	0.925830	0.948296	0.964590	0.976834	0.996645
0.050	0.766915	0.857568	0.898763	0.921854	0.946556	0.963837	0.976532	0.996639

Furthermore, when we use the *equal tail* quantiles of the χ^2_{n-1} distribution: $0 < \theta^*(\alpha, n) < 1$, as illustrated by Table 1, for $\alpha = 0.001, 0.002, 1/370.4, 0.003, 0.004, 0.005, 0.010, 0.050$ and $n = 2, 3, 4, 5, 7, 10, 15, 100$; the maximum of $ARL_\sigma(\theta)$ is associated with a value of σ that is about $[1 - \theta^*(\alpha, n)] \times 100\,\%$ less than the target value σ_0; and, for instance, when $\theta \in [\theta^*(\alpha, n), 1)$ the standard S^2- chart is less likely to trigger a valid signal within the x first samples than in the absence of assignable causes. What Pignatiello et al. (1995) and Morais (2002, pp. 20–22) failed to mention is that the range of the interval where $ARL(\theta) > ARL(1)$ decreases with n, as shown in Fig. 2. This is essentially due to the fact that the interquantile range $b(\alpha, n) - a(\alpha, n)$ (resp. the ratio $b(\alpha, n)/a(\alpha, n)$) increases (resp. decreases) with n, for fixed α (see Saunders and Moran 1978)—thus, $[\theta^*(\alpha, n)]^{\frac{n-1}{2}} = \frac{b(\alpha,n)-a(\alpha,n)}{\ln[b(\alpha,n)]-\ln[a(\alpha,n)]}$ increases with n, not to mention the fact that chi-square distributions look more and more "symmetrical" as the sample size grows.

3 Revisiting the ARL-Unbiased S^2-Chart

Pignatiello et al. (1995), certainly inspired by Ramachandran (1958), proposed alternative control limits for the S^2-chart in order to achieve an *ARL-unbiased* chart. Let

$$\frac{\sigma_0^2}{n-1} \times \tilde{a}(\alpha, n) \quad \text{and} \quad \frac{\sigma_0^2}{n-1} \times \tilde{b}(\alpha, n) \tag{15}$$

be the lower and upper control limits of what we shall term the *ARL-unbiased* S^2-chart. In this case,

$$\tilde{\xi}_\sigma(\theta) = 1 - \left\{ F_{\chi^2_{n-1}}\left[\tilde{b}(\alpha, n)/\theta^2\right] - F_{\chi^2_{n-1}}\left[\tilde{a}(\alpha, n)/\theta^2\right] \right\}, \ \theta \in (0, +\infty), \tag{16}$$

and $\widetilde{ARL}_\sigma(\theta) = 1/\tilde{\xi}_\sigma(\theta)$. Then the critical values $\tilde{a}(\alpha, n)$ and $\tilde{b}(\alpha, n)$, associated with the *ARL-unbiased* S^2-chart, are determined by solving the system of equations

$$F_{\chi^2_{n-1}}\left[\tilde{b}(\alpha, n)\right] - F_{\chi^2_{n-1}}\left[\tilde{a}(\alpha, n)\right] = 1 - \alpha \tag{17}$$

$$\tilde{b}(\alpha, n) \times f_{\chi^2_{n-1}}\left[\tilde{b}(\alpha, n)\right] - \tilde{a}(\alpha, n) \times f_{\chi^2_{n-1}}\left[\tilde{a}(\alpha, n)\right] = 0 \tag{18}$$

numerically, namely using $a(\alpha, n)$ and $b(\alpha, n)$ as the initial values in the numerical search. Suffice to say that Eqs. (17) and (18) follow immediately from $\widetilde{ARL}_\sigma(1) = 1/\alpha$ and the condition for unbiasedness $d\,\widetilde{ARL}_\sigma(\theta)/d\theta\,|_{\theta=1} = 0$, respectively.

Since the p.d.f. of the chi-square distribution with $(n-1)$ degrees of freedom is proportional to $x^{(n-3)/2}\, e^{-x/2}$, (18) can be rewritten as

$$f_{\chi^2_{n+1}}[\tilde{b}(\alpha, n)] - f_{\chi^2_{n+1}}[\tilde{a}(\alpha, n)] = 0. \tag{19}$$

According to Tate and Klett (1959), conditions (17) and (19) were introduced by Neyman and Pearson (1936, pp. 18, 19, 25). Furthermore, invoking once again such proportionality, we can add that (18) is equivalent to

$$\left[\frac{\tilde{b}(\alpha, n)}{\tilde{a}(\alpha, n)}\right]^{\frac{n-1}{2}} = \exp\left[\frac{\tilde{b}(\alpha, n) - \tilde{a}(\alpha, n)}{2\theta^2}\right], \tag{20}$$

as mentioned by Fertig and Proehl (1937), Ramachandran (1958), Kendall and Stuart (1979, p. 219), and Pignatiello et al. (1995), or put in equivalent equations by Uhlmann (1982, p. 213) and Krumbholz and Zoeller (1995).

As for proving that argmax $\widetilde{ARL}_\sigma(\theta) = 1$, we have to simply proceed either as in the proof of Proposition 1, by replacing $a(\alpha, n)$ and $b(\alpha, n)$ with $\tilde{a}(\alpha, n)$ and $\tilde{b}(\alpha, n)$ in $k(\theta)$; or in a similar fashion to Ramalhoto and Morais (1995, Appendix A), Ramalhoto and Morais (1999, Proof of Prop. 1)[2] or Kendall and Stuart (1979, p. 219).

A large set of critical values $\tilde{a}(\alpha, n)$ and $\tilde{b}(\alpha, n)$ is provided by Table 2. Other sets of these critical values can be found in: Ramachandran (1958, Table 744) with 2 decimal places, for $\alpha = 0.05$ and $n - 1 = 2(1)8(2)24, 30, 40, 60$; Tate and Klett (1959, Table 680) with 4 decimal places, for $\alpha = 0.001, 0.005, 0.01, 0.05, 0.1$ and $n - 1 = 2(1)29$; Pachares (1961, Table I) with 5 significant figures, for $\alpha = 0.01, 0.05, 0.1$ and $n - 1 = 1(1)20, 24, 30, 40, 60, 120$; Kendall and Stuart

[2]We ought to add that Ramalhoto and Morais (1995) and Ramalhoto and Morais (1999) proposed *ARL-unbiased* charts for the scale parameter of a Weibull quality characteristic (thus, dealing with control statistic with a χ^2_{2n} distribution) and proved not only that the ARL has a maximum at $\theta = 1$, but also that it strictly increases (resp. decreases) with $\theta \in (0, 1)$ (resp. $\theta \in (1, +\infty)$).

Table 2 $a(\alpha, n)$, $b(\alpha, n)$, $\tilde{a}(\alpha, n)$ and $\tilde{b}(\alpha, n)$ for standard and *ARL-unbiased* S^2-charts, with $\sigma_0^2 = 1$, $\alpha = 0.001, 0.002, 1/370.4, 0.003, 0.004, 0.005, 0.010$ and $n = 2, 3, 4, 5, 7, 10, 15, 100$

	n							
α	2	3	4	5	7	10	15	100
0.001	$a(\alpha, n) = 3.9270 \times 10^{-7}$	0.0010	0.0153	0.0639	0.2994	0.9717	2.6967	59.1282
	$b(\alpha, n) = 12.1157$	15.2018	17.7300	19.9974	24.1028	29.6658	38.1094	151.9340
	$\tilde{a}(\alpha, n) = 1.4027 \times 10^{-6}$	0.0018	0.0221	0.0831	0.3520	1.0745	2.8651	59.5624
	$\tilde{b}(\alpha, n) = 16.2662$	18.4677	20.5239	22.4856	26.2141	31.4692	39.6502	152.8876
0.002	1.5708×10^{-6}	0.0020	0.0243	0.0908	0.3811	1.1520	3.0407	61.1365
	10.8276	13.8155	16.2662	18.4668	22.4577	27.8772	36.1233	148.2304
	5.5529×10^{-6}	0.0036	0.0349	0.1174	0.4466	1.2716	3.2277	61.5832
	14.7955	16.9255	18.9214	20.8306	24.4664	29.5988	37.6015	149.1649
1/370.4	2.8623×10^{-6}	0.0027	0.0297	0.1058	0.4234	1.2412	3.2060	62.0565
	10.2730	13.2155	15.6306	17.8006	21.7392	27.0933	35.2498	146.5811
	1.0067×10^{-5}	0.0048	0.0425	0.1365	0.4955	1.3691	3.4019	62.5089
	14.1564	16.2538	18.2224	20.1079	23.7018	28.7784	36.7002	147.5071
0.003	3.5343×10^{-6}	0.0030	0.0319	0.1116	0.4394	1.2744	3.2667	62.3878
	10.0786	13.0046	15.4068	17.5659	21.4857	26.8164	34.9407	145.9943
	1.2407×10^{-5}	0.0053	0.0456	0.1439	0.5140	1.4053	3.4659	62.8422
	13.9314	16.0170	17.9759	19.8530	23.4318	28.4884	36.3811	146.9172

0.004	6.2832×10^{-6}	0.0040	0.0387	0.1292	0.4864	1.3702	3.4398	63.3140
	9.5495	12.4292	14.7955	16.9238	20.7912	26.0564	34.0913	144.3730
	2.1936×10^{-5}	0.0071	0.0551	0.1662	0.5682	1.5098	3.6481	63.7741
	13.3164	15.3692	17.3011	19.1547	22.6917	27.6925	35.5043	145.2875
0.005	9.8175×10^{-6}	0.0050	0.0449	0.1449	0.5266	1.4501	3.58202	64.0561
	9.1406	11.9829	14.3203	16.4239	20.2494	25.4625	33.4260	143.0938
	3.4117×10^{-5}	0.0088	0.0639	0.1859	0.614451	1.5969	3.7979	64.5208
	12.8382	14.8647	16.7753	18.6103	22.1138	27.0701	34.8173	144.0017
0.010	3.9270×10^{-5}	0.0100	0.0717	0.2070	0.6757	1.7349	4.0747	66.5101
	7.8794	10.5966	12.8382	14.8603	18.5476	23.5894	31.3193	138.9868
	1.3422×10^{-4}	0.0175	0.1011	0.2640	0.7857	1.9068	4.3161	66.9898
	11.3450	13.2854	15.1269	16.9013	20.2956	25.1056	32.6412	139.8733
0.050	9.8207×10^{-4}	0.0506	0.2158	0.4844	1.2373	2.7004	5.6287	73.3611
	5.0239	7.3778	9.3484	11.1433	14.4494	19.0228	26.1189	128.4220
	3.1593×10^{-3}	0.0847	0.2962	0.6070	1.4250	2.9532	5.9477	73.8821
	7.8168	9.5303	11.1915	12.8024	15.8966	20.3049	27.2631	129.2527

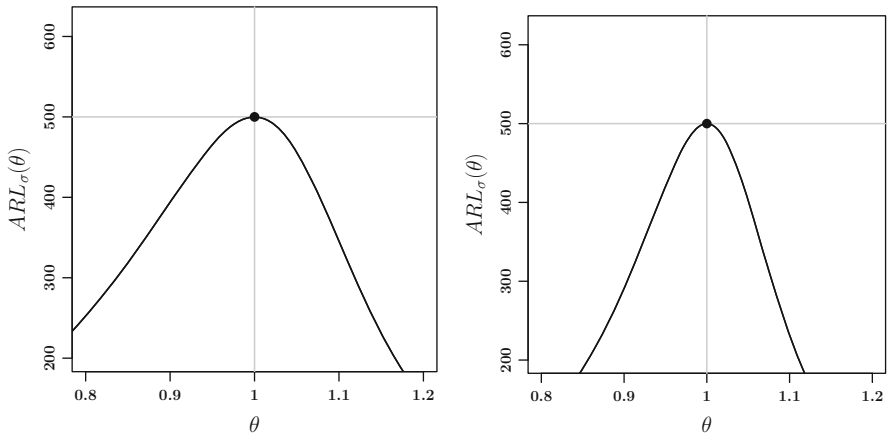


Fig. 3 $\widetilde{ARL}_\sigma(\theta)$ of *ARL-unbiased* S^2-charts with $\sigma_0^2 = 1$ and $\alpha = 0.002$, for $n = 5$ (*left*) and $n = 10$ (*right*)

(1979, Table 23.1, p. 219) with 2 decimal places, for $\alpha = 0.05$ and $n - 1 = 2, 5, 10, 20, 30, 40, 60$; Pignatiello et al. (1995, Table 3) with 4 decimal places, for $\alpha = 0.005, 0.00286, 0.0020$ and $n = 3(2)15(10)55$.

It goes without saying that the use of the critical values $\tilde{a}(\alpha, n)$ and $\tilde{b}(\alpha, n)$, such as the ones in Table 2, leads to *ARL-unbiased* S^2-charts which offer a more balanced protection[3] against both increases and decreases in the process dispersion, as depicted in Fig. 3. Moreover,

$$\widetilde{RL}_\sigma(\theta) \uparrow_{st} \text{ with } \theta \in (0, 1] \tag{21}$$

$$\widetilde{RL}_\sigma(\theta) \downarrow_{st} \text{ with } \theta \in [1, +\infty). \tag{22}$$

4 The ARL-Unbiased EWMA-S^2-Chart

If we dismiss the Shewhart charts with run rules suggested by Page (1955), WECO (1956), and Roberts (1958), we can add that Shewhart charts only use the last observed value of their control statistics to trigger (or not) a signal—and simply ignore any information contained in the previous samples. This fact is responsible for a serious and well-known limitation: Shewhart charts are not effective in the detection of small to moderate changes in the parameter being monitored. This limitation led to the proposal of alternative charts, such as the

[3] Regretfully, this chart has not uniformly smaller out-of-control ARLs than any other S^2-chart, as noted by Zhang et al. (2005). This result prompted these authors to propose two alternatives to the *ARL-unbiased* S^2-chart.

cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) control charts introduced by Page (1954) and Roberts (1959), respectively. For example, EWMA control charts make use of recursive control statistics that account for the information contained in every collected sample of the process—in fact, the EWMA charts for the process mean apply successively decreasing weights to the past sample means (Nelson 1982, p. 181), and certainly prove to require a smaller average number of samples than the $\bar{X}$-chart to detect small to moderate shifts in the process mean. Unsurprisingly, the EWMA chart for the mean of Gaussian output—with control limits symmetric around the target value μ_0—is an *ARL-unbiased* chart because the control statistic has a symmetric distribution around μ_0 when the process mean is on target.

Pioneering papers on the use of the EWMA smoothing to monitor the process variance are Wortham and Ringer (1971), Sweet (1986), MacGregor and Harris (1993) and Mittag et al. (1998) are more recent contributions on the subject. For more details about monitoring normal variance with EWMA charts, see Knoth (2010) and further references therein.

Setting up an EWMA-S^2-chart is straightforward. Recall that it makes use of the following control statistic:

$$Z_i = \begin{cases} z_0, & i = 0 \\ (1-\lambda)\, Z_{i-1} + \lambda\, S_i^2, & i \in \mathbb{N}. \end{cases} \tag{23}$$

The quantity z_0 represents the starting value, often the target value σ_0^2. λ is a suitable constant chosen from the interval $(0, 1]$ to ensure a specific in-control and out-of-control signaling behavior: small (resp. large) values of λ should be used to efficiently signal the occurrence of small (resp. large) shifts (Capizzi and Masarotto 2003).

Since the distribution of this control statistic is obviously not symmetrical, adopting control limits that are symmetrical with respect to the target value σ_0^2 might not provide proper protection against both increases and decreases in spread. In early papers such as Wortham and Ringer (1971) and MacGregor and Harris (1993), where EWMA was applied to squared individual observations, simple approximations of the two control limits in the spirit of the above classical Shewhart limits were constructed. To be precise, both papers use chi-square percentiles based limits whose degrees of freedom are a function of the smoothing parameter λ. Besides, in Ng and Case (1989) similar ideas were used for EWMA R (sample range) charts. In all cases neither the overall on-target condition is met nor an ARL-unbiased design is obtained. Note that the problem of choosing the control limits of EWMA charts meant to monitor both increases and decreases in the process variance and based on asymmetrically distributed control statistics is not properly discussed in literature.

Here, we check two design approaches: (i) Control limits that are symmetrical with respect to the target value σ_0^2 (*vanilla*). (ii) Mimic the Shewhart S^2 setup by taking these limits and use them as upper and lower control limits, respectively, of

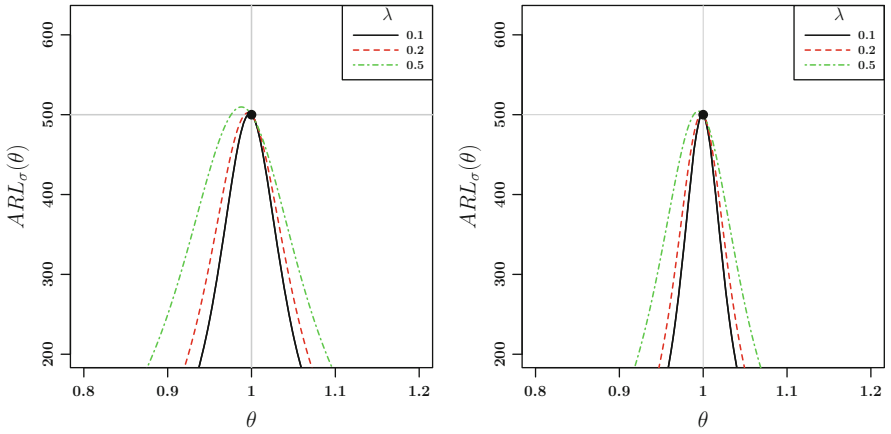


Fig. 4 $ARL_\sigma(\theta)$ of *equal tails* EWMA-S^2-schemes, with $\sigma_0^2 = 1$ and on-target ARL equal to 500, for $n = 5$ (*left*) and $n = 10$ (*right*)

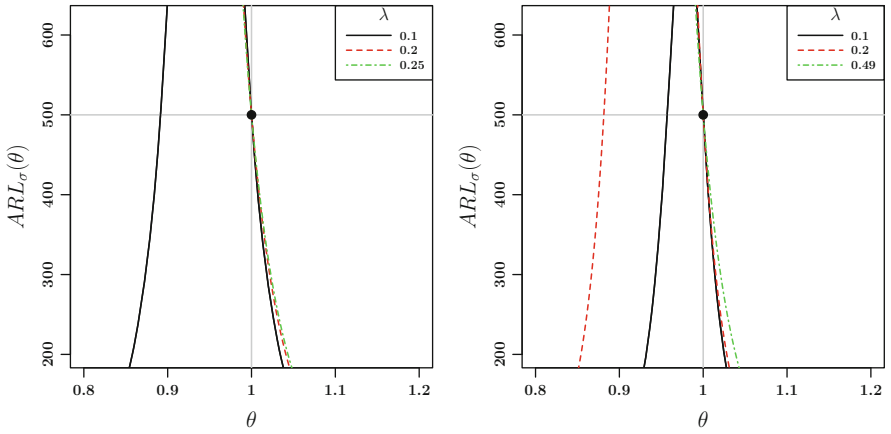


Fig. 5 $ARL_\sigma(\theta)$ of *vanilla* EWMA-S^2-schemes, with $\sigma_0^2 = 1$ and on-target ARL equal to 500, for $n = 5$ (*left*) and $n = 10$ (*right*)

one-sided EWMA charts yielding the same on-target ARL (*equal tails*). Recall that for values of λ larger than a certain real number it is not possible at all to construct *vanilla* designs—see Fig. 5.

To start with the more promising (and more demanding) *equal tails* case, (ii) is evaluated for $\lambda \in \{0.1, 0.2, 0.5\}$. The related Fig. 4 illustrates not only how biased is the ARL profile of the *equal tails* EWMA-S^2-schemes, but also how the bias is much less pronounced as we decrease the value of the smoothing constant λ. It looks much worse for the *vanilla* case, that is with symmetric control limits—see Fig. 5. Firstly, for $\lambda \geq 0.26$ and $\lambda \geq 0.5$ symmetric designs do not exist in case of $n = 5$ and $n = 10$, respectively. Secondly, the ARL profiles are heavily biased—only for

$\lambda = 0.1$ we get tolerable profiles: however, compared to the *equal tails* results in Fig. 4 the *vanilla* design approach fails completely if ARL-unbiased designs are requested.

All these ARL profiles can be certainly improved. However, the obtention of an *ARL-unbiased* EWMA (or CUSUM) chart for the variance of Gaussian output is more complex than the derivation of its Shewhart counterpart. This problem has been tackled previously, namely by: Pignatiello et al. (1995), who proposed an *ARL-unbiased* scheme consisting of two one-sided CUSUM charts whose control statistics deploy S; Acosta-Mejía and Pignatiello (2000), who discussed *ARL-unbiased* EWMA dispersion charts for subgroups of size one; Knoth (2010), who analyzed different dispersion EWMA schemes (S^2, S, R, $\ln S^2$) in their *ARL-unbiased* versions. In any of these papers, the authors obtained several combinations of pairs of control limits that yield the same in-control ARL, thus obtaining what Acosta-Mejía and Pignatiello (2000) called the Iso-ARL curve. Subsequently, they adopted a search procedure to find the point on the Iso-ARL curve that produces the *ARL-unbiased* CUSUM or EWMA chart.

The *ARL-unbiased* EWMA-S^2-chart triggers a signal at sample i iff

$$Z_i < \sigma_0^2 - \tilde{a}_E(\alpha, n)\sqrt{\frac{2\lambda\sigma_0^2}{(2-\lambda)(n-1)}} \quad \text{or} \quad Z_i > \sigma_0^2 + \tilde{b}_E(\alpha, n)\sqrt{\frac{2\lambda\sigma_0^2}{(2-\lambda)(n-1)}}, \tag{24}$$

where the constants $\tilde{a}_E(\alpha, n)$ and $\tilde{b}_E(\alpha, n)$ are chosen in such way that the ARL takes its maximum value at $\theta = 1$. These constants can be found in Table 3 for $\sigma_0^2 = 1$, $\lambda = 0.1$, on-target ARL equal to α^{-1}, $\alpha = 0.001, 0.002, 1/370.4$, $0.003, 0.004, 0.005, 0.010$ and $n = 2, 3, 4, 5, 7, 10, 15, 100$.

Choosing the *ARL-unbiased* chart not only makes the ARL values decrease as we move away from $\theta = 1$, but also decrease most rapidly in the vicinity of $\theta = 1$, as illustrated in Fig. 6. The use of the constants $\tilde{a}_E(\alpha, n)$ and $\tilde{b}_E(\alpha, n)$, whose obtention is described in detail in the appendix, is indeed to good advantage when it comes to the detection of both decreases and increases in the process variance.

Further examples of the ARL profiles of variance charts can be found in Fig. 7. They refer to the vanilla and ARL-unbiased versions of the EWMA-log S^2-, EWMA-S-charts and the two-sided[4] CUSUM$-S^2$-scheme. It is apparent that, for the vanilla version of the EWMA-log S^2-chart, the maximum of ARL profile is to the right of the in-control value $\theta_0 = 1$. As for the ARL of the EWMA-S vanilla chart, Fig. 6 leads to the conclusion that it is less biased than the ARL of the vanilla version of the EWMA-S^2 we have previously discussed. Interestingly enough,

[4]The two-sided CUSUM$-S^2$-scheme consists of two separate lower and upper one-sided charts. Their reference values k_L and k_U are implicitly given by the chosen out-of-control values $\sigma_1 < \sigma_0$ and $1/\sigma_1$.

Table 3 $\tilde{a}_E(\alpha, n)$ and $\tilde{b}_E(\alpha, n)$ for *ARL-unbiased* EWMA-S^2-charts, with $\sigma_0^2 = 1$, $\lambda = 0.1$, on-target ARL equal to α^{-1}, $\alpha = 0.001, 0.002, 1/370.4, 0.003, 0.004, 0.005, 0.010$ and sample sizes $n = 2, 3, 4, 5, 7, 10, 15, 100$

	n							
α	2	3	4	5	7	10	15	100
0.001	$\tilde{a}_E(\alpha, n) = 1.9905$	2.2541	2.3840	2.4653	2.5656	2.6504	2.7274	2.9304
	$\tilde{b}_E(\alpha, n) = 4.4786$	4.0470	3.8579	3.7463	3.6152	3.5094	3.4174	3.1906
0.002	1.8891	2.1218	2.2354	2.3061	2.3929	2.4661	2.5322	2.7057
	3.9840	3.6345	3.4798	3.3880	3.2798	3.1921	3.1155	2.9257
1/370.4	1.8390	2.0578	2.1641	2.2302	2.3111	2.3792	2.4406	2.6012
	3.7659	3.4505	3.3102	3.2268	3.1282	3.0482	2.9781	2.8040
0.003	1.8204	2.0342	2.1380	2.2024	2.2812	2.3475	2.4073	2.5634
	3.6886	3.3850	3.2497	3.1692	3.0739	2.9966	2.9287	2.7601
0.004	1.7664	1.9666	2.0633	2.1232	2.1963	2.2577	2.3130	2.4568
	3.4764	3.2042	3.0822	3.0094	2.9231	2.8529	2.7912	2.6372
0.005	1.7213	1.9107	2.0018	2.0582	2.1269	2.1845	2.2363	2.3707
	3.3105	3.0618	2.9499	2.8829	2.8034	2.7385	2.6814	2.5387
0.010	1.5583	1.7140	1.7881	1.8337	1.8889	1.9350	1.9762	2.0824
	2.7903	2.6088	2.5258	2.4757	2.4158	2.3666	2.3231	2.2133
0.050	1.0124	1.1000	1.1413	1.1664	1.1965	1.2215	1.2437	1.3000
	1.6350	1.5570	1.5196	1.4963	1.4680	1.4442	1.4229	1.3678

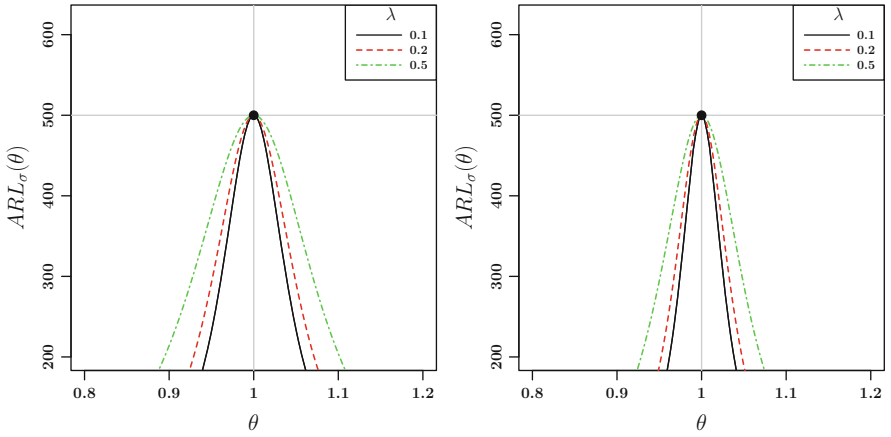


Fig. 6 $ARL_\sigma(\theta)$ of *ARL-unbiased* EWMA-S^2-charts with $\sigma_0^2 = 1$ and on-target ARL equal to 500, for $n = 5$ (*left*) and $n = 10$ (*right*)

the two-sided CUSUM vanilla scheme exhibits only slightly biased ARL profiles and should be preferred to all the remaining vanilla charts. For a more elaborated comparison among different EWMA-type schemes for the variance, please refer to Knoth (2010).

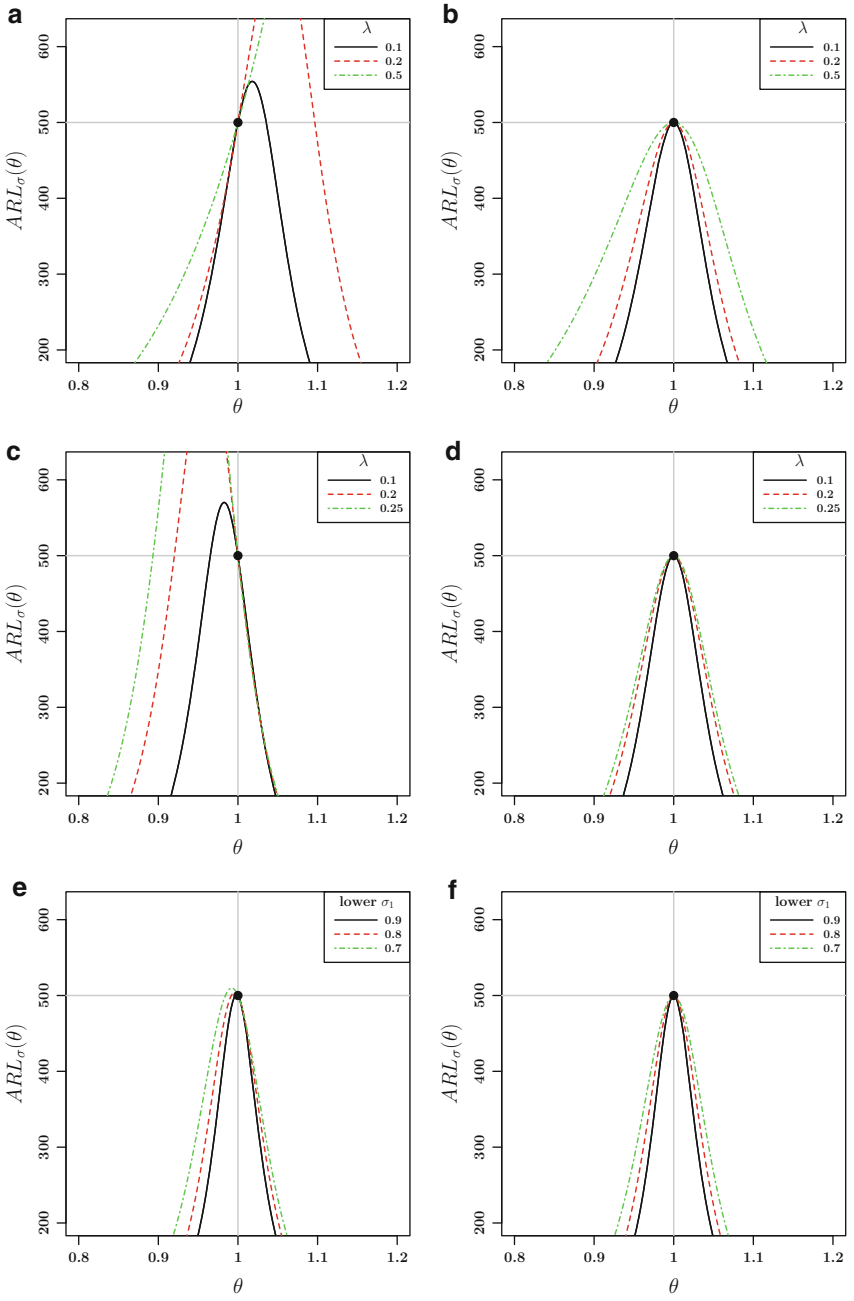


Fig. 7 $ARL_\sigma(\theta)$ of vanilla (*left*) and *ARL-unbiased* (*right*) designs for EWMA-log S^2 and EWMA-S charts and two-sided CUSUM-S^2 scheme, for $n = 5$. (**a**) EWMA-log S^2, vanilla, (**b**) EWMA-log S^2, unbiased, (**c**) EWMA-S, vanilla, (**d**) EWMA-S, unbiased, (**e**) CUSUM-S^2, vanilla, (**f**) CUSUM-S^2, unbiased

5 Relating ARL-Unbiased Charts and UMPU Tests

The concept of an *ARL-unbiased* Shewhart-type chart is certainly related to the notion of unbiased test, which is due to Neyman and Pearson (1936, 1938), according to Lehmann (2006, p. 8821). Let us remind the reader that:

- a size α test for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_0^c$, with power function $\xi(\theta)$, is said to be *unbiased* if $\xi(\theta) \leq \alpha$, for $\theta \in \Theta_0$, and $\xi(\theta) \geq \alpha$, for $\theta \in \Theta_0^c$ (Shao 2003, p. 404), thus, the test is at least as likely to reject under any alternative as under the null hypothesis (Lehmann 2006, p. 8820);
- if we consider $\mathcal{C}$ a class of tests for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_0^c$, then a test in $\mathcal{C}$, with power function $\xi(\theta)$, is a uniformly most powerful (UMP) class $\mathcal{C}$ test if $\xi(\theta) \geq \xi'(\theta)$, for every $\theta \in \Theta_0^c$ and every $\xi'(\theta)$ that is a power function of a test in class $\mathcal{C}$ (Casella and Berger 1990, p. 365).

Lehmann (2006, p. 8820) also notes that in many situations in which no UMP test exists, there is a test which is UMP among the class of all unbiased tests—the uniformly most powerful unbiased (UMPU) test.

Proposition 2, inspired by Kendall and Stuart (1979, pp. 224–226), Lehmann (1986, pp. 145–147) and Shao (2003, pp. 406–407), provides a systematic way of obtaining an UMPU test concerning a real-valued parameter in an exponential family, with the remaining parameters occurring as unspecified nuisance parameters.

Proposition 2 *Let $X = (X_1, \ldots, X_n)$ be distributed according to*

$$f_{\eta,\varphi}(x) = \exp\left\{\eta\, Y(x) + \varphi^\top U(x) - \zeta(\theta, \varphi)\right\}, \tag{25}$$

where: η is a real-valued parameter; φ is a vector-valued parameter; Y a real-valued statistic; U a vector-valued statistic. Then for testing $H_0 : \eta = \eta_0$ and $H_1 : \eta \neq \eta_0$, the UMPU test of size α is characterized by the following critical function:

$$T(Y,U) = \begin{cases} 1, & Y < c_1(U) \text{ or } Y > c_2(U) \\ \gamma_i(U), & Y = c_i(U),\ i = 1,2 \\ 0, & c_1(U) < Y < c_2(U), \end{cases} \tag{26}$$

where $c_i(u)$ and $\gamma_i(u)$ are functions determined by

$$E_{\eta_0}[T(Y,U) \mid U = u] = \alpha \tag{27}$$

$$E_{\eta_0}[T(Y,U) \times Y \mid U = u] = \alpha \times E_{\theta_0}(Y \mid U = u), \tag{28}$$

for every u.

Proposition 2 can be applied in a straightforward manner to derive an UMPU test for $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$ (see (Kendall and Stuart 1979, pp. 227–228),

(Lehmann 1986, p. 194) and (Shao 2003, pp. 411–412)), hence to define the control limits of what could be called an *UMPU* Shewhart-type chart for σ^2. Let $X_1, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$, where both $\mu \in IR$ and $\sigma^2 \in IR^+$ are unknown and $n \in \{2, 3, \ldots\}$. Then

$$f_{\eta,\varphi}(x) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{n\mu}{\sigma^2}\bar{x} - \frac{n\mu^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2)\right\},$$

$$\eta = -\frac{1}{2\sigma^2}, \quad \varphi = \frac{n\mu}{\sigma^2},$$

$$Y(X) = \sum_{i=1}^{n} X_i^2, \quad U(X) = \bar{X}.$$

Moreover:

- by Basu's theorem, $(n-1)S^2 = Y - n\,U^2$ is independent of U;
- $V = (n-1)S^2/\sigma_0^2 \sim \chi^2_{n-1}$;
- $\gamma_i(u) = 0$, $i = 1, 2$ because Y is a continuous r.v., thus, there is no need to randomize the test;
- $c_i(u)$, $i = 1, 2$, do not depend on u and are related to quantiles of the χ^2_{n-1};
- the critical function can be written solely in terms of V and should be interpreted as follows:
  - reject $H_0 : \sigma^2 = \sigma_0^2$ (i.e., $\eta = \eta_0$) if $V < \tilde{a}$ or $V > \tilde{b}$,
  - not reject H_0, otherwise;
- conditions (27) and (28) can be written only in terms of V

$$E_{\sigma_0}\left[1 - I_{[\tilde{a},\tilde{b}]}(V)\right] = \alpha \tag{29}$$

$$E_{\sigma_0}\left\{[1 - I_{[\tilde{a},\tilde{b}]}(V)] \times V\right] = \alpha \times E_{\sigma_0}(V), \tag{30}$$

  where $I_{[\tilde{a},\tilde{b}]}$ is the indicator function of the interval $[\tilde{a}, \tilde{b}]$.

Finally, capitalizing on the distribution of V and on fact that

$$v f_{\chi^2_{n-1}}(v) = (n-1) f_{\chi^2_{n+1}}(v), \tag{31}$$

we are able to rewrite conditions (29) and (30) and add that the constants $\tilde{a}$ and $\tilde{b}$ are determined by solving

$$F_{\chi^2_{n-1}}(\tilde{b}) - F_{\chi^2_{n-1}}(\tilde{a}) = 1 - \alpha \tag{32}$$

$$F_{\chi^2_{n+1}}(\tilde{b}) - F_{\chi^2_{n+1}}(\tilde{a}) = 1 - \alpha. \tag{33}$$

These constants are not those corresponding to the *equal tail* quantiles or to the likelihood ratio test (Lehmann 2006, p. 8820). Interestingly enough, $\tilde{a}$ and $\tilde{b}$ coincide with $\tilde{a}(\alpha, n)$ and $\tilde{b}(\alpha, n)$, respectively, even though condition (18) is equivalent to

$$f_{\chi^2_{n+1}}[\tilde{b}(\alpha, n)] - f_{\chi^2_{n+1}}[\tilde{a}(\alpha, n)] = 0, \tag{34}$$

and does not seem to agree with (33).

6 Going Beyond ARL-Based Chart Designs

Since the control statistics of EWMA charts are dependent, the probability of signalling at sample i and the probability of triggering a signal at sample i given that the chart has not signalled prior to sample i are time-varying functions. As a result, we fail to relate *ARL-unbiased* EWMA charts to unbiased or UMPU tests and we are bound to explore other criteria—other than the *ARL-unbiasedness*—to set up EWMA-S^2-charts to monitor dispersion.

The remainder of the paper is devoted to finding the control limits of EWMA charts for σ^2 that simultaneously yield

$$P[RL_\sigma(1) \leq RL^{(0)}] \leq \alpha \tag{35}$$

$$P[RL_\sigma(\theta) \leq RL^{(0)}] \geq \alpha, \text{ for all } \theta \neq 1, \tag{36}$$

where $RL^{(0)}$ and α are pre-specified constants. The resulting chart is going to be simply termed *unbiasedProb* EWMA-S^2-chart. We ought to comment on these two conditions.

The first one parallels with the condition $ARL_\sigma(1) \geq \alpha^{-1} \Leftrightarrow \xi_\sigma(1) \leq \alpha$ (i.e., the probability of a false alarm never exceeds α) in the Shewhart setting. To motivate the choice of $RL^{(0)}$ and α, assume $RL^{(0)}$ is the planned monitoring horizon for the EWMA chart and α represents the probability of at least one false alarm during this monitoring horizon. As for the second condition, it corresponds somehow to the *ARL-unbiased* condition $ARL_\sigma(\theta) \leq 1/\alpha \Leftrightarrow \xi_\sigma(\theta) \geq \alpha$, for all $\theta \neq 1$, in the Shewhart scenario.

An example with $RL^{(0)} = 500$ and $\alpha = 0.25$ is provided to illustrate the properties of an *unbiasedProb* EWMA-S^2-chart. The performance of this chart—in terms of the probability that at least one signal is triggered in the monitoring horizon, $P[RL_\sigma(\theta) \leq RL^{(0)}]$—is confronted with the one of a *equaltailsProb* EWMA-S^2-scheme that makes use (setting it up) of a lower one-sided and an upper one-sided EWMA-S^2-chart set in such way that they both have the same on-target probability that at least a false alarm during the monitoring horizon, from now on called the *equaltailsProb*. Fig. 8 has the profiles of $P[RL_\sigma(\theta) \leq 500]$, for the three values of the smoothing constant and the two sample sizes previously used, $\lambda = 0.1, 0.2, 0.5$

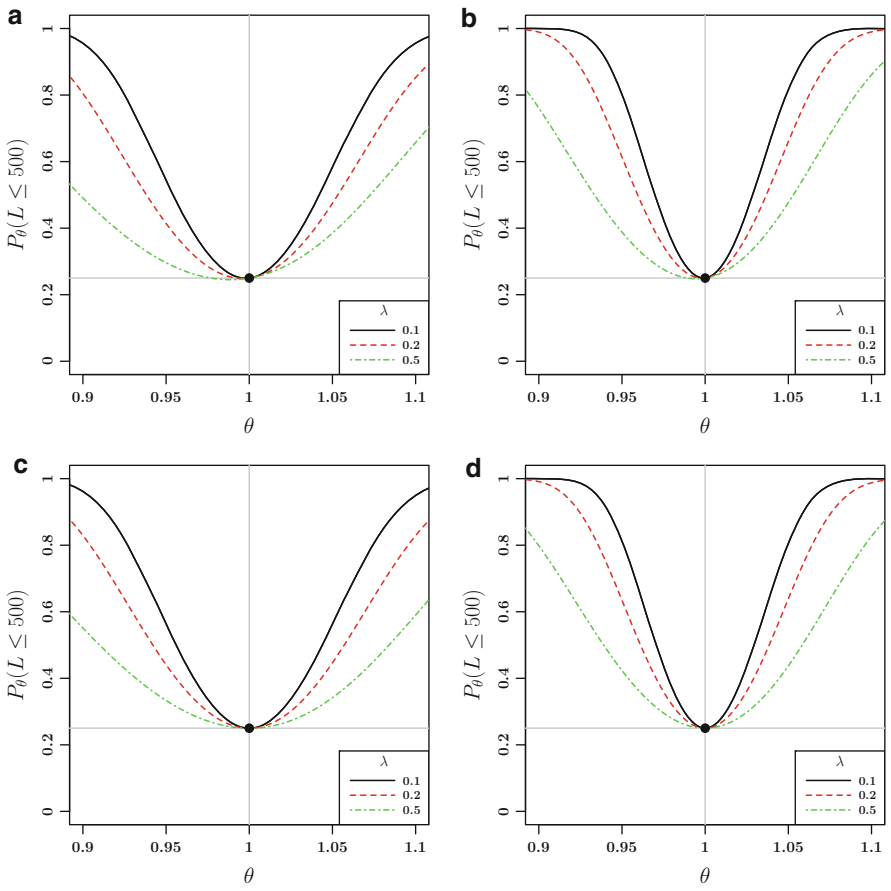


Fig. 8 $P[RL_\sigma(\theta) \leq 500]$ for sample sizes $n = 5$ (*left*) and $n = 10$ (*right*). (**a**) $n = 5$, *equaltailsProb* setup, (**b**) $n = 10$, *equaltailsProb* setup, (**c**) $n = 5$, *unbiasedProb* setup, (**d**) $n = 10$, *unbiasedProb* setup

and $n = 5, 10$. The *bias* of the function $P[RL_\sigma(\theta) \leq 500]$ is clearly visible when the *equaltailsProb* EWMA-S^2-scheme is used, in particular for the smallest of the sample sizes and the largest value of λ. In contrast, the *unbiasedProb* EWMA-S^2-chart yields nearly symmetrical profiles of $P[RL_\sigma(\theta) \leq 500]$ in the vicinity of $\theta = 1$.

We strongly believe that this approach to design an EWMA-S^2-chart might reconcile the classical control chart users from the Shewhart chart community with the users of control charts with recursive control statistics, such as the EWMA. Moreover, it is a plausible direction to setting up charts when the quality control practitioner has additional knowledge of the monitoring horizon of the chart.

7 Concluding Remarks

The aim of this paper is to draw the attention of practitioners who tend to use control charts which are not *ARL-unbiased* because the adopted control limits were based, namely on *equal tail* quantiles, disregarding the skewness of the distribution of the control statistic and the fact that this customary procedure is far from satisfactory.

We feel bound to point out that a decrease in the process variance, the percentage defective or the average number of defects should be taken seriously as it is a synonym for a quality improvement. Thus, quick detection and identification of such changes are rather important, particularly if we desired to maximize the production of better quality items (Pignatiello et al. 1995), and the adoption of *ARL-unbiased* charts (or an UMPU test, etc.) plays an absolutely crucial role in achieving this while using charts to signal to both decreases and increases in a parameter.

A final note: Several programs for the statistical software system R (refer to the URL http://www.r-project.org/, the R package `spc` (see the URL http://cran.at.r-project.org/web/packages/spc/), and Mathematica (http://reference.wolfram.com/legacy/v5/) were used to produce the graphs and obtain the values in the tables in this paper; programs will be made available to those who request them from the authors.

Acknowledgements The second author gratefully acknowledges the financial support received from CEMAT (Centro de Matemática e Aplicações) to attend the XIth International Workshop on Intelligent Statistical Quality Control, Sydney, Australia, August 20–23, 2013.

Appendix

What follows are some notes on how to obtain the critical values $a_E(\alpha, n)$ and $b_E(\alpha, n)$, for the *equal tails* EWMA-S^2-scheme, and $\tilde{a}_E(\alpha, n)$ and $\tilde{b}_E(\alpha, n)$, for the *ARL-unbiased* EWMA-S^2-chart.

We ought to mention that the EWMA-S^2-charts, described in Sect. 6, are set up taking advantage of the algorithms used to obtain $a_E(\alpha, n)$, $b_E(\alpha, n)$, $\tilde{a}_E(\alpha, n)$, and $\tilde{b}_E(\alpha, n)$.

Numerics for the Equal Tails EWMA-S^2-Scheme

The critical constants, $a_E(\alpha, n)$ and $b_E(\alpha, n)$, should ensure that

$$ARL_\sigma(1) = ARL^{(0)}$$
$$ARL_\sigma^{\text{upper}}(1) = ARL_\sigma^{\text{lower}}(1),$$

where: $ARL^{(0)}$ is the desired on-target ARL (e.g., 500); and $ARL_\sigma^{\text{upper/lower}}(1)$ are the on-target ARL values of the corresponding one-sided charts to be considered for the *equal tails* EWMA-S^2-scheme.

$a_E(\alpha, n)$ and $b_E(\alpha, n)$ are obtained numerically—the algorithm is implemented in the R package `spc`—by applying a two-dimensional secant rule as follows. To start with, initial values for $a_E(\alpha, n)$ and $b_E(\alpha, n)$ are derived for the one-sided charts with on-target ARL equal to $2 \times ARL^{(0)}$. Secondly, a slightly decreased $a_E(\alpha, n)$ and increased $b_E(\alpha, n)$ serve as the second pair to initialize the secant rule iteration process. Then 7 ARL values (two for each one-sided chart, three for the two-sided chart) are calculated to allow the computation of the discrete, two-dimensional, and symmetric Jacobi matrix. Finally, after less than ten iterations the procedure is stopped. The stopping rule is: (i) the above equalities are fulfilled with an error less than 10^{-6}; or (ii) the changes in the constants are smaller than 10^{-8}.

This is the output for $\lambda = 0.1$ and $n = 5$:

```
lower limit upper limit lower ARL upper ARL 2-sided ARL
0.623811 1.541675 1000.0000 1000.0000 491.5735
0.617333 1.567712 1179.1669 1349.0616 620.6845
0.621418 1.543226 1062.0820 1017.7417 511.2787
0.623059 1.543026 1019.0277 1015.4410 500.1801
0.623142 1.543148 1016.9033 1016.8459 500.0018
0.623143 1.543150 1016.8711 1016.8711 500.0000
0.623143 1.543150 1016.8711 1016.8711 500.0000
```

The same algorithm is utilized to determine the control limits of a *equaltailsProb* EWMA-S^2-chart fulfilling the condition $P[RL_\sigma(1) \leq RL^{(0)}] = \alpha$.

Numerics for the ARL-Unbiased EWMA-S^2-Chart

The critical constants $\tilde{a}_E(\alpha, n)$ and $\tilde{b}_E(\alpha, n)$ should yield

$$ARL_\sigma(1) = ARL^{(0)}$$
$$\frac{|ARL_\sigma(1+\varepsilon) - ARL_\sigma(1-\varepsilon)|}{2\varepsilon} \leq 10^{-6},$$

where ε is set to 10^{-4} and the ratio in the second condition is an approximation of the derivative that should be zero at $\theta = 1$.

The numerical algorithm is a nested secant rule. The outer procedure updates the upper control limit until: (i) the magnitude of the difference quotient does not exceed 10^{-6}; or (ii) the change in that control limit is smaller than 10^{-8}. Given a fixed upper control limit, a similar inner procedure searches for a lower control limit that leads to $ARL_\sigma(1) = ARL^{(0)}$. The initial upper control limit is the one from the upper one-sided chart. Then this control limit is increased until the quotient above is positive. The last two values of this loop are used as starting values for the outer secant rule.

The corresponding output is, for $\lambda = 0.1$ and $n = 5$:

```
lower limit upper limit difference quotient
0.000000 1.478111 -15356.622543
0.615248 1.528111 -3457.229820
0.635062 1.578111 3607.692047
0.627064 1.552578 423.304834
0.625727 1.549184 -61.990991
0.625902 1.549618 0.884083
0.625900 1.549612 0.001812
0.625900 1.549612 -0.000000
0.625900 1.549612 0.000000
```

Note that for each line in the above output one inner secant rule is utilized and two further ARL values ($ARL_\sigma(1-\varepsilon)$ and $ARL_\sigma(1+\varepsilon)$) are calculated. Thus, the calculation of the critical constants $\tilde{a}_E(\alpha, n)$ and $\tilde{b}_E(\alpha, n)$ requires more time than the obtention of $a_E(\alpha, n)$ and $b_E(\alpha, n)$ of the *equal tails* EWMA-S^2-chart. The same numerical algorithm is used to set the *unbiasedProb* EWMA-S^2-chart.

References

Acosta-Mejía, C. A., & Pignatiello, J. J. (2000). Monitoring process dispersion without subgrouping. *Journal of Quality Technology, 32*, 89–102.

Barbosa, E. P. P., Gneri, M. A., & Meneguetti, A. (2012). Range control charts revisited: Simpler Tippett-like formulae, its practical implementation, and the study of false alarm. *Communications in Statistics, 42*, 247–262.

Capizzi, G., & Masarotto, G. (2003). An adaptive exponentially weighted moving average control chart. *Technometrics, 45*, 199–207.

Casella, G., & Berger, R.L. (1990). *Statistical inference*. Belmont, CA: Duxberry.

Fertig, J. W., & Proehl, E. A. (1937). A test of a sample variance based on both tail ends of the distribution. *The Annals of Mathematical Statistics, 8*, 193–205.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics: Vol. 2, Inference and relationship (Fourth edition)*. London/High Wycombe: Charles Griffin & Company Limited.

Knoth, S. (2010). Control charting normal variance—reflections, curiosities, and recommendations. In: H. -J. Lenz, & P. -T. Wilrich (Eds.), *Frontiers in Statistical Quality Control* (Vol. 9, pp. 3–18). Heidelberg: Physica.

Krumbholz, W., & Zöller, A. (1995). p-Karten vom Shewhartschen Typ für die messende Prüfung. *Allgemeines Statistisches Archiv, 79*, 347–360.

Lehmann, E. L. (1986). *Testing statistical hypotheses* (2nd edn). Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Lehmann, E. L. (2006). Unbiasedness. In: S. Kotz, N. Balakrishan, C. Read, B. Vidakovic (Eds.), & N. L. Johnson (Founder and Chairman of Ed. Board), *Encyclopedia of Statistical Sciences* (2nd edn, Vol. 14, pp. 8818–8823). New York: Wiley.

MacGregor, J. F., & Harris, T. J. (1993). The exponentially weighted moving variance. *Journal of Quality Technology, 25*, 106–118.

Mittag, H. J., Stemann, D., & Tewes, B. (1998). EWMA-Karten zur Überwachung der Streuung von Qualitätsmerkmalen. *Allgemeines Statistisches Archiv, 82*, 327–338.

Morais, M. J. C. (2002). *Stochastic ordering in the performance analysis of quality control schemes*. Ph.D. thesis, Instituto Superior Técnico, Technical University of Lisbon.

Nelson, L. S. (1982). Control charts. In: S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 2, pp. 176–183). New York: Wiley.

Neyman, J., & Pearson, E. S. (1936). Contributions to the theory of testing statistical hypotheses. *Statistical Research Memoirs, 1*, 1–37.

Neyman, J., & Pearson, E. S. (1938). Contributions to the theory of testing statistical hypotheses. *Statistical Research Memoirs, 2*, 25–57.

Ng, C. H., & Case, K. E. (1989). Development and evaluation of control charts using exponentially weighted moving averages. *Journal of Quality Technology, 21*, 242–250.

Pachares, J. (1961). Tables for unbiased tests on the variance of a normal population. *The Annals of Mathematical Statistics, 32*, 84–87.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika, 41*, 100–115.

Page, E. S. (1955). Control charts with warning lines. *Biometrics, 42*, 243–257.

Pignatiello, J. J., Acosta-Mejía, C. A., & Rao, B. V. (1995). The performance of control charts for monitoring process dispersion. In *4th Industrial Engineering Research Conference* (pp. 320–328).

Ramachandran, K. V. (1958). A test of variances. *Journal of the American Statistical Society, 53*, 741–747.

Ramalhoto, M. F., & Morais, M. (1995). Cartas de controlo para o parâmetro de escala da população Weibull tri-paramétrica. (Control charts for the scale parameter of the Weibull population.) In *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística (Proceedings of the Annual Congress of the Portuguese Statistical Society)* (pp. 345–371).

Ramalhoto, M. F., & Morais, M. (1999). Shewhart control charts for the scale parameter of a Weibull control variable with fixed and variable sampling intervals. *Journal of Applied Statistics, 26*, 129–160.

Roberts, S. W. (1958). Properties of control chart zone tests. *The Bell System Technical Journal, 37*, 83–114.

Roberts, S. W. (1959). Control charts tests based on geometric moving averages. *Technometrics, 1*, 239–250.

Saunders, I. W., & Moran, P. A. P. (1978). On the quantiles of the gamma and F distributions. *Journal of Applied Probability, 15*, 426–432.

Shaked, M., & Shanthikumar, J. G. (1994). *Stochastic orders and their applications*. London: Academic.

Shao, J. (2003). *Mathematical statistics* (2nd edn). New York: Springer.

Sweet, A. L. (1986). Control charts using coupled exponentially weighted moving averages. *IIE Transactions, 18*, 26–33.

Tate, R. F., & Klett, G. W. (1959). Optimal confidence intervals for the variance of a normal distribution. *Journal of the American Statistical Association, 54*, 574–582.

Uhlmann, W. (1982). *Statistische Qualitätskontrolle (2. Aufl.)*. Stuttgart: Teubner.

Western Electrical Company (1956). *Statistical quality control handbook*. Indianopolis, IN: Western Electrical Company.

Wortham, A. W., & Ringer, L. J. (1971). Control via exponential smoothing. *The Logistics Review, 7*, 33–40.

Zhang, L., Bebbington, M. S., Lai, D. C., & Govindaraju, K. (2005). On statistical design of the S^2 chart. *Communications in Statistics, 34*, 229–244.

Optimal Cumulative Sum Charting Procedures Based on Kernel Densities

Jessie Y. Su, Fah Fatt Gan, and Xu Tang

Abstract The cumulative sum (CUSUM) charting procedure is an important online monitoring procedure which is especially effective for detecting small and moderate shifts. In the design and implementation of an optimal CUSUM chart, the probability density function of an in-control process distribution is assumed to be known. If the density is not known or cannot be approximated using a known density, an optimal CUSUM chart cannot be implemented. We propose a CUSUM chart which does not require the density to be known. Kernel density estimation method will be used to estimate the density of an in-control process distribution. The performance of this chart is investigated for unimodal distributions. The results obtained reveal that this chart works well if we can obtain sufficient observations from an in-control process for kernel density and average run length estimations. An example is given to illustrate the design and implementation of this chart.

Keywords Average run length • Collocation method • Integral equation • Sensitivity analysis • Sequential probability ratio test • Statistical process control • Unimodal distribution

1 Introduction

A key tool in statistical process control is the control chart. The first control chart was introduced by Shewhart (1931). Even though the insensitivity of the Shewhart chart in detecting small and moderate shifts of a process parameter was realized from the beginning, the chart has been widely used. Many improvements of the Shewhart chart have been developed. One such improvement is the cumulative

J.Y. Su • F.F. Gan (✉) • X. Tang
Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore
e-mail: jessie.su90@gmail.com; staganff@nus.edu.sg; g0900761@nus.edu.sg

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_8

sum (CUSUM) chart introduced by Page (1954). A CUSUM chart is able to detect small and moderate shifts more quickly. In addition, a shift is usually revealed more prominently on a CUSUM chart.

The CUSUM chart is well researched for a few probability density functions (pdf's) like normal (Page 1961; Goel and Wu 1971; Gan 1991), exponential (Vardeman and Ray 1985; Gan 1994), gamma (Hawkins and Olwell 1998; Huang et al. 2013) and inverse Gaussian (Hawkins and Olwell 1997). The current practice of implementing an optimal CUSUM chart is to examine whether the underlying distribution is one of those already studied. Parametric density estimation method is usually used but this method requires the form of the distribution to be known a priori. If the density is not known or cannot be approximated using a known density, an optimal CUSUM chart cannot be implemented. The distribution of a sample mean can be approximated using a normal distribution because of the central limit theorem. Although a CUSUM chart based on sample mean can be used, Hawkins and Olwell (1998) emphasized the usefulness of a CUSUM chart based on individual observations and stated "…just as the default for Shewhart charts is rational groups with individual readings as exception, so in CUSUM charting individual readings are the norm and rational groups are the exception".

In this paper, we develop a near optimal CUSUM chart without having to specify the density of the in-control process. Our approach is to use a CUSUM chart using the density estimated using a nonparametric method. Nonparametric methods of density estimation do not require any assumption of the form of an underlying density. The estimated density is driven entirely by a given data set. Lehmann (1990) remarked that the nonparametric method is favoured over the parametric method due to its greater flexibility and its insensitivity to specification bias. In this paper, we will use the nonparametric kernel estimation method developed by Rosenblatt (1956), Parzen (1962) and Cencov (1962). Kernel density estimation is also known as Parzen-Rosenblatt window method, although the former nomenclature is more common within the statistical community. Silverman (1986) provided a comprehensive treatment of kernel density estimation. Scott (1992) and Duda et al. (2000) showed that kernel density estimation techniques are powerful methods as their estimators converge to any density function asymptotically and are therefore appropriate and useful for many problems where the forms of the underlying distributions are not known a priori.

In this paper, we examine the CUSUM charting procedure in Sect. 2, kernel density estimation in Sect. 3 and develop the CUSUM charting procedure based on kernel density estimation in Sect. 4. The performance of the proposed procedure is studied in Sect. 5. An example to illustrate the design and implementation of the procedure is given in Sect. 6, and a conclusion is given in Sect. 7.

2 Cumulative Sum Charting Procedure

The CUSUM chart is based on the sequential probability ratio test (SPRT). Wald (1943, 1945) conceptualized and developed the SPRT to increase the efficiency in monitoring the quality of military equipment manufactured during World War II. The SPRT uses information from each observation once it becomes available. In SPRT, we test the density using a simple null hypothesis $H_0 : f(x) = f_0(x)$ against a simple alternative hypothesis $H_1 : f(x) = f_1(x)$. According to Wald (1945), for a sequence of independent observations $X_1, X_2, \ldots, X_n$ from the density $f(x)$, the log-likelihood ratio test statistic is computed sequentially as

$$\log(\Lambda_n) = \sum_{t=1}^{n} W_t, \tag{1}$$

where

$$W_t = W(X_t) = \log\left[\frac{f_1(X_t)}{f_0(X_t)}\right]. \tag{2}$$

When H_0 is true, the W_t's are expected to be negative, shifting the sum in Eq. (1) downwards. Similarly, when H_1 is true, the sum will shift upwards. In SPRT, there are two cutoff constants, A and B, where $A < B$. The SPRT accepts H_0 if $\log(\Lambda_n) \leq \log(A)$ and accepts H_1 if $\log(\Lambda_n) \geq \log(B)$. Otherwise, it does not accept any hypothesis and calls for another observation. The values of A and B are determined by fixing the probabilities of Type I and II errors. Wald hypothesized (1947) and later proved (Wald and Wolfowitz 1948) that the SPRT is optimal in testing these hypotheses.

For process monitoring, we are only interested in signalling when H_1 is accepted because this provides evidence of an out-of-control process. If a process is in-control, no signal shall be given. The CUSUM chart introduced by Page (1954) plots

$$S_t = \max(0, S_{t-1} + W_t) \tag{3}$$

against the time t. The initial value of S_t is chosen such that $0 \leq S_0 < h$ and is often set to be zero. The chart signals when S_t exceeds h where h is commonly known as the control chart limit. The value of h is prespecified based on the in-control average run length (ARL). The time period chosen is arbitrary, although it is best to use the smallest possible time period so that a shift in process parameter can be detected earlier. Moustakides (1986) proved that the CUSUM chart is optimal.

The ARL is a common measure of the performance of a control chart. The run length is the number of observations sampled until a signal is given and the ARL is the expectation of this random variable. The integral equation developed by Page (1954) can be used to approximate the ARL. The integral equation is given as

$$L(z) = 1 + L(0)P(W < -z) + \int_0^h L(x) f_W(x - z)dx, \tag{4}$$

where $L(z)$ is the ARL of a CUSUM chart with an initial value $S_0 = z$ and $f_W(\cdot)$ is the pdf of W. The integral in Eq. (4) can be approximated using the Gauss-Legendre quadratures (Abramowitz and Stegun 1972) to yield

$$L(z) \approx 1 + L(0)P(W < -z) + \sum_{i=1}^{M} w_i L(x_i) f_W(x_i - z), \tag{5}$$

where x_i and w_i, $i = 1, 2, \ldots, M$ are the Gauss-Legendre abscissas, and weights respectively. The system of $M + 1$ linear equations in $L(0)$, $L(x_1)$, …, $L(x_M)$ can then be solved to obtain an approximation to $L(z)$.

The integral equation method only gives accurate results when the interval $[0, h]$ lies within the range of $W(x)$. Knoth (2005, 2006) demonstrated that the collocation method is a more accurate method. The collocation method approximates $L(z)$ by $\sum_{j=1}^{N} c_j T_j(z)$, where c_j's are unknown constants and $T_j(z) = \cos[(j - 1)\arccos((2z - h)/h)]$, $j = 1, \ldots, N$. Let the minimum and maximum of W be u and l, respectively. The constants c_j's can be found by solving the following system of N linear equations

$$\sum_{j=1}^{N} c_j T_j(z_i) = 1 + P(W < -z_i) \sum_{j=1}^{N} c_j T_j(0) + \sum_{j=1}^{N} c_j \int_{l^*}^{u^*} T_j(x) f_W(x - z_i)dx, \tag{6}$$

where $z_i = h[1 + \cos\{(2i - 1)\pi/(2N)\}]/2, i = 1, \ldots, N$. The lower limit $l^* = 0$ if $0 \geq l + z_i$ and $l^* = l + z_i$ if $0 < l + z_i$, while $u^* = h$ if $h \leq u + z_i$ and $u^* = u + z_i$ if $h > u + z_i$. The integral on the right can be approximated using the Gauss-Legendre quadratures.

3 Kernel Density Estimation

Let $\{X_1, \ldots, X_m\}$ be a set of independent and identically distributed univariate random variables sampled from $f(x)$. According to Silverman (1986), the kernel density estimation method estimates $f(x)$ using the formula

$$\hat{f}(x) = \frac{1}{md} \sum_{i=1}^{m} K\Big(\frac{x - X_i}{d}\Big), \tag{7}$$

where $K(\cdot)$ is a kernel function and d is the bandwidth. The kernel function is usually unimodal and symmetrical around the origin. The bandwidth d controls the degree of smoothness of an estimated density. Much attention on kernel density estimation has been on the computation of an optimal d. The choice of d is usually based on the minimization of the discrepancy between an estimated density and its actual density. The density estimate $\hat{f}$ depends on the observations $X_1, X_2, \ldots, X_m$ and as such is a random variable. This dependence will not be expressed explicitly. According to Silverman (1986), when considering estimation at a single point x, the discrepancy of the density estimate $\hat{f}(x)$ from the true density $f(x)$ can be measured by the mean squared error which is defined as

$$\text{MSE}_x(\hat{f}) = E[\hat{f}(x) - f(x)]^2, \tag{8}$$

where E denotes the expected value with respect to the sample. When measuring the global discrepancy of $\hat{f}$ from f, the mean integrated squared error developed by Rosenblatt (1956) is commonly used. It is defined as

$$\text{MISE}(\hat{f}) = E\Big(\int [\hat{f}(x) - f(x)]^2 dx\Big). \tag{9}$$

Silverman (1986) emphasized the importance of choosing a suitable bandwidth d as the bias of $\hat{f}(x)$ depends directly on d. He suggested that when estimating a normal distribution using a Gaussian kernel, the following bandwidth can be used,

$$d_1 = 1.06\sigma m^{-1/5}, \tag{10}$$

where σ can be estimated from the given observations. While Eq. (10) may work well for normal or unimodal distributions, it may over smooth the density of a multimodal distribution. For skewed and long-tailed distributions, Silverman (1986) suggested the following bandwidth,

$$d_2 = 0.79Rm^{-1/5}, \tag{11}$$

where R represents the interquartile range of the distribution to be estimated. Although Eq. (11) takes a more flexible measure of spread into account, it over-smooths the density of a multimodal distribution more than the bandwidth in Eq. (10). A bandwidth that combines the characteristics of Eqs. (10) and (11) is defined as

$$d_3 = 0.9Am^{-1/5}, \tag{12}$$

where

$$A = \min\left(\sigma, \frac{R}{1.34}\right), \tag{13}$$

where σ and R can be estimated from the given observations. Equation (12) is also known as the Silverman's reference bandwidth. It is shown by Silverman (1986) that Eq. (12) can be used for the estimation of a variety of densities. Therefore, Eq. (12) will be used in this paper.

Scott (1992) noted that the choice of a kernel has little effect on the MISE. Hence, without undue concern for the loss of efficiency, a kernel can be chosen for other reasons such as the ease of computation. Silverman (1986) noted that an estimated density will inherit all continuity and differentiability properties from its kernel. In this paper, the standard normal kernel will be used due to its symmetry, smoothness and infinite support. The ease in estimating both pdf and cumulative distribution function (cdf) due to its differentiability is also why the Gaussian kernel is chosen instead of the Epanechnikov kernel despite the latter's higher efficiency.

The accuracy of an estimated density also depends on the number of observations m. The accuracy, especially at the extreme points of an estimated density, generally increases with an increasing number of observations. Silverman (1986) explained that the bias in estimating $f(x)$ depends indirectly on m through its dependence on d. The effects of m on density estimation and hence ARL estimation will be investigated in Sect. 5. To summarize this section, the following formulae will be used for the estimation of pdf and cdf

$$\hat{f}(x) = \frac{1}{md}\sum_{i=1}^{m}\phi\left(\frac{x - X_i}{d}\right), \tag{14}$$

$$\hat{F}(x) = \frac{1}{m}\sum_{i=1}^{m}\Phi\left(\frac{x - X_i}{d}\right), \tag{15}$$

where d is defined in Eq. (12) and $\phi(\cdot)$ and $\Phi(\cdot)$ represent the standard normal pdf and cdf, respectively.

4 Cumulative Sum Chart Using Kernel Densities

In order to implement a CUSUM chart, Eqs. (2) and (3) show that the monitoring statistic $W(x)$ must be known and it is a function of $f_0(x)$ and $f_1(x)$. If the densities are not known, an optimal CUSUM chart cannot be implemented. To overcome this problem, we will develop a CUSUM chart using kernel densities. A closely related problem is that of setting an appropriate chart limit when the parameters and in-control distribution are estimated. By bootstrapping the data used to estimate the in-control state, Gandy and Kvaløy (2013) propose an adjustment to a charting procedure such that the procedure will have a high probability of a certain conditional performance given the estimated in-control state.

Suppose we have a sample of observations $X_1, \ldots, X_m$ taken from an unknown in-control distribution $f_0(x)$. The in-control pdf can be estimated using the kernel density in Eq. (14). Consider a CUSUM chart for detecting an upward shift in process mean. In the design of a CUSUM chart, the shift that is most important to be detected quickly must also be specified. Without loss of generality, we will let this be a one σ shift in the mean,

$$f_1(x) = f_0(x - \sigma), \tag{16}$$

where σ is the standard deviation of $f_0(x)$. When $\hat{f}_0(x)$ is estimated using a kernel density, the corresponding $\hat{f}_1(x)$ can be estimated using

$$\hat{f}_1(x) = \hat{f}_0(x - s), \tag{17}$$

where s is an estimate of σ based on the m observations. The monitoring statistic can then be estimated as

$$\hat{W}(x) = \log\left(\frac{\hat{f}_1(x)}{\hat{f}_0(x)}\right). \tag{18}$$

Hence, using $\hat{W}(x)$ in Eq. (3), we can implement a CUSUM chart.

To complete the design of this CUSUM chart, we will also need to determine the control limit h for a given ARL. Unfortunately, the ARL of this chart is mathematically intractable and we propose estimating the in-control ARL using simulation based on sampling with replacement from the same historical data set. This means a sufficiently large in-control historical data set is required for estimating the ARL accurately. This will be investigated in the next section.

The CUSUM chart using kernel densities is highly flexible because it does not require the in-control pdf to be known. How well this chart performs when compared to the CUSUM chart using known $f_0(\cdot)$ and $f_1(\cdot)$ depends on how well $W(x)$

is estimated. Since $\hat{f}_1(x)$ is derived from $\hat{f}_0(x)$, this ultimately depends on the accuracy of $\hat{f}_0(x)$ obtained. As stated in Sect. 3, accuracy of $\hat{f}_0(x)$ depends on a few factors such as the kernel chosen, bandwidth and the size of historical sample. With the recommended kernel function and bandwidth, we will investigate the effect of the size of historical sample on the performance of CUSUM chart in the next section.

5 Performance of CUSUM Chart Using Kernel Densities

We are interested in developing a CUSUM chart for monitoring a unimodal distribution because many quality measures follow such a distribution. We will investigate the performance of the CUSUM chart using kernel densities for the normal distribution and the t-distribution with 10 degrees of freedom. The latter distribution is chosen because it represents a distribution that is more spread out than the normal.

We will first consider the normal distribution. Consider the hypotheses $H_0 : f(x) = f_0(x) \sim N(0, 1)$ and $H_1 : f(x) = f_1(x) \sim N(1, 1)$. Using Eq. (2), $W(x)$ can be shown to follow the distribution $N(-1/2, 1)$. The range of $W(x)$ is $(-\infty, \infty)$. Hence, the integral equation method can be used to obtain accurate ARL for the CUSUM chart using known pdf since $[0, h]$ is a subset of this range. The collocation method also yields accurate ARL. Asymptotically, the ARL of the CUSUM chart using kernel densities will approach that of the CUSUM chart using known pdf's as the sample size m increases. This is because $\hat{f}_0(\cdot)$ converges to $f_0(\cdot)$ as m increases. Table 1 contains the ARLs of CUSUM charts using kernel densities estimated using historical data of size $m = 500$, 1,000, 3,000, 5,000 and 10,000. The chart limits of these charts were approximated using a simple method as described in Sect. 6. An ARL in the first row is simulated using observations based on sampling with replacement from the same historical data set used to estimate the kernel densities. The following rows are the in-control and out-of-control ARLs simulated using observations from the true process densities. The values in parentheses refer to the standard deviations of the respective ARLs obtained.

Table 1 shows that a CUSUM chart based on a historical data set of size $m =$ 5,000 is adequate to produce an in-control ARL that is close to 100. In addition, the out-of-control run length profiles of CUSUM charts using kernel densities based on $m =$ 5,000 and 10,000 are similar to those of the CUSUM charts using known densities. This means that the CUSUM chart using kernel densities has near optimal run length properties for m at least 5,000. A sensitivity analysis is also done by replicating the results three more times for $m =$ 5,000 and 10,000. These results, together with the ones obtained in Table 1, are displayed in Table 2. The results are consistent across the samples. In order to investigate the performance of the

Table 1 ARLs of CUSUM charts using Kernel densities for monitoring the normal distribution

m	500	1,000	3,000	5,000	10,000	Exact
h	2.626	2.604	2.712	2.748	2.767	2.849
Shift						
0.00	*100.1	99.9	100.6	99.6	100.2	
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	
0.00	**53.5	86.7	77.1	99.3	100.9	100.0
	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	
0.10	37.2	57.4	51.3	63.8	64.7	64.1
	(0.11)	(0.17)	(0.15)	(0.20)	(0.20)	
0.20	26.5	39.2	35.2	42.7	43.2	42.7
	(0.07)	(0.11)	(0.10)	(0.12)	(0.13)	
0.30	19.8	27.7	25.2	29.7	30.0	30.0
	(0.05)	(0.07)	(0.07)	(0.08)	(0.08)	
0.40	15.1	20.2	18.8	21.5	21.7	21.4
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	
0.50	11.8	15.5	14.4	16.1	16.3	16.1
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	
0.75	7.3	8.8	8.4	9.2	9.2	9.2
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
1.00	5.1	5.9	5.7	6.2	6.1	6.1
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
1.25	3.8	4.4	4.3	4.6	4.5	4.5
	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	
1.50	3.1	3.5	3.4	3.7	3.6	3.6
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	
2.00	2.2	2.5	2.5	2.5	2.5	2.6
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	
2.50	1.9	2.1	2.0	2.1	1.9	2.0
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
3.00	1.5	1.8	1.8	1.7	1.5	1.7
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	

*An ARL in the first row is simulated using observations based on sampling with replacement from the same historical data set

**An ARL from this row onwards is simulated using observations from the true process density

CUSUM chart for other unimodal distributions, we consider $H_0 : f_0(x) \sim t_\nu$, t-distribution with $\nu = 10$ degrees of freedom and $H_1 : f_1(x) = f_0(x - \sigma)$. Using Eq. (2), it can be shown that $W(x) = -[(\nu+1)/2] \log[\{\nu+(x-\sigma)^2\}/(\nu+x^2)]$. With the pdf and cdf of $W(X)$ derived, the exact ARL of the CUSUM chart using known densities can be approximated accurately using the collocation method. Table 3 contains the ARLs of the CUSUM charts using kernel densities for monitoring the

Table 2 Sensitivity analysis for the CUSUM charts using Kernel densities for monitoring the normal distribution

m	5,000	5,000	5,000	5,000	10,000	10,000	10,000	10,000
h	2.748	2.771	2.771	2.752	2.767	2.780	2.798	2.760
Shift								
0.00	*100.3	99.7	99.5	99.4	100.2	100.4	100.0	99.9
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
0.00	**99.3	102.3	101.9	84.5	100.9	99.5	102.3	98.1
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
0.10	63.8	65.1	65.3	55.3	64.7	64.2	65.6	63.5
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
0.20	42.7	43.4	43.7	37.5	43.2	42.7	43.0	42.5
	(0.12)	(0.13)	(0.13)	(0.11)	(0.13)	(0.13)	(0.13)	(0.12)
0.30	29.7	30.2	30.1	26.4	30.0	29.8	29.8	29.7
	(0.08)	(0.08)	(0.09)	(0.07)	(0.08)	(0.08)	(0.09)	(0.09)
0.40	21.5	21.8	21.8	19.4	21.7	21.5	21.6	21.5
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
0.50	16.1	16.4	16.3	14.8	16.3	16.1	16.2	16.2
	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
0.75	9.2	9.3	9.3	8.6	9.2	9.2	9.2	9.1
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
1.00	6.2	6.2	6.2	5.9	6.1	6.1	6.2	6.1
	(0.010)	(0.012)	(0.012)	(0.011)	(0.010)	(0.012)	(0.012)	(0.012)
1.25	4.6	4.6	4.6	4.4	4.5	4.5	4.6	4.5
	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
1.50	3.7	3.7	3.7	3.5	3.6	3.6	3.6	3.5
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
2.00	2.5	2.7	2.7	2.6	2.5	2.6	2.6	2.5
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
2.50	2.2	2.2	2.2	2.1	1.9	2.0	2.1	2.0
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
3.00	1.7	1.9	1.9	1.7	1.5	1.6	1.8	1.6
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

*An ARL in the first row is simulated using observations based on sampling with replacement from the same historical data set

**An ARL from this row onwards is simulated using observations from the true process density

t-distribution. The table shows that a CUSUM chart based on $m = 5{,}000$ is adequate to produce an in-control ARL that is close to 100. In addition, the out-of-control run length profiles of CUSUM charts using kernel densities based on $m = 5{,}000$ and 10,000 are similar to those of the CUSUM charts using known densities. A similar

Table 3 ARLs of CUSUM charts using Kernel densities for monitoring the t-distribution

m	500	1,000	3,000	5,000	10,000	Exact
h	2.7287	2.7960	2.8114	2.8483	2.9142	2.9920
Shift						
0.00	*99.1	100.2	99.8	100.1	100.1	
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	
0.00	**69.0	96.3	92.2	111.2	97.5	100.0
	(0.2)	(0.3)	(0.3)	(0.3)	(0.3)	
0.10	49.4	61.9	61.5	72.1	63.1	68.0
	(0.14)	(0.18)	(0.18)	(0.21)	(0.18)	
0.20	35.5	40.6	41.3	48.2	42.4	47.2
	(0.10)	(0.11)	(0.12)	(0.14)	(0.12)	
0.30	26.0	28.0	29.1	32.9	29.3	33.7
	(0.07)	(0.08)	(0.08)	(0.09)	(0.08)	
0.40	19.2	20.0	21.0	23.4	21.0	24.7
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	
0.50	14.7	14.8	15.7	17.4	15.7	18.7
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	
0.75	8.5	8.4	8.9	9.5	9.9	10.6
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
1.00	5.7	5.6	5.9	6.2	5.9	7.0
	(0.011)	(0.010)	(0.011)	(0.012)	(0.011)	
1.25	4.3	4.2	4.4	4.6	4.5	5.2
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	
1.50	3.5	3.5	3.6	3.7	3.6	4.1
	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	
2.00	2.7	2.6	2.7	2.8	2.7	3.0
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	
2.50	2.4	2.3	2.3	2.4	2.3	2.5
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	
3.00	2.4	2.1	2.2	2.3	2.2	2.2
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	

*An ARL in the first row is simulated using observations based on sampling with replacement from the same historical data set

**An ARL from this row onwards is simulated using observations from the true process density

sensitivity analysis is done and the results are displayed in Table 4. For $m = 5{,}000$, the estimated ARL using the historical data set tends to underestimate the true ARL by about 10 %. The estimation improves when $m = 10{,}000$.

Table 4 Sensitivity analysis for the CUSUM charts using Kernel densities for monitoring the t-distribution

m	5,000	5,000	5,000	5,000	10,000	10,000	10,000	10,000
h	2.8483	2.8890	2.8483	2.8758	2.9162	2.8729	2.9151	2.8992
Shift								
0.00	*100.1	100.2	100.1	100.1	100.1	100.4	99.7	99.9
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
0.00	**111.2	115.3	110.8	110.5	97.5	108.1	104.4	96.2
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
0.10	72.1	73.0	72.0	71.5	63.1	70.0	66.3	62.2
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
0.20	48.2	47.8	48.0	47.3	42.4	46.7	44.0	41.9
	(0.14)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)
0.30	32.9	32.2	33.2	32.2	29.3	31.9	30.2	28.9
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.08)
0.40	23.4	22.6	23.4	22.9	21.0	22.8	21.6	21.0
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)
0.50	17.4	16.7	17.3	17.0	15.7	16.9	16.1	15.6
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
0.75	9.5	9.1	9.5	9.4	8.9	9.3	9.0	8.8
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
1.00	6.2	6.1	6.2	6.2	5.9	6.1	6.0	5.9
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
1.25	4.6	4.5	4.6	4.6	4.5	4.5	4.6	4.5
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
1.50	3.7	3.7	3.7	3.7	3.6	3.6	3.7	3.6
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
2.00	2.8	2.8	2.8	2.8	2.7	2.7	2.8	2.7
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
2.50	2.4	2.3	2.4	2.3	2.3	2.3	2.4	2.3
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
3.00	2.3	2.1	2.2	2.1	2.1	2.1	2.1	2.1
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

*An ARL in the first row is simulated using observations based on sampling with replacement from the same historical data set

**An ARL from this row onwards is simulated using observations from the true process density

6 Application

We illustrate the implementation of the CUSUM chart using kernel densities in this section. We simulate an in-control data set of size 5,000 from $N(\mu = 74, \sigma = 0.01)$ assumed to be the in-control process distribution of the diameter of a piston ring. Kernel density estimation is used to obtain $\hat{f}_0(x)$ given in Eq. (14). The bandwidth

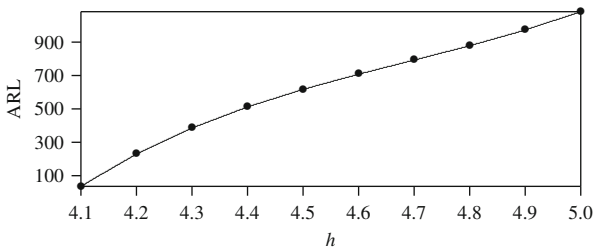


Fig. 1 Plot of simulated ARL against control chart limit h and the fitted cubic regression line

stated in Eq. (12) is used. For this application, it is decided that one sigma shift is most important to be detected. For detecting an upward shift in mean, $\hat{f}_1(x)$ and $\hat{W}(x)$ are obtained using Eqs. (17) and (18), respectively. For detecting a downward shift, we set $\hat{f}_1(x) = \hat{f}_0(x + s)$. To obtain an estimated in-control ARL for a particular control limit h, the run length of the CUSUM chart is repeated 100,000 times using observations randomly selected with replacement from the in-control data set. The average of these 100,000 run lengths is an estimated in-control ARL. This is repeated for several values of h. To estimate the value of h for a specified in-control ARL, a plot of estimated in-control ARL against h is shown in Fig. 1. We use a cubic function to fit the points and obtained

$$\text{ARL} = -132379.2 + 83568.8h - 17653.8h^2 + 1255.7h^3. \tag{19}$$

Using Eq. (19), an approximate control limit for a one-sided CUSUM chart can be determined approximately as $h = 4.64$ for an in-control ARL of 740. A combined CUSUM scheme consisting of two one-sided charts will have an approximate ARL of 370. To illustrate the implementation, the two CUSUM charts are plotted for a data set for which the first 50 observations are generated from the in-control distribution $N(74, 0.01)$ and the last 50 observations are generated from an out-of-control distribution $N(74.008, 0.01)$ which represents a shift of 0.008 in the mean. The charts are displayed as charts (a) and (b) in Fig. 2. Both charts do not show any unusual fluctuations for the first half of the data when the process is in-control. For the second half of the data, the CUSUM chart for detecting upward shift shows more fluctuations and signals at 68th, 89th and 98th observations. The other CUSUM chart becomes much quieter. In order to compare the CUSUM chart using kernel densities and the CUSUM chart using known densities $f_0(\cdot)$ and $f_1(\cdot)$, we plot the latter as the charts (c) and (d) in Fig. 2. Note that the CUSUM charts using kernel densities are almost identical to the CUSUM charts using actual densities. This shows that there is little practical difference between the two types of charts.

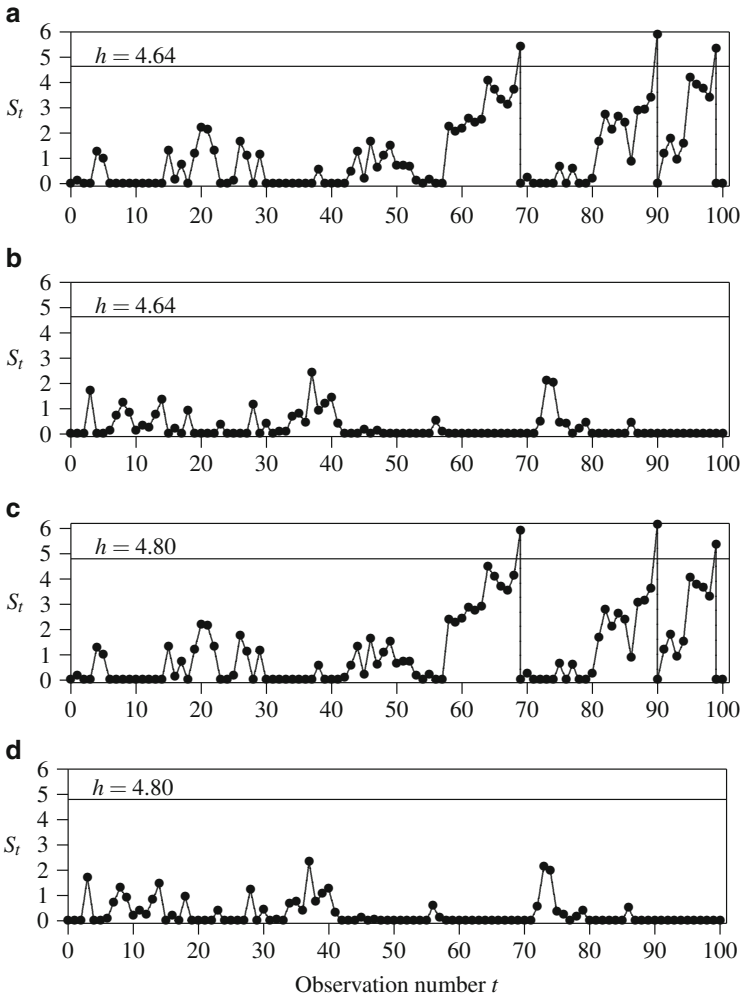


Fig. 2 CUSUM charts using kernel and actual densities for detecting a shift in process mean where the the first 50 observations are generated from the in-control distribution $N(74, 0.01)$ and the last 50 observations are generated from an out-of-control distribution $N(74.008, 0.01)$ which represents a shift of 0.008 in the mean. (**a**) CUSUM chart using Kernel densities to detect upward shift. (**b**) CUSUM chart using Kernel densities to detect downward shift. (**c**) CUSUM chart using actual densities to detect upward shift. (**d**) CUSUM chart using actual densities to detect downward shift

7 Conclusion

In this paper, CUSUM chart using kernel densities is developed for monitoring a unimodal distribution without having to specify the process density. This is a big advantage in application because current implementation of an optimal CUSUM

chart requires the in-control process density to be known and this may not be feasible for many processes. In addition, many process densities are assumed to have infinite support and this is not realistic in practice. Thus, the chart using kernel densities is even more appropriate than the chart using known densities for practical use. Our investigation shows that the chart requires an in-control data set of at least 5000 observations for adequate density and ARL estimations. The chart can be expected to have run length properties similar to the optimal CUSUM chart using known densities if sufficient historical data are available for estimation of the densities.

Acknowledgements The second and third authors are supported by the Academic Research Fund Tier 1 (R-155-000-137-112), Ministry of Education, Singapore.

References

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover Publications.

Cencov, N. N. (1962). Evaluation of an unknown density from observations. *Soviet Mathematics, 3*, 1559–1562.

Duda, R. O., Stork, D. G., & Hart, P. E. (2000). *Pattern classification*. New York: Wiley.

Gan, F. F. (1991). An optimal design of CUSUM quality control charts. *Journal of Quality Technology, 23*, 279–286.

Gan, F. F. (1994). Design of optimal exponential CUSUM control charts. *Journal of Quality Technology, 26*, 109–124.

Gandy, A., & Kvaløy, J. T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics*. doi:10.1002/sjos.12006.

Goel, A. L., & Wu, S. M. (1971). Determination of A.R.L. and a contour nomogram for CUSUM charts to control normal mean. *Technometrics, 13*, 221–230.

Hawkins, D. M., & Olwell, D. H. (1997). Inverse Gaussian cumulative sum control charts for location and shape. *Journal of the Royal Statistical Society: Series D (The Statistician), 46*, 323–335.

Hawkins, D. M., & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. New York: Springer.

Huang, W., Shu, L., Jiang, W., & Tsui, K. -L. (2013). Evaluation of run-length distribution for CUSUM charts under gamma distributions. *IIE Transaction, 45*, 981–994.

Knoth, S. (2005). Accurate ARL computation for EWMA-S^2 control charts. *Statistics and Computing, 15*, 341–352.

Knoth, S. (2006). Computation of the ARL for CUSUM-S^2 schemes. *Computational Statistics & Data Analysis, 51*, 499–512.

Lehmann, E. L. (1990). Model specifiction: the views of Fisher and Neyman, and later developments. *Statistical Science, 5*, 160–168.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distribution. *The Annals of Statistics, 14*, 1379–1387.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika, 41*, 100–115.

Page, E. S. (1961). Cumulative sum charts. *Technometrics, 3*, 1–9.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*, 1065–1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistic, 27*, 832–837.

Scott, D. W. (1992). *Multivariate density estimation*. New York: Wiley-Interscience.
Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. New York: D. Van Nostrand.
Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.
Vardeman, S., & Ray, D. (1985). Average run lengths for CUSUM schemes when observations are exponentially distributed. *Technometrics, 27*, 145–150.
Wald, A. (1943). *Sequential analysis of statistical data: Theory*. A Report Submitted by the Statistical Research Group, Columbia University to the Applied Mathematics Panel, National Defense Research Committee, Sept 1943.
Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics, 16*, 117–186.
Wald, A. (1947). *Sequential analysis*. New York: Wiley.
Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics, 19*, 326–339.

A Simple Approach for Monitoring Process Mean and Variance Simultaneously

Su-Fen Yang and Barry C. Arnold

Abstract Control charts are effective tools for signal detection in both manufacturing processes and service processes. Much of the data in service industries comes from a process having non-normal or unknown distributions. The commonly used Shewhart variable control charts, which depend heavily on the normality assumption, are not appropriately used here. In this paper, we propose a new EWMA-V Chart and EWMA-M Chart based on two simple independent statistics to monitor process mean and variance shifts simultaneously. Further, we explore the sampling properties of the new monitoring statistics, and calculate the average run lengths when using both of the proposed EWMA Charts. A numerical example involving non-normal service times from the service system of a bank branch in Taiwan is used to illustrate the applications of the new EWMA-M and EWMA-V Charts, and to compare them with the existing mean and variance (or standard deviation) charts. The proposed new EWMA-M and EWMA-V Charts show superior detection performance compared to the existing mean and variance charts. The new EWMA-M and EWMA-V Charts are thus recommended.

Keywords Arcsine transformation • EWMA charts

1 Introduction

Control charts are commonly used tools in process signal detection to improve the quality of manufacturing processes and service processes. In the past few years, more and more statistical process control techniques are applied to the service industry, and control charts are also becoming an effective tool in improving service quality. There have been a few studies in this area, like those of Tsung et al. (2008) and Ning et al. (2009). Much service process data comes from processes

S.-F. Yang (✉)
National Chengchi University, Taipei, Taiwan
e-mail: yang@nccu.edu.tw

B.C. Arnold
University of California, Riverside, CA 92521, USA
e-mail: barry.arnold@ucr.edu

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_9

with variables having non-normal or unknown distributions so the commonly used Shewhart variables control charts, which depend on the normality assumption, are not suitable. Hence the question arises: "How to monitor a process with non-normal or unknown distribution data?" Some research has been done to deal with such a situation; see, for example, Ferrell (1953), Amin et al. (1995), Chakraborti et al. (2001), Altukife (2003), Bakir (2004, 2006), Li et al. (2010), Zou and Tsung (2010), and Graham et al. (2011). Little research has been done to deal with process variability monitoring; see, for example, Das and Bhattacharya (2008) and Jones-Farmer and Champ (2010).

A major drawback of the previous nonparametric approaches is that they are not easy for practitioners to apply because they are not statisticians and do not quite understand the proper way to implement the schemes. Yang et al. (2011) proposed a new Sign Chart for variables data to monitor the deviation of the process measurement from the target without the assumption of a normal process distribution or a distribution of known form. Yang and Cheng (2011) proposed a CUSUM Mean Chart to monitor small shifts in the process mean. Yang et al. (2012) addressed a new Mean Chart based on a simple statistic to monitor the shifts of the process mean. Their approaches are quite easy to use, and even easier than some of the above published nonparametric approaches. However, Yang and Cheng (2011), and Yang et al. (2011, 2012) did not consider a variance chart.

In this paper, we propose using both an EWMA-V Chart and an EWMA-M Chart for variables data to monitor the process variance and mean, extending Yang et al.'s approach (2012). The approach is still quite easy to use, and has better detection ability than the existing standard deviation and mean charts. The paper is organized as follows: In Sect. 2, we describe the EWMA-M Chart, and illustrate its detection performance. In Sect. 3, we discuss the construction of a newly proposed EWMA-V Chart, and measure its detection performance. In Sect. 4, we measure the detection performance of using both the new EWMA-V and EWMA-M Charts simultaneously. In Sect. 5, we describe the estimates for unknown population parameters. In Sect. 6, a numerical example of a service system in a bank branch was used to construct the proposed new EWMA-V and EWMA-M Charts to monitor the quality of service time, and their performance compared with those of some existing charts. Section 7 summarizes the findings and provides a recommendation.

2 The Proposed EWMA-M Chart

Assume that a critical quality characteristic, X, has a mean μ and variance σ^2. Following Yang and Cheng (2011), let $Y = X - \mu$ and $p = P(Y > 0)$ = the "Process Proportion." If the process were in-control, then $p = p_{m0}$, and if the process were out-of-control, that is if μ had shifted, then $p = p_{m1} \neq p_{m0}$. If p_{m0} is not given, it will be estimated using a preliminary data set.

To monitor the process mean, a random sample of size n_1, $X_1, X_2, \ldots, X_{n_1}$, is taken from X. Define

$$Y_j = X_j - \mu \quad \text{and} \quad I_j = \begin{cases} 1, & \text{if } Y_j > 0, \\ 0, & \text{otherwise,} \end{cases}, \; j = 1, 2, \ldots, n_1 . \tag{1}$$

Let M_t be the total number of $Y_j > 0$ at time t, then $M_t = \sum_{j=1}^{n_1} I_j$ would follow a binomial distribution with parameters (n_1, p_{m0}) for an in-control process.

2.1 The Control Limits of EWMA-M Chart

Monitoring the process mean shifts is equivalent to monitoring the changes in process proportion. However, the binomial distribution is discrete and is asymmetric for $p_{m0} \neq 0.5$. In addition the values of out-of-control average run length (ARL) (ARL_{M1}) of the M_t Chart do not change inversely with sample size as they normally should (see Yang et al. 2012). To rectify this problem, they propose an "arcsine transformed EWMA-M chart." Each of these EWMA charts has the usual value of 370 for in-control ARL, and they are sensitive for monitoring small shifts in the process mean quickly and effectively.

Let $T_M = \sin^{-1}\left(\sqrt{M/n_1}\right)$, then the distribution of T_M is approximately normal with a mean $\sin^{-1}\left(\sqrt{p_M}\right)$ and variance $1/(4n_1)$ (see Mosteller and Youtz 1961). The EWMA-M statistic is

$$\text{EWMA}_{T_{M_t}} = \lambda_1 T_{M_t} + (1 - \lambda_1)\text{EWMA}_{T_{M_{t-1}}}, \; 0 < \lambda_1 \leq 1, \; t = 1, 2, \ldots . \tag{2}$$

The New EWMA-M chart is constructed as follows. Define

$$\text{UCL}_1 = \sin^{-1}\left(\sqrt{p_{m0}}\right) + L_1\sqrt{\frac{\lambda_1}{4n_1(2-\lambda_1)}}, \tag{3}$$

$$\text{CL}_1 = \sin^{-1}\left(\sqrt{p_{m0}}\right), \tag{4}$$

$$\text{LCL}_1 = \sin^{-1}\left(\sqrt{p_{m0}}\right) - L_1\sqrt{\frac{\lambda_1}{4n_1(2-\lambda_1)}}, \tag{5}$$

and plot $\text{EWMA}_{T_{M_t}}$. If any $\text{EWMA}_{T_{M_t}} \leq \text{LCL}_1$ or $\text{EWMA}_{T_{M_t}} \geq \text{UCL}_1$, an out-of-control signal is issued. The two parameters, L_1 and λ_1, are chosen to yield an in-control ARL, $\text{ARL}_{T_{M0}} = 370$, using the Markov chain approach proposed by Lucas and Saccucci (1990).

Table 1 The $\mathrm{ARL}_{T_{M1}}$ of the EWMA-M chart when $p_{m0} = 0.5$

	p_{m1}				
n_1	0.30	0.35	0.40	0.45	0.50
8	7.6	12.7	28.0	100.4	370.5
9	7.0	11.5	25.0	91.4	370.5
10	6.4	10.5	22.5	83.8	370.5
11	6.0	9.6	20.6	77.3	370.5
12	5.6	9.0	18.9	71.7	370.5
13	5.3	8.4	17.5	66.8	370.5
14	5.0	7.9	16.3	62.5	370.5
15	4.8	7.5	15.3	58.7	370.5
16	4.6	7.1	14.4	55.3	370.5
17	4.4	6.8	13.6	52.2	370.5
18	4.3	6.5	12.9	49.5	370.5
19	4.1	6.2	12.3	47.0	370.5
20	4.0	6.0	11.8	44.8	370.5

2.2 The In-Control and Out-of-Control Average Run Lengths of the EWMA-M Chart

The $\mathrm{ARL}_{T_{M0}}$ of the new EWMA-M chart is a function of (n_1, L_1, λ_1). Adopting the in-control process proportion $p_{m0} = 0.5$, $\mathrm{ARL}_{T_{M0}} \approx 370$ with $\lambda_1 = 0.2$ and $L_1 = 2.86$, the out-of-control ARL, $\mathrm{ARL}_{T_{M1}}$, of the EWMA-M chart is listed for $n_1 = 8(1)20$ and $p_{m1} = 0.30(0.05)0.50$ in Table 1. From Table 1, we found that $\mathrm{ARL}_{T_{M1}}$ decreases when the out-of-control value of p_{m1} is far away from the in-control value of p_{m0}, and when n_1 increases.

3 The Proposed EWMA-V Chart

To monitor the process variance, another random sample of size n_2, $X_1, X_2, \ldots, X_{n_2}$, is taken from the process, X. Assume that the sample size n_2 is even for convenience (if not, delete one observation). Define

$$Y_1^* = (X_2 - X_1)^2/2\,,\ Y_2^* = (X_4 - X_3)^2/2\,,\ \ldots\,,\ Y_{\frac{n_2}{2}}^* = (X_{n_2} - X_{n_2-1})^2/2\,, \tag{6}$$

$$E(Y_{j'}^*) = \sigma^2 \quad,\ j' = 1, 2, \ldots, n_2/2\,, \text{ and} \tag{7}$$

$$I_{j'} = \begin{cases} 1\,, & \text{if } Y_{j'}^* > \sigma^2 \\ 0\,, & \text{otherwise} \end{cases}, \quad j' = 1, 2, \ldots, n_2/2\,. \tag{8}$$

Let V be the total number of $Y_{j'}^* > \sigma^2$, then $V = \sum_{j'=1}^{0.5n_2} I_{j'}$ will have a binomial distribution with parameters $(0.5n_2,\ p_{v0})$ for an in-control process, where $p_{v0} = P(Y_{j'}^* > \sigma^2)$. The value of p_{v0} will depend on the distribution of the X_i's. For example, if the X_i's are normally distributed, then $p_{v0} = P(Y_{j'}^* > \sigma^2) = P(Z^2 > 1)$ where $Z \sim N(0, 1)$. Thus in this case $p_{v0} = 0.3147$. If the distribution of $X_{n_2} - X_{n_2-1}$ is unimodal, as it frequently is, the version of the Chebychev inequality for unimodal variables implies that the quantity p_{v0} is bounded above by 4/9. The value of p_{v0} can be arbitrarily small but it usually will be in the range 0.25–0.50.

Similar to the M_t Chart, the V_t Chart is a new chart in that the binomial variable is not the count of nonconforming units in the sample but rather the number of pairs of X values in a sample that are in-control with respect to the process variance. Monitoring process variance shifts is equivalent to monitoring the changes in process proportion, p_{v0}. The V_t Chart is also asymmetric for $p_{v0} \neq 0.5$, and the values of out-of-control ARL (ARL_{V1}) of the V_t Chart do not change inversely with sample size as they normally should. Hence, we propose an "arcsine transformed EWMA-V chart." Each of these EWMA charts has the usual value of 370 for in-control ARL, and they are sensitive for monitoring small shifts in the process variance quickly and effectively.

3.1 The Control Limits of EWMA-V Chart

Let $T_V = \sin^{-1}\left(\sqrt{V/0.5n_2}\right)$, then the distribution of T_V is approximately normal with a mean $\sin^{-1}\left(\sqrt{p_V}\right)$ and variance $1/(2n_2)$ (see Mosteller and Youtz 1961). We define the New EWMA-V statistic as:

$$\text{EWMA}_{T_{V_t}} = \lambda_2 T_{V_t} + (1 - \lambda_2)\text{EWMA}_{T_{V_{t-1}}}, \quad 0 < \lambda_2 \leq 1\,. \tag{9}$$

Analogous to the derivation of the arcsine transformed EWMA-M chart, we can construct the new EWMA-V chart as follows. Define

$$\text{UCL}_2 = \sin^{-1}\left(\sqrt{p_{v0}}\right) + L_2\sqrt{\frac{\lambda_2}{2n_2(2-\lambda_2)}}\,, \tag{10}$$

$$\text{CL}_2 = \sin^{-1}\left(\sqrt{p_{v0}}\right), \tag{11}$$

$$\text{LCL}_2 = \sin^{-1}\left(\sqrt{p_{v0}}\right) - L_2\sqrt{\frac{\lambda_2}{2n_2(2-\lambda_2)}}\,, \tag{12}$$

and plot $\text{EWMA}_{T_{V_t}}$. If any $\text{EWMA}_{T_{V_t}} \geq \text{UCL}_2$ or $\text{EWMA}_{T_{V_t}} \leq \text{LCL}_2$, an out-of-control signal is issued. Here again, the two parameters, L_2 and λ_2, are chosen to yield the desired in-control ARL ($\text{ARL}_{T_{V0}} \approx 370$) using the Markov chain approach.

Table 2 The $\mathrm{ARL}_{T_{V1}}$ of the EWMA-V chart when $p_{v0} = 0.3$

	p_{v1}			
$0.5n_2$	0.2	0.4	0.6	0.8
4	41.9	50.5	7.0	3.4
5	33.7	40.8	6.0	3.0
6	28.2	34.2	5.2	2.7
7	24.2	29.4	4.7	2.5
8	21.3	25.8	4.3	2.3
9	19.0	23.0	4.0	2.2
10	17.2	20.7	3.7	2.1
11	15.7	18.9	3.5	2.0
12	14.5	17.4	3.3	2.0
13	13.4	16.1	3.2	1.9
14	12.6	15.1	3.1	1.8
15	11.8	14.1	2.9	1.8

3.2 The In-Control and Out-of-Control Average Run Lengths of the EWMA-V Chart

The $\mathrm{ARL}_{T_{V0}}$ of the new EWMA-V chart is a function of (n_2, L_2, λ_2). Adopting the in-control process proportion $p_{v0} = 0.3$, $\mathrm{ARL}_{T_{V0}} \approx 370$ with $\lambda_2 = 0.2$ and $L_2 = 2.86$, the out-of-control ARL, $\mathrm{ARL}_{T_{V1}}$, of the EWMA-V chart is listed for $0.5n_2 = 4(1)15$ and $p_{v1} = 0.2(0.2)0.8$ in Table 2. From Table 2, we found that $\mathrm{ARL}_{T_{V1}}$ decreases when the out-of-control value of p_{v1} is far away from the in-control value of p_{v0}, and when n_2 increases.

4 Performance Measurement of Using the EWMA-V Chart and the EWMA-M Chart Simultaneously

Use of both the EWMA-V Chart and EWMA-M Chart permits monitoring of the process variance and mean simultaneously. We will use the ARL to measure the out-of-control detection performance of using both the EWMA-V Chart and EWMA-M Chart.

4.1 The In-Control Average Run Lengths of the EWMA-V Chart and the EWMA-M Chart

In a production process, we take a sample of size $n_1 + n_2$, the first n_1 observations are used to calculate the statistic EWMA_{T_M}, then the remaining n_2 observations are taken to calculate the statistic EWMA_{T_V}. The statistics EWMA_{T_M} and EWMA_{T_V} are

independent since the two groups of observations are independent. The in-control overall ARL, ARL_0, of the newly proposed charts is well approximated as follows (see Hawkins 1992),

$$\begin{aligned}\mathrm{ARL}_0 &= \frac{1}{\frac{1}{\mathrm{ARL}_{T_{M0}}} + \frac{1}{\mathrm{ARL}_{T_{V0}}} - \left(\frac{1}{\mathrm{ARL}_{T_{M0}}}\frac{1}{\mathrm{ARL}_{T_{V0}}}\right)} \\ &\approx \frac{1}{\frac{1}{\mathrm{ARL}_{T_{M0}}} + \frac{1}{\mathrm{ARL}_{T_{V0}}}}. \end{aligned} \tag{13}$$

The ARL_0 of using both the EWMA-M Chart and EWMA-V Chart with any combinations of (n_1, n_2) and (p_{m0}, p_{v0}) are all approximately 185 because of

$$\mathrm{ARL}_0 \approx \frac{1}{\frac{1}{\mathrm{ARL}_{T_{M0}}} + \frac{1}{\mathrm{ARL}_{T_{V0}}}} = \frac{1}{\frac{1}{370} + \frac{1}{370}}.$$

Usually, the most efficient way of using a combined charting procedure is to use all the observations for both the mean and variance charts but this was not considered in our scheme. The reason is that the two monitoring statistics EWMA-V and EWMA-M are dependent if all the observations are used, and this will complicate the calculation of the overall ARLs of the two proposed control charts. Of course, we may use simulation to estimate the overall ARLs.

4.2 *The Out-of-Control Average Run Lengths of the EWMA-V Chart and the EWMA-M Chart*

When the process is out of control due to a shift in the process mean, μ, the process proportion becomes p_{m1} ($\neq p_{m0}$). For an out-of-control process whose variance σ^2 has changed, the process proportion becomes p_{v1} ($\neq p_{v0}$). The out-of-control overall ARL, ARL_1, of using the EWMA-V Chart and EWMA-M Chart simultaneously could be calculated approximately using

$$\mathrm{ARL}_1 \approx \frac{1}{\frac{1}{\mathrm{ARL}_{T_{M1}}} + \frac{1}{\mathrm{ARL}_{T_{V1}}}}. \tag{14}$$

The ARL_1s of using both the EWMA-M Chart and EWMA-V Chart with the combinations of $n_1 = 8(2)24$, $0.5n_2 = 4(1)12$, $p_{v0} = 0.1$, $p_{m0} = 0.5$, $p_{v1} = 0.2$, and $p_{m1} = 0.25(0.05)0.45$ are listed in Table 3. In Table 3, we observe that the ARL_1 changes inversely with n_1 and n_2, and the ARL_1 decreases when p_{m1} is far away from p_{m0} and/or p_{v1} is far away from p_{v0}. The results are much more reasonable than those corresponding to use of both the V Chart and M Chart.

Table 3 The ARL_1 of the EWMA-V and EWMA-M charts when $p_{v0} = 0.1$, $p_{m0} = 0.5$, and $p_{v1} = 0.2$

	p_{m1}					
$0.5n_2$	n_1	0.25	0.30	0.35	0.40	0.45
4	8	6.8	8.4	12.0	19.0	26.0
5	10	5.9	7.4	10.4	15.9	21.8
6	12	5.1	6.6	9.0	13.4	18.1
7	14	4.9	5.8	7.7	11.5	15.3
8	16	4.3	5.6	7.2	8.0	14.4
9	18	4.1	5.0	6.5	9.4	12.6
10	20	4.0	4.8	6.0	8.7	11.6
11	22	3.4	4.3	5.7	8.0	10.7
12	24	3.3	4.1	5.2	7.4	9.7

5 When the Population Mean and Variance Are Unknown

When the in-control process mean, μ, and the process variance, σ^2, are unknown, and hence the in-control process proportions, p_{m0} and p_{v0}, are unknown, we can use the following two preliminary independent sample data sets

$$X_{t,1}, X_{t,2}, \ldots, X_{t,n_1}\ , \quad t = 1, 2, \ldots, k\,.$$
$$X_{t,n_1+1}, X_{t,n_1+2}, \ldots, X_{t,n_1+n_2}\ , \quad t = 1, 2, \ldots, k\,.$$

from k sampling periods, each with an even number of observations, n_1 and n_2, to estimate them (see, e.g., Montgomery 2009), i.e.

$$\hat{\mu} = \bar{\bar{x}} = \frac{\sum_{t=1}^{k}\sum_{j=1}^{n_1} x_{t,j}}{kn_1}\ , \quad \hat{\sigma} = \frac{\bar{S}}{c_4} = \frac{\sum_{t=1}^{k} S_t}{c_4 k}\ ,$$

where

$$\hat{p}_{m0} = \frac{\sum_{t=1}^{k} \frac{M_t}{n_1}}{k}\ , \quad \hat{p}_{v0} = \frac{\sum_{t=1}^{k} \frac{V_t}{0.5n_2}}{k}\ , \quad S_t = \sqrt{\frac{\sum_{j=n_1+1}^{n_1+n_2} \left(X_{t,j} - \bar{X}_t\right)^2}{n_2 - 1}}\ , \quad t = 1, \ldots, k\,,$$

$$c_4 = \left(\frac{2}{n_2 - 1}\right)^{0.5} \frac{\Gamma(0.5n_2)}{\Gamma\big(0.5(n_2 - 1)\big)}\ , \quad \text{and} \quad \bar{S} = \frac{\sum_{t=1}^{k} S_t}{k}\,.$$

The EWMA-V and EWMA-M Charts are thus constructed using these estimated values of p_{m0} and p_{v0}. The statistics $EWMA_{T_M}$ and $EWMA_{T_V}$ corresponding to the samples of sizes n_1 and n_2 are plotted on the resulting EWMA-V and EWMA-M Charts simultaneously. If no points fall outside their control limits, then we would deem the process to be in-control.

6 Example

We will use an example of service time from Yang et al. (2012) to illustrate the use of the new EWMA-V and EWMA-M Charts. Service time is an important quality characteristic in the banking industry. To measure the efficiency in the service system of a bank branch, the in-control sampling service times (unit: minutes) is measured from twenty counters every day for 15 days. That is, fifteen samples of size $n_1 + n_2 = 20$, where $n_1 = n_2 = 10$ are available. This in-control data has been analyzed assuming a non-normal distribution. For each sample, the first ten observations illustrated in Table 4 are used to calculate the EWMA_{T_M} statistic and the last ten observations illustrated in Table 5 are used to calculate the EWMA_{T_V} statistic.

To construct the EWMA-V and EWMA-M charts, the variance and mean of the service time are estimated by $(\bar{S}/c_4)^2$ and $\bar{\bar{x}}$ using the fifteen samples in Tables 5 and 4, respectively. The estimate of the variance is $\hat{\sigma}^2 = (\bar{S}/c_4)^2 = 30.159$ and the estimate of the mean is $\hat{\mu} = \bar{\bar{x}} = 5.77$. For each sample in Table 5, the monitoring statistic EWMA_{T_V} is calculated. For each sample in Table 4, the monitoring statistic EWMA_{T_M} is calculated. Hence, the estimates of proportions (p_{m0}, p_{v0}) are $(\hat{p}_{m0} = \frac{\sum_{t=1}^{15} M_t/10}{15} = 0.39, \hat{p}_{v0} = \frac{\sum_{t=1}^{15} V_t/5}{15} = 0.24)$, where M_t = Total number of $(Y_j > 5.77)$ and $j = 1, 2, \ldots, 10$, V_t = Total number of $(Y_j^* > 30.159)$ and $j = 11, 12, \ldots, 20$. The EWMA-V and EWMA-M Charts with $\lambda_1 = \lambda_2 = 0.2$ are constructed as follows based on the fifteen in-control samples, respectively.

Table 4 The service times from the first ten counters in a bank branch

t	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	M_t	$\text{EWMA}_{T_{M_t}}$
1	0.88	0.78	5.06	5.45	2.93	6.11	11.59	1.20	0.89	3.21	2	0.63
2	3.82	13.4	5.16	3.20	32.27	3.68	3.14	1.58	2.72	7.71	3	0.62
3	1.40	3.89	10.88	30.85	0.54	8.40	5.10	2.63	9.17	3.94	4	0.63
4	16.8	8.77	8.36	3.55	7.76	1.81	1.11	5.91	8.26	7.19	7	0.71
5	0.24	9.57	0.66	1.15	2.34	0.57	8.94	5.54	11.69	6.58	4	0.70
6	4.21	8.73	11.44	2.89	19.49	1.20	8.01	6.19	7.48	0.07	6	0.74
7	15.08	7.43	4.31	6.14	10.37	2.33	1.97	1.08	4.27	14.08	5	0.74
8	13.89	0.30	3.21	11.32	9.90	4.39	10.5	1.70	10.74	1.46	5	0.76
9	0.03	12.76	2.41	7.41	1.67	3.70	4.31	2.45	3.57	3.33	2	0.70
10	12.89	17.96	2.78	3.21	1.12	12.61	4.23	6.18	2.33	6.92	5	0.71
11	7.71	1.05	1.11	0.22	3.53	0.81	0.41	3.73	0.08	2.55	1	0.64
12	5.81	6.29	3.46	2.66	4.02	10.95	1.59	5.58	0.55	4.10	3	0.62
13	2.89	1.61	1.30	2.58	18.65	10.77	18.23	3.13	3.38	6.34	4	0.64
14	1.36	1.92	0.12	11.08	8.85	3.99	4.32	1.71	1.77	1.94	2	0.60
15	21.52	0.63	8.54	3.37	6.94	3.44	3.37	6.37	1.28	12.83	5	0.64

Table 5 The service times from the last ten counters in a bank branch

t	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	V_t	$\text{EWMA}_{T_{V_t}}$
1	3.82	6.29	10.88	30.85	9.9	3.99	1.59	1.71	8.26	4.1	1	0.5
2	0.24	12.76	11.44	3.2	3.53	0.57	18.23	2.45	2.72	6.92	3	0.58
3	3.82	7.43	0.12	3.37	1.12	12.61	1.59	1.08	0.89	0.07	1	0.56
4	13.89	3.89	5.16	11.32	4.02	0.57	8.01	6.19	1.77	6.58	1	0.54
5	5.81	12.76	2.41	1.15	3.53	0.81	11.59	5.91	4.27	3.33	0	0.43
6	12.89	8.73	10.88	2.89	18.65	10.95	0.41	3.13	4.27	7.71	1	0.44
7	2.89	0.63	0.12	0.22	4.02	10.95	8.01	1.08	10.74	4.1	0	0.35
8	16.8	1.05	1.3	3.2	2.34	0.81	4.32	3.13	0.08	1.46	1	0.37
9	4.21	17.96	5.06	0.22	4.02	3.99	8.01	5.91	0.55	3.33	1	0.39
10	12.89	8.77	11.44	7.41	1.12	1.81	4.32	5.58	0.89	14.08	1	0.41
11	0.88	8.77	5.06	3.55	8.85	10.95	18.23	5.54	2.33	6.58	2	0.46
12	7.71	7.43	0.12	2.58	1.12	2.33	4.23	2.63	4.27	3.33	0	0.37
13	7.71	9.57	0.12	30.85	7.76	1.81	3.14	1.71	2.72	14.08	2	0.43
14	2.89	1.05	2.41	11.32	32.27	8.4	1.97	2.45	11.69	12.83	2	0.48
15	1.36	0.63	3.46	11.32	0.54	10.95	4.23	2.45	2.33	6.34	2	0.52

The EWMA-V Chart:

$$\text{UCL}_2 = 0.725\,, \quad \text{LCL}_2 = 0.299\,.$$

The EWMA-M Chart:

$$\text{UCL}_1 = 0.85\,, \quad \text{LCL}_1 = 0.47\,.$$

The monitoring statistics EWMA-V and EWMA-M are calculated (see Tables 4 and 5). The EWMA-V and EWMA-M Charts show no signals (see Fig. 1a, b).

For comparison, we constructed the corresponding Shewhart standard deviation and mean (S-$\bar{X}$) charts, EWMA-S and EWMA-$\bar{X}$ charts, and transformed S and $\bar{X}$ charts by applying $X^{0.278}$ transformation (see Montgomery 2009), respectively. The Shewhart S-$\bar{X}$ charts are constructed with bounds as follows:

$$\text{UCL}_S = 9.66\,, \ \text{LCL}_S = 1.597\,, \quad \text{UCL}_{\bar{X}} = 10.98\,, \ \text{LCL}_{\bar{X}} = 0.55\,.$$

The Shewhart S-$\bar{X}$ charts also show no signals (see Fig. 1c, d). The EWMA-S and EWMA-$\bar{X}$ charts are constructed with bounds as follows:

$$\text{UCL}_{\text{EWMA-}S} = 7.495\,, \ \text{LCL}_{\text{EWMA-}S} = 2.485\,,$$
$$\text{UCL}_{\text{EWMA-}\bar{X}} = 11.02\,, \ \text{LCL}_{\text{EWMA-}\bar{X}} = 0.53\,.$$

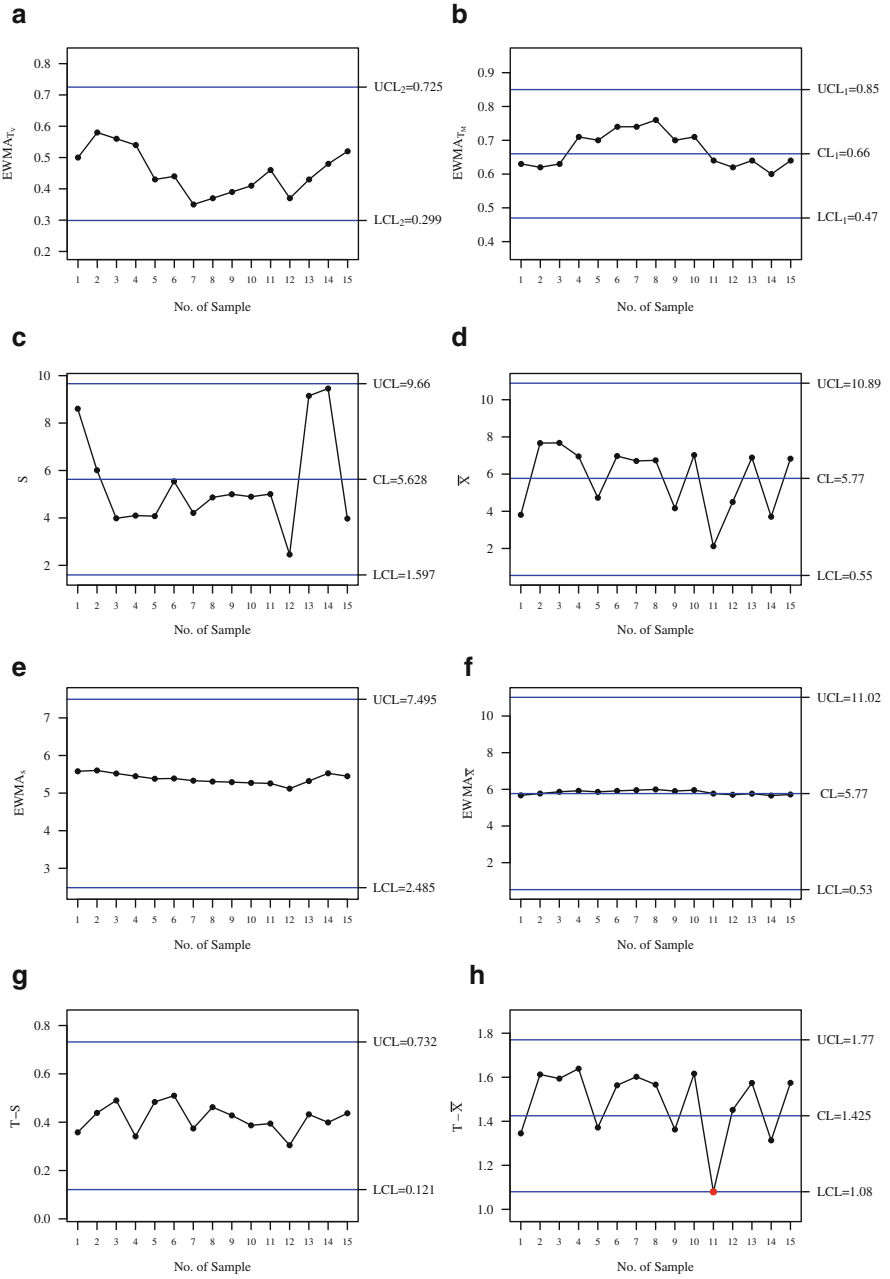


Fig. 1 First 15 control chart points. (**a**) The EWMA-V chart. (**b**) The EWMA-M chart. (**c**) Shewhart S chart. (**d**) Shewhart $\bar{X}$ chart. (**e**) The EWMA-S chart. (**f**) The EWMA-$\bar{X}$ chart. (**g**) The transformed Shewhart S chart. (**h**) The transformed Shewhart $\bar{X}$ chart

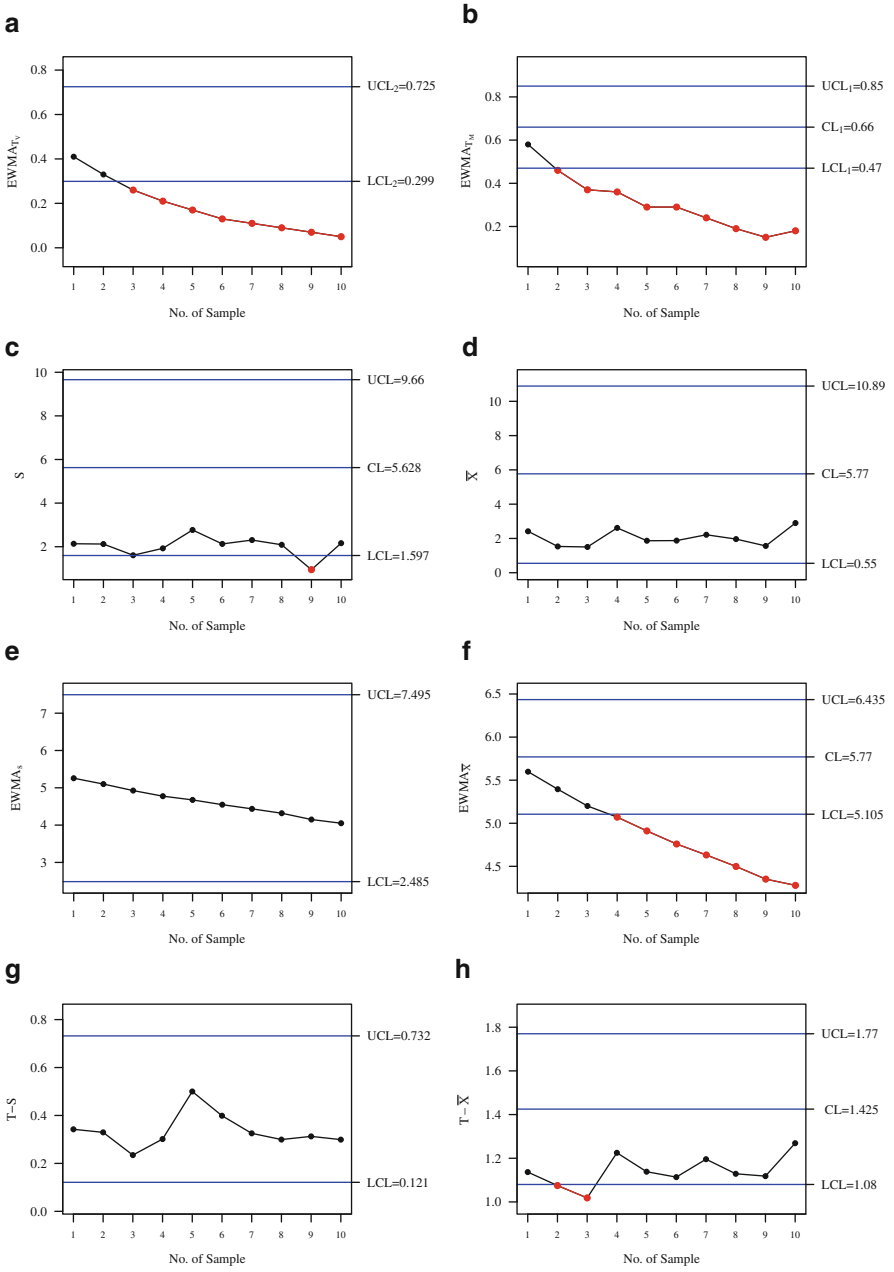


Fig. 2 Next 10 control chart points. (**a**) The EWMA-V chart. (**b**) The EWMA-M chart. (**c**) Shewhart S chart. (**d**) Shewhart $\bar{X}$ chart. (**e**) The EWMA-S chart. (**f**) The EWMA-$\bar{X}$ chart. (**g**) The transformed Shewhart S chart. (**h**) The transformed Shewhart $\bar{X}$ chart

The EWMA-S and EWMA-$\bar{X}$ charts also show no signals (see Fig. 1e, f). The transformed S and $\bar{X}$ charts are constructed with bounds as follows:

$$\mathrm{UCL}_{TS} = 0.732\,,\ \mathrm{LCL}_{TS} = 0.121\,,\ \ \mathrm{UCL}_{T\bar{X}} = 1.77\,,\ \mathrm{LCL}_{T\bar{X}} = 1.08\,.$$

The transformed S chart shows no signal but the transformed $\bar{X}$ chart shows a false signal (see Fig. 1g, h).

A data set consisting of ten samples of ten service times from an improved new automatic service system of the bank branch was collected and shown in Fig. 2. The service times should be reduced because of the improved new automatic service system. The new service times are regarded as out-of-control data set. The new proposed EWMA-V and EWMA-M Charts are used to monitor the new service times to illustrate their out-of-control detection ability

The EWMA-V and EWMA-M Charts detected out-of-control signals from the third sample and the second sample onward, respectively (samples 3–10 on EWMA-V Chart and samples 2–10 on EWMA-M Chart) (see Fig. 2a, b). That is, the variance and mean of the new service times are significantly reduced because of the improved new automatic service system. However, the corresponding Shewhart S-$\bar{X}$ charts produced only one true out-of-control signal (sample 9 on S chart) (see Fig. 2c, d). The EWMA-S and EWMA-$\bar{X}$ charts produced six true out-of-control signals (samples 4–10 on EWMA-$\bar{X}$ chart) (see Fig. 2e, f), and transformed S and $\bar{X}$ charts produced two true out-of-control signals (samples 2–3 on transformed $\bar{X}$ chart) (see Fig. 2g, h).

To construct the Shewhart S-$\bar{X}$ charts, EWMA-S and EWMA-$\bar{X}$ charts and transformed S and $\bar{X}$ charts one requires the normality assumption but this is not the case for the EWMA-V and EWMA-M Charts. In this example, the EWMA-V and EWMA-M charts detected most of the out-of-control signals. The New EWMA-V and EWMA-M Charts showed better detection ability than the existing charts in monitoring and detecting process variance and mean shifts. The new EWMA-V and EWMA-M Charts are thus recommended.

7 Conclusions

In this paper, we propose using both the EWMA-M and EWMA-V Charts, based on two simple independent statistics to monitor the mean and variance shifts in the process simultaneously when the distribution of a quality characteristic is not known or is not believed to be normal. A numerical example of service times from a bank branch with a right skewed distribution illustrated the application of the new EWMA-M and EWMA-V Charts which were compared with some existing charts. The proposed new EWMA-M and EWMA-V Charts showed better detection ability than the existing charts in monitoring and detecting both the process mean and variance shifts. The new EWMA-M and EWMA-V Charts are thus recommended.

The M-V charts have the advantage of simplicity, but this is counter weighed by some anomalies in their performance due to the discrete nature of the monitoring variables. In an analysis based on two data sets, the EWMA-M and EWMA-V charts appear to have good performance. Knoth and Morais (2013) deal with what they termed an ARL-unbiased chart. We may consider ARL-unbiased EWMA charts based on non-transformed statistics in a future study.

Acknowledgements The research was partially supported by the National Science Council of the Republic of China, grant no. NSC 100-2118-M-004-003-MY2, Commercial College of National Chengchi University, Taiwan and National Center for Theoretical Sciences, Taiwan.

References

Amin, R., Reynolds, M. R., Jr., & Baker, S. (1995). Nonparametric quality control charts based on the sign statistic. *Communications in Statistics – Theory and Methods, 24*, 1597–1624.

Altukife, P. F. (2003). A new nonparametric control charts based on the observations exceeding the grand median. *Pakistan Journal of Statistics, 19*(3), 343–351.

Bakir, S. T. (2004). A distribution-free Shewhart quality control chart based on signed-ranks. *Quality Engineering, 16*(4), 613–623.

Bakir, S. T. (2006). Distribution free quality control charts based in sign rank like statistics. *Communication in Statistics: Theory and Methods, 35*, 743–757.

Chakraborti, S., Lann, P., & Van der Wiel, M. A. (2001). Nonparametric control charts: an overview and some results. *Journal of Quality Technology, 33*, 304–315.

Das, N., & Bhattacharya, A. (2008). A new non-parametric control chart for controlling variability. *Quality Technology & Quantitative Management, 5*(4), 351–361.

Ferrell, E. B. (1953). Control charts using midranges and medians. *Industrial Quality Control, 9*, 30–34.

Graham, M. A., Chakraborti, S., & Human, S. W. (2011). A nonparametric exponentially weighted moving average signed-rank chart for monitoring location. *Computational Statistics and Data Analysis, 55*(8), 2490–2503.

Hawkins, D. M. (1992). A fast approximation for average run length of CUSUM control charts. *Journal of Quality Technology, 24*, 37–43.

Jones-Farmer, L. A., & Champ, C. W. (2010). A distribution-free Phase I control chart for subgroup scale. *Journal of Quality Technology, 42*(2), 373–387.

Knoth, S., & Morais, M. C. (2013). On ARL-unbiased charts. In *Proceedings of the XIth International Workshop on Intelligent Statistical Quality Control* (pp. 31–50).

Li, S., Tang, L., & Ng, S. (2010). Nonparametric CUSUM and EWMA control charts for detecting mean shifts. *Journal of Quality Technology, 42*(2), 209–226.

Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics, 32*, 1–12.

Montgomery, D. C. (2009). *Introduction to statistical quality control*. New York: Wiley.

Mosteller, F., & Youtz, C. (1961). Tables of the Freeman-Tukey transformations for the binomial and Poisson distributions. *Biometrika, 48*(3,4), 433–440.

Ning, X., Shang, Y., & Tsung, F. (2009). Statistical process control techniques for service processes: a review. In *The 6th International Conference on Service Systems and Service Management* (pp. 927–931), Xiamen, China, Apr 2009.

Tsung, F., Li, Y., & Jin, M. (2008). Statistical process control for multistage manufacturing and service operations: a review and some extensions. *International Journal of Services Operations and Informatics, 3*, 191–204.

Yang, S. F., & Cheng, S. (2011). A new nonparametric CUSUM mean chart. *Quality and Reliability Engineering International, 38*(5), 867–875.

Yang, S. F., Cheng, C. T., Hung, Y. C., & Cheng, S. (2012). A new chart for monitoring service process mean. *Quality and Reliability Engineering International, 28*(4), 377–386.

Yang, S. F., Lin, J. S., & Cheng, S. (2011). A new nonparametric EWMA sign chart. *Expert Systems with Applications, 38*(5), 6239–6243.

Zou, C., & Tsung, F. (2010). Likelihood ratio-based distribution-free EWMA control charts. *Journal of Quality Technology, 42*(2), 1–23.

Comparison of Phase II Control Charts Based on Variable Selection Methods

Giovanna Capizzi and Guido Masarotto

Abstract In recent years, control charts based on variable selection (VS) algorithms have been suggested for monitoring multivariate data. These charts share the common idea that process faults usually affect a small fraction of the monitored quality characteristics. Thus, VS methods can be used to identify the subset of the variables for which the shift may have occurred. However, the suggested VS-based control charts differ in many aspects such as the particular VS algorithm and the type of control statistic. In this paper, we compare VS-based control charts in various out-of-control scenarios characterizing modern manufacturing environments such as high-dimensional data, profile, and multistage process monitoring. The main aim of this paper is to provide practical guidelines for choosing a suitable VS-based monitoring scheme.

Keywords Generalized least squares • LAR • LASSO • Multivariate EWMA

1 Introduction

Nowadays, the performance of modern processes depends on several related quality characteristics. The statistical monitoring of "high-dimensional" processes is known as multivariate statistical process control (MSPC, see Bersimis et al. 2007, for a comprehensive review of the MSPC literature). A critical task for an MSPC control scheme is assessing whether the multidimensional process is in-control (IC) or not. Although it is unlikely that all the quality characteristics shift simultaneously, it is more common that only a subset of variables experiences abnormal changes. Thus, it could be more efficient to monitor only the potential out-of-control (OC) variables, which, however, are not known in advance. Thus, recent developments in the MSPC framework propose using variable selection (VS) algorithms to identify the suspected variables and then charting only these characteristics to test whether the multidimensional process is in-control or not

G. Capizzi (✉) • G. Masarotto
Department of Statistical Sciences, University of Padua, Padua, Italy
e-mail: capizzi@stat.unipd.it; masarotto@stat.unipd.it

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_10

(see Wang and Jiang 2009; Zou and Qiu 2009; Zou et al. 2010; Capizzi and Masarotto 2011; Jiang et al. 2012). These VS-based approaches seem attractive since they offer a very satisfactory performance in OC scenarios involving a shift in one, two,..., all the monitored quality characteristics. They are also coupled with diagnostic tools to accurately identify the variables responsible for the change.

These recent proposals combine different multivariate control charts with different VS procedures. A "forward" selection algorithm (FVS) has been combined with a Shewhart-type and a multivariate EWMA (MEWMA) by Wang and Jiang (2009) and Jiang et al. (2012), respectively. Other VS algorithms such as Least Absolute Shrinkage and Selection Operator (LASSO, see Tibshirani 1996) and Least Angle Regression (LAR, see Efron et al. 2004) have been proposed combined with an MEWMA-based control chart, by Zou and Qiu (2009), Zou et al. (2010) and Capizzi and Masarotto (2011), respectively. The suggested monitoring schemes differ not only in the VS algorithm but also in other aspects. In particular, the control charts based on stepwise regression assume that the number of variables that can be potentially OC is fixed a priori; this condition has been relaxed for the LASSO- and LAR-based schemes (LEWMA and LAR-EWMA, hereafter). Indeed, these control charts assume that any, proper or improper, subset of the monitored variables can potentially shift. Further, LEWMA and LAR-EWMA are based on two slightly different control statistics. In addition, LAR-EWMA is developed not only for testing the status of the process mean but also for detecting an increase in the total variability.

To provide some guidelines on how to choose between different VS-based multivariate control charts and give some suggestions for further research, we here compare and discuss some VS-based control charts recently proposed in the SPC literature. For a more objective comparison, we use for all the investigated control charts the general regression model introduced in Capizzi and Masarotto (2011) for the LAR-EWMA. Indeed, this more general regression framework allows to handle a wide a variety of multivariate scenarios not only involving shifts in the component of a multivariate mean vector but also those related to changes in a profile or in a multistage process.

The paper is organized as follows. Section 2 briefly describes the procedures based on the variable selection algorithms. Section 3 presents the main results concerning comparisons, in terms of average run length (ARL), between some control schemes based on different VS-based algorithms. Details on the multivariate OC scenarios, discussed in the comparisons, are given in the Appendix. Concluding remarks are given in Sect. 4.

2 Statistical Monitoring Based on Variable Selection Algorithms

2.1 Generalities

Assume that at each time t, $t = 1, 2, \ldots$, independent observations on $\boldsymbol{y}_t$, an $n \times 1$ vector of quality characteristics, are available, and consider the following Gaussian change-point model

$$\boldsymbol{y}_t \sim \begin{cases} N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) & \text{if } t < \tau \text{ (in-control)} \\ N_n(\boldsymbol{\mu} + \boldsymbol{\delta}, \boldsymbol{\Omega}) & \text{if } t \geq \tau \text{ (out-of-control)} \end{cases} \tag{1}$$

that is, at τ, an unknown instant of time, the mean vector and the covariance matrix shift leading the process to an OC state. Further, we suppose that the IC mean vector $\boldsymbol{\mu}$ and the IC covariance matrix $\boldsymbol{\Sigma}$ are known.

Concerning the OC mean vector, we assume that, at least approximately, the mean shift $\boldsymbol{\delta}$ takes the form

$$\boldsymbol{\delta} = \boldsymbol{F}\boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\boldsymbol{F}$ a suitable $n \times p$ matrix of known constants. Thus, the mean vector may shift along any vector in the subspace spanned by the columns of $\boldsymbol{F}$, allowing for a multitude of potential shift directions. As shown in Capizzi and Masarotto (2011), formulation (2) is sufficiently flexible to encompass a wide variety of change-point scenarios. Further, suppose there is a practical interest only in detecting an increase in the total dispersion and assume that $\boldsymbol{\Omega} - \boldsymbol{\Sigma}$ is a positive definite matrix.

Suppose process observations are accumulated in the following MEWMA

$$\boldsymbol{z}_t = (1 - \lambda)\boldsymbol{z}_{t-1} + \lambda(\boldsymbol{y}_t - \boldsymbol{\mu}) \tag{3}$$

with $\boldsymbol{z}_0 = \boldsymbol{0}_n$, $0 < \lambda \leq 1$. Assuming the following (approximated) linear model

$$\boldsymbol{z}_t = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{a}_t,$$

with $\boldsymbol{a}_t \sim N_n(\boldsymbol{0}_n, \lambda/(1 - \lambda)\boldsymbol{\Sigma})$, the stability of the process mean can be checked by testing the hypothesis system

$$\begin{cases} H_0 : \boldsymbol{\beta} = \boldsymbol{0}_p, \\ H_1 : \boldsymbol{\beta} \neq \boldsymbol{0}_p. \end{cases} \tag{4}$$

Unfortunately, the standard test, described in any regression textbook, for the hypothesis system (4) can show a very low sensitivity when only a few components of $\boldsymbol{\beta}$ effectively shift, and a much more efficient approach should consider alternative hypothesis systems on reduced subsets of the parameters.

A promising solution consists of using a suitable VS algorithm for determining subsets, having different sizes, of suspected variables, i.e., subsets of columns of $\boldsymbol{F}$ corresponding to nonzero coefficients. In particular, for $k = 1, \ldots, p$, denote with $J_k = \{j_{k,1}, \ldots, j_{k,k}\}$ the indices of the selected predictors. Since the set of coefficients $\{\beta_{j_{k,1}}, \ldots, \beta_{j_{k,k}}\}$ correspond to a plausible subset of possible out-of-control parameters, the VS-based control statistics, for $k = 1, \ldots, p$, test the following hypothesis systems:

$$\begin{cases} H_0': \quad \beta_j = 0 \text{ for } j = 1, \ldots, p, \\ H_{1,k}': \beta_j \neq 0 \text{ if } j \in J_k \text{ and } \beta_j = 0 \text{ if } j \notin J_k. \end{cases} \tag{5}$$

2.2 Three Different Approaches

Three distinct methods have been suggested for testing the hypothesis system (5).

In Wang and Jiang (2009) and Jiang et al. (2012), users are requested to choose in advance a suitable value for k. Then, for $t = 1, 2, \ldots$, a standard forward search algorithm is used to select J_k, and an OC alarm is signaled when the following control statistic

$$S_{t,k} = \hat{\boldsymbol{\beta}}_{t,k}' \boldsymbol{F}' \boldsymbol{\Sigma}^{-1} \boldsymbol{F} \hat{\boldsymbol{\beta}}_{t,k} \tag{6}$$

is greater than the control limit chosen for giving a desired IC performance. Here, $\hat{\boldsymbol{\beta}}_{t,k}$ denotes the GLS estimate of $\boldsymbol{\beta}$ obtained under $H_{1,k}'$, i.e., constraining to zero the coefficients of the predictors not in J_k.

In Zou and Qiu (2009), $J_1, \ldots, J_p$ are determined using the LASSO algorithm. Then, for $k = 1, \ldots, p$, the authors suggest to compute the control statistic

$$V_{t,k} = \frac{(\boldsymbol{z}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{F} \tilde{\boldsymbol{\beta}}_{t,k})^2}{\tilde{\boldsymbol{\beta}}_{t,k}' \boldsymbol{F}' \boldsymbol{\Sigma}^{-1} \boldsymbol{F} \tilde{\boldsymbol{\beta}}_{t,k}}, \tag{7}$$

where $\tilde{\boldsymbol{\beta}}_{t,k}$ denotes the LASSO estimator of $\boldsymbol{\beta}$ obtained under $H_{1,k}'$. An OC alarm is given when the overall control statistic

$$W_t = \max_{k=1,\ldots,p} \frac{V_{t,k} - E[V_{t,k}]}{\sqrt{Var[V_{t,k}]}} \tag{8}$$

is greater than a suitable control limit. In (8), the mean and standard deviation of (7) are computed under the null hypothesis.

Alternatively, Capizzi and Masarotto (2011) suggest selecting $J_1, \ldots, J_k$ using the LAR algorithm and, for each $k = 1, \ldots, p$, to compute the statistic $S_{t,k}$. Since it is important to detect not only changes in the process mean but also increases in

the dispersion, Capizzi and Masarotto (2011) also consider the additional alternative hypothesis

$$H'_{1,p+1} : \boldsymbol{\beta} = \mathbf{0}_p \text{ and } E\{(\boldsymbol{y}_t - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_t - \boldsymbol{\mu})\} > n,$$

and the related one-sided EWMA statistic

$$S_{t,p+1} = \max\left(1, (1-\lambda) S^{(1)}_{t-1,p+1} + \lambda \frac{(\boldsymbol{y}_t - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_t - \boldsymbol{\mu})}{n}\right), \tag{9}$$

with $S_{0,p+1} = 1$. Then, the LAR-based EWMA, for jointly monitoring the process mean and dispersion, is given by the aggregation of the $p+1$ statistics

$$M_t = \max_{k=1,\ldots,p+1} \frac{S_{t,k} - E[S_{t,k}]}{\sqrt{\text{Var}[S_{t,k}]}}, \tag{10}$$

where $S_{t,k}$ is given by (6) for $k = 1, \ldots, p$, and by (9) for $k = p+1$. The combined control statistic (10) triggers an alarm when it exceeds a suitable control limit.

2.3 First Recommendations and Open Questions

As shown in Zou and Qiu (2009), Capizzi and Masarotto (2011), and Jiang et al. (2012), control charts like W_t and M_t offer a good protection against shifts occurring in one, two,..., all components. Although the resulting scheme is not necessarily the best for detecting a shift occurring in a fixed number of components, it is usually close to the best. Conversely, control charts using a fixed value of k, such as those proposed by Wang and Jiang (2009) and Jiang et al. (2012), offer the best protection when shifts involve exactly k variables and unavoidably inferior protection when a shifts occur in a number of components different from the fixed value. Further, statistics such as W_t and M_t do not need an a priori choice of k. Thus, we suggest using an aggregated control statistic.

In addition, we strongly recommend including a control statistic, like $S_{t,p+1}$, designed for detecting a change in the dispersion. Indeed, joint monitoring of the process mean and dispersion is relevant *per se* but also provides some level of robustness against modeling errors and unforeseen behaviors. Further, as shown in the univariate case by Reynolds and Stoumbos (2005, 2006), the inclusion of a variance control statistic can be helpful for efficiently detecting large changes in the mean.

In the following, studying by simulation the ARL performance of VS-based control charts, we address the following additional issues: (1) Which variable selection algorithm should be used? (2) Which is better to use for monitoring, the elementary control statistic $S_{t,k}$ or $V_{t,k}$?

3 A Simulation Study

To address some of the issues discussed in the previous section, we compare five VS-based monitoring schemes. As recommended, all the schemes are based on a combination, similar to M_t, of p elementary control statistics used for detecting a mean shift and of the control statistic $S_{t,p+1}$, given by (9), for detecting increases in the total variation.

Details of the five control charts are given in Table 1. Note that, when a forward stepwise search is used, we have, for each k, that $S_{t,k} = V_{t,k}$. Thus, we present only one scheme for the forward VS algorithm. However, the control statistics $V_{t,k}$, given in (7), are here based on the LASSO- and LAR-based estimators of the vector $\boldsymbol{\beta}$. Observe that in (6) and (7), at each stage k, the nonzero elements obtained via these three different VS algorithms are not necessarily the same.

Concerning the choice of the smoothing constant, as suggested in the literature (Lucas and Saccucci 1990; Prabhu and Runger 1997; Zou and Qiu 2009; Capizzi and Masarotto 2011; Jiang et al. 2012), a reasonable choice for normally distributed observations is between 0.1 and 0.3. The performance of the different VS-based schemes has been investigated for different values of λ and τ. Because results are comparable for all the choices of these tuning constants, in the following results will be referred only to $\lambda = 0.1$ and $\tau = 1$. The five VS-based control charts are compared in terms of out-of-control ARL evaluated using 500,000 Monte Carlo replications. The control limits, giving an in-control ARL equal to 500, have been computed using a stochastic approximation algorithm (Ruppert 1991; Polyak and Juditsky 1992). Within a reasonable number of iterations, the algorithm estimates the control limits with a given level of accuracy. Table 1 lists the estimates of the control limits for the five VS-based control charts. In addition, the mean and standard deviation of the elementary statistics $S_{t,k}$ and $V_{t,k}$, for $k = 1, \ldots, p + 1$, were computed by simulation.

Suitable choices of the matrix F lead to several change-point models, such as the "unstructured" scenario, when changes directly involve the components of the multivariate mean vector, and several "structured" scenarios, such those involving changes in a profile, that is, in the relationship between a response variable and

Table 1 Five VS-based control charts

	FORWARD	LASSO/S	LASSO/V	LAR/S	LAR/V
VS algorithm	FORWARD	LASSO	LASSO	LAR	LAR
Elementary statistics	$S_{t,k}$	$S_{t,k}$	$V_{t,k}$	$S_{t,k}$	$V_{t,k}$
Critical values					
Unstructured	4.766477	4.578106	4.680325	4.722629	4.715809
Linear profile	5.033247	5.201960	5.011673	4.898843	5.061300
Cubic profile	4.877482	5.140879	4.889752	4.821646	4.870619
Nonparametric profile	5.268651	5.602346	5.546736	5.112253	5.156662
Multistage process	4.677853	4.603569	4.758084	4.726339	4.915792

one or more explanatory variables, and in a multistage process. Details for the components of the vector $\boldsymbol{\beta}$ that are supposed to change are listed, for each change-point model, in the Appendix. In every case, several possible mean shifts are considered, including shifts in a single parameter, equal and different shifts in a pair of parameters, shifts of the same size in either even or odd components and shifts in variance. Here, we briefly describe the scenarios examined in the simulation study.

3.1 Unstructured

In this case $p = n$, the matrix $\boldsymbol{F}$ reduces to the identity matrix $\boldsymbol{F} = \boldsymbol{I}_n$ and the i-th element of $\boldsymbol{\beta}$ directly points to a mean shift of the i-th quality characteristic, i.e., $\delta_i = \beta_i$. Following the example in Zou and Qiu (2009), we consider $p = n = 15$ and assume that the IC distribution is $N_n(\mathbf{0}_n, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (\sigma_{ij}) = (0.75^{|i-j|})$ for $i, j = 1, 2, \ldots, n$ and the OC distribution $N_n(\boldsymbol{\beta}, \omega^2\boldsymbol{\Sigma})$ with $\omega > 1$.

3.2 Linear and Cubic Profiles

Under this scenario, we assume that

$$y_{t,i} = \begin{cases} \epsilon_{t,i} & \text{if } t < \tau \\ \beta_1 + \beta_2 x_i + \cdots + \beta_p x_i^{p-1} + \epsilon_{t,i} & \text{if } t \geq \tau \end{cases}$$

with $x_i = (2i - n - 1)/(n - 1)$, for $i = 1, \ldots, n$. Here, $\epsilon_{t,i}$ are independent, zero-mean, Gaussian random variables, with the IC and OC variance equal to one and $\omega^2 > 1$, respectively. Thus, in the described scenario, $\boldsymbol{F} = (f_{i,j}) = (x_i^{j-1})$, $\boldsymbol{\Sigma} = \boldsymbol{I}_n$ and $\boldsymbol{\Omega} = \omega^2\boldsymbol{I}_n$. In particular, we consider linear ($p = 2$) and cubic ($p = 4$) profiles with $n = 4$ and $n = 8$ observations, respectively.

3.3 Nonparametric Profiles

To investigate the performance of the VS-based control chart for nonparametric monitoring of non-linear profiles, we use the same IC model considered by Zou et al. (2008), $y_{t,i} = 1 - \exp(-x_i) + \epsilon_{t,i}$, where $x_i = (i - 0.5)/20$, $i = 1, \ldots, 20$, and the following three OC models:

I. $y_{t,i} = 1 - \beta_1 \exp(-x_i^{\beta_2}) + \epsilon_{ti}$;
II. $y_{t,i} = 1 - \exp(-x_{t,i}) + \beta_1 \cos(\beta_2 \pi (x_{t,i} - 0.5)) + \epsilon_{t,i}$;
III. $y_{t,i} = 1 - \exp(-x_{t,i} - \beta_1 \max(0, (x_{t,i} - \beta_2)/(1 - \beta_2))^2) + \epsilon_{t,i}$.

Here, $\epsilon_{t,i}$ are independent, zero-mean, Gaussian random variables, with the IC and OC variance equal to one and $\omega^2 > 1$, respectively. In this case, we set $\boldsymbol{F}$ equal to the basis matrix of a cubic spline with four equispaced knots within the interval $[0, 1]$.

3.4 Multistage Processes

We consider an n-state process representable by the linear state-space model

$$\begin{cases} y_{t,i} = \mu_i + c_i x_{t,i} + v_{t,i} \\ x_{t,i} = d_i x_{t,i-1} + \beta_i I_{\{t \geq \tau\}} + w_{t,i} \end{cases} \quad (i = 1, \ldots, n)$$

where $v_{t,i}$ and $w_{t,i}$ are independent normal random variables with zero mean. The n elements of the $\boldsymbol{\beta} = (\beta_i)$ vector define the magnitude of the shifts and the stages at which the shifts occur. It is easy to show that this model is a particular case of (1). See Capizzi and Masarotto (2011) for the details and, in particular, for the structure of the $\boldsymbol{F}$ and $\boldsymbol{\Sigma}$ matrices. In the simulation, we fix the number of stages to $n = 10$ and investigate the performance for different shift locations, occurring in one, two, five and all stages assuming that $\mu_i = 0$, $c_i = d_i = \text{var}(w_{t,i}) = \text{var}(v_{t,i}) = 1$ for every i.

3.5 Results

Results are summarized in Fig. 1, which shows the following percent relative differences

$$100 \cdot \frac{\text{ARL}_{rs} - \text{MARL}_r}{\text{MARL}_r}, \quad s = 1, \ldots, 5, \tag{11}$$

i.e., the percent relative differences between ARL_{rs}, the OC ARL of s-th control chart in the r-th OC scenario, and MARL_r, the mean of the five out-of-control ARL values, one for each control chart, obtained for the r-th OC scenario. Observe that the number of the OC scenarios is different for the different cases. In particular, $r = 1, \ldots, 18$ for the case of nonparametric and multistage process monitoring, $r = 1, \ldots, 23$ for the monitoring of linear and cubic profiles and $r = 1, \ldots, 30$ in the unstructured case (see the Appendix for a detailed description of each OC scenario).

A negative (positive) value of (11) can be interpreted as a quicker reaction (slower) reaction of the s-th control chart to the r-th OC situation, when compared to the other VS-based control charts.

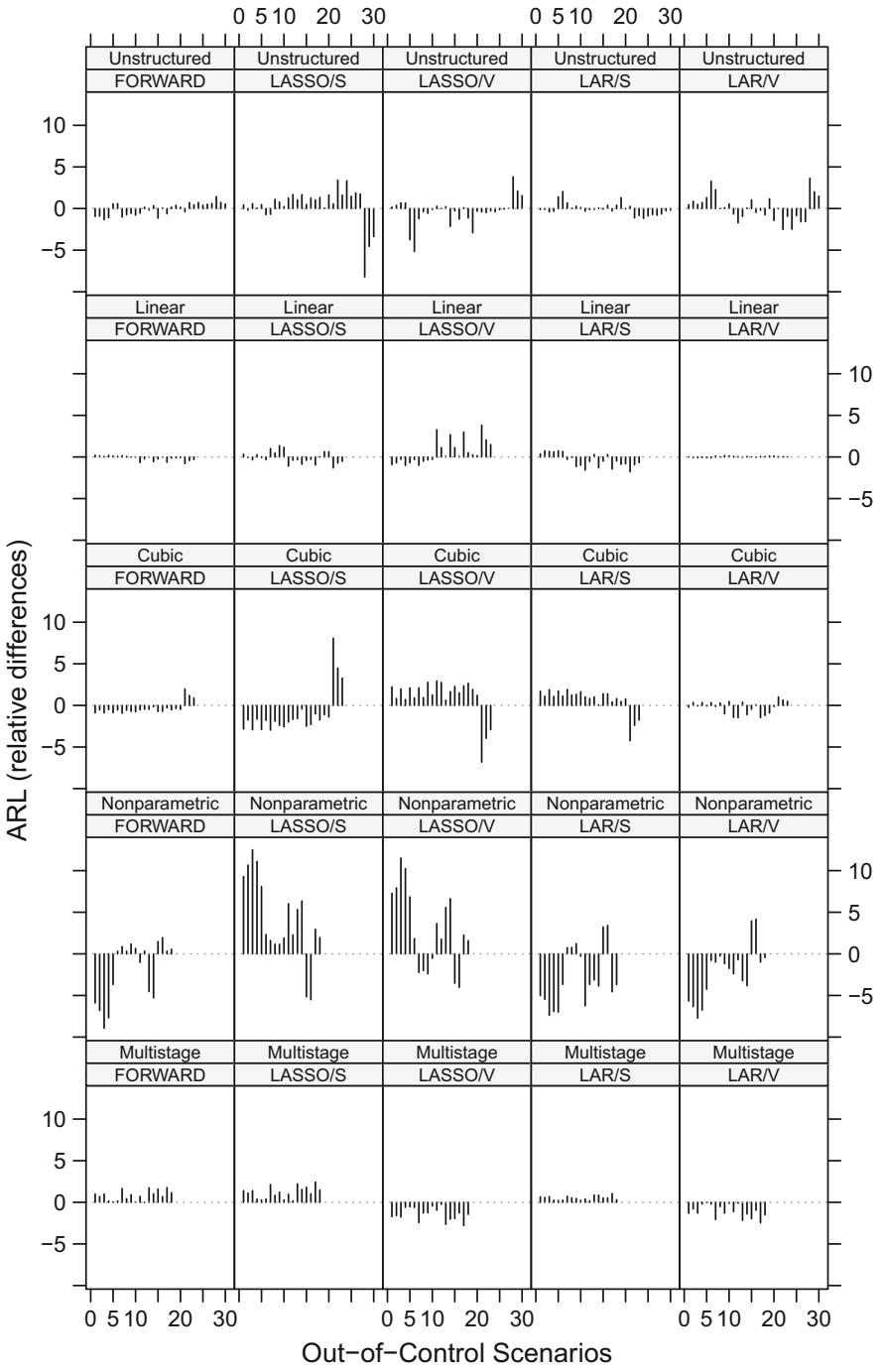


Fig. 1 Relative ARL differences of five VS-based control charts for several OC scenarios (see the Appendix for labels in the x axis)

Results show that, independently from the multivariate control charts, the forward- and the LAR-based schemes show a similar behavior that seems to be quite stable throughout all the different practical contexts. However, while LASSO shows a substantially negligible advantage for some out-of-control situations, it can also show a relatively large degradation for other applications, such as nonparametric profile monitoring. Further, monitoring schemes based on the same VS algorithm but on different elementary control statistics, i.e., on $S_{t,k}$ or $V_{t,k}$, offer essentially the same performance.

4 Conclusions

In this paper, we compared the performance of multivariate control charts based on three different variable selection procedures. In particular, the compared multivariate control charts have been implemented for detecting many out-of-control conditions, even involving increases in process variation, for several MSPC frameworks. Results show that whereas control charts consisting of the aggregation of several control statistics, such as those proposed by Zou and Qiu (2009) and Capizzi and Masarotto (2011), behave quite similarly for different OC applications, suggestions can be given to practitioners concerning the particular variable selection procedure to use. As discussed before, the LASSO-based control charts can show an unsatisfactory performance in detecting some particular OC situations. However, the forward and LAR-based procedures can be considered substantially equivalent in terms of OC ARL performance for the investigated practical applications. Thus, from a practical point of view, multivariate control charts based on forward selection could be more appealing to users since these charts are more intuitive and simpler to implement.

Appendix

In the following, we provide details on the OC scenarios listed in the x axes of Fig. 1. When only variables or stages with an even (odd) index are subject to a shift of size δ, the OC scenario is indicated with either Even[δ] or Odd[δ].

1. *Unstructured model*: $1 = (\beta_1 = 0.5)$, $2 = (\beta_1 = 1)$, $3 = (\beta_3 = 0.5)$, $4 = (\beta_3 = 1)$, $5 = (\beta_1 = 0.5,\ \beta_2 = 0.25)$, $6 = (\beta_1 = 0.5,\ \beta_2 = 0.5)$, $7 = (\beta_1 = 0.5,\ \beta_2 = 0.75)$, $8 = (\beta_1 = 0.5,\ \beta_3 = 0.25)$ $9 = (\beta_1 = 0.5,\ \beta_3 = 0.5)$, $10 = (\beta_1 = 0.5,\ \beta_3 = 0.75)$, $11 = (\beta_3 = 0.5,\ \beta_8 = 0.25)$, $12 = (\beta_3 = 0.5,\ \beta_8 = 0.5)$, $13 = (\beta_3 = 0.5,\ \beta_8 = 0.75)$, $14 = \beta_1 = 0.5,\ \beta_2 = 0.25,\ \beta_3 = 0.25)$, $15 = (\beta_1 = 0.25,\ \beta_2 = 0.25,\ \beta_3 = 0.5)$, $16 = (\beta_2 = 0.5,\ \beta_3 = 0.25,\ \beta_8 = 0.25)$, $17 = (\beta_2 = 0.25,\ \beta_3 = 0.25,\ \beta_8 = 0.5)$, $18 = (\beta_7 = 0.5,\ \beta_8 = 0.25,\ \beta_9 = 0.5)$, $19 = (\beta_7 = 0.25,\ \beta_8 = 0.75,\ \beta_9 = 0.5)$, $20 = (\beta_6 = 0.5,\ \beta_8 = 0.25,\ \beta_{10} = 0.5)$, $21 = (\beta_6 = 0.25,\ \beta_8 = 0.75,\ \beta_{10} = 0.5)$,

22 = (Even[0.25]), 23 = (Even[0.5]), 24 = (Odd[0.25]), 25 = (Even[0.5]), 26 = (Even[0.5], Odd[0.25]), 27 = (Odd[0.5], Even[0.25]), 28 = ($\omega = 1.2$), 29 = ($\omega = 1.5$), 30 = ($\omega = 2$).

2. *Linear profiles*: 1 = ($\beta_1 = 0.1$), 2 = ($\beta_1 = 0.3$), 3 = ($\beta_2 = 0.2$), 4 = ($\beta_2 = 0.5$), 5 = ($\beta_3 = 0.2$), 6 = ($\beta_3 = 0.5$), 7 = ($\beta_1 = 0.1,\ \beta_2 = 0.1$), 8 = ($\beta_1 = 0.4,\ \beta_2 = 0.2$) 9 = ($\beta_1 = 0.4,\ \beta_2 = 0.6$), 10 = ($\beta_1 = 0.6,\ \beta_2 = 0.6$), 11 = ($\beta_1 = 0.1,\ \omega = 1.2$), 12 = ($\beta_1 = 0.3,\ \omega = 1.2$), 13 = ($\beta_1 = 1,\ \omega = 1.2$), 14 = ($\beta_2 = 0.1,\ \omega = 1.2$), 15 = ($\beta_2 = 0.4,\ \omega = 1.2$), 16 = ($\beta_1 = 1.2,\ \omega = 1.2$), 17 = ($\beta_1 = 0.1,\ \beta_2 = 0.1,\ \omega = 1.2$), 18 = ($\beta_1 = 0.4,\ \beta_2 = 0.2,\ \omega = 1.2$), 19 = ($\beta_1 = 0.4,\ \beta_2 = 0.6,\ \omega = 1.2$), 20 = ($\beta_1 = 0.6,\ \beta_2 = 0.6,\ \omega = 1.2$), 21 = ($\omega = 1.2$), 22 = ($\omega = 1.5$), 23 = ($\omega = 2$).
3. *Cubic profiles*: 1 = ($\beta_1 = 0.1$), 2 = ($\beta_1 = 0.3$), 3 = ($\beta_1 = 1$), 4 = ($\beta_2 = 0.2$), 5 = ($\beta_2 = 0.4$), 6 = ($\beta_2 = 1.2$), 7 = ($\beta_4 = 0.2$), 8 = ($\beta_4 = 0.5$), 9 = ($\beta_1 = 0.1,\ \beta_2 = 0.1$), 10 = ($\beta_1 = 0.1,\ \beta_3 = 0.2$) 11 = ($\beta_1 = 0.1,\ \beta_4 = 0.2$), 12 = ($\beta_2 = 0.2,\ \beta_3 = 0.2$), 13 = ($\beta_2 = 0.4,\ \beta_4 = 0.2$), 14 = ($\beta_3 = 0.5,\ \beta_4 = 0.5$), 15 = ($\beta_1 = 0.1,\ \beta_2 = 0.1,\ \beta_3 = 0.1$), 16 = ($\beta_1 = 0.1,\ \beta_3 = 0.2,\ \beta_4 = 0.1$), 17 = ($\beta_2 = 0.1,\ \beta_3 = 0.3,\ \beta_4 = 0.2$), 18 = ($\beta_1 = 0.1,\ \beta_2 = 0.1,\ \beta_3 = 0.1,\ \beta_4 = 0.1$), 19 = ($\beta_1 = 0.1,\ \beta_2 = 0.2,\ \beta_3 = 0.1,\ \beta_4 = 0.2$), 20 = ($\beta_1 = 0.2,\ \beta_2 = 0.1,\ \beta_3 = 0.2,\ \beta_4 = 0.1$), 21 = ($\omega = 1.2$), 22 = ($\omega = 1.5$), 23 = ($\omega = 2$).
4. *Non parametric profiles*. The following OC scenarios are referred to possible shifts in the regression coefficients of models I, II and III. 1 = (I, $\beta_1 = 1.00,\ \beta_2 = 1.30$), 2 = (I, $\beta_1 = 1.00,\ \beta_2 = 1.50$), 3 = (I, $\beta_1 = 1.10,\ \beta_2 = 1.00$) 4 = (I, $\beta_1 = 1.30,\ \beta_2 = 1.00$), 5 = (I, $\beta_1 = 1.20,\ \beta_2 = 1.00,\ \omega = 1.10$), 6 = (I, $\beta_1 = 1.00,\ \beta_2 = 1.20,\ \omega = 1.30$), 7 = (II, $\beta_1 = 0.10,\ \beta_2 = 3.00$), 8 = (II, $\beta_1 = 0.30,\ \beta_2 = 3.00$), 9 = (II, $\beta_1 = 0.10,\ \beta_2 = 2.00$) 10 = (II, $\beta_1 = 0.30,\ \beta_2 = 2.00$), 11 = (II, $\beta_1 = 0.20,\ \beta_2 = 4.00,\ \omega = 1.10$), 12 = (II, $\beta_1 = 0.20,\ \beta_2 = 4.00,\ \omega = 1.30$) 13 = (III, $\beta_1 = 2.00,\ \beta_2 = 0.90$), 14 = (III, $\beta_1 = 4.00,\ \beta_2 = 0.90$), 15 = (III, $\beta_1 = 2.00,\ \beta_2 = 0.75$), 16 = (III, $\beta_1 = 4.00,\ \beta_2 = 0.75$), 17 = (III, $\beta_1 = 2.00,\ \beta_2 = 0.90,\ \omega = 1.20$), 18 = (III, $\beta_1 = 4.00,\ \beta_2 = 0.75,\ \omega = 1.20$).
5. *Multistage process*. In the following, β_j indicates the shift of size β, occurring at the j-th stage, $j = 1, \ldots, 20$. 1 = ($\beta_1 = 0.75$), 2 = ($\beta_5 = 0.75$), 3 = ($\beta_{10} = 0.75$), 4 = ($\beta_1 = 1.5$), 5 = ($\beta_5 = 1.5$), 6 = ($\beta_{10} = 1.5$), 7 = ($\beta_2 = 0.6,\ \beta_8 = 0.6$), 8 = ($\beta_4 = 0.6,\ \beta_5 = 0.6$), 9 = ($\beta_2 = 1.2,\ \beta_8 = 1.2$), 10 = ($\beta_4 = 1.2,\ \beta_5 = 1.2$), 11 = ($\beta_2 = 1.8,\ \beta_8 = 1.8$), 12 = ($\beta_4 = 1.8,\ \beta_5 = 1.8$), 13 = ($\beta_1 = \beta_5 = \beta_{10} = 0.2,\ \beta_3 = \beta_7 = 0.4$), 14 = ($\beta_3 = \beta_5 = \beta_7 = 0.2,\ \beta_4 = \beta_6 = 0.4$), 15 = ($\beta_1 = \beta_5 = \beta_{10} = 0.6,\ \beta_3 = \beta_7 = 0.4$), 16 = ($\beta_3 = \beta_5 = \beta_7 = 0.6,\ \beta_4 = \beta_6 = 0.4$), 17 = (Even[0.1], Odd[0.2]), 18 = (Even[0.5], Odd[0.25]).

References

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International, 23*, 517–543.

Capizzi, G., & Masarotto, G. (2011). A least angle regression control chart for multidimensional data. *Technometrics, 53*, 285–296.

Efron, E., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32*, 407–499.

Jiang, W., Wang, K., & Tsung, F. (2012) A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis. *Journal of Quality Technology, 44*, 209–230.

Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted movin average control schemes: properties and enhancements. *Technometrics, 32*, 1–29.

Polyak, B. T., & Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal of Control Optmization, 30*, 838–855.

Prabhu, S. S., & Runger, G. C. (1997). Designing a multivariate EWMA control chart. *Journal of Quality Technology, 29*, 8–15.

Reynolds, M. R., & Stoumbos, Z. G. (2005). Should exponentially weighted moving average and cumulative sum charts be used with Shewhart limits? *Technometrics, 47*(4), 409–424.

Reynolds, M. R., & Stoumbos, Z. G. (2006). Comparisons of some exponentially weighted moving average control charts for monitoring the process mean and variance. *Technometrics, 48*(4), 550–567.

Ruppert, D. (1991). Stochastic approximation. In B. K. Ghosh, & P. K. Sen (Eds.), *Handbook of sequential analysis* (pp. 503–529). New York, NY: Marcel Dekker.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B, 58*, 267–288.

Wang, K., & Jiang, W. (2009). High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology, 41*, 247–258.

Zou, C., Ning, X., & Tsung, F. (2010). Lasso-based multivariate linear profile monitoring. *Annals of Operation Research, 192*, 3–19.

Zou, C., & Qiu, P. (2009). Multivariate statistical process control using LASSO. *Journal of American Statistical Association, 104*, 1586–1596.

Zou, C., Tsung, F., & Wang, Z. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics, 50*, 512–526.

The Use of Inequalities of Camp-Meidell Type in Nonparametric Statistical Process Monitoring

Rainer Göb and Kristina Lurz

Abstract A few authors have used the classical Camp-Meidell inequality for the nonparametric analysis of statistical process monitoring. The following issues have not received sufficient attention. (i) The use of moments of order higher than 2 in the inequalities provides tighter bounds. (ii) The problem of estimating the moments in the bounds, e.g., from a phase 1 sample, cannot be neglected. The present study analyses both aspects (i) and (ii). Appropriate estimators, their properties, and the effect of estimation on the properties of process monitoring charts are investigated. In particular, the use of empirical Camp-Meidell bounds in quantile control charts is studied.

Keywords Quantile bounds • Shewhart charts • Zero inflated data

1 Introduction

Historically, the mathematical theory of statistical process monitoring has strongly been concentrating on the paradigm of normally distributed observations. Process monitoring emerged from the manufacturing industries. However, the normality assumption is far from being universally valid for data from manufacturing processes as already noted by Shewhart (1931) in his seminal work on process monitoring. The normal distribution is even less common in more recent targets of process monitoring like service, finance, or logistics, see Pyzdek (1995) analysis, for instance. The abundant use of the normality assumption in the process monitoring literature seems to be disproportionate to the actual importance of the normal distribution in field practice. Several authors have investigated the fallacious effect of erroneous normality assumptions on the properties of process monitoring

R. Göb (✉) • K. Lurz
Institute for Applied Mathematics and Statistics, University of Würzburg, Sanderring 2, 97070 Würzburg, Germany
e-mail: goeb@mathematik.uni-wuerzburg.de; kristina.lurz@uni-wuerzburg.de

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_11

procedures, e.g., Burr (1967), Schilling and Nelson (1976), Yourstone and Zimmer (1992), Spedding and Rawlings (1994), Samanta and Bhattacherjee (2004). Distrust against the normality assumption has motivated two lines of research: (1) the so-called parametric approach studies process monitoring procedures under particular parametric distribution families other than the normal, whereas (2) the so-called nonparametric approaches do not rely on particular distribution assumptions.

Early instances of the approach (1) are Ferrell's (1958) and Morrison's (1958) studies on control charts under the lognormal distribution. Later, the topic of process monitoring under specific distributions has generated a large number of literature contributions. The Weibull distribution received particular interest, see the studies by Johnson (1966), Nelson (1979), Ramalhoto and Morais (1998, 1999), Kanji and Arif (2001), Zhang and Chen (2004), Erto and Pallotta (2007), Erto et al. (2008), Guo and Wang (2012). Yang and Xie (2000) consider the exponential distribution. Further studies on the lognormal distribution were provided by Joffe and Sichel (1968), Kotz and Lovelace (1988), Cheng and Xie (2000), Areepong and Sukparungsee (2010). González and Viles (2000, 2001) and Kantam and Sriram (2001) consider the gamma distribution. Hardly any distribution seems to have escaped the attention of interested authors: Subba and Kantam (2008) study the double exponential distribution, Edgeman (1989) considers the inverse Gaussian, Kantam et al. (2006) assume the log-logistic distribution, Betul and Yaziki (2006) discuss control charts for the Burr distribution.

The practical usefulness of process monitoring procedures tailored for particular distribution classes must be doubted. Such procedures provide a benefit only if the specific distribution type is correctly identified. The latter is often impossible in practice, due to a lack of statistical expertise, or due to small sample sizes. Such problems are avoided by *nonparametric* or *distribution-free* monitoring schemes. These do not require to identify a particular distribution family but only some basic properties of the underlying distribution, e.g., skewed or symmetric. Various ideas have come up to avoid the use of a particular reference distribution in the design of process monitoring procedures, see the partial survey by Chakraborti et al. (2001). Basically, two lines of distribution-free inference can be distinguished: (i) inference based on small online samples from the process, and (ii) inference from training samples drawn under guaranteed in-control operation of the process, e.g., in phase I of a control chart, see Woodall (2000). Most of the approaches of type (i) are based on sign or rank statistics, see the studies by Bakir and Reynolds (1979), Hackl and Ledolter (1992), Amin et al. (1995), Bakir (2004, 2006), Chakraborti and Eryilmaz (2007). A large group of approaches of type (ii) use bootstrapping, see Leger et al. (1992), Seppala et al. (1995), Liu and Tang (1996), Teyarachakul et al. (2007), Chatterjee and Qiu (2009), and the discussion by Jones and Woodall (1998). Some further heterogeneous approaches of type (ii) based on in-control training samples are suggested by Park and Reynolds (1987), Hackl and Ledolter (1991), Willemain and Runger (1996), Qiu and Li (2011). Winterbottom (1993), Chen (1998), Chan and Cui (2003) consider p charts; Wu and Wang (2009) suggest control limits based on the quantile expansion developed by Cornish and Fisher (1938). The parameters of the expansion, namely the third and fourth standardised cumulant, have to be

estimated from in-control reference samples. However, the referenced studies give an oversimplified account of the underlying estimation tasks. In particular, the problem of the high variance of the standardised cumulant estimators are ignored.

Stochastic inequalities, in particular Chebyshev's (1867) and the inequality established by Camp (1922) and Meidell (1922), have often been referenced to evaluate and justify the use of simple 3-sigma control limits under arbitrary in-control distributions, see Shewhart (1931), Florac and Carleton (1999), Grant and Leavenworth (1999), for instance. For the proper design of control limits, however, Chebyshev's inequality and the standard form of the Camp-Meidell inequality are too unprecise. Apparently, the only closer analysis of the Camp-Meidell inequality in the context of nonparametric statistical process control is due to Ion (2001) thesis. However, higher order versions of the inequality remain unconsidered in this thesis.

The present study investigates the use of higher order Camp-Meidell inequalities in the design of control charts. The Camp-Meidell inequality rigorously holds for symmetric unimodal distributions only. In the nonparametric context, many studies have been accounting for skewness, as a property which marks a clear difference from the classical Gaussian paradigm, see the contributions by Choobineh and Ballard (1987), Bai and Choi (1995), Wu (1996), Chan and Cui (2003), Chang and Bai (2001), Derya and Canan (2012), for instance. Nevertheless, symmetric distributions play an important role for process modeling. In particular, symmetric distributions are often suitable models for measurement error, see the "Guide to the Expression of Uncertainty in Measurement", ISO (2008), or for financial returns. In the class of unimodal symmetric distributions, the normal distribution can be an inadequate model because of its light tails, see Haas and Pigorsch (2009). Most of the classical unimodal symmetric distribution models like Laplace, logistic, Student's t have considerably higher probability mass on the extremes than the normal distribution. The quantiles of the true underlying distribution are thus often seriously underestimated from a normal distribution model. Control limits designed for an in-control normal distribution will lead to a large number of false alarms.

The study also approaches the problem of *point-inflated data*, i.e., random observations with a positive probability mass at some point of the real line. In the case of symmetric unimodal distributions, the interesting case is a positive probability mass occurring at the expectation μ_X of the considered random variable X, so that $X - \mu_X$ is *zero-inflated*. Such data occur particularly in the context of financial auditing where deviations of the true audit values from the stipulated book values of inventories, debtor or creditor accounts in either direction are infrequent.

The study is organised in the following sections. Section 2 provides theoretical quantile bounds of Camp-Meidell type. The estimation problems in the empirical use of the Camp-Meidell quantile bounds are discussed by Sect. 3. Section 4 studies an estimator for the Camp-Meidell quantile bounds. The usage of the empirical bounds in the design of a Shewhart individual observation control chart is described by Sect. 5. Section 6 discusses the achieved results and outlines remaining research topics.

2 The Camp-Meidell Inequality and Resulting Quantile Bounds

The subsequent theorem 1 presents a version of the inequality established by Camp (1922) and Meidell (1922) which is generalised in two respects: (i) it includes distributions with a zero inflation at the mean, and (ii) it clearly describes the type of distributions for which the upper bound of the inequality is exact. The theorem considers *mean-modal random variables X with finite central moments* with the following detailed properties: M1) there is a real a and independent random variables I, X_0 with $X = a + IX_0$; M2) I has a binomial distribution $Bi(1, 1-q)$ where $0 \leq q < 1$; M3) X_0 is absolutely continuous, $E[X_0] = 0$, and the density f_{X_0} is increasing on $(-\infty; 0)$ and decreasing $(0; +\infty)$; M4) X_0 has finite central moments. M1) through M4) imply: $\mu_X = E[X] = a$; the central moments $\mathrm{m}_{X,s} = E[(X - \mu_X)^s] = (1-q)\mu_{X_0,s}$ are finite; for $q > 0$, the distribution of X is inflated at its mean μ_X with probability mass $\mathrm{P}(X = \mu_X) = q$; the conditional distribution of X under $X \neq \mu_X$ is absolutely continuous with a density f_X increasing on $(-\infty; \mu_X)$ and decreasing on $(\mu_X; +\infty)$. In particular, mean-modality holds for symmetric unimodal random variables.

Theorem 1 (Two-Sided Camp-Meidell Inequality) *Let X be a mean-modal random variable with finite central moments $\mathrm{m}_{X,s} = E[(X - \mu_X)^s]$, see the above definition in the introduction of Sect. 2. For $0 \leq p \leq 1$, $\vartheta > 0$ let $Y_{p,\vartheta}$ be a random variable with the following properties: (i) $P(Y_{p,\vartheta} = \mu_X) = 1 - p$, (ii) $P(Y_{p,\vartheta} < \mu_X) = p/2 = P(Y_{p,\vartheta} > \mu_X)$, (iii) the conditional distribution of $Y_{p,\vartheta}$ under $Y_{p,\vartheta} < \mu_X$ is the rectangular distribution on the support $[\mu_X - \vartheta; \mu_X]$, i.e.,*

$$P(Y_{p,\vartheta} \leq y | Y_{p,\vartheta} < \mu_X) = \frac{p}{2} \frac{y - (\mu_X - \vartheta)}{\vartheta} \quad \text{for } \mu_X - \vartheta < y \leq \mu_X, \tag{1}$$

(iv) the conditional distribution of $Y_{p,\vartheta}$ under $Y_{p,\vartheta} > \mu_X$ is the rectangular distribution on the support $[\mu_X; \mu_X + \vartheta]$, i.e.,

$$P(Y_{p,\vartheta} \geq y | Y_{p,\vartheta} > \mu_X) = \frac{p}{2} \left(1 - \frac{y - \mu_X}{\vartheta}\right) \quad \text{for } \mu_X < y \leq \mu_X + \vartheta. \tag{2}$$

The CDF of $Y_{p,\vartheta}$ is symmetric about μ_X and satisfies

$$F_{Y_{p,\vartheta}}(y) = \begin{cases} \frac{p}{2} \frac{y - (\mu_X - \vartheta)}{\vartheta} & \text{for } \mu_X - \vartheta < y \leq \mu_X, \\ 1 - \frac{p}{2} & \text{for } y = \mu_X, \\ 1 - \frac{p}{2} + \frac{p}{2} \frac{y - \mu_X}{\vartheta} & \text{for } \mu_X \leq y \leq \mu_X + \vartheta, \end{cases} \tag{3}$$

and we have $\mu_{Y_{p,\vartheta}} = \mu_X$. The following assertions hold for $x > 0$:

(a) In the case $x < [\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}$ *we have*

$$\begin{aligned} \frac{\mathrm{m}_{X,2r}}{\left(x\frac{2r+1}{2r}\right)^{2r}} \;&>\; 1 - \frac{x}{[\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}} \;= \\ P(|Y_{p_0,\vartheta_0} - \mu_{Y_{p_0,\vartheta_0}}| \geq x) \;&\geq\; P(|X - \mu_X| \geq x), \end{aligned} \tag{4}$$

where $p_0 = 1$, $\vartheta_0 = [\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}$. *Then* $E[Y_{p_0,\vartheta_0}^{2r}] = \mathrm{m}_{X,2r}$.

(b) In the case $x \geq [\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}\frac{2r}{2r+1}$ *we have*

$$\begin{aligned} 1 - \frac{x}{[\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}} \;&\geq\; \frac{\mathrm{m}_{X,2r}}{\left(x\frac{2r+1}{2r}\right)^{2r}} \;= \\ P(|Y_{p_0,\vartheta_0} - \mu_{Y_{p_0,\vartheta_0}}| \geq x) \;&\geq\; P(|X - \mu_X| \geq x) \end{aligned} \tag{5}$$

where

$$p_0 \;=\; \frac{\mathrm{m}_{X,2r}(2r+1)}{\left(x\frac{2r+1}{2r}\right)^{2r}}, \quad \vartheta_0 \;=\; \frac{x(2r+1)}{2r}, \quad E[Y_{p_0,\vartheta_0}^{2r}] \;=\; \mathrm{m}_{X,2r}.$$

A proof of Theorem 1 is provided by Göb and Lurz (2013a). Inverting the inequalities stated by Theorem 1 provides the quantile bounds of the subsequent theorem 2.

Theorem 2 (Camp-Meidell Quantile Bounds) *Let* X *be a mean-modal random variable* X *with finite central moments, see the above definition in the introduction of Sect. 2. Let* $r > 0$, $0 < \alpha < 1$,

$$x_{C,r,\alpha} = \begin{cases} (1-\alpha)[\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}, & \text{if } \alpha \geq \frac{1}{2r+1}, \\ \left(\frac{\mathrm{m}_{X,2r}}{\alpha}\right)^{\frac{1}{2r}}\frac{2r}{2r+1} = \sigma_X\left(\frac{c_{X,2r}}{\alpha}\right)^{\frac{1}{2r}}\frac{2r}{2r+1}, & \text{if } \alpha \leq \frac{1}{2r+1}. \end{cases}$$

For $0 \leq p \leq 1$, $\vartheta > 0$ *let* $Y_{p,\vartheta}$ *be defined as in Theorem 1. Then we have for* $0 < \alpha < 1$

$$P(|X - \mu_X| \geq x_{C,r,\alpha}) \;\leq\; \alpha \;=\; P(|Y_{p,\vartheta} - \mu_{Y_{p,\vartheta}}| \geq x_{C,\alpha}) \tag{6}$$

where in the case $\alpha \geq \frac{1}{2r+1}$ *we have* $p = 1$, $\vartheta = [\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}$, *and where in the case* $\alpha \leq \frac{1}{2r+1}$ *we have* $p = \alpha(2r+1)$, $\vartheta = [\mathrm{m}_{X,2r}/\alpha]^{1/(2r)}$. *If the distribution of* X *is symmetric around* μ_X, *we obtain by letting* $z_{C,r,X}(\rho) = \mu_X + x_{C,2(1-\rho)}$ *for*

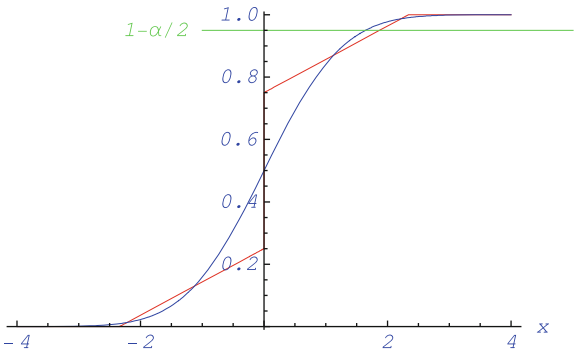


Fig. 1 Normal distribution $N(0,1)$, $\alpha = 0.10$

$0 < \rho < 1$ *the inequality*

$$z_X(\rho) \leq z_{C,r,X}(\rho) = \begin{cases} \mu_X + (2\rho - 1)[\mathrm{m}_{X,2r}(2r+1)]^{1/(2r)}, & \text{if } \rho \leq \frac{4r+1}{4r+2}, \\ \mu_X + \left(\frac{\mathrm{m}_{X,2r}}{2(1-\rho)}\right)^{\frac{1}{2r}} \frac{2r}{2r+1}, & \text{if } \rho \geq \frac{4r+1}{4r+2}, \end{cases} \tag{7}$$

where $z_X(\rho)$ *is the one-sided level* ρ *quantile defined by* $F_X(z_X(\rho)) = \rho$.

Figure 1 displays the CDF of the maximising variable $Y_{p,\vartheta}$ from Theorem 2 for $\alpha = 0.1$, i.e., $\rho = 1 - \alpha/2 = 0.95$, $2r = 4$, under the standard normal distribution $N(0,1)$. In this case, we have $\mathrm{m}_{X,2r} = 3$, $\rho \geq 0.9 = (2r-1)/(2r+1)$. The distribution of $Y_{p,\vartheta}$ collects the weight $p = \alpha(2r+1) = 0.5$ at $\mu_X = 0$. The exact $N(0,1)$ quantile is $z_X(\rho) = z_X(0.95) = 1.64485$, the bound from the inequality (7) is 1.87228. The maximising distribution established by Theorem 2 has a discrete point mass $\mathrm{P}(Y_{p,\vartheta} = \mu_X) = 1 - p$ at $y = \mu_X$, i.e., the distribution is inflated at the point $y = \mu_X$. The closer a distribution to inflation at the mean, the more accurate are the quantile bounds. Such shapes occur particularly in the context of financial auditing where deviations of the true audit values from the stipulated book values of inventories, debtor or creditor accounts in either direction are infrequent.

Subsequently, we evaluate the accuracy of the Camp-Meidell bounds for some classical symmetric unimodal distributions which are often assumed in SPC literature. Figure 2 compares the accuracy of the one-sided quantile bound for the parameter choices $r = 1, 2, 3, 4$ for the normal, Laplace, logistic, and central t-distribution. The graphs plot the ratio $z_{C,r,X}(\rho)/z_X(\rho)$ for quantile levels $0.90 < \rho$. The standard form of the Camp-Meidell inequality considers the value $r = 1$ only, see Patel et al. (1976) or Ion (2001). However, for the examples considered in Fig. 2, $r = 1$ performs poorly, particularly for large quantile levels ρ. $r = 2$ performs consistently well, except for very large values of ρ above 0.995, where $r \geq 3$ can achieve better results. The discussion in Sects. 3 and 4 show that the empirical use of

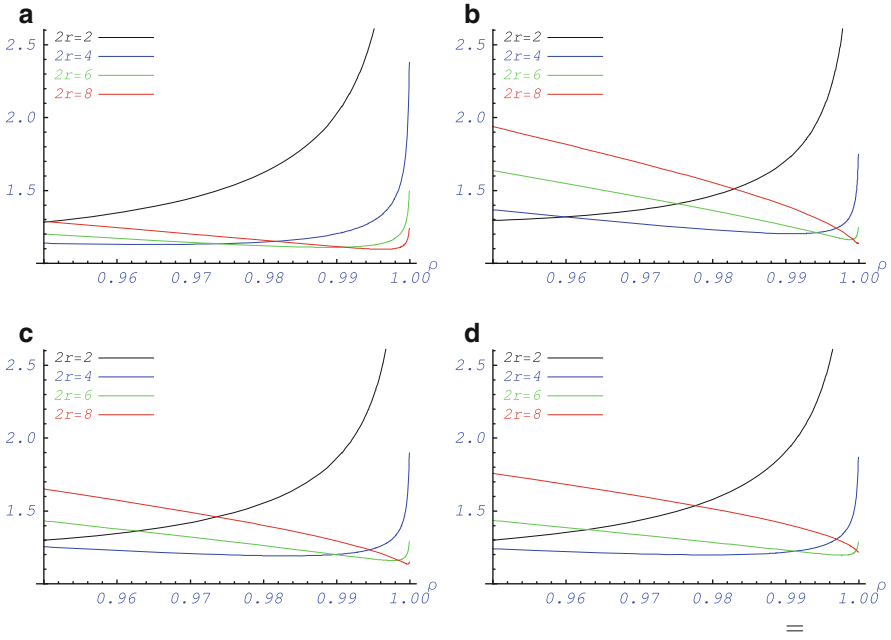


Fig. 2 Accuracy of Camp quantile bounds from Theorem 2 for symmetric unimodal absolutely continuous distributions. (**a**) Normal distribution $N(\mu, \sigma^2)$. (**b**) Laplace distribution $LAP(\mu, \beta)$. (**c**) Logistic distribution $LGT(\mu, \beta)$. (**d**) Central t-distribution $t(\nu)$, $\nu = 10.0$

Camp-Meidell quantile bounds with $r > 2$ involves particular estimation problems. All in all, the use of $r = 2$ is recommended in practice.

3 Estimation Problems in the Empirical Use of Camp-Meidell Quantile Bounds

The empirical use of the Camp-Meidell quantile bounds established by Theorem 2 requires the estimation of the roots $\mathrm{m}_{X,2r}^{1/(2r)}$ of the central moments $\mathrm{m}_{X,2r}$ from observations $X_i, \ldots, X_n$. At first glance, it seems obvious to use the root $M_{n,X,2r}^{1/(2r)}$ of the sample central moment $M_{n,X,2r} = \frac{1}{n}\sum(X_i - \overline{X})^{2r}$ as an estimator. However, this estimator is unsatisfactory. First, the sample central moment $M_{n,X,2r} = \frac{1}{n}\sum(X_i - \overline{X})^{2r}$ is not unbiased for $\mathrm{m}_{X,2r}$, see Cramér (1946). Particularly in the case $2r = 4$, $M_{n,X,2r}$ seriously underestimates $\mathrm{m}_{X,2r}$ for all classical symmetric unimodal distribution models, see Göb and Lurz (2013a). Second, the function $y \mapsto y^{1/(2r)}$ is strictly concave. Hence the well-known Jensen inequality implies $E[M_{n,X,2r}^{1/(2r)}] < E[M_{n,X,2r}]^{1/(2r)}$ for random variables X which are not almost

surely constant. Thus the observed underestimation of $M_{n,X,2r}$ for $\mathrm{m}_{X,2r}$ is further intensified when estimating $\mathrm{m}_{X,2r}^{1/(2r)}$ by $M_{n,X,2r}^{1/(2r)}$.

Let us first consider the estimation of the central moment $\mathrm{m}_{X,2r}$. A distribution-free unbiased estimator can be found by expressing moments through cumulants, and using Fisher's (1928) k-statistics as the unbiased cumulant estimators. For the case $2r = 4$, this leads to an estimator $a(n)M_{n,X,4} + b(n)M_{n,X,2}^2$ where $a(n)$ and $b(n)$ are rational functions of the sample size n, see Cramér (1946, p.352) . For larger r, the unbiased estimator rests on a considerably involved expression.

For the normal distribution, we can adjust the sample moment by a simple coefficient to obtain

$$K_{n,X,2r} = \left(\frac{n}{n-1}\right)^r M_{n,X,2r} \tag{8}$$

as an unbiased estimator for $\mathrm{m}_{X,2r}$, see Göb and Lurz (2013a) for a proof. For standardised moments $\mathrm{m}_{X,s}/\sigma_X^s$, it is often recommended to use estimators which are unbiased under the normal distribution also under other distribution families, see Joanes and Gill (1998). Following this strategy, Göb and Lurz (2013a) compare the performance of the distribution-free unbiased estimator with the performance of the estimator $K_{n,X,2r}$ for the case $2r = 4$ under several symmetric distributions. For small sample sizes n in the range of $n \leq 10$, $K_{n,X,4}$ underestimates $\mathrm{m}_{X,4}$ at 10 to 20 %, for sample sizes $n \geq 25$ the underestimation is rather negligible. However, the estimator $K_{n,X,4}$ strikingly outperforms the distribution-free unbiased estimator with respect to the mean square error (MSE) $E[(K_{n,X,4} - \mathrm{m}_{X,4})^2]$. Particularly for small n, the distribution-free unbiased estimator $a(n)M_{n,X,4} + b(n)M_{n,X,2}^2$ has an excessive variance, resulting from the strong correlation between the sample moments $M_{n,X,4}$ and $M_{n,X,2}$. The results are illustrated in Fig. 3 for the Laplace distribution. Concluding from this pattern, the simple estimator $K_{n,X,4}$ is a recommendable alternative to the distribution-free unbiased estimator under symmetric distributions. An unbiased or nearly unbiased estimator for $\mathrm{m}_{X,2r}$ does not provide directly a suitable solution for the estimation of $\mathrm{m}_{X,2r}^{1/(2r)}$ since the root induces underestimation on the average, see the above remarks. Distribution-free unbiased estimators for $\mathrm{m}_{X,2r}^{1/(2r)}$ are not known. Approximately unbiased estimators are investigated by

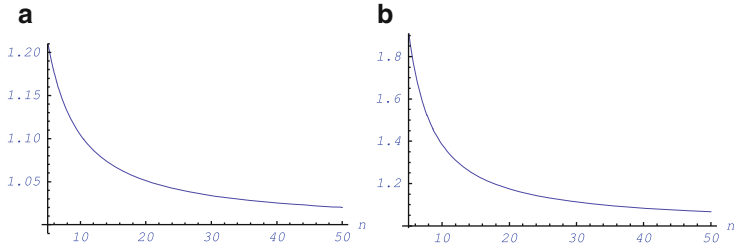


Fig. 3 Comparison of the distribution-free unbiased estimator for $\mathrm{m}_{X,4}$ with the estimator $K_{n,X,4}$ under the Laplace distribution. (**a**) $\mathrm{m}_{X,4}/E[K_{n,X,4}]$. (**b**) MSE of unbiased estimator/MSE$[K_{n,X,4}]$

Göb and Lurz (2013a). For the purpose of control charting, a simple and easily implementable solution is required. Motivated by the good performance of the estimator $K_{n,X,2r}$ for $\mathrm{m}_{X,2r}$ we consider an analogous solution for the estimation of $\mathrm{m}_{X,2r}^{1/(2r)}$: correct the root $M_{n,X,2r}^{1/(2r)}$ by a coefficient $c_{n,2r}$ to obtain an estimator

$$R_{n,X,2r} = c_{n,2r} M_{n,X,2r}^{1/(2r)} \tag{9}$$

where the coefficient $c_{n,2r}$ is chosen to make $R_{n,X,2r}$ unbiased for a baseline symmetric unimodal distribution. Asymptotically, all choices of the baseline distribution are equivalent: the central limit theorem and the delta method show that $M_{n,X,2r}^{1/(2r)}$ is asymptotically distributed by a normal distribution with mean $\mathrm{m}_{X,2r}^{1/(2r)}$. The problem is to choose a baseline distribution which provides good results also for small sample sizes. One reasonable choice for the baseline model is the normal distribution. If interest is rather in distributions with stronger weights on the tails, the Laplace distribution or hyperbolic distributions are suitable alternatives.

Under the normal distribution, the unbiasing coefficient for the case $r = 1$ is

$$c_{n,2r} = c_{n,2} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}. \tag{10}$$

The latter coefficients are tabulated in most statistical quality control textbooks. For $r \geq 2$, the unbiasing coefficients cannot be expressed in closed form. The coefficients for $r = 2, 3, 4$ displayed in Table 1 were calculated by simulation. The number of simulation runs was chosen so as to keep the length of the empirical 99 % confidence interval smaller than 2.5×10^{-3}. N varies from $N = 1.5 \times 10^6$ for small sample size n down to $N = 2.5 \times 10^5$ for large sample size $n > 50$. Burr (1967) demonstrates that the unbiasing coefficients $c_{n,2r} = c_{n,2}$ for $r = 1$ given by (10) are considerably robust against deviation from normality, as long as the true distribution is not markedly skewed. For $r = 2$, Fig. 4 evaluates the bias of the uncorrected root $M_{n,X,4}^{1/4}$ and the bias of the corrected estimator $R_{n,X,4}$ by the ratios $\mathrm{m}_{X,4}^{1/4}/E[M_{n,X,4}^{1/4}]$ and $\mathrm{m}_{X,4}^{1/4}/E[R_{n,X,4}]$ under the symmetric unimodal distributions listed in Table 2. The logistic, Laplace, and generalised hyperbolic distribution families are closed under affine transformations. For these distributions, the study of the ratios can be restricted to the parameters of standardised representatives with $E[X] = 0$, $V[X] = \mathrm{m}_{X,2} = 1$ such that the central fourth moment $\mathrm{m}_{X,4}$ equals the excess coefficient $\mathrm{m}_{X,4}/\mathrm{m}_{X,2}^2$. For the two-parameter logistic and Laplace families, the study of the standardised representatives is representative for the entire family. The standardised symmetric version of $GH(\lambda, \alpha, \beta, \delta, \mu)$ has the parameters $\mu = 0 = \beta$, but still depends on the parameters α and λ which affect the fourth moment. Figure 4 considers a standardised symmetric generalised hyperbolic distribution with a very large excess at $\mathrm{m}_{X,4} = \mathrm{m}_{X,4}/\mathrm{m}_{X,2}^2 = 14.96$.

The results displayed in Fig. 4 are based on simulation. The number of simulation runs was chosen so as to keep the length of the empirical 99 % confidence interval

Table 1 Coefficients $c_{n,2r}$ unbiasing the estimator (9) under the normal distribution

n	$r=2$	$r=3$	$r=4$	n	$r=2$	$r=3$	$r=4$
5	1.32435	1.45293	1.57817	53	1.02747	1.05116	1.08288
6	1.26098	1.37422	1.48164	54	1.02681	1.04959	1.08224
7	1.21822	1.31849	1.41691	55	1.02678	1.04924	1.08079
8	1.18909	1.27857	1.36926	56	1.02615	1.04789	1.07991
9	1.16460	1.24819	1.33204	57	1.02593	1.04763	1.07883
10	1.14873	1.22465	1.30444	58	1.02575	1.04684	1.07742
11	1.13517	1.20647	1.28093	59	1.02480	1.04630	1.07615
12	1.12271	1.18853	1.25966	60	1.02404	1.04576	1.07578
13	1.11255	1.17614	1.24398	61	1.02398	1.04505	1.07501
14	1.10502	1.16468	1.22954	62	1.02369	1.04448	1.07388
15	1.09700	1.15338	1.21789	63	1.02342	1.04397	1.07322
16	1.09112	1.14551	1.20667	64	1.02278	1.04308	1.07197
17	1.08610	1.13750	1.19693	65	1.02234	1.04250	1.07149
18	1.08147	1.13075	1.18844	66	1.02231	1.04194	1.07044
19	1.07647	1.12461	1.18099	67	1.02170	1.04119	1.06968
20	1.07323	1.11884	1.17478	68	1.02169	1.04039	1.06896
21	1.06985	1.11395	1.16749	69	1.02152	1.04019	1.06842
22	1.06594	1.10945	1.16130	70	1.02096	1.03985	1.06741
23	1.06340	1.10539	1.15534	71	1.02080	1.03902	1.06693
24	1.06057	1.10096	1.15069	72	1.02021	1.03871	1.06643
25	1.05841	1.09832	1.14622	73	1.02025	1.03826	1.06505
26	1.05571	1.09448	1.14196	74	1.02011	1.03797	1.06472
27	1.05389	1.09172	1.13778	75	1.01983	1.03748	1.06405
28	1.05176	1.08909	1.13377	76	1.01918	1.03720	1.06358
29	1.05022	1.08640	1.13061	77	1.01904	1.03672	1.06291
30	1.04863	1.08361	1.12704	78	1.01904	1.03574	1.06245
31	1.04666	1.08085	1.12346	79	1.01870	1.03595	1.06185
32	1.04604	1.07841	1.12183	80	1.01852	1.03536	1.06119
33	1.04440	1.07639	1.11827	81	1.01817	1.03503	1.06073
34	1.04304	1.07471	1.11553	82	1.01764	1.03458	1.05930
35	1.04143	1.07280	1.11282	83	1.01787	1.03401	1.05924
36	1.04065	1.07090	1.11041	84	1.01776	1.03407	1.05862
37	1.03967	1.06917	1.10830	85	1.01744	1.03328	1.05786
38	1.03875	1.06789	1.10721	86	1.01726	1.03332	1.05813
39	1.03764	1.06649	1.10407	87	1.01701	1.03292	1.05750
40	1.03671	1.06510	1.10271	88	1.01631	1.03230	1.05692
41	1.03572	1.06375	1.10032	89	1.01668	1.03272	1.05593
42	1.03454	1.06162	1.09814	90	1.01632	1.03191	1.05587
43	1.03416	1.06131	1.09695	91	1.01592	1.03167	1.05494
44	1.03336	1.06049	1.09573	92	1.01601	1.03143	1.05483
45	1.03291	1.05853	1.09407	93	1.01600	1.03121	1.05473

(continued)

Table 1 (continued)

n	$r = 2$	$r = 3$	$r = 4$	n	$r = 2$	$r = 3$	$r = 4$
46	1.03168	1.05746	1.09204	94	1.01567	1.03068	1.05370
47	1.03078	1.05626	1.09042	95	1.01556	1.03030	1.05366
48	1.03066	1.05548	1.08942	96	1.01554	1.03058	1.05321
49	1.02978	1.05444	1.08791	97	1.01505	1.03029	1.05306
50	1.02926	1.05303	1.08724	98	1.01505	1.02949	1.05211
51	1.02828	1.05274	1.08489	99	1.01480	1.02907	1.05231
52	1.02848	1.05142	1.08385	100	1.01478	1.02914	1.05173

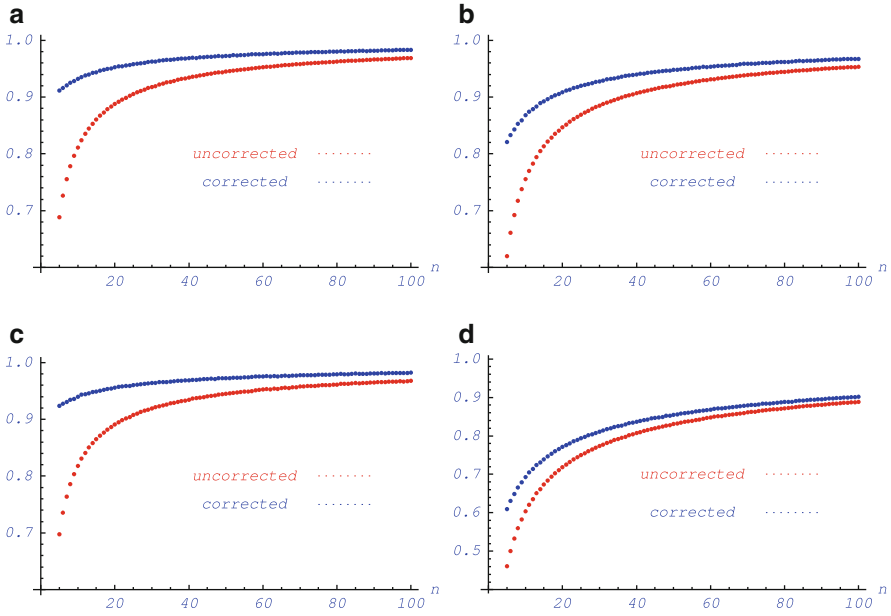


Fig. 4 Ratios $E[M_{n,X,4}^{1/4}]/\mathrm{m}_{X,4}^{1/4}$ and $E[R_{n,X,4}]/\mathrm{m}_{X,4}^{1/4}$ for the uncorrected root $M_{n,X,4}^{1/4}$ of the fourth sample moment and for the corrected estimator $R_{n,X,4} = c_{n,4} M_{n,X,4}^{1/4}$, see (9). (**a**) Logistic distribution. (**b**) Laplace distribution. (**c**) Central $t(\nu)$, $\nu = 10$. (**d**) Generalised hyperbolic

smaller than 2.5×10^{-3}. N varies from $N = 1.5 \times 10^6$ for small sample size n downto $N = 2.5 \times 10^5$ for large sample size $n > 50$. All considered distributions have larger probability mass of the tails than the normal distribution which has the excess $\mathrm{m}_{X,4}/\mathrm{m}_{X,2}^2 = 3.0$. In particular, the generalised hyperbolic was introduced by Barndorff-Nielsen and Blæsild (1981) as a *semi-heavy tails* distribution which places markedly larger probability mass on the tails than the normal distribution, but still has finite moments of every order. The generalised hyperbolic distributions has been used in modelling financial data, see Necula (2009) for a literature overview. In spite of the large excesses, the corrected estimator $R_{n,X,4}$ performs well

Table 2 Symmetric distributions considered in Fig. 4

Distribution	Density	Considered parameters
Logistic LGT(μ, β)	$\dfrac{\exp\left(\frac{-(x-\mu)}{\beta}\right)}{\beta\left[1+\exp\left(\frac{-(x-\mu)}{\beta}\right)\right]^2}$	$\mu = 0,\ \beta = \frac{\sqrt{3}}{\pi},\ \mathrm{m}_{X,4} = 4.2$
Laplace LAP(μ, β)	$\dfrac{1}{2\beta}\exp\left(\dfrac{-\lvert x-\mu\rvert}{\beta}\right)$	$\mu = 0,\ \beta = \frac{1}{\sqrt{2}},\ \mathrm{m}_{X,4} = 6$
Central t $t(\nu)$	$\dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\dfrac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\nu = 10,\ \mathrm{m}_{X,4} = 6.25,\ \frac{\mathrm{m}_{X,4}}{\mathrm{m}_{X,2}^2} = 4.0$
Generalised hyperbolic GH($\lambda, \alpha, \beta, \delta, \mu$)	$a(\lambda,\alpha,\beta,\delta,\mu)\left[\delta^2+(x-\mu)^2\right]^{2\lambda-1}$ $\times K_{\lambda-0.5}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)$ $\times \exp\left(\beta(x-\mu)\right)$	$\mu = 0 = \beta,\ \lambda = 0,\ \alpha = 0.6371795,$ $\delta = 0.1569416,\ \mathrm{m}_{X,4} = 14.96$

for all considered distributions, except for the considered generalised hyperbolic distribution with its very high excess. The performance is good even for small sample sizes, and in any case strikingly better than the uncorrected root $M_{n,X,4}^{1/4}$ of the fourth sample moment. Both statistics underestimate $\mathrm{m}_{X,4}^{1/4}$ on the average with $E[M_{n,X,4}^{1/4}] < E[R_{n,X,4}] < \mathrm{m}_{X,4}^{1/4}$. Particularly for small sample sizes, the bias of $M_{n,X,4}^{1/4}$ is excessive whereas $E[R_{n,X,4}]$ comes reasonably close to $\mathrm{m}_{X,4}^{1/4}$.

4 Empirical Camp-Meidell Quantile Bounds

Concluding from the results of the previous Sect. 3, the corrected central sample moment root $R_{n,X,2r}$ defined by (9) is a suitable estimator for the central moment root $\mathrm{m}_{X,2r}^{1/(2r)}$. Hence an empirical Camp-Meidell bound $\hat{z}_{\mathrm{C},r,X}(\rho)$ for the one-sided quantile $z_X(\rho)$ is obtained by replacing in the formula (7) for the theoretical bound $z_{\mathrm{C},r,X}(\rho)$ the occurrences of $\mathrm{m}_{X,2r}^{1/(2r)}$ by the estimator $R_{n,X,2r}$ and the mean μ_X by the sample mean so as to obtain

$$\hat{z}_{\mathrm{C},r,X}(\rho) = \begin{cases} \overline{X} + (2\rho-1)R_{n,X,2r}(2r+1)^{1/(2r)}, & \text{if } \rho \le \frac{4r+1}{4r+2}, \\ \overline{X} + \frac{R_{n,X,2r}}{[2(1-\rho)]^{\frac{1}{2r}}}\frac{2r}{2r+1}, & \text{if } \rho \ge \frac{4r+1}{4r+2}. \end{cases} \tag{11}$$

A suitable benchmark for the usefulness of the quantile bound $\hat{z}_{\mathrm{C},r,X}(\rho)$ is a nonparametric quantile estimator. An exhaustive comparison with the large variety

of nonparametric quantile estimators suggested in literature is far beyond the scope of the present study. We confine the comparison to the elementary sample quantile defined by as the order statistic $X_{(\lceil n\rho \rceil,n)}$ among the ordered sample $X_{(1,n)} \leq \ldots \leq X_{(n,n)}$. Various improvements of the sample quantile have been suggested in literature, see the discussion by Hyndman and Fan (1996), but the basic properties of all these variants are the same.

We compare the quantile bound $\hat{z}_{C,r,X}(\rho)$ and the sample quantile $X_{(\lceil n\rho \rceil,n)}$ in two respects: (i) the expectations $E[\hat{z}_{C,r,X}(\rho)]$, $E[X_{(\lceil n\rho \rceil,n)}]$, and (ii) the coverages $\mathrm{P}(\hat{z}_{C,r,X}(\rho) \geq z_X(\rho))$, $\mathrm{P}(X_{(\lceil n\rho \rceil,n)} \geq z_X(\rho))$ with respect to the true quantile $z_X(\rho)$. Figures 5 and 6 show graphs of the characteristic quantities as functions of the sample size n ranging from $n = 5$ to $n = 100$ for the quantile levels $\rho = 0.975$ and $\rho = 0.995$. We consider the same distributions as in Fig. 4, see Table 2. The results displayed in Figs. 5 and 6 were calculated by simulation based on the same technique as used for calculating the results provided in Fig. 4.

The estimator $\hat{z}_X(\rho)$ performs well for both quantile levels ρ and for all considered distributions. The bounding relation $\hat{z}_X(\rho) > z_X(\rho)$ holds on the average, even for very small sample sizes downto $n = 5$. The coverage $\mathrm{P}(\hat{z}_{C,r,X}(\rho) \geq z_X(\rho))$ exceeds 0.5 for small sample sizes, and rapidly increases to values of 0.8 to 0.9 for sample sizes $n \geq 50$.

Considered as a quantile estimator, the empirical bound $\hat{z}_{C,r,X}(\rho)$ has some appealing features in comparison with the sample quantile, in particular a high coverage with an only moderate overestimation for small sample sizes, and the absence of discontinuities in expectation and coverage as functions of the sample size, as visible for $\rho = 0.975$ in Fig. 5.

5 Shewhart Individual Observation Charts Based on Camp-Meidell Quantile Bounds

We consider the statistical design of a two-sided Shewhart individual observation control chart under i.i.d. mean-modal measurements (X_i), see the introduction of Sect. 2 for the definition of mean-modality. The individual measurement X_i signals an alarm if $X_i < \mathrm{LCL}$ or if $X_i > \mathrm{UCL}$. A bound α is prescribed for the false alarm probability, i.e., the probability of an alarm in the in-control state. To guarantee the bound, a parametric design based on a specific symmetric distribution uses the control limits

$$\mathrm{LCL} = z_X\left(\frac{\alpha}{2}\right), \quad \mathrm{UCL} = z_X\left(1-\frac{\alpha}{2}\right). \tag{12}$$

In a nonparametric design, the bound should hold for arbitrary symmetric mean-modal in-control distributions, without referring to a specific distribution. Concluding from the results of the preceding paragraphs, the use of Camp-Meidell quantile

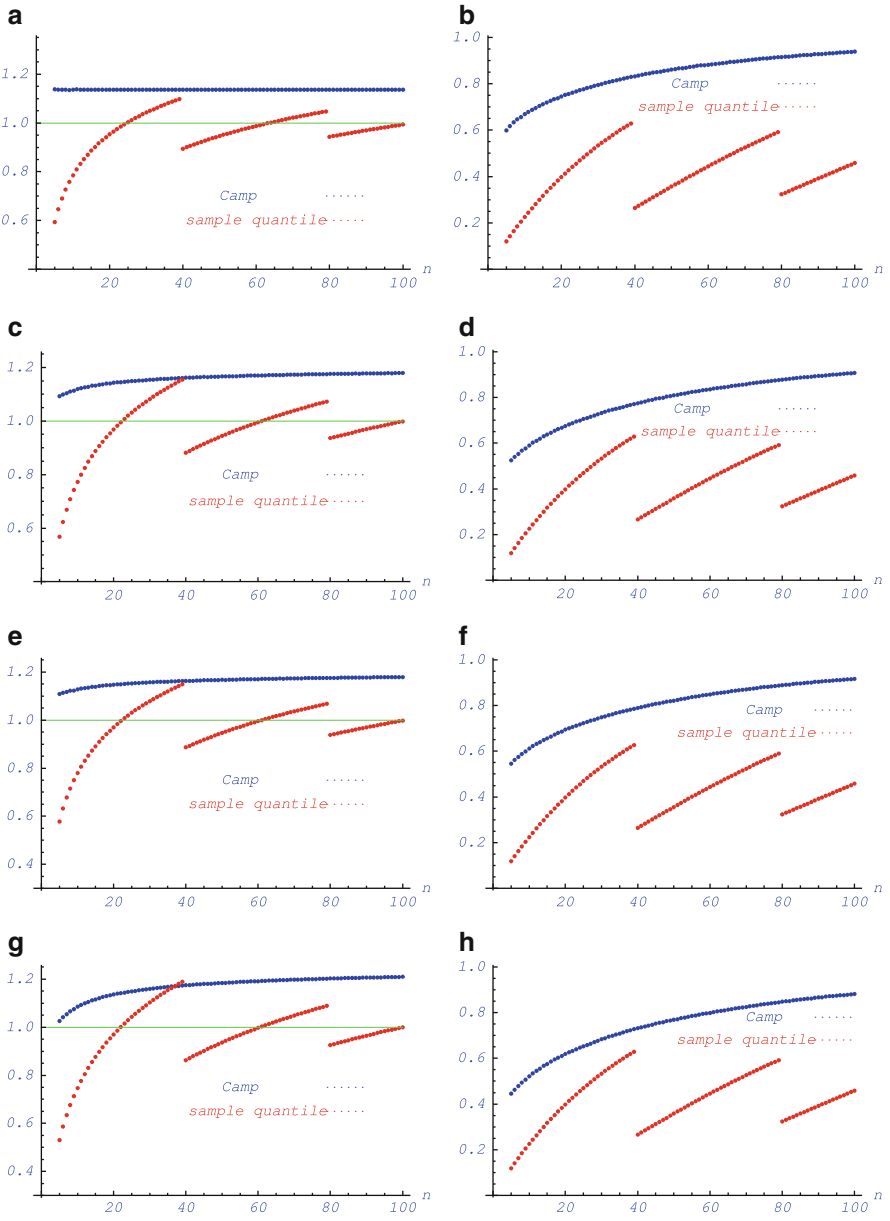


Fig. 5 Left-hand graphs: ratios $E[\hat{z}_{C,r,X}(\rho)]/z_X(\rho)$, $E[X_{(\lceil n\rho\rceil,n)}]/z_X(\rho)$ for $\rho = 0.975$; right-hand graphs: coverages $P(\hat{z}_{C,r,X}(\rho) \geq z_X(\rho))$, $P(X_{(\lceil n\rho\rceil,n)} \geq z_X(\rho))$ for $\rho = 0.975$. (**a**) Normal distribution. (**b**) Normal distribution. (**c**) Logistic distribution. (**d**) Logistic distribution. (**e**) Central t-distribution $t(10)$ (**f**) Central t-distribution $t(10)$. (**g**) Laplace distribution. (**h**) Laplace distribution

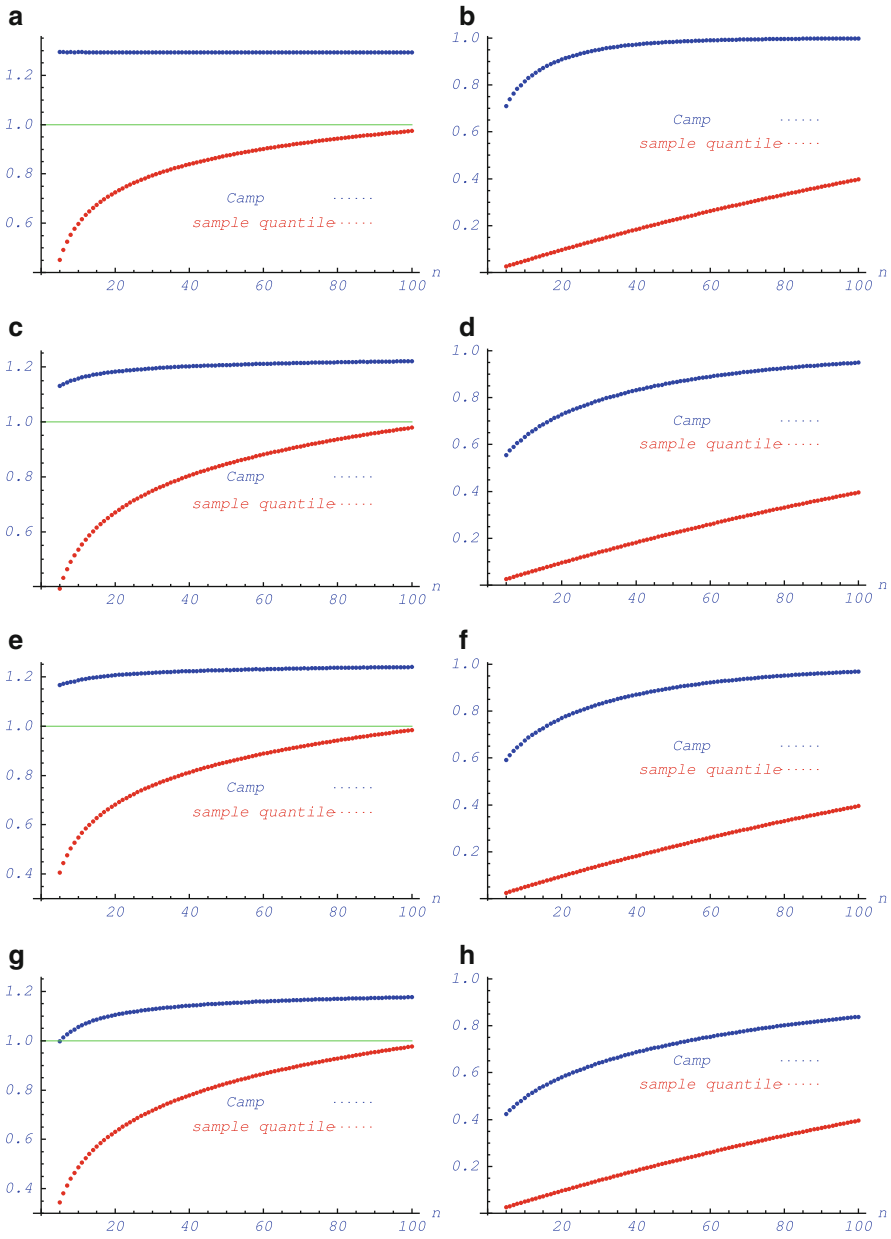


Fig. 6 Left-hand graphs: ratios $E[\hat{z}_{C,r,X}(\rho)]/z_X(\rho)$, $E[X_{(\lceil n\rho \rceil,n)}]/z_X(\rho)$ for $\rho = 0.995$; right-hand graphs: coverages $P(\hat{z}_{C,r,X}(\rho) \geq z_X(\rho))$, $P(X_{(\lceil n\rho \rceil,n)} \geq z_X(\rho))$ for $\rho = 0.995$. (**a**) Normal distribution. (**b**) Normal distribution. (**c**) Logistic distribution. (**d**) Logistic distribution. (**e**) Central t-distribution $t(10)$. (**f**) Central t-distribution $t(10)$. (**g**) Laplace distribution. (**h**) Laplace distribution

bounds $z_{C,r,X}(\rho)$ with $\rho = 1 - \alpha$ from Theorem 2 is a reasonable approach to the nonparametric design task. For the implementation we distinguish two cases.

Case 1: Known Camp-Meidell quantile bounds, i.e., known central moment $m_{X,2r}$. In practice, the "known" case means that the moments and the Camp-Meidell bound, respectively, can be estimated with good precision from a sufficient amount n of in-control data, e.g., from a phase 1 study. The results of Sect. 4 show that an in-control data set $X_1, \ldots, X_n$ of size $n \approx 100$ already provides a precise estimate of Camp-Meidell quantile bounds. One would then calculate the estimate $\hat{z}_X\left(1 - \frac{\alpha}{2}\right)$ from $X_1, \ldots, X_n$ and use

$$\text{UCL} = \hat{z}_X\left(1 - \frac{\alpha}{2}\right), \quad \text{LCL} = 2\overline{X} - \text{UCL} \tag{13}$$

as the control limits in the subsequent implementation of the chart.

Case 2: Unknown Camp-Meidell quantile bounds, i.e., unknown central moment $m_{X,2r}$. In this case, the quantile bound estimations have to be updated from successive observations $X_1, X_2, \ldots$. If X_{n+1} is the present observation, and past observations $X_1, \ldots, X_n$ have not generated an alarm and can be considered as in-control, the quantile estimator $\hat{z}_X\left(1 - \frac{\alpha}{2}\right)$ can be calculated from $X_1, \ldots, X_n$ and the control limits for the test on X_{n+1} can be set as in (13). The results of Sect. 4 show that even for a small number n of in-control rated observations the obtained bounds cover the true quantiles with a high probability, in any case with a much higher probability than achieved by the empirical quantile.

Relative to the classification established in Sect. 1, the Case 1 procedure is a variant of the approach (ii) which uses inference from training samples, whereas the Case 2 procedure is an instance of the approach (i) based on potentially small online samples from the process. The empirical Camp-Meidell quantile bounds can be successfully used for both approaches.

Under the above assumptions on (X_i), the in-control average run length (ARL) of the individual observations chart under known quantile $z_X(\rho)$ is $(1 - \rho)^{-1}$. If an estimate $\hat{z}(\rho)$ is used as an estimate for $z_X(\rho)$, the actual average run length for the in-control symmetric distribution function F_X is $[1 - F_X(\hat{z}_X(\rho))]^{-1}$. Since $(0; 1) \mapsto (1 - x)^{-1}$ is strictly increasing, we have

$$\text{P}\left(\frac{1}{1 - F_X(\hat{z}_X(\rho))} \geq \frac{1}{1 - \rho}\right) = \text{P}\Big(\hat{z}(\rho) \geq z_X(\rho)\Big).$$

Hence the properties of the estimated ARL $[1 - F_X(\hat{z}(\rho))]^{-1}$ under the estimates $\hat{z}_X(\rho) = z_{C,r,X}(\rho)$ based on the Camp-Meidell bound and the sample quantile $\hat{z}_X(\rho) = X_{(\lceil n\rho \rceil, n)}$ are illustrated by the right-hand columns in Figs. 5 and 6. For the considered in-control distributions, the Camp-Meidell approach provides with high probability an upper bound for the true ARL with a moderate overestimation even for small sample sizes. For industrial quality control, Fig. 6 corresponding to a theoretical in-control ARL of 200 is more relevant than Fig. 5 which corresponds to a theoretical in-control ARL of 40.

6 Conclusive Discussion and Outlook

Higher order Camp-Meidell inequalities provide bounds which approximate the true quantiles of levels $\rho \geq 0.95$ of the familiar symmetric mean-modal distributions up to a moderate relative error. We have developed an estimator of the bound which essentially preserves the bounding property. The estimator is even competitive as a quantile estimator, particularly for small sample sizes. It makes sense to use the empirical bound for the design of an individual observation chart. The particular advantage of the technique is the very modest amount of data required for a sufficiently precise estimation of the control limits.

The following issues remain to be studied. (i) We have studied the classical symmetric mean-modal distributions. Further distributions should be studied to corroborate the results, in particular, more distributions with large probability mass on the tails, and mean-inflated distributions. (ii) It is to be conjectured that slight deviations from symmetry or mean-modality will not substantially alter the pattern. This should be studied in detail. (iii) The design based on empirical Camp-Meidell bounds is conservative with respect to the false alarm rate. The overestimation of the theoretical quantile is moderate, but the effects on the rate of undetected shifts should be explored in more detail. (iv) The in-control and out-of-control run length of the chart should be studied in more detail.

References

Amin, R. W., Reynolds, M. R., & Saad, B. (1995). Nonparametric quality control charts based on the sign statistic. *Communications in Statistics – Theory and Methods, 24*(6), 1597–1623.

Areepong, Y., & Sukparungsee, S. (2010). An integral equation approach to EWMA chart for detecting a change in lognormal distribution. *Thailand Statistician, 8*(1), 47–61.

Bai, D. S., & Choi, I. S. (1995). $\bar{X}$ and R control charts for skewed populations. *Journal Of Quality Technology, 27*, 120–131.

Bakir, S. T. (2004). A distribution-free Shewhart quality control charts based on signed-ranks. *Quality Engineering, 16*(4), 613–623.

Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics. *Communications in Statistics – Theory and Methods, 35*(4), 743–757.

Bakir, S. T., & Reynolds, M. R. (1979). A nonparametric procedure for process control based on within-group ranking. *Technometrics, 21*(2), 175–183.

Barndorff-Nielsen, O.E., & Blæsild, P. (1981). Hyperbolic distributions and ramifications: Contributions to theory and application. In C. Taillie, G. Patil, & B. Baldessari (Eds.), *Statistical distributions in scientific work* (Vol. 4, pp. 19–44). Dordrecht: Reidel.

Betul, K., & Yaziki, B. (2006). The individual control charts for BURR distributed data. In *Proceedings of the Ninth WSEAS International Conference on Applied Mathematics* (pp. 645–649). Istanbul, Turkey.

Burr, I. W. (1967). The effects of non-normality on constants for $\bar{X}$ and R charts. *Industrial Quality Control, 24*, 563–569.

Camp, B. H. (1922). A new generalization of Tchebycheff's inequality. *Bulletin of the American Mathematical Society, 28*, 427–432.

Chakraborti, S., & Eryilmaz, S. (2007). A nonparametric Shewhart-type signed-rank control chart based on runs. *Communications in Statistics – Simulation and Computation, 36*(2), 335–356.
Chakraborti, S., van der Laan, P., & Bakir, S. T. (2001). Nonparametric control charts: An overview and some results. *Journal Of Quality Technology, 33*(3), 304–315.
Chan, L. K., & Cui, H. J. (2003). Skewness correction $\bar{X}$ and R charts for skewed distributions. *Naval Research Logistics, 50*, 1–19.
Chang, Y. S., & Bai, D. S. (2001). Control charts for positively skewed populations with weighted standard deviations. *Quality and Reliability Engineering International*, 17(5), 397–406.
Chatterjee, S., & Qiu, P. (2009) . Distribution-free cumulative sum control charts using bootstrap-based control limits. *The Annals of Applied Statistics, 3*(1), 349–369.
Chebyshev, B. L. (1867). Des valeurs moyennes. *Journal de Mathématique Pure et Appliquée, 12*, 177–184.
Chen, G. (1998). An improved p chart through simple adjustments. *Journal of Quality Technology, 30*, 142–151.
Cheng, S. W., & Xie, H. (2000). Control charts for lognormal data. *Tamkang Journal of Science and Engineering, 3*, 131–137.
Choobineh, F., & Ballard, J. L. (1987). Control-limits of QC charts for skewed distributions using weighted variance. *IEEE Transactions on Reliability, 36*, 473–477.
Cornish, E. A., & Fisher, R. A. (1938). Moments and cumulants in the specifications of distributions. *Revue de l'Institut Internationale de Statistique, 5*(4), 307–320.
Cramér, H. (1946). *Mathematical methods of statistics.* Princeton, NJ, Princeton University Press.
Derya, K., & Canan, H. (2012). Control charts for skewed distributions: Weibull, gamma, and lognormal. *Metodološki zvezki, 9*(2), 95–106.
Edgeman, R. L. (1989). Inverse Gaussian control charts. *Australian Journal of Statistics, 31*(1), 435–446.
Erto, P., & Pallotta, G. (2007) . A new control chart for Weibull technological processes. *Quality Technology and Quantitative Management, 4*(4), 553–567.
Erto, P., Pallotta, G., & Park, S. H. (2008) . An example of Data Technology product: A control chart for Weibull processes. *International Statistical Review, 76*(2), 157–166.
Ferrell, E. B. (1958). Control charts for lognormal universe. *Industrial Quality Control, 15*, 4–6.
Fisher, R. A. (1928). Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society, 30*, 199–238.
Florac, W. A., & Carleton, A. D. (1999). *Measuring the software process: Statistical process control for software process improvement.* Boston: Addison-Wesley.
Göb, R., & Lurz, K. (2013a). Estimating the fourth central moment for unimodal distributions. *Preprints of the Institute of Mathematics*, 334. Würzburg: University of Würzburg.
Göb R, Lurz K (2013b). Empirical quantile bounds based on the Camp-Meidell inequality. *Preprints of the Institute of Mathematics*, 335. Würzburg: University of Würzburg.
González, I., & Viles, E. (2000). Semi-economic design of X-control charts assuming Gamma distribution. *Economic Quality Control, 15*, 109–118.
González, I., & Viles, E. (2001). Design of R control chart assuming a Gamma distribution. *Economic Quality Control, 16*(2), 199–204.
Grant, E. L., & Leavenworth, R. S. (1999). *Statistical Quality Control* 7th Ed. Boston: McGraw-Hill.
Guo, B. C., & Wang, B. X. (2012). Control charts For monitoring the Weibull Shape parameter based on type-II censored sample. *Quality and Reliability Engineering International, 30*, 13–24. http://dx.doi.org/10.1002/qre.1473
Haas, M., & Pigorsch, C. (2009). Financial economics, fat-tailed distributions. In R. A. Meyers (Ed.) *Encyclopedia of Complexity and Systems Science* (pp. 3404–3435). New York: Springer.
Hackl, P., & Ledolter, J. (1991). A control chart based on ranks. *Journal of Quality Technology, 23*(2), 117–124.
Hackl, P., & Ledolter, J. (1992). A new nonparametric quality control technique. *Communications in Statistics – Simulation and Computation, 21*(2), 423–443.

Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician, 50*, 361–365.

International Organisation for Standardisation (ISO) (2008). *SO/IEC Guide 98–3:2008: Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement.* ISO: Geneva, Switzerland.

Ion, R. A. (2001). *Nonparametric Statistical Process Control.* Korteweg-de Vries Instituut voor Wiskunde, Faculteit der Natuurwetenschappen, Wiskunde en Informatica. Amsterdam: Universiteit van Amsterdam.

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician, 47*(1), 183–189.

Joffe, A. D., & Sichel, H. S. (1968). A chart for sequentially testing observed arithmetic means from lognormal populations against a given standard. *Technometrics, 10*, 605–612.

Johnson, N. L. (1966). Cumulative sum control charts and the Weibull distribution. *Technometrics, 8*(3), 481–491.

Jones, L. A., & Woodall, W. H. (1998). The performance of bootstrap control charts. *Journal of Quality Technology, 30,* 362–375.

Kanji, G. K., & Arif, O. H. (2001). Median rankit control chart for Weibull distribution. *Total Quality Management, 12(5)*, 629–642.

Kantam, R. R. L., & Sriram, B. (2001). Variable control charts based on Gamma distribution. *IAPQR Transactions, 26*(2), 63–67.

Kantam, R. R. L., Vasudeva, R. A., & Srinivasa, R. G. (2006). Control charts for log-logistic distribution. *Economic Quality Control, 21*(1), 77–86.

Kotz, S., & Lovelace, C. R. (1988). *Process capability indices in theory and practice.* London: Edward Arnold.

Leger, C., Politis, D., & Romano, J. (1992). Bootstrap technology applications. *Technometrics, 34*, 378–398.

Liu, R. Y., & Tang, J. (1996). Control charts for dependent and independent measurements based on bootstrap methods. *Journal of the American Statistical Association, 91*, 1694–1700.

Meidell, M. B. (1922). Sur un problème du calcul des probabilités et les statistiques mathématiques. *Comptes Rendus, 175*, 806–808.

Morrison, J. (1958). The lognormal distribution in quality control. *Applied Statistics, 7*, 160–172.

Necula, C. (2009). Modelling heavy-tailed stock index returns using the generalised hyperbolic distribution. *Romanian Journal of Economic Forecasting, 2*, 118–131.

Nelson, P. R. (1979). Control charts for Weibull processes with standards given. *IEEE Transactions on Reliability, 28*, 383–387.

Park, C., & Reynolds, M. R. (1987). Nonparametric procedures for monitoring a location parameter based on linear placement statistics. *Sequential Analysis, 6*(4), 303–323.

Patel, J. K., Kapadia, C. H., & Owen, D. B. (1976). *Handbook of statistical distributions.* New York/Basel: Marcel Dekker Inc.

Pyzdek, I. (1995). Why normal distributions aren't [all that normal]. *Quality Engineering, 7*(4), 769–777.

Qiu, P., & Li, Z. (2011). On nonparametric statistical process control of univariate processes. *Technometrics, 53*(4), 390–405.

Ramalhoto, M. F., & Morais, M. (1998). EWMA control charts for the scale parameter of a Weibull control variable with fixed and variable sampling intervals. *Economic Quality Control, 13*, 23–46.

Ramalhoto, M. F., & Morais, M. (1999). Shewhart control charts for the scale parameter of a Weibull control variable with fixed and variable sampling intervals. *Journal of Applied Statistics, 26(1)*, 129–160.

Samanta, B., & Bhattacherjee, A. (2004). Problem of nonnormality in statistical quality control. *Journal of the South African Institute of Mining and Metallurgy, 104*, 257–264.

Schilling, E. G., & Nelson, P. R. (1976). The effect of nonnormality on the control limits of charts. *Journal of Quality Technology, 8*(4), 179–193.

Seppala, T., Moskowitz, H., Plante, R., & Tang, J. (1995). Statistical process control via the subgroup bootstrap. *Journal of Quality Technology, 27*, 139–153.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. London: Macmillan.

Spedding, T. A., & Rawlings, P. L. (1994). Nonnormality in statistical process control measurements. *International Journal of Quality & Reliability Management, 11(6)*, 27–37.

Subba, R. R., & Kantam, R. R. L. (2008). Variable control charts for process mean with double exponential distribution. *Acta Cinica Indica, 34(4)*, 1925–1930.

Teyarachakul, S., Chand, S., & Tang, J. (2007). Estimating the limits for statistical process control charts: A direct method improving upon the bootstrap. *European Journal of Operational Research, 178*, 472–481.

Willemain, T. R., & Runger, G. C. (1996). Designing control charts using an empirical reference distribution. *Journal of Quality Technology, 28*(1), 31–38.

Winterbottom, A. (1993). Simple adjustments to improve control limits on attribute charts. *Quality and Reliability Engineering International, 9*, 105–109.

Woodall, W. H. (2000) . Controversies and contradictions in statistical process control. *Journal of Quality Technology, 32*(4), 341–350.

Wu, C., & Wang, Z. (2009). A synthetic control chart based on the Cornish-Fisher expansion. *Chinese Journal of Applied Probability and Statistics, 25*(3), 258–265.

Wu, Z. (1996). Asymmetric control limits of the X-bar chart for skewed process distributions. *International Journal of Quality & Reliability Management, 13*(9), 49–60.

Yang, Z., & Xie, M. (2000). Process monitoring of exponentially distributed characteristics through an optimal normalizing transformation. *Journal of Applied Statistics, 27*, 1051–1063.

Yourstone, S. A., & Zimmer, W. J. (1992). Nonnormality and the design of control charts for averages. *Decision Sciences, 23*, 1099–1113.

Zhang, L., & Chen, G. (2004). EWMA charts for monitoring the mean of censored Weibull lifetimes. *Journal of Quality Technology, 36*, 321–328.

Strategies to Reduce the Probability of a Misleading Signal

Manuel Cabral Morais, Patrícia Ferreira Ramos, and António Pacheco

Abstract Standard practice in statistical process control (SPC) is to run two individual charts, one for the process mean and another one for the process variance. The resulting scheme is known as a simultaneous scheme and it provides a way to satisfy Shewhart's dictum that proper process control implies monitoring both location and dispersion.

When we use a simultaneous scheme, the quality characteristic is deemed to be out-of-control whenever a signal is triggered by either individual chart. As a consequence, the misidentification of the parameter that has changed can occur, meaning that a shift in the process mean can be misinterpreted as a shift in the process variance and vice versa. These two events are known as misleading signals (MS) and can occur quite frequently.

We discuss (necessary and) sufficient conditions to achieve values of probabilities of misleading signals (PMS) smaller than or equal to 0.5, explore, for instance, alternative simultaneous Shewhart-type schemes and check if they lead to PMS which are smaller than the ones of the popular $(\bar{X}, S^2)$ simultaneous scheme.

Keywords Misleading signals • Simultaneous schemes • Statistical process control

1 Control Charts and the Phenomenon of Misleading Signals

Concerns about quality can be traced back to the Babylonian Empire (1830BC–539BC). Browsing the Code of Hammurabi (Code of Hammurabi/Wikipedia 2012)—a well-preserved Babylonian law code, dating back to about 1772BC, and aptly named after the sixth Babylonian king, Hammurabi, who enacted

M.C. Morais (✉) • A. Pacheco
CEMAT & Department of Mathematics, Instituto Superior Técnico, Av. Rovisco Pais, Lisbon, Portugal
e-mail: maj@math.ist.utl.pt; apacheco@math.ist.utl.pt

P.F. Ramos
CEMAT, Instituto Superior Técnico, Av. Rovisco Pais, Lisbon, Portugal
e-mail: patriciaferreira@ist.utl.pt

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_12

it—can be an "eye-opening" experience. This code consists of 282 laws dealing with matters of contracts, terms of transactions or addressing household and family relationships such as inheritance, divorce, paternity and sexual behavior (Code of Hammurabi/Wikipedia 2012). For instance, its laws 229 through 233 illustrate "an eye for an eye" approach to quality and certainly prove that quality and warranty were matters of importance to Babylonians.

Although quality has long been considered utterly relevant, we have to leap to the twentieth century to meet the father of modern quality control and the founder of statistical process control (SPC), Walter A. Shewhart (1891–1967). This physicist, engineer and statistician for Bell Laboratories acknowledged that any production process, no matter how well designed, always presents a certain amount of variability (Wieringa 1999, p. 2) and distinguished between acceptable and undesirable variability, arising from what is usually termed as common causes and assignable causes, respectively.

Shewhart proposed, in 1924, the quality control chart, a graphical device to detect the presence of assignable causes of variation in the process by plotting the observed value of a (control) statistic against time and comparing it with a pair of control limits. Points lying outside the control limits indicate potential assignable causes that should be investigated and eliminated (Nelson 1982, p. 178).

Shewhart suggested the use of the sample mean and a pair of control limits, LCL_μ and UCL_μ, thus, defining a $\bar{X}$-chart to monitor the process mean of a continuous quality characteristic X. If the process mean (resp. standard deviation) shifts from its target value μ_0 (resp. σ_0) to $\mu = \mu_0 + \delta\sigma_0/\sqrt{n}$, where $\delta \neq 0$ (resp. $\sigma = \theta\sigma_0$, where $\theta > 1$), then the number of samples taken until a signal is triggered by the $\bar{X}$-chart, its run length (RL), say $\mathrm{RL}_\mu(\delta, \theta)$, has a geometric distribution with parameter $\xi_\mu(\delta, \theta) = P_{\delta,\theta}(\bar{X} \notin [\mathrm{LCL}_\mu, \mathrm{UCL}_\mu])$ and the average run length (ARL) equals $\mathrm{ARL}_\mu(\delta, \theta) = 1/\xi_\mu(\delta, \theta)$.

To satisfy *Shewhart's famous dictum* (Hawkins and Maboudou-Tchao 2008) that proper process control implies monitoring both location and dispersion, standard (and popular) practice is to simultaneously run one individual chart for μ and another one for σ^2. If we are only concerned about increases in the dispersion of a process, then we should make use of the control statistic $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ along with an upper control limit UCL_σ. The RL of the S^2-chart, say $\mathrm{RL}_\sigma(\theta)$, has a geometric distribution with parameter $\xi_\sigma(\theta) = P_\theta(S^2 \notin [0, \mathrm{UCL}_\sigma])$ and $\mathrm{ARL}_\sigma(\theta) = 1/\xi_\sigma(\theta)$.

Since it is not realistic to believe that only one of the two parameters is subject to shifts, a simultaneous scheme triggers a signal whenever one (or both) of the individual charts triggers a signal, suggesting a potential shift in μ, in σ^2, or in both μ and σ^2. In other words, the RL of a simultaneous scheme, say $\mathrm{RL}_{\mu,\sigma}(\delta, \theta)$, is the minimum of the RL of the two individual control charts for μ and σ^2: $\mathrm{RL}_{\mu,\sigma}(\delta, \theta) = \min\{\mathrm{RL}_\mu(\delta, \theta), \mathrm{RL}_\sigma(\theta)\}$. Moreover, for Gaussian output, $\mathrm{RL}_{\mu,\sigma}(\delta, \theta)$ has a geometric distribution with parameter $\xi_{\mu,\sigma}(\delta, \theta) = \xi_\mu(\delta, \theta) + \xi_\sigma(\theta) - \xi_\mu(\delta, \theta) \times \xi_\sigma(\theta)$.

Additionally, it is possible that:

- a signal is triggered by the chart for μ before the chart for σ^2 signals, even though μ is on-target and σ^2 is off-target;
- μ has changed and σ^2 is on-target, but the chart for σ^2 signals before the chart for μ.

These are instances of what St. John and Bragg (1991) called *misleading signals* (MS), namely MS of types III and IV, respectively.

Since the assignable causes on charts for μ can differ from those on charts for σ^2, the diagnostic procedures that follow a signal can differ depending on whether the signal is given by the chart for μ or the one for σ^2 (Morais 2002, p. 109, Knoth et al. 2009). Thus, misleading signals are valid signals that may well have different consequences: they can namely lead the quality control operator or engineer to misdiagnose assignable causes and deploy incorrect actions to bring the process back to target.

2 Looking Closely at the Probability of a Misleading Signal

The main question is not whether there will be MS but rather how frequent they are (Morais 2002, p. 112). Therefore, the impact of MS in the performance of the simultaneous scheme should be assessed by calculating the probabilities of misleading signal (PMS).

According to the definition of MS of types III and IV, the corresponding PMS can be written as:

$$\mathrm{PMS}_{III}(\theta) = P[\mathrm{RL}_{\mu}(0,\theta) < \mathrm{RL}_{\sigma}(\theta)], \quad \theta > 1; \tag{1}$$

$$\mathrm{PMS}_{IV}(\delta) = P[\mathrm{RL}_{\sigma}(1) < \mathrm{RL}_{\mu}(\delta,1)], \quad \delta \neq 0. \tag{2}$$

Simple expressions of the PMS can be derived if the individual charts of the simultaneous scheme are Shewhart-type charts based on independent control statistics (Morais 2002, p. 118), such as $\bar{X}$ and S^2. Moreover, as noted by Knoth et al. (2009):

$$\mathrm{PMS}_{III}(\theta) = \frac{\xi_{\mu}(0,\theta) \times [1 - \xi_{\sigma}(\theta)]}{\xi_{\mu}(0,\theta) + \xi_{\sigma}(\theta) - \xi_{\mu}(0,\theta) \times \xi_{\sigma}(\theta)}, \quad \theta > 1; \tag{3}$$

$$\mathrm{PMS}_{IV}(\delta) = \frac{[1 - \xi_{\mu}(\delta,1)] \times \xi_{\sigma}(1)}{\xi_{\mu}(\delta,1) + \xi_{\sigma}(1) - \xi_{\mu}(\delta,1) \times \xi_{\sigma}(1)}, \quad \delta \neq 0. \tag{4}$$

Furthermore, the PMS of Type III (resp. IV) can be obviously interpreted as a conditional probability—it corresponds to the probability that the chart for μ (resp. σ^2) triggers a signal and the chart for σ^2 (resp. μ) fails to do so, given that the simultaneous scheme was responsible for an alarm.

Awareness of the phenomenon of MS stretches back to the seminal work of St. John and Bragg (1991), and the PMS, proposed by Morais and Pacheco (2000), has been already addressed for:

- i.i.d. Gaussian output by Morais and Pacheco (2000); Morais (2002); Morais and Pacheco (2006);
- autocorrelated Gaussian output by Antunes (2009); Knoth et al. (2009); Ramos et al. (2012, 2013);
- i.i.d. bivariate and multivariate normal output by Ramos et al. (2013a,b,c).

These authors have presented some striking and instructive examples that show that:

- the occurrence of MS should be a cause of concern in practice;
- Shewhart-type simultaneous schemes compare unfavourably to the ones with more sophisticated schemes, namely of the exponentially weighted moving average (EWMA) type;
- simultaneous schemes that ignore autocorrelation are far from being reliable in identifying which parameter has changed;
- the fact that the control statistics of the individual charts for the mean vector ($\underline{\mu}$) are based on quadratic forms—and therefore tend to confound shifts in the mean vector with shifts in the covariance matrix ($\underline{\Sigma}$)—aggravates the prevalence of MS.

More important, in certain instances, specifically when we deal with simultaneous schemes that falsely assume that the output is i.i.d. or with simultaneous schemes for $\underline{\mu}$ and $\underline{\Sigma}$, PMS can take values which exceed 50 %.

The next proposition provides necessary and sufficient conditions—written in terms of the ARL of the individual charts of the Shewhart-type simultaneous scheme—to obtain PMS which do not exceed 50 %.

Proposition 1 *If the individual charts of the simultaneous scheme for μ and σ^2 are of the Shewhart-type and are based on independent control statistics, then:*

- $\text{PMS}_{III}(\theta) \leq 0.5$ *iff* $\text{ARL}_{\mu}(0,\theta) + 1 \geq \text{ARL}_{\sigma}(\theta)$*;*
- $\text{PMS}_{IV}(\delta) \leq 0.5$ *iff* $\text{ARL}_{\sigma}(1) + 1 \geq \text{ARL}_{\mu}(\delta, 1)$.

Proof Capitalizing on formulas (3) and (4), and on the fact that the ARL function of a Shewhart-type chart is the reciprocal of the probability of a signal and that ARL is larger than or equal to the unit, we can provide necessary and sufficient condition to yield PMS of Type III not larger than 50 %:

$$\begin{aligned}
\text{PMS}_{III}(\theta) \leq 0.5 &\Leftrightarrow \frac{\xi_{\mu}(0,\theta)[1-\xi_{\sigma}(\theta)]}{\xi_{\mu}(0,\theta)+\xi_{\sigma}(\theta)-\xi_{\mu}(0,\theta)\xi_{\sigma}(\theta)} \leq 0.5 \\
&\Leftrightarrow \xi_{\mu}(0,\theta)-\xi_{\mu}(0,\theta)\xi_{\sigma}(\theta) \leq \xi_{\sigma}(\theta) \\
&\Leftrightarrow \frac{1}{\xi_{\sigma}(\theta)} - 1 \leq \frac{1}{\xi_{\mu}(0,\theta)} \\
&\Leftrightarrow ARL_{\mu}(0,\theta) + 1 \geq ARL_{\sigma}(\theta). \qquad (5)
\end{aligned}$$

The equivalence

$$\text{PMS}_{IV}(\delta) \leq 0.5 \Leftrightarrow \text{ARL}_{\sigma}(1) + 1 \geq \text{ARL}_{\mu}(\delta, 1) \tag{6}$$

is proven in a similar fashion. □

A sufficient condition for $\text{PMS}_{III}(\theta) \leq 0.5$ (resp. $\text{PMS}_{IV}(\delta) \leq 0.5$) is $\text{ARL}_{\mu}(0, \theta) \geq \text{ARL}_{\sigma}(\theta)$ (resp. $\text{ARL}_{\sigma}(1) \geq \text{ARL}_{\mu}(\delta, 1)$) or, equivalently,

$$\xi_{\mu}(0, \theta) \leq \xi_{\sigma}(\theta) \quad (\text{resp. } \xi_{\sigma}(1) \leq \xi_{\mu}(\delta, 1)). \tag{7}$$

(7) sounds rather reasonable: If the chart for μ (resp. σ^2) triggers valid signals less or as frequently as the chart for σ^2 (resp. μ), when $\delta = 0$ and $\theta > 1$ (resp. $\delta \neq 0$ and $\theta = 1$), then the PMS of type III (resp. IV) does not exceed 50 %.

Even though Proposition 1 was stated considering that the location and dispersion are univariate parameters (resp. the RL of the chart for σ^2 does not depend on the shift in μ), results (5) and (6) are also valid while dealing with simultaneous Shewhart-type schemes for the control of the mean vector and covariance matrix of multivariate normal output (resp. simultaneous Shewhart-type residual schemes for the mean and variance of autocorrelated output, where the RL of the chart for σ^2 depends on δ).

As far as the PMS of simultaneous EWMA schemes are concerned, it is not easy to extend Proposition 1 due to the Markovian character of the control statistics of the individual charts and consequently the phase-type distributions of the RL of the individual charts. Nevertheless, we can add that the approximate results for the PMS, obtained by Ramos et al. (2013c) for simultaneous EWMA-type schemes for $\underline{\mu}$ and $\underline{\Sigma}$, led to very few values of the PMS of Type III larger than 0.5 even though condition (5) was valid. Unsurprisingly, these values refer to the PMS of Type III when the unitary variances increase and the correlation coefficients shift from 0 to a non-null value, a scenario that corresponds to a radical change in the joint behaviour of the quality characteristics and obviously results in a more frequent misinterpretation of the shift in $\underline{\Sigma}$ as a shift in $\underline{\mu}$ by the chart for the mean vector.

3 A Few Strategies to Reduce the Probability of a Misleading Signal

Simultaneous schemes for μ and σ^2 are usually set in such way that the ARL of their individual control charts are matched in-control—i.e., $\text{ARL}_{\mu}(0, 1) = \text{ARL}_{\sigma}(1)$.

Interestingly enough, matching the charts for μ and σ^2 (or $\underline{\mu}$ and $\underline{\Sigma}$), in the absence of assignable causes, seems to lead to values of the PMS of Type IV which do not exceed 50 %. In fact, if ARL_{σ} does not depend on δ and $\text{ARL}_{\mu}(\delta, 1)$ decreases with δ, such as in the case of simultaneous Shewhart schemes for the mean and variance of i.i.d. output, we get $\text{ARL}_{\sigma}(1) \geq \text{ARL}_{\mu}(\delta, 1)$, then $\text{PMS}_{IV}(\delta) \leq 0.5$.

Regretfully, when we deal with multivariate normal output, the $\mathrm{ARL}_{\underline{\mu}}$ is often much more sensitive to changes in the covariance matrix than $\mathrm{ARL}_{\underline{\Sigma}}$, leading to a violation of (5) and, thus, to PMS of Type III larger than 50 %, even if the individual charts were matched in-control, as illustrated by Ramos et al. (2013a,b,c).

The choice of control statistics also has an impact on the values of PMS. Having said this, we ought to refer that replacing the $\bar{X}$ and S^2 charts with their EWMA counterparts proves to reduce the PMS of types III and IV considerably, as extensively reported by Morais (2002) and many other authors in different settings (i.i.d. Gaussian output, autocorrelated Gaussian output and i.i.d. bivariate and multivariate normal output).

Since there are quality control practitioners who are still reluctant to use EWMA charts, we explore now the use of alternative control statistics for the location and the spread of i.i.d. univariate Gaussian output to decrease the prevalence of MS. In Walsh (1952), we can find two other possibilities of control statistics for the simultaneous Shewhart-type schemes for μ and σ^2:

- $\sqrt{n}(\bar{X}-\mu_0)/S$ and S^2;
- $\bar{X}$ and $\sum_{i=1}^{n}(X_i-\mu_0)^2/n$.

Walsh (1952) justifies the use of the control statistics $\sqrt{n}(\bar{X}-\mu_0)/S$ and $\sum_{i=1}^{n}(X_i-\mu_0)^2/n$ quite carefully. Although $\bar{X}$ is an estimator of μ, the probability that $\bar{X}$ lies beyond the control limits depends not only on how μ compares with its target value μ_0 but also on the how σ^2 compares with σ_0^2, hence a value of $\bar{X}$ outside the control limits cannot necessarily be attributed to the discrepancy between μ and μ_0; unlike $\bar{X}$, the $t-$statistic $\sqrt{n}(\bar{X}-\mu_0)/S$ provides a rigorous measure of how μ deviates from its target value. Walsh (1952) continues: When $\mu=\mu_0$, $\sum_{i=1}^{n}(X_i-\mu_0)^2/n$ is an estimator of σ^2 which is more efficient than S^2; thus, using $\bar{X}$ and $\sum_{i=1}^{n}(X_i-\mu_0)^2/n$ might be preferable to $\bar{X}$ and S^2 or $\sqrt{n}(\bar{X}-\mu_0)/S$ and S^2.

It is important to notice that the two pairs of alternative control statistics suggested by Walsh (1952) are not independent; consequently, the corresponding PMS have different expressions than the ones in Eqs. (3) and (4). However, by capitalizing on expressions (1) and (2) and on Walsh's mathematical derivations of the associated operating characteristic (OC) functions, we were able to obtain expressions for the PMS of types III and IV, as shown in Proposition 2.

Notation wise, let

$$(T^{(i)},U^{(i)})=\begin{cases}\left(\bar{X},\,S^2\right), & i=1\\ \left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}},\,S^2\right), & i=2\\ \left(\bar{X},\,\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2\right), & i=3\end{cases} \tag{8}$$

be the three pairs of control statistics used to monitor both the mean and variance of normal i.i.d. output.

All the charts for μ depend on a lower and an upper control limit, say $\text{LCL}_{\mu}^{(i)}$ and $\text{UCL}_{\mu}^{(i)}$, $i = 1, 2, 3$. However, unlike Walsh (1952), our aim is solely the detection of increases in σ^2, thus, all charts for σ^2 only rely on their upper control limit, $\text{UCL}_{\sigma}^{(i)}$, $i = 1, 2, 3$. Having in mind the distributions of $\bar{X}$, $\sqrt{n}(\bar{X} - \mu_0)/S$, S^2 and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$ in the absence of assignable causes, we shall deal with

$$\text{LCL}_{\mu}^{(i)} = \begin{cases} \mu_0 - \gamma_{\mu}^{(i)} \times \frac{\sigma_0}{\sqrt{n}}, & i = 1, 3 \\ -\gamma_{\mu}^{(2)}, & i = 2 \end{cases} \tag{9}$$

$$\text{UCL}_{\mu}^{(i)} = \begin{cases} \mu_0 + \gamma_{\mu}^{(i)} \times \frac{\sigma_0}{\sqrt{n}}, & i = 1, 3 \\ \gamma_{\mu}^{(2)}, & i = 2 \end{cases} \tag{10}$$

$$\text{UCL}_{\sigma}^{(i)} = \begin{cases} \gamma_{\sigma}^{(i)} \times \frac{\sigma_0^2}{n-1}, & i = 1, 2 \\ \gamma_{\sigma}^{(3)} \times \frac{\sigma_0^2}{n}, & i = 3, \end{cases} \tag{11}$$

where:

- $\gamma_{\mu}^{(i)} = \Phi^{-1}(1 - \beta^{(i)})$, $i = 1, 3$; $\gamma_{\mu}^{(2)} = F_{t_{(n-1)}}^{-1}(1 - \beta^{(2)})$;
- $\gamma_{\sigma}^{(i)} = F_{\chi^2_{(n-1)}}^{-1}(1 - 2\beta^{(i)})$, $i = 1, 2$; $\gamma_{\sigma}^{(3)} = F_{\chi^2_{(n)}}^{-1}(1 - 2\beta^{(3)})$.

Let $\text{RL}_{\mu,\sigma}^{(i)}(\delta, \theta)$ be the RL of the simultaneous scheme that makes use of the control statistics $(T^{(i)}, U^{(i)})$. Since these RL are geometrically distributed with parameter equal to

$$\xi_{\mu,\sigma}^{(i)}(\delta, \theta) = 1 - P_{\delta,\theta}\left(T^{(i)} \in [\text{LCL}_{\mu}^{(i)}, \text{UCL}_{\mu}^{(i)}],\ U^{(i)} \in [0, \text{UCL}_{\sigma}^{(i)}]\right) \tag{12}$$

and expected value $\text{ARL}_{\mu,\sigma}^{(i)}(\delta, \theta) = 1/\xi_{\mu,\sigma}^{(i)}(\delta, \theta)$, the values of $\beta^{(i)}$—corresponding to a given false alarm rate $\xi_{\mu,\sigma}^{(i)}(0, 1) = \alpha$—are found by equating

$$\text{ARL}_{\mu,\sigma}^{(i)}(0, 1) = \alpha^{-1}. \tag{13}$$

Suffice to say that: the individual charts are matched in-control and $\text{ARL}_{\mu}^{(i)}(0, 1) = \text{ARL}_{\sigma}^{(i)}(0, 1) = [2\beta^{(i)}]^{-1}$; $\beta^{(1)} = \frac{1-\sqrt{1-\alpha}}{2}$; $\beta^{(2)}$ and $\beta^{(3)}$ are determined by numerical search.

Proposition 2 *The MS of types III and IV occur in the simultaneous Shewhart-type schemes, with constituent charts for μ and σ^2 with control statistics $T^{(i)}$ and $U^{(i)}$, with probabilities:*

$$\text{PMS}_{III}^{(i)}(\theta) = \frac{P_{0,\theta}\left(T^{(i)} \notin [\text{LCL}_{\mu}^{(i)}, \text{UCL}_{\mu}^{(i)}],\ U^{(i)} \in [0, \text{UCL}_{\sigma}^{(i)}]\right)}{\xi_{\mu,\sigma}^{(i)}(0, \theta)},\ \theta > 1; \tag{14}$$

$$\mathrm{PMS}_{IV}^{(i)}(\delta) = \frac{P_{\delta,1}\left(T^{(i)} \in [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U^{(i)} \notin [0, \mathrm{UCL}_{\sigma}^{(i)}]\right)}{\xi_{\mu,\sigma}^{(i)}(\delta, 1)},\ \delta \neq 0. \quad (15)$$

The full expressions of the numerators and denominators can be found in the appendix.

Proof Taking advantage of the fact that the $\{(T_j^{(i)}, U_j^{(i)}) : j \in IN\}$ are i.i.d. bivariate r.v., we successively get:

$$\begin{aligned}
\mathrm{PMS}_{III}^{(i)}(\theta) &= \sum_{m=1}^{\infty} P\left(\mathrm{RL}_{\mu}^{(i)}(0,\theta) = m,\ \mathrm{RL}_{\sigma}^{(i)}(\theta) > m\right) \\
&= \sum_{m=1}^{\infty} P_{0,\theta}\left(T_j^{(i)} \in [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U_j^{(i)} \in [0, \mathrm{UCL}_{\sigma}^{(i)}],\ j = 1, \ldots, m-1;\right. \\
&\qquad \left. T_m^{(i)} \notin [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U_m^{(i)} \in [0, \mathrm{UCL}_{\sigma}^{(i)}]\right) \\
&= P_{0,\theta}\left(T^{(i)} \notin [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U^{(i)} \in [0, \mathrm{UCL}_{\sigma}^{(i)}]\right) \\
&\qquad \times \sum_{m=1}^{\infty} \left[1 - \xi_{\mu,\sigma}^{(i)}(0,\theta)\right]^{m-1} \\
&= \frac{P_{0,\theta}\left(T^{(i)} \notin [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U^{(i)} \in [0, \mathrm{UCL}_{\sigma}^{(i)}]\right)}{\xi_{\mu,\sigma}^{(i)}(0,\theta)}.
\end{aligned}$$

$\mathrm{PMS}_{IV}^{(i)}(\delta)$ is derived similarly. □

Table 1 contains the critical values $\gamma_{\mu}^{(i)}$ and $\gamma_{\sigma}^{(i)}$, $i = 1, 2, 3$, of the individual charts of the three simultaneous Shewhart-type schemes for μ and σ^2. These values were determined for $n = 3, 5, 7$, by finding the value of $\beta^{(i)}$ which yields a specified probability of false alarm $\alpha = 1/1000, 1/500, 1/370.4$ for scheme$^{(i)}$, $i = 1, 2, 3$.

Please note that $\beta^{(1)}$ and $\gamma_{\mu}^{(1)}$ do not depend on n. Since this fact has implications in the behaviour of the PMS, namely of Type IV, it is going to be addressed in the next section.

4 Discussion, Numerical Results and Final Thoughts

The PMS of Type III (resp. IV) represents a means of studying the (in)ability of a simultaneous scheme for the process mean and spread to misinterpret a shift in σ (resp. μ) as a shift in μ (resp. σ).

In this section we provide values of the PMS of types III and IV, for the $(\bar{X},\ S^2)$, $(\sqrt{n}(\bar{X} - \mu_0)/S,\ S^2)$ and $(\bar{X},\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2)$ simultaneous schemes for μ

Table 1 Parameters of three simultaneous Shewhart-type schemes for μ and σ^2 ($\alpha = 1/1000, 1/500, 1/370.4$ and $n = 3, 5, 7$)

α	n	Statistics	i	$\beta^{(i)}$	$\gamma_{\mu}^{(i)}$	$\gamma_{\sigma}^{(i)}$	$\text{ARL}_{\mu,\sigma}^{(i)}(0,1)$
1/1000	3	$(\bar{X},\ S^2)$	1	0.000250	3.480689	15.201305	1000.0
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000250	44.704587	15.201805	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000281	3.449785	17.487214	
	5	$(\bar{X},\ S^2)$	1	0.000250	3.480689	19.996805	
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000250	10.306255	19.997355	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000266	3.464442	21.966495	
	7	$(\bar{X},\ S^2)$	1	0.000250	3.480689	24.102209	
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000250	6.788340	24.102799	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000260	3.469857	25.918977	
1/500	3	$(\bar{X},\ S^2)$	1	0.000500	3.290386	13.814510	500.0
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000500	31.599055	13.815511	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000566	3.255733	16.005584	
	5	$(\bar{X},\ S^2)$	1	0.000500	3.290386	18.465718	
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000500	8.610302	18.466827	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000535	3.271520	20.359516	
	7	$(\bar{X},\ S^2)$	1	0.000500	3.290386	22.456550	
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000500	5.958816	22.457744	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000524	3.277483	24.207792	

Table 1 (continued)

α	n	Statistics	i	$\beta^{(i)}$	$\gamma_{\mu}^{(i)}$	$\gamma_{\sigma}^{(i)}$	$\text{ARL}_{\mu,\sigma}^{(i)}(0,1)$
1/370.4	3	$(\bar{X},\ S^2)$	1	0.000675	3.204962	13.214110	370.4
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000675	27.190078	13.215461	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000766	3.168440	15.361211	
	5	$(\bar{X},\ S^2)$	1	0.000675	3.204962	17.799087	
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000675	7.959194	17.800591	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000724	3.184766	19.657645	
	7	$(\bar{X},\ S^2)$	1	0.000675	3.204962	21.737619	
		$(\sqrt{n}(\bar{X}-\mu_0)/S,\ S^2)$	2	0.000675	5.624694	21.739242	
		$(\bar{X},\ \sum_{i=1}^{n}(X_i-\mu_0)^2/n)$	3	0.000709	3.190994	23.458359	

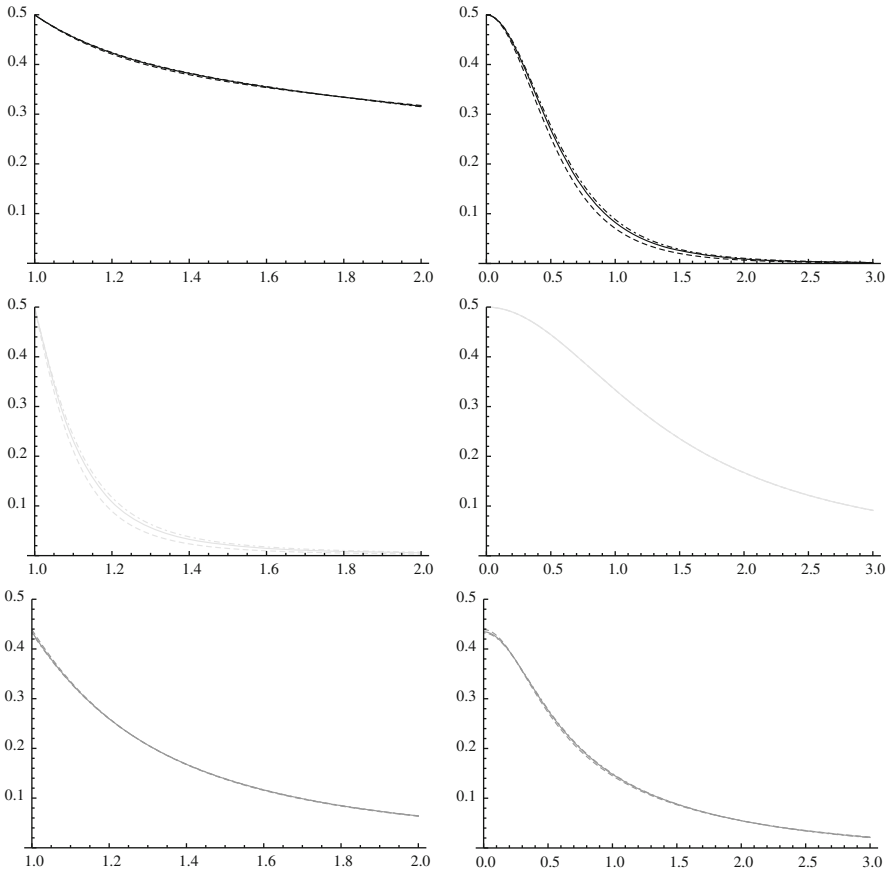


Fig. 1 Plots of $\mathrm{PMS}_{III}^{(i)}(\theta)$ (*left*) and $\mathrm{PMS}_{IV}^{(i)}(\delta)$ (*right*), for $i = 1, 2, 3$ (in *black, light grey, grey; top to bottom*, resp.), $\alpha = 1/1000, 1/500, 1/370.4$ (*dashed, solid and dot dashed lines*, resp.) and $n = 3$

and σ^2, considering: $\theta = 1.02, 1.1, 1.2, 2$; $\delta = 0.05, 0.5, 1, 2$.[1] Plots have been added, namely to make the comparison between these three simultaneous schemes possible.

Firstly, it is worthy of note that we obtained the graphs of the PMS of types III and IV for all three simultaneous schemes, sample sizes ($n = 3, 5, 7$) and significance levels ($\alpha = 1/1000, 1/500, 1/370.4$); interestingly enough, these PMS seem to be somewhat insensitive to the value of the in-control ARL, α^{-1}, as depicted in Fig. 1 for $n = 3$. In Table 2 we can find the values of the PMS of types III and IV only for $\alpha = 1/500$.

[1] Since the PMS of Type IV has the same value for $\delta = -c$ as for $\delta = c$, only positive values of δ are plotted or tabulated.

Table 2 $\mathrm{PMS}_{III}^{(i)}(\theta)$ and $\mathrm{PMS}_{IV}^{(i)}(\delta)$ for simultaneous Shewhart-type schemes for μ and σ^2 ($\alpha = 1/500$ and $n = 3, 5, 7$)

		$\mathrm{PMS}_{III}^{(i)}(\theta)$				$\mathrm{PMS}_{IV}^{(i)}(\delta)$		
		n				n		
i	θ	3	5	7	δ	3	5	7
1	1.02	0.489460	0.475799	0.465561	0.05	0.496112	0.496112	0.496112
2		0.433340	0.419974	0.410012		0.499377	0.498806	0.498367
3		0.411395	0.431322	0.433766		0.431778	0.462197	0.473257
1	1.1	0.454931	0.397725	0.357110	0.5	0.269291	0.269291	0.269291
2		0.231677	0.192758	0.167371		0.444525	0.400651	0.370967
3		0.332637	0.319955	0.297960		0.276739	0.289937	0.292367
1	1.2	0.423132	0.331381	0.271748	1	0.082472	0.082472	0.082472
2		0.108061	0.075944	0.058490		0.333556	0.238787	0.189438
3		0.259552	0.225470	0.193665		0.148157	0.144792	0.139310
1	2	0.315981	0.169326	0.102093	2	0.009078	0.009078	0.009078
2		0.005592	0.003031	0.002133		0.167056	0.072439	0.043153
3		0.064100	0.034090	0.019236		0.054896	0.047982	0.042719

Secondly, the examination of Table 2 shows that the PMS of Type IV does not depend on the sample size, when we are dealing with the $(\bar{X}, S^2)$ simultaneous scheme. As far as we know, this behaviour has not been reported before, essentially because the illustrations in the literature refer to a fixed sample size. The fact that $\text{PMS}_{IV}^{(1)}(\delta)$ is independent of n, for fixed α and δ, is easily justified. By "incorporating" the sample size in $\delta = \sqrt{n}(\mu - \mu_0)/\sigma_0$, the probability that the $\bar{X}$ chart signals in the absence of a shift in σ, $\xi_\mu(\delta, 1)$, no longer depends on n. Furthermore, by fixing the in-control ARL of the simultaneous scheme and demanding that both individual charts have the same in-control ARL, we get $\xi_\mu(\delta, 1) = 2\beta^{(1)} = \xi_\sigma(1)$, thus, also independent of the sample size. Finally, a close inspection of (4) leads to the conclusion that $\text{PMS}_{IV}^{(1)}(\delta)$ is a mere function of $\xi_\mu(\delta, 1)$ and $\xi_\sigma(1)$, hence independent of n.

Thirdly, we are bound to refer that collecting larger samples may cause slight increases in the PMS of Type III (resp. IV), when the $(\bar{X}, \sum_{i=1}^{n}(X_i - \mu_0)^2/n)$ simultaneous scheme is at use in the presence of small shifts in σ (resp. μ). On the one hand, this seems to be a startling result because the misinterpretation of a shift in σ as a shift in μ and vice versa should become less frequent as we collect more information. On the other hand, the values of (OC function) = (1 − Power function), in Table 3 of Walsh (1952), anticipated somehow such a surprising result:[2] they increase occasionally with the sample size (they should decrease instead!), when $\theta = 1$ and the $(\bar{X}, \sum_{i=1}^{n}(X_i - \mu_0)^2/n)$ simultaneous scheme is used.

Finally, other plots of the PMS of types III and IV have been added to this section and are assembled in Fig. 2, which gives a more accurate idea than Table 2 of how the PMS of the three simultaneous schemes in this paper compare.

Figure 2 gives considerable evidence that the simultaneous scheme based on $(\sqrt{n}(\bar{X} - \mu_0)/S,\ S^2)$ (light grey lines) tends to be superior (resp. inferior) to those based on $(\bar{X},\ S^2)$ (black lines) or $(\bar{X},\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2)$ (grey lines), as far as the PMS of Type III (resp. IV) is concerned. Furthermore, Fig. 2 tells us that the adoption of the $(\bar{X},\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2)$ simultaneous scheme might be a fairly good decision when it comes to reducing the PMS of types III and IV if large shifts in μ are not anticipated.

The over-all consensus seems to be that none of the two alternative simultaneous schemes are able to reduce the PMS of types III and IV of the simultaneous scheme based on $(\bar{X},\ S^2)$ and no pair of simultaneous schemes are roughly equivalent when it comes to the PMS of both types, as portrayed in Fig. 2. Consequently, the choice of one of the three simultaneous Shewhart-type schemes for μ and σ^2 should also depend on the amount of effort and money spent while attempting to identify and correct non-existing causes of variation in μ (resp. σ^2), i.e., when an MS of Type III (resp. IV) occurs, as put by Morais and Pacheco (2006). For example, it seems reasonable to favour a simultaneous scheme based on $(\sqrt{n}(\bar{X} - \mu_0)/S,\ S^2)$ if the

[2]Even if those values refer to a simultaneous scheme used to monitor both decreases and increases in the process mean or variance.

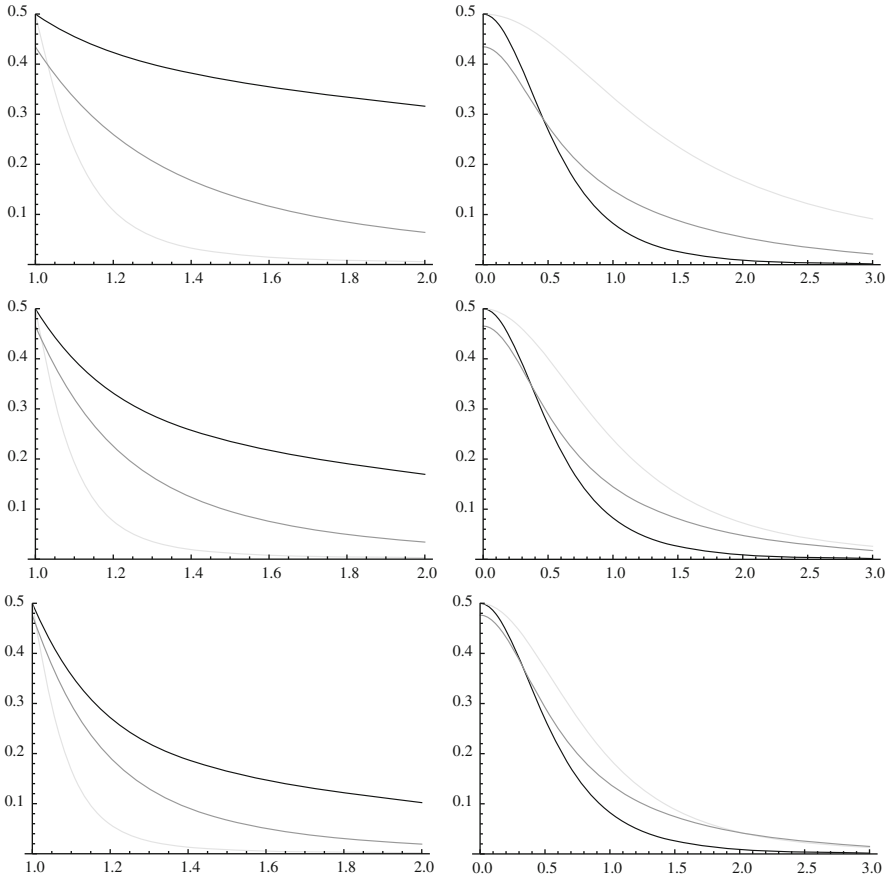


Fig. 2 Plots of $\mathrm{PMS}_{III}^{(i)}(\theta)$ (*left*) and $\mathrm{PMS}_{IV}^{(i)}(\delta)$ (*right*), for $i = 1, 2, 3$ (in *black, light grey, grey*, resp.), $\alpha = 1/500$ and $n = 3, 5, 7$ (*top to bottom*)

implications of MS of Type IV are unimportant and the occurrence of MS of Type III are a great cause of concern.

We would like to note in passing that using the EWMA counterpart of $(\bar{X},\ S^2)$ is a strategy to reduce the PMS of types III and IV that quality control practitioners should ponder too, as suggested by the results in Morais (2002, p. 125) and Morais and Pacheco (2006).

A possibility for future work is to investigate how the adoption of a control statistic (or a set of control statistics) for $\underline{\mu}$, which is not based on a quadratic form, can influence the PMS of simultaneous schemes for the mean vector and covariance matrix of multivariate normal output. The analysis of the effect of the adoption of other control charts for the covariance matrix on the PMS, like the

rather complex one proposed by Tang and Barnett (1996a,b), also deserves some consideration. These authors showed that this chart and a few other alternative outperformed the |S| chart; the control statistics of these alternative charts have a crippling disadvantage: even though they have well-known distributions under control, it is very difficult to derive the corresponding out-of-control distributions. As a consequence, we would have to resort to simulation, as Tang and Barnett (1996b) did, to investigate the impact of the adoption of these charts for $\underline{\Sigma}$ on both PMS.

Acknowledgements The first author gratefully acknowledges the financial support received from CEMAT (Centro de Matemática e Aplicações) to attend the XIth International Workshop on Intelligent Statistical Quality Control, Sydney, Australia, August 20–23, 2013.

Appendix

In this section we present the necessary results to compute the probabilities in the denominator and numerator of the PMS of types III and IV:

- $P_{\delta,\theta}\left(T^{(i)} \in [\mathrm{LCL}_\mu^{(i)}, \mathrm{UCL}_\mu^{(i)}],\ U^{(i)} \in [0, \mathrm{UCL}_\sigma^{(i)}]\right)$ equals

$$\begin{cases} \left[\Phi\left(\frac{\gamma_\mu^{(1)}}{\theta} - \frac{\delta}{\theta}\right) - \Phi\left(-\frac{\gamma_\mu^{(1)}}{\theta} - \frac{\delta}{\theta}\right)\right] \times F_{\chi^2_{(n-1)}}\left(\frac{\gamma_\sigma^{(1)}}{\theta^2}\right), & i = 1 \\ \int_0^{\frac{\gamma_\sigma^{(2)}}{\theta^2}} \left[\Phi\left(\frac{\gamma_\mu^{(2)}}{\sqrt{n-1}}\sqrt{x} - \frac{\delta}{\theta}\right) - \Phi\left(-\frac{\gamma_\mu^{(2)}}{\sqrt{n-1}}\sqrt{x} - \frac{\delta}{\theta}\right)\right] \times f_{\chi^2_{(n-1)}}(x)\,dx, & i = 2 \\ \int_{-\frac{\gamma_\mu^{(3)}}{\theta} - \frac{\delta}{\theta}}^{\frac{\gamma_\mu^{(3)}}{\theta} - \frac{\delta}{\theta}} \phi(z) \times F_{\chi^2_{(n)}}\left(\max\left\{0, \gamma_\sigma^{(3)}/\theta^2 - (z + \delta/\theta)^2\right\}\right) dz, & i = 3; \end{cases}$$

- $P_{\delta,\theta}\left(T^{(i)} \notin [\mathrm{LCL}_\mu^{(i)}, \mathrm{UCL}_\mu^{(i)}],\ U^{(i)} \in [0, \mathrm{UCL}_\sigma^{(i)}]\right)$ is given by

$$\begin{cases} \left\{1 - \left[\Phi\left(\frac{\gamma_\mu^{(1)}}{\theta} - \frac{\delta}{\theta}\right) - \Phi\left(-\frac{\gamma_\mu^{(1)}}{\theta} - \frac{\delta}{\theta}\right)\right]\right\} \times F_{\chi^2_{(n-1)}}\left(\frac{\gamma_\sigma^{(1)}}{\theta^2}\right), & i = 1 \\ \int_0^{\frac{\gamma_\sigma^{(2)}}{\theta^2}} \left\{1 - \left[\Phi\left(\frac{\gamma_\mu^{(2)}}{\sqrt{n-1}}\sqrt{x} - \frac{\delta}{\theta}\right) - \Phi\left(-\frac{\gamma_\mu^{(2)}}{\sqrt{n-1}}\sqrt{x} - \frac{\delta}{\theta}\right)\right]\right\} \times f_{\chi^2_{(n-1)}}(x)\,dx, & i = 2 \\ \int_{-\infty}^{-\frac{\gamma_\mu^{(3)}}{\theta} - \frac{\delta}{\theta}} \phi(z) \times F_{\chi^2_{(n-1)}}\left(\max\{0, \gamma_\sigma^{(3)}/\theta^2 - (z - \delta/\theta)^2\}\right) dz & \\ \quad + \int_{\frac{\gamma_\mu^{(3)}}{\theta} - \frac{\delta}{\theta}}^{\infty} \phi(z) \times F_{\chi^2_{(n)}}\left(\max\left\{0, \gamma_\sigma^{(3)}/\theta^2 - (z + \delta/\theta)^2\right\}\right) dz, & i = 3; \end{cases}$$

- $P_{\delta,\theta}\left(T^{(i)} \in [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U^{(i)} \notin [0, \mathrm{UCL}_{\sigma}^{(i)}]\right)$ is equal to

$$\begin{cases} \left[\Phi\left(\frac{\gamma_{\mu}^{(1)}}{\theta} - \frac{\delta}{\theta}\right) - \Phi\left(-\frac{\gamma_{\mu}^{(1)}}{\theta} - \frac{\delta}{\theta}\right)\right] \times \left[1 - F_{\chi^2_{(n-1)}}\left(\frac{\gamma_{\sigma}^{(1)}}{\theta^2}\right)\right], & i = 1 \\ \int_{\frac{\gamma_{\sigma}^{(2)}}{\theta^2}}^{\infty} \left[\Phi\left(\frac{\gamma_{\mu}^{(2)}}{\sqrt{n-1}}\sqrt{x} - \frac{\delta}{\theta}\right) - \Phi\left(-\frac{\gamma_{\mu}^{(2)}}{\sqrt{n-1}}\sqrt{x} - \frac{\delta}{\theta}\right)\right] \times f_{\chi^2_{(n-1)}}(x)\, dx, & i = 2 \\ \int_{-\frac{\gamma_{\mu}^{(3)}}{\theta} - \frac{\delta}{\theta}}^{\frac{\gamma_{\mu}^{(3)}}{\theta} - \frac{\delta}{\theta}} \phi(z) \times \left[1 - F_{\chi^2_{(n)}}\left(\max\left\{0, \gamma_{\sigma}^{(3)}/\theta^2 - (z + \delta/\theta)^2\right\}\right)\right] dz, & i = 3. \end{cases}$$

The derivation of $P_{\delta,\theta}(T^{(i)} \in [\mathrm{LCL}_{\mu}^{(i)}, \mathrm{UCL}_{\mu}^{(i)}],\ U^{(i)} \in [0, \mathrm{UCL}_{\sigma}^{(i)}])$ follows closely Walsh (1952). The apparent differences are essentially due to the fact that Walsh (1952) represented the shift in the process mean (resp. standard deviation) by $a = \sqrt{n}(\mu_0 - \mu)/\sigma$ (resp. $b = \sigma_0/\sigma$), i.e., $\delta = -a/b$ (resp. $\theta = 1/b$), and considered a chart for σ with a lower control limit.

The derivation of the remaining probabilities follows in a straightforward manner.

References

Antunes, C. (2009). *Avaliação do impacto da correlação em sinais Erróneos de esquemas Conjuntos para o valor esperado e variância (Assessment of the impact of the correlation on misleading signals in joint Schemes for the mean and variance)*. Master's thesis, Instituto Superior Técnico, Technical University of Lisbon.

Hawkins, D. M., & Maboudou-Tchao, E. M. (2008). Multivariate exponentially weighted moving covariance matrix. *Technometrics, 50*, 155–166.

Knoth, S., Morais, M. C., Pacheco, A., & Schmid, W. (2009). Misleading signals in simultaneous residual schemes for the mean and variance of a stationary process. *Communications in Statistics - Theory and Methods, 38*, 2923–2943.

Morais, M. J. C. (2002). *Stochastic ordering in the performance analysis of quality control schemes*. Ph.D. thesis, Instituto Superior Técnico, Technical University of Lisbon.

Morais, M. C., & Pacheco, A. (2000). On the performance of combined EWMA schemes for μ and σ: a Markovian approach. *Communications in Statistics—Simulation and Computation*, 29, 153–174.

Morais, M. C., & Pacheco, A. (2006). Misleading signals in joint schemes for μ and σ. In H. J. Lenz, & P. Th. Wilrich (Eds.), *Frontiers in Statistical Quality Control* (Vol. 8, pp. 100–122). Heidelberg: Physica-Verlag.

Nelson, L. S. (1982). Control charts. In S. Kotz, & N. L. Johnson (Eds.-in-Chief), *Encyclopedia of Statistical Sciences,* (Vol. 2, pp. 176–183). New York: Wiley.

Ramos, P. F., Morais, M. C., & Pacheco, A. (2013). Misleading signals in simultaneous residual schemes for the process mean and variance of AR(1) processes: a stochastic ordering approach. In J. L. Silva, F. Caeiro, I. Natário, & C. A. Braumann (Eds.), *Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications* (pp. 161–170). Berlin: Springer-Verlag.

Ramos, P. F., Morais, M. C., Pacheco, A., & Schmid, W. (2012). Assessing the impact of autocorrelation in misleading signals in simultaneous residual schemes for the process mean

and variance: a stochastic ordering approach. In H. J. Lenz, P. Th. Wilrich, & W. Schmid (Eds.). *Frontiers in Statistical Quality Control 10* (pp. 35–52). Heidelberg: Physica-Verlag.

Ramos, P. F., Morais, M. C., Pacheco, A., & Schmid, W. (2013a). Stochastic ordering in the qualitative assessment of the performance of simultaneous schemes for bivariate processes. *Sequential Analysis, 32*, 214–229.

Ramos, P. F., Morais, M. C., Pacheco, A., & Schmid, W. (2013b). Misleading signals in simultaneous schemes for the mean vector and the covariance matrix of a bivariate process. In P. E. Oliveira, M. G. Temido, M. Vichi, & C. Henriques (Eds.). *Recent Developments in Modeling and Applications in Statistics – Studies in Theoretical and Applied Statistics* (pp. 225–235). Berlin: Springer.

Ramos, P. F., Morais, M. C., Pacheco, A., & Schmid, W. (2013c). On the misleading signals in simultaneous schemes for the mean vector and covariance matrix of multivariate i.i.d. output. (Accepted for publication in Statistical Papers; published on line in Feb. 2015).

St. John, R. C. & Bragg, D. J. (1991). Joint X-bar R charts under shift in mu or sigma. *ASQC Quality Congress Transactions* (pp. 547–550). Milwaukee: American Society for Quality Control.

Tang, P. F., & Barnett, N. S. (1996a). Dispersion control for multivariate processes. *Australian Journal of Statistics, 38*, 235–251.

Tang, P. F., & Barnett, N. S. (1996b). Dispersion control for multivariate processes — some comparisons. *Australian Journal of Statistics, 38*, 253–273.

Walsh, J. E. (1952). Operating characteristics for tests of the stability of a normal population. *Journal of the American Statistical Association, 47*, 191–202.

Wieringa, J. E. (1999). *Statistical Process Control for Serially Correlated Data*. Ph.D. thesis, Faculty of Economics, University of Groningen.

Wikipedia (2012). *Code of Hammurabi* — Wikipedia, The Free Encyclopedia, Accessed from http://en.wikipedia.org/wiki/Code_of_hammurabi on 2012–11-21.

Characteristics of Economically Designed CUSUM and $\bar{X}$ Control Charts

Erwin Saniga, Darwin Davis, Alireza Faraz, Thomas McWilliams, and James Lucas

Abstract In this paper we investigate the characteristics of economic control chart designs for both Shewhart ($\bar{X}$) and CUSUM control charts. Authors in the past have made some suggestions regarding the design of these charts, where design is defined as finding the values of sample size, intersample interval and control limit (Shewhart chart) or control parameters (k and h) for the CUSUM chart. Here, we run a large number of experiments consisting of many configurations of the parameters and describe and model the results in terms of the actual economic designs.

Keywords Control charts • Economic design • CUSUM charts • Shewhart $\bar{X}$ charts

1 Introduction

Previous research by Saniga et al. (2012) and Saniga et al. (2006a,b) has shown that the statistical advantage of CUSUM control charts for controlling a sample mean versus the Shewhart $\bar{X}$ chart translates to an economic advantage provided the fixed cost of sampling is small or when there are two components of variance as is common in process industries. These results were obtained experimentally by

E. Saniga (✉) • D. Davis
Department of Business Administration, Alfred Lerner College of Business & Economics, University of Delaware, Newark, DE 19716, USA
e-mail: saniga@udel.edu; dd@udel.edu

A. Faraz
HEC-management School, University of Liège, 4000 Liège, Belgium
e-mail: Alireza.Faraz@ulg.ac.be

T. McWilliams
Decision Sciences Department, LeBow College of Business, Drexel University, Philadelphia, PA 19104, USA
e-mail: tmcwilliams@drexel.edu

J. Lucas
James Lucas and Associates, Wilmington, DE, USA
e-mail: James.Lucas@verizon.net

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_13

finding economic designs of CUSUM charts and $\bar{X}$ charts for a wide range of input parameters, a wide range of shift sizes and for situations in which there is one or possibly two components of variance. In this paper we delve into the experimental results to investigate the characteristics of these designs. More precisely, we investigate the distributions of the design parameters over the experimental range, determine what input parameters seem to most significantly affect the design, look at ARL_0 contours by k and h for the CUSUM chart where ARL_0 is the in control average run length of the CUSUM chart and k is the reference value and h is the decision interval. We also look at regions of these contours to show that in some cases an $\bar{X}$ chart may be employed while in most cases a CUSUM chart should be employed because of its economic advantage. This chart also gives some design guidelines for k and h for the economically designed CUSUM chart.

Comparison of CUSUM and $\bar{X}$ charts is important as each has certain advantages and disadvantages in controlling a process mean. Montgomery (2001) notes that one should use the CUSUM control chart if one wants to detect a shift that is expected to be small because the CUSUM chart has a marked statistical advantage for small shifts. While the CUSUM chart is optimal (see Moustakides 1986) in the sense that the average run length (*ARL*) of the CUSUM is minimum for detecting any particular shift for a fixed in control *ARL*, this advantage is larger for small shifts. In fact, it is well known that the Shewhart $\bar{X}$ chart is a CUSUM chart for larger shifts. The ease of design, use, maintenance and lack of abstractness might point to an advantage of using the $\bar{X}$ chart in this situation.

This is contrary to the conclusion of Reynolds and Stoumbos (2004) who recommend the use of $n = 1$ CUSUM or EWMA charts in general if the criterion is statistical. At times, though, the simplicity of the $\bar{X}$ chart in terms of ease of use and design might make it the control chart of choice, especially when one wants to guard against especially large shifts.

Some authors including Goel (1968), von Collani (1987), Saniga et al. (2006a,b) have compared the CUSUM chart to the $\bar{X}$ chart in terms of cost under the assumption that each was designed in an economically optimal fashion and there was a single component of variance process. Saniga et al. (2012) have done the same comparison for the two components of variance process.

The first two studies found that there is no real cost advantage to the CUSUM chart and that the design of the optimal CUSUM chart is one in which the reference value k is large and the decision interval h is small which makes it more or less equivalent to the $\bar{X}$ chart (i.e., the $\bar{X}$ chart is a special case of the CUSUM chart where $k = 3$ and $h = 0$).

The Saniga et al. (2006b) and Saniga et al. (2012) studies contradict the previous studies, finding that first, consideration of fixed costs may make the CUSUM chart with $n = 1$ far from economically optimal; second, there are regions where the CUSUM chart is economically advantageous; third, optimal economic designs of CUSUM charts are unlike the $\bar{X}$ chart in terms of k and h; and fourth, even if small shifts are expected the CUSUM chart may not have an economic advantage. A fifth

conclusion is that in the presence of two components of variance the CUSUM chart dominates the $\bar{X}$ chart.

In this paper, we employ the same experimental results of Saniga et al. (2012) but we now investigate in detail the characteristics of economically designed CUSUM and $\bar{X}$ control charts. In the next section we discuss the cost model, and in Sect. 3 we present the experiment. Section 4 describes the results, and Sect. 5 presents some conclusions.

2 The Cost Model

One of the first to model a process where quality is monitored by a control chart was Duncan (1956). This model was expanded to cover a wider range of industrial applications by Lorenzen and Vance (1986). Most models in this realm assume that the data are available to estimate the average shift size, that the time in control follows the negative exponential distribution, that all system and cost parameters are known, and that the process is in state of statistical control. That is, we are monitoring a stable process. In this paper we will use the Lorenzen and Vance (1986) and the terms defined in that paper. Some of the terms specific to this paper are:

n = sample size
g = hours between samples
k = number of standard deviations from control limits to center line (for the $\bar{X}$ chart only)
k = reference value for the CUSUM chart
h = decision interval for the CUSUM chart
δ = number of standard deviations shift when out of control
E = time to sample and chart one item
Y = cost per false alarm
W = cost to locate and repair the assignable cause
A = fixed cost per sample
B = variable cost of sampling
T_0 = expected search time for a false alarm
T_1 = expected time to find an assignable cause
T_2 = expected time to repair the assignable cause
C_0 = cost per hour when in control
C_1 = cost per hour when an assignable cause exists
λ = arrival rate of an assignable cause
σ_B^2 = between sample variance
σ_W^2 = within sample variance
D_1 = 1 if production continues during searches
= 0 if production ceases during searches
D_2 = 1 if production continues during repair
= 0 if production ceases during repair

We use this model to find the economic design for the CUSUM chart and the $\bar{X}$ chart. Our algorithm to find the optimal economic designs is discussed in Saniga et al. (2006b). The discussion of standardized shifts when there are two components of variance is contained in Saniga et al. (2012).

3 The Experiments

We designed four experiments—three are essentially the same as in Saniga et al. (2006b) after example problems in the literature provided by Chiu (1974) and Lorenzen and Vance (1986). These are defined in Saniga et al. (2012), but we provide some definition here as well because there is a slight difference. In each experiment we find optimal $\bar{X}$ control chart designs as well as optimal CUSUM control chart designs.

The four experiments are as follows:

Experiment A: Here we find designs for 7,680 configurations of cost and system parameters derived from Chiu's (1974) first example. In this experiment we allow $C_0 = 0$; $C_1 = 100, 500, 1,000$; $b = 0.1, 1$; $a = 0.5, 10$; $\lambda = 0.01, 0.05$; $\delta = 0.5, 1, 2, 3$; $\delta_1 = \delta_2 = 0, 1$; $Y = W = 75, 500$; $t_0 = t_1 = 0.1, 0.5$; and $t_2 = 0.2, 0.5$. We set $E_0 = 0$. Also, we vary $\sigma_B^2 = 0, 0.25, 0.5, 1$ and 2. In all experiments we fix $\sigma_W^2 = 1$ and let $\sigma_B^2 = 0, 0.25, 0.5, 1$ and 2.

Experiment B: This experiment is the same as the previous experiment except that $E = 0.5$. Thus, it includes 7,680 configurations of input parameters.

Experiment C: This experiment is the same as the previous two experiments except that $E = 1.0$. It also includes 7,680 configurations of input parameters.

Experiment D: This experiment alters the parameters of the example in Lorenzen and Vance (1986). In this case $C_0, C_1 = (0, 835)(114.2, 949)$; $b = 0.1, 0.5$; $a = 0, 10, 50, 200$; $\lambda = 0.02, 0.05$; $\delta = 0.5, 0.86, 1.5$; $Y = W = (200, 200)(977, 977)(1,500, 1,500)$; $t_0 = t_1 = 0.0833, 0.5$; $t_2 = 0.75, 1.5$; $E_0 = 0.0833, 0.5$. We fix $\delta_1 = 1$, $\delta_2 = 0$ and we vary $\sigma_B^2 = 0, 0.25, 0.5, 1$ and 2. There are 17,280 configurations of parameters in this experiment.

In each of the experiments we also allow an alternative process control strategy to be employed if it is economically better than either the CUSUM or $\bar{X}$ chart. This policy is a regular search policy. In this case the process is shut down every g h and a search for an assignable cause takes place.

4 Analysis of Results

Finding an economic design for a practical scenario is a difficult problem for several reasons. First, some knowledge of cost accounting is necessary to accurately determine the various costs specified in the Lorenzen and Vance (1986) model. Second, some care must be taken to accurately determine the other system parameters; this

Table 1 Significant values ($\alpha = 0.05$) for various design parameters ✓ = Experiment A, ∗ = Experiment B, X = Experiment C, 0 = Experiment D

$\bar{X}$ CHART

	ARL_0				ARL_1				g				k				n			
Y	✓	*	X	0	✓	*	X	0		*	X	0	✓	*	X	0	✓	*	X	0
A	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
B	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
λ	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0		*	X	0
δ	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
σ_B^2	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
C_0								0				0				0				0
C_1		*				*	X	0				0				0			X	0
T_0		*	X									0			X					
D_1										*		0		*						
E				0				0				0				0			X	0

CUSUM CHART

	ARL_0				ARL_1				g				k				h				n			
Y	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0		*	X	0			X	0
A	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
B	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
λ	✓	*	X					0		*	X	0	✓	*	X	0	✓	*		0		*	X	0
δ	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
σ_B^2	✓	*	X	0	✓	*		0	✓	*	X	0	✓	*	X	0	✓	*	X	0	✓	*	X	0
C_0				0				0				0				0				0				0
C_1				0				0			X	0		*	X	0				0			X	0
T_0		*	X							*		0						*	X					
T_1																								
E			X	0				0				0			X	0				0			X	0

can be done by a systematic data collection strategy. Third, nonlinear programming must be employed to find these designs although there are published algorithms that allow these to be calculated for $\bar{X}$ and R charts (see, e.g., McWilliams et al. 2001).

Our first analysis was to run regression models of the design parameters (g, k, and h for the $\bar{X}$ chart and g, k, h and n for the CUSUM chart) and ARL_0 and ARL_1 for both charts. All models run were linear with respect to all parameters. The model for n, say, is $n = f(Y, A, B, \lambda, \delta, \sigma_B^2, C_0, C_1, T_0, T_1, E) + \epsilon$.

Table 1 contains the results for the four experiments. It is interesting to note that, generally, the first six variables Y, A, B, λ, δ, σ_B^2 are significant (at $\alpha = 0.05$) in most of the 11 regression models estimated. Experiment D, which is the Lorenzen and Vance (1986) example also finds E to be a significant variable for all of the design and design related parameters. We caution that the determination of significance is based upon an assumption of normality. The actual models are available upon request.

Future work involves finding designs based upon a model based upon just these parameters (or fixing the others at a constant level) and investigating the robustness of this reduced size model on the estimation of design parameters and the errors in cost estimation. Of course, the latter is of secondary importance if the design parameters are correctly determined.

Table 2 shows the means and ranges of the various input parameters for Experiment A broken down as follows: experiments in which $n > 0$ for the CUSUM and $\bar{X}$ chart *and* $ARL_0 \geq 500$, experiments in which $n > 0$ for the CUSUM and $\bar{X}$ chart, and experiments in which $n = 0$ for the CUSUM and $\bar{X}$ chart. The first case is one in which an economic design would be feasible where we define a feasible

Table 2 Input parameter means and ranges for Experiment A

		E	T0	T1	T2	C0	C1	Y	W	A	B	λ	δ	D1	D2	σ^2_b	
n=0	x-bar	0.000	0.303	0.303	0.350	0.00	523.64	230.30	230.30	6.58	0.626	0.031	0.577	0.494	0.494	1.073	Mean
		0	0.4	0.4	0.3	0	900	425	425	9.5	0.9	0.04	0.5	1	1	2	Range
	CUSUM	0.000	0.305	0.305	0.350	0.00	513.43	162.87	162.87	8.24	0.693	0.031	0.612	0.488	0.488	1.085	Mean
		0	0.4	0.4	0.3	0	900	425	425	9.5	0.9	0.04	0.5	1	1	2	Range
		E	T0	T1	T2	C0	C1	Y	W	A	B	λ	δ	D1	D2	σ^2_b	
n>0	x-bar	0.000	0.299	0.299	0.350	0.00	535.74	301.69	301.69	4.92	0.531	0.030	1.885	0.501	0.501	0.670	Mean
		0	0.4	0.4	0.3	0	900	425	425	9.5	0.9	0.04	2.5	1	1	2	Range
	CUSUM	0.000	0.299	0.299	0.350	0.00	536.43	306.91	306.91	4.78	0.528	0.030	1.783	0.502	0.502	0.698	Mean
		0	0.4	0.4	0.3	0	900	425	425	9.5	0.9	0.04	2.5	1	1	2	Range
		E	T0	T1	T2	C0	C1	Y	W	A	B	λ	δ	D1	D2	σ^2_b	
n>0 ARL_0>500	x-bar	0.000	0.300	0.300	0.350	0.00	535.68	417.30	417.30	3.53	0.358	0.030	2.546	0.500	0.500	0.420	Mean
		0	0.4	0.4	0.3	0	900	425	425	9.5	0.9	0.04	2	1	1	2	Range
	CUSUM	0.000	0.299	0.299	0.350	0.00	535.20	445.02	445.02	2.32	0.463	0.030	1.939	0.502	0.502	0.637	Mean
		0	0.4	0.4	0.3	0	900	425	425	9.5	0.9	0.04	2.5	1	1	2	Range

design as one in which the charts have an $ARL_0 \geq 500$, which is an arbitrary but plausible number for an in control *ARL*. Keep in mind the in control *ARL* for Shewhart's $\bar{X}$ chart is about 370. For the CUSUM, 12,227 of the 40,320 runs in the four experiments were feasible. For the $\bar{X}$ chart, 4,690 of the 40,320 runs were feasible. Results for all four experiments are available upon request.

Woodall (1986) has argued that there are practical problems with economic designs including the possibility of having an economically optimal design with a low in control *ARL*. This implies that there would be frequent false searches for false alarms causing a lack of trust in the control chart and possibly incorrect process adjustments leading to an increase in the variability of the process.

The results in Table 2 show that generally, with smaller shift sizes and larger fixed costs of sampling, a regular search policy (i.e., $n = 0$) might be advantageous to using a control chart for both the CUSUM and $\bar{X}$ chart. A regular search policy is one in which the process is investigated at every g interval of time without sampling to see if an assignable cause has occurred. Now, feasible economic designs are those in which the variables A and B are smaller and the variables Y and W are larger. Of course, it is quite easy to find a feasible design by adding various constraints to the optimization problem as suggested by Saniga (1989).

Figures 1 and 2 show the quantile plots for the design parameters and *ARLs* for all CUSUM designs, and Figs. 3 and 4 show the quantile plots for the design parameters and *ARLs* for all $\bar{X}$ chart designs. The "c" in the figures refers to the CUSUM chart and the "x" refers to the $\bar{X}$ chart. Note from Figs. 2 and 4 that the median ARL_0 is less than 500 for the CUSUM designs and for the $\bar{X}$ chart designs. The observed high frequency of low ARL_0 values is consistent with Woodall's observation regarding economic designs. Also, n can be quite large for both designs, especially for experiment A, possibly causing the loss of the advantage of subgrouping in later data analysis. Note from the $\bar{X}$ chart designs that k seems to have a median value around 2.6, which is, of course, less than Shewhart's recommended 3σ control limit. This is another way of stating that the ARL_0 may be too low for most of the cases we investigated where too low is defined as an

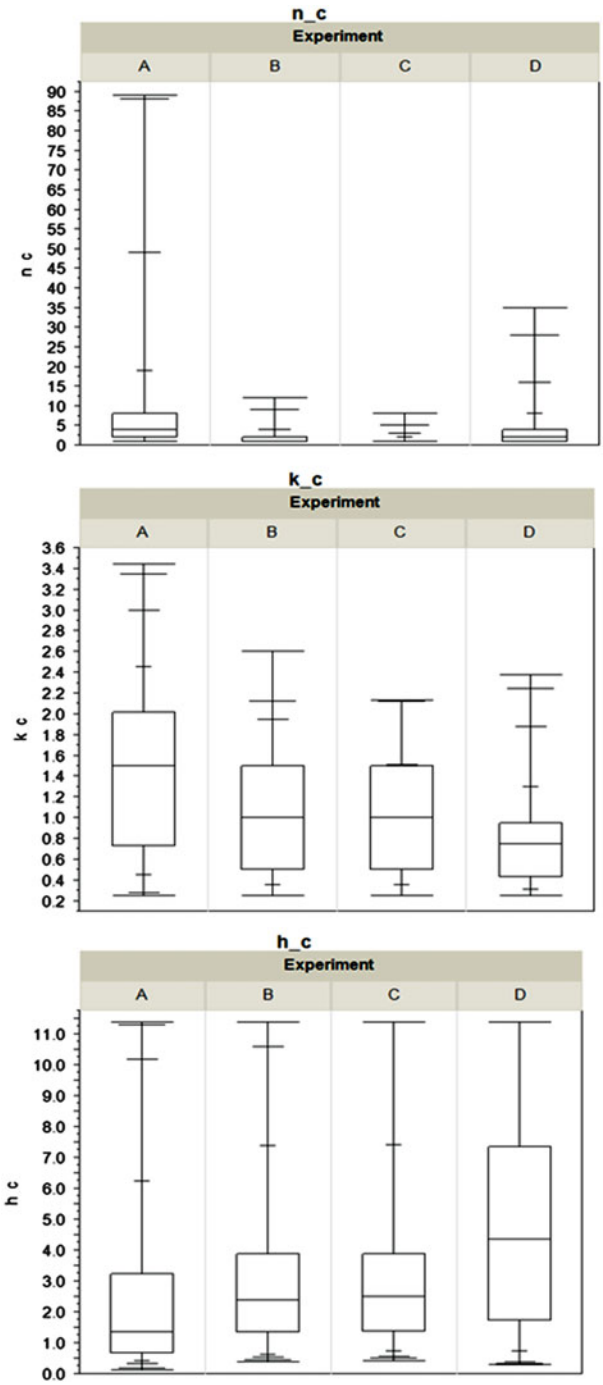


Fig. 1 Quantile plots for CUSUM designs

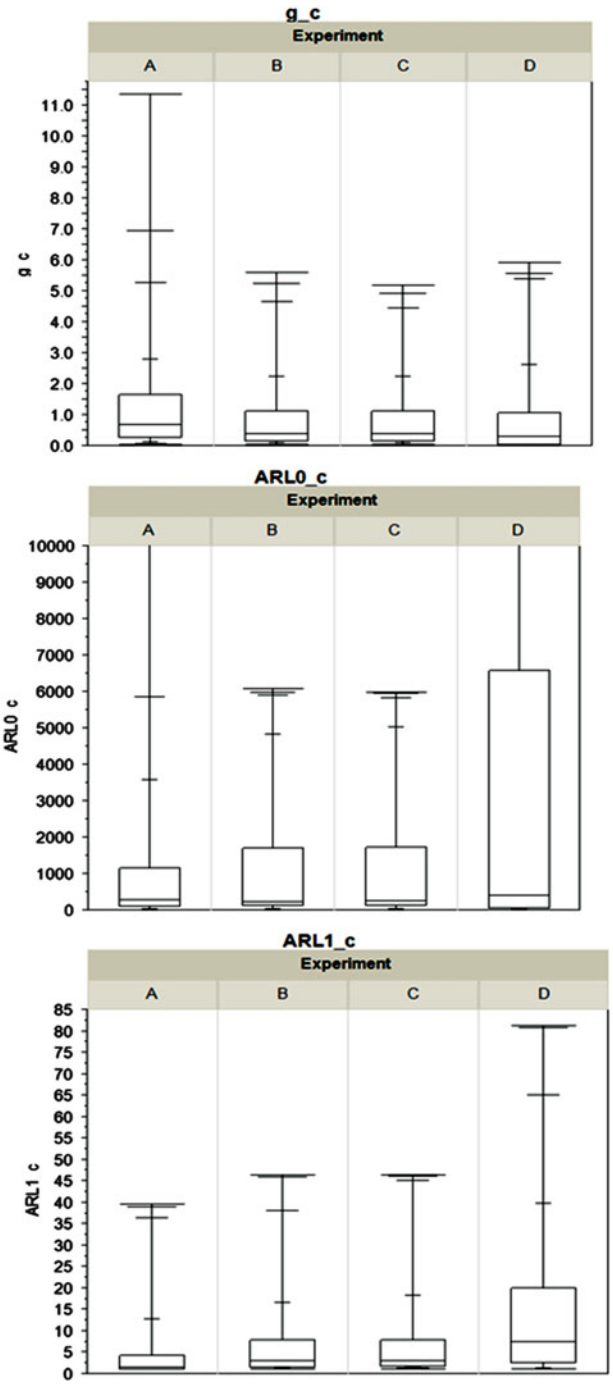


Fig. 2 Quantile plots for CUSUM designs

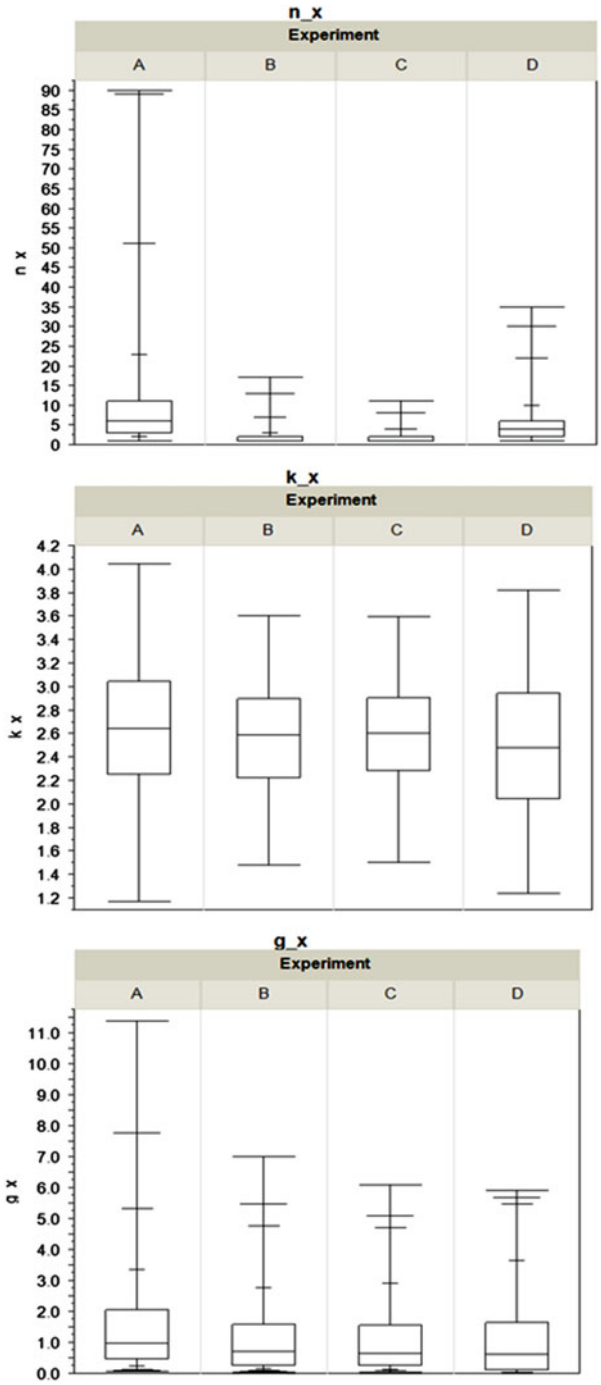


Fig. 3 Quantile plots for $\overline{X}$ designs

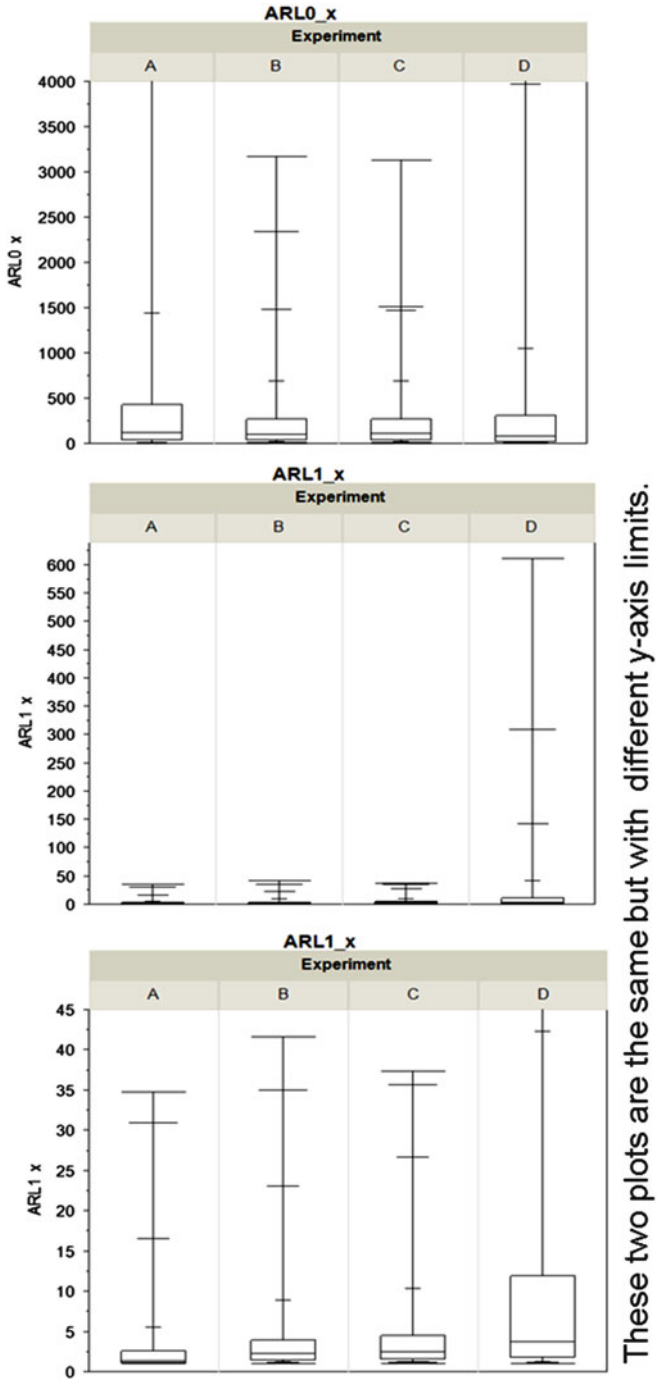


Fig. 4 Quantile plots for $\overline{X}$ designs

$ARL_0 < 370$, say. Of course, one can add *ARL* constraints to solve this problem as done by Saniga (1989). We have also done quantile plots for the CUSUM chart and the $\bar{X}$ chart with $ARL_0 \geq 500$. These are available upon request.

Sampling frequencies seem to have a median around one for most experiments for both charts as well, indicating that the recommendation of taking one sample every hour is a good one in general, although there are many examples where a higher (and lower) sampling frequency should be used. Keep in mind that one can add a constraint here as well to ensure a sampling frequency that meets the temporal imperative of the organization.

Figures 5 and 6 show scatter plots for various design and input parameters for CUSUM designs, and Fig. 7 shows scatter plots for various design and input parameters for $\bar{X}$ designs. For the first set of plots given in Figs. 5 and 6 one sees the relationship of k and h for the CUSUM; as h increases k decreases at a decreasing rate. Perhaps there is a positive relationship between k and n as well for two of the experiments while g and n do not seem to be related. The rule of thumb that $k = \delta/2$ seems to be violated when discussing economic design rather than statistical design as well, except for the larger shift sizes. These results also hold true for the scatter plots when $ARL_0 \geq 500$ for both charts; these plots are available upon request.

For the $\bar{X}$ chart there seem to be no relationships between the design parameters.

Figure 8 shows contour plots for various in control *ARL*s as a function of k and h for the CUSUM chart. Note that many of these are nonmonotonic; perhaps, the best explanation of this is that we are plotting a range of *ARL*s in each case and the plotting algorithm in JMP causes this outcome. For a particular h though, it is apparent that a higher k must be used to ensure feasible *ARL*s. The same is true for a fixed k; higher h's result in a higher *ARL*.

We have isolated the extreme regions of the highest *ARL* contour in this graph; these regions are circled and labeled AREA 1 and AREA 2. In the figure we present the cost ratios of the $\bar{X}$ chart to the CUSUM chart in the upper right-hand side of the figure. One can easily see that in AREA 1 one could employ an $\bar{X}$ chart without a cost disadvantage; i.e., the cost ratio is about 1 which mean both charts cost the same. In AREA 2 though, one sees the marked advantage of the CUSUM chart economically, having an average cost advantage of 68 %. Moreover, optimal economic CUSUM designs, especially when they dominate $\bar{X}$ designs, are dissimilar in terms of k and h to an $\bar{X}$ design.

We also present the mean and range of the input parameters for all outcomes in these areas. Note that some of the same conclusions drawn by Saniga et al. (2012) are apparent here. In particular, note that the CUSUM chart is cost dominant when the shift size δ is small (and the $\bar{X}$ chart is not disadvantageous in terms of cost when δ is large ($\delta = 3$)). A larger Y and W and especially a small fixed cost of sampling, A, as well as a two component of variance process, makes the CUSUM a definite choice economically.

Figure 9 shows the same graph as Fig. 8 except that we show the *ARL* contours separately for each of the experiments.

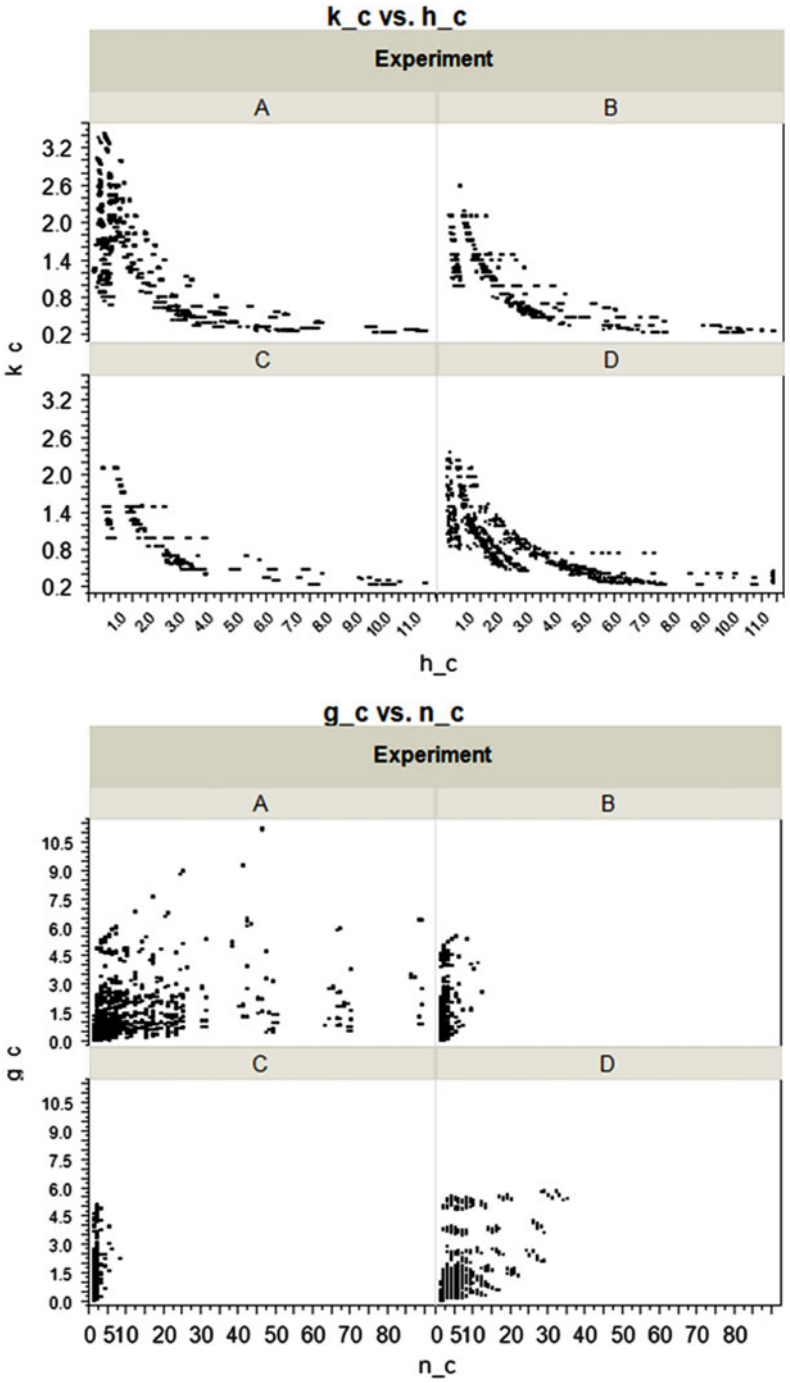


Fig. 5 Scatter plots for CUSUM designs

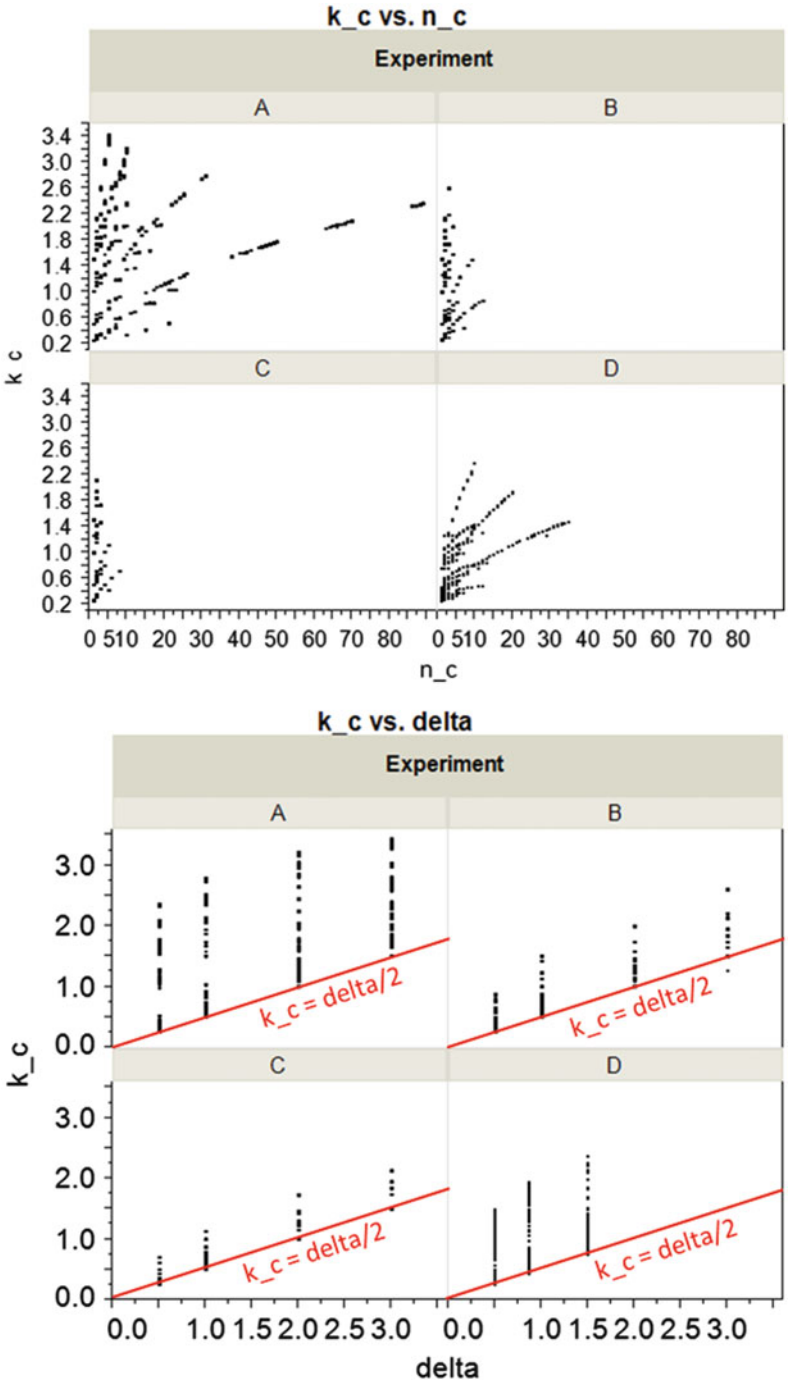


Fig. 6 Scatter plots for CUSUM designs

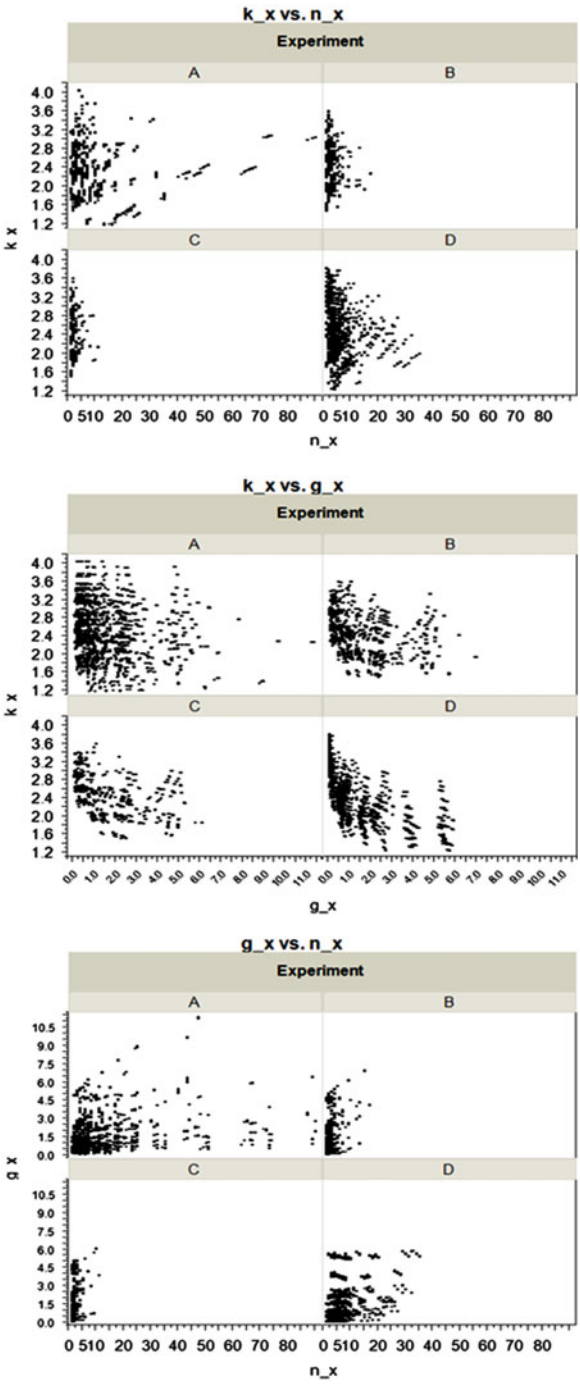
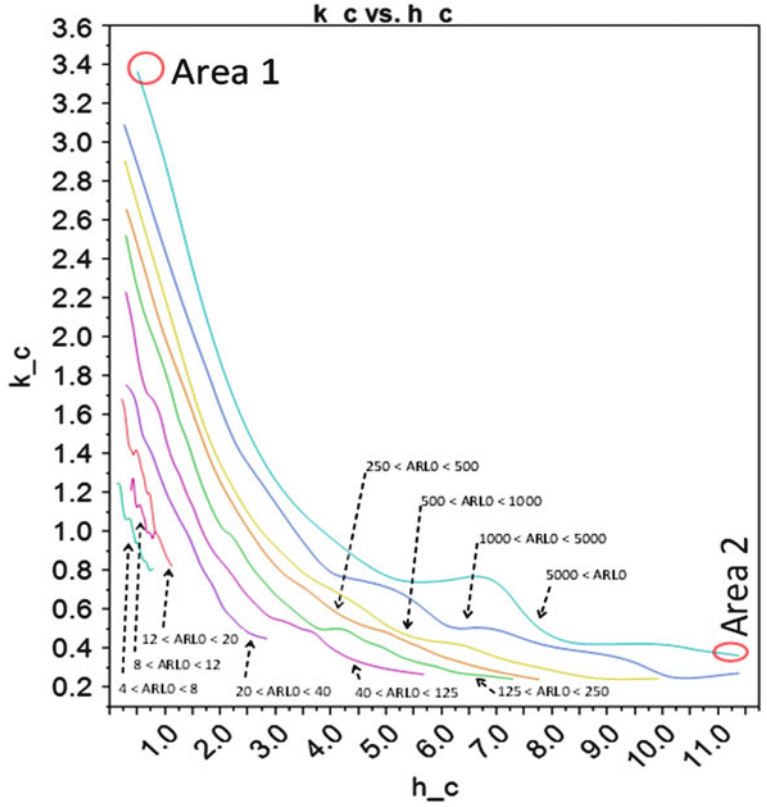


Fig. 7 Scatter plots for $\overline{X}$ designs



	Area 1		Area 2	
E	0.000	0	0.307	0.417
T0	0.290	0.4	0.298	0.417
T1	0.290	0.4	0.298	0.417
T2	0.350	0.3	1.114	0.75
C0	0.00	0	55.49	114.2
C1	511.90	900	890.39	114
Y	500.00	0	1316.38	1300
W	500.00	0	1316.38	1300
A	10.00	0	0.00	0
B	0.100	0	0.873	0.9
λ	0.031	0.04	0.038	0.03
δ	3.000	0	0.500	0
D1	0.548	1	1.000	0
D2	0.548	1	0.000	0
σ^2_b	0.000	0	0.784	2
	Mean	Range	Mean	Range

Cost Ratio (x-bar / CUSUM)

	Area 1	Area 2
	3.32 ≤ k ≤ 3.44	0.30 ≤ k ≤ 0.30
	0.48 ≤ h ≤ 0.61	11.34 ≤ h ≤ 11.36
ARL0 > 5000	Ave = 0.99998	Ave = 1.68417
	Min = 0.99995	Min = 1.17114
	Max = 1.00000	Max = 2.25989
	n = 42	n = 177

Fig. 8 Data from Experiments A, B, C & D

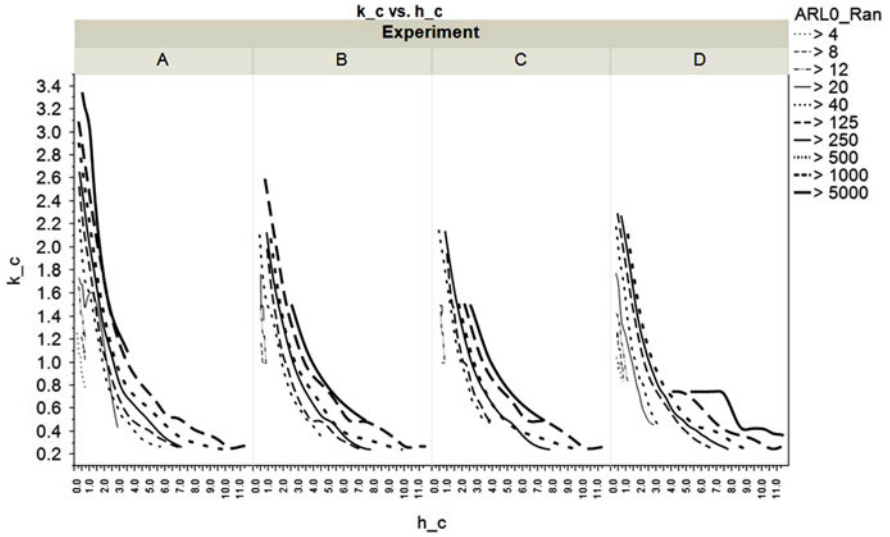


Fig. 9 ARL_0 ranges by Experiment

5 Conclusions

We have done extensive data analysis on many economic control chart designs for both the $\bar{X}$ chart and the CUSUM chart. These are presented in various forms in this paper. Some general conclusions are that economic control chart designs may not be a good choice unless constraints on *ARL*s and other requirements of particular processes are carefully managed; economically optimal CUSUM designs do not mirror $\bar{X}$ chart designs in terms of k and h; optimal CUSUM designs do not follow the statistical rule of thumb that $k = \delta/2$; and that CUSUM designs should be used when certain process parameters are at certain levels.

Acknowledgements The authors wish to thank the referees for their helpful comments.

References

Chiu, W. K. (1974). The economic design of cusum charts for controlling normal means. *Applied Statistics*, *23*, 420–433.

Duncan, A. J. (1956). The economic design of X bar charts used to maintain current control of a process. *Journal of the American Statistical Association*, *51*, 228–242.

Goel, A. L. (1968). A comparative and economic investigation of X bar and cumulative sum control charts. Unpublished Ph.D Dissertation, University of Wisconsin, Madison.

Lorenzen, T. J., & L. C. Vance (1986). The economic design of control charts: a unified approach. *Technometrics*, *28*, 3–10.

McWilliams, T., Saniga, E., & Davis, D. (2001). Economic-statistical design of $\bar{X}$ and R charts or $\bar{X}$ and S charts. *Journal of Quality Technology*, *33*(2), 234–241.

Montgomery, D. C. (2001). *Introduction to statistical quality control* (11th ed.). Hoboken, N.J.: Wiley.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, *14*, 1379–1387.

Reynolds, M. Jr., & Stoumbos, Z. (2004). Control charts and the efficient allocation of sampling resources. *Technometrics*, *46*, 200–214.

Saniga, E. (1989). Economic statistical control chart designs with an application to $\bar{X}$ and R charts. *Technometrics*, *31*, 313–320.

Saniga, E., Lucas, J., Davis, D., & McWilliams, T. (2012). Economic control chart policies for monitoring variables when there are two components of variance. In H.-J. Lenz, P.-T. Wilrich, & W. Schmid (Eds.) *Frontiers in statistical quality control, 10* (pp. 85–95). Heidelberg: Physica Verlag.

Saniga, E., McWilliams, T., Davis, D., & Lucas, J. (2006a). Economic advantages of CUSUM control charts for variables. In H.-J. Lenz, & P.-T. Wilrich (Eds.) *Frontiers in Statistical Quality Control 8* (pp. 185–198). Heidelberg: Physica Verlag.

Saniga, E., McWilliams, T., Davis, D., & Lucas, J. (2006b). Economic control chart policies for monitoring variables. *International Journal of Productivity and Quality Management*, *1*(1), 116–138.

von Collani, E. (1987). Economic process control. *Statistica Neerlandica*, *41*, 89–97.

Woodall, W. H. (1986). Weaknesses of the economic design of control charts (Letter to the Editor). *Technometrics*, *28*, 408–410.

SPC of Processes with Predicted Data: Application of the Data Mining Methodology

Olgierd Hryniewicz

Abstract SPC procedures are usually designed to control stability of directly observed parameters of a process. However, when quality parameters of interest are related to reliability characteristics it is practically hardly possible to monitor such characteristics directly. Instead, we use some training data in order to build a model that is used for the prediction of the value of an unobservable variable of interest basing on the values of observed explanatory variables. Such prediction models have been developed for normally distributed characteristics, both observable and unobservable. However, when reliability is concerned the random variables of interest are usually described by non-normal distributions, and their mutual dependence may be quite complicated. In the paper we consider the model of a process when traditionally applied assumptions are violated. We show that in such a case some non-statistical prediction models proposed in the area of data-mining, such as Quinlan's C4.5 decision tree, perform better than popular linear prediction models. However, new problems have to be considered when shifts in the levels of process parameters may influence the performance of applied classification algorithms.

Keywords Monitoring of unobserved variables • Quality prediction • Shewhart charts

1 Introduction

For many years procedures of Statistical Quality Control (SQC) have been used for the analysis of independent, and usually normally distributed, quality characteristics. With the development and automatization of new measurement techniques production processes can be now described by many, usually interdependent, characteristics. For many years the T^2 control chart, introduced by Hotelling in the 1947, was the only SPC tool used for SPC of processes described by multivariate data, see

O. Hryniewicz (✉)
Systems Research Institute, Newelska 6, 01-447 Warsaw, Poland
e-mail: hryniewi@ibspan.waw.pl

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_14

Montgomery (2011). However, during the last 20 years some new techniques have been proposed for dealing with interdependent statistical quality data. For example, control charts for the parameters of the so-called profiles have been introduced in order to control not only numerical values of quality characteristics, but the structure of their mutual dependence as well, see the paper by Woodall et al. (2004), the book by Noorsana et al. (2011), and recent papers by Xu et al. (2012), and by Wang and Huwang (2012) for more information. These methods can be used for the analysis of different dependencies of a regression type, both linear and non-linear. However, in practically all cases the proposed models have been obtained under the assumption of normality of measured characteristics. Moreover, it is assumed that all important quality characteristics of interest are *directly* measurable.

In contemporary production processes important parameters of produced objects are usually measured for all produced items. When specification limits are set for the values of these measurements one can say that a 100 % quality inspection has been implemented for this process. However, the parameters that can be measured during the production process are not necessarily the same as the quality characteristics that determine the quality of produced items. For example, important reliability characteristics such as the lifetime cannot be measured during a production process. The same is with all quality characteristics whose measurement may have a negative impact on the quality of inspected items. There are also other quality characteristics whose measurements are costly (e.g. when the time of measurement is too long for a production process), and thus infeasible. In all such cases there are attempts to measure these characteristics indirectly by the measurements of other characteristics.

The problem of an indirect inspection of important quality characteristics attracted the attention of relatively few authors for the last more than 50 years. There exist two general approaches. In the first approach, introduced by Owen and collaborators, see Owen and Su (1977), a multivariate probability distribution of the random vector $(Z, X_1, \ldots, X_k)$ is built, where Z is the quality characteristic of interest, and $X_1, \ldots, X_k$ are the characteristics which are directly measurable in a production process. This approach gives acceptable results only in the case of the multivariate (usually bivariate) normal (Gaussian) distribution describing $(Z, X_1, \ldots, X_k)$. Another approach is based on the assumption that the relation between the random variable Z and the variables $X_1, \ldots, X_k$ is described by a certain (usually linear) regression model. Also in this case the normality assumption about Z is usually used in practice. In both cases there is a direct link of the proposed methods to the multivariate SPC tools mentioned in the first paragraph of this section.

Unfortunately, the models mentioned above are of rather limited applicability when the actual multivariate probability distribution of $(Z, X_1, \ldots, X_k)$ is different from the normal (Gaussian) one, and when the number of predictors (explanatory variables) $X_1, \ldots, X_k$ is not small. In such cases building of a well-established probabilistic model is rather infeasible. Instead, one can think about the usage of the data mining methodology for a simple classification of inspected items. In the first step used in this approach some (usually two: conforming and nonconforming) classes of

inspected items are defined in relation to the possible values of Z. Then, a classifier (e.g. linear classifier, decision tree or artificial neural network) is built using a training data consisted of the limited number of observations $(Z, X_1, \ldots, X_k)$. Finally, the classifier is used in the inspection process for "labeling" the produced items.

Classifiers used in the inspection process are usually built using small amount of data, named training data. Thus, the results of classification are not error-free. What is more important, however, that the relation between the results of classification and the actual level of the quality characteristic of interest may be quite complicated. Therefore, there is a need to investigate the impact the quality of the classification procedures on the efficiency of SPC procedures used in production processes.

The remaining part of this paper is organized as follows. Section 2 is devoted to the problem of the prediction of directly unobserved quality characteristics using data mining techniques. The simulation model used for the evaluation of different prediction algorithms is described. Then, the performance of two data mining algorithms, namely the Linear Discrimination Analysis (LDA) and the Classification Decision Tree Quinlan's C4.5 algorithm, is evaluated in terms of prediction errors for both non-shifted and shifted process levels. The properties of classical Shewhart control charts for attributes used for the predicted quality data are discussed in Sect. 3. Some conclusions and indication for future work are presented in the last section of the paper.

2 Quality Prediction of Indirectly Observed Processes: Simulation Experiments

The problem of process control when quality characteristics of interest are not directly observable, but only assessed on the basis of observations of other, possibly related, variables is much more complicated than the classical one when all variables of interest are directly observable. The variability of such processes consists of two parts. One is related to the variability of the process itself, and the second one is related to unavoidable uncertainty of classification (prediction) procedures. What is more important, these two types of variability are practically inseparable. In this section we are using the results of simulation experiments with the aim to evaluate possible extent to which these two types of variability (and especially the second one) may influence the performance of SPC procedures.

As it has been noted in the previous section, the majority of statistical procedures used for the prediction of unobservable quality characteristics is based on the assumption of multivariate normality. Usually this assumption is reduced to the case that the quality characteristic of interest and its observable predictor are jointly distributed according to a bivariate normal distribution. When several possible predictors are available it is also often assumed that these predictors are statistically independent. Under such assumptions simple regression models (usually linear) are

built and used for the purpose of quality evaluation. Unfortunately, when quality characteristics of interest describe reliability these simple assumptions are hardly acceptable. First of all, reliability characteristics, such as the lifetime, are usually described by strongly skewed distributions. Moreover, predictors are frequently modeled by random variables defined on subsets of positive real numbers, and their distributions can be quite far from the normal distribution. Finally, behind the values of observed predictors there are some common physical and chemical phenomena which often make them strongly statistically dependent. One can also add another dimension by assuming strong non-linearity of the relations describing physical phenomena with the observed lifetimes, described, e.g., by models of the so-called competitive risks. Thus, the real models describing the process of, e.g., reliability prediction may be very complicated, and usually extremely difficult to identify.

In order to investigate the impact of some of the problems mentioned above on the efficiency of prediction (classification) process we have built a simulation model consisted of three levels. This construction reflects the situation when an observed failure is a result of the activation of one or more possible hidden mechanisms. On the first level we have four random variables, denoted by A, B, C, D, respectively, which describe observable characteristics. These variables may be described by several probability distributions (normal, uniform, exponential, Weibull, log-normal) chosen by an experimenter. Observed variables may be pairwise dependent, and their dependence may be described by several copulas (normal, Clayton, Gumbel, Frank) chosen by an experimenter. The strength of dependence is defined by the value of Kendall's coefficient of association τ. Detailed information about the usage of copulas for the description of complex dependence structures can be found in the monograph by Nelsen (2006). These assumptions allow to simulate quite complicated structures of interdependent predictors. On the second level we have four hidden (unobservable) random variables H_A, H_B, H_C, H_D defined on a positive part of the real line, and having the interpretation of the activation times of hidden failure mechanisms. Their probability distributions may be chosen from the set of distributions used in the theory of reliability (exponential, Weibull, log-normal). Each of the hidden random variables is related to its respective observed variable, i.e. H_A to A, H_B to B, etc., and this relation is described by a chosen copula (with a given value of Kendall's τ) describing their joint probability distribution, and a certain linear dependence between their expected values. Finally, on the third level, hidden random variables are transformed to the final random variable E that describes the lifetime that can be observed only in specially designed experiments. The relation between H_A, H_B, H_C, H_D and E is strongly non-linear, and is described by operators of a "min-max" type. This type of non-linear relations is observed in practice when an observed failure can be considered as the result of the activation of the so-called competing risks.

The simulation system described above allows to simulate sets of data with a very complex, and practically impossible to be predicted in advance, structure. In this paper we show the results of experiments of only one of the investigated models. In this model A is distributed according to the normal distribution, B has the exponential distribution, C is distributed according to the log-normal distribution,

and D has the Weibull distribution. The dependence between A and B has been described by the Clayton copula with $\tau = 0,8$. The joint distribution of B and C is described by the normal copula with $\tau = -0,8$ (Notice that this is bivariate "normal" distribution, but with non-normal marginals!), and the joint distribution of C and D has been described by the Frank copula with $\tau = 0,8$. The hidden variable H_A is described by the log-normal distribution, and its joint probability distribution with A has been described by the normal copula with $\tau = -0,8$. The joint distribution of H_B and B has been described by the Frank copula with $\tau = 0,9$, and the marginal distribution of H_B is assumed to be the exponential. The joint model of H_C and C is similar, but the copula describing the dependence in this case is the Gumbel copula. Finally, the hidden variable H_D is described by the Weibull distribution, and its joint probability distribution with D has been described by the Clayton copula with $\tau = -0,8$. The random variable E that describes the lifetime has been defined as $E = min[max(H_A, H_B), H_C, H_D]$. The parameters of the aforementioned distributions have been found experimentally in such a way, that unreliable items have their lifetimes E smaller than 5. Moreover, the relation between the observed variables A, B, C, D and their hidden counterparts H_A, H_B, H_C, H_D is such that a shift in the expected value of each observed variable, measured in terms of its standard deviation, results with the similar shift of the expected value of its hidden counterpart, measured in terms of its own standard deviation.

The ultimate goal of the performed simulation experiments is to evaluate the efficiency of several classifiers which can be used for the prediction of reliability of produced items. As it has been written in the Introduction we are looking for a "labeling" classification procedure which assigns labels for potentially reliable and unreliable items. In the practice of data mining such classification procedures are designed using training data. It is rather obvious that the amount of training data significantly influences the efficiency of classification. In the majority of practical examples described in the literature training data sets have several hundreds units. Unfortunately, such numerous data sets are hardly possible in reliability tests. Our experience shows that $n = 100$ can be regarded as an upper practical limit for the size of training data. In the simulation experiment described in this paper we have used this value for the size of the training data sets in order to show how different classifiers may perform in the most favorable, when the efficiency of classification is taken into account, situation. In our simulation experiments we have also taken into account the random variability of training data sets used for the construction of classifiers. We have randomly generated ten different training data sets each of $n = 100$ elements. For these different data sets we designed different classification procedures (classifiers). The simulation experiment designed this way should reflect the impact of random variability in choosing the training data on the efficiency of quality inspection processes.

2.1 The Case of Constant Process Levels

There exist dozens of methods used for solving classification problems. Some of them, based on some statistical assumptions, have certain optimal properties. The properties of other methods, mainly based on a data mining approach, can be only assessed experimentally. In our research we have considered the performance of several classification methods in the analysis of data generated by our simulation system. Because of limited space we restrict ourselves to only two of them.

First considered classifier is based on classical statistical results of Fisher. It is known as the Linear Discrimination Analysis (LDA), and is described in many textbooks on multivariate statistical analysis, and data mining (see, e.g. Hastie et al. (2008). In this method statistical data are projected on a certain hyperplane estimated from the training data. Those data points who are closer to the mean value of the projected on this hyperplane training data representing the class 1 than to the mean value of training data representing the remaining class 2 are classified to the class 1. Otherwise, they are classified to the class 2. The equation of the hyperplane is given by the following formula:

$$L = y_A A + y_B B + y_C C + y_D D + y_F, \tag{1}$$

where L is the value of the transformed data point calculated using the values of explanatory variables A, B, C, D, and y_A, y_B, y_C, y_D, y_F are respective coefficients of the LDA equation estimated from a training set of n elements. If Z_L denote the decision point, a new item is classified to the class 1 if $L \geq Z_L$, and to the class 2, otherwise. In our research we considered several methods of the calculation of Z_L, but finally we present the results when this point is just the average of the mean values of the transformed data points from the training set that belong to the class 1 and the class 2, respectively. The calculation of the coefficients in the LDA equation (1) is not so simple. However, it can be done using basic versions of many statistical packages such as SPSS, STATISTICA, etc.

The second considered classification method is based on the implementation of the one of the most popular data mining classification algorithms, namely the classification decision tree (CDT) algorithm C4.5 introduced by Quinlan (1993), and described in many textbooks on data mining, such as Witten et al. (2011). In our experiments we used its version (known as J48) implemented in the WEKA software, available from the University of Waikato, Hamilton, New Zealand, under the GNU license. The decision tree is constructed using "IF..THEN..ELSE" rules, deducted from the training data. In the description of this classification method in this paper we use the notation of the MS Excel function $IF(lt, t, f)$, where lt is a logical condition (e.g. $C < 50$), t is the action when $lt = true$, and f is the action when $lt = false$. The actions t and f can be implementations of other IF functions, or—finally—the assignments of classes to the considered items.

The elements of decision rules (LDA linear equations or decision tree "IF.. THEN..ELSE" rules) are estimated from training data sets. In our experiment these

training data sets have been generated by our simulation program. In the artificial intelligence community it is assumed that good training data sets should consist of several hundreds of items. In our reliability prediction problem such large data sets are absolutely infeasible. In our simulation experiments for the generation of training data sets we have taken an upper feasible limit on the number n of observed items, namely $n = 100$. In order to estimate the effect of the randomness of the training data sets on the classification decision rules, and finally on the results of classification during a production process, we have generated several different data sets. For each of these data sets we have built LDA and CDT classification rules. Because of the restricted space of this paper we present the results for only 10 such sets.

A comparison of the decision model parameters for different training data sets in the LDA case is presented in Table 1. In this case we cannot say about statistical significance of the parameters of the decision rule. However, the general impression is that some predictors are of limited importance for the classification purposes. The particular models look completely different depending on the training data set. However, in all the cases the explanatory variable C seems to have no effect (very low values of the coefficient describing this variable) on the classification.

Now, let us consider different decision rules estimated for the CDT algorithm. Because of a completely different structure of decision rules presented in Table 2 we cannot compare directly these rules with the rules described by the Eq. (1). They also look completely different for different training data sets, but in nearly all cases (except for the Set 9) decision are predominantly (and in one case exclusively) based on the value of the explanatory variable C. One can notice that the weight assigned to the explanatory variables in the CDT algorithm is nearly exactly opposite to the weights assigned in the LDA classification model (1). In order to explain this shocking difference one should look at Fig. 1.

The dependence between the lifetime E and the explanatory variables C and D is not only non-linear, but non-monotonic as well. This dependence cannot be captured by the measures of linear correlation in the linear model (1). However, it

Table 1 Linear discrimination analysis—different sets of training data

Dataset	y_A	y_B	y_C	y_D	y_F	Midpoint
Set 1	0.687	0.174	0.001	0.133	−6.338	0.628
Set 2	−0.045	0.178	−0.001	0.151	−2.710	−0.014
Set 3	−0.646	0.148	<0.0005	−0.001	1.663	0.464
Set 4	0.880	0.152	0.002	0.087	−7.499	0.585
Set 5	1.500	0.091	0.003	−0.008	−9.121	0.254
Set 6	−0.342	0.219	<0.0005	0.107	−1.399	0.706
Set 7	0.703	0.196	0.002	0.037	−6.044	0.784
Set 8	−0.501	0.173	<0.0005	0.006	0.636	0.629
Set 9	1.458	0.127	0.001	0.143	−10.048	0.344
Set 10	0.008	0.087	0.002	−0.206	0.272	0.771

Table 2 Decision trees—different sets of training data

Dataset	Decision rule
Set 1	$IF(C <= 70,0181; IF(C <= 56,1124; 1; IF(C <= 63,2962; 2; 1)); IF(D <= 16,4381; 2; IF(A <= 4,3509; 2; 1)))$
Set 2	$IF(C <= 56,4865; 1; IF(D <= 17,3301; 2; IF(A <= 4,0217; 2; 1)))$
Set 3	$IF(C <= 73,6148; IF(C <= 57,1355; 1; IF(A <= 5,0876; 2; IF(D <= 4,497; 2; 1))); IF(D <= 17,3499; 2; 1))$
Set 4	$IF(C <= 70,2191; 1; IF(D <= 15,9098; 2; 1))$
Set 5	$IF(C <= 73,1584; 1; (IF(D <= 17,0516; (IF(C <= 87,8921; (IF(D <= 5,0679; 2; 1)); 2)); 1)))$
Set 6	$IF(C <= 60,3912; 1; IF(D <= 16,3504; 2; 1))$
Set 7	$IF(C <= 71,8184; 1; 2)$
Set 8	$IF(C <= 71,4456; 1; (IF(C <= 983,0929; 2; (IF(D <= 18,8213; 2; 1)))))$
Set 9	$IF(B <= 14,7339; (IF(D <= 16,7482; (IF(D <= 4,527; 1; 2)); 1)); 1)$
Set 10	$IF(C <= 60,5044; 1; 2)$

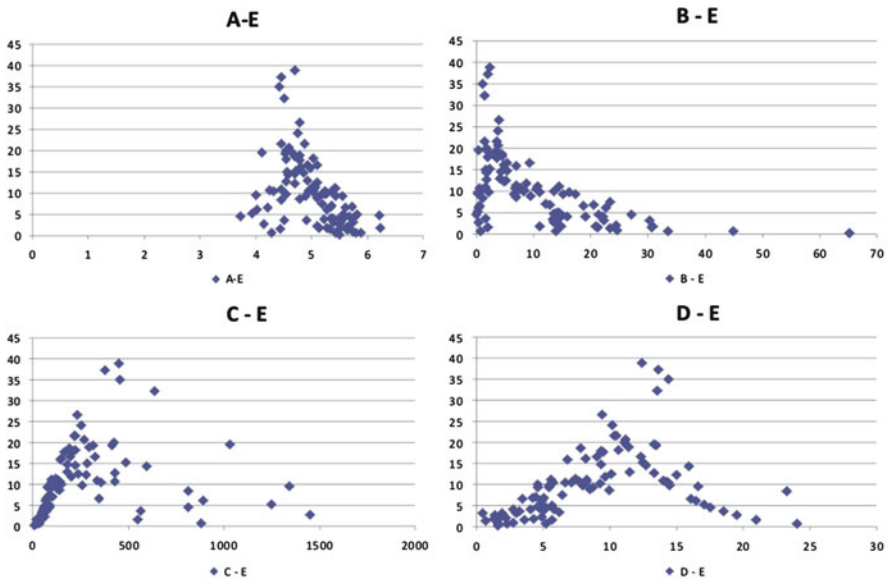


Fig. 1 Dependencies between the lifetime E and explanatory variables A, B, C, D

seems that the explanatory potential of these two variables is much greater than the potential of the variables A and B. We have investigated this problem in our further simulation experiments.

The simplest method for the comparison of classification algorithms is the comparison of their false classification rates. In this comparison one should make distinction between false non-detection of unreliable items (i.e. those with the lifetime in our simulation experiment smaller than 5), and the false classification

of reliable items. In many application the false classification of the first type is much more dangerous, as, e.g., unreliable components may be installed in technical systems that are important for safety of people. In Table 3 we compare false classification rates estimated from samples of 100,000 elements. The subscripts *nd* denote the false non-detection of unreliable items, and the subscripts *fd* denote the false classification of reliable items. The results for the regression algorithm are denoted by L for the LDA algorithm and by T for the CDT algorithm, respectively.

A close look at Table 3 reveals some interesting features. First, one can see that the false detection rates vary significantly depending upon the training data set. It shows that the practical necessity to use small training data sets has a negative impact on the stability of the classification procedures. Second, the classification algorithm based on the decision tree approach is visibly more accurate than the LDA one. Additional interesting information about the efficiency of compared algorithms can be found in Table 4 where falsely classified items have been presented in terms of percentages of all items belonging to a given class. Subscripts *pf* denote the percentages of unreliable items that have been classified as reliable ones, and subscripts *pnf* denote the percentages of reliable items that have been falsely classified as unreliable ones. When we compare the considered two algorithms we

Table 3 Percentages of misclassified data—different sets of training data

Dataset	L_{nd}	L_{fd}	T_{nd}	T_{fd}
Set 1	7.29	7.13	6.56	3.08
Set 2	4.71	14.14	6.78	1.86
Set 3	8.23	3.89	3.55	4.07
Set 4	9.94	4.15	2.01	7.46
Set 5	5.89	9.83	0.74	9.26
Set 6	8.28	4.73	4.62	4.57
Set 7	6.40	7.37	4.97	2.93
Set 8	7.95	4.21	4.36	3.14
Set 9	6.94	9.36	2.34	9.42
Set 10	6.63	4.53	7.58	0.87

Table 4 Relative percentages of misclassified data—different sets of training data

Dataset	L_{pf}	L_{pnf}	T_{pf}	T_{pnf}
Set 1	28.5	9.6	25.7	4.1
Set 2	18.8	18.9	27.1	2.5
Set 3	32.1	5.2	13.9	5.5
Set 4	39.3	5.6	8.0	10.0
Set 5	23.5	13.1	2.9	12.4
Set 6	32.3	6.4	18.0	6.1
Set 7	25.1	9.9	19.9	3.9
Set 8	31.4	5.7	17.2	4.2
Set 9	27.7	12.5	9.3	12.6
Set 10	26.6	6.0	30.2	1.2

Table 5 Observed fraction of potentially non-reliable items—different classifiers

Dataset	Actual	LDA	CDT
Set 1	0.256	0.254	0.221
Set 2	0.251	0.345	0.201
Set 3	0.256	0.213	0.261
Set 4	0.253	0.195	0.397
Set 5	0.251	0.290	0.336
Set 6	0.256	0.221	0.256
Set 7	0.255	0.264	0.234
Set 8	0.253	0.216	0.241
Set 9	0.251	0.275	0.321
Set 10	0.251	0.230	0.183

can see that in the visible majority of cases the CDT algorithm outperforms the LDA algorithm.

If one looks at percentages of misclassified data it is immediately seen that the observed quality level of a process may be, depending on a classifier used, completely different from the actual one. In Table 5 we present the estimated, from samples of 100,000 items, values of process levels for two considered classifiers. In the first column we present the value of the actual process level, as if it were actually observed in different samples of that size.

The differences between the observed process levels and the actual one are really striking. Even for classifiers of the same type the differences between the observed process levels resulting from the randomness of training data sets are very large. Therefore, the observed fractions nonconforming can be very different from the actual one. Hence, by applying SPC tools for such data one can only monitor the stability of the process, but not the value the actual process level.

2.2 The Case of Shifted Process Levels

In the previous section we have considered the case when observations of explanatory (predictive) and decision variables are governed by the same random mechanisms as those forming the training data, and used for building classifiers. However, considered processes may vary in time in many different ways. In our experiments we consider only the simplest case when the expected value of only one of explanatory variables is shifted (up or down) by the value equal to one half of the standard deviation of this variable. This shift has been chosen deliberately small in order to reveal possible differences in the change of the proportion of unreliable items resulted from the change of process parameters.

First of all, we have to note that the shift of the expected value of an explanatory variable may result in two mechanisms of the change of the observed process level. First, it changes, but to usually unknown extent (because of very complicated

relations between the values of considered variables), the actual fraction of unreliable items. Second, it results in changing the efficiency of classification rules. After such a shift has occurred the existing classification rules do not fit to the actual data, and the rates of false classification may change quite dramatically.

In our experiments we have considered the impact of the process shift on the false classification rates of several classifiers, but in this paper we present the results for only two of them: the LDA linear classifier, and the decision tree based classifier (CDT). This impact has been evaluated on samples of 100,000 items, for all considered sets of classifiers (obtained for different training data sets).

First, let us consider the positive shifts of the process levels, and suppose that each shift is observed for only one of the explanatory variables. In all cases these shifts result in the decrease of the actual (not observed in a real process!) fraction of unreliable items. However, only for the variables C and D the fraction of unreliable items has decreased significantly. The differences between the No shift, Shift A and Shift B cases may be attributed to random errors of simulation. In Table 6 we show how these fractions are changing following the shift of the process level for a given explanatory variable.

In Tables 7 and 8 we show how the probabilities of wrong classification are changing for such shifts of the expected values of the explanatory variables that result in the decrease of the actual fraction nonconforming. In the first column of these tables we present these probabilities when process data have the same distribution as in the case of no shift.

In the case of the LDA classifier, and positive small shifts in the expected values of the explanatory variables, the average false classification rate (probabilities), averaged over all considered sets of classifiers with different classification rules, have similar values for the case of no-shift, and for the cases of all considered explanatory variables. Therefore, small positive shifts of the expected values of the explanatory variables usually do not change the average efficiency of the LDA linear

Table 6 Fraction of potentially non-reliable items—shifts of 0.5σ

No shift	Shift A	Shift B	Shift C	Shift D
0.253	0.252	0.252	0.234	0.244

Table 7 LDA—probability of wrong classification (shift 0.5σ)

Dataset	No shift	Shift A	Shift B	Shift C	Shift D
Set 1	14.4	15.4	16.6	15.1	16.0
Set 2	18.9	18.6	24.1	18.6	31.8
Set 3	12.1	12.1	13.2	11.7	11.1
Set 4	14.1	14.0	14.1	16.0	14.1
Set 5	15.7	20.5	17.0	22.4	14.8
Set 6	13.0	12.8	14.6	13.0	12.9
Set 7	13.8	14.5	16.2	15.9	13.3
Set 8	12.2	11.9	13.5	12.0	11.2
Set 9	16.3	20.8	15.7	17.3	20.1
Set 10	11.2	11.2	11.6	13.4	10.4

Table 8 CDT—probability of wrong classification (shifts of 0.5σ)

Dataset	No shift	Shift A	Shift B	Shift C	Shift D
Set 1	9.6	9.8	9.6	15.1	17.2
Set 2	8.6	8.4	8.8	15.6	18.4
Set 3	7.6	7.6	7.7	14.7	18.3
Set 4	9.5	9.5	9.5	14.6	22.7
Set 5	10.0	9.9	9.8	11.0	22.1
Set 6	9.2	9.3	9.1	16.2	22.1
Set 7	7.9	7.7	7.9	12.7	6.9
Set 8	7.5	7.5	7.5	12.2	7.7
Set 9	11.8	11.7	10.7	12.3	23.2
Set 10	8.4	8.5	8.6	15.7	7.5

Table 9 Fraction of potentially non-reliable items—shifts of -0.5σ

No shift	Shift A	Shift B	Shift C	Shift D
0.253	0.252	0.253	0.303	0.269

classifier. Only in the case of the shift in the expected value of the variable D the variability of the observed false classification rates is larger than in the remaining cases.

The situation is completely different in the case of the decision tree classifier (CDT). The average values of false classification rates are nearly the same in the cases of no-shift, and shifts in the levels of A and B. However, in the case of the variables C and, especially, D the probabilities of false classification are visibly higher. It is due to the fact that the CDT classifiers are built mainly on the observed values of these variables. Moreover, in the cases of shifts in the levels of C and D these probabilities may vary in a wide range, depending on the randomly chosen classification rules.

Now, let us consider the negative shifts of the process levels when the shift is observed for only one of the explanatory variables. For the explanatory variables A and B such shifts do not significantly change the actual non-observed value of the process level (the differences are due to a random simulation error). However, for explanatory variables C and D such shifts result in the significant increase of the actual (not observed in a real process!) fraction of unreliable items. In Table 9 we show how these fractions are changing following the shift of the process level for a given explanatory variable.

In Table 10 we present probabilities of wrong classification according to the LDA algorithm when we observe small negative shift. Similarly to the case of positive small shifts these probabilities are very similar with some exceptions in the case of the shift of the variable C.

The situation changes significantly if we use the CDT algorithm for a negatively shifted process. As it is seen from Table 11 the probability of wrong classification increases dramatically if we decrease the expected value of the variable C. As we see from Table 2 the classification rules obtained using this algorithm mainly depend upon the value of this particular variable. The only exception is for Set 9 where the

Table 10 LDA—probability of wrong classification (shifts of -0.5σ)

Dataset	No shift	Shift A	Shift B	Shift C	Shift D
Set 1	14.4	13.7	13.9	15.0	14.9
Set 2	18.9	19.1	15.2	18.4	16.4
Set 3	12.1	12.2	13.1	13.9	13.8
Set 4	14.1	16.4	18.2	20.0	17.7
Set 5	15.7	13.6	14.6	13.9	17.0
Set 6	13.0	13.1	13.6	35.6	14.3
Set 7	13.8	13.0	13.0	13.2	14.9
Set 8	12.2	12.4	13.0	13.6	13.7
Set 9	16.3	14.6	15.7	16.3	16.2
Set 10	11.2	11.2	11.5	12.6	14.6

Table 11 CDT—probability of wrong classification (shifts of -0.5σ)

Dataset	No shift	Shift A	Shift B	Shift C	Shift D
Set 1	9.6	9.4	9.6	33.8	11.0
Set 2	8.6	9.0	9.0	29.0	11.1
Set 3	7.6	7.9	7.5	30.9	10.9
Set 4	9.5	9.5	9.4	40.1	9.1
Set 5	10.0	9.9	9.8	45.7	10.0
Set 6	9.2	9.3	9.2	56.8	10.1
Set 7	7.9	7.8	8.0	39.1	9.5
Set 8	7.5	7.6	7.4	38.8	9.4
Set 9	11.8	11.7	10.9	11.4	17.5
Set 10	8.4	8.6	8.4	32.5	10.2

decision rule does not depend upon the value of C. One may say, that the CDT classification algorithm in some cases may be completely unacceptable when the values of explanatory variables are shifted.

To conclude this section we can say that the classification rules obtained using the LDA algorithm are characterized by larger classification errors, but their behavior remains stable in the presence of small shifts of the expected values of the explanatory variables. On the other hand, the classification rules obtained using the CDT algorithm perform much better in the case of stable processes. However, when the values of some crucial parameters are changing their performance may drastically deteriorate.

3 Application of a Shewhart Control Chart for Monitoring the Process

The ultimate goal of any SPC procedure is to keep the process at an acceptable level. Even if we observe all items in a process we can still use SPC procedures for monitoring the process quality. For example, we can divide the entire process

Table 12 Control limits of p chart—different classifiers

Dataset	Actual-L	Actual-U	LDA-L	LDA-U	CDT-L	CDT-U
Set 1	0.130	0.394	0.133	0.397	0.093	0.341
Set 2	0.112	0.368	0.205	0.491	0.082	0.324
Set 3	0.120	0.379	0.091	0.337	0.128	0.392
Set 4	0.123	0.384	0.085	0.327	0.171	0.449
Set 5	0.107	0.361	0.151	0.423	0.182	0.462
Set 6	0.123	0.385	0.091	0.337	0.117	0.378
Set 7	0.133	0.397	0.160	0.334	0.122	0.382
Set 8	0.118	0.376	0.088	0.332	0.111	0.370
Set 9	0.123	0.384	0.152	0.424	0.183	0.464
Set 10	0.112	0.368	0.093	0.341	0.062	0.290

into consecutive segments of n elements, and treat these segments as samples for charting purposes. Alternative approaches, such as using a sliding window for monitoring the process, are also possible. In our research we considered two approaches: division of the process into segments of $n = 100$ items considered as samples for charting the Shewhart p-chart, and using a moving average chart (MAV) with a sliding window of $n = 100$ items. In this paper we describe only the first of the both considered methods.

For the construction of Shewhart p-charts we used the data sets consisted of samples of 1,000 items each. These samples represented the Phase 1 sampling period, and were used for estimating control chart parameters for different classification rules derived from the samples of the training data used for the setting of classification rules. The calculated control limits are presented in Table 12.

The columns labeled Actual-L and Actual-U give the values of the lower and upper control limits of the p-chart as if the actual (not predicted!) binary values of the variable of interest were observed. One can notice that they visibly vary depending upon the data set used for the estimation of the fraction nonconforming p. In the remaining columns the control limits have been calculated from the predicted binary values observed from the Phase 1 samples. It has to be noticed that these control limits are quite different for classifiers of the considered types (LDA and CDT), and also vary significantly for different training data sets used for the construction of classifiers. This significant variation stems mainly from the variation of the probabilities of wrong classifications described in the previous section of this paper.

First, let us consider the case of a stable process. Stability in this case means not only the stability of the actual process level, but the stability of probability distributions of the explanatory variables used for classification purposes. In our experiments we have generated 100 runs of the process, and have calculated the average run length $ARL0$. The number of simulation runs is too small to evaluate the accurate values of $ARL0$, but provides a general idea about the performance of the monitoring process. The results of this experiments are summarized in Table 13.

Table 13 Average run length $ARL0$—different classifiers

Dataset	Actual	LDA	CDT
Set 1	535.1	344.7	416.1
Set 2	150.7	251.8	390.0
Set 3	263.4	385.8	681.0
Set 4	372.5	221.8	338.2
Set 5	145.8	196.0	253.1
Set 6	403.8	294.1	269.3
Set 7	326.7	180.3	274.3
Set 8	298.6	448.2	462.2
Set 9	380.6	353.0	366.4
Set 10	178.9	211.6	143.9

Table 14 Average run length ARL—actual (not observed) values

Dataset	Shift B	Shift C	Shift D
Set 1	513.5	42.1	275.9
Set 2	161.7	41.4	58.6
Set 3	281.3	13.3	95.0
Set 4	401.7	23.2	180.0
Set 5	173.4	10.5	59.7
Set 6	409.8	24.3	141.0
Set 7	368.8	45.6	274.9
Set 8	296.3	15.9	115.4
Set 9	423.2	29.2	199.6
Set 10	170.4	10.2	60.4

In the first column of this table we display the values of ARLs as if the values of the actual process were observed. Therefore, the variability observed in Table 13 is not only due to the variability of the characteristics of classifiers, but to the variability of chart's control limits as well. One should note, looking at the first column of Table 13, that even in the case of 1000 items used for the estimation of control limits, the variability of the control charts may be strikingly large.

In Table 14 we give the actual (if the real times to failure were observed) values of the ARLs when the process is deteriorating due to small negative shifts (-0.5σ) of the explanatory variables B, C, and D. We have not presented the data for the explanatory variable A as they are similar to those obtained for B. From Table 9 we know that a small negative shift of B does nor change the actual fraction nonconforming (potentially unreliable) items. This is confirmed by the data presented in the second column of Table 14 where the observed values of ARL are similar to those given in the second column of Table 13. A similar (in terms of standard deviations) shift of C causes severe deterioration of the process, and this deterioration is reflected in the third column of Table 14. The impact of the shift of D is also visible (see the fourth column of this table), but is not so significant as in the case of C. The respective observed values of the ARLs for LDA and CDT classifiers are presented in Table 15. The results presented in this table are really

Table 15 Average run length ARL—different shifts and classifiers

Dataset	ShiftB-LDA	ShiftB-CDT	ShiftC-LDA	ShiftC-CDT	ShiftD-LDA	ShiftD-CDT
Set 1	18.7	467.4	260.3	1.0	75.9	279.7
Set 2	6.3	427.6	4.2	1.0	11.0	143.9
Set 3	17.8	577.8	416.8	1.0	432.6	15.4
Set 4	13.3	309.4	68.5	1.0	95.7	14.9
Set 5	195.6	246.1	74.5	1.0	164.3	17.0
Set 6	16.5	237.3	275.6	1.0	256.6	56.5
Set 7	9.5	288.7	56.6	1.0	103.9	281.6
Set 8	45.8	488.0	491.7	1.0	481.9	537.8
Set 9	33.0	135.8	205.4	367.6	37.3	156.6
Set10	321.1	142.3	566.2	1.0	6.3	176.0

alarming. When we use the LDA classifier we observe unnecessary alarms (small values of the ARL) in the cases of shifts of B, and the lack of necessary alarms (large values of the ARL) in the cases of shifts of C. Moreover, these results are completely unpredictable. For example, the classifier used in the analysis of the Set 10 reacts correctly for the shift of B, but completely wrongly for the shift of C. On the other hand, the classifier used in the analysis of the Set 2 reacts incorrectly for the shift of B, but completely correctly for the shift of C. The CDT classifier performs quite well in the case of shifts of B (large values of the ARL) and C (very small values of the ARL, but with a noticeable exception of Set 9 for which the classification rule does not depend upon the values of this particular parameter). However, in the case of the shift of D the behavior of the CDT classifier seems to be completely unpredictable.

To sum up, from the results presented in Tables 14 and 15 we can conclude that the p control charts based on predicted classification data may trigger alarms when the actual impact of shifts in explanatory variables on actual quality is negligible, and—vice versa—do not trigger alarms when it is needed. This behaviour strongly depends upon the type of a classifier, and its parameters estimated from a training data. In the simulations described in this section we assumed that alarms are triggered by crossing either the lower or the upper control limit. When only the upper control limit of the p-chart is active, the respective values of the ARL are much larger, especially in the case of no-shift or if the shift in the explanatory variable has a small effect on the quality variable of interest.

4 Conclusions

The results presented in this paper are of very preliminary character. They show that in the case of non-normal distributions of characteristics of interest, and non-linear dependencies between observable (explanatory) and not directly observable

(only predicted!) values of processes the properties of control charts designed using the standard methodology may be not satisfactory. The most popular classifiers (statistical one—LDA, and typical for data mining—CDT) used for prediction purposes may not perform well. Moreover, their performance is difficult to predict in advance. Further research is needed with the aim to analyze the impact of the size of training data sets, and the size of the Phase 1 samples, on the characteristics of control charts. Additional research on the possible application of more complicated classifiers is also needed. The results presented in this paper show that the application of modern data mining techniques for SPC purposes, which is strongly advocated by some specialists, is not so promising as it may look like.

References

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction* (2nd edn.). New York: Springer.

Montgomery, D. C. (2011). *Introduction to statistical quality control* (6th edn.). New York: Wiley.

Nelsen, R. B. (2006). *An introduction to copulas* (2nd edn.). New York: Springer.

Noorsana, R., Saghaei, A., & Amiri, A. (2011). *Statistical analysis of profile monitoring.* Hoboken, NJ: Wiley.

Owen, D. N., & Su, Y. H. (1977). Screening based on normal variables. *Technometrics, 19*, 65–68.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning.* Los Altos, CA: Morgan Kaufmann.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd edn.). Amsterdam: Elsevier.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product profiles. *Journal of Quality Technology, 36*, 309–320.

Wang, Y. T., & Huwang, L. (2012). On the monitoring of simple linear Berkson profiles. *Quality and Reliability Engineering International, 28*, 949–965.

Xu, L., Wang, S., Peng, Y., Morgan, J. P., Reynolds Jr., M. R., & Woodall, W. H. (2012). The monitoring of linear profiles with a GLR control chart. *Journal of Quality Technology, 44*, 348–362.

Shewhart's Idea of Predictability and Modern Statistics

Alessandro Di Bucchianico and Edwin R. van den Heuvel

Abstract Shewhart's view on statistical control as presented in his 1931 book is connected to predictability and it seems to be inspired by philosophical theories. At that time, there was no proper statistical framework available when Shewhart implemented his ideas on statistical control. This was not a problem for standard settings for which the original Shewhart control chart was developed, but there are currently several much more complicated situations where the standard tools of Shewhart do not suffice without modification. We will discuss whether current statistical notions like hypothesis testing (both the standard Neyman-Pearson theory and other forms like sequential statistics and equivalence testing), prediction intervals and tolerance intervals can be useful in these other settings. We will also discuss alternative settings of statistical control proposed in the literature including Bayesian settings.

Keywords Charting versus testing • Neyman/Pearson tests • Statistical control • Statistical intervals

1 Introduction

Shewhart's view on statistical control as presented in Shewhart (1931) is connected to predictability. His ideas are deep and are remarkably relevant even today. However, when Shewhart developed his ideas in the 1920s and 1930s, the development of basic statistical theory had just started (e.g., the Neyman-Pearson theory of hypothesis testing was developed in the early 1930s) and it was not until the 1950s that the basic statistical theory of estimation and hypothesis testing was

A. Di Bucchianico (✉)
Department of Mathematics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: a.d.bucchianico@tue.nl

E.R. van den Heuvel
Department of Mathematics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: e.r.v.d.heuvel@tue.nl

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_15

established. So, there was no proper statistical framework available to Shewhart when he implemented his ideas and for that matter he relied on simple statistical tools. This was not a problem for the conventional manufacturing of parts since in those settings the pragmatic approach of the original Shewhart control chart does suffice (as exemplified by, for instance, the view on hypothesis testing by Deming and others quoted in Woodall 2000). Nowadays, however, there are several other application domains in which statistical process control (SPC) is applied with a variety of domain specific goals. Examples of such application domains include batch process manufacturing, health care, environmental surveillance, computer network surveillance, biosurveillance and finance (see, e.g., MacCarthy and Wasusri 2002; Frisén 2009; Ning et al. 2009; Okhrin and Schmid 2007 for more detailed overviews and several explicit examples). Obviously one cannot expect that the standard tools of Shewhart suffice in all such cases. In view of this, we feel it is worthwhile to revisit the ideas of Shewhart. We will investigate to which extent modern statistical theory may be used to extend Shewhart's ideas in the above-mentioned wider range of application domains. In particular, we will discuss whether current statistical notions like hypothesis testing (both the standard Neyman-Pearson theory and the other forms like sequential statistics and equivalence testing), prediction intervals, and tolerance intervals can be applied or implemented. We will also discuss alternative settings of statistical control proposed in the literature (see, e.g., Alwan and Roberts 1988; Andersson et al. 2007; Chakhunashvili and Bergman 2007; Crowder et al. 1997; Di Bucchianico et al. 2004; Frisén and De Maré 1991; Hawkins et al. 2003).

Our discussion in this paper is limited to univariate SPC. Even in this restricted case there are several issues that one has to address properly before moving on to more complicated cases. We also note that the current theory of multivariate SPC has a strong algorithmic flavour, opposed to univariate SPC where statistical modelling is prominent (cf. Breiman 2001).

2 Shewhart's Definitions of Statistical Control

As mentioned in the introduction, Shewhart's view on statistical control as presented in Shewhart (1931) is connected to predictability. On page 6 of Shewhart (1931) it says:

> a phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here it is understood that the prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within given limits

The link with predictability is somewhat lost in current discussions of the notion of statistical control. Indeed, the original Shewhart chart still seems to be the most widely used control chart in industrial practice (see, e.g., the survey

in MacCarthy and Wasusri 2002), and new research papers keep appearing on modifications of this control chart. Therefore, the notion of statistical control is somewhat restricted since it is implicitly or explicitly tied to this specific control chart. Shewhart charts (without additions like runs rules) only use the current data (either a rational subgroup or an individual observation) which makes the decision events independent. Thus the performance of this particular chart can be fully described in terms of type I errors of a statistical test instead of performance measures based on time until detection. Before we continue let us remark that in Shewhart (1939) there is another definition of statistical control in terms of joint distributions that are invariant under permutations. This definition seems to be inspired by a philosophical theory called pragmatism, see Bergman (2009), Mauléon and Bergman (2009) and Wilcox (2004) for modern discussions on this topic. As remarked in Barlow and Irony (1992) and Chakhunashvili and Bergman (2007), this definition boils down to exchangeability, which is a weaker notion than independence. Exchangeability was used by De Finetti in his approach to subjective probability and as such has a definite Bayesian flavour (see Bergman 2009). We will not discuss this definition in the sequel because we feel that it does not apply to all application domains and it is more difficult to translate into practice.

3 Application Scenarios

Before we start our review of existing approaches let us mention some scenarios that could be viewed as being in statistical control, but might cause many false signals in the original Shewhart control chart. In order to focus ideas, every scenario consists of a practical situation together with an appropriate statistical model. The scenarios that we have in mind are:

1. Tool wear: linear regression (see, e.g., Noorossana et al. 2012) or isotonic regression (see, e.g., Chang and Fricker 1999; Kang and Albin 2000)
2. Processes that are sampled at high frequency: time series models (see, e.g., Faltin et al. 1997)
3. Processes that cannot be kept at a setpoint (chemical reactors): time series models (see, e.g., Vander Wiel et al. 1992; Vander Wiel 1996)
4. Multiple random components: random effect models or variance component models (see, e.g., Chakhunashvili and Bergman 2007; Woodall and Thomas 1995; Yashchin 1994)
5. Trend reversals in business cycles: two-state Hidden Markov model (see, e.g., Andersson et al. 2007)
6. Multiple suppliers: mixture of distributions (see, e.g., Chakhunashvili and Bergman 2007)
7. Processes with feedback controllers (integration of SPC and APC/EPC): time series models (see, e.g., Box and Kramer 1992; Göb et al. 2001; Nembhard and Chen 2007)

Of course, combined versions of these scenarios are also possible (see, e.g., Göb 2000). All these scenarios can be interpreted as being in statistical control since the process outcomes can in principle all be predicted, but they do not correspond to the classical notion of stability. Note that nonstationary models may be required for some of these scenarios in order to formulate the notion of statistical control.

4 General Frameworks

As an introduction let us remark that a discussion on control charts should bear in mind that there are three aspects: process model, observations and decision strategy. Often these aspects are confused, since one often defines the concept of being in statistical control in terms of the control chart itself. Reasons for distinguishing between process model and observations include the joint use of control charts for mean and variance (e.g., a Shewhart $\overline{X}$ chart used alongside with an R or S chart) or the use of control charts with variable sampling intervals or variable sampling sizes. In spite of its inherent importance, we will not discuss such monitoring strategies because they do not have an added value for our discussion.

Shewhart clearly distinguished between the economic operation of a process and its predictability. However, as stated in Chakhunashvili and Bergman (2007), Shewhart did not clearly transfer this distinction to a definition of being in statistical control. Processes that are predictable within limits but possibly have special (assignable) causes of variations are said to be "in weak statistical control" in Chakhunashvili and Bergman (2007). Some of our scenarios also appear in that paper to illustrate that we should rethink the basic notion of being in statistical control.

In Frisén and De Maré (1991) a very general framework is described. The state of the process is described as a general stochastic process. In-control situations are a subset of realizations of this stochastic process, while out-of-control situations are seen as the complement of the set of in-control realizations. This framework includes all scenarios except Scenarios 3 and 7, but it is too generic to be seen as a real framework (although Andersson et al. (2007) gave an example of the use of this framework for a specific model). The framework that is behind the control charts based on time series as put forward in Alwan and Roberts (1988) (see especially Sect. 4) is similar, but assumes time series modelling of the underlying stochastic process. It can handle Scenarios 2, 3, 4 and 7 but it cannot handle Scenarios 1, 5 and 6.

The weak statistical control definition of Chakhunashvili and Bergman (2007) extends the stationarity framework and is supposed to include Scenario 3. It has the disadvantage of not being explicit in a mathematical sense. Our contribution is to investigate whether modern statistical techniques and theory can be used to develop actual monitoring strategies that correspond to these notions of being in statistical control.

5 Prediction and Tolerance Intervals

Since Shewhart explicitly mentions prediction in his definition of being in statistical control, it seems logical to involve prediction intervals or prediction regions in making his definition operational. Instead we see widely recommended practices like jointly monitoring the mean μ and the standard deviation σ with the $\overline{X}$ and S for all kinds of situations. Of course, monitoring the variance is sensible in view of process improvement. For example, under Scenario 4 a process could exhibit two sources of variation, e.g. within and between batch variation. The between variation may come from sources that are still in statistical control, but just add to the variability (Roes and Does 1995). The sum of the variation sources may remain approximately constant and thus predictability is guaranteed, but the ratio of the two individual sources may change due to circumstances which would indicate an out of control situation if the parameters would be monitored. The use of prediction intervals is common practice in regression analysis where one constructs such intervals for individual or sets of observations (Hahn and Meeker 1991, especially in engineering and the medical sciences. Note that the control chart for individual observations can be seen in view of predictability as well as in view of monitoring distributional parameters. However, this dual view is lacking in our seven practical scenarios which indicates again the need for a re-evaluation of the concepts of in statistical control.

A related concept to predictability is of course tolerance intervals, that unlike confidence intervals for parameters are defined in terms of coverage of a probability distribution (see, e.g., Guttman 1970; Krishnamoorthy and Mathew 2009). There are two main types of tolerance intervals: guaranteed content tolerance intervals and mean coverage tolerance intervals. The former type contains with a certain confidence level a proportion of the distribution, while the latter contains this proportion on average. It is not widely known that mean coverage tolerance intervals for independent and identically distributed random variables coincide with prediction intervals (a result due to Paulson 1943). Tolerance and prediction intervals have been derived for many statistical models (see, e.g., Patel 1986; Hahn and Meeker 1991) and there is also a generalization of these classical concepts due to Weerahandi that is useful for situations in which traditional intervals cannot be computed (see, e.g., Weerahandi 2003 for a general overview and Krishnamoorthy and Mathew 2009 for tolerance intervals in particular).

However, there is a problem in using these intervals for checking whether a process is in statistical control. Should one check whether only the next observation falls within predictable limits or should one do so for some or all future observations? In the settings of our practical scenarios, the parameters of the statistical model may alter over time, but this may not necessarily severely affect the prediction or tolerance interval. The process remains predictable, but the process does seem to have changed and could also be viewed out of control. Furthermore, simultaneous prediction intervals do exist, but these intervals do not have a natural associated way to judge performance when used for monitoring. In a sense, the variance

parameters are indirectly monitored, since the prediction and tolerance intervals require a precise and unbiased estimate of the variance components.

Checking whether a value lies within a confidence interval of a distribution parameter is related to hypothesis testing. Hypothesis testing has the advantage that it provides natural performance measures. Therefore we turn to hypothesis testing in the next section.

6 Hypothesis Testing

To start our discussion on the role of hypothesis testing let us consider what in the SPC literature is called Phase II (on-line monitoring). The standard implicit framework of the Shewhart control chart is monitoring a process modelled as a sequence of independent, normally distributed random variables or vectors with constant mean and variance. Let us also make the restrictive assumption that the mean and variance are known and that we are only applying an $\overline{X}$ chart. The goal is then to quickly detect an out-of-control situation, which is usually assumed to be a persistent shift of the mean. To be more precise, it is assumed that there is an unknown, deterministic time τ (usually called changepoint) such that the means μ_i at times $i = 1, 2, \ldots$ are equal to a known value μ_0 before time τ and equal to an unknown value $\mu_0 + \delta$ from time τ onwards. This situation can be put in a hypothesis testing framework which is typical of the so-called changepoint literature (see, e.g., Chen and Gupta 2012; Csörgő and Horváth 1997; Gombay 2003; Lai 1995; Yashchin 1997):

$$H_0 : \mu_1 = \ldots = \mu \text{ versus } H_1 : \mu_1 = \ldots = \mu_{\tau-1} = \mu \neq \mu + \delta = \mu_\tau = \mu_{\tau+1} = \ldots \tag{1}$$

This hypothesis framework obviously cannot handle any of the scenarios. Mixtures could be handled if one drops the normality assumption since they are usually multimodal (and thus not normally distributed). Scenario 2 can also be handled if one assumes a parametric time series model and/or a parametric regression model and let the hypothesis refer to one or more parameters. Woodall (2000) reviews statements of several authors on the relation between hypothesis testing and control charting. Several of these quoted authors state that control charting is quite different from hypothesis testing. Some of these criticisms seem to apply to overtheoreticizing standard uses of control charting. As mentioned above, control charting is being applied more and more in new, complex situations that do require careful study of underlying statistical notions. Therefore we choose to concentrate on technical drawbacks of the standard hypothesis framework.

A major drawback is that on-line use of control charts involves repeated decisions so that we should have a sequential point of view. This is especially important for control charts like EWMA and CUSUM charts that use not only current observations but also observations from the past, since these charts have decisions

based on overlapping sets of data. For the standard Shewhart chart this does not play a role, since we then use non-overlapping data (only the current observation or rational subgroup), and decisions can be described without any problems in terms of type I errors. However this is not appropriate even in simple cases like adding run rules to the standard Shewhart chart. This criticism is in fact not related to the use of hypotheses (process model) but to the decision strategy. Statistical techniques like Sequential Probability Ratio Tests and Generalized Likelihood Ratio tests exist to properly deal with the sequential nature of repeated decisions (see, e.g., Lai 1995; Stoumbos and Reynolds Jr 1997).

A similar criticism refers to the asymmetry between the null hypothesis and the alternative hypothesis in the standard Neyman-Pearson approach to hypothesis testing. We will return to this later when we discuss the relation with equivalence testing. The major criticism against the use of hypothesis testing is that the standard hypotheses are simple hypotheses and that as a consequence, any deviation from the single value of the parameter(s) under the null hypothesis has to be interpreted as being out-of-control. This is a much too strict formulation in general and does not match the scenarios of Sect. 3. It is also not a convenient translation of Shewhart's "predictability".

One should not conclude from the criticisms mentioned in previous paragraphs, that hypothesis testing is not an appropriate framework. An important advantage of hypothesis testing (static or sequential) is that it connects with concepts to describe the performance of control charts (see, e.g., Frisén 2007; Kenett and Pollak 2012 for recent discussions that go far beyond the traditional average run length notions). It is thus worthwhile to try to explore more suitable ways of formulating hypotheses. In Does and Schriever (1992) the hypotheses of (1) were slightly generalized by using parameterized cumulative distribution functions so that any parameter (not just the mean) is allowed as well as changes in more than one parameter:

$$H_0 : F_1 = F_2 = \ldots = F \text{ versus } H_1 : F_1 = F_2 = \ldots = F_{\tau-1} = F \neq G = F_\tau = F_{\tau+1} \ldots \tag{2}$$

The case of a mean shift as in (1) is included as the special choice $G(x) = F(x-\delta)$. This type of hypothesis can handle Scenario 6 only. However, it is possible to extend this framework. For instance, in Andersson et al. (2007) the following hypotheses are used:

$$H_0 : \mu_1 \leq \mu_2 \leq \ldots \text{ versus } H_1 : \mu_1 \leq \mu_2 \leq \ldots \mu_{\tau-1} \geq \mu_\tau \geq \mu_{\tau+1} \geq \ldots \tag{3}$$

Although these hypotheses are in terms of means and hence generalize (1) rather than (2), one could use cumulative distribution functions under the stochastic ordering $F \leq G$ if and only if $F(x) \leq G(x)$ for all x. This would lead to hypotheses of the form

$$H_0 : F_1 \leq F_2 \leq \ldots \text{ versus } H_1 : F_1 \leq F_2 \leq \ldots \leq F_{\tau-1} \geq F_\tau \geq F_{\tau+1} \geq \ldots \tag{4}$$

In order to accommodate Scenario 3 one could use composite hypotheses. In Di Bucchianico et al. (2004) GLR procedures are presented for several composite hypotheses. E.g., non-monotone threshold crossing is included by considering the following hypotheses (similar monotone versions can be found in Chang and Fricker (1999)):

$$H_0 : \mu_i \leq \delta \text{ for } i = 1, \ldots \text{ versus } H_1 : \mu_i \leq \delta \text{ for } i < \tau \text{ and } \mu_i > \delta \text{ for } i \geq \tau \tag{5}$$

Of course, two-sided versions are also possible so that one gets rid of the unrealistic assumption of a single true value of a process parameter (e.g., the mean):

$$H_0 : \delta_1 \leq \mu_i \leq \delta_2 \text{ for } i = 1, \ldots \text{ versus } H_1 : \delta_1 \leq \mu_i \leq \delta_2 \text{ for } i < \tau \text{ and } \mu_i < \delta_1 \text{ or } \mu_i > \delta_2 \text{ for } i \geq \tau \tag{6}$$

Note that many variations are possible. For instance, in so-called epidemic alternatives there is only a temporary change of parameters. This is relevant for Scenario 7, since shifts in the underlying process may be seen temporarily in the process output due to the compensating effect of feedback controllers. Scenario 4 seems to provide some difficulty in formulation of hypothesis testing. The reason is that the mean values over time in some cases are random variables themselves due to one particular source of variation that may vary naturally but is not measured itself over time.

These types of hypotheses bring us to another general issue: the so-called equivalence testing (see, e.g., Wellek 2010). Neyman-Pearson's theory of hypothesis testing is asymmetric in the sense that one does not reject the null hypothesis instead of accepting it. This becomes a problem in confirmatory tests in pharmaceutics when one wishes to demonstrate that a new drug has the same effect as another one ("bio-equivalence"). A similar situation occurs when one uses a goodness-of-fit test where one wishes to accept the null hypothesis of a certain assumed distribution to be true. In SPC we do not wish to "accept" or "not reject" a null hypothesis of being in statistical control, but reject the hypothesis that the process is out-of-control and continue sampling. The essence of equivalence testing is to quantify a range of parameter values that may be considered (almost) equivalent processes. Whenever the process parameter remains within this range the process is considered under control. This view is more natural, since small variations in the model parameters may not severely alter the range of individual outcomes of a process and therefore the process remain predictable. Equivalence has been introduced essentially by quality improvement programs like Six Sigma. Indeed, in these type of programs the range of acceptable values for the process mean was set at approximately 1.5 times the short-term standard deviation, although these programs did not present this flexibility in terms of equivalence.

As pointed out in Wellek (2010), one can adapt the Neyman-Pearson framework and its related likelihood ratio tests to accommodate this. One does therefore not need a new statistical theory, but it does yield different decisions than applying standard statistical theory. This also holds for the hypotheses discussed in Di Bucchianico et al. (2004) and Woodall (1985): the methods still work if one interchanges them and adapt the likelihood ratios accordingly. Note that conventional sequential testing theory also requires one to choose between two alternative hypotheses or continue collecting data (cf. Lai 1995; Lai 2001).

All of the above is based on the frequentist approach to hypothesis testing which assumes true, but unknown values of process parameters. Bayesian methods assume distributions on parameters and hence are more flexible than hypotheses like (6) since they also allow randomness in the unknown parameters involved like the changepoint parameter τ or the shift parameter δ in (1). The randomness is formulated as a prior distribution on the parameters involved, while the alarm is triggered using a rule based on the a posteriori distribution. The idea of using a Bayesian framework goes back to at least the 1960s (see, e.g., the nice overview of Shiryaev (2010) on the origins and mathematical background of the Shiryaev-Roberts procedure). Early overviews of Bayesian procedures for changepoint problems and statistical process control can be found in Zacks (1982) and Zacks (1983). The development of the Markov Chain Monte Carlo methods like the Metropolis and the Gibbs samplers have made the practical use of Bayesian statistics feasible (see, e.g., Carlin et al. 1992; Colosimo and Del Castillo 2010 for extensive modern discussions of Bayesian approaches to changepoint problems and statistical process control that include both computational issues and modelling issues, e.g., multiple changepoints). These developments give an extra dimension to the discussion in Bergman (2009) on the connection of Bayesian statistics with conceptual pragmatism, the philosophy that has inspired Shewhart to develop his ideas on quality control.

The hypothesis testing framework described above is strongly related to outlier detection in a set of independent observations. In this research area hypothesis testing is typically referred to as discordance testing (Barnett and Lewis 1994). In the above formulations of process control the null hypothesis is formulated as $H_0 : X_i \sim F, \forall i \in \{1, 2, \ldots, n\}$, while the alternative hypothesis is $H_1 : \exists j : X_j \sim G, G \neq F$. In outlier detection the observation X_j will be called contaminated, but it is unknown which observation would be from the alternative distribution G. Therefore, the alternative hypothesis for discordant testing is, for instance, reformulated as: $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$ belong to F and $X_{(n)}$ belong to G for a single outlier in the right tail, with $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ being the order statistics. Under normality assumptions the difference with the alternative hypothesis in (1) is that the process has changed indefinitely, while in outlier detection only one observation may have slipped into a change in mean, which is also referred to as the slippage hypothesis. Depending on the type of assumptions of the distributions F and G and the imposed characteristics of the test statistic, many outlier detection tests have been developed in the past (Barnett and Lewis 1994). We believe that the outlier detection hypothesis is strongly related to the concept of predictability, since

an outlier would indicate that the process is not within predictable limits anymore. It relates to an out-of-control signal, similar to the one used in Shewhart's control charts without the additional run rules, but more formally, it can be shown that outlier detection fits within the concept of tolerance intervals (Hawkins 1980).

7 Conclusion

Shewhart defined being in statistical control in terms of a process being predictable. We reviewed his concept in view of the major developments in statistical methodology that have taken place since then. Several non-standard scenarios in which SPC are being used currently have been presented. These scenarios go beyond the simple manufacturing scenario for which Shewhart developed his methods. Prediction intervals and the related concept of tolerance intervals have been developed for several statistical models, but they lack a natural way to assess the performance of using them to monitor processes. The related concept of hypothesis testing is much better suited for this. We discussed several useful extensions of the standard setup with simple hypotheses. These extensions can accommodate the non-standard scenarios that we presented and may also take away the criticism to using hypothesis testing as the basic statistical methodology underlying the use of control charts. More work is needed here to obtain a general framework.

Bayesian methods offer additional flexibility in modelling uncertainty and may thus yield more realistic approaches for practical use of control charts. These methods have been relatively unexplored in the field of statistical process control. An additional advantage of Bayesian methods is that they seem to match well with the philosophical theory (pragmatism) that underlies the ideas of Shewhart. This is also an area that is worthwhile to explore.

Acknowledgements We would like to thank the reviewer of our paper for his detailed reading of our paper. His comments led to several clarifications in our paper.

References

Alwan, L. C., & Roberts, H. V. (1988). Time-series modeling for statistical process control. *Journal of Business and Economic Statistics*, *6*, 87–95.

Andersson, E., Bock, D., & Frisén, M. (2007). Some statistical aspects of methods for detection of turning points in business cycles. *Journal of Applied Statistics*, *33*(3), 257–278.

Barlow, R. E., & Irony, T. Z. (1992). Foundations of statistical quality control. *Lecture Notes-Monograph Series*, *17*, 99–112.

Barnett, V., & Lewis, T.: *Outliers in statistical data* 3rd edn. New York: Wiley

Bergman, B. (2009). Conceptualistic pragmatism: a framework for Bayesian analysis? *IIE Transactions 41*(1), 86–93.

Box, G. E. P., & Kramer, T. (1992). Statistical process monitoring and feedback adjustment—a discussion. *Technometrics*, *34*, 251–267.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Journal of Royal Statistical Society C*, *41*(2), 389–405.

Chakhunashvili, A., & Bergman, B. (2007). In weak statistical control? *International Journal of Six Sigma Computer Advantage*, *3*(1), 91–102.

Chang, J. T., & Fricker Jr., R. D. (1999). Detecting when a monotonically increasing mean has crossed a threshold. *Journal of Quality Technology 31*, 217–234.

Chen, J., & Gupta, A. K. (2012). *Parametric statistical change point analysis* 2nd edn. Boston: Birkhäuser.

Colosimo, B. M., & Del Castillo, E. (Eds.) (2010). *Bayesian Process Monitoring, Control and Optimization*. Boca Raton: Chapman and Hall/CRC,

Crowder, S. V., Hawkins, D. M., Reynolds Jr. M. R., & Yashchin, E. (1997). Process control and statistical inference. *Journal of Quality Technology*, *29*, 134–139.

Csörgő, M., & Horváth, L. (1997). *Limit theorems in change-point analysis*. Chichester: Wiley

Di Bucchianico, A., Hušková, M., Kláštereckỳ, P., & Van Zwet, W. R. (2004). Performance of control charts for specific alternative hypotheses. In: J. Antoch, (Ed.), *COMPSTAT 2004 Symposium* (pp. 903–910). Heidelberg: Physica Verlag.

Does, R. J. M. M., & Schriever, B. F. (1992). Variables control chart limits and tests for special causes. *Statistica Neerlandica*, *46*, 229–245.

Faltin, F. W., Mastrangelo, C. M., Runger, G. C., & Ryan, T. P. (1997). Considerations in the monitoring of autocorrelated and independent data. *Journal of Quality Technology*, *29*(2), 131–133.

Frisén, M. (2007). Properties and use of the Shewhart method and its followers. *Sequential Analysis*, *26*(2), 171–193.

Frisén, M. (2009). Optimal sequential surveillance for finance, public health, and other areas. *Sequential Analysis*, *28*(3), 310–337.

Frisén, M., & De Maré, J. (1991). Optimal surveillance. *Biometrika*, *78*, 271–280.

Göb, R. (2000). Shewhart charts for the detection of linear trends and of shifts in an autoregression model. *International Journal of Reliability Quality and Safety Engineering*, 7, 309–330.

Göb, R., Del Castillo, E., & Ratz, M. (2001). Run length comparisons of Shewhart charts and most powerful test charts for the detection of trends and shifts. *Communication in Statistics Simulation Computation*, *30*(2), 355–376.

Gombay, E. (2003). Sequential change-point detection and estimation. *Sequential Analysis*, *22*(3), 203–222.

Guttman, I. (1970). *Statistical tolerance regions: classical and Bayesian*. London: Charles Griffin

Hahn, G. J., & Meeker, W. Q. (1991). *Statistical intervals: a guide for practitioners*. New York: Wiley.

Hawkins, D. M. (1980). *Identification of outliers*. London: Chapman and Hall.

Hawkins, D. M., Qiu, P., & Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of Quality Technology*, *35*(4), 355–366.

Kang, L., & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, *32*, 418–426.

Kenett, R., & Pollak, M. (2012). On assessing the performance of sequential procedures for detecting a change. *Quality And Reliability Engineering International*, *28*(5), 500–507.

Krishnamoorthy, K., & Mathew, T. (2009). *Statistical tolerance regions: theory, applications, and computation*. New York: Wiley

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, Series B*, *57*, 613–658.

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, *11*, 303–350.

MacCarthy, B. L., & Wasusri, T. (2002). A review of nonstandard applications of statistical process control (SPC) charts. *International Journal of Quality and Reliability Management*, *19*(3), 295–320.

Mauléon, C., & Bergman, B. (2009). Exploring the epistemological origins of Shewhart's and Deming's theory of quality: Influences from CI Lewis' conceptualistic pragmatism. *International Journal Quality and Service Science*, *1*(2), 160–171.

Nembhard, H. B., & Chen, S. (2007). Cuscore control charts for generalized feedback-control systems. *Quality and Reliability Engineering International*, *23*(4), 483–502.

Ning, X., Shang, Y., & Tsung, F. (2009). Statistical process control techniques for service processes: a review. In *6th International Conference on Service Systems and Service Management (ICSSSM'09)* (pp. 927–931). IEEE.

Noorossana, R., Saghaei, A., & Amiri, A. (2012). *Statistical analysis of profile monitoring*. New York: Wiley

Okhrin, Y., & Schmid, W. (2007). Surveillance of univariate and multivariate linear time series. In: M. Frisén (Ed.), *Financial surveillance* (pp. 115–152). New York: Wiley.

Patel, J. K. (1986). Tolerance limits-a review. *Communications in Statistics: A Theory and Methods*, 15, 2719–2762.

Paulson, E. (1943). A note on control limits. *The Annals of Mathematical Statistics*, *14*, 90–93.

Roes, K. B., & Does, R. J. M. M. (1995). Shewhart-type Charts for Nonstandard Situations. *Technometrics*, *37*(1), 15–24.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. London: Macmillan

Shewhart, W. A. (1939). *Statistical method : from the viewpoint of quality control*. Washington D.C.: Graduate School of the Department of Agriculture

Shiryaev, A. N. (2010). Quickest detection problems: fifty years later. *Sequential Analysis*, *29*(4), 345–385.

Stoumbos, Z. G., & Reynolds Jr., M. R. (1997). Control charts applying a sequential test at fixed sampling intervals. *Journal of Quality Technology*, *29*(1), 21–40.

Vander Wiel, S. A. (1996). Monitoring process that wander using IMA models. *Technometrics*, *38*(2), 139–151.

Vander Wiel, S. A., Tucker, W. T., Faltin, F. W., & Doganaksoy, N. (1992). Algorithmic statistical process control: Concepts and an application. *Technometrics*, *34*, 286–297.

Weerahandi, S. (2003). *Exact statistical methods for data analysis*. New York: Springer.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority* 2nd edn. Boca Raton: CRC Press.

Wilcox, M. (2004). Prediction and pragmatism in Shewhart theory of statistical control. *Management Decision*, *42*(1), 152–165.

Woodall, W. H. (1985). The statistical design of quality control charts. *The Statistician*, *34*(2), 155–160.

Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, *32*, 341–378.

Woodall, W. H., & Thomas, E. V. (1995). Statistical process control with several components of common cause variability. *IIE Transaction*, 27(6), 757–764.

Yashchin, E. (1994). Monitoring variance components. *Technometrics*, *36*(4), 379–393.

Yashchin, E. (1997). Change-point models in industrial applications. *Nonlinear Analysis: Theory, Methods & Applications*, *30*(7), 3997–4006.

Zacks, S. (1982). Classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures. *Statistique et Analyse Des Données*, *7*(1), 48–81.

Zacks, S. (1983). Survey of classical and Bayesian approaches to the change-point problem: Fixed sample and sequential procedures of testing and estimation. In: M. H. Rivzi, J. Rustagi, & D. Siegmund (Eds.), *Recent advances in statistics. Papers in honor of Herman Chernoff's 60th birthday* (pp. 245–269). New York: Academic Press

Part II
Acceptance Sampling

Sampling Inspection by Variables with an Additional Acceptance Criterion

Peter-Th. Wilrich

Abstract We deal with sampling inspection by variables, i.e. acceptance sampling procedures wherein the acceptability of a lot is statistically established from the measurement results of a specified continuous variable X obtained at the items in a sample from the lot. An item is qualified as nonconforming if its measured quality characteristic x is larger than a defined upper specification limit U. We accept the lot if $\bar{x} + k\sigma \leq U$ or $\bar{x} + ks \leq U$ in the case of known or unknown lot standard deviation σ, respectively, where $\bar{x}$ and s are mean and standard deviation of a sample of size n drawn at random from the lot and k is an acceptance constant given as a parameter of the sampling plan.

In some cases the acceptance procedure is extended by an additional limit $U^\star = U + \Delta$ with $\Delta \in \mathbb{R}$ that must not be exceeded by any of the measurements $x_1, x_2, \ldots, x_n$, i.e. for acceptance of the lot the largest measurement result $x_{(n)} = \max(x_1, x_2, \ldots, x_n)$ must be less or equal to $U^\star$, $x_{(n)} \leq U^\star = U + \Delta$. Of course, with this additional requirement for acceptance the probability of acceptance of the lot is smaller than without it for each fraction p of nonconforming items in the lot.

Such extended sampling plans are, e.g., used for the evaluation of bacterial contamination in foods, the amount of active ingredient used in formulating drug products and the strength of concrete.

The OC function of these extended sampling plans for inspection by variables is derived and the advantages/disadvantages in comparison with unextended sampling plans are discussed. It turns out that especially in the case of "known" σ plans the extended sampling plans protect against a true standard deviation that is larger than the value being used in the acceptance criterion $\bar{x} + k\sigma \leq U$.

Keywords Content uniformity test • Inspection by variables • Sampling inspection • Three-class sampling • Variables sampling

P.-Th. Wilrich (✉)
Institut für Statistik und Ökonometrie, Freie Universität Berlin, Garystrasse 21, D-14195 Berlin, Germany
e-mail: wilrich@wiwiss.fu-berlin.de

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_16

1 Introduction

We deal with sampling inspection by variables, i.e. acceptance sampling procedures wherein the acceptability of a lot is statistically established from the measurement results of a specified continuous variable X obtained at the items in a sample from the lot. An item is qualified as nonconforming when its measured quality characteristic x is larger than a defined upper specification limit U.

Under the assumption that the quality characteristic X is normally distributed with mean μ and standard deviation σ the fraction of nonconforming items in the lot is

$$p = P(X > U) = 1 - P(X \le U) = 1 - P\left(\frac{X-\mu}{\sigma} \le \frac{U-\mu}{\sigma}\right)$$
$$= 1 - P\left(Z \le \frac{U-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{U-\mu}{\sigma}\right) \tag{1}$$

where Z is the standardized normal variable and $\Phi(\cdot)$ is the cumulative distribution function of the standardized normal distribution. Hence, $(U-\mu)/\sigma$ is equal to the $(1-p)$-quantile z_{1-p} of the standardized normal distribution,

$$\frac{U-\mu}{\sigma} = z_{1-p}. \tag{2}$$

A sample of n items is randomly drawn from the lot and the quality characteristic X is measured at each of the sample items. The sample average

$$\bar{x} = \sum_{i=1}^{n} x_i \tag{3}$$

and (eventually) the sample standard deviation

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4}$$

of the measurements $x_1, x_2, \ldots, x_n$ are used for the acceptance decision.

In the case of known standard deviation σ we accept the lot if

$$\bar{x} + k\sigma \le U; \tag{5}$$

the operating characteristic function (OC), i.e. the probability of acceptance of a lot as a function of the fraction of nonconforming items in the lot, p, is

$$P_A(p) = P(\bar{X} + k\sigma \le U) = \Phi(\sqrt{n}(z_{1-p} - k)). \tag{6}$$

In the case of unknown σ we accept the lot if

$$\bar{x} + ks \leq U; \tag{7}$$

the OC is

$$P_A(p) = P(\bar{X} + kS \leq U) = 1 - F_{T_{n-1,\delta}}(k\sqrt{n}) \tag{8}$$

where $F_{n-1,\delta}(\cdot)$ is the cumulative distribution function of the noncentral t-distribution with $f = n - 1$ degrees of freedom and noncentrality parameter $\delta = \sqrt{n}z_{1-p}$; see Schilling and Neubauer (2009). We denote such an acceptance procedure as a single sampling plan (n, k) for inspection by variables.

In some cases the acceptance procedure is extended by an additional limit

$$U^\star = U + \Delta \tag{9}$$

with $\Delta \in \mathbb{R}$ that must not be exceeded by any of the measurements $x_1, x_2, \ldots, x_n$, i.e. for acceptance of the lot the largest measurement result $x_{(n)} = \max(x_1, x_2, \ldots, x_n)$ must be less or equal to $U^\star$,

$$x_{(n)} \leq U^\star = U + \Delta. \tag{10}$$

A lot is accepted if the requirements (5) or (7), respectively, and (10) are fulfilled.[1] Of course, with this additional requirement for acceptance the probability of acceptance of the lot, $P_A(p)$, is smaller than without it for each fraction p of nonconforming items in the lot.

In Bray et al. (1973) such extended sampling plans were introduced as Three-class sampling plans and their application concerning hazardous substances in foods or the amount of active ingredient used in formulating drug products were discussed. The quality characteristic X is the concentration for which two specification limits U and $U^\star > U$ are defined. By this definition the items in the lot are classified into three classes: conforming items ($x \leq U$), marginally conforming items ($U < x \leq U^\star$) and totally nonconforming ($x > U^\star$) items. The application of Three-class sampling plans for inspection by attributes for the evaluation of bacterial contamination have been discussed in Hildebrandt et al. (1995), Milchverordnung (1995), Dahms and Hildebrandt (1998), Legan et al. (2001), ICMSF (2002), Dahms (2003, 2004), and Wilrich and Weiss (2011). In Wilrich and Weiss (2011) it is proposed to use Three-class sampling plans for inspection by variables instead of Three-class sampling plans for inspection by attributes in order to increase the discrimination power and at the same time maintain the advantages of the Three-class sampling plan.

[1] If a lower specification limit L is defined, the additional limit is $L^\star = L - \Delta$. The lot is accepted if $\bar{x} - k\sigma \geq L$ or $\bar{x} - ks \geq L$, respectively, and $x_{(1)} = \min(x_1, \ldots, x_n) \geq L^\star$

An interesting modification of an extended sampling plan for inspection by variables is the Content uniformity test of the European Pharmacopoeia (and identically of the United States Pharmacopeia and of the Japanese Pharmacopeia) EDQM (2010). It starts with the measurement of the doses of a first sample of $n = 10$ dosage units (e.g., tablets), i.e. for each dosage unit the concentration and the weight are measured and multiplied. These results are expressed as percentage with the claimed content = 100 % . Mean $\bar{x}$ and standard deviation s of these doses are calculated and the acceptance value is obtained as $AV = \max(\min(\bar{x}, 101.5), 98.5) + k \cdot s$ with $k = 2.4$. If AV is not larger than 15 (percent) the content uniformity test is passed. If $AV > 15$ a second sample of $n = 20$ dosage measurements is obtained and combined with the first sample. With $\bar{x}$ and s of the $n = 30$ doses the acceptance value is obtained as $AV = \max(\min(\bar{x}, 101.5), 98.5) + k \cdot s$ with $k = 2.0$. If AV is not larger than 15 (percent) the content uniformity test is passed; otherwise not. In addition to this acceptance criterion it is required for acceptance that all measured doses fall within the interval $100 \pm 25\,\%$.

In CEB (1978) extended sampling plans are introduced for the evaluation of the strength of concrete, and they are now standardized in EN (2000). Taerwe (1988) presents OC curves of these plans obtained by simulation.

The upper limit U is a specification limit because it specifies whether an item is conforming ($X \leq U$) or not. However, at the same time it is a decision limit because the lot is rejected if $\bar{x} + k\sigma > U$ or $\bar{x} + ks > U$. In the application in microbiology the additional limit $U^\star = U + \Delta$ is also a specification limit because it classifies the items of the lot into three classes (conforming, marginally conforming and nonconforming items), and it is a decision limit because the lot is rejected if $x_{(n)} > U^\star$. However, in many applications as in the inspection of concrete this additional limit $U^\star$ is only a decision limit because it does not specify the conformance/nonconformance of items but is only used as an additional acceptance criterion. In this case the additional limit $U^\star$ can be smaller than the specification limit U, i.e. $\Delta < 0$.

The OC function of the extended sampling plans for known standard deviation σ is derived in Sect. 2 and for unknown σ in Sect. 3. In Sect. 4 we investigate how the extended sampling plans protect against a standard deviation being larger than the assumed value, and in Sect. 5 how they protect against deviations from the assumed normal distribution. Section 6 summarizes the results.

2 The OC Function of the Extended Sampling Plans with Known Standard Deviation σ

The probability of acceptance of a lot with the fraction p of nonconforming items is

$$P_A(p) = P(\bar{X} + k\sigma \leq U \wedge X_{(n)} \leq U + \Delta \mid p). \tag{11}$$

Since $\bar{X} \leq X_{(n)}$ the acceptance criterion $X_{(n)} \leq U + \Delta$ implies $\bar{X} \leq U + \Delta$. Hence, the acceptance criterion for $\bar{X}$ is $\bar{X} \leq U - k\sigma \wedge \bar{X} \leq U + \Delta$ or $\bar{X} \leq \min(U - k\sigma, U + \Delta)$. This is only relevant if $\Delta < -k\sigma$, i.e. if $\delta < -k$.

$X_{(n)}$ and $\bar{X}$ are not independent and hence, we look at the random variable $X_{(n)} - \bar{X}$ which is independent of $\bar{X}$. The random variables $X_i - \bar{X};\ i = 1, \ldots, n$ are normally distributed with mean 0 and standard deviation $\sigma\sqrt{(n-1)/n}$ and hence, their cumulative distribution function is

$$F_{X_i - \bar{X}}(x) = \Phi\left(\frac{x}{\sigma\sqrt{(n-1)/n}}\right); \tag{12}$$

however, the distribution of their maximum $X_{(n)} - \bar{X}$ is difficult to obtain because the random variables $X_i - \bar{X}$ and $X_j - \bar{X}; i, j = 1, \ldots, n; i \neq j$ are correlated (with the small correlation coefficient $-1/n$). In McKay (1935) the probability density function of $(X_{(n)} - \bar{X})/\sigma$ for a given n was derived as a function of the cumulative distribution function for $n-1$ and hence, principally, the distribution can be obtained by recursion. However, for $5 \leq n \leq 100$ the distribution of $(X_{(n)} - \bar{X})/\sigma$ (that does not depend on μ and σ) can be approximated by a logarithmic normal distribution. For each n we have performed 10^5 simulation runs and obtained estimates of the mean and the standard deviation of this distribution. The dependence of these means on n can be approximately described by the function

$$\mu_n = \exp(-9.26n^{-0.95}) \tag{13}$$

and that of the standard deviations by

$$\sigma_n = 0.71n^{-0.33} \tag{14}$$

so that we get the cumulative distribution function of $(X_{(n)} - \bar{X})/\sigma$ as approximately

$$P\left(\frac{X_{(n)} - \bar{X}}{\sigma} \leq x\right) \approx \Phi\left(\frac{\log(x) - \mu_n}{\sigma_n}\right). \tag{15}$$

Now we write (11) as

$$\begin{aligned} P_A(p) &= \int_{-\infty}^{U+\min(\Delta,-k\sigma)} f_{\bar{X}}(\bar{x}) F_{(X_{(n)}-\bar{X})/\sigma}\left(\frac{U + \Delta - \bar{x}}{\sigma}\right) d\bar{x} \\ &= \int_{-\infty}^{U+\min(\Delta,-k\sigma)} f_{\bar{X}}(\bar{x}) F_{(X_{(n)}-\bar{X})/\sigma}\left(z_{1-p} + \frac{\Delta}{\sigma} - \frac{z}{\sqrt{n}}\right) d\bar{x} \end{aligned} \tag{16}$$

and with (15), $z = \sqrt{n}(\bar{x} - \mu)/\sigma$, $f_{\bar{X}}(\bar{x})d\bar{x} = \phi(z)dz$ and the integration limit $U + \min(\Delta, -k\sigma)$ for $\bar{x}$ becoming $\sqrt{n}(U + \min(\Delta, -k\sigma) - \mu)/\sigma = \sqrt{n}(z_{1-p} + \min(\Delta/\sigma, -k))$ for z,

$$P_A(p) \approx \int_{-\infty}^{\sqrt{n}(z_{1-p}+\min(\Delta/\sigma,-k))} \phi(z) \cdot \Phi\left(\frac{\log(z_{1-p} + \Delta/\sigma - z/\sqrt{n}) - \mu_n}{\sigma_n}\right) dz. \tag{17}$$

The probability of acceptance of a lot does not only depend on the two parameters of the sampling plan, n and k, but also on the distance Δ between the specification limit U and the additional limit $U^\star$ divided by σ. We denote the ratio

$$\delta = \frac{\Delta}{\sigma} \tag{18}$$

as *standardized limit distance*. For $\Delta \to \infty$, i.e. if the additional limit $U^\star$ becomes larger, $P_A(p)$ tends towards that of the unextended sampling plan, given in (6); for $\Delta \to 0$, i.e. if the additional limit $U^\star$ is not much larger than U, $P_A(p)$ tends towards the probability of $X_{(n)}$ not exceeding the additional limit, $P(X_{(n)} \leq U + \Delta)$.

Figure 1 shows OC curves of the sampling plan ($n = 9, k = 1.83$) for known σ. The black curve is that for the unextended plan ($\delta \to \infty$), the red curve is that for the case of the additional limit equal to the specification limit ($\delta = 0$). OC curves for $\delta > 0$ lie between these two curves; for $\delta < 0$, the OC curves lie below the curve for $\delta = 0$. Apparently, the additional criterion for acceptance decreases the probability

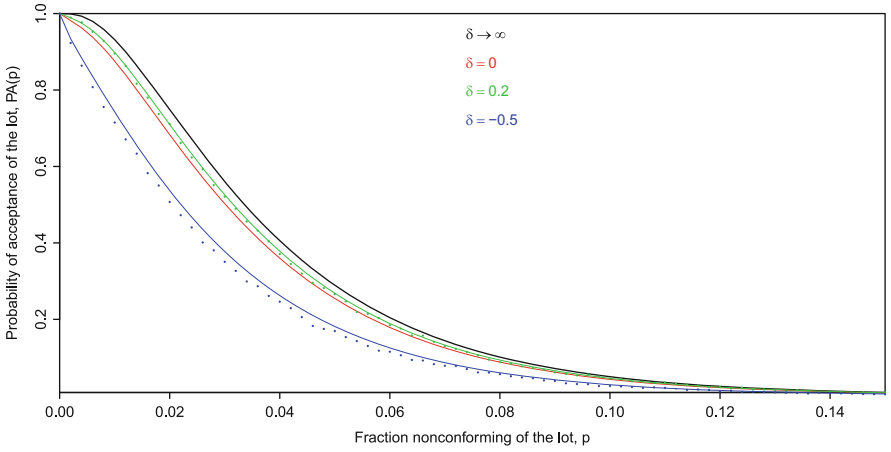


Fig. 1 The OC curves of the sampling plan $(n, k) = (9, 1.83)$ for inspection by variables with known σ for various standardized limit distances δ. Theoretical OC curves are presented as *solid lines*, simulation results as points (each *point* represents the average of 10^4 simulation runs)

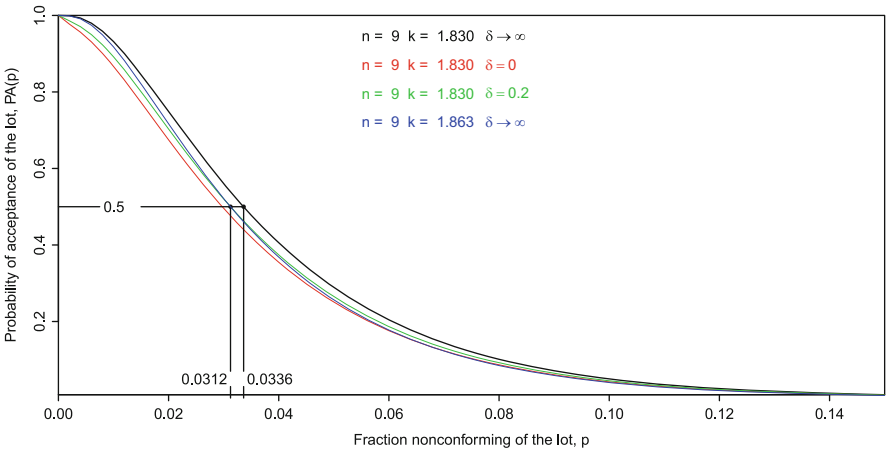


Fig. 2 The OC curves of the sampling plan $(n, k) = (9, 1.83)$ for inspection by variables with known σ for $\delta \to \infty$ (unextended, plan 1, *black line*) and $\delta = 0.2$ (extended, plan 2, *green line*) and the unextended plan that matches the OC of plan 2 at the indifference point $(p_0^\star = 0.0312, 0.5)$. The latter one (plan 3, *blue line*, $k = 1.858$) has a steeper OC than plan 2

of acceptance for each fraction p of nonconforming items in the lot. That this is not an advantage can be further demonstrated by the following comparison presented in Fig. 2: It repeats the OCs of the sampling plans for $(n = 9, k = 1.83)$. The OC of the unextended plan (plan 1, black line) has its indifference point (probability of acceptance equal to 0.5) when in the OC function according to (6) $z_{1-p_0} = k = 1.83$ and hence, $p_0 = 1 - \Phi(k) = 0.0336$. The extended plan with the standard limit distance $\delta = 0.2$ (plan 2, green line) has its indifference point at $p_0^\star = 0.0312$. The acceptance constant of the sampling plan 3 (blue line) without an additional decision limit that coincides in the point $(p_0^\star, 0.5)$ with plan 2 is $k^\star = 1 - z_{1-p_0^\star} = 1.858$. The OC of this sampling plan $(n, k^\star)$ is steeper than that of plan 2: Increasing the acceptance constant of our plan from $k = 1.83$ to $k = 1.857$ is a better strategy than using the extended plan with $\delta = 0.2$ because, compared with plan 2, it improves the discrimination power between lots with small and lots with large fractions of nonconforming items. This is not surprising, because, given the sample average $\bar{x}$ the distribution of $X_{(n)} - \bar{X}$ does not depend on μ and hence, not on p, i.e. it is uninformative concerning the fraction of nonconforming items in the lot. However, in Sects. 4 and 5 we show that the additional acceptance criterion partly protects against the case that the true standard deviation σ_{true} is larger than the "known" standard deviation σ used in the acceptance criterion (5) and against violations of the normality assumption, especially against outliers.

3 The OC Function of the Extended Sampling Plans with Unknown Standard Deviation σ

The probability of acceptance of a lot with the fraction p of nonconforming items is

$$P_A(p) = P(\bar{X} + kS \le U \wedge X_{(n)} \le U + \Delta \mid p). \tag{19}$$

$X_{(n)}$ is not independent of $\bar{X}$ and S and hence, we look at the random variable $(X_{(n)} - \bar{X})/S$ which is independent of $\bar{X}$ and S. In analogy to Sect. 3 the distribution of $(X_{(n)} - \bar{X})/S$ can be approximated by a logarithmic normal distribution. For each n between $n = 5$ and $n = 100$ we have performed 10^5 simulation runs and obtained estimates of the mean and the standard deviation of this distribution. The dependence of these means on n can be approximately described by the function

$$\mu_n = \exp(-6.48n^{-0.87}) \tag{20}$$

and that of the standard deviations by

$$\sigma_n = 0.30n^{-0.15} \tag{21}$$

so that we get the cumulative distribution function of $(X_{(n)} - \bar{X})/S$ as approximately

$$P\left(\frac{X_{(n)} - \bar{X}}{S} \le x\right) \approx \Phi\left(\frac{\log(x) - \mu_n}{\sigma_n}\right). \tag{22}$$

Now we write (19) as

$$\begin{aligned} &P_A(p) \qquad\qquad (23)\\ &= \int_{\bar{x}=-\infty}^{U} f_{\bar{X}}(\bar{x}) \left[\int_{s=0}^{(U-\bar{x})/k} f_S(s) F_{(X_{(n)}-\bar{X})/S}\left(\frac{U + \Delta - \bar{x}}{s}\right) ds\right] d\bar{x} \\ &= \int_{\bar{x}=-\infty}^{U} f_{\bar{X}}(\bar{x}) \left[\int_{s=0}^{(U-\bar{x})/k} f_S(s) F_{(X_{(n)}-\bar{X})/S}\left(\frac{z_{1-p} + \Delta/\sigma - (\bar{x} - \mu)/\sigma}{s/\sigma}\right) ds\right] d\bar{x}. \end{aligned}$$

With $z = \sqrt{n}(\bar{x} - \mu)/\sigma$, $f_{\bar{X}}(\bar{x})d\bar{x} = \phi(z)dz$ and the integration limit U for $\bar{x}$ becoming $\sqrt{n}(U - \mu)/\sigma = \sqrt{n}z_{1-p}$ for z, $y = (n-1)s^2/\sigma^2$, $f_S(s)ds = f_{\chi^2_\nu}(y)dy$ and the integration limit $(U - \bar{x})/k$ for s becoming $(n-1)(z_{1-p} - z/\sqrt{n})^2/k^2$ for y we get

$$\begin{aligned} P_A(p) = \int_{z=-\infty}^{\sqrt{n}z_{1-p}} \phi(z) \Bigg[\int_{y=0}^{(n-1)(z_{1-p}-z/\sqrt{n})^2/k^2} & f_{\chi^2_\nu}(y) \qquad (24)\\ & \cdot F_{(X_{(n)}-\bar{X})/S}\left(\frac{\sqrt{n-1}(z_{1-p} + \Delta/\sigma - z/\sqrt{n})}{\sqrt{y}}\right) dy\Bigg] dz. \end{aligned}$$

With (22) we finally have

$$P_A(p) \approx \int_{z=-\infty}^{\sqrt{n}z_{1-p}} \phi(z) \left[\int_{y=0}^{(n-1)(z_{1-p}-z/\sqrt{n})^2/k^2} f_{\chi^2_\nu}(y) \right. \tag{25}$$
$$\left. \cdot \Phi\left(\frac{\log(\sqrt{n-1}(z_{1-p} + \Delta/\sigma - z/\sqrt{n})/\sqrt{y}) - \mu_n}{\sigma_n} \right) dy \right] dz.$$

As in the case of known σ the probability of acceptance of a lot does not only depend on the two parameters of the sampling plan, n and k, but also on the standardized limit distance $\delta = \Delta/\sigma$.

For $\Delta \to \infty$, i.e. if the additional limit $U^\star$ becomes larger, $P_A(p)$ tends towards that of the unextended sampling plan, given in (8); for $\Delta \to 0$, i.e. if the additional limit $U^\star$ is not much larger than U, $P_A(p)$ tends towards the probability of $X_{(n)}$ not exceeding the additional limit, $P(X_{(n)} \le U + \Delta)$.

Figure 3 shows OC curves for the sampling plan ($n = 25, k = 1.83$) . The black curve is that for the unextended plan ($\delta \to \infty$); its OC curve is almost identical to that of the unextended plan ($n = 9, k = 1.83$) of Fig. 1 for known σ. The red curve is that for the case of the additional limit being equal to the specification limit ($\delta = 0$). OC curves for $\delta > 0$ lie between these two curves; for $\delta < 0$, the OC curves lie below the curve for $\delta = 0$. As for the case of known σ the additional criterion for acceptance decreases the probability of acceptance for each fraction p of nonconforming items in the lot. In analogy to Fig. 2, Fig. 4 demonstrates

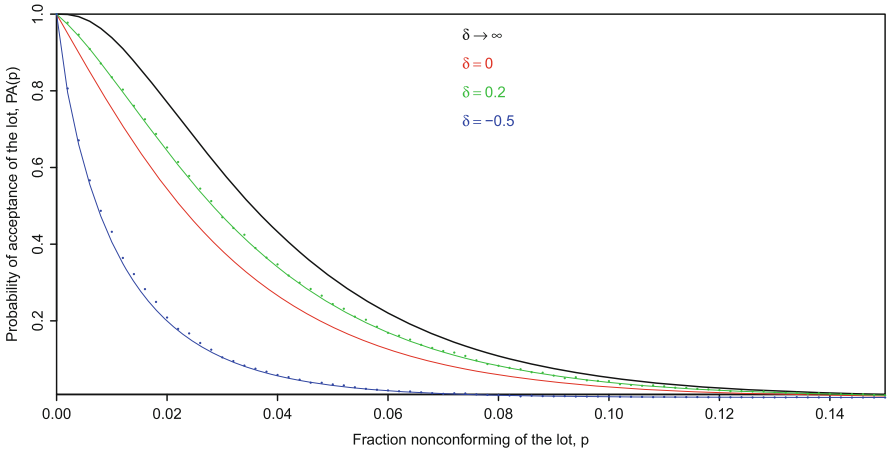


Fig. 3 The OC curves of the sampling plan $(n, k) = (25, 1.83)$ for inspection by variables with unknown σ for various standardized limit distances δ. Theoretical OC curves are presented as *solid lines*, simulation results as points (each *point* represents the average of 10^4 simulation runs)

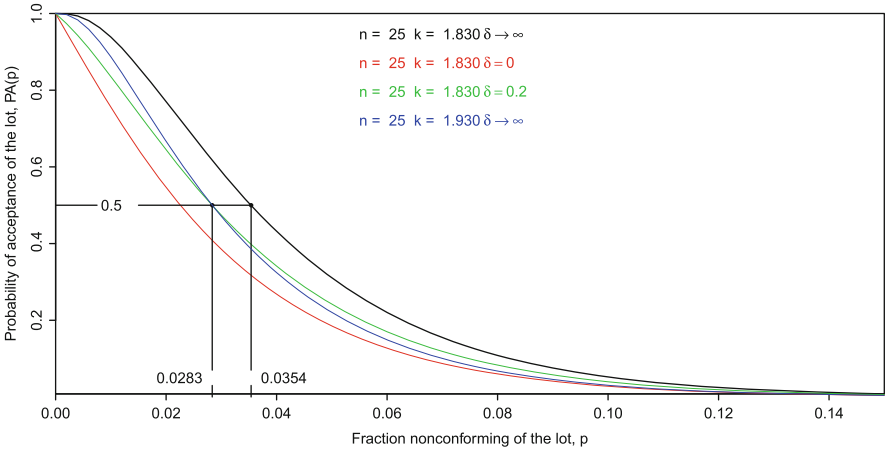


Fig. 4 The OC curves of the sampling plan $(n, k) = (25, 1.83)$ for inspection by variables with unknown σ for $\delta \to \infty$ (unextended, plan 1, *black line*) and $\delta = 0.2$ (extended, plan 2, *green line*) and the unextended plan that matches that the OC of plan 2 at the indifference point $(p_0^\star = 0.0283, 0.5)$. The latter one (plan 3, *blue line*, $k^\star = 1.931$) has a steeper OC than plan 2

that increasing the acceptance constant of the unextended plan from $k = 1.83$ to $k = 1.931$ is a better strategy than using the extended plan with $\delta = 0.2$ because, compared with plan 2, it improves the discrimination power between lots with small and lots with large fractions of nonconforming items.

It should be noted that the OC function (25) depends on the standardized limit distance $\delta = \Delta/\sigma$ and hence, for any Δ that is used in the application of the plan, on the unknown standard deviation σ. Therefore, the OC of an extended sampling plan for inspection by variables with unknown σ is principally unknown. In order to overcome this difficulty we recommend to assume a reasonable value of σ, calculate $\delta = \Delta/\sigma$ and the OC for this δ. This OC is valid if the true standard deviation, σ_{true}, is equal to σ. We discuss this problem further in Sect. 4.

4 The OC of the Sampling Plan if the Standard Deviation Differs from the Assumed Value

4.1 "Known" Standard Deviation

The "known" standard deviation σ, used in the application of the sampling plan for known standard deviation, is the value that is assumed to be equal to the true standard deviation σ_{true}. However, this assumption does not necessarily hold. If we

distinguish between σ and σ_{true} the OC function (17) becomes

$$
\begin{aligned}
&P_A(p)\\
&= P(\bar{X} + k\sigma \le U \wedge X_{(n)} \le U + \Delta \mid p)\\
&= P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma_{\text{true}}} \le \sqrt{n}\frac{U-\mu}{\sigma_{\text{true}}} - k\frac{\sigma}{\sigma_{\text{true}}}) \wedge X_{(n)} \le U + \Delta\right)\\
&= P(Z \le \sqrt{n}(z_{1-p} - k/\tau) \wedge X_{(n)} \le U + \Delta)\\
&= \int_{-\infty}^{\sqrt{n}(z_{1-p}+\min(\delta,-k/\tau))} \phi(z)) F_{(X_{(n)}-\bar{X})/\sigma_{\text{true}}}\left(\frac{U+\Delta-\bar{x}}{\sigma_{\text{true}}}\right) d\bar{x}\\
&\approx \int_{-\infty}^{\sqrt{n}(z_{1-p}+\min(\delta,-k/\tau))} \phi(z)\Phi\left(\frac{\log(z_{1-p}+\delta/\tau - z/\sqrt{n}) - \mu_n}{\sigma_n}\right) dz \qquad (26)
\end{aligned}
$$

where $\tau = \sigma_{\text{true}}/\sigma$. For $\sigma_{\text{true}} = \sigma$ ($\tau = 1$) Eq. (26) is equal to (17). Figure 5 presents the OC curves of the unextended sampling plan ($n = 9, k = 1.83$) and the extended plan for $\delta = 0.2$ for various values of τ. If σ_{true} is larger than σ ($\tau > 1$), then for each value of the fraction nonconforming, p, the probabilities of acceptance of both the unextended and the extended sampling plan are larger than those of the respective plan for $\sigma_{\text{true}} = \sigma$ ($\tau = 1$). This is an unavoidable disadvantage of the sampling plans for inspection by variables, however, the effect is smaller for the extended sampling plan than for the unextended plan. The additional limit $U^\star = U + \Delta = U + \delta\sigma$ reduces the negative effect of σ_{true} being larger

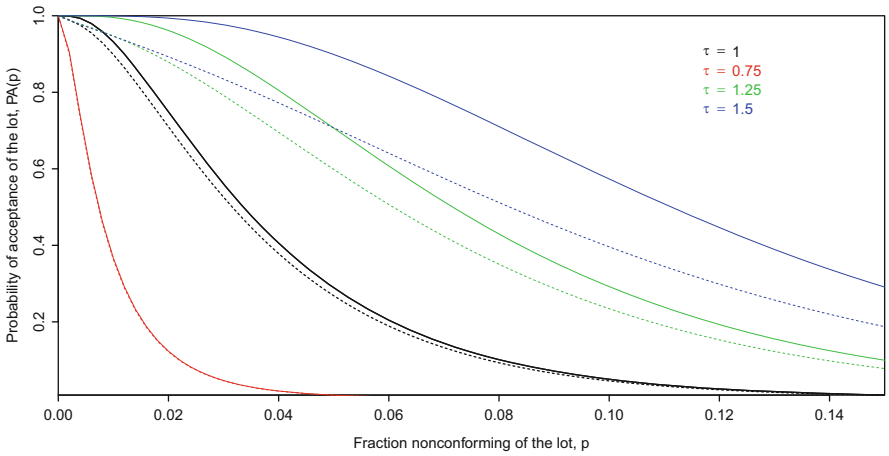


Fig. 5 The OC curves of the sampling plan $(n, k) = (9, 1.83)$ for inspection by variables with known σ, unextended (*solid lines*) and extended with $\delta = 0.2$ (*dashed lines*), for various ratios $\tau = \sigma_{\text{true}}/\sigma$

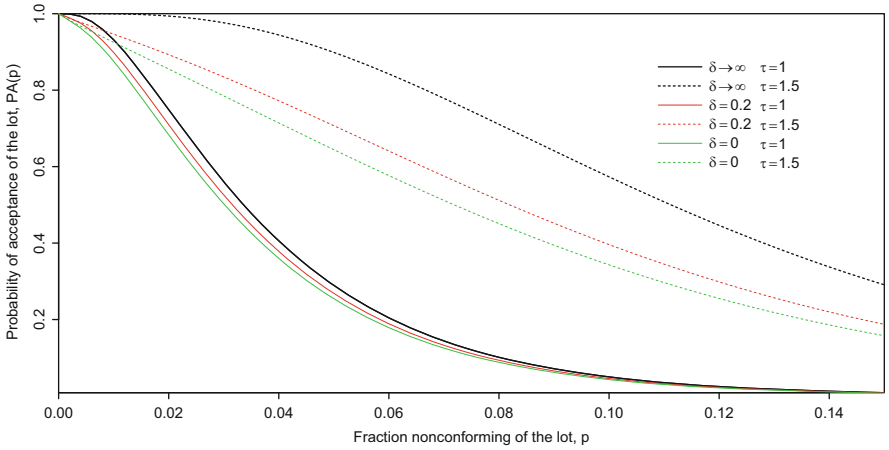


Fig. 6 The OC curves of the sampling plan $(n, k) = (9, 1.83)$ for inspection by variables with known σ, unextended ($\delta \to \infty$, *black lines*), extended with $\delta = 0.2$ (*red lines*) and extended with $\delta = 0$ (*green lines*) for ratios $\tau = \sigma_{\text{true}}/\sigma = 1$ and 1.5

than σ. Figure 6 shows that the negative effect of $\sigma_{\text{true}} > \sigma$ becomes smaller if the additional limit δ tends towards 0. The solid lines are the OCs for $\tau = 1$ and $\delta \to \infty$ (unextended plan, i.e. no additional limit, black line), $\delta = 0.2$ (red line) and $\delta = 0$ (additional limit equal to the specification limit), and the dashed lines are the respective lines for $\tau = 1.5$. Apparently, the distance between the OC of the extended and the unextended plan is smallest for $\delta = 0$ and largest for $\delta \to \infty$.

4.2 Unknown Standard Deviation

The OC (8) of the unextended sampling plan for unknown standard deviation does not depend on the unknown standard deviation. However, as (25) shows, the OC of the extended plan depends on σ via $\delta = \Delta/\sigma$. In Sect. 3 it was recommended to overcome this difficulty by starting with the desired value Δ and then to assume a reasonable value of σ, to calculate $\delta = \Delta/\sigma$ and the OC for this δ. This OC is valid if the true value of the standard deviation, σ_{true}, is equal to σ. If $\sigma_{\text{true}} \neq \sigma$, i.e. $\tau = \sigma_{\text{true}}/\sigma \neq 1$, the OC (25) becomes

$$P_A(p) \approx \int_{z=-\infty}^{\sqrt{n}z_{1-p}} \phi(z) \left[\int_{y=0}^{(n-1)(z_{1-p}-z/\sqrt{n})^2/k^2} f_{\chi^2_\nu}(y) \right. \tag{27}$$

$$\left. \cdot \Phi\left(\frac{\log(\sqrt{n-1}(z_{1-p} + \delta/\tau - z/\sqrt{n})/\sqrt{y}) - \mu_n}{\sigma_n} \right) dy \right] dz.$$

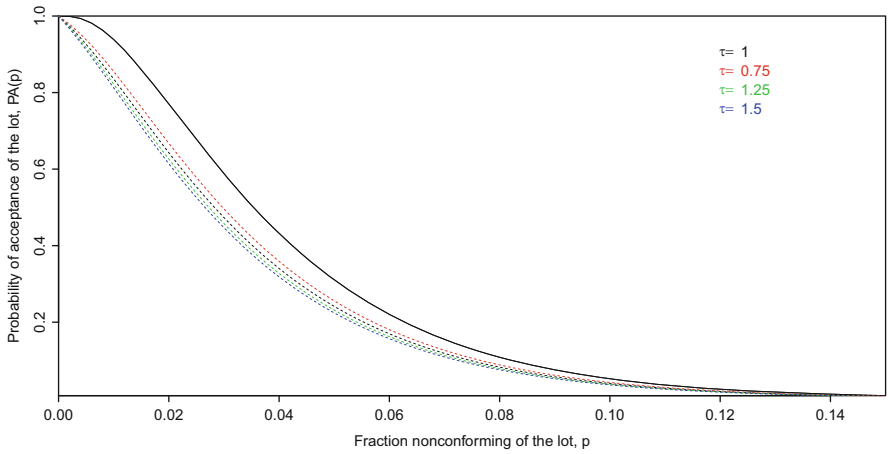


Fig. 7 The OC curves of the sampling plan $(n, k) = (25, 1.83)$ for inspection by variables with unknown σ, unextended (*solid black line*) and extended with $\delta = 0.2$ (*dashed lines*), for various ratios $\tau = \sigma_{\text{true}}/\sigma$

Figure 7 corresponds to Fig. 5. It presents, for $\tau = \sigma_{\text{true}}/\sigma = 1, 0.75, 1.25, 1.5$, the OC curve of the unextended plan (that does not depend on τ) as solid black line and the OC curves of the extended plan as dashed lines. Compared with the case where $\sigma_{\text{true}} = \sigma$ ($\tau = 1$, black dashed line) the probabilities of acceptance are decreased for $\sigma_{\text{true}} > \sigma$ ($\tau > 1$, green and blue line) and increased for $\sigma_{\text{true}} < \sigma$ ($\tau < 1$, red line).

5 The Performance of the Sampling Plans If the Distributional Assumption Is Violated

If X is distributed with unknown $E(X) = \mu$, known $V(X) = \sigma^2$, cumulative distribution function $F_X(x)$ and $(1 - p)$-quantiles x_{1-p} defined by $F_X(x_{1-p}) = 1 - p$, then the fraction nonconforming in the lot is, given the upper limit U,

$$P(X > U) = 1 - P(X \leq U) = 1 - F_X(U) = p, \tag{28}$$

i.e. $U = x_{1-p}$. The OC function of the unextended sampling plan (n, k) for known σ is

$$P_A(p) = P(\bar{X} + k\sigma \leq U \mid p) = P(\bar{X} \leq x_{1-p} - k\sigma). \tag{29}$$

According to the central limit theorem $\bar{X}$ is approximately normal distributed with $E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma^2/n$. Hence, $\sqrt{n}(\bar{X}-\mu)/\sigma$ is approximately standardized normal distributed and

$$\begin{aligned} P_A(p) &= P\left(\sqrt{n}(\bar{X}-\mu)/\sigma \le \sqrt{n}(x_{1-p}-k\sigma-\mu)/\sigma\right) \\ &\approx \Phi\left(\sqrt{n}\left(\frac{x_{1-p}-\mu}{\sigma}-k\right)\right). \end{aligned} \tag{30}$$

The OC functions of the extended plans for known σ and of the unextended and extended plans for unknown σ cannot be obtained analytically because the distributions of $(X_{(n)} - \bar{X})/\sigma$, S and $(X_{(n)} - \bar{X}/S$ are unknown and their independence cannot generally be assumed. Therefore, we obtain these OCs by simulation.

A comparison of (30) with the OC function under normality, (6), shows that the $(1-p)$-quantile z_{1-p} of the standardized normal distribution is replaced by the standardized $(1-p)$-quantile $(x_{1-p}-\mu)/\sigma$ of X and we immediately see that $P_A(p)$ will be larger under the non-normal distribution of X than under normality if $(x_{1-p}-\mu)/\sigma > z_{1-p}$ and smaller otherwise.

As an example of a symmetric heavy-tailed (leptocurtic) distribution we choose the t_ν-distribution with ν degrees of freedom. Its expectation and variance are $E(T_\nu) = 0$ and $V(T_\nu) = \nu/(\nu-2)$; $\nu = 3, \ldots$, respectively. We find the standardized $(1-p)$-quantile of T_ν as

$$\frac{x_{1-p}-\mu}{\sigma} = \frac{t_{\nu;1-p}}{\sqrt{\nu/(\nu-2)}} \tag{31}$$

where $t_{\nu;1-p}$ is the $(1-p)$-quantile of T_ν. Figure 8 shows the OCs of the sampling plan $(n = 9, k = 1.83)$, unextended ($\delta \to \infty$, solid lines) and extended ($\delta = 0$, dotted lines) for the normal distribution (black) and the t_ν-distribution with $\nu = 10$ (green) and $\nu = 3$ (blue) degrees of freedom. The standardized quantile function $t_{\nu;1-p}/\sqrt{\nu/(\nu-2)}$ of the t_ν-distribution with $\nu = 10$ and of the quantile function z_{1-p} of the standardized normal distribution intersect at $p = p_0 = 0.036$ for $\nu = 10$ and at $p = p_0 = 0.018$ for $\nu = 3$. For $p \le p_0$ the former one and hence, $P_A(p)$ is larger, for $p \ge p_0$ it is smaller. As heavier the tails of the underlying distribution are, as more is the acceptance probability decreased (for not too small p) as compared with the normal distribution. This is also true for the extended sampling plan with $\delta = 0$, however, the extension of the sampling plan has no additional effect. Figure 9 corresponds to Fig. 8, however now it gives the OCs of the matching sampling plan $(n = 25, k = 1.83)$ for unknown σ. Both the unextended and the extended sampling plan are rather robust with regard to a symmetric heavy-tailed distribution instead of the normal distribution.

As an example of a distribution that is skewed to the right we choose the χ^2_ν-distribution with ν degrees of freedom. Its expectation and variance are $E(\chi^2_\nu) = \nu$

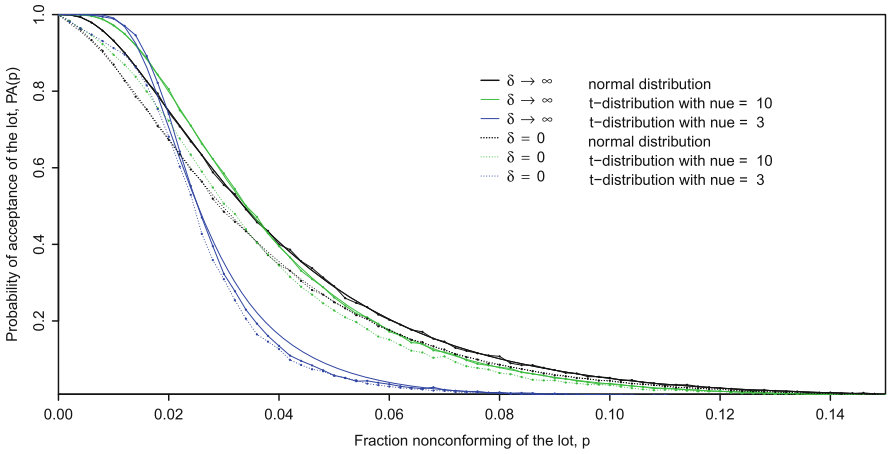


Fig. 8 OCs of the sampling plan ($n = 9, k = 1.83$) for known σ, unextended ($\delta \to \infty$, *solid lines*) and extended ($\delta = 0$, *dotted lines*) for the normal distribution (*black*) and the t_ν-distribution with $\nu = 10$ (*green*) and $\nu = 3$ (*blue*) degrees of freedom. The smooth *curves* are the theoretical OCs. Each point represents the average over 10^4 simulation runs

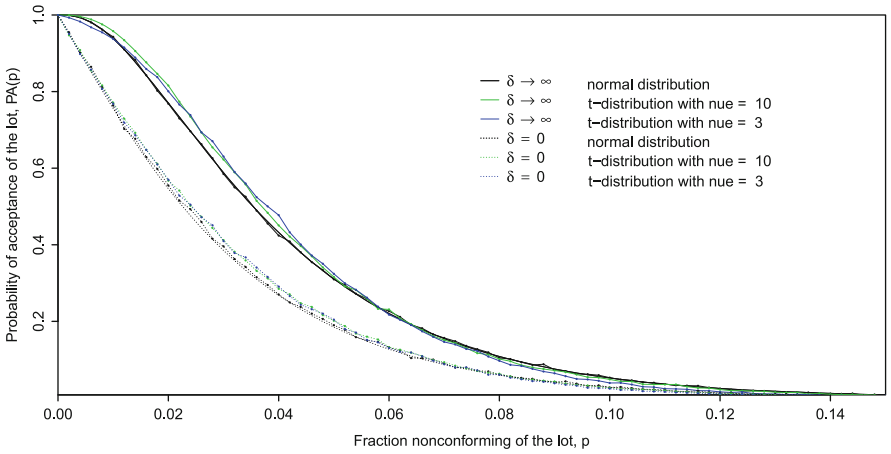


Fig. 9 OCs of the sampling plan ($n = 25, k = 1.83$) for unknown σ, unextended ($\delta \to \infty$, *solid lines*) and extended ($\delta = 0$, *dotted lines*) for the normal distribution (*black*) and the t_ν-distribution with $\nu = 10$ (*green*) and $\nu = 3$ (*blue*) degrees of freedom. Each *point* represents the average over 10^4 simulation runs

and $V(\chi^2_\nu) = 2\nu$; $\nu = 1, \ldots$, respectively. We find the standardized $(1-p)$-quantile of T_ν as

$$\frac{x_{1-p} - \mu}{\sigma} = \frac{\chi^2_{\nu;1-p} - \nu}{\sqrt{2\nu}} \tag{32}$$

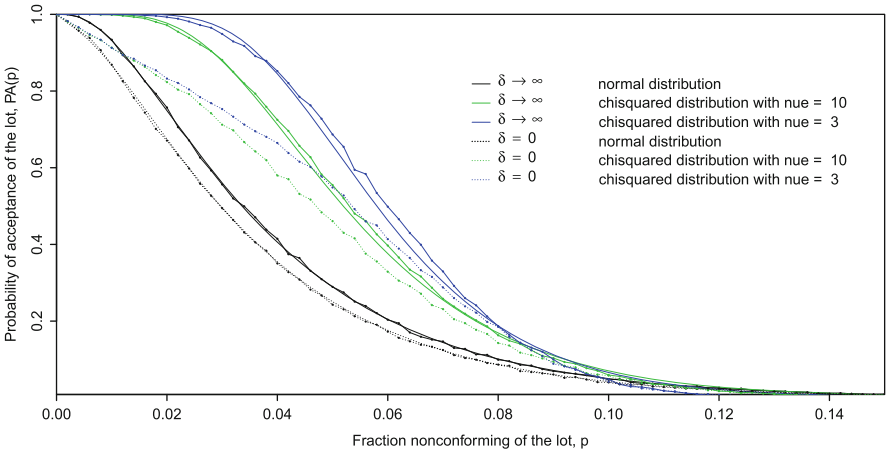


Fig. 10 OCs of the sampling plan ($n = 9, k = 1.83$) for known σ, unextended ($\delta \to \infty$, *solid lines*) and extended ($\delta = 0$, *dotted lines*) for the normal distribution (*black*) and the χ^2_ν-distribution with $\nu = 10$ (*green*) and $\nu = 3$ (*blue*) degrees of freedom. Each *point* represents the average over 10^4 simulation runs

where $\chi^2_{\nu;1-p}$ is the $(1-p)$-quantile of χ^2_ν. Figure 10 shows the OCs of the sampling plan ($n = 9, k = 1.83$), unextended ($\delta \to \infty$, solid lines) and extended ($\delta = 0$, dotted lines) for the normal distribution (black) and the χ^2_ν-distribution with $\nu = 10$ (green) and $\nu = 3$ (blue) degrees of freedom. The standardized quantile function $(\chi^2_{\nu;1-p} - \nu)/\sqrt{2\nu}$ of the χ^2_ν-distribution with $\nu = 10$ and of the quantile function z_{1-p} of the standardized normal distribution intersect at $p = p_0 = 0.114$ for $\nu = 10$ and at $p = p_0 = 0.133$ for $\nu = 3$. For $p \le p_0$ the former one and hence, $P_A(p)$ is (much) larger, for $p \ge p_0$ it is smaller. As more skewed to the right the underlying distribution is, as more is the acceptance probability increased as compared with the normal distribution. This is also true for the extended sampling plan with $\delta = 0$, however, the extension of the sampling plan has no additional effect. Figure 11 corresponds to Fig. 10, however now it gives the OCs of the matching sampling plan ($n = 25, k = 1.83$) for unknown σ. The acceptance probabilities of the unextended plan are heavily increased under the skewed distribution whereas they remain almost stable if the extended plan with $\delta = 0$ is used.

As examples of mixed distributions we choose two distributions with the densities $p_1 f_1(\mu_1, \sigma_1^2) + p_2 f_2(\mu_2, \sigma_2^2)$; $p_1 + p_2 = 1$, $\mu_2 > \mu_1$ where f_1 and f_2 are normal densities. We set $p_1 = 0.8$ and $\sigma_1^2 = \sigma_2^2 = 0.5$ (for the mixed distribution 1) and $\sigma_1^2 = \sigma_2^2 = 0.3$ (for the mixed distribution 2). In order to get a mixed distribution with $E(X) = \mu = 0$ and $V(X) = \sigma^2 = 1$ we have to fix $\mu_1 = -p_2 d$ and $\mu_2 = p_1 d$ where $d = \mu_2 - \mu_1 = \sqrt{\frac{1-\sigma_1^2}{p_1 p_2}}$. This gives $\mu_1 = -0.354$ and $\mu_2 = 1.414$ for the mixed distribution 1 and $\mu_1 = -0.418$ and $\mu_2 = 1.673$ for the mixed distribution 2. Both mixed distributions have a large component ($p_1 = 0.8$) on the left side and a small component on the right side. In the mixed distribution

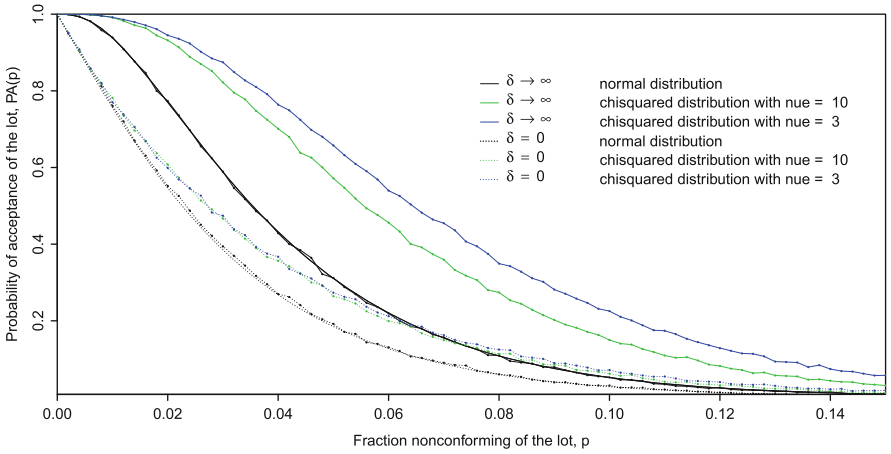


Fig. 11 OCs of the sampling plan ($n = 25, k = 1.83$) for unknown σ, unextended ($\delta \to \infty$, *solid lines*) and extended ($\delta = 0$, *dotted lines*) for the normal distribution (*black*) and the χ^2_ν-distribution with $\nu = 10$ (*green*) and $\nu = 3$ (*blue*) degrees of freedom. Each *point* represents the average over 10^4 simulation runs

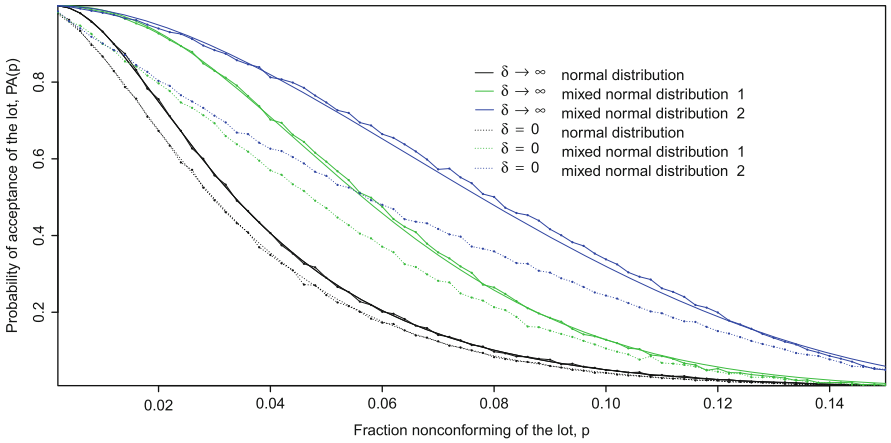


Fig. 12 OCs of the sampling plan ($n = 9, k = 1.83$) for known σ, unextended ($\delta \to \infty$, *solid lines*) and extended ($\delta = 0$, *dotted lines*) for the normal distribution (*black*) and the mixed distribution 1 (*green*) and 2 (*blue*). Each *point* represents the average over 10^4 simulation runs

2 the distance between the means of the two components is larger in relation to their common standard deviation than in the mixed distribution 1. Figure 12 shows the OCs of the sampling plan ($n = 9, k = 1.83$), unextended ($\delta \to \infty$, solid lines) and extended ($\delta = 0$, dotted lines) for the normal distribution (black) and the mixed distribution 1 (green) and 2 (blue). The standardized quantile functions of both mixed distributions are larger than that of the standardized normal distribution (z_{1-p}) and hence, $P_A(p)$ is (much) larger than under normal distribution. As more

distant the means of the mixed distribution (with a large component on the left side and a small component on the right side), as more is the acceptance probability increased as compared with the normal distribution.

We do not present a similar graph for the case of sampling plans with unknown σ because their behaviour under mixed distributions is principally equal to that of the plans with known σ.

6 Conclusions

- The additional criterion $X_{(n)} \leq U + \Delta$ for acceptance decreases the probability of acceptance for each fraction p of nonconforming items in the lot. However, such an extended sampling plan is less efficient than the unextended sampling plan with identical indifference point.
- The "known" standard deviation σ, used in the application of the sampling plan for known standard deviation, is the value that is assumed to be equal to the true standard deviation σ_{true}. If σ_{true} is larger than σ, then for each value of the fraction nonconforming, p, the probabilities of acceptance of both the unextended and the extended sampling plan are larger than those of the respective plan for $\sigma_{\text{true}} = \sigma$ and vice versa. The additional limit $U^\star = U + \Delta = U + \delta\sigma$ reduces the negative effect of σ_{true} being larger than σ.
- The OC function of the unextended sampling plan for unknown standard deviation does not depend on the unknown standard deviation. However, for the extended sampling plan it depends on the standardized limit distance $\delta = \Delta/\sigma$ and hence, for any $\Delta \neq 0$ on the unknown σ. Therefore, the OC of an extended sampling plan for inspection by variables with unknown σ is principally unknown. In order to overcome this difficulty we recommend to assume a reasonable value of σ, to calculate $\delta = \Delta/\sigma$ and the OC for this δ. Compared with the case where $\sigma_{\text{true}} = \sigma$ the probabilities of acceptance are increased for $\sigma_{\text{true}} > \sigma$ and decreased for $\sigma_{\text{true}} < \sigma$.
- Both the unextended and the extended sampling plan are rather robust with regard to a symmetric heavy-tailed distribution instead of the normal distribution.

  The acceptance probabilities of the unextended plan are heavily increased under a skewed distribution whereas they remain almost stable if the extended plan with $\delta = 0$ is used.

  As more distant the means of a mixed distribution (with a large normal component on the left side and a small normal component on the right side) are, as more is the acceptance probability increased as compared with the normal distribution. This effect is similar for the unextended and the extended sampling plan.
- Each outlying observation x above the additional limit $U + \Delta$ causes an immediate rejection of the lot.

The essential advantages of the extended sampling plan as compared to the unextended sampling plan are

- a good protection against outlying observations (that might indicate a contamination of the distribution of X with a component with larger mean or/and larger standard deviation) and
- a partial compensation of the negative effect of the true standard deviation being larger than the "known" or assumed value σ.

References

Bray, D. F., Lyon, D. A., & Burr, I. W. (1973). Three class attributes plans in acceptance sampling. *Technometrics*, *15*, 575–585.

CEB-FIP (1978). *Model code for concrete structures* Vol's I and II, (3rd ed.). (pp. 124–125/273) Paris: Bulletin d'information CEB

Dahms, S. (2003). Microbiological sampling plans–statistical aspects. *SGLH tagung microbiological food safety*, Zürich.

Dahms, S. (2004). Stichprobenpläne und mikrobiologische Kriterien als Risikomanagement-Option in neuen Konzepten zur Lebensmittelsicherheit. *Berl Münch Tierärztl Wschenschr*, *117*, 193–200.

Dahms, S., & Hildebrandt, G. (1998). Some remarks on the design of three-class sampling plans. *Journal of Food Protection*, *61*, 757–761.

European Directorate for the Quality of Medicines, EDQM (2010). European Pharmacopoeia 7.4, Section 2.9.40 "Uniformity of Dosage Units", pp. 4101–4103.

European Standard EN 206-1 (2000). Concrete – Part 1: Specification, performance, production and conformity. Brussels: European Committee for Standardization (CEN).

Hildebrandt, G., Böhmer, L., & Dahms, S. (1995). Three-class attributes plans in microbiological quality control: a contribution to the discussion. *Journal of Food Protection*, *58*, 784–790.

ICMSF (2002). Microrganisms in foods 7. microbiological testing in food safety management. *Chapter 7 (Sampling Plans)* pp. 123–165. New York: Kluwer Academic/Plenum Publishers

Legan, J. D., Vandeven, M. H., Dahms, S., & Cole, M. B. (2001). Determining the concentration of microorganisms controlled by attributes sampling plans. *Food Control*, *12*, 137–147.

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, *27*, 466–471.

Schilling, E. G., & Neubauer, D. V. (2009). *Acceptance sampling in quality control* (2nd ed.), Boca Raton: Chapman & Hall/CRC

Taerwe, L. (1988). Evaluation of compound compliance criteria for concrete strength. *Materials and Structures*, *21*, 13–20.

Verordnung über Hygiene- und Qualitätsanforderungen an Milch und Erzeugnisse auf Milchbasis (Milchverordnung) vom 24. April 1995, (BGBl. I, 544).

Wilrich, P.-Th., Weiss, H. (2011). Three-class sampling plans for the evaluation of bacterial contamination. *Milchwissenschaft*, *66*, 413–416.

Fractional Acceptance Numbers for Lot Quality Assurance

K. Govindaraju and G. Jones

Abstract Fractional acceptance numbers such as one-quarter fraction nonconforming are used in the acceptance sampling literature. The concept of fractional acceptance number is particularly useful for short-run food manufacturing processes involving a measurable quality characteristic such as the percentage sugar or fat content. Attribute method of inspection is desirable for small sample sizes because of the difficulty in identifying the underlying distribution. Analytical testing of fat content, etc. also involves considerable measurement uncertainty, often up to half of the observed variation. However the distribution of the measurement errors can be fairly well ascertained using past calibration studies. An observed measurement X is classified with certainty as conforming or not for given specification limits only when there are no measurement errors. Measurement error uncertainty results only in an estimated probability of conformance of a unit. The probability of nonconformance of an individual unit based on the error-prone measurement is defined as the 'fractional' nonconforming unit. A new fractional acceptance number sampling plan, which is a mix of attribute and variables methods, is introduced. The operating characteristics of the proposed plan are evaluated using common error distributions and the incomplete beta function. The fractional acceptance numbers also extend to broken acceptance numbers such as 1.6.

Keywords Measurement error correction • OC function • Sampling plans

1 Introduction

Measurement of quality by the variables method is usually considered as more informative than the attribute method of counting nonconforming units. Let a random sample of size n, say $(X_1, X_2, \ldots, X_n)$, be available for a continuous quality characteristic X. Let $\bar{X}$ and S be the sample mean and standard deviation ($n - 1$ in the divisor). The uniformly minimum variance unbiased estimate (UMVUE) of

K. Govindaraju (✉) • G. Jones
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
e-mail: k.govindaraju@massey.ac.nz; g.jones@massey.ac.nz

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_17

the fraction nonconforming p for a normally distributed quality characteristic with a lower specification limit L is obtained using the statistic

$$z_{pL} = \sqrt{\frac{n}{n-1}}\left(\frac{L-\bar{X}}{S}\right)$$

which follows the non-central t distribution (see (Lieberman and Resnikoff 1955) or (Schilling and Neubauer 2009) for details). For Maximum Likelihood (ML) and other methods of estimation of p, see Guenther (1971), Brown and Rutemiller (1973), Boullion et al. (1985), and Md-Yusof and Rigdon (1992).

Let the variable measurements be instead used to classify the sampled items as conforming or nonconforming based on a given lower specification limit L (say). The observed number of nonconforming units d becomes a binomial random variable giving an estimated fraction nonconforming $\hat{p} = \frac{d}{n}$. Consider the (fictitious) sample measurements (50.01, 50.04, 50.07, 50.10, 50.15, 50.20, 50.29, 50.42, 50.45, 50.48, 50.55, 50.60, 50.80, 51.20, 51.30) for $L = 50$ which passes various normality tests. The estimated fraction nonconforming under the normal distribution for X is $\hat{p} = 0.13$. On the other hand, there are no nonconforming units in the sample and hence $\hat{p} = 0$ under the attribute method. Such huge differences in the estimated fraction nonconforming leads to contradictory conclusions depending on whether an attribute or variables sampling plan is employed. A lot will always be accepted if there are no nonconforming units in the sample under the attribute method (irrespective of the acceptance number Ac). A lot rejection with no apparent nonconforming units in the sample is not viewed favorably by the producer. The disadvantage of variables plans rejecting lots with no apparent nonconforming units is well known and discussed in Woods (1955), Duncan (1986), and Collani (1991).

Small sample sizes can be blamed for the large standard errors in the estimated fraction nonconforming but the justification of a known probability distribution for X is important to correctly estimate the fraction nonconforming. It is well known that the fraction nonconforming estimates based on the normal distribution are not robust. The effect of non-normality on variables sampling plans is studied by Das and Mitra (1964), Singh (1966), Owen (1969), Takagi (1972), Schneider and Wilrich (1981), and Montgomery (1985). The producer's and consumer's risks as well as the fraction nonconforming are incorrectly estimated when the distribution of the lot quality characteristic is non-normal. Process capability indices are also not robust to non-normality, see Clements (1989), Kotz and Johnson (1993), Somerville and Montgomery (1996), Ahmed (2005), and others. Tail area estimation depends on the assumed probability distribution, and hence it is not surprising that the parametric approach to the estimation of fraction nonconforming, process capability, etc. does not work well for small sample sizes. The validity of the normal distribution for finite size lots is investigated by Collani (1991), Seidel (1997), and others. When compared to attribute plans, variables plans are found to incorrectly estimate the fraction nonconforming and risks for finite lot sizes. In the absence of quality history, such as the Phase I process control data,

the distributional uncertainty in the quality characteristic X forces us to take a nonparametric approach to quality assurance. While non-normality is not critical for the $\bar{X}$ chart (see, Burr (1967), Schilling and Nelson (1978), and Balakrishnan and Kocherlakota (1986)), the distributional uncertainty cannot be ignored for control charts for dispersion such as the S chart (see Abbasi and Miller (2012) and the references cited therein).

Food manufacturing processes are often short-run or batch production processes. Such processes involve frequent start-ups due to the need to clean the machinery for safety reasons. Frequent sampling from the process is also not practical. Hourly samples are often taken for a production run whose length may extend to 20–30 h. Food quality characteristics such as the percentage sugar and fat are determined using analytical methods which involve considerable measurement uncertainty. This paper deals with the quality assurance problem when only small samples are available, and half of the observed variation or even more is due to measurement uncertainty. In Sect. 2, a review of fractional acceptance numbers is given. In Sect. 3, the use of fractional and broken acceptance numbers for lot quality assurance after correction for measurement errors is discussed. The case of known distribution for the quality characteristic of interest is then discussed in Sect. 4. The last section provides the summary and conclusions.

2 Fractional Acceptance Numbers

In a single sampling attribute plan, the acceptance number Ac is the maximum allowable number of nonconforming units in the sample of size n for lot acceptance purposes. In order to compare the single sampling plan with other types of sampling plans such as the double sampling plan, Hamaker (1950) introduced the concept of fractional and broken acceptance numbers (and sample sizes). For example, an acceptance number of 1.6 is equivalent to using the $Ac = 1$ plan 40 % of the time and the $Ac = 2$ plan 60 % of the time using a randomization (lottery) device. This approach derives its support from the randomized hypothesis test procedure in the statistical inference literature (Young and Smith 2005, p.66). Govindaraju (1991) showed the equivalence of fractional acceptance number single sampling plans to double sampling plans with different sample sizes for the first and second samples. This approach operationalizes the fractional acceptance numbers to achieve a desired level of discrimination between acceptable and rejectable levels of fraction nonconforming. Fractional acceptance numbers are also part of the ISO Standard ISO 2859-1:1999(E) (1999).

A random sample of continuous measurements $X_1, X_2, \ldots, X_n$ becomes available under the variables method of inspection. For a given lower specification limit L, the indicator variable $I_{X_i<L}$ becomes a Bernoulli random variable with $p_i = Pr(X_i < L)$. That is, the variable measurements are transformed to a binary quality measure (one for $X_i < L$ or zero for $X_i \geq L$) with the number of nonconforming units $d = \sum_{i=1}^{n} I_{X_i<L}$ becoming a binomial random variable

for $p_1 = p_2 = \ldots = p_n = p$. The single sampling attribute plan having an acceptance number Ac accepts the lot when $d \leq Ac$. The assumption of identically and independently distributed X_i is critical for using the binomial distribution for obtaining the OC function of the single sampling attribute plan as

$$P_a(p|n, Ac) = \sum_{d=0}^{Ac} \binom{n}{d} p^d (1-p)^{n-d}.$$

The above OC function giving the probability of acceptance for a given lot fraction nonconforming p, known as the Type B OC function, is valid for large lot sizes as well as for a series of lots; see Duncan (1986). The above OC function is equivalent to the incomplete beta function

$$B(p|n, Ac) = \frac{\Gamma(n-1)}{\Gamma(Ac-1)\Gamma(n-Ac)} \int_0^{1-p} t^{n-Ac-1}(1-t)^{Ac} dt$$

which allows for fractional Ac values. For $np < 5$, the incomplete gamma function

$$G(x|Ac) = \frac{1}{\Gamma(Ac+1)} \int_x^{\infty} e^{-t} t^{Ac} dt$$

provides a good approximation where $x = np$. Single sampling plans are commonly designed (i.e. n and Ac are determined) for given Acceptance Quality Limit (AQL) $p_{1-\alpha}$, producer's risk $\alpha = 1 - P_a(p_{1-\alpha})$, Limiting Quality (LQ) p_β, and consumer's risk $\beta = P_a(p_\beta)$. Fractional acceptance numbers are useful in achieving a desired operating ratio $p_\beta / p_{1-\alpha}$ such as 8. For example, let $p_{1-\alpha} = 0.01$, $\alpha = 0.05$, $p_\beta = 0.08$, $\beta = 0.1$. The resulting equations $B(0.01|n, Ac) = 0.95$ and $B(0.08|n, Ac) = 0.1$ can be solved for n and Ac using software such as R (R Development Core Team 2012) as $Ac = 1.444531$ and $n = 55.16412$ (see the Appendix for R codes). Rounding the sample size n to an integer value such as 55 does not affect the achieved producer's and consumer's risks much. Such rounding cannot be done to Ac because the $p_\beta / p_{1-\alpha}$ is purely a function of Ac. Table 1 shows the operating ratios $R = \frac{np_\beta}{np_{1-\alpha}} = \frac{p_\beta}{p_{1-\alpha}}$ for various fractional acceptance numbers where p_β is the fraction nonconforming for which $G(np_\beta|Ac) = \beta$ and $p_{1-\alpha}$ is the fraction nonconforming for which $G(np_{1-\alpha}|Ac) = 1 - \alpha$. In other words, the operating ratios are purely a function of the fractional acceptance numbers under the gamma (Poisson) model for the expected number of nonconforming units in a sample of size n. This result is also true for broken acceptance numbers exceeding 1 such as 1.6. The R codes given in the Appendix can be used to obtain such broken acceptance numbers $Ac > 1$. Fractional and broken acceptance numbers are particularly useful for lot quality assessment in the presence of measurement errors because of the inherent uncertainty in assessing whether a particular unit conforms to the specifications or not. The probability distribution of the measurement errors

Table 1 Operating ratios of fractional acceptance number Ac for given α and β

Ac	$np_{0.95}$	$np_{0.1}$	$R = \frac{np_{0.1}}{np_{0.95}}$	$np_{0.99}$	$np_{0.05}$	$R = \frac{np_{0.05}}{np_{0.99}}$
0	0.0513	2.3026	44.8908	0.0101	2.9957	297.2171
0.02	0.0549	2.3372	42.5641	0.0111	3.0344	273.2344
0.04	0.0587	2.3716	40.4097	0.0122	3.0727	251.6212
0.06	0.0626	2.4059	38.4423	0.0134	3.1109	232.2854
0.08	0.0666	2.4400	36.6410	0.0146	3.1489	215.0587
0.10	0.0708	2.4740	34.9667	0.0160	3.1867	199.6580
0.12	0.0750	2.5077	33.4358	0.0174	3.2242	185.6969
0.14	0.0794	2.5414	32.0205	0.0188	3.2616	173.4350
0.16	0.0838	2.5749	30.7091	0.0203	3.2988	162.2469
0.18	0.0884	2.6083	29.4925	0.0219	3.3358	152.1442
0.20	0.0931	2.6415	28.3604	0.0236	3.3727	143.4700
0.22	0.0979	2.6745	27.3056	0.0254	3.4093	134.3994
0.24	0.1029	2.7075	26.3210	0.0272	3.4459	126.8273
0.26	0.1079	2.7403	25.4000	0.0291	3.4822	119.8340
0.28	0.1130	2.7731	24.5377	0.0310	3.5184	113.4246
0.30	0.1183	2.8056	23.7161	0.0331	3.5545	107.4634
0.32	0.1236	2.8381	22.9569	0.0352	3.5903	102.0146
0.34	0.1291	2.8705	22.2422	0.0373	3.6261	97.1426
0.36	0.1346	2.9027	21.5687	0.0396	3.6617	92.4704
0.38	0.1402	2.9349	20.9330	0.0419	3.6972	88.1764
0.40	0.1459	2.9669	20.3318	0.0443	3.7325	84.2003
0.42	0.1518	2.9989	19.7613	0.0468	3.7678	80.4894
0.44	0.1577	3.0307	19.2225	0.0494	3.8028	77.0459
0.46	0.1637	3.0625	18.7118	0.0520	3.8378	73.8355
0.48	0.1698	3.0941	18.2270	0.0547	3.8726	70.8364
0.50	0.1759	3.1257	17.7664	0.0574	3.9074	68.0318
0.52	0.1822	3.1572	17.3285	0.0603	3.9420	65.4049
0.54	0.1885	3.1885	16.9115	0.0632	3.9765	62.9413
0.56	0.1950	3.2198	16.5141	0.0662	4.0109	60.6276
0.58	0.2015	3.2511	16.1351	0.0691	4.0452	58.5037
0.60	0.2081	3.2822	15.7732	0.0723	4.0793	56.4517
0.62	0.2148	3.3132	15.4274	0.0755	4.1134	54.5178
0.64	0.2215	3.3442	15.0967	0.0787	4.1474	52.6931
0.66	0.2284	3.3751	14.7802	0.0820	4.1813	50.9695
0.68	0.2353	3.4059	14.4771	0.0854	4.2150	49.3397
0.70	0.2422	3.4367	14.1865	0.0889	4.2488	47.7992
0.72	0.2493	3.4673	13.9094	0.0924	4.2824	46.3396
0.74	0.2564	3.4979	13.6403	0.0960	4.3158	44.9549
0.76	0.2636	3.5284	13.3844	0.0997	4.3493	43.6402

(continued)

Table 1 (continued)

Ac	$np_{0.95}$	$np_{0.1}$	$R = \frac{np_{0.1}}{np_{0.95}}$	$np_{0.99}$	$np_{0.05}$	$R = \frac{np_{0.05}}{np_{0.99}}$
0.78	0.2709	3.5589	13.1363	0.1034	4.3826	42.3907
0.80	0.2783	3.5893	12.8988	0.1072	4.4158	41.2022
0.82	0.2857	3.6196	12.6704	0.1110	4.4490	40.0706
0.84	0.2931	3.6499	12.4506	0.1149	4.4821	38.9924
0.86	0.3007	3.6801	12.2366	0.1189	4.5151	37.9641
0.88	0.3083	3.7102	12.0331	0.1230	4.5480	36.9828
0.90	0.3160	3.7403	11.8358	0.1271	4.5808	36.0455
0.92	0.3238	3.7703	11.6453	0.1313	4.6136	35.1496
0.94	0.3316	3.8002	11.4619	0.1355	4.6463	34.2927
0.96	0.3394	3.8301	11.2845	0.1398	4.6789	33.4724
0.98	0.3474	3.8599	11.1111	0.1441	4.7114	32.6867
1	0.3554	3.8897	10.9455	0.1486	4.7439	31.9336

results in fractional nonconformance of a given measurement depending on its closeness to the specification. This is discussed in the next section.

3 Measurement Error Correction for Attribute Inspection

Type I (false positive) misclassification error places a true conforming unit as (apparently) nonconforming, while Type II (false negative) misclassification error places a true nonconforming unit as (apparently) conforming. Let e_1 and e_2 be the Type I and Type II error probabilities, respectively. The relationship between true fraction nonconforming p and the apparent fraction nonconforming units p_e is given by $p_e = e_1(1-p)+(1-e_2)p$, see Lavin (1946). It is established in the literature that the misclassification errors affect the producer more than the consumer in general. A good introduction to the topic of inspection errors in attribute sampling is available in the text by Johnson et al. (1991). The main weakness in the theory of inspection errors for attribute inspection is that the error probabilities e_1 and e_2 are assumed to be fixed and known.

Fuzzy set theory is also used to adjust for the inspection errors. Hryniewicz (2008) provided an excellent review and discussion on handling fuzzy data in statistical quality control. Jamkhaneh (2011) provided a recent review on the effect of inspection on single sampling attribute plans with fuzzy parameters. Imprecise or fuzzy quality measurements are modelled using membership functions. It is common to model the inspection of the whole sample using a trinomial or multinomial distribution, see the discussion in Ohta (1988) and Hryniewicz (2008). A major limitation of the fuzzy approach to the design of acceptance sampling plans is that the fuzzy bounds (numbers) must be known and fixed. Fixed fuzzy numbers may not be adequate for measurable characteristics which are

converted to attributes after comparing the measured value with the specifications. A measurement closer to the specification is more likely to be nonconforming than a measurement away from the specification. The fuzzy approach allows for only possibilities and not precise probabilities for the degree of nonconformance. A flexible approach that allows for varying probabilities for nonconformance within the attribute inspection is desirable in practice. The probability distribution of the measurement errors is usually determined by repeat testing of the same unit or the fully blended composite specimen. In the absence of other nuisance variables, a controlled repeat testing experiment will justify a probability distribution model for measurement errors. The measurement error distributions are usually unimodal and symmetrical distributions such as triangular, normal, etc. For small samples, the distributional uncertainty in the main quality characteristic of interest cannot be resolved satisfactorily, and hence an attribute assessment of lot quality is desirable. However the distributional uncertainty is resolvable for measurement errors and hence the degree of nonconformance of a measurement can be quantified as probabilities.

As an example, let the measurement error t follows a triangular distribution $T(\min = a, \max = b, \text{mode} = c)$. A single observation x from a unimodal distribution estimates the mode of the unknown underlying distribution. When the triangular error distribution is shifted to the observation x (i.e. setting the mode c at x), $T(\min = a + (x - c), \max = b + (x - c), \text{mode} = x)$ gives an estimate of the fraction conforming $q_c = \Pr(t > L | \min, \max, \text{mode})$ and an estimate of the fraction nonconforming $p_c = 1 - q_c$. For example, assume that the triangular distribution $T(a = -0.1, b = 0.3)$ with mode at $c = 0.05$ was fitted for measurement errors after repeated testing of several specimens. The fitted error distribution is asymmetrical and hence some instrument bias is also present. For a given apparent measurement $x = 50.05$, the error distribution $T(\min = 49.9, \max = 50.3, \text{mode} = 50)$ gives $q_c = 0.75$ being the probability of falling above the lower specification limit $L = 50$. It should be noted that a measurement of 50.05 yields an estimate of 1 (certainty) for fraction conforming under the attribute method (assuming Bernoulli trials) when measurement errors are ignored but this estimate falls to 0.75 after correcting for the measurement error because the apparent measurement is closer to $L = 50$.

When error-prone variables measurements are converted into attribute measurements, the fractional and broken acceptance numbers are suitable. A new fractional acceptance number plan is introduced in the following steps:

1. Obtain the apparent sample measurements $x_1, x_2, \ldots, x_n$.
2. Calculate the estimated probabilities of nonconformance $\widehat{p_1}, \widehat{p_2}, \ldots, \widehat{p_n}$ under the known measurement error distribution.
3. Accept the lot if $\delta = \sum_{i=1}^{n} \widehat{p_i} \leq Ac$. Reject the lot when $\delta > Ac$.

For the fictitious sample of 15 measurements considered in Sect. 1, we compute $\delta = \sum_{i=1}^{n} \widehat{p_i} = 0.958$ under the triangular $T(a = -0.1, b = 0.3, c = 0.05)$ error distribution. For a fractional acceptance number $Ac = 0.75$ (say), the lot will be rejected. Sample protein percentages found for each of the ten lots of a milk

Table 2 Protein percentage data for ten lots

Lot 1	Lot 2	Lot 3	Lot 4	Lot 5	Lot 6	Lot 7	Lot 8	Lot 9	Lot 10
24.76	24.81	24.79	24.41	24.47	24.25	24.34	24.23	24.29	24.43
24.74	24.79	24.81	24.43	24.46	24.24	24.31	24.37	24.33	24.45
24.76	24.77	24.80	24.48	24.42	24.26	24.32	24.24	24.29	24.39
24.77	24.79	24.82	24.48	24.37	24.27	24.33	24.16	24.30	24.44
24.76	24.81	24.78	24.42	24.39	24.23	24.27	24.21	24.34	24.50
24.80	24.74	24.79	24.49	24.40	24.26	24.33	24.21	24.33	24.49
24.77	24.75	24.78	24.54	24.50	24.24	24.26	24.15	24.27	24.47
24.97	24.71	24.73	24.19	24.26	24.31	24.32	24.25	24.39	24.46
25.01	24.65	24.64	24.22	24.23	24.29	24.29	24.30	24.39	24.30
24.95	24.66	24.66	24.16	24.16	24.35	24.24	24.17	24.57	24.27
24.94	24.65	24.62	24.20	24.18	24.33	24.23	24.32	24.35	24.42
24.89	24.70	24.59	24.23	24.17	24.34	24.32	24.44	24.32	24.35
25.07	24.64	24.60	24.16	24.22	24.35	24.27	24.37	24.28	24.34
24.98	24.65	24.54	24.19	24.26	24.27	24.24	24.49	24.32	24.30
24.91	24.67	24.49	24.22	24.28	24.30	24.27	24.40	24.26	24.40
25.04	24.63	24.44	24.26	24.24	24.25	24.20	24.47	24.33	24.30
25.05	24.67	24.47	24.19	24.27	24.20	24.22	24.43	24.30	24.46
25.03	24.72	24.59	24.44	24.25	24.26	24.26	24.32	24.30	24.43
24.94	24.75	24.50	24.45	24.28	24.21	24.20	24.29	24.23	24.37
24.97	24.83	24.48	24.42	24.23	24.28	24.17	24.25	24.21	24.45
24.81	24.65	24.45	24.38	24.27	24.25	24.15	24.31	24.23	24.31
24.87	24.81	24.41	24.40	24.31	24.30	24.23	24.21	24.24	24.35
24.86	24.70	24.45	24.31	24.26	24.32	24.23	24.35	24.38	24.41
24.75	24.80	24.49	24.43	24.27	24.30	24.24	24.35	24.39	24.40

product is given in Table 2. For a lower specification limit of 24 % (say), there are no apparent nonconforming measurements and hence all the ten lots will be accepted under an attribute plan. On the other hand, the δ values for the ten lots are (0.0006, 0.0065, 0.1071, 1.8414, 2.1028, 2.0893, 2.4988, 2.0521, 1.5759, 0.7296) for the known error distribution $N(0, 0.2)$ (say). Evidently the overall nonconformance as measured by δ is different for the ten lots. The first lot was subject to heavy guard-banding while the quality of lots 6 to 8 is somewhat closer to the lower specification suggesting minimal guard-banding ($\delta > 2$). If a broken acceptance number of 1.6 (say) is set for the δ statistic, then only 5 of the 10 lots will be accepted. It should also be noted that the fractional and broken acceptance number plans do not require a randomization procedure as originally contemplated by Hamaker (1950). The measurement error distribution acts as a natural randomizer resulting in a nonconformance probability, particularly when the measurement is closer to the specification.

If we denote the apparent measurement of the true (unobserved) quality characteristic X by Y, then $Y = X + Z$ where Z is the measurement error, assumed to have a known distribution function $F_Z(\cdot)$. For a given lower specification limit L, the probability that $X < L$ conditional on the observed value Y is

$$p_Y = Pr(X < L|Y) = Pr(Z > Y - L|Y) = 1 - F_Z(Y - L).$$

Suppose now that, following the usual procedure for a randomized test, we choose with probability p_Y to decide that this observation is nonconforming. Then the binary decision variable A, with values 0 (conforming) and 1 (nonconforming), follows a Bernoulli distribution:

$$A|Y \sim \text{Bernoulli}(p_Y).$$

The unconditional distribution of A will also be Bernoulli, with probability

$$p^* = \int [1 - F_Z(Y - L)]dF_Y$$

where $F_Y(.)$ is the (unknown) distribution of the observed measurements. Thus A has mean p^* and variance $p^*(1 - p^*)$. Now consider the unconditional distribution of the fractional nonconformance p_Y. We have

$$E[A] = E\,[E[A|Y]] = E[p_Y]$$

and

$$V[A] = V\,[E[A|Y]] + E\,[V[A|Y]] = V[p_Y] + E[p_Y(1 - p_Y)].$$

Thus the fractional nonconformance p_Y has the same mean, but a smaller variance, than the randomized nonconformance A of Hamaker (1950). The same will be true for the total nonconformance from a random sample of size n. The randomized nonconformance $\sum_{i=1}^{n} A_i$ will have a binomial distribution, so for this the R codes given in the Appendix can be used to accurately control the producer's and consumer's risks at desired levels. However, if the decision is based on the total fractional nonconformance $\delta = \sum_{i=1}^{n} p_Y$, without randomization, then using the same sample size and acceptance number will, because of the smaller variance, be conservative leading to smaller producer's and consumer's risks than planned. Table 3 compares the probability acceptance values of the plan based on the δ statistic with the plan based on the randomized conformance procedure of Hamaker (1950) for a sample of size $n = 10$. The fractional acceptance number of 0.15 is set for the Hamaker (1950) plan applied on the observed (random) Y data from the standard normal distribution $N(0, 1)$. Assuming $N(0, 0.3)$ distribution for the measurement errors (Z), the fractional acceptance number of 0.65 is employed for the δ statistic. Both plans achieve a producer's risk of about 5 % at $p = 0.005$.

Table 3 Comparison of fractional nonconformance plan with Hamaker's randomized *Ac* plan

True p	$P_a(p)$ (Randomized $Ac = 0.15$)	$P_a(p)$ ($\delta \leq 0.65$ plan)
0.001	0.9870	0.9889
0.002	0.9760	0.9795
0.005	0.9455	0.9499
0.007	0.9243	0.9290
0.010	0.8947	0.8981
0.020	0.8097	0.8036
0.030	0.7299	0.7117
0.050	0.6048	0.5655
0.070	0.4964	0.4387
0.10	0.3705	0.2969
0.20	0.1329	0.0748
0.30	0.0423	0.0153

Table 4 Fractional nonconformance plan matching Hamaker's randomized *Ac* plan

AQL	LQL/AQL	n	Ac	Producer's risk	Consumer's risk
$X \sim N(0,1)$, $Z \sim N(0,0.3)$					
0.005	35	11 (12.89)	0.75 (1.10)	0.046	0.098
0.005	25	17 (21.53)	0.9 (1.26)	0.050	0.087
0.005	15	40 (62.45)	1.3 (1.97)	0.043	0.042
0.010	10	35 (48.86)	1.8 (2.42)	0.043	0.058
0.010	6	62 (94.71)	2.5 (3.33)	0.047	0.090
0.010	4	130 (212.35)	4.3 (5.51)	0.040	0.094
$X \sim G(1,2)$, $Z \sim N(0,0.3)$					
0.005	35	11 (12.06)	0.75 (1.08)	0.046	0.111
0.005	25	17 (17.82)	0.9 (1.12)	0.060	0.113
0.005	15	40 (51.76)	1.3 (2.10)	0.023	0.097
0.010	10	35 (40.57)	1.8 (2.32)	0.035	0.102
0.010	6	62 (75.36)	2.5 (3.32)	0.026	0.194
0.010	4	130 (160.64)	4.3 (5.47)	0.014	0.270

However the δ statistic provides better discrimination at other good and poor quality levels when compared to the randomized procedure of Hamaker (1950). Table 4 provides matched sets of broken and fractional acceptance numbers plans based on the δ statistic proposed in this paper with the randomized conformance plans (in brackets) of Hamaker (1950) for various AQL and LQL/AQL values assuming $X \sim N(0,0)$ as well as $X \sim G(1,2)$ for $Z \sim N(0,0.3)$ error distribution. Evidently the smaller variance associated with the δ statistic leads to a smaller sample size in general. The $G(1,1)$ distribution is right skewed towards the upper specification limit and hence results in higher producer's and consumer's risks. In other words, the δ statistic will lead to smaller risks when compared to the Hamaker (1950) plan depending on how large $E[p_Y(1-p_Y)]$ is.

4 Measurement Error Correction for Variables Inspection

A parametric approach is taken in the variables sampling inspection plan literature to correct for the measurement errors. The apparent measurements (Y) subject to errors (Z) are assumed to follow a probability distribution such as $N(\epsilon, \tau)$ instead of the $N(0, 1)$ distribution for the true measurements X. In the presence of (instrument) bias in measurements, the condition $\epsilon \neq 0$ holds. When apparent measurements are imprecise, the condition $\tau \neq 1$ holds. Assuming that Y and Z are normally distributed, Mei et al. (1975), Owen and Chou (1983), Basnet and Case (1992), Fang and Zhang (1995), Wilrich (2000), and Melgaard and Thyregod (2001) studied the single sampling variables plans. Mei et al. (1975), Owen and Chou (1983), and Wilrich (2000) suggested an increase in sample size to mitigate the effect of inaccuracy in measurements. The acceptability constant k is also adjusted for a known level of bias. Grzegorzewski (2002) considered fuzzy theory for modelling imprecise variables data. This approach also requires the fuzzy bounds to be prefixed.

The case of non-normal errors for the variables sampling plan has not received much attention because the convolutions of normal and other non-normal distribution cannot be obtained analytically. The estimation of the true distribution of measurements can be made by the deconvolution process given the apparent measurements and the error distribution, see Hazelton and Turlach (2009) and Hazelton and Turlach (2010). Estimation of tail areas using non-parametric kernel density estimation methods needs reasonably large sample sizes and is also more complex to administer. A simple procedure of estimating the true fraction nonconforming p from the apparent fraction nonconforming p_e is desirable.

Consider the fictitious data discussed in Sect. 1. Assuming that the quality characteristic follows a normal distribution, the estimate of the apparent fraction nonconforming is found as $\hat{p} = 0.13$. After correcting for the measurement errors using the triangular error distribution but without the normal assumption, this estimate becomes $\hat{p} = 0.064$ (see Sect. 3). The estimated excess fraction nonconforming due to the measurement errors is of the order 0.06, which is not small. A formal treatment to the problem of estimating the true fraction nonconforming can be given if X and Z are known to be independent and reproductive random variables such as normal. Let $X \sim N(\mu_X, \sigma_X^2)$ and $Z \sim N(0, \sigma_Z^2)$. This implies that $Y \sim N(\mu_X, \sigma_X^2 + \sigma_Z^2)$, assuming that there is no instrument bias or $\mu_X = \mu_Y$. Let $\tau_X = 1/\sigma_X^2$ and $\tau_Z = 1/\sigma_Z^2$. The conditional probability density $f(Z|Y = y) = \frac{f_X(y-z) f_Z(z)}{f_Y(y)}$ is proportional to

$$\exp\left[-\frac{1}{2}\left(\tau_Z z^2 + \tau_X (y - z - \mu_X)^2\right)\right].$$

After simplification, we obtain

$$f\,(Z|Y=y) \propto \exp\left[-\frac{1}{2}\left((\tau_Z+\tau_X)\left(z-\frac{\tau_X}{\tau_X+\tau_Z}(y-\mu_X)\right)^2\right)\right].$$

That is, the measurement error Z conditional on the given apparent measurement value y also follows the normal distribution $N\left(\frac{\tau_X}{\tau_X+\tau_Z}(y-\mu_X), \frac{1}{\tau_X+\tau_Z}\right)$ or $N\left(k(y-\mu_X), (1-k)\sigma_Z^2\right)$ where $k=\sigma_Z^2/\sigma_Y^2$. Here k is the ratio of measurement error variance to the total observed variance, which is in the range 0.1 to 0.5 for many bulk materials, particularly when measurements are made by analytical methods.

For a given upper specification limit U, $P[x>U|Y=y]=P[z<y-U]=\Phi(\omega_U)$ where

$$\omega_U=\left(\frac{(y-U)-k(y-\mu_X)}{\sigma_Z\sqrt{1-k}}\right)$$

If we estimate $\mu_X\ (=\mu_Y)$ by $\bar{y}$, then $\hat{p}_N=\frac{1}{n}\sum_1^n\Phi(\widehat{\omega_U})$, where $\widehat{\omega_U}=\left(\frac{(y-U)-k(y-\bar{y})}{\sigma_Z\sqrt{1-k}}\right)$, gives the estimated true fraction nonconforming in a sample of size n while $\Phi(\widehat{\omega_U})$ gives the fractional nonconforming estimate for given U, k, and σ_Z. Without loss of generality, let $\sigma_X^2=1$, $\sigma_Y^2=\frac{1}{1-k}$ and $\sigma_Z^2=\frac{k}{1-k}$ leading to $\widehat{\omega_U}=\left(\frac{(y-U)-k(y-\bar{y})}{\sqrt{k}}\right)$.

The traditional estimate of the true fraction nonconforming is $\widehat{p_U}=\Phi(\kappa_U)$ where $\kappa_U=\left(\frac{\bar{x}-U}{\sigma_x}\right)=(\bar{x}-U)$. This estimate is recommended by Hahn (1982) as a way of adjusting for the measurement error when the repeatability standard deviation σ_Z is known. It should be noted that $\widehat{p_U}$ estimate is not based on the individual estimates of fractional nonconformance of individual units. A comparison of the smoothed density of $\widehat{p_U}$ and $\hat{p}_N$ (obtained by simulation) for $n=30$, $p=0.03$, and $k=0.5$ is shown in Fig. 1. The traditional $\widehat{p_U}$ estimate is more variable than the $\hat{p}_N$ estimate. We are not particularly concerned with the point estimation properties such as unbiasedness in acceptance sampling. This is because an estimate of the fraction nonconforming is finally compared with the maximum allowable fraction nonconforming for lot acceptance/rejection decisions, and hence any bias in the estimation of p cancels out. However the precision of the estimate will affect the producer's and consumer's risks. This issue is discussed using an example in the next paragraph.

Let the maximum allowable fraction nonconforming p_* be 0.05. That is, the lot is accepted when $\widehat{p_U}\le p_*$ and $\widehat{p_N}\le p_*$ under the two methods of operation (estimation). By varying the unknown fraction nonconforming p, the probability

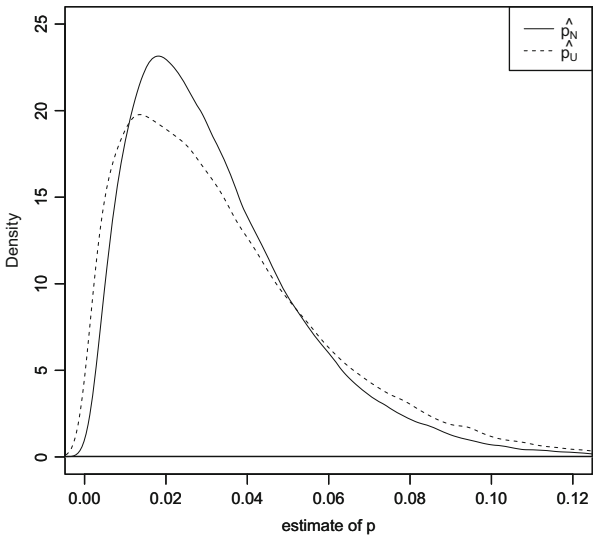


Fig. 1 Comparison of $\widehat{p_U}$ and $\hat{p}_N$ estimates

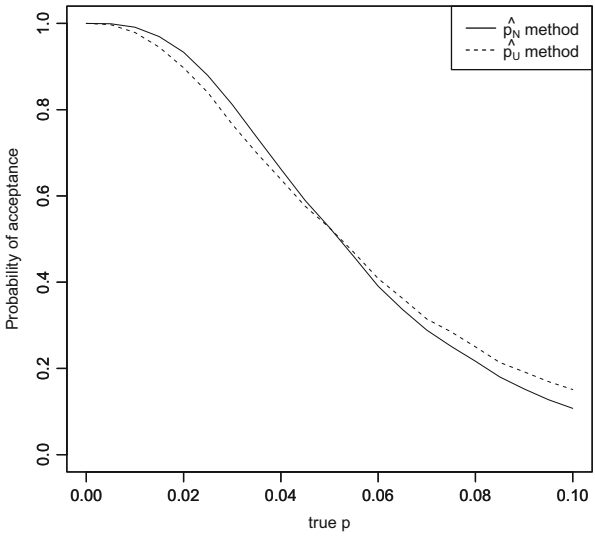


Fig. 2 OC curves of $\widehat{p_U} \leq p_*$ and $\hat{p}_N \leq p_*$ Plans

of acceptance of the lot can be found as the proportion of lots accepted using simulation methods. For example, let $n = 30$ and $k = 0.5$. Figure 2 shows that the $\widehat{p_N} \leq p_*$ method of estimation and operation of the variables plan results in a more discriminatory OC curve when compared to the $\widehat{p_U} \leq p_*$ plan.

5 Summary and Conclusions

In the presence of measurement errors, an individual measurement cannot be regarded as conforming or not with certainty. If the measurement error distribution is known, the probabilities of conformance and nonconformance of a measurement become relevant. For a given sample of n measurements, the sum of probabilities of nonconformance can be used to assess the lot acceptance based on fractional or broken acceptance numbers. Fractional acceptance numbers are particularly suitable for practical applications where the fraction nonconforming is expected to be small. A fully parametric approach based for fractional nonconformance is also introduced. The fractional and broken acceptance number plans are particularly useful for lots produced by short run production processes.

Acknowledgements The constructive suggestions of the Reviewer are gratefully acknowledged. Thanks to Mr. Roger Kissling and Dr. Haizhen Wu for their comments. This work is partly supported by the Primary Growth Partnership (PGP) scheme of Fonterra Co-operative Group and New Zealand Government.

Appendix

```
p1=0.01; p2=0.08; alpha=0.05; beta=0.1
require("BB")
fn <- function(p) {
  r <- rep(NA, length(p))
  r[1] <- pbeta(p1, exp(p[1])+1, exp(p[2])+1)-alpha
  r[2] <- pbeta(p2, exp(p[1])+1, exp(p[2])+1)-(1-beta)
  r}
p0 <- c(2,2)
res = BBsolve(par = p0, fn = fn, control = list(trace = FALSE))
(a= exp(res$par[1])+1)
(b= exp(res$par[2])+1)
```

References

Abbasi, S. A., & Miller, A. (2012). On proper choice of variability control chart for normal and non-normal processes. *Quality and Reliability Engineering International*, *28*, 279–296.

Ahmed, S. (2005). Assessing the process capability index for non-normal processes. *Journal of Statistical Planning and Inference*, *129*, 195–206.

Balakrishnan, N., & Kocherlakota, S. (1986). Effects of nonnormality on $\bar{X}$ charts single assignable cause model. *Sankhya: The Indian Journal of Statistics, Series B (1960–2002)*, *48*, 439–444.

Basnet, C., & Case, K. E. (1992). The effect of measurement error on accept/reject probabilities for homogeneous products. *Quality Engineering*, *4*, 383–397.

Boullion, T. L., Cascio, G. C., & Keating, J. P. (1985). Comparison of estimators of the fraction defective in the normal distribution. *Communications in Statistics - Theory and Methods*, *14*, 1511–1529.

Brown, G. G., & Rutemiller, H. C. (1973). The efficiencies of maximum likelihood and minimum variance unbiased estimators of fraction defective in the normal case. *Technometrics*, *15*, 849–855.

Burr, I. W. (1967). The effect of non-normality on constants for $\bar{X}$ and R charts. *Industrial Quality Control*, *23*, 563–569.

Clements, J. A. (1989). Process capability calculations for non-normal distributions. *Quality Progress*, *22*, 98–100.

Collani, E. V. (1991). A note on acceptance sampling for variables. *Metrika*, *38*, 19–36.

Das, N. G., & Mitra, S. K. (1964). Effect of non-normality on plans for sampling inspection by variables. *Sankhya: The Indian Journal of Statistics, Series A (1961–2002)*, *26*, 169–176.

Duncan, A. J. (1986). *Quality Control and Industrial Statistics* (5th Ed.). Homewood, Illinois: Richard D. Irwin Inc.

Fang, X. D., & Zhang, Y. (1995). Adjusting plans of acceptance sampling based on sample mean for undesired measurement conditions. *Quality Engineering*, *8*, 57–73.

Govindaraju, K. (1991). Fractional acceptance number single sampling plan. *Communications in Statistics - Simulation and Computation*, *20*, 173–190.

Grzegorzewski, P. (2002). Acceptance sampling plans by variables for vague data. In P. Grzegorzewski, O. Hryniewicz, & M. Gil (Eds.), *Soft methods in probability, statistics and data analysis*. Advances in Intelligent and Soft Computing (Vol.16 pp. 197–206). Heidelberg: Physica-Verlag HD.

Guenther, W. C. (1971). A note on the minimum variance unbiased estimate of the fraction of a normal distribution below a specification limit. *The American Statistician*, *25*, 18–20.

Hahn, G. J. (1982). Removing measurement error in assessing conformance to specifications. *Journal of Quality Technology*, *14*, 117–121.

Hamaker, H. C. (1950). Some notes on lot-by-lot inspection by attributes. *Review of the International Statistical Institute*, *18*, 179–196.

Hazelton, M., & Turlach, B. (2009). Nonparametric density deconvolution by weighted kernel estimators. *Statistics and Computing*, *19*, 217–228.

Hazelton, M. L., & Turlach, B. A. (2010). Semiparametric density deconvolution. *Scandinavian Journal of Statistics*, *37*, 91–108.

Hryniewicz, O. (2008). Statistics with fuzzy data in statistical quality control. *Soft computing*, *12*, 229–234.

ISO 2859-1:1999(E) (1999). *Sampling schemes indexed by Acceptance Quality Limit (AQL) for lot-by-lot inspection, Part 1*. Geneva, Switzerland: International Organization for Standardization.

Jamkhaneh, E. B., & Sadeghpour-Gildeh, B. & Yari, G. (2011). Inspection error and its effects on single sampling plans with fuzzy parameters. *Structural and Multidisciplinary Optimization*, *43*, 555–560.

Johnson, N. L., Kotz, S., & Wu, X. (1991). *Inspection errors for attributes in quality control*. New York: Chapman and Hall.

Kotz, S., & Johnson, N. (1993). *Process Capability Indices*. London: Chapman and Hall.

Lavin, M. (1946). Inspection efficiency and sampling inspection plans. *Journal of the American Statistical Association*, *41*, pp. 432–438.

Lieberman, G. J., & Resnikoff, G. J. (1955). Sampling plans for inspection by variables. *Journal of the American Statistical Association*, *50*, 457–516.

Md-Yusof, S., & Rigdon, S. (1992). A comparison of estimators for the proportion in the tail of a normal distribution. *Statistical Papers*, *33*, 33–38.

Mei, W. -H., Case, K. E., & Schmidt, J. W. (1975). Bias and imprecision in variables acceptance sampling effects and compensation. *International Journal of Production Research*, *13*, 327–340.

Melgaard, H., & Thyregod, P. (2001). Acceptance sampling by variables under measurement uncertainty. In H.-J. Lenz, & P.-T. Wilrich (Eds.), *Frontiers in statistical quality control* (Vol.6, pp. 47–57). Heidelberg: Physica-Verlag.

Montgomery, D. C. (1985). The effect of nonnormality on variables sampling plans. *Naval Research Logistics Quarterly*, *32*, 27–33.

Ohta H., & Ichihashi H. (1988). Determination of Single-Sampling Attribute Plans Based on Membership Functions. *International Journal of Production Research*, *26*, 1477–1485.

Owen, D., & Chou, Y. (1983). Effect of measurement error and instrument bias on operating characteristics for variables sampling plans. *Journal of Quality Technology*, *15*, 107–117.

Owen, D. B. (1969). Summary of recent work on variables acceptance sampling with emphasis on non-normality. *Technometrics*, *11*, 631–637.

R Development Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. ISBN 3-900051-07-0.

Schilling, E. G., & Nelson, P. R. (1978). The effect of nonnormality on the control limits of $\bar{X}$ charts. *Journal of Quality Technology*, *8*, 183–188.

Schilling, E. G., & Neubauer, D. V. (2009). *Acceptance sampling in quality control*. (2nd Ed.). Boca Raton, FL: Chapman & Hall.

Schneider, H., & Wilrich, P. (1981). The robustness of sampling plans for inspection by variables. In H. B. Ning, & P. Naeve (Eds.), *Computational Statistics: Wolfgang Wetzel Zur Vollendung Seines 60.* (pp. 281–295). Walter de Gruyter. ISBN-13: 9783110084191.

Seidel, W. (1997). Is sampling by variables worse than sampling by attributes? A decision theoretic analysis and a new mixed strategy for inspecting individual lots. *Sankhya: The Indian Journal of Statistics, Series B*, *59*, 96–107.

Singh, H. R. (1966). Producer and consumer risks in non-normal populations. *Technometrics*, *8*, 335–343.

Somerville, S. E., & Montgomery, D. C. (1996). Process capability indices and non-normal distributions. *Quality Engineering*, *9*, 305–316.

Takagi, K. (1972). On designing unknown-sigma sampling plans based on a wide class of non-normal distributions. *Technometrics*, *14*, 669–678.

Wilrich, P.-T. (2000). Single sampling plans for inspection by variables in the presence of measurement error. *AStA. Allgemeines Statistisches Archiv*, *84*, 239–250.

Woods, W. M. (1955). *Variables sampling inspection procedures which guarantee acceptance of perfectly screened lots*. Technical Report 47 Applied Mathematics and Statistics Laboratory, Stanford University, California. http://statistics.stanford.edu/~ckirby/reports/1950_1959/reports1955.html.

Young, G., & Smith, R. (2005). *Essentials of statistical inference*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

Sampling Plans for Control-Inspection Schemes Under Independent and Dependent Sampling Designs with Applications to Photovoltaics

Ansgar Steland

Abstract The evaluation of produced items at the time of delivery is, in practice, usually amended by at least one inspection at later time points. We extend the methodology of acceptance sampling for variables for arbitrary unknown distributions when additional sampling information is available to such settings. Based on appropriate approximations of the operating characteristic, we derive new acceptance sampling plans that control the overall operating characteristic. The results cover the case of independent sampling as well as the case of dependent sampling. In particular, we study a modified panel sampling design and the case of spatial batch sampling. The latter is advisable in photovoltaic field monitoring studies, since it allows to detect and analyze local clusters of degraded or damaged modules. Some finite sample properties are examined by a simulation study, focusing on the accuracy of estimation.

Keywords Acceptance sampling • Dependence • Quality control • Renewable energies • Sampling design

1 Introduction

The acceptance sampling problem deals with the construction of sampling plans for inspection, in order to decide, using a minimal number of measurements, whether a lot (or shipment) of produced items should be accepted or rejected. Our approach is motivated by applications in photovoltaics, where the distribution of the relevant quality features, in particular the power output of solar panels, is typically non-normal, unknown and cannot be captured appropriately by parametric models. However, in photovoltaics additional measurements from the production line are available and can be used to construct acceptance sampling plans.

A. Steland (✉)
Institute of Statistics, RWTH Aachen University, Aachen, Germany
e-mail: steland@stochastik.rwth-aachen.de

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_18

Let X represent a control measurement of produced item, with distribution function F that possesses a finite fourth moment, e.g. the power output of a photovoltaic module. It is classified as non-conforming, defective or out-of-spec, if $X \leq \tau$, where τ is a (one-sided) specification limit, usually defined as $\tau = \mu^*(1-\varepsilon)$, where μ^* is the target (or nominal) mean and $\varepsilon \in (0,1)$ the tolerance. If there were no randomness, $X = \mu$, where $\mu = E(X)$ is the true mean of X. Then items are non-conforming if $\mu - \tau \leq 0$ and, clearly, we should reject the lot, if and only if $\mu - \tau \leq 0$. But if the distribution of X is not degenerated, it is reasonable to replace μ by its unbiased canonical estimator $\hat{\mu} = \overline{X}$ to form a decision rule, and thus to reject the lot if $\overline{X} - \tau \leq c$, where the critical value $c > 0$ accounts for the estimation error.

The fraction of non-conforming (out-of-spec) modules (or items) corresponding to the above definition is then given by

$$p = P(X_1 \leq \tau) = F(\tau).$$

It is usually regarded as the quantity of interest in quality control, although it makes no assertion about how far away from the specification the non-conforming items are. However, it is worth mentioning that the fraction of non-conforming modules is directly related to the resulting costs, since it determines the number of modules one has to repair or replace in case of a total inspection. For these reasons, one aims at the determination of control procedures that allow to infer whether the fraction p, also called quality level, is acceptable, i.e. $p <$ AQL, or not, i.e. $p >$ RQL. Here $0 <$ AQL $<$ RQL < 1 denote the *acceptance quality limit* (AQL) and the *rejectable quality limit* (RQL).

It is known that the probability of acceptance for the above rule based on $\overline{X} - \tau$ is a function of the fraction defectives p, but it depends on the unknown distribution function F. In photovoltaics and presumably other areas as well, we are given additional data from the production line that can be used to estimate unknowns.

In this paper, we construct sampling plans for the following situation: We assume that a control sample is drawn at the time of delivery of the modules, i.e., when the modules are new and unused, in order to ensure that shipments that are out-of-spec are identified and are not delivered to the customers. The sampling plan used at this first stage of the procedure is constructed using an additional sample from the production line, as discussed above. We further assume that shipments that passed the first-stage acceptance sampling procedure are inspected at a later time point, in order to check whether they are still in agreement with the quality requirements after some defined period of operation. At this second stage a further sample is taken, which is combined with the sample information from the first stage. This is done in order to avoid that a close decision in favor of acceptance, i.e. when the first stage control statistic has attained a value close to the critical value, results again in an acceptance due to the fact that the second stage sample size is relatively small.

At first glance, it seems that the approach is a double sampling procedure. Recall that the idea of double sampling plans is to give a questionable lot a second chance or, put differently, allow for quick acceptance of very good lots: If the test statistic (say, e.g., the number of defectives) is small, say smaller than a, one accepts, if it is too large, say larger than b, one rejects, and if it is between a and b one draws a second sample and bases the decision on the enlarged sample. Our approach is related in that we aim at re-using the sample information from the first-stage control sample, but in our approach the second sample is not taken at the same time instant, but at the inspection time, and the decision to accept or reject at the first stage is only based on the first sample. Indeed, we have in mind that there may be a substantial time lag between the two stages.

The fact that in practice one prefers to take repeated measures at inspection time, i.e. of those items already selected, complicates the design of appropriate procedures, since now the samples at different time points are not independent. Thus, we extend the required theoretical results to the case of dependent sampling under quite general conditions. We propose a panel-based sampling scheme, where the items selected at the first stage form the basis of the second stage sample, which is enlarged by new items if necessary. A further important sampling design is spatial batch sampling. Here the batches of observations may be correlated, for example since their spatial closeness implies that they carry the same factors that may affect quality measurements. We show that our results are general enough to apply to such a sampling design as well.

The rest of the paper is organized as follows. Section 2 discusses related work and applications. Section 3 introduces the two-stage acceptance sampling framework, in particular the operating characteristic curves that define the statistical behavior of our two-stage procedure, and discusses our model for the two-stage setting with a control sample, an inspection sample and an additional sample from the production line. In Sect. 4, we provide the asymptotic results that allow us to construct valid acceptance sampling plans that control the overall operating characteristic curve. Those results cover expansions of the control statistics, their joint asymptotic normality, and approximations of the operating characteristics. We provide results for the case that the control sample and the inspection sample are independent as well as for the more general and realistic case that the samples are dependent. Computational issues are discussed in Sect. 5. Lastly, Sect. 6 presents results from a simulation study.

2 Preliminaries

2.1 Related Work

The acceptance sampling problem dates back to the seminal contributions of Dodge and has been studied since then to some extent. For a general overview

of classical procedures and their implementations in standards, we refer to recent monograph Schilling and Neubauer (2009). For the Gaussian case, optimal plans have been constructed by Liebermann and Resnikoff (1955), see also Brun-Suhr and Krumbholz (1991), Feldmann and Krumbholz (2002), where the latter paper studies double sampling plans for normal and exponential data, and the references given there. Their lack of robustness with respect to departures from normality has been discussed in Kössler and Lenz (1997). Kössler (1995) used a Pareto-type tail approximation of the operating characteristic combined with maximum likelihood estimation, in order to estimate the fraction of defectives and then constructed sampling plans using the asymptotic distribution of that estimate, when the lot is accepted if the estimated fraction of defectives is too small. The methods works, if the tails are not too short. Since in industrial applications large production lots are usually classified in classes, the case of non-normal but compactly supported distributions deserves attention. For such distributions, approximations based on the asymptotic normality of sample means are a convenient and powerful tool for the construction of sampling plans, having in mind that t-type statistics are a natural choice to decide in favor of acceptance or rejection of a lot, as discussed above. Thus, recent works focused on t-type test statistics resembling the statistic used by the optimal procedure under normality.

Sampling plans for variables inspection when the underlying quality variable has an arbitrary continuous distribution with finite fourth moment and the related estimation theory based on the sample quantile function of an additional sample has been studied in Steland and Zähle (2009) employing empirical process theory. For historic samples, i.e. samples having the same distribution as the control sample, a simplified proof using the Bahadur representation can be found in Meisen et al. (2012). In the present work, it is shown that this method of proof extends to the case of a difference in location between the additional sample and the control sample as studied in Steland and Zähle (2009). Further results and discussions on acceptance sampling for photovoltaic data and applications can be found in Herrmann et al. (2006, 2010). Herrmann and Steland (2010) have shown that the accuracy of such acceptance sampling plans using additional samples from the production line can be substantially improved by using smooth quantile estimators such as numerically inverted integrated cross-validated kernel density estimators. The construction of procedures using the singular spectral analysis (SSA) approach with adaptively estimated parameters has been recently studied by Golyandina et al. (2012). Bernstein-Durrmeyer polynomial estimators are well known as a general purpose approach to estimation that provides quite smooth estimators. The relevant theory as well as their application to the construction of acceptance sampling plans with one-sided specification limits has been investigated in Pepelyshev et al. (2014a). The extension of the methodology to two-sided specification limits and numerical results focusing on photovoltaic applications can be found in Pepelyshev et al. (2014b).

2.2 Applications in Photovoltaics

The quality control of photovoltaic systems has become a key application area for recent developments and extensions of the acceptance sampling methodology, although the results can certainly be adopted to many other areas in industry.

The production of solar panels has become a highly complex high-throughput production process. Today's cell technologies rely on sophisticated solar cell designs. A solar cell can be regarded as a stack of thin layers, in order to trap as many photons as possible, transform them into electron-hole pairs and then ease the electrons' movement through the cell to the wires. Anti-reflective coatings of the glass covering have been introduced recently, in order to maximize the amount of sunlight trapped by the solar cells by channelling the photons to the lower layers of the cell. Optical filters are used in order to ensure that only those wavelengths pass that can be processed by the semiconductor to form electron-hole pairs. Let us briefly recall how a solar cell works: The p-type silicon layer consists of silicon, which has four electrons, doped with a compound (such as Phosphorous) that contains one more valence electron, such that this layer is positively charged. The n-type silicon layer is silicon doped with compounds (such as Boron) having one less valence electron than silicon, such that only three electrons are available for binding with four adjacent silicon atoms. Thus, the n-type layer is negatively charged. An incomplete bond (hole) of the n-type layer can attract an electron to fill the hole, in which case the hole moves. At the np-junction where both layers meet, electrons from the n-type layer being freed by the photons' energy move to the p-layer and from there to the back contact, and the corresponding holes move to the contact grid at the top of the cell. This results in a current I. Combined with the internal electrical field of the cell due to the differently charged p- and n-silicon layers leads to power ($P = UI$), which can be used by an external load attached to the cell.

Each layer of a cell makes use of specific physical and chemical properties of the base material and the added compounds. In a multijunction (tandem) design two cells are mechanically stacked on top of each other. The second cell at the bottom absorbs the higher energy photons not absorbed by the top cell. Such designs can increase the efficiency substantially. The physical and chemical interaction of those materials and particles is a complex dynamic process driven by the sun's irradiance, the associated heat and the weather conditions that may range from extreme cold to extreme heat. Even in the absence of manufacturing faults, these facts cause serious changes of the physical and chemical properties due to ageing, and, as a consequence, also of the electrical properties of a solar cell, leading to what could be called degradation by design.

Manufacturing faults add to those unavoidable sources of degradation. For example, even minor defects in the encapsulation of a PV module may result in leakage after a couple of years, thus leading to internal corrosion and other effects that degrade or even destroy the module. Micro cracks in the crystalline semiconductor arising by improper handling of the modules at the production line, stress during transport to the site of construction or improper handling during assembly of the

photovoltaic system, are invisible by eye but visible in electro luminescence (EL) images. Usually they have no effect on the electrical characteristics. However, it is conjectured that such micro cracks could have serious impact on long-term degradation and result in failures after several years of operation. As a consequence, insurance companies routinely make EL images from insured lots or systems and are highly interested in the long-term influence of micro cracks. Antireflective coatings have been shown experimentally to degrade after damp-heat tests leading to a loss of power. Driven by the high potential between the cell's surface and the ground, a likely source of potential induced degradation (PID) is the wandering of Na^+ ions from the glass surface through the cell to the *np*-junction, where they short-circuit the emitter. The fact that the emerging markets for photovoltaics are in countries such as India or Saudi Arabia, the degradation of the glass surface due to sand is an important issue for the reliability of PV systems.

As a consequence, there is a need for proper inspection plans that combine available information from the production line, quality assessments and audits at the time of delivery and construction of the solar systems and later inspections.

3 Method

3.1 Two-Stage Acceptance Sampling

To proceed, let us fix some notions more rigorously. Let T_n be a statistic (decision function) depending on a sample $X_1, \ldots, X_n$ constructed in such a way that large values of $\overline{X}_n - \tau$ indicate that the lot should be accepted. A pair $(n, c) \in \mathbb{N} \times [0, \infty)$ is called a *(acceptance) sampling plan*, if one draws a sample of n observations and accepts the lot if $\overline{X}_n - \tau > c$. Then the probability that the lot is accepted,

$$\mathrm{OC}(p) = P(T_n > c|p), \qquad p \in [0, 1],$$

is called *operating characteristic*. Here $P(\bullet|p)$ indicates that the probability is calculated under the assumption that the true fraction of non-conforming items equals p. Given specifications of the AQL and RQL and error probabilities α and β, a sampling plan is called *valid*, if

$$\mathrm{OC}(p) \geq 1 - \alpha, \qquad \text{for all } p \leq \mathrm{AQL}, \tag{1}$$

and

$$\mathrm{OC}(p) \leq \beta, \qquad \text{for all } p \geq \mathrm{RQL}. \tag{2}$$

In this article, we consider two-stage acceptance procedures where a lot is examined at two time points. At time t_1, usually the time of production, delivery

or construction of the system that uses the delivered items (PV modules), a control sample is taken, in order to decide whether the lot or shipment can be accepted. If the lot is rejected, we stop. If the lot is accepted, one proceeds and at time instant t_2 the system is inspected again. One applies a further acceptance sampling plan, based on an inspection sample, in order to conclude whether the shipment is still in agreement with the specifications.

Let us denote the test statistic used at time t_i using a sampling plan (n_i, c_i) by $T_{ni} = T_{n_i,i}$, $i = 1, 2$. Notice that here and in what follows, with some abuse of usual notation, we indicate the dependence on n_i by n, in order to keep notation simple and clean; this will cause neither confusion nor conflict.

Then the corresponding operating characteristics are given by

$$\mathrm{OC}_1(p) = P(T_{n1} > c_1|p), \qquad p \in [0, 1],$$

and, since the sampling plan (n_2, c_2) is constructed given $T_{n1} > c_1$,

$$\mathrm{OC}_2(p) = P(T_{n2} > c_2|T_{n1} > c_1, p), \qquad p \in [0, 1].$$

Since a lot is accepted if and only if it is accepted at stage 1 and stage 2, the *overall operating characteristic*, $\mathrm{OC}(p) = P(\text{'lot acceptance'}|p)$, is given by

$$\mathrm{OC}(p) = \mathrm{OC}_1(p)\mathrm{OC}_2(p), \qquad p \in [0, 1]. \tag{3}$$

Of course, one may design the procedure such that at both stages the operating characteristics are valid for the same error probabilities. However, then one cannot control the error probabilities of the overall procedure, since its operating characteristic is given by (3).

Thus, we propose to design the procedure by controlling the overall operating characteristics. This means, the stage-wise sampling plans are determined in such a way that both of them ensure (1) and (2) for stage-specific error probabilities α_1, β_1 and α_2, β_2, i.e.

$$\mathrm{OC}_i(p) \geq 1 - \alpha_i, \qquad p \leq \mathrm{AQL}, \qquad (i = 1, 2), \tag{4}$$

and

$$\mathrm{OC}_i(p) \leq \beta_i, \qquad p \leq \mathrm{RQL}, \qquad (i = 1, 2). \tag{5}$$

If (4) and (5) can be ensured with equality for $p \in \{\mathrm{AQL}, \mathrm{RQL}\}$, then we obtain

$$\mathrm{OC}(\mathrm{AQL}) = (1 - \alpha_1)(1 - \alpha_2), \qquad \mathrm{OC}(\mathrm{RQL}) = \beta_1\beta_2$$

for the overall OC curve. If we want that it represents an (overall) valid acceptance sampling plan, i.e.

$$\mathrm{OC}(p) \geq 1-\alpha,\ p \leq \mathrm{AQL}, \quad \mathrm{OC}(p) \leq \beta,\ p \geq \mathrm{RQL}, \tag{6}$$

for given global error probabilities α and β, we have to design the procedures at both stages appropriately. Treating the producer risk and the consumer risk symmetrically imposes the constraints

$$\alpha_1 = \beta_1, \qquad \alpha_2 = \beta_2,$$

such that it remains to select α_1 and α_2 in such a way that the resulting procedures guarantees a valid overall sampling plan. For example, if one additionally imposes the constraint $\alpha_1 = \alpha_2$, one obtains a valid (overall) acceptance sampling plan, if (α_1, β_1) is selected such that $(1-\alpha_1)^2 = 1-\alpha, \beta_1^2 = \beta$. If now the global error probabilities α and β are given and one puts $\alpha_2 = \alpha_1 = 1-\sqrt{1-\alpha}$ and $\beta_1 = \beta_2 = \sqrt{\beta}$, then (6) holds with equalities, but the stage-wise consumer risks β_1, β_2 may be too high in practice—observe that $\sqrt{0.1} \approx 0.31623$. Another approach is to use plans with $\alpha_1 < \alpha_2$, such that the procedure is, in terms of the producer risk, more restrictive at the first stage than at the second stage. In the simulations, we specified $\alpha(= 0.1)$ and α_1, determined the corresponding α_2, i.e. $\alpha_2 = 1-\frac{1-\alpha}{1-\alpha_1}$, and put $\beta_i = \alpha_i, i = 1, 2$. This means, at each stage producer and consumer risks are symmetric. As a consequence, the procedure will work on a small global consumer risk, since $\beta = \beta_1\beta_2$ with $\beta_1, \beta_2 \leq \alpha_1$. For example, the choices $\alpha = 0.1, \alpha_1 = 0.03$ lead to $\alpha_2 = \beta_2 \approx 0.072$ such that $\beta \approx 0.00216$, yielding a valid acceptance sampling plan.

3.2 A Two-Stage Procedure Using Additional Data

We assume that we are given an additional data set of size m, usually quite large, consisting of independent and identically distributed measurements,

$$X_0, X_{01}, \ldots, X_{0m} \overset{i.i.d.}{\sim} F_0,$$

sampled at time $t_0 < t_1$, with

$$\text{mean } \mu_0 = E(X_0) \text{ and variance } \sigma_0^2 = \mathrm{Var}(X_0),$$

which can be used for the construction of the decision procedures. Recalling from our above discussion that those additional measurements usually represent historic data or are taken using a different measurement system, we allow for difference

in location with respect to the independent and identically distributed control measurements,

$$X_1, X_{11}, \ldots, X_{1n_1} \overset{i.i.d.}{\sim} F_1,$$

taken at time instant t_1. At the inspection time point t_2, we have additional measurements

$$X_2, X_{21}, \ldots, X_{2n_2} \overset{i.i.d.}{\sim} F_2.$$

Here we shall allow for a degradation effect leading to smaller measurements. Concretely, our distributional model relating the marginal distributions of the three samples is now as follows.

$$X_{ji} \sim \begin{cases} F(\bullet - \Delta), & j = 0, \\ F(\bullet), & j = 1, \\ F\left(\frac{\bullet}{d}\right), & j = 2. \end{cases} \tag{7}$$

for constants $\Delta \in \mathbb{R}$ and $0 < d < \infty$. Equivalently, in terms of equality in distribution,

$$\begin{aligned} X_0 &\overset{d}{=} \Delta + X_1, \\ X_1 &\sim F, \\ X_2 &\overset{d}{=} dX_1. \end{aligned}$$

The constant d determines the degree of degradation (if $d < 1$) and is assumed to be known. We work with a simple degradation model, since in photovoltaics there is not yet enough knowledge about the degradation of photovoltaic modules, which would justify to go beyond the assumption that degradation acts like a damping factor on the power output measurement. We also assume that d is known, since even the estimation of the mean yearly degradation is a difficult practical problem and requires large data sets over long time periods. Hence it is not realistic to estimate d within our framework.

Further, we may and shall assume that

$$F(x) = G\left(\frac{x - \mu}{\sigma}\right), \qquad x \in \mathbb{R},$$

for some fixed but unknown d.f. G with $\int x dG(x) = 0$ and $\int x^2 dG(x) = 1$, such that

$$\mu = E(X_1), \qquad \sigma^2 = \mathrm{Var}(X_1),$$

are the mean and the variance of the control measurements taken at time t_1.

The two-stage acceptance sampling procedure to be studied is now as follows. At stage 1, i.e. at time t_1, based on a sampling plan (n_1, c_1) we accept the lot or shipment, if

$$T_{n1} = \sqrt{n_1}\frac{\overline{X}_1 - \tau}{S_m} > c_1, \tag{8}$$

where $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$ is the average of the observations taken at time t_1 and

$$S_m = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(X_{0i} - \overline{X}_0)^2}$$

is the sample standard deviation calculated from the observation taken at time t_0. It is worth mentioning that standardizing with the sample standard deviation calculated from the time t_0 measurements is crucial; indeed, S_m cannot be replaced by, say, $\hat{\sigma}_1 = (n_1 - 1)^{-1}\sum_{j=1}^{n_i}(X_{1j} - \overline{X}_i)^2$.

If the lot is accepted at time t_1, we draw the additional observations $X_{21}, \ldots, X_{2n_2}$ for inspection and calculate the corresponding statistic

$$T_{n2} = \sqrt{n_2}\frac{D\overline{X}_2 - \tau}{S_m},$$

where $D = 1/d$ and $\overline{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}$. At inspection time t_2 the lot is accepted if

$$T_{n1} + T_{n2} > c_2. \tag{9}$$

Notice that here we aggregate the available information by summing up T_{n1} and T_{n2}. The rationale behind this rule is as follows. We reach the inspection time, only if we passed the quality control at time t_1. The value of the statistic T_{n1} comprises the evidence in favor of acceptance and rejection, respectively. But even if the lot is accepted, the decision could be a close one, i.e. $T_{n1} > c_1$ but $T_{n1} \approx c_1$. In such cases, the probability is relatively large that the lot is accepted at the inspection time again, if one drops the information already obtained at stage 1. Thus, it makes sense to aggregate all available information to come to a decision, i.e. to take the sum of the statistics T_{n1} and T_{n2} and compare with a new critical value c_2.

4 Approximations of the Operating Characteristics

In order to calculate concrete acceptance sampling plans, we need to calculate the true operating characteristics, which is impossible without knowing the true underlying distributions. Thus, we shall derive appropriate approximations of the

operating characteristics that will allow us to construct asymptotically optimal sampling plans.

Let us introduce the following notations. The standardized arithmetic averages will be denoted by

$$\overline{X}_i^* = \sqrt{n_i}\frac{\overline{X}_i - \mu}{\sigma}, \qquad i = 1, 2.$$

Here and in what follows, we assume that $D = 1$, since otherwise one may replace the X_{2i} by DX_{2i}. It turns out that the asymptotically optimal acceptance sampling plans depend on the quantile function $G^{-1}(p)$ of the standardized observations

$$X_i^* = (X_i - \mu)/\sigma, \qquad i = 0, 1.$$

Thus, we shall assume that we have an *arbitrary* consistent quantile estimator $G_m^{-1}(p)$ of $G^{-1}(p)$ at our disposal. It will be calculated from the additional sample taken at time t_0. We only need the following regularity assumption.

Assumption Q One of the following two conditions holds.

(i) $G_m^{-1}(p)$ is a quantile estimator of the quantile function $G^{-1}(p)$ of the standardized measurements satisfying the central limit theorem

$$\sqrt{m}(G_m^{-1}(p) - G^{-1}(p)) \xrightarrow{d} V,$$

as $m \to \infty$, for some random variable V.

(ii) $F_m^{-1}(p)$ is a quantile estimator of the quantile function $F_0^{-1}(p)$ of the measurements taken at time t_0, satisfying the central limit theorem

$$\sqrt{m}(F_m^{-1}(p) - F_0^{-1}(p)) \xrightarrow{d} U,$$

as $m \to \infty$, for some random variable U.

Remark 1 Notice that under condition (ii) of Assumption Q, one may construct a quantile estimator for $G^{-1}(p)$ by

$$\tilde{G}_m^{-1}(p) := \frac{F_m^{-1}(p) - \mu_0}{\sigma_0},$$

if μ_0 and σ_0 are known, since the quantiles of F and G are related by

$$F^{-1}(p) = \mu + \sigma G^{-1}(p).$$

If μ and σ are unknown, one should take the estimator

$$G_m^{-1}(p) := \frac{F_m^{-1}(p) - \overline{X}_0}{S_m}, \tag{10}$$

where $(\overline{X}_0, S_m)$ consistently estimates (μ_0, σ_0) under our assumption of an i.i.d. sample $X_{01}, \ldots, X_{0m}$ with a finite second moment.

Let us consider two examples.

Example 1 A natural candidate for $F_m^{-1}(p)$ is the corresponding sample quantile,

$$\tilde{F}_m^{-1}(p) = X_{0,(\lceil mp \rceil)}, \qquad p \in (0,1),$$

where $X_{0,(1)} \leq \cdots \leq X_{0,(m)}$ is the order statistic associated with $X_{01}, \ldots, X_{0m}$. However, it is known that the sample quantiles perform poorly for the type of acceptance sampling plans to be studied here, see Meisen et al. (2012).

Example 2 Suppose that the distribution of the measurements is concentrated on a finite interval $[a, b]$ that can be assumed to be $[0, 1]$. The Bernstein-Durrmeyer polynomial estimator of degree $N \in \mathbb{N}$ for $F_0^{-1}(p)$ is then defined as

$$F_{m,N}^{-1}(p) = (N+1) \sum_{i=0}^{N} a_i \, B_i^{(N)}(p), \qquad p \in (0,1),$$

with coefficients

$$a_i = \int_0^1 \tilde{F}_m^{-1}(u) B_i^N(u) \, du, \quad \text{where} \quad B_i^{(N)}(x) = \binom{N}{i} x^i (1-x)^{N-i}$$

for $i = 0, \ldots, N$ are the Bernstein polynomials. In Pepelyshev et al. (2014a) it has been shown that $\hat{F}_{m,N}^{-1}(p)$ is consistent in the MSE and MISE sense and almost attains the optimal parametric rate of convergence if F_0^{-1} is smooth. The degree N can be chosen in a data-adaptive way by controlling the number of modes of the density associated with the estimator as well as the closeness of the associated estimator of the distribution function $\hat{F}_{m,N}(x)$ in the sense that the maximal distance between $\hat{F}_{m,N}(\hat{F}_{m,\hat{N}}^{-1}(x))$ and the identity function $i(x) = x$ is uniformly less or equal to $1/R_m$ where $R_m = 2\sqrt{m}/\sqrt{2 \log \log m}$; for details of the algorithm leading to the estimate $\hat{N}$, we refer to Pepelyshev et al. (2014a). For the resulting estimate $\hat{F}_{m,\hat{N}}^{-1}(p)$ an uniform error bound can be established,

$$\sup_q |\hat{F}_{m,\hat{N}}^{-1}(q) - F_0^{-1}(q)| \leq 2\sqrt{2 \log \log m}/\sqrt{m},$$

see Pepelyshev et al. (2014a, Theorem 3.1).

Example 3 From previous studies it is known that quantile estimators obtained by (numerically) inverting a kernel density estimator

$$\hat{f}_m(x) = \frac{1}{mh} \sum_{i=1}^{n} K_h(x - X_{0i}), \qquad x \in \mathbb{R},$$

provide accurate results for sampling plan estimation, see Herrmann and Steland (2010). Here $K(\bullet)$ is a regular kernel, usually chosen as a density with mean 0 and unit variance, $h > 0$ the bandwidth and $K_h(z) = K(z/h)/h$, $z \in \mathbb{R}$, its rescaled version. The associated quantile estimator is obtained by solving for given $p \in (0, 1)$ the nonlinear equation

$$F_m(x_p) = \int_{-\infty}^{x_p} \hat{f}_m(x)\, dx \stackrel{!}{=} p.$$

For a kernel density estimator one has to select the bandwidth h. If the resulting estimate is consistent for $f(x)$ for each $x \in \mathbb{R}$, which requires to select $h = h_n$ such that $h \to 0$ and $nh \to \infty$, it follows that the corresponding estimator of the distribution function, $F_m(x) = \int_{-\infty}^{x} \hat{f}_m(u)\, du$, $x \in \mathbb{R}$, is consistent as well, see Glick (1974), if the kernel used for smoothing is a density function.

In many cases, the central limit theorem for a quantile estimator, say, $\hat{Q}_m(p)$, of a quantile function $F^{-1}(p)$ can be strengthened to a functional version that considers the scaled difference $\sqrt{m}(\hat{Q}_m(p) - F^{-1}(p))$ as a function of p.

Assumption Q' Assume that $F_m^{-1}(p)$ is a quantile estimator of the quantile function $F_0^{-1}(p)$, $0 < p < 1$, satisfying

$$\sqrt{m}(F_m^{-1}(p) - F_0^{-1}(p)) \stackrel{d}{\to} \mathscr{F}(p),$$

as $m \to \infty$, for some random process $\{\mathscr{F}(p) : 0 < p < 1\}$, in the sense of weak convergence of such stochastic processes indexed by the unit interval.

Assumption Q' holds true for the sample quantiles $F_m^{-1}(p)$ as well as, for example, the Bernstein-Durrmeyer estimator, if the underlying distribution function attains a positive density. For the latter results, further details and discussion, we refer to Pepelyshev et al. (2014a).

Having defined the decision rules for acceptance and rejection at both stages by (8) and (9), the overall OC curve is obviously given by

$$\mathrm{OC}(p) = P(T_{n1} > c_1, T_{n1} + T_{n2} > c_2 | p), \qquad p \in [0, 1].$$

Further, the operating characteristic of stage 2 is a conditional one and given by

$$\mathrm{OC}_2(p) = P(T_{n1} + T_{n2} > c_2 | T_{n1} > c_1, p), \qquad p \in [0, 1].$$

Those probabilities cannot be calculated explicitly under the general assumptions of this paper. Hence, we need appropriate approximations in order to construct valid sampling plans.

4.1 Independent Sampling

The case of independent sampling is, of course, of some relevance. In particular, it covers the case of destructive testing or, more generally, testing methods that may change the properties. Examples are accelerated heat-damp tests of PV modules. Let us assume that the samples $X_{i1}, \ldots, X_{in_i}$, $i = 1, 2$, are independent. Our OC curve approximations are based on the following result, which provides expansions of the test statistics involving the quantile estimates $G_m^{-1}(p)$.

Proposition 1 *Under independent sampling, we have*

$$T_{n1} = \overline{X}_1^* - \sqrt{n_1} G_m^{-1}(p) + o_P(1),$$

as $n_1 \to \infty$ *and* $m/n_1 = o(1)$, *and*

$$T_{n2} = \overline{X}_2^* - \sqrt{n_2} G_m^{-1}(p) + o_P(1),$$

as $n_2 \to \infty$ *and* $m/n_2 = o(1)$. *If a quantile estimator* $F_m^{-1}(p)$ *of* $F_0^{-1}(p)$ *is available, both expansions still hold true with* $G_m^{-1}(p)$ *defined by (10).*

In what follows, $\Phi(x)$ denotes the distribution function of the standard normal distribution. We obtain the following approximation of the overall OC curve. The approximation holds in the following sense: We say A approximates A_n and write $A_n \approx A$, if $A_n = A + o_P(1)$, as $\min(n_1, n_2) \to \infty$.

Theorem 1 *Under independent sampling and Assumption Q we have*

$$\mathrm{OC}_2(p) \approx \frac{1}{\sqrt{2\pi}} \frac{\int_{c_1 + \sqrt{n_1} G_m^{-1}(p)} \left[1 - \Phi(c_2 - z + (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p))\right] e^{-z^2/2}\, dz}{1 - \Phi(c_1 + \sqrt{n_1} G_m^{-1}(p))},$$

for any fixed $p \in (0, 1)$.

4.2 Dependent Sampling

If it is not necessary to rely on independent samples for quality control at time t_0 and inspection at time t_1, i.e. to test different modules at inspection, it is better to take the same modules. This means, one should rely on a panel design, where at time t_0 or t_1 a random sample from the lot is drawn, the so-called panel, and that panel is

also analyzed at the time of inspection, i.e. the modules are remeasured. To simplify the technical proofs, we shall assume in the sequel that the panel is established at time t_1 and that the sample taken at time t_0 is independent from the observations taken at later time points.

The control-inspection scheme studied in this paper aims at the minimization of the costs of testing by aggregating available information. Therefore, the inspection sample should be (and will be) considerably smaller than the first stage sample, i.e. $n_2 << n_1$, although it also may happen occasionally that $n_2 > n_1$, since the sample sizes are random.

In order to deal with this issue, the following dependent sampling scheme is proposed. If $n_2 < n_1$, one draws a subsample of size n_2 from the items drawn at time t_1 to obtain the control sample of size n_2. Those n_2 items are remeasured at time instant t_2 yielding the sample $X_{21}, \ldots, X_{2n_2}$. Notice that for fixed item (solar panel) i the corresponding measurements, denoted by X_{1i} and X_{2i}, are dependent, since they are obtained from the same item. Thus, we are given a paired sample

$$(X_{1i}, X_{2i}), \qquad i = 1, \ldots, n_2,$$

which has to be taken into account.

It remains to discuss how to proceed, if $n_2 > n_1$. Then one remeasures all n_1 items already sampled at time t_1 yielding n_1 paired observations (X_{1i}, X_{2i}), $i = 1, \ldots, n_1$, and draws $n_2 - n_1$ additional items from the lot.

As a consequence, n_1 observations from the stage 2 sample are stochastically dependent from the stage 1 observations, whereas the others are independent. In order to proceed, let us assume that the sample sizes n_1 and n_2 satisfy

$$\lim \frac{n_1}{n_2} = \lambda. \tag{11}$$

Notice that

$$\mathrm{Cov}(\overline{X}_1^*, \overline{X}_2^*) = \frac{1}{\sqrt{n_1 n_2}} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathrm{Cor}(X_{1i}, X_{2j}) = \sqrt{\frac{n_1}{n_2}} \rho',$$

where

$$\rho' = \mathrm{Cor}(X_1, X_2) \neq 0.$$

Thus, if $\rho' \neq 0$, the approximation results of the previous subsection are no longer valid, since even asymptotically $\overline{X}_1$ and $\overline{X}_2$ are correlated, and thus the standardized versions are correlated as well under this condition.

The following results provide the extensions required to handle this case of dependent sampling. Proposition 2 provides the asymptotic normality of the sample averages, which share the possibly non-trivial covariance.

Proposition 2 *Suppose that the above sampling scheme at stages 1 and 2 is applied and assume that one of the following assumptions is satisfied.*

(i) $X_{01}, \ldots, X_{0m}$ *is an i.i.d. sample with common distribution function* $F_0(x) = F(x - \Delta)$ *and Assumption Q holds.*
(ii) Assumption Q' is satisfied.

Then we have

$$\begin{pmatrix} \overline{X}_1^* \\ \overline{X}_2^* \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right),$$

as $\min(n_1, n_2) \to \infty$ *with* $n_1/n_2 \to \lambda$*, where the asymptotic covariance matrix is given by*

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with $\rho = \sqrt{\lambda}\,\mathrm{Cor}(X_1, X_2)$.

The following theorem now establishes expansions of the test statistics, which hold jointly.

Theorem 2 *Suppose that the above sampling scheme at stages 1 and 2 is applied and assume that one of the following assumptions is satisfied.*

(i) $X_{01}, \ldots, X_{0m}$ *is an i.i.d. sample with common distribution function* $F_0(x) = F(x - \Delta)$ *and Assumption Q holds.*
(ii) Assumption Q' is satisfied.

Then we have

$$\begin{pmatrix} T_{n1} \\ T_{n2} \end{pmatrix} = \begin{pmatrix} \overline{X}_1^* \\ \overline{X}_2^* \end{pmatrix} - \begin{pmatrix} \sqrt{n_1} G_m^{-1}(p) \\ \sqrt{n_2} G_m^{-1}(p) \end{pmatrix} + o_P(1),$$

as $\min(n_1, n_2) \to \infty$ *with* $n_1/n_2 \to \infty$ *and* $\max(n_1, n_2)/m = o(1)$.

The approximation of the OC curve $\mathrm{OC}_2(p)$ is now more involved. Recall at this point the well-known fact that for a random vector (X, Y) that is bivariate normal with mean vector $(\mu_X, \mu_Y)'$, variances $\sigma_X^2 = \sigma_Y^2 = 1$ and correlation ρ_{XY}, the conditional distribution of, say, Y given $X = z$ attains the Gaussian density

$$x \mapsto \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x - \rho z)^2}{2\sqrt{1 - \rho_{XY}^2}} \right), \qquad x \in \mathbb{R}.$$

The following theorem now provides us with the required approximation of the operating characteristic for the second stage sampling plan. It will be established in the following sense, slightly modified compared to the previous subsection: We say A approximates A_n and write $A_n \approx A$, if $A_n = A + o_P(1)$, as $\min(n_1, n_2) \to \infty$ with $n_1/n_2 \to \lambda$, $\max(n_1, n_2)/m = o(1)$ and $n_1 \geq n \to \infty$.

Theorem 3 *Suppose that the above sampling scheme at stages 1 and 2 is applied and assume that one of the following assumptions is satisfied.*

(i) $X_{01}, \ldots, X_{0m}$ *is an i.i.d. with common distribution function* $F_0(x)$ *and Assumption Q holds.*
(ii) Assumption Q' is satisfied.

If, additionally, $|\rho| < 1$*, then we have*

$$\mathrm{OC}_2(p) \approx \frac{1}{\sqrt{2\pi}} \frac{\int_{c_1+\sqrt{n_1}G_m^{-1}(p)} \left[1 - \Phi\left(\frac{c_2 - z + (\sqrt{n_1}+\sqrt{n_2})G_m^{-1}(p) - \hat{\rho}z}{\sqrt{1-\hat{\rho}^2}}\right)\right] e^{-z^2/2}\, dz}{1 - \Phi(c_1 + \sqrt{n_1}G_m^{-1}(p))}, \tag{12}$$

where

$$\hat{\rho} = \sqrt{\frac{n_1}{n_2}} \frac{\hat{\gamma}}{\hat{\sigma}_1 \hat{\sigma}_2} \quad \textit{with} \quad \hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^{n} (X_{ji} - \overline{X}_i)^2, \ j = 1, 2,$$

and

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^{n} (X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2).$$

The above result deserves some discussion.

Remark 2 Observe that the unknown correlation coefficient is estimated from n pairs (X_{1i}, X_{2i}), $i = 1, \ldots, n$. Since the sampling plan (n_2, c_2) cannot be determined without an estimator $\hat{\rho}$, one should fix $n \leq n_1$ and remeasure n items at inspection time t_2, in order to estimate ρ.

Remark 3 The fact that the approximation also holds true under the general probabilistic assumption Q' points to the fact that the results generalize the acceptance sampling methodology to the case of dependent sampling, for example when it is not feasible to draw randomly from the lot and instead one has to rely on consecutive produced items that are very likely to be stochastically dependent due to the nature of the production process.

Remark 4 The condition (11) can be easily ensured by replacing n_2 by n_1/λ, i.e. put $n_2(\lambda) = n_1/\lambda$ and determining λ such that a valid sampling plan (n_2, c_2) results. However, the procedure is not reformulated in this way for the sake of clarity.

4.3 Sampling in Spatial Batches

In photovoltaic quality control, it is quite common to sample in spatial batches. Here one selects randomly a solar panel from the photovoltaic system, usually arranged as a grid spread out over a relatively large area. Then the selected module and $b-1$ neighboring modules are measured on site. Of course, observations from neighboring modules are correlated, since they share various factors that affect variables relevant for quality and reliability. Among those are the frame on which they are installed, so that they share risk factors due to wrong installation, the local climate within the area (wind and its direction that leads to stress due to vibrations, see Assmus et al. (2011)), the wires as well as the inverter to which they are connected. Further, one cannot assume that during installation the modules are randomly spread over the area, so that their ordering may be the same as on the production line.

So let us assume that one substitutes n_i by $\lceil n_i/b \rceil b$ and c_i by the re-adjusted critical value (see step 6 of the algorithm in Sect. 5). Thus, we may and will assume that

$$n_i = r_i b, \qquad i = 1, 2, \quad \text{and} \quad r_1 \le r_2,$$

where b is the batch size and r_i the number of randomly selected batches. Suppose that the observations are arranged such that

$$(X_{i1}, \ldots, X_{in_i}) = (X_1^{(1)}, \ldots, X_b^{(1)}, \ldots, X_1^{(r_i)}, \ldots, X_b^{(r_i)}),$$

where $X_\ell^{(j)}$ is the ℓth observation from batch j, $\ell = 1, \ldots, b$, $j = 1, \ldots, r_i$.

Let us assume the following spatial-temporal model:

$$X_{i,(\ell-1)b+j} = \mu_i + B_\ell + \epsilon_{ij},$$

for $i = 1, 2$, $\ell = 1, \ldots, r$ and $j = 1, \ldots, b$. Here $\{\epsilon_{ij} : 1 \le j \le b, i = 1, 2\}$ are i.i.d. $(0, \sigma_\epsilon^2)$ error terms, $\{B_\ell : \ell = 1, \ldots, r_2\}$ are i.i.d. $(0, \sigma_B^2)$ random variables representing the batch effect. It is assumed that $\{\epsilon_{ij}\}$ and $\{B_\ell\}$ are independent.

Then the covariance matrix of the random vector $\mathbf{X}_i = (X_{i1}, \ldots, X_{1n_i})$ is given by

$$\mathrm{Cov}(\mathbf{X}_i) = \bigoplus_{i=1}^{r_i} [\sigma_B^2 \mathbf{J}_b + \sigma_\epsilon^2 \mathbf{I}_b],$$

for $i = 1, 2$, where $\mathbf{J}_b$ denotes the $(b \times b)$-matrix with entries 1, $\mathbf{I}_b$ is the b-dimensional identity matrix and $A \oplus B = \mathrm{diag}(A, B)$ is the direct sum of two

matrices A and B, i.e. the block-diagonal matrix $\left(\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & B \end{array}\right)$. Observing that

$$\mathrm{Cov}(\sqrt{n_1}\overline{X}_1, \sqrt{n_2}\overline{X}_2) = \frac{S}{\sqrt{n_1}\sqrt{n_2}}$$

where S is the sum of all elements of $\mathrm{Cov}(\mathbf{X}_1)$, we obtain

$$\mathrm{Cov}(\sqrt{n_1}\overline{X}_1, \sqrt{n_2}\overline{X}_2) = \frac{r_1 b^2\sigma_B^2 + r_1 b\sigma_\epsilon^2}{\sqrt{r_1 r_2}b} = \sqrt{\frac{r_1}{r_2}}b\sigma_B^2 + \sqrt{\frac{r_1}{r_2}}\sigma_\epsilon^2.$$

It can be shown that the method of proof used to show the above results extends to that spatial batch sampling, if one additionally assumes that b is fixed and

$$\lim \frac{r_1}{r_2} = r^* > 0.$$

5 Computational Aspects

It is worth discussing some computational aspects. We confine ourselves to the case of independent sampling, since the modifications for the dependent case are then straightforward.

The calculation of the two-stage sampling plan is now as follows. At stage 1, one solves the equations

$$\mathrm{OC}_1(\mathrm{AQL}) = 1 - \alpha_1, \qquad \mathrm{OC}_1(\mathrm{RQL}) = \beta_1,$$

leading to the explicit solutions

$$n_1 = \left\lceil \frac{(\Phi^{-1}(\alpha_1) - \Phi^{-1}(1-\beta_1))^2}{(G_m^{-1}(\mathrm{AQL}) - G_m^{-1}(\mathrm{RQL}))^2} \right\rceil, \tag{13}$$

$$c_1 = -\frac{\sqrt{n_1}}{2}(G_m^{-1}(\mathrm{AQL}) + G_m^{-1}(\mathrm{RQL})). \tag{14}$$

The sampling plan (n_2, c_2) for stage 2 has to be determined such that

$$\mathrm{OC}_2(\mathrm{AQL}) = 1 - \alpha_2, \qquad \mathrm{OC}_2(\mathrm{RQL}) = \beta_1,$$

which is done by replacing OC_2 by its approximation, thus leading us to the nonlinear equations

$$\frac{1}{\sqrt{2\pi}} \frac{\int_{c_1+\sqrt{n_1}G_m^{-1}(\mathrm{AQL})} \left[1-\Phi(c_2-z+(\sqrt{n_1}+\sqrt{n_2})G_m^{-1}(\mathrm{AQL}))\right] e^{-z^2/2}\,dz}{1-\Phi(c_1+\sqrt{n_1}G_m^{-1}(\mathrm{AQL}))} = 1-\alpha_2$$

and

$$\frac{1}{\sqrt{2\pi}} \frac{\int_{c_1+\sqrt{n_1}G_m^{-1}(\mathrm{RQL})} \left[1-\Phi(c_2-z+(\sqrt{n_1}+\sqrt{n_2})G_m^{-1}(\mathrm{RQL}))\right] e^{-z^2/2}\,dz}{1-\Phi(c_1+\sqrt{n_1}G_m^{-1}(\mathrm{RQL}))} = \beta_2,$$

which have to be solved numerically. Notice that the integrals appearing at the left side also have to be calculated numerically.

In order to calculate the sampling plan (n_2, c_2), the following straightforward algorithm performed well and was used in the simulation study.
ALGORITHM:

1. Select $\varepsilon > 0$.
2. Calculate (n_1, c_1) using (13) and (14).
3. Perform a grid search minimization of the OC curve over $(n, c) \in \{(n', c') : c' = 1, \ldots, c^*(n'),\ n' = 1, \ldots, 200\}$, where $c^*(n') = \min\{1 \le c'' \le 60 : (\mathrm{OC(AQL)} - (1-\alpha_2))^2 + (\mathrm{OC(RQL)} - \beta_2)^2 \le \varepsilon\}$ for given n'. Denote the grid-minimizer by (n^*, c^*).
4. Use the grid-minimizer (n^*, c^*) as a starting value for numerically solving the nonlinear equations up to an error bound ε for the sum of squared deviations from the target. Denote the minimizer by (n_2^*, c_2^*).
5. Put $n_2 = \lceil n_2 \rceil$.
6. For fixed $n = n_2$ minimize numerically the nonlinear equations with respect to c_2 up to an error bound ε for the sum of squared deviations from the target. Denote the minimizer by c_2^*.
7. Output $(n_2, c_2) = (n_2, c_2^*)$.

It turned out that the combination of a grid search to obtain starting values and a two-pass successive invocation of a numerical optimizer to minimize with respect to the sample size and the control limit in the first stage and, after rounding up the sample size, minimizing with respect to the control limit results in a stable algorithm.

6 Simulations

The simulation study has been conducted, in order to get some insights into the final sample statistical properties of the procedures. It was designed to mimic certain distributional settings that are of relevance in photovoltaic quality control.

It is known from previous studies that the standard deviation of the estimated sample size is often quite high even when a large data set $X_{01}, \ldots, X_{0m}$ can be used to estimate it, see Meisen et al. (2012), Golyandina et al. (2012), and Pepelyshev et al. (2014a). The question arises how accurately the second stage sampling plan can be estimated, having in mind that the estimated first stage sample size affects the operating characteristic at the second stage.

For the simulations the following parameters were used: $\alpha = \beta = 0.1$ (global error probabilities), AQL $= 2\,\%$ and RQL $= 5\,\%$. The error probabilities $\alpha_1 = \beta_1$ for the first stage acceptance sampling procedure were selected from the set $\{0.03, 0.05, 0.07\}$ and the corresponding value $\alpha_2 = 1 - (1-\alpha)/(1-\alpha_1)$ was then calculated for the second stage inspection, cf. our discussion in Sect. 3. The sample size m of the additional sample from the production line was chosen as 250 and 500.

Data sets according to the following models were simulated:

$$\text{Model 1:} \quad X_0 \sim F^1 = N(220, 4),$$

$$\text{Model 2:} \quad X_0 \sim F^2 = 0.9N(220, 4) + 0.1N(230, 8).$$

$$\text{Model 3:} \quad X_0 \sim F^3 = 0.2N(210, 8) + 0.6N(220, 4) + 0.2N(230, 8).$$

$$\text{Model 4:} \quad X_0 \sim F^4 = 0.2N(212, 4) + 0.6N(220, 8) + 0.2N(228, 6).$$

The required quantiles for methods based on the kernel density estimator for the construction of the sampling plans were estimated by numerically inverting an integrated kernel density estimator $\hat{f}_m(x)$ calculated from the standardized sample $X_{01}^*, \ldots, X_{0m}^*$. The following methods of quantile estimation were used, where the first four approaches employ the kernel estimator with different bandwidth selectors:

1. Biased cross-validated (BCV) bandwidth.
2. Sheather-Johnson bandwidth selection (SJ), Sheather and Jones (1991).
3. Golyandina-Pepeyshev-Steland method (GPS), Golyandina et al. (2012).
4. Indirect cross-validation (ICV), Savchuk et al. (2010).
5. Bernstein-Durrmeyer polynomial (BDP) quantile estimator, Pepelyshev et al. (2014a).

The following tables summarize the simulation results. Each case was simulated using 10,000 repetitions.

Table 1 provides results for normally distributed measurements with mean 220 and variance 4. The results show that even for such small sample sizes as 250 and 500, respectively, the second-stage sampling plan (n_2, c_2) can be estimated with comparable accuracy as the first-stage plan. Further, it can be seen that the GPS bandwidth selector provides on average the smallest sampling plan numbers n_2 and the highest accuracy.

Table 1 Characteristics of the sampling plans for Model 1

α_1	α_2	m	Type	$E(n_1)$	sd(n_1)	c_1	sd(c_1)	$E(n_2)$	sd(n_2)	c_2	sd(c_2)
3 %	7.22 %	250	BCV	79.76	22.47	17.39	2.14	18.33	8.63	26.30	3.34
		250	SJ	82.13	25.42	17.43	2.38	19.97	9.97	26.41	3.42
		250	GPS	78.92	21.82	17.37	2.11	17.68	8.29	26.23	3.33
		250	ICV	80.10	22.89	17.40	2.17	18.50	8.81	26.31	3.34
		250	BDP	90.58	34.08	16.91	2.64	26.97	13.95	26.28	3.29
7 %	3.23 %	250	BCV	49.29	13.84	13.67	1.68	22.35	8.46	23.71	3.07
		250	SJ	50.76	15.67	13.71	1.86	23.85	9.60	23.93	3.38
		250	GPS	48.78	13.43	13.66	1.65	21.80	8.22	23.61	2.99
		250	ICV	49.50	14.09	13.68	1.70	22.52	8.62	23.73	3.11
		250	BDP	55.98	21.01	13.29	2.07	31.04	13.83	23.99	3.76
3 %	7.22 %	500	BCV	80.21	17.99	17.26	1.72	19.11	6.86	26.55	2.94
		500	SJ	81.60	19.51	17.27	1.84	20.09	7.52	26.66	3.06
		500	GPS	79.49	17.42	17.25	1.67	18.50	6.64	26.44	2.88
		500	ICV	80.36	18.22	17.26	1.73	19.18	6.98	26.56	2.97
		500	BDP	93.90	24.27	17.43	1.95	27.32	9.90	27.38	2.83
7 %	3.23 %	500	BCV	49.58	11.07	13.57	1.34	23.17	7.05	23.68	2.52
		500	SJ	50.45	12.01	13.58	1.44	24.09	7.57	23.81	2.69
		500	GPS	49.15	10.72	13.56	1.31	22.62	6.84	23.60	2.44
		500	ICV	49.67	11.21	13.57	1.35	23.27	7.19	23.70	2.54
		500	BDP	58.01	14.93	13.70	1.53	31.97	10.12	24.70	2.89

For Model 2, a mixture model where for 10 % of the items the mean is reduced by ten units, the situation is now different. Here biased cross-validation and indirect cross-validation perform best and produce the most accurate estimates, see Table 2. Again, the stage-two plan can be estimated with comparable accuracy.

Model 3 represents a symmetric distribution with two smaller subpopulations whose mean is larger or smaller, such that there are notable local minima of the density between the three corresponding local maxima. The results are given in Table 3. Whereas for Models 1, 2, and 4 the GDP method leads to larger expected sample sizes and larger standard deviations than the other methods, it outperforms all other methods under Model 3, when $m = 250$.

Of considerable interest in photovoltaic applications, and presumable other areas as well, is Model 4, a kind of head-and-shoulders distribution resulting in relatively short tails. The results in Table 4 demonstrate that in this case the GPS method provides the best results in all cases, both in the sense of smallest expected sample sizes for both stages and in the sense of highest accuracy of estimation (i.e., smallest standard deviations).

Table 2 Characteristics of the sampling plans for Model 2

α_1	α_2	m	Type	$E(n_1)$	sd(n_1)	c_1	sd(c_1)	$E(n_2)$	sd(n_2)	c_2	sd(c_2)
3 %	7.22 %	250	BCV	281.52	88.60	21.80	2.45	115.68	46.15	29.34	2.28
		250	SJ	296.12	94.82	22.07	2.60	126.44	45.39	29.92	2.71
		250	GPS	297.60	97.68	22.06	2.63	128.48	48.30	30.38	2.82
		250	ICV	274.44	83.41	21.67	2.33	111.32	39.88	30.22	2.50
		250	BDP	320.28	123.21	21.35	3.20	136.48	57.84	29.55	1.97
7 %	3.23 %	250	BCV	173.44	54.63	17.11	1.93	110.88	38.94	31.64	3.62
		250	SJ	182.56	58.42	17.33	2.04	118.28	38.59	32.22	3.66
		250	GPS	183.32	60.02	17.32	2.06	122.88	42.15	32.34	3.74
		250	ICV	169.20	51.33	17.01	1.83	108.20	33.66	31.42	3.30
		250	BDP	197.52	75.82	16.77	2.51	129.72	50.85	31.20	4.59
3 %	7.22 %	500	BCV	280.24	56.48	21.94	2.04	116.00	26.61	28.84	2.82
		500	SJ	289.00	61.28	22.14	2.19	122.44	30.35	28.62	2.08
		500	GPS	283.32	62.64	22.02	2.14	118.40	30.42	28.49	2.09
		500	ICV	276.88	53.26	21.86	1.99	114.52	23.33	28.37	2.13
		500	BDP	331.44	91.83	22.40	2.73	138.20	42.49	30.36	3.29
7 %	3.23 %	500	BCV	172.84	34.83	17.23	1.60	110.52	21.92	31.89	2.96
		500	SJ	178.12	37.56	17.38	1.71	115.16	25.58	32.14	3.19
		500	GPS	174.68	38.62	17.29	1.68	113.04	29.44	31.68	2.84
		500	ICV	170.68	32.78	17.16	1.56	108.96	20.69	31.78	2.89
		500	BDP	204.36	56.50	17.59	2.14	133.32	35.27	32.30	3.34

Table 3 Characteristics of the sampling plans for Model 3

α_1	α_2	m	Type	$E(n_1)$	sd(n_1)	c_1	sd(c_1)	$E(n_2)$	sd(n_2)	c_2	sd(c_2)
3 %	7.22 %	250	BCV	206.12	78.11	26.99	4.46	75.92	33.60	29.05	2.57
		250	SJ	210.84	78.68	27.26	4.55	77.44	34.14	30.39	2.96
		250	GPS	203.44	75.73	26.86	4.26	72.64	33.43	29.36	2.99
		250	ICV	202.96	76.32	26.83	4.36	74.12	33.60	29.57	3.01
		250	BDP	171.00	66.69	23.82	3.70	58.76	29.80	29.15	2.07
7 %	3.23 %	250	BCV	127.16	48.05	21.20	3.50	73.36	35.32	33.94	2.66
		250	SJ	129.96	48.50	21.40	3.57	75.60	35.86	33.73	2.48
		250	GPS	125.36	46.58	21.09	3.33	70.88	34.31	33.98	2.63
		250	ICV	125.12	47.01	21.07	3.42	72.44	36.28	34.05	2.31
		250	BDP	105.64	41.11	18.73	2.90	58.64	28.75	32.19	3.20

(continued)

Table 3 (continued)

α_1	α_2	m	Type	$E(n_1)$	$\mathrm{sd}(n_1)$	c_1	$\mathrm{sd}(c_1)$	$E(n_2)$	$\mathrm{sd}(n_2)$	c_2	$\mathrm{sd}(c_2)$
3 %	7.22 %	500	BCV	190.80	58.17	26.28	3.54	68.00	25.37	28.45	2.04
		500	SJ	191.68	60.03	26.30	3.67	67.92	25.75	29.17	2.57
		500	GPS	188.96	56.97	26.20	3.44	66.52	24.81	28.79	1.74
		500	ICV	190.68	57.17	26.30	3.48	67.80	25.28	28.74	1.92
		500	BDP	194.84	49.04	25.89	2.67	71.80	22.06	28.73	1.62
7 %	3.23 %	500	BCV	117.64	35.78	20.64	2.77	65.76	27.16	33.69	2.20
		500	SJ	118.40	36.91	20.67	2.87	65.12	27.68	33.44	2.44
		500	GPS	116.48	34.94	20.57	2.69	65.36	26.14	33.62	2.11
		500	ICV	117.56	35.21	20.65	2.72	66.32	26.38	33.79	2.23
		500	BDP	120.12	30.31	20.33	2.10	67.88	22.75	34.64	2.05

Table 4 Characteristics of the sampling plans for Model 4

α_1	α_2	m	Type	$E(n_1)$	$\mathrm{sd}(n_1)$	c_1	$\mathrm{sd}(c_1)$	$E(n_2)$	$\mathrm{sd}(n_2)$	c_2	$\mathrm{sd}(c_2)$
3 %	7.22 %	250	BCV	171.76	49.49	24.52	2.58	59.79	23.18	28.55	1.85
		250	SJ	230.81	65.47	27.29	3.28	87.28	29.17	29.45	1.91
		250	GPS	173.57	46.82	24.62	2.49	60.70	22.25	28.57	1.74
		250	ICV	173.32	44.49	24.61	2.37	60.56	21.12	28.47	1.71
		250	BDP	251.94	80.19	26.98	3.67	96.91	34.33	29.63	2.04
7 %	3.23 %	250	BCV	105.94	30.47	19.26	2.03	58.85	24.02	33.12	2.37
		250	SJ	142.30	40.30	21.43	2.57	87.94	32.09	33.85	2.33
		250	GPS	107.05	28.84	19.33	1.96	59.57	22.68	33.32	2.63
		250	ICV	106.91	27.38	19.33	1.86	59.49	21.82	33.39	2.43
		250	BDP	155.31	49.36	21.19	2.88	100.03	39.10	33.95	2.47
3 %	7.22 %	500	BCV	231.72	52.48	27.43	2.58	88.00	23.35	29.27	1.61
		500	SJ	254.06	55.76	28.41	2.72	97.89	24.41	29.85	1.72
		500	GPS	209.60	43.00	26.43	2.20	78.13	19.35	28.69	1.36
		500	ICV	224.87	49.94	27.13	2.48	84.83	22.18	29.07	1.52
		500	BDP	295.24	72.06	29.34	3.10	115.42	30.61	30.47	2.01
7 %	3.23 %	500	BCV	142.85	32.31	21.54	2.02	88.49	26.36	34.24	2.23
		500	SJ	156.60	34.33	22.31	2.13	99.64	27.51	33.92	2.13
		500	GPS	129.24	26.48	20.76	1.73	77.04	21.69	34.58	2.21
		500	ICV	138.64	30.75	21.30	1.95	84.79	25.11	34.38	2.25
		500	BDP	181.98	44.37	23.04	2.43	120.50	35.80	33.86	2.09

7 Discussion

A sampling plan methodology for a control-inspection policy is established that allows for independent as well as dependent sampling. Relying on a decision rule based on a t-type test statistic, sampling plans are constructed based on quantile estimates calculated from an additional sample taken from the production line. The

new methodology applies to independent samples as well as dependent ones, under general conditions. When aggregating the available sampling information in order to minimize the required additional sampling costs at inspection time, it turns out that the relevant operating characteristics are relatively involved nonlinear equations that have to be solved numerically. Monte-Carlo simulations show that the approach works well and that the second stage sampling plan can be estimated with an accuracy that is comparable to the accuracy for the known formulas applicable for the first stage sampling plan. It also turns out that there is no uniformly superior method of bandwidth selection when relying on quantile estimates using inverted kernel density estimators. However, ICV and the GPS bandwidth selectors provide better results in many cases than more classical approaches.

The extension of the acceptance sampling methodology to the case of $L \geq 2$ number of inspection time points, preferably allowing for dependent cluster sampling, requires further investigation. Firstly, the question arises whether or not one should design such procedures such that the overall type I and type II error rates are under control. Further, it remains an open issue to which extent one should aggregate data and to which extent time effects can be modelled stochastically. Lastly, for large L appropriate procedures could resemble sequential (closed-end) procedures.

Having in mind that in many cases present day quality control is based on high dimensional data arising from measurement curves and images such as IV curves or EL images in photovoltaics, the extension of the acceptance sampling methodology to high dimensional and functional data deserves future research efforts as well; a deeper discussion is beyond the scope of the present article.

Acknowledgements The author thanks M.Sc. Andreas Sommer and M.Sc. Evgenii Sovetkin for proof-reading. Part of this work has been supported by a grant from the German Federal Ministry for Economic Affairs and Energy (BMWi), PV-Scan project (grant no. 0325588B).

Appendix: Proofs

The results are obtained by refinements of the results obtained in Steland and Zähle (2009) and Meisen et al. (2012) and their extension to the two-stage setup with possibly dependent samples. First, we need the two following auxiliary results, which are proved in Meisen et al. (2012) for independent observations. However, it can be easily seen that the proofs work under more general conditions.

Lemma 1 *If $X_1, X_2, \ldots$ have mean μ, variance $\sigma^2 \in (0, \infty)$ and satisfy a central limit theorem, i.e. $\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{d} N(0,1)$, as $n \to \infty$, then*

$$R_n = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}\frac{\sigma - S_m}{S_m} = o_P(1),$$

as $\min(n, m) \to \infty$, if S_m is a weakly consistent estimator for σ.

Lemma 2 *Suppose that*

$$\begin{pmatrix} \sqrt{m}(F_m^{-1}(p) - F^{-1}(p)) \\ \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \tag{15}$$

as $m \to \infty$*, for a pair* $(V_1, V_2)'$ *of random variables. Then*

$$V_n = \sqrt{\frac{n}{m}} \frac{\sqrt{m}(F_m^{-1}(p) - F^{-1}(p))}{S_m} = o_P(1),$$

as $\min(n, m) \to \infty$ *such that* $n/m = o(1)$.

Proof (Theorem 1) In order to establish the approximations, first notice that the well-known Skorohod/Dudley/Wichura representation theorem allows us to assume that all distributional convergences can be assumed to hold a.s. and that all $o_P(1)$ terms are $o(1)$; we leave the details to the reader. In particular, we may and shall assume that, almost surely,

$$(\overline{X}_1^*, \overline{X}_2^*)' \to (Z_1, Z_2) \quad \Leftrightarrow \quad \overline{X}_1^* - Z_1 = o(1),\ \overline{X}_2^* - Z_2 = o(1), \tag{16}$$

as $\min(n_1, n_2) \to \infty$, where (Z_1, Z_2) are i.i.d standard normal random variables. Let us consider the probability $q = P(T_{n1} > c_1, T_{n1} + T_{n2} > c_2)$. As shown in detail below in the proof of Theorem 2 for the more involved case of dependent sampling, we have the asymptotic expansions

$$\begin{pmatrix} T_{n1} \\ T_{n2} \end{pmatrix} = \begin{pmatrix} \overline{X}_1^* \\ \overline{X}_2^* \end{pmatrix} - \begin{pmatrix} \sqrt{n_1} G_m^{-1}(p) \\ \sqrt{n_2} G_m^{-1}(p) \end{pmatrix} + o_P(1),$$

as $\min(n_1, n_2) \to \infty$ with $n_1/n_2 \to \infty$ and $\max(n_1, n_2)/m = o(1)$, and both coordinates are independent given $G_m^{-1}(p)$. Combing these expansions with (16), we obtain, by plugging in the above expansions and (Z_1, Z_2) for $(\overline{X}_1^*, \overline{X}_2^*)$,

$$\begin{aligned} q &= P(\overline{X}_1^* - \sqrt{n_1} G_m^{-1}(p) + o(1) > c_1, \overline{X}_1^* + \overline{X}_2^* - (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) + o(1) > c_2) \\ &= P(Z_1 - \sqrt{n_1} G_m^{-1}(p) + o(1) > c_1, Z_1 + Z_2 - (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) + o(1) > c_2) \end{aligned}$$

Conditioning on $Z_2 = z_2$ and $X_{01}, \ldots, X_{0m}$ leads to the expression

$$\int P(z > c_1 + \sqrt{n_1} G_m^{-1}(p) + o(1), Z_2 > c_2 - z + (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) + o(1))\, d\Phi(z)$$

for q. Using $E(1_A 1_B) = 1_A E(1_B)$, if A is non-random with respect to P, we obtain

$$q = \int_{c_1+\sqrt{n_1}G_m^{-1}(p)+o(1)}^{\infty} [1 - \Phi(c_2 - z + (\sqrt{n_1} + \sqrt{n_2})G_m^{-1}(p) + o(1))]\, d\Phi(z) + o(1)$$
$$= \int_{c_1+\sqrt{n_1}G_m^{-1}(p)+o(1)}^{\infty} [1 - \Phi(c_2 - z + (\sqrt{n_1} + \sqrt{n_2})G_m^{-1}(p))]\, d\Phi(z) + o(1),$$

where we used the continuity of the integral. Further, the $o(1)$ term in the integrand can be dropped by virtue of the Lipschitz continuity of Φ. Combing the above results with the approximation $P(T_{n1} > c_1) = 1-\Phi(c_1+\sqrt{n_1}G_m^{-1}(p))+o(1)$, establishes the result. □

We are now in a position to show Theorem 2. If $X_{01}, \ldots, X_{0m}$ and $X_{11}, \ldots, X_{1n_1}$ are independent, then (15) follows easily. Otherwise, Assumption Q' ensures the validity of the joint asymptotic normality for independent as well as a large class of dependent sampling schemes.

Proof (Theorem 2) Recall that $E(\overline{X}_i) = \mu$ and $\mathrm{Var}(\overline{X}_i) = \sigma/n_i$, $i = 1, 2$. We may closely follow the arguments given in Meisen et al. (2012), since we have

$$T_{ni} = \sqrt{n_i}\frac{\overline{X}_i - \tau}{S_m} = \sqrt{n_i}\frac{\overline{X}_i - \mu}{\sigma} + R_{ni} + \sqrt{n_i}\frac{\mu - \tau}{\sigma} + V_{ni},$$

where

$$R_{ni} = \sqrt{n_i}\frac{\overline{X}_n - \mu}{\sigma}\frac{\sigma - S_m}{S_m} = o_P(1),$$
$$V_{ni} = \sqrt{n_i}\frac{\mu - \tau}{\sigma}\left(\frac{\sigma}{S_m} - 1\right) = o_P(1),$$

as $\min(n_i, m) \to \infty$, by virtue of Lemma 1, since $\sqrt{m}(S_m - \sigma)$ is asymptotically normal (by an application of the Δ-method, if the fourth moment is finite, and $n_i/m = o(1)$, also see Steland and Zähle (2009). Thus, it remains to consider

$$\sqrt{n_i}\frac{\mu - F^{-1}(p)}{\sigma} = -\sqrt{n_i}G^{-1}(p) = \sqrt{\frac{n_i}{m}}\sqrt{m}[G_m^{-1}(p) - G^{-1}(p)] - \sqrt{n_i}G_m^{-1}(p),$$

where, by virtue of Assumption Q, the first term is $o_P(1)$, if $\min(m, n_i) \to \infty$ and $n_i/m = o(1)$. This shows the first assertion which is relevant when a quantile estimator of the standardized observations is available. Recall that $\mu_0 = \mu + \Delta = E(X_0)$ and $\sigma_0^2 = \mathrm{Var}(X_0) = \sigma^2$. If a quantile estimator F_m^{-1} for the quantile function $F_0^{-1}(p) = \mu_0 + \sigma_0 G^{-1}(p)$ of the additional sample taken at time t_0 is

available, one proceeds as follows. Noting that $\frac{\mu - F^{-1}(p)}{\sigma} = G^{-1}(p) = \frac{\mu_0 - F_0^{-1}(p)}{\sigma_0}$, we have

$$\sqrt{n_i}\frac{\mu - F^{-1}(p)}{\sigma} = \sqrt{\frac{n_i}{m}}\frac{\sqrt{m}[F_m^{-1}(p) - F_0^{-1}(p)]}{\sigma_0} - \sqrt{n_i}\frac{F_m^{-1}(p) - \overline{X}_0}{S_m}$$
$$- \sqrt{n_i}\frac{F_m^{-1}(p) - \overline{X}_0}{S_m}\left(\frac{S_m}{\sigma_0} - 1\right) + \sqrt{m}\frac{\mu_0 - \overline{X}_0}{\sigma_0}\sqrt{\frac{n_i}{m}}.$$

In this decomposition at the right side the first, third, and fourth term are $o_P(1)$, as $\min(n_i, m) \to \infty$ with $n_i/m = o(1)$, $i = 1, 2$. Notice that the fourth term is $o(1)$, since

$$\sqrt{m}(\overline{X}_0 - \mu_0)/\sigma_0 \xrightarrow{d} N(0, 1),$$

if $X_{01}, \ldots, X_{0m}$ are i.i.d. $\sim F((\bullet - \mu_0)/\sigma_0)$ or as a consequence of Assumption Q'. Thus,

$$\sqrt{n_i}\frac{\mu - F^{-1}(p)}{\sigma} = \sqrt{n_i}G_m^{-1}(p) + o_P(1),$$

where now $G_m^{-1}(p) = \frac{F_m^{-1}(p) - \overline{X}_0}{S_m}$ is an estimator of the quantile function $G^{-1}(p)$ of the standardized observations, see Remark 1. □

Proof (Proposition 2) We consider the case $n_1 < n_2$. W.l.o.g. we can assume that $X_{21}, \ldots, X_{2n_1}$ are the time t_2 measurements from those n_1 items (modules) already drawn at time t_1, and $X_{2,n_1+1}, \ldots, X_{2n_2}$ are $n_2 - n_1$ measurements taken from newly selected items from the lot. By virtue of the Cramér-Wold device, to prove the proposition, it suffices to show that for all constants $d_1, d_2 \in \mathbb{R}$ with $(d_1, d_2) \neq (0, 0)$

$$d_1\overline{X}_1^* + d_2\overline{X}_2^* \xrightarrow[n\to\infty]{d} N(0, d_1^2 + d_2^2 + 2d_1d_2\sqrt{\lambda}\varrho'),$$

since $E(d_1\overline{X}_1^* + d_2\overline{X}_2^*) = 0$ and

$$Var(d_1\overline{X}_1^* + d_2\overline{X}_2^*) = d_1^2 + d_2^2 + 2d_1d_2\sqrt{\frac{n_1}{n_2}}\varrho'.$$

Write $d_1\overline{X}_1^* + d_2\overline{X}_2^* = A_n + B_n$, where

$$A_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n_1}\left[d_1\frac{X_{1j} - \mu_1}{\sigma_1} + d_2\sqrt{\frac{n_1}{n_2}}\frac{X_{2j} - \mu_1}{\sigma_1}\right],$$

$$B_n = \frac{d_2}{\sqrt{n_2}}\sqrt{n_2 - n_1}\frac{1}{\sqrt{n_2 - n_1}}\sum_{j=n_1+1}^{n_2}\frac{X_{2j} - \mu_2}{\sigma_2}.$$

The summands of A_n form an array of row-wise independent random variables $\xi_{n_1,j}, 1 \le j \le n_1, n_1 \ge 1$, with mean zero and variance

$$Var(\xi_{n1,j}) = d_1^2 + d_2^2 \frac{n_1}{n_2} + 2d_1 d_2 \sqrt{\frac{n_1}{n_2}} \varrho' \to d_1^2 + d_2^2 \lambda + 2d_1 d_2 \sqrt{\lambda} \varrho',$$

as $n_1 \to \infty$. Further, it is easy to verify that $B_n \xrightarrow{d} N(0, d_2^2(1-\lambda))$, as $n_1 \to \infty$. By independence of A_n and B_n, we obtain

$$\begin{pmatrix} A_n \\ B_n \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_1^2 + d_2^2\lambda + 2d_1 d_2 \sqrt{\lambda}\varrho' & 0 \\ 0 & d_2^2(1-\lambda) \end{pmatrix} \right),$$

as $n_1 \to \infty$. Now the continuous mapping theorem entails

$$A_n + B_n \xrightarrow{d} N(0, \sigma_{AB}^2),$$

as $n_1 \to \infty$, where $\sigma_{AB}^2 = d_1^2 + d_2^2\lambda + 2d_1 d_2 \sqrt{\lambda}\varrho' + d_2^2(1-\lambda) = d_1^2 + d_2^2\lambda + 2d_1 d_2 \sqrt{\lambda}\varrho'$, which establishes the assertion. □

Proof (Theorem 3) The proof goes along the lines of the proof for the independent case. Again we may and shall assume that the distributional convergence is a.s. and $o_P(1)$ are $o(1)$ a.s. Therefore, $\left(\overline{X}_1^*, \overline{X}_2^*\right) \xrightarrow{a.s.} (Z_1, Z_2)$, as $\min(n_1, n_2) \to \infty$. Here (Z_1, Z_2) is a bivariate random vector that is jointly normal with mean 0, unit variances and correlation ρ. The probability $q = P(T_{n1} > c_1, T_{n1} + T_{n2} > c_2)$ can now be calculated as follows. We have

$$\begin{aligned} q &= P(\overline{X}_1^* - \sqrt{n_1} G_m^{-1}(p) + o(1) > c_1, \overline{X}_1^* + \overline{X}_2^* - (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) + o(1) > c_2) \\ &= P(Z_1 > c_1 + \sqrt{n_1} G_m^{-1}(p) + o(1), Z_2 > c_2 - z + (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) + o(1)) \\ &= \int 1(z > c_1 + \sqrt{n_1} G_m^{-1}(p) + o(1)) \\ &\qquad P(Z_2 > c_2 - z + (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) + o(1) | Z_1 = z)\, d\Phi(z) + o(1). \end{aligned}$$

However, now we have to take into account that the conditional law of Z_2 given $Z_1 = z$ is a normal distribution that depends on z, namely with mean ρz and variance $1 - \rho^2$. Therefore, we may conclude that, up to an $o(1)$ term,

$$q = \frac{1}{\sqrt{2\pi}} \int_{c_1 + \sqrt{n_1} G_m^{-1}(p)}^{\infty} \left[1 - \Phi\left(\frac{c - z + (\sqrt{n_1} + \sqrt{n_2}) G_m^{-1}(p) - \rho z}{\sqrt{1-\rho^2}} \right) \right] e^{-z^2/2}\, dz.$$

The unknown correlation parameter ρ may be replaced by its consistent estimator $\hat{\rho}$, since the integrand is Lipschitz continuous, if $|\rho| < 1$. Indeed, observing that

$$\frac{d}{d\rho}\Phi\left(\frac{c - z + (\sqrt{n_1} + \sqrt{n_2})G_m^{-1}(p) - \rho z}{\sqrt{1-\rho^2}}\right)$$
$$= \varphi_{(0,1)}\left(\frac{c - z + (\sqrt{n_1} + \sqrt{n_2})G_m^{-1}(p) - \rho z}{\sqrt{1-\rho^2}}\right)$$
$$\cdot \frac{-z}{\sqrt{1-\rho^2}} + \rho\frac{c - z + (\sqrt{n_1} + \sqrt{n_2})G_m^{-1}(p) - \rho^2}{(\sqrt{1-\rho^2})^3},$$

where $\varphi_{(0,1)}$ denotes the density of the $N(0,1)$-distribution, we can find $0 < c < \infty$, such that the above expression is not larger than $c|z|$, as a function of z. Hence, replacing ρ by its estimator $\hat{\rho}_n$ results in an error term that can be bounded by $(2\pi)^{-1}c\int |z|e^{-z^2/2}\,dz|\hat{\rho}_n - \rho| = o_P(1)$. Putting things together, we arrive at the assertion of the theorem. □

References

Assmus, M., Jack, S., Weiss, K.-A., & Koehl, M. (2011). Measurement and simulation of vibrations of PV-modules by dynamic mechanical loads. *Progress in Photovoltaics, 19*, 688–694.

Bruhn-Suhr, M., & Krumbholz, W. (1991). Exact two-sided Liebermann-Resnikoff sampling plans. *Statistische Hefte, 32*, 233–241.

Feldmann, B. & Krumbholz, W. (2002). ASN-minimax double sampling plans. *Statistical Papers, 43*, 361–377.

Glick, N. (1974). Consistency conditions for probability densities and integrands of density estimators. *Utilitas Mathematica, 6*, 75–86.

Golyandina, N., Pepelyshev, A. & Steland, A. (2012). New approaches to nonparametric density estimation and selection of smoothing parameters. *Computational Statistics and Data Analysis, 56*(7), 2206–2218.

Herrmann, W. & Steland, A. (2010). Evaluation of photovoltaic modules based on sampling inspection using smoothed empirical quantiles. *Progress in Photovoltaics, 18(1)*, 1–9.

Herrmann, W., Althaus, J., Steland, A. & Zähle, H. (2006). Statistical and experimental methods for assessing the power output specification of PV modules. In *Proceedings of the 21st European Photovoltaic Solar Energy Conference*, (pp. 2416–2420).

Herrmann, W., Steland, A. & Herff, W. (2010). Sampling procedures for the validation of PV module output specification. In *Proceedings of the 24th European Photovoltaic Solar Energy Conference*, Hamburg, Germany, ISBN 3-936338-25-6, 3540–3547. doi:10.4229/24thEUPVSEC2009-4AV.3.70.

Kössler, W. & Lenz, H. -J. (1997). On the non-robustness of maximum-likelihood sampling plans by variables. In H.-J. Lenz & P.-T. Wilrich (Eds.), *Frontiers in statistical quality control* (Vol. 5, pp.38–51). Berlin: Springer.

Kössler, W. (1995). A new one-sided variable inspection plan for continuous distribution functions. *Allgemeines Statistisches Archiv*, 83(4), 416–433.

Liebermann, G. J. & Resnikoff, G. J. (1995). Sampling plans for inspection by variables. *Journal of the American Statistical Association, 50*, 457–516.

Meisen, S., Pepelyshev, A. & Steland, A. (2012). Quality assessment in the presence of additional data in photovoltaics. In H.-J. Lenz, W. Schmid & P.-T. Wilrich (Eds.), *Frontiers in Statistical Quality Control* (Vol. 10, pp. 249–274). Berlin: Springer.

Pepelyshev, A., Rafajłowicz, R. & Steland, A. (2014a). Estimation of the quantile function using Bernstein-Durrmeyer polynomials. *Journal of Nonparametric Statistics, 145*, 49–73.

Pepelyshev, A., Steland, A., & Avellan-Hampe, A. (2014b). Acceptance sampling plans for photovoltaic modules with two-sided specification limits. *Progress in Photovoltaics, 22*(6), 603–611.

Savchuk, O. Y., Hart, J. D. & Sheather, S. J. (2010). Indirect cross-validation for density estimation. *Journal of the American Statistical Association, 105*(489), 415–423.

Schilling, D. G. & Neubauer, D. V. (2009). *Acceptance sampling in quality control.* Boca Raton: Chapman & Hall/CRC.

Sheather, S.J. & Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society B, 53*, 683–690.

Steland, A. & Zähle, H. (2009). Sampling inspection by variables: Nonparametric setting. *Statistica Neerlandica, 63*(1), 101–123.

Part III
Design of Experiments

An Overview of Designing Experiments for Reliability Data

G. Geoffrey Vining, Laura J. Freeman, and Jennifer L.K. Kensler

Abstract The reliability of products and processes will become increasing important in the near future. One definition of reliability is "quality over time." Customers increasing will make their purchasing decisions on how long they can expect their products and processes to deliver high quality results. As a result, there will be increasing demands for manufacturers to design appropriate experiments to improve reliability. This paper begins with a review of the current practice for planning reliability experiments. It then reviews some recent work that takes into proper account the experimental protocol. A basic issue is that most reliability engineers have little training in planning experiments while most experimental design experts have little background in reliability data.

Keywords Censored data • Skewed distributions

1 Introduction

Consumers are demanding more reliable products and services. A popular definition of reliability is "quality over time." Consumers expect that products will continue to meet or exceed their expectations for at least the advertised lifetime, if not longer. One reason for the rise of the Japanese automotive industry within North America since the 1980s is the far better reliability of their cars and trucks.

Just as companies needed to apply experimental design concepts to improve quality, so too will they need to apply these same concepts to improve reliability. Current practice almost exclusively restricts the use of experimental protocols in

G.G. Vining (✉)
Department of Statistics, Virginia Tech, Blacksburg, VA 24060, USA
e-mail: vining@vt.edu

L.J. Freeman
Institute for Defense Analyses, Alexandria, VA, USA
e-mail: lfreeman@ida.org

J.L.K. Kensler
Shell Global Solutions (USA) Inc., Houston,TX 77210, USA
e-mail: Jennifer.Kensler@shell.com

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_19

reliability testing to completely randomized accelerated life tests. The future will see more broad scale use of basic experimental designs, analyses, protocols, and concepts.

A major problem facing this transition to more use of experimental design in reliability is the nature of reliability data. Classical experimental designs and analyses assume that the data at least roughly follow normal distributions. Reliability data tend to follow highly skewed distributions, better modeled by such distributions as the Weibull. Another complication is that typical reliability experiments censor large amounts of the data, which stands in stark contrast to classical experimental design and analysis. The issue then becomes that the people who routinely work with reliability data apply very different tools and methods than people who routinely do classical experimental design and analysis. The only proper way to apply classical experimental design approaches is to understand at a fundamental level the nature of reliability data. Unfortunately, very few people understand both fields well enough to bridge the gap.

This paper first presents an introduction to reliability data, with a special focus on the Weibull distribution and censoring. It then gives an overview of designing experiments for lifetime data. The next section introduces a motivating example analyzed in Meeker and Escobar (1998), who analyze the results as if they came from a completely randomized design. However, the data actually reflect sub-sampling. We then introduce a naive two-stage analysis that takes into account the actual experimental protocol. The next section discusses a more statistically rigorous approach to the data analysis. We then extend these basic results to the situation where we have sub-sampling within random blocks. The final section offers some conclusions and some future research directions.

2 Introduction to Reliability Data

Typically, reliability data focus on lifetimes. In some cases, these data are cycles until failure, which is a surrogate for time. Engineering examples include extremely complex systems, such as aircraft engines, as well as relatively simple parts such as metal braces. Often, engineers must build reliability models on the relatively simple components in order to develop a reasonable model for the complex system.

The most common distributions used by reliability engineers to model reliability data are the lognormal, the exponential, and the Weibull. Of these three, the Weibull tends to dominate, especially since the exponential is a special case. The lognormal distribution transforms highly skewed data to the normal distribution. The exponential distribution has a constant hazard function, which is associated with true random failure behavior, i.e. there is no specific failure mechanism associated with the failure. The biggest value of the Weibull distribution is its ability to model the times to failure for specific failure mechanisms.

Some textbooks discuss the gamma distribution for reliability applications. The gamma distribution is extremely important in queueing theory for modeling inter-arrival times. Most reliability engineers reject the basic concept of modeling times to failures as an inter-arrival problem. The primary reason is that failures have fundamental causes as opposed to truly random events. The physics based interpretation of the Weibull distribution is that it models the time to failure when the failure is due to the "weakest link," which is a common failure mechanism. This paper purely focuses on the Weibull distribution for modeling reliability data because of its overwhelming popularity among reliability engineers.

Most reliability data involve censoring, which does complicate the analysis. The basic types of censoring are:

- Right, where the test stops before all specimens fail
  - Type I, where the test stops at a pre-specified time,
  - Type II, where the test stops after a pre-specified number of failures,
- Left, where the first failure time recorded is after failures have begun to occur,
- Interval, where the analyst only knows that the failure occurred between two times.

Censoring reflects the reality that failures typically are rare events, even under accelerated conditions. By far, the most common approaches for estimation and inference for reliability data are maximum likelihood and log-likelihood.

The likelihood for Type I and Type II censored data is:

$$L(\beta, \gamma) = \mathscr{C} \prod_{i=1}^{N} [f(t_i)]^{\delta_i} \, [1 - F(t_i)]^{1-\delta_i} \, ,$$

where $\delta_i = 1$ if the data point is observed and $\delta_i = 0$ if the data point is right censored. Additionally, $f(t_i)$ is the probability distribution function (PDF) for the assumed distribution, $F(t_i)$ is the cumulative distribution function (CDF), and $\mathscr{C}$ is a constant which varies based on the censoring type. However, $\mathscr{C}$ does not impact the maximum likelihood estimators (MLEs). Therefore we use $\mathscr{C} = 1$ for simplicity. The log-likelihood for the right censored data case is then given by:

$$\ell_i(\beta, \gamma) = \sum_{i=1}^{N} \delta_i \log(f(t_i)) + \sum_{i=1}^{N} [1 - \delta_i] \log(1 - F(t_i)).$$

The PDF for the Weibull distribution is

$$f\,(t, \beta, \eta) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta - 1} e^{-\left(\frac{t}{\eta}\right)^{\beta}} ,$$

where $\beta > 0$ is the shape parameter, $\eta > 0$ is the scale parameter, and $t > 0$ is the time to failure. The Weibull distribution is popular because the shape parameter allows it to model several different mechanisms of failure. The CDF is

$$F(t, \beta, \eta) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}.$$

The hazard function represents the instantaneous probability of failure, which is quite important for reliability engineers. The Weibull hazard function is

$$h(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}.$$

We note that the hazard function is a constant when $\beta = 1$ (the exponential distribution). As a result, reliability engineers view the exponential distribution as modeling purely random failure, which often is of limited interest. For $\beta < 1$, the hazard function is monotonically decreasing, which corresponds to infant mortality. For $\beta > 1$, the hazard function is monotonically increasing, which corresponds to wear-out. As a result, the Weibull shape parameter, β, has a specific relationship to the specific failure mechanism.

The Weibull distribution is a special case of the smallest extreme value distribution, which is a member of the log-location-scale family of distributions. Let $\mu = \log(\eta)$, and let $z_i = \beta[\log(t_i) - \mu]$. We note that

$$\log[f(t_i)] = \log\left[\frac{\beta}{t_i}\right] + z_i - e^{z_i} \quad \text{and}$$

$$\log[1 - F(t_i)] = -e^{z_i}.$$

As a result, the log-likelihood for right censored Weibull data reduces to:

$$\ell(\beta, \gamma) = \sum_{i=1}^{N}\left(\delta_i\left[\log\left(\frac{\beta}{t_i}\right) + z_i\right] - e^{z_i}\right).$$

3 Current Approaches to Planning Experiments with Reliability Data

Reliability engineers conduct life tests to develop models for the product/process lifetimes at the use conditions. In some cases, the use conditions produce sufficient failures to estimate the model well. In most cases, however, the engineers must use a stress factor (in some cases, more than one stress factor) to increase the

probability of failures. Such experiments are called accelerated life tests. Common stress factors include temperature, voltage, humidity, etc. The engineer uses the estimated model to project back to the use conditions. Inherently, accelerated life tests involve extrapolation. As a result, the experimenter must exercise caution in choosing the appropriate levels for these stress factors. If the level is too extreme, the failure mechanism can change, which then nullifies the ability to extrapolate back to the expected behavior at the use conditions. In such a case, the cause of the failure may never occur at the use conditions, which is a problem.

The current literature for designed experiments with reliability data almost exclusively uses a completely randomized design, even when the actual protocol is something different. The focus of the current literature is on planning optimal designs. The basic issues are how many levels to use for the stress factors, how to allocate the available units to these levels, and how long to run the test.

Typically, accelerated life tests use a single stress factor with at least three levels. The linear models theory underlying the analysis suggests that the optimal design should use only two levels. The rationale for at least three is practical. Often the level for the stress factor closest to the use conditions does not produce enough failures to estimate the model well. Using more than two levels helps to mitigate that risk. Similarly, using more than two levels can help to mitigate the risk of inducing a new failure mechanism.

The typical analysis of life and accelerated life tests uses the reparameterization of the Weibull distribution to the smallest extreme value distribution. The basic idea is to use a linear model for the log-location parameter, μ. As a result, the basic model is

$$\mu_i = \boldsymbol{x}_i' \boldsymbol{\theta},$$

where μ_i is the log-location parameter for the ith experimental run, $\boldsymbol{x}_i$ is the specific value for the ith setting for the experimental factors (taking the model into account), and $\boldsymbol{\theta}$ is the corresponding vector of regression coefficients. Typically, the model does include the y-intercept term. Engineers then use maximum likelihood to estimate the model parameters and log-likelihood to perform inference.

4 Motivating Example

Zelen (1959) discusses a factorial experiment to determine the effect of voltage and temperature on the lifespan of a glass capacitor. Zelen describes the experiment as "n components are simultaneously placed on test." Table 1 summarizes the experimental results. Zelen uses eight capacitors per test stand and Type II censoring after the fourth failure. Each test stand receives a different combination

Table 1 Life test results of capacitors, adapted from Zelen (1959)

	Applied voltage			
Temperature (C)	200	250	300	350
170	439	572	315	258
	904	690	315	258
	1,092	904	439	347
	1,105	1,090	628	588
180	439	572	315	258
	904	690	315	258
	1,092	904	439	347
	1,105	1,090	628	588

of temperature and voltage. It is quite clear that the actual experimental protocol involves sub-sampling. The actual experimental units are the test stands since the treatment combinations are applied to the stand, not to the individual capacitors. Each capacitor in a test stand receives the exact same combination of the two factors. Two capacitors within a stand cannot have different temperatures or voltages. As a result, the capacitors within a stand are correlated with each other.

Meeker and Escobar (1998) use these data to illustrate how to analyze a reliability experiment using regression. They treat each capacitor as independent, thus ignoring the fact that capacitors within cells are correlated. Meeker and Escobar analyze the experiment as if there are 64 experimental units when in fact there are only 8. As a result, they treat the data as if they came from a completely randomized design in the capacitors replicated a total of eight times. The problem is that the actual protocol is an unreplicated completely randomized design in the test stands.

5 Naive Two-Stage Analysis of Reliability Data with Sub-Sampling

Freeman and Vining (2010) propose a naive two-step approach to this problem that assumes:

- Lifetimes within a test stand follow the same Weibull distribution.
- The failure mechanism remains the same across the test stands.
- The impact of treatments is through the scale parameter.
- Test stands are independent and, given the scale parameter, the observations within a test stand are independent.
- The experimental variability between scale parameters is log-normal.

5.1 First Stage of the Naive Analysis

Let t_{ij} be the observed lifetime for the jth item within the ith test stand. The failure times follow a Weibull distribution within a test stand, therefore:

$$f\left(t_{ij}\right) = \frac{\beta}{\eta_i}\left(\frac{t_{ij}}{\eta_i}\right)^{\beta-1} e^{-\left(\frac{t_{ij}}{\eta_i}\right)^{\beta}},$$

where $\beta > 0$ is the constant shape parameter and η_i is the scale parameter for i^{th} test stand. The likelihood for an individual test stand with right censoring present is:

$$\mathscr{L}(\beta, \mu_i) = \mathscr{C}\prod_{j=1}^{n}\left[f(t_{ij})\right]^{\delta_{ij}}\left[1 - F(t_{ij})\right]^{1-\delta_{ij}},$$

where $\delta_{ij} = 1$ if the item fails and $\delta_{ij} = 0$ if the item is censored. Again, $\mathscr{C}$ is a constant dependent on the type of censoring but can be taken as $\mathscr{C} = 1$ when calculating maximum likelihood estimates. The joint log-likelihood for data with right censoring then becomes:

$$\ell(\beta, \mu_1, \ldots, \mu_m) = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\delta_{ij}\log\left(\frac{\beta}{t_{ij}}\right) + \delta_{ij}z_{ij} - e^{z_{ij}}\right).$$

One then can find the MLEs for β and the η_is by maximizing the joint likelihood function. Many standard statistical software packages such as Minitab and SAS-JMP do this analysis.

Meeker and Escobar (1998) show that the Weibull distribution meets the regularity conditions to derive the asymptotic variance–covariance matrix from the maximum likelihood estimates. The resulting estimated variance matrix for the maximum likelihood estimates is:

$$\hat{\Sigma}_{\hat{\theta}} = \begin{bmatrix} \widehat{\text{Var}}(\hat{\beta}) & \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_1) & \cdots & \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_m) \\ \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_1) & \widehat{\text{Var}}(\hat{\mu}_1) & & \vdots \\ \vdots & & \ddots & \widehat{\text{Cov}}(\hat{\mu}_{m-1}, \hat{\mu}_m) \\ \widehat{\text{Cov}}(\hat{\beta}, \hat{\mu}_m) & \cdots & \widehat{\text{Cov}}(\hat{\mu}_m, \hat{\mu}_{m-1}) & \widehat{\text{Var}}(\hat{\mu}_m) \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\beta^2} & -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\beta\partial\mu_1} & \cdots & -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\beta\partial\mu_m} \\ -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\beta\partial\mu_1} & -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\mu_1^2} & & \vdots \\ \vdots & & \ddots & -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\mu_{m-1}\partial\mu_m} \\ -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\beta\partial\mu_m} & \cdots & -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\mu_m\partial\mu_{m-1}} & -\frac{\partial^2\ell(\beta,\mu_1,\ldots,\mu_m)}{\partial\mu_m^2} \end{bmatrix}^{-1}.$$

From the log-likelihood one can establish:

$$-\frac{\partial^2 \ell(\beta, \mu_1, \ldots, \mu_m)}{\partial \beta^2} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\frac{\delta_{ij}}{\beta^2} + \left(\frac{z_{ij}}{\beta} \right)^2 \exp(z_{ij}) \right]$$

$$-\frac{\partial^2 \ell(\beta, \mu_1, \ldots, \mu_m)}{\partial \beta \partial \mu_i} = -\sum_{j=1}^{n} \left(2 \frac{z_{ij}}{\beta} \exp(z_{ij}) \right)$$

$$-\frac{\partial^2 \ell(\beta, \mu_1, \ldots, \mu_m)}{\partial \mu_i^2} = \sum_{j=1}^{n} \left(\beta^2 \exp(z_{ij}) \right).$$

Additionally, the second derivatives between all pairs of μ_i and μ_j are zero. This variance matrix will be used in the second stage of the model.

5.2 The Second Stage: The Model Between Experimental Units

This step uses the estimates for the shape parameter and the scale parameters and their corresponding variances for each experimental unit to model the estimates of the scale parameters as a linear function of the factors. The most appropriate way to estimate the model takes into account the variances on the scale parameter estimates through weighted least squares. In this case, the second stage model is:

$$\hat{\boldsymbol{\mu}} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}$ is the matrix containing the treatment levels of the factors, $\boldsymbol{\theta}$ is the vector of the corresponding regression coefficients, and $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \boldsymbol{V})$. The variance matrix, $\boldsymbol{V}$, accounts for the scale parameter variance estimates. Since the covariances are essentially 0, we can simplify the analysis by assuming that $\boldsymbol{V}$ is diagonal with the non-zero elements simply being $< \widehat{\mathrm{Var}}(\widehat{\mu}_i) >$. The resulting parameter estimates are:

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{V}^{-1} \hat{\boldsymbol{\mu}}.$$

The big advantage to this approach is that one can correctly model the experimental error in current statistical packages that have the ability to fit lifetime distributions and linear models.

The key to this analysis is a proper understanding of how one can deal with subsampling under normal theory. Once again, the observations within the experimental unit are correlated. However, one can take into proper consideration the correlation by replacing the individual observations within the experimental unit by their average, as long as each experimental unit has the same number of observations. The

Table 2 Stage one analysis results

Voltage	Temperature	$\hat{\eta}$	$\widehat{\mu_i} = \log(\widehat{\eta_i})$	$\widehat{\text{Var}}\,(\hat{\mu}_i)$
200	170	1,262.35	7.141	0.1387
200	180	1,292.78	7.165	0.1390
250	170	1,207.58	7.096	0.1386
250	180	532.85	6.278	0.1387
300	170	683.61	6.527	0.1385
300	180	431.04	6.066	0.1388
350	170	633.86	6.452	0.1384
350	180	510.10	6.235	0.1386

Table 3 Analysis from minitab, stage two analysis result

Predictor	Coefficient	Standard error	T	p
Constant	14.613	3.249	4.50	0.056
Voltage	−0.005638	0.001644	−3.43	0.019
Temperature	−0.03682	0.01838	−2.00	0.102

proposed two-stage analysis extends this basic idea to the sub-sampling situation with Weibull data. The first step uses the Weibull distribution to estimate the common shape parameter and the different scale parameters, one for each test stand. The second step models the log transform of the different scale parameters using weighted least squares where the experimental error terms are given by the asymptotic experimental error derived in the first step for the log-scale parameters.

Table 2 presents the results from Minitab estimating the eight different scale parameter assuming constant β. The estimate of the shape parameter is $\hat{\beta}_{\text{New}} = 3.62$, which is a dramatically different estimate from the shape parameter estimate in the traditional reliability analysis, which is $\hat{\beta}_{\text{Trad.}} = 2.75$. This difference in the shape parameter estimate is the first practical implication of taking the experimental design into account.

The second step of our proposed new analysis models the resulting maximum likelihood estimates for the μ_i's using a weighted regression linear model where the weights are determined by the asymptotic variances from the first step of this model. The second stage of this analysis can be done in any standard statistical package. Note that the variance estimates on the different μ_i are essentially equal in Table 2. This is a nice result because in the second stage of the model using a weighted regression is essentially equivalent to standard least squares regression further simplifying this two-stage analysis method. This is the case because we have assumed a constant shape parameter, β and the shape parameter is the driving parameter in the Fisher Information matrix calculations for the variances on the scale parameters. The results from running the analysis in Minitab are displayed in Table 3. Table 4 gives the analysis from Meeker and Escobar (1998), which assumes that all the capacitors are independent.

Several practical differences emerge comparing the results of the new analysis back to the traditional analysis. First, the Meeker and Escobar analysis overstates the true experimental degrees of freedom by treating each observation as an independent

Table 4 Analysis from Meeker and Escobar

Predictor	Coefficient	Standard error	Z	p
Intercept	13.4070	2.29584	5.84	0.000
Voltage	−0.0059	0.0010398	−5.68	0.000
Temperature	−0.0289	0.0129	−2.24	0.0250
Weibull shape	2.74869	0.418739		

data point. As a result, their standard errors of the coefficients are all smaller. The increase in standard error results in the temperature not being a significant factor at $\alpha = 0.05$ level for the new analysis. Additionally, the estimates of the shape parameter are dramatically different between the two analysis methods. The coefficient estimates for the linear relationship between the log scale parameter and temperature and pressure are also slightly different.

6 Joint Likelihood Approach

Kensler (2012) performed a simulation study comparing the two stage approach to the Meeker and Escobar approach. The type I error rate for the two stage was very close to the nominal. On the other hand, the Meeker and Escobar approach produced Type I error rates much higher than the nominal. In many cases, so high, in fact, as to invalidate the analysis.

A major problem with the naive two-stage analysis is that it cannot generate a joint likelihood for β and the coefficients for temperature and voltage. As a result, one cannot generate confidence or prediction intervals for such predicted values as percentiles, which often are of prime importance to a reliability analysis.

Freeman and Vining (2013) propose a joint likelihood approach that requires a variance component to take into proper account the test stand-to-test stand variability. If we have $i = 1, \ldots, m$ independent experimental units and $j = 1, \ldots, n_i$ sub-samples or observational units per experimental unit, one can specify the nonlinear mixed model for the Weibull distribution with sub-sampling as:

$$
\begin{aligned}
t_{ij}|u_i &\sim \text{Indep. Weib}(\beta, \eta_i) \\
F_1(t_{ij}|\beta, \eta_i, u_i) &= 1 - \exp\left[-\left(\frac{t_{ij}}{\eta_i}\right)^{\beta}\right] \\
\log(\eta_i) &= \mu_i = \boldsymbol{x}_i^T\boldsymbol{\theta} + u_i \\
f_2(u_i) &\sim iidN\ (0, \sigma_u^2),
\end{aligned}
$$

where $\boldsymbol{x}_i$ is the p x 1 vector of fixed factor levels, $\boldsymbol{\theta}$ is the vector of fixed effect coefficients,and u_i are $i = 1, \ldots, m$ independent random effects. Since the random

effects are independent, we can write the likelihood as:

$$\mathscr{L}(\beta, \boldsymbol{\theta}\,|\text{Data}) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} \left[f_1(t_{ij}|u_i) \right]^{\delta_{ij}} \left[1 - F_1(t_{ij}|u_i) \right]^{1-\delta_{ij}} f_2(u_i) \right] du_i,$$

where $f_1(t_{ij}|u_i)$ is the Weibull PDF for the data within an experimental unit and $f_2(u_i)$ is the normal PDF for the random effect.

Random effects models, especially nonlinear models, pose computational issues since it is necessary to integrate over the random effect u_i to maximize the likelihood. Gauss-Hermite (G-H) quadrature is an especially effective technique when the random effect follows a normal distribution. Quadrature involves weighting the sum of a function values at specific points over the domain of integration. G-H quadrature uses the roots of the Hermite polynomials to provide these specific points. G-H quadrature requires the random effect to have the form e^{-x^2}. As a result, a change of variables is necessary to apply G-H quadrature to our likelihood function. Let $u_i = \sqrt{2}\sigma_u v_i$, then the likelihood before the change of variables is:

$$\mathscr{L}(\beta, \boldsymbol{\theta}\,|\text{Data}) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} g(t_{ij}|u_i) \frac{e^{\frac{-u_i^2}{2\sigma_u^2}}}{\sqrt{2\pi\sigma_u^2}} \right] du_i,$$

where $g(t_{ij}|u_i) = \left[f_1(t_{ij}|u_i) \right]^{\delta_{ij}} \left[1 - F_1(t_{ij}|u_i) \right]^{1-\delta_{ij}}$ for right censored data. Executing the change in variables results in the following likelihood:

$$\mathscr{L}(\beta, \boldsymbol{\theta}\,|\text{Data}) = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n_i} g(t_{ij}|\sqrt{2}\sigma_u v_i) \frac{e^{-v_i^2}}{\sqrt{\pi}} \right] dv_i$$

G-H quadrature results in the following approximation of the likelihood:

$$\mathscr{L}(\beta, \boldsymbol{\theta}\,|\text{Data}) \approx \prod_{i=1}^{m} \frac{1}{\sqrt{\pi}} \left\{ \sum_{k=1}^{n_k} \left[\prod_{j=1}^{n_i} g(t_{ij}|\sqrt{2}\sigma_u q_{k_i}) w_k \right] \right\},$$

where n_k is the number of quadrature points, q_k are the evaluation points found from the roots of the Hermite polynomials, and w_k are the corresponding weights to the evaluation points:

$$w_k = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(q_k)]^2}.$$

A common recommendation for the number of quadrature points to minimize bias is 20 points. Pinheiro and Bates (1995) show that G-H quadrature with 100 points

is as good as any other solution they investigated to the numerical optimization problem. In this research, we use 20 quadrature points in all of our analyses, unless otherwise stated, as an appropriate compromise on computation time, especially in the simulation studies. The log-likelihood is:

$$\ell(\beta, \boldsymbol{\theta}\,|\text{Data}) \approx \sum_{i=1}^{m} \log\left(\frac{1}{\sqrt{\pi}} \sum_{k=1}^{n_k}\left[\prod_{j=1}^{n_i} g(t_{ij}|\sqrt{2}\sigma_u q_{k_i})w_k\right]\right).$$

The approximate log-likelihood is maximized through standard maximization techniques.

A major advantage of using G-H quadrature is that it results in a closed-form approximate log-likelihood which allows for one to derive a Hessian matrix and the corresponding asymptotic covariance matrix. Maximum likelihood theory states that under certain regularity conditions that $\sqrt{(n)}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges in distribution to a multivariate normal. Let $\boldsymbol{\theta}^{*T} = [\beta, \boldsymbol{\theta}]$, then $\hat{\boldsymbol{\theta}}^{*} \sim$ Asymptotically $MVN\ (\boldsymbol{\theta}^{*}, I(\boldsymbol{\theta}^{*})^{-1})$, where

$$I(\boldsymbol{\theta}^{*}) = \begin{bmatrix} -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\beta^2} & -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\beta\partial\theta_1} & \cdots & -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\beta\partial\theta_p} \\ -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\beta\partial\theta_1} & -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\theta_1^2} & & \vdots \\ \vdots & & \ddots & -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\theta_{p-1}\partial\theta_p} \\ -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\beta\partial\theta_p} & \cdots & -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\theta_p\partial\theta_{p-1}} & -\frac{\partial^2 \ell(\beta,\theta_1,\ldots,\theta_p)}{\partial\theta_p^2} \end{bmatrix}.$$

The estimated covariance for the parameter estimates can be found by substituting the MLEs for the parameters they estimate into the information matrix, $I(\boldsymbol{\theta}^{*})$. Meeker and Escobar (1998, p. 622) note that the regularity conditions hold for the Weibull distribution. See Freeman (2010) for the derivation of the information matrix for the random effects Weibull model. Alternatively, the standard errors for the model parameters could be calculated through a bootstrapping procedure.

Table 5 summarizes the analysis of the Zelen data using the joint likelihood approach. The standard errors for the intercept, voltage, and temperature are similar to those from the naive two-stage analysis. However, the estimate of the shape parameter from the joint analysis is very similar to that from the Meeker and Escobar analysis rather than the two-stage analysis, which is much higher. This result suggests that the two-stage method may be susceptible to bias in the estimation of the shape parameter.

Freeman and Vining (2013) perform a simulation study to investigate the properties of the joint likelihood analysis. Their basic conclusions are

- The joint likelihood approach results in Weibull shape parameter estimates that are robust to model misspecification and random effect variation.
- Weibull scale parameter estimates are consistently good for both the joint likelihood analysis and the Meeker and Escobar—independent analysis.

Table 5 Joint likelihood analysis of the Zelen data

Parameter	Estimate	Standard error	T	p-value
Weibull shape	2.7753	0.6622	4.19	0.0041
Intercept	13.5257	3.0636	4.42	0.0031
Voltage	−0.00589	0.001154	−5.10	0.0014
Temperature	−0.02964	0.01808	−1.64	0.1451
Log(σ_u)	−3.0184	9.0655	−0.33	0.7489

- The joint likelihood approach poorly estimates the true value of σ_μ, primarily due to the small number of degrees of freedom available in realistic reliability experiments to estimate this error term.
- The two-stage analysis provides a ready solution for practitioners with sub-sampling data.
- The joint likelihood approach provides a comprehensive solution for analyzing data from life test designs that are not completely randomized.
- Both analyses provide a motivation for thinking about design in a reliability context more comprehensively.

7 Extensions to Random Blocks with Sub-sampling

Kensler et al. (2014) extend the two-stage approach to the situation where we have test stands within random blocks. Like Freeman and Vining (2010), Kensler et al. estimate the log scale parameter for each test stand assuming a constant shape parameter across all the test stands in the first stage. They then perform a standard random block analysis using the estimated log scale parameters as the response. The model for the second stage analysis is

$$\hat{\boldsymbol{\mu}} = \boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{Z}\boldsymbol{\rho} + \boldsymbol{\omega},$$

where $\boldsymbol{\mu}$ is the vector of estimated log-scale parameters from the first stage, $\boldsymbol{X}$ is the model matrix for the treatment effects, $\boldsymbol{\gamma}$ is the vector of model coefficients, $\boldsymbol{Z}$ is the incidence matrix for the blocks, $\boldsymbol{\rho}$ is the vector of random block effects, and $\boldsymbol{\omega}$ is the vector of random test stand errors. The second stage analysis assumes that the ρ's are independent and normally distributed with a mean of 0 and a variance of σ_ρ^2, that the ω's are independent and normally distributed with a mean of 0 and a variance of σ_ω^2, and that the ρ's and the ω's are independent.

They adapted the battery data from Montgomery (2005, p. 165) as the basis for their illustrative example. The experimental objective was to determine the effect of operating temperature on battery life. The batteries came from three batches, assumed to be random. Each test stand had eight batteries. The researchers used Type II censoring after the fourth failure.

Table 6 The battery life data (in hours)

Batch	Temperature		
	15	70	125
1	74	34	20
	130	40	58
	155	75	80
	180	80	82
2	126	106	25
	150	115	45
	159	122	58
	188	136	70
3	110	120	25
	138	139	82
	160	150	96
	168	174	104

Table 7 The first stage analysis of the battery life data

Batch	Temperature	η_{ij}	$\mu_{ij} = \log(\eta_{ij})$	$\widehat{\text{Var}}(\mu_{ij})$
1	15	197.79	5.287	0.0158
1	70	87.98	4.477	0.0158
1	125	89.85	4.498	0.0158
2	15	208.88	5.342	0.0160
2	70	153.56	5.034	0.0160
2	125	76.28	4.334	0.0158
3	15	189.22	5.243	0.0160
3	70	193.85	5.267	0.0160
3	125	116.05	4.754	0.0159

Table 8 The second stage analysis of the battery life data

Source	df	MS	F	p-value
Temperature	1	0.8704	16.96	0.009
Block	2	0.0839	1.63	0.284
Residual	5	0.05132		

Table 6 summarizes the battery data. Table 7 summarizes the estimates of the scale parameters, the log scale parameters, and the variances of the log scale parameters. The estimate of the shape parameter, β is 4.03, which indicates a wear-out failure mode. Once again, the estimated variances for the estimated log scale parameters are essentially constant like the example in Freeman and Vining (2010). As a result, Kensler et al. use standard ordinary least squares in stage two of their analysis. Table 8 summarizes the second stage regression analysis from Minitab. The estimates of the two variance components are $\sigma_{\omega}^2 = 0.05132$ and $\sigma_{\omega}^2 = 0.01086$. The second stage analysis does show that the temperature does have an important effect on the battery life.

Kensler et al. (2014) performed a simulation study to examine the performance of the two-stage method. They found that the two-stage method did an excellent job preserving the Type I error rate. They also found that power for the two-stage method was close to the nominal for larger β and number of failures; however, the actual power was less than the nominal for small β's and only a few failures. Just as in the Freeman and Vining (2010) paper, the estimates for β from the first stage tend to be biased.

8 Conclusions and Future Research

The basic conclusions reached so far by this research are:

- The two-stage approach provides a straightforward basis for analyzing reliability experiments with sub-sampling that practitioners can apply with current standard statistical software.
- The two-stage approach does a good job preserving the nominal Type I error rates and a much better job than the traditional analyses.
- The two-stage approach does produce biased estimates of β.
- The two-stage approach does not allow the analyst to compute confidence or prediction intervals around predicted values.
- The joint likelihood approach has much less bias in its estimates of β.
- The joint likelihood approach does allow analysts to generate appropriate confidence and prediction intervals.

Future research includes

- submitting the paper that uses the joint likelihood approach for the random blocks case.
- extending the two-stage and joint likelihood approaches to split-plot experiments.

References

Freeman, L. J. (2010). *Statistical methods for reliability data from designed experiments.* Unpublished Ph.D. dissertation, Virginia Tech, Department of Statistics.

Freeman, L. J., & Vining, G. G. (2010). Reliability data analysis for life test experiments with subsampling. *Journal of Quality Technology, 42*(3), 233–241.

Freeman, L. J., & Vining, G. G. (2013). Reliability data analysis for life test designed experiments with subsampling. *Quality and Reliability Engineering International, 29*, 509–519.

Kensler, J. L. K. (2012). *Analyzing reliability experiments with random blocks and subsampling.* Unpublished Ph.D. dissertation, Virginia Tech, Department of Statistics.

Kensler, J. L. K., Freeman, L. J., & Vining, G. G. (2014). A practitioner's guide to analyzing reliability experiments with random blocks and subsampling. *Quality Engineering, 26*(3), 359–369.

Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. Hoboken, NJ: Wiley.

Montgomery, D. C. (2005). *Design and analysis of experiments*. Hoboken, NJ: Wiley.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics, 4*(1), 12–35.

Zelen, M. (1959). Factorial experiments in life testing. *Technometrics, 1*(3), 269–288.

Bayesian D-Optimal Design Issues for Binomial Generalized Linear Model Screening Designs

Edgar Hassler, Douglas C. Montgomery, and Rachel T. Silvestrini

Abstract Bayesian D-optimal designs have become computationally feasible to construct for simple prior distributions. Some parameter values give rise to models that have little utility to the practitioner for effect screening. For some generalized linear models such as the binomial, inclusion of such models can cause the optimal design to spread out toward the boundary of the design space. This can reduce the D-efficiency of the design over much of the parameter space and result in the Bayesian D-optimal criterion's divergence from the concerns of a practitioner designing a screening experiment.

Keywords Challenger data set • Confidence intervals • Non-linear designs

1 Introduction

Screening experiments address the situation when a process is not well understood and the experimenter wishes to learn which effects are important and which are unimportant with respect to controlling the response over the design space. The screening problem is well understood for linear models, but little attention has been paid to this problem in a non-linear setting. This may be due, in part, to the fact that most non-linear models require a level of understanding of the system to be able to define the model, and such a level of understanding may obviate the need for screening. Yet the need for screening designs arises in practice for generalized linear models (GLMs).

Screening designs, and experimental designs in general, must address many concerns that sometimes conflict. Many numerical criteria exist describing desirable qualities of experimental designs. One strategy for constructing designs is to focus on finding a design that is best with respect to just one such criterion, and these

E. Hassler (✉) • D.C. Montgomery
Arizona State University, Tempe, AZ, USA
e-mail: ehassler@asu.edu; doug.montgomery@asu.edu

R.T. Silvestrini
Naval Postgraduate School, Monterey, CA, USA
e-mail: rtsilves@nps.edu

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_20

are referred to as *optimal designs*. To create an optimal design one must specify a region of interest for a set of model input variables (called the design space), specify a mathematical model, select the sample size of the design, and choose the optimality criterion. A computer algorithm chooses the design points in the design space based on an optimization scheme that maximizes (or minimizes) the criterion.

Criteria for optimal designs are often functions of the Fisher information matrix $\mathbb{M}$. Common "alphabetic criteria" for linear models are given by Atkinson et al. (2007):

- A-optimality where trace $[\mathbb{M}^{-1}]$ is minimized.
- D-optimality where $|\mathbb{M}|$ is maximized.
- E-optimality where $\sup_{\boldsymbol{a}'\boldsymbol{a}=1} \mathrm{Var}[\boldsymbol{a}'\hat{\boldsymbol{\beta}}]$ is minimized.

Khuri et al. (2006) reviewed issues associated with designing experiments for GLMs, principal among which was what they termed the *design dependence problem*—the information matrix for non-linear models is dependent on the parameter values of the model. This leaves the practitioner seeking a screening design in the uncomfortable position of having to have a good estimate of the parameter values in order to design an experiment to get a good estimate of the parameter values.

Bayesian and pseudo-Bayesian approaches provide a way to address the design dependence problem. These solutions can range from simply integrating a criterion over a prior distribution to a full decision-theoretic treatment of the screening decision problem. Chaloner and Larntz (1989) searched for an exact design ξ in the space of all candidate designs that maximizes the Bayesian D-optimality (D_{B}-optimality) which we denote as $\Phi(\xi)$. This criterion is with respect to a prior distribution $p(\boldsymbol{\theta})$ over a parameter space Θ:

$$\Phi(\xi) = \int_{\Theta} \log |\mathbb{M}(\xi, \boldsymbol{\theta})| p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}. \tag{1}$$

Evaluating (1) often requires numerical integration methods, but when the dimension of Θ becomes larger than 1 or 2 the number of samples required by Monte Carlo integration techniques makes such techniques prohibitively expensive. A cubature scheme from Gotwalt et al. (2009) has made possible the evaluation of and search over this criterion. It is implemented in the commercial software JMP, Version 10 (1989–2013).

The D_{B} criterion has an appealing interpretation when parameters are fit by maximum likelihood. It can be regarded as simply integrating a weighted D criterion over the parameter space, in which case it can be thought of as minimizing the volume of the expected asymptotic confidence region of the parameters. In terms of a full Bayesian formulation, the D_{B} criterion can be seen as maximizing the expected Shannon information (among other interpretations).

For screening experiments, minimizing the length of parameter estimate confidence intervals can be more important than minimizing the covariance determinant because the confidence intervals are related to tests of active effects for which the screening experiment is conducted. Reliance on asymptotic normal approximations

can result in the nominal confidence intervals that are much wider than calculated due to the inaccuracy of the asymptotic result for small samples and changes depending on the location of the design points. This can happen for the following reasons:

- Compromises in how the prior is defined (i.e. as a product of independent priors) needed for computation, overly broad priors, and/or the curse of dimensionality can lead to designs where the number of distinct points collapses to very few stuck on or near the boundary of the design space.
- When the GLM is for a lattice distribution then the small sample behavior can change dramatically across small regions. This can cause the design that optimizes D_{B}-criterion to be very near a design that has superior performance in terms of both the parameter estimate variances and the parameter confidence intervals over a majority of the mass of the prior distribution.
- For the binomial GLM, the s-shaped link function means that designs that differ greatly in parameter values may have very similar shapes over the design space, and this requires attention when defining the prior to avoid over-weighting certain models that tend to draw the D_{B}-optimal design points toward the boundaries of the design space.

The same small sample behavior that results in poor confidence interval coverage can also result in an incorrect determinant of the covariance matrix. We examine these issues for the binomial GLM to demonstrate the effect on optimal screening designs.

Consider the following motivating example. A D_{B}-optimal screening design was constructed for the model from the Challenger data set in Faraway (2005) with a binomial GLM where $x \in [30, 80]$ with i.i.d. samples $X_i \sim \mathrm{Bin}(6, \pi(x_i))$ and the canonical link. The model in Faraway (2005, p. 27) has $\theta_0 = 11.66$ and $\theta_1 = -0.2$. Using a very broad independent prior for each parameter

$$\theta_0 \sim \mathrm{Unif}(-72, 72) \quad \text{and} \quad \theta_1 \sim \mathrm{Unif}(-0.4, 0.4)$$

yielded the D_{B}-optimal design

$$\left[\begin{array}{cccccccc} 30 & 30 & 33.87 & 33.87 & 76.05 & 76.05 & 80 & 80 \end{array}\right].$$

Using the given design, the fidelity of the asymptotic approximation for parameter estimate variance was measured by looking at the ratio of the determinant of the Wald covariance matrix versus the determinant of the numerically estimated parameter estimate covariance matrix. This was then compared to the same procedure for a design (discussed later, and in Fig. 5) where the design points were manually spread out.

Parameter values were randomly sampled from uniform distributions around the true value of the model, with $\theta_0 \sim \mathrm{Unif}(8.66, 14.66)$ and $\theta_1 \sim \mathrm{Unif}(-0.6, -0.2)$. This generated 500 pairs, and for each pair 5,000 samples were simulated from the binomial GLM. The D_{B}-optimal design given above (which will be referred

to as design 1) had poorer accuracy for the Wald approximation over 80 % of the parameter space when compared to a manually generated design (where the design points are more spread out, as is described later in Sect. 4). This demonstrates that the accuracy of the D_B-criterion in reflecting the parameter estimate variances can be influenced by the design.

Perhaps more important for a screening experiment, the same phenomenon can be seen for the confidence intervals. Holding θ_1 at 11.66, the coverage probability of a 95 % confidence interval was examined as θ_1 varied over 500 equally spaced values in $[-0.4, 0]$. At each parameter setting 5,000 samples were generated according to the true distribution and the coverage of Wald-type confidence intervals were measured.

The true coverage of the 95 % confidence interval ranged between 87.2 % and 100 %. The coverage estimates are presented in Fig. 1. The amount of over-estimation of the design's local performance exceeds the under-estimation raising questions about the overall performance of the design in terms of estimating the value of the parameters. When the prior for θ_1 is mainly contained within $[-0.3, -0.2]$ then the criterion systematically overestimates the design quality for screening experiments. When the prior for θ_1 is mainly contained within $[-0.4, -0.3]$ then the criterion systematically underestimates the quality of the design for screening experiments.

In Sect. 2 we discuss approaches to screening designs for non-linear models. In Sect. 3 we discuss concerns for choosing a prior distribution and problems that can develop from compromises in the specification. In Sect. 4 we discuss how the difference between $\log|\mathbb{M}|$ and the covariance of the parameter estimates at each $\boldsymbol{\theta}$ can lead to a D_B-optimal design that is near a design with better D-efficiency over

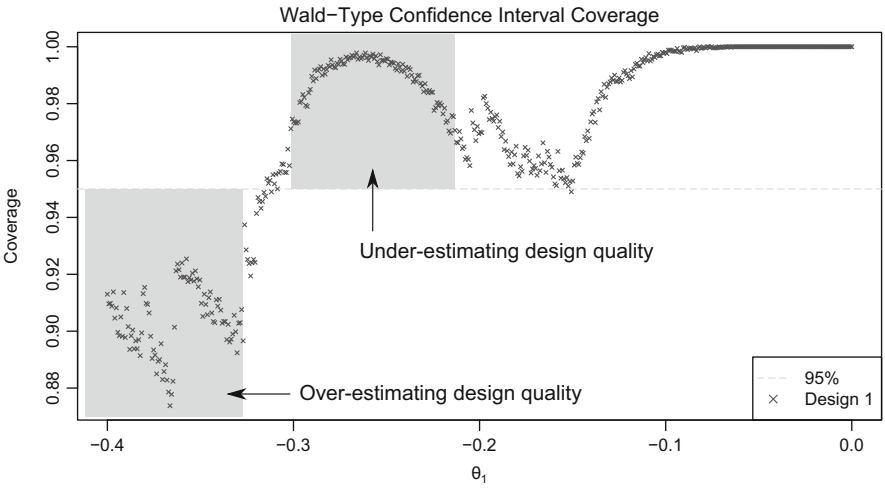


Fig. 1 Wald-type confidence interval coverage for 95 % confidence intervals as θ_1 in the Challenger data example from Faraway (2005) is varied

a majority of the parameter space. Section 4 also describes how the small sample behavior of the parameter estimate variances can deviate from the asymptotic value in the same way as the confidence intervals.

2 Optimal Non-linear Design

As reviewed in Khuri et al. (2006) and Atkinson et al. (2007), the design dependence problem has been addressed several ways in the literature. Perhaps the easiest approach to create a design is the use of *locally optimal* designs where the practitioner simply guesses at the true value of the parameters, perhaps by relying on historical data or subject matter experts. Locally optimal designs based on $\boldsymbol{\theta}'$ when the true parameter value is $\boldsymbol{\theta}$ may have poor performance the larger the distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. The *maximin* design is

$$\arg\max_{\xi} \min_{\boldsymbol{\theta}\in\Theta} \log|\mathbb{M}(\xi,\boldsymbol{\theta})|,$$

where Θ is the bounded space of parameters that are to be considered. Atkinson et al. (2007) noted that this approach is not always desirable as it can tend to overemphasize the importance of models with parameters on the edge of the parameter space.

When a full prior is defined over Θ then we may use a Bayesian approach. Chaloner and Verdinelli (1995) and DasGupta (1995) discuss Bayesian experimental design in terms of a decision theoretic approach from Lindley (1972). A utility function $U(d,\boldsymbol{\theta},\xi,Y)$ is defined in terms of the design ξ, possible data Y, parameter values $\boldsymbol{\theta}$, and terminal decision d. Following Clyde (2001) the *posterior* expected utility is

$$\max_{d}\int_{\Theta} U(d,\boldsymbol{\theta},\xi,Y)\,p(\boldsymbol{\theta}|Y,\xi)\,\mathrm{d}\boldsymbol{\theta} = U(\xi,Y)$$

and the *pre-posterior* expected utility $U(\xi)$ is

$$\begin{aligned} U(\xi) &= \int_{\Omega} U(\xi,Y)\,p(Y|\xi)\,\mathrm{d}Y \\ &= \int_{\Omega}\max_{d}\int_{\Theta} U(d,\boldsymbol{\theta},\xi,Y)\,p(\boldsymbol{\theta}|Y,\xi)\,p(Y|\xi)\,\mathrm{d}\boldsymbol{\theta}\,\mathrm{d}Y. \end{aligned}$$

The optimal design is then $\xi^{\star} = \arg\max_{\xi} U(\xi)$.

Chaloner and Verdinelli (1995) note a number of utility functions that yield the D_{B} criterion in (1), including maximizing the expected Shannon information

$$U(x,\boldsymbol{\theta},\xi,Y) = \log\{p(\boldsymbol{\theta}|Y,\xi)\},$$

where they rely on a normal approximation

$$\boldsymbol{\theta}|Y,\xi \sim \mathrm{N}(\hat{\boldsymbol{\theta}}, [n\mathbb{M}(\hat{\boldsymbol{\theta}}, \xi)]^{-1}).$$

They suggest another approximation

$$\theta|Y,\xi \sim \mathrm{N}(\hat{\boldsymbol{\theta}}, [\mathbb{R} + n\mathbb{M}(\hat{\boldsymbol{\theta}}, \xi)]^{-1}),$$

where $\mathbb{R}$ is the matrix of second derivatives of the logarithm of the prior density function, and represents the contribution of the prior to the information matrix of the joint distribution. The information matrix from the first approximation is only with respect to the likelihood and ignores the prior, whereas the information matrix in the second approximation is with respect to the joint distribution of the response and parameters. This second approximation yields a different criterion

$$\int_{\Theta} \log|\mathbb{R} + \mathbb{M}(\xi, \boldsymbol{\theta})|\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}. \tag{2}$$

Chaloner and Verdinelli (1995) presented strengths of both approximations but recommend using (1) for reasons including avoiding specification of $\mathbb{R}$, the appeal of such criterion in non-Bayesian frameworks, and less dependence on sample size.

Atkinson et al. noted that the approach of (1) is sometimes called a pseudo-Bayesian approach since $\mathbb{M}$ does not incorporate $p(\boldsymbol{\theta})$ in its expectation as in (2). This independence of $\mathbb{M}$ from the prior, in the case of GLMs, allows for an easy modification to make (1) robust to link choice. Let link g_i have prior probability π_i. Then letting $\mathbb{M}_{g_i}$ be the information matrix with respect to link g_i we can use

$$\sum_i \pi_i \int_{\Theta_i} \log|\mathbb{M}_{g_i}(\xi, \boldsymbol{\theta})|\, p(\boldsymbol{\theta}|g_i)\, \mathrm{d}\boldsymbol{\theta}$$

as a robust criterion. Firth and Hinde (1994) and Lohr (1995) showed that (1) is invariant to reparameterization. Atkinson et al. (2007) also described many other criteria called Bayesian D-optimality, Sándor and Wedel (2001) used the kth root instead of log for scaling within the integral in (1), and other small changes to (1) exist in the literature. In this paper we rely on (1) defined in Chaloner and Larntz (1989) as it occurs often in the literature and is implemented in JMP's Nonlinear Design Platform.

An advantage that D_{B}-optimal designs have over locally optimal designs is that they tend to spread their design points around the design space more so than the locally optimal design, making the D_{B}-optimal design more robust to poor guesses for $\boldsymbol{\theta}$ than the locally optimal design. There are two cases where there will be very few distinct design points in the D_{B}-optimal design: when the prior is very narrow, or when the prior is very broad. When the prior is narrow, the D_{B}-optimal design nears locally optimal designs. Section 3 describes how broad priors can cause design points to be forced to a few places on the boundary.

D_B-optimal designs are computationally expensive to generate because the integral in (1) is very difficult to calculate, more so as Θ increases beyond 1 or 2 dimensions. Several approaches to this problem have been proposed. Woods et al. (2006) used simulated annealing to search the design space and estimates based on the sum of local optimality criteria. Dror and Steinberg (2006) sampled from the prior and used clustering on the generated locally optimal designs to achieve good designs. Gotwalt et al. (2009) relied on a cubature scheme that provides high accuracy with very few evaluations for several sample problems, and this method is used throughout this paper to calculate D_B-optimal designs.

Simple coordinate exchange (Meyer and Nachtsheim 1995) is used in the method to search the design space for an optimal design. The cubature itself is related to the *radial-spherical* cubature from Monahan (2001) used for calculating normal-prior Bayesian integrals. In the implementation by Gotwalt et al. (2009), a set of radii are chosen according to a Radau-Gauss-Legendre rule (see Davis and Rabinowitz 2007). Random rotations are then applied to spherical cubatures at each radii to estimate (1). This approach shows surprising accuracy for very few cubature points, increases in complexity only polynomially with the parameter space, and has very few tuning parameters. Also, the speed of the algorithm makes this method preferable to the other two methods for several sample problems.

3 Utility and Prior Specification

There is little benefit to the practitioner performing a screening experiment where the mean response does not change by some practical minimum over the design space. In the identity link case with standardized design space, the size of the parameters is additively related to the change in the mean response, such that if the mean response does not change by some minimum then none of the parameters may be very large. However, this is not generally the case, and there are situations when models with large parameters may have little utility to the practitioner.

In general there are two situations when the mean response can be mainly flat. The first is when the non-intercept parameters are all nearly zero. In this case the lack of utility is clear. The second situation arises when a link function is very flat in one of its tails. For binomial GLMs the link functions are all necessarily flat in both tails. When a link has a flat tail then a large parameter value of the intercept or a non-centered factor may shift the linear predictor into this tail area of the link function.

This second situation has little utility because changes in any of the factor levels has almost no effect on the mean response over the entire design space. If a practitioner uses an intercept only model, then this will not deviate very far from the mainly flat model. Statistical tests will have a difficult time differentiating the non-intercept parameters from zero, and improving the sensitivity of such tests will serve only to make clear that the associated factor has little effect on the mean response over the design space.

The negligibility of factors with large parameter magnitudes can be seen even in the normal linear model: If a parameter is not exactly zero but small, we can make it as large as we wish simply by dividing the factor levels by a large constant. Standardizing the factor levels prevents this from happening in the normal linear model, and can prevent this phenomenon from happening to the non-intercept terms in GLMs but not the intercept itself.

The cost of ignoring these mainly flat models is that tests that the parameters are non-zero are less likely to differentiate the parameter from zero. Yet, since the intercept-only model is very close to the true model, very little utility is lost.

This is in stark contrast to the benefit when some effect is practically significant. In such a case, properly identifying an effect as practically significant is the goal of the screening experiment. False identification of an effect as important (a type I error) causes a slight inconvenience for the practitioner in the presence of some correctly identified important effects. Only failure to identify an important effect (a type II error) greatly reduces the benefit of the experiment to the practitioner.

These values are not reflected in $U(d, \boldsymbol{\theta}, \xi, Y)$. Instead, $U(d, \boldsymbol{\theta}, \xi, Y)$ as in (1) can be seen as relentlessly focusing on reducing the determinant of the covariance of the parameter estimates, even over models where better estimates do no good. This issue may be addressed in two equivalent ways: either the utility function can be weighted to discount parameter values that give rise to models where the response does not change by some practical minimum, or the prior may be restricted away from such parameters.

Taking a more liberal view of the prior in (1) beyond a representation of how believable certain parameters are to the practitioner lends perspective to the performance of the design. Given a prior p_1 with bounded support, expanding that support to p_2 has the effect of sacrificing the performance of the design over the support of p_1 to improve the performance over p_2. Thus, the practitioner should ask themselves if they would be willing to sacrifice efficiency over part of the parameter space to gain efficiency over another part. Put another way, it may make sense to craft a prior to cover certain models where the efficiency of the estimators is more important than in other models.

In addition to a loss of efficiency, some non-linear model D_{B}-optimal designs can have the number of distinct points collapse in the presence of a large proportion of models that are very flat over the design space. This may mean that models where change is focused within the design space will have lower efficiency than alternative designs. This can cause further issues as discussed in Sect. 4.

Most GLMs require the linear predictor to be nearly constant for the response to be nearly constant, but the binomial GLM can exhibit nearly constant responses when the parameters are large in magnitude. If a practitioner constructs a prior that is overly broad, either intentionally or as a simplification for computational concerns, then the D_{B}-optimal design can give too much importance to such nearly constant models. Computational difficulties arise when the prior covers models where the value of $\log|\mathbb{M}|$ can vary widely.

3.1 GLMs with Near Constant Responses

GLMs with a near constant linear predictor have very flat responses and tend to bias the D_{B}-optimal design by drawing points to the boundary of the design space. In the case of the binomial GLM, when the linear predictor is nearly zero everywhere the relative importance of the model in (1) is maximal. We may write the information matrix for GLMs as $d(\boldsymbol{\theta})\mathbb{X}'\mathbb{V}\mathbb{X}$ where $d(\boldsymbol{\theta})$ is related to dispersion (which is constant for the binomial model) and $\mathbb{V}$ is a diagonal matrix with elements that are functions of the mean for each design point (Myers et al. 2012). The entries of $\mathbb{V}$ for the binomial GLM are proportional to $\pi(x_i)\big(1-\pi(x_i)\big)$ which is maximal when the linear predictor is zero.

The binomial GLM also admits models where the parameters of the linear predictor are large but the response over the design space is very near 0 or 1. These models have an intercept in the linear predictor that is very different from the other parameters, causing the response's s-shape to shift out of the design space. These models have very small values of $|\mathbb{M}|$ as samples from within the design space can tell us very little about the shape of the response. The log-scaling of (1) tends to inflate the relative importance of these shifted models, as well as the benefit of small improvements in their determinant.

Figure 2 shows the value of the determinant of the information matrix in terms of fraction of the parameter space from a design over $[-1, 1]$ and with the prior noted on the plot. The D_{B}-criterion is the integral over (the log of) these values, and so examining these plots can give some sense of the relative importance of the various magnitudes of determinants of the local information matrices.

Plots (a1) and (a2) include flat-in-the-design-space models in the prior's support and show a sharp spike in the value of the determinant. The large number of near-zero determinants at regular scale become very large negative values at log scale. Such large negative values make up a proportionately large part of the D_{B}-criterion's value. As such, a design that has higher efficiency for such low-utility models will be attractive to the criterion. Plots (b1), (b2), (c1), and (c2) do not include very flat models, and plots (b1) and (b2) demonstrate the greater regularity of determinant values when flat-in-the-design-space models are ignored. The log-scaling that so inflates the relative importance of low-utility models may be addressed by using an alternative scale (such as using the kth root instead of log as in Sándor and Wedel 2001).

Another feature of (a2) is the fast bend about 30 % that may indicate complexity of the log-determinant surface beyond the relatively low-order polynomial approximation of the cubature scheme. In comparison with (a1) and (a2), the change in determinant for (b1) and (b2) over the parameter space is much less, and the smoothness over more of the parameter space suggests that the cubature scheme may perform better. Empirically, we found numerical issues with several designs that considered many flat-in-the-design-space models, yet had no such issues when such models were not considered.

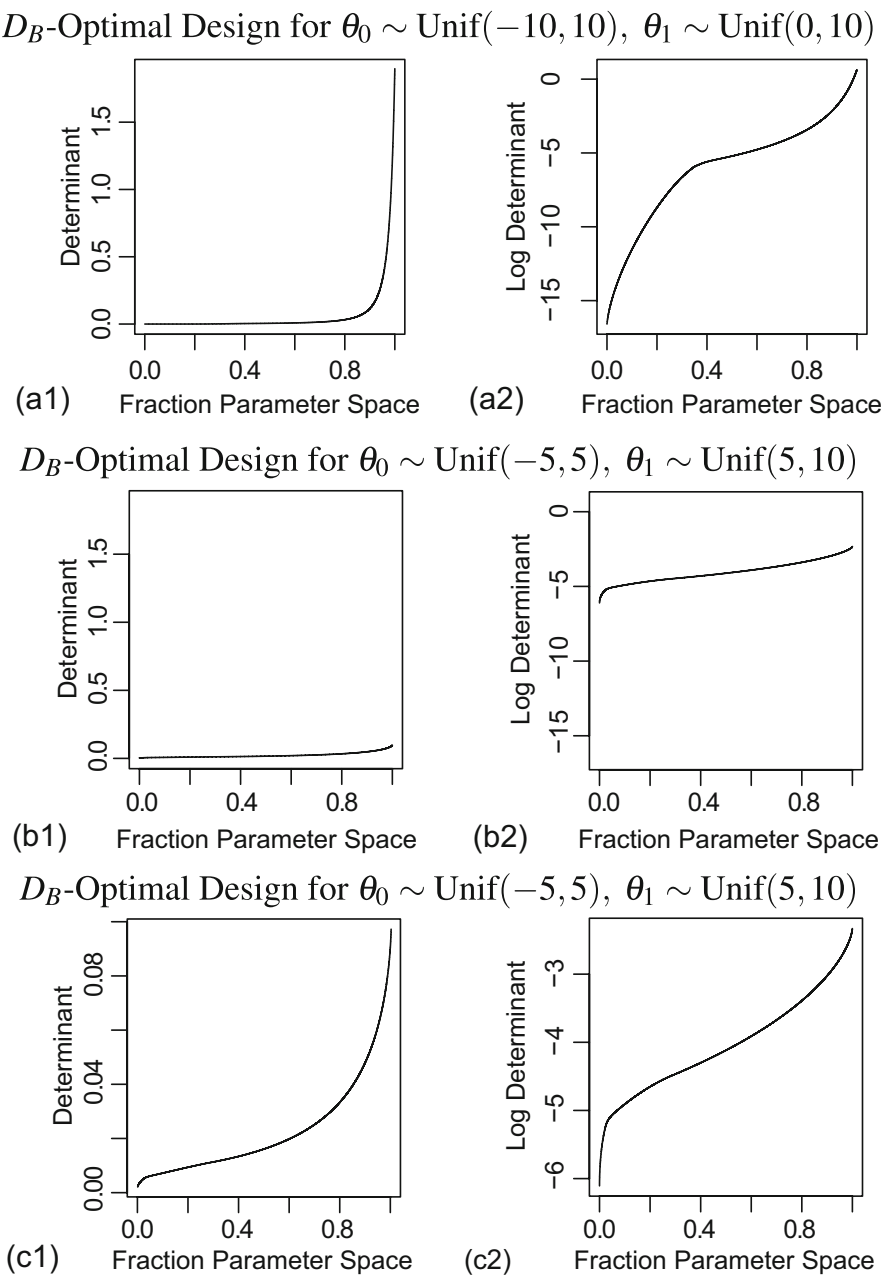


Fig. 2 Fraction of parameter space plots for the D_{B}-optimal designs under various priors. Plots (*a1*), (*b1*), and (*c1*) show the fraction of the parameter space at or below that determinant, whereas plots (*a2*), (*b2*), (*c2*) show the fraction of the parameter space at or below that log determinants. Plots (*a1*) and (*a2*) show the behavior of the determinant when models that do not vary in the design space are in the prior's support. Plots (*b1*), (*b2*), (*c1*), and (*c2*) represent the same design, that only includes in the prior models that vary in the design space, plotted at different scales

3.2 Simplified Priors and Shifted Models

A practitioner constructing the prior for a binomial GLM screening design may include a single effect that is dramatic or several effects that are more modest yet important. The algorithm from Gotwalt et al. (2009) requires the prior to be of a specific form, and the implementation in JMP 10 requires independent prior distributions for each parameter. If the practitioner simplifies their prior to maintain coverage, then they will include many shifted models in the periphery of the simplified parameter space. This can quickly become a problem as the number of parameters can increase exponentially with the number of factors, and the proportion of the peripheral to the center of the parameter space increases exponentially in the number of parameters.

Shifted models in locally optimal designs force the design points toward the boundary of the design space, and the effect of including such models in the prior is D_{B}-optimal designs with points focused near the boundary of the design space. Consider regression on a simple binomial GLM with the canonical link and mean response $\pi(x)$. The parameterization in terms of θ_0 and θ_1 is not as intuitive as a parameterization in terms of the 50 % quantile ϕ_0 and a scale parameter ϕ_δ. Let ϕ_δ be the value of t so that $\pi(\phi_0 + t/2) - \pi(\phi_0 - t/2) = \delta$. We may express the relationship of these parameterizations as

$$\phi_0 \propto \tfrac{\theta_0}{\theta_1} \; \phi_\delta \propto \tfrac{1}{\theta_1} \; \theta_0 \propto \tfrac{\phi_0}{\phi_\delta} \; \theta_1 \propto \tfrac{1}{\phi_\delta} \; .$$

Note that increasing θ_1 affects both the center and scale of the link function. Suppose that our best representation of the prior is $\phi_0 \sim \mathrm{Unif}(30, 80)$ and $\phi_\delta \sim \mathrm{Unif}(5, 50)$ independently. Figure 3 shows the prior in the reparameterized space (a) and the prior in the original parameterization (b). Surrounding these are the models and locally optimal designs at those models (c1), (c2), (c3), and (c4) that are included in a simplified independent uniform prior over the θ's but not included in the original reparameterization.

Locally optimal designs for these shifted models have very small determinants and are highly inefficient for practically significant models. The D_{B}-optimal design with the following prior

$$\theta_0 \sim \mathrm{Unif}(-8, 8) \quad \text{and} \quad \theta_1 \sim \mathrm{Unif}(1, 10)$$

is less efficient over 4 % of the parameter space $[-10, 10] \times [1, 10]$ than the D_{B}-optimal design with uniform prior over that space. This 4 % includes other models that vary outside of the design space. By removing 20 % of the models in the parameter space we observed only a slight effect on the efficiency of the D_{B}-optimal design when Θ is low dimensional.

As the dimension of the parameter space increases, the curse of dimensionality and the requirement of a simple prior can quickly increase the fraction of shifted models in the parameter space. The motivating example in the introduction is a case

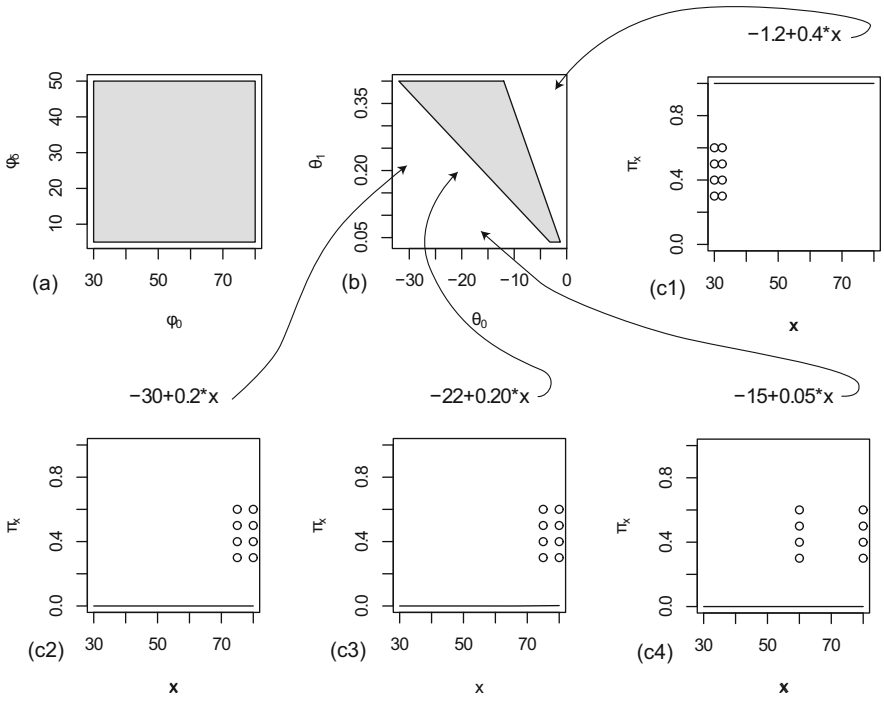


Fig. 3 Plot (*a*) represents an independent prior on the reparameterized space. Plot (*b*) is the region in (*a*) converted into the original parameterization. Plots (*c1*), (*c2*), (*c3*), and (*c4*) show models with eight-point locally optimal designs denoted with *circles*

where the proportion of models varying primarily outside of the design space is large. When the parameter space is of a higher dimension then the $\log|\mathbb{M}|$ diverges further and further from the nearest low-order polynomial. The high variability of the nearly flat binomial GLM models can exacerbate errors in the computation scheme for (1).

3.3 Computational Difficulty

The log-determinant of the information matrix is very smooth everywhere except as the design points collide with the boundary of the design space. Even though it is smooth, the log-determinant can change very quickly when the determinant is near zero. These features can make numerical calculation of an optimal design difficult, and can result in odd behavior. Several times the algorithm generated designs that, when the uniform prior was narrowed, the design became worse over a majority of the narrower prior's support.

The procedure by Gotwalt et al. (2009) calculates weights and abscissae based on spherical cubatures with radii according to the abscissae of a Radau-Gauss-Laguerre scheme. Though we can calculate these radial abscissae and weights by root finding (Press et al. 2007) or by eigenvalue-based (Gautschi 2004) approaches, both approaches quickly degrade due to numerical issues as the number of abscissae increases (e.g., in 64-bit floating point format the weights and abscissae for a 16-point quadrature begin to lose accuracy). Since the spherical cubatures in the procedure have a fixed number of abscissae, we may increase the number of spheres at random rotation, but this does not address the issues of finer features being missed due to the spacing of the radii.

Increasing the number of random starting locations also does not solve the numerical issue since the cubature scheme remains fixed except for the random rotations. The practitioner might consider using a more expensive estimation technique for (1) once a design has been selected, comparing the two estimates of the criterion. Adjusting the prior by moving its center and changing its spread may provide different designs. If the design fails to be robust to such manipulation, then the practitioner could order the designs based on higher fidelity approximations of (1).

4 Poor Coverage and Few Design Points

Wald-type confidence intervals can have poor coverage for lattice distributions like the binomial GLM, and this is related to higher order terms in the asymptotic expansion for the parameter estimates. Parameter estimation for GLMs is often done by maximum likelihood (McCullagh and Nelder 1989), and the maximum likelihood estimates (MLEs) are asymptotically normally distributed. If $\hat{\boldsymbol{\theta}}$ is the MLE for $\boldsymbol{\theta}$ and for a design ξ, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Rightarrow \mathrm{N}\big(\mathbf{0}, \mathbb{M}(\xi, \boldsymbol{\theta})^{-1}\big).$$

The series of papers (Brown et al. 2001, 2002, 2003) describe the issues of coverage related to estimating the proportion in a binomial random variable. They found that for $X_i \sim \mathrm{Bin}(n, p)$ i.i.d. certain "unlucky" combinations of n and p produced radically different coverage for confidence intervals, and that difference between the presumed and realized coverage became larger as p neared 0 or 1. They further showed that the main culprit was the discrete nature of the binomial response. These same issues effect binomial regression, as shown in Fig. 1.

The D_{B}-optimality criterion can be viewed as minimizing the volume of the confidence region around the asymptotically normal parameter estimates. Methods such as that of Firth (1993) exist to correct for some of the bias, but the D_{B}-optimality criterion is blind to such bias, and the act of correcting this bias may introduce new problems for the optimization scheme. It may be that the D_{B}-optimal is a design that is very close to a design with a better capacity to estimate parameters,

both in terms of overall variance of the estimate and in terms of any particular confidence interval. Again, for a practitioner performing screening experiments, the width of particular confidence intervals may be most important.

Let the design from the motivating example be design 1, and consider the new design, design 2, as

$$\left[\begin{array}{cccccccc} 30 & 33.87 & 37 & 45 & 54 & 73 & 76.05 & 80 \end{array}\right].$$

Figure 4 shows the true coverage of the 95 % confidence interval on θ_1 for the two designs. The better spread of the points in design 2 prevented the design points from all accumulating on "bad-luck" values for the various parameters. This suggests that the spread of points which often occurs in D_{B}-optimal designs may serve as a buffer against the bias of (1) with respect to the true parameter estimate confidence intervals.

Figure 5 shows the difference in the relative magnitude of the determinant of the approximate covariance matrix divided by the true parameter estimate covariance matrix (directly estimated numerically, using 5,000 samples per estimate) from unity. A total of 500 samples of the parameters were taken uniformly distributed about the true model from the motivating example with

$$\theta_0 \sim \mathrm{Unif}(8.66, 14.66) \quad \text{and} \quad \theta_1 \sim \mathrm{Unif}(-0.6, -0.2)\ .$$

When the ratio is near one then D-criterion for the Wald approximation is a good approximation of the D-criterion for the parameter estimate covariance, and when the ratio is away from one then the approximation is poor. The fraction of the

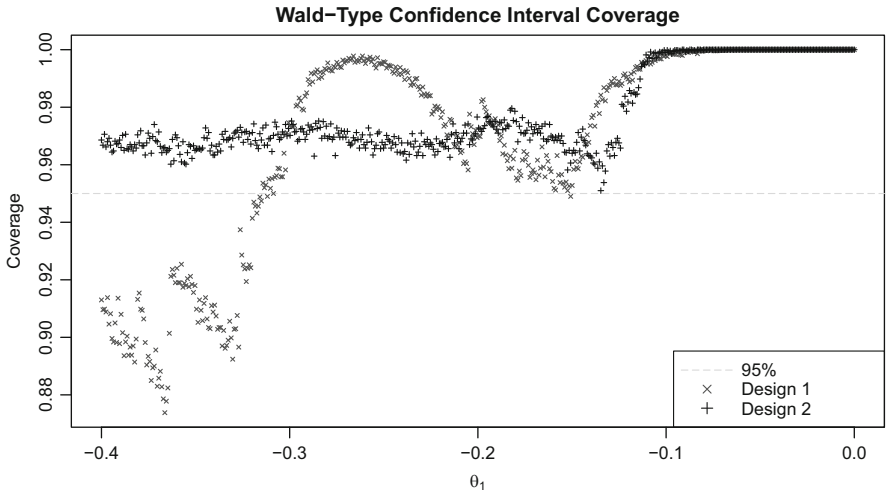


Fig. 4 Wald-type confidence interval coverage for 95 % confidence intervals as θ_1 in the Challenger data example from Faraway (2005) is varied

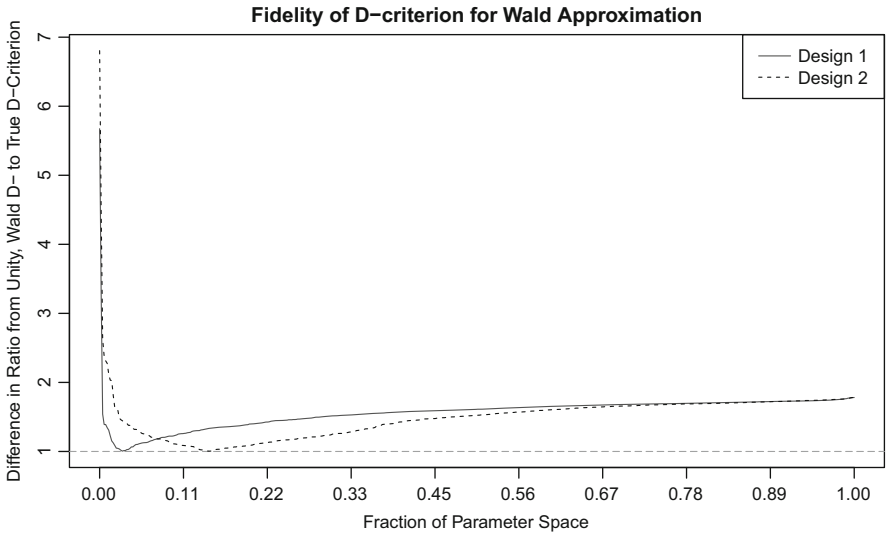


Fig. 5 Fraction of parameter space with the difference from unity of the ratio of Wald asymptotic covariance matrix determinant versus numerically estimated covariance determinant for parameter estimates over the parameter space $\theta_0 \sim \text{Unif}(8.66, 14.66)$ and $\theta_1 \sim \text{Unif}(-0.6, -0.2)$ centered around the values in the Challenger data example from Faraway (2005)

parameter space plot demonstrates that the Wald approximation for Design 2 has better fidelity over a majority of the parameter space when compared to Design 1. Note that no claim is being made about the efficiency of Design 2 versus Design 1, only that the Wald approximation is better for Design 2 versus Design 1 over a majority of the parameter space. The bias that is so prominently displayed in Fig. 4 is present in the parameter estimate variances as well.

Taken together, Figs. 4 and 5 suggest that D_{B}-optimal designs with very few distinct points may have nearby designs that have better performance, both for parameter estimate variances and for parameter estimate confidence intervals. Alternatively, the chance of getting strong bias due to the small sample behavior of the asymptotics is reduced as the number of distinct points increases.

5 Conclusion

We have identified three concerns for D_{B}-optimal designs. First, the prior may be used to modify the measure of the utility of models. We identified two types of models that may reside in high density regions of the prior but that have little utility. These models introduce bias in the D_{B}-optimal design that, under conditions such as extreme misspecification or simplification of priors in higher dimensional parameter spaces, can have a noticeable effect on the design. In the binomial GLM with

canonical link the effect was to force design points towards extremes of the design space. Down-weighting or eliminating such models from the prior can improve the D-efficiency of the design over a majority of the design space.

Second, we described computational issues of the D_{B}-optimal criterion. Priors that include a large proportion of nearly-flat-in-the-design-space models and parameter values may contribute to numerical problems in estimating (1).

Third, we showed how the well-known problem of bias due to relying on asymptotics for lattice and skewed distributions can have an effect on D_{B}-optimal designs with priors that cause the number of distinct design points to be small. When (1) is evaluated over few distinct design points then (1) may diverge from the goals of screening and result in a design that is sub-optimal.

Acknowledgements The authors would like to thank Christopher M. Gotwalt for his helpful correspondence in implementing the algorithm described in Gotwalt et al. (2009).

References

Atkinson, A., Donev, A., & Tobias, R. (2007). *Optimum experimental designs, with SAS*. Oxford: Oxford University Press.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science, 16*(2), 101–117.

Brown, L. D., Cai, T. T., & DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics, 30*(1), 160–201.

Brown, L. D., Cai, T. T., & DasGupta, A. (2003). Interval estimation in exponential families. *Statistical Sinica, 13*, 19–49.

Chaloner, K., & Larntz, K. (1989). Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference, 21*(2), 191–208.

Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science, 10*, 273–304.

Clyde, M. (2001). Experimental design: a bayesian perspective. In N. J. Smelser, et al. (Eds.), *International encyclopedia of the social and behavioral science* (pp. 5075–5081). New York: Elsevier Science.

DasGupta, A. (1995). *Review of optimal bayes designs*. Technical Report #95-4, Department of Statistics, Purdue University.

Davis, P., & Rabinowitz, P. (2007). *Methods of numerical integration*. New York: Dover.

Dror, H. A., & Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics, 48*(4), 520–529.

Faraway, J. J. (2005). *Extending the linear model with R*. Boca Raton: CRC Press.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), 27–38.

Firth, D., & Hinde, J. P. (1994). Invariant optimum design for nonlinear models. Unpublished.

Gautschi, W. (2004). Orthogonal polynomials: computation and approximation. In G. H. Golub, C. Schwab, & E. Süli (Eds.), *Numerical mathematics and scientific computation*. Oxford: Oxford University Press.

Gotwalt, C. M., Jones, B. A., & Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics, 51*(1), 88–95.

JMP, Version 10 (1989–2013). Cary, NC: SAS Institute Inc.

Khuri, A. I., Mukherjee, B., Sinha, B. K., & Ghosh, M. (2006). Design issues for generalized linear models: a review. *Statistical Science, 21*, 376–399.

Lindley, D. V. (1972). *Bayesian statistics - A review*. Philadelphia: SIAM.
Lohr, S. L. (1995). Optimal Bayesian design of experiments for the one-way random effects model. *Biometrika, 82*, 175–186.
McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). Boca Raton: Chapman and Hall.
Meyer, R. K., & Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics, 37*(1), 60–69.
Monahan, J. (2001). *Numerical methods in statistics.* Cambridge: Cambridge University Press.
Myers, R., Montgomery, D., Vining, G., & Robinson, T. (2012). *Generalized linear models: with applications in engineering and the sciences*. New York: Wiley.
Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge: Cambridge University Press.
Sándor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research, 38*(4), 430–444.
Woods, D. C., Lewis, S. M., Eccleston, J. A., & Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics, 48*(2), 284–292.

Bayesian Lasso with Effect Heredity Principle

Hidehisa Noguchi, Yoshikazu Ojima, and Seiichi Yasui

Abstract The Bayesian Lasso is a variable selection method that can be applied in situations where there are more variables than observations; thus, both main effects and interaction effects can be considered in screening experiments. To apply the Bayesian framework to experiments involving the effect heredity principle, which governs the relationships between interactions and their corresponding main effects, several initial tunings of the Bayesian framework are required. However, it is rather unnatural to specify these tuning values before running an experiment. In this paper, we propose models that do not require the initial tuning values to be specified in advance. The proposed methods are demonstrated with screening examples such as Plackett–Burman and mixed-level design.

Keywords Factor interaction • Hierarchical models • Variable selection

1 Introduction

A screening experiment is conducted to identify important factors from a large number of candidates. The purpose of the experiment is to identify main effects of the important factors and to obtain some insights about which factors may be involved in two-factor interaction. Therefore, designs with complex aliasing patterns such as two-level non-geometric Plackett–Burman (Plackett and Burman 1946) are used (Hamada and Wu 1992). When a subset of active factors is selected in this situation, two typical variable selection problems might arise. The first problem is that the number of factors might exceed the number of observations. The second problem is selected subset might not be interpretable, for example, it consists of only interactions.

Bayesian variable selection (Mitchell and Beauchamp 1988; George and McCulloch 1993), Lasso (Tibshirani 1996), and Bayesian Lasso (Park and Casella 2008) can handle the first problem. To manage the second problem, Chipman et al. (1997)

H. Noguchi (✉) • Y. Ojima • S. Yasui
Department of Industrial Administration, Tokyo University of Science, 2641 Yamazaki, Noda, Chiba 278-8510, Japan
e-mail: ngc1120@gmail.com; ojima@rs.tus.ac.jp; yasui@rs.tus.ac.jp

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_21

extended Bayesian variable selection, and Nam et al. (2010) and Noguchi et al. (2012) extended the Lasso. Both the extensions of these techniques were accomplished by incorporating the effect heredity principle, which governs the relationship between an interaction and its corresponding main effects.

Chipman et al. (1997) proposed that Bayesian variable selection combined with the heredity principle can manage both of the variable selection problems outlined above, if proper initial tuning values are incorporated. However, based on which tuning values are incorporated, different subsets of factors might be selected by their method, and it is not always practical to specify tuning values before designing screening experiments.

In this paper, we extend the Bayesian Lasso such that the heredity principle can be taken into account in the variable selection without selecting arbitrary initial tunings. In other words, this method is intended to facilitate the selection of an interpretable and well-fitted subset of factors without incorporating initial tunings.

This paper is organized as follows. In Sect. 2, we review the Lasso and the Bayesian Lasso. In Sect. 3, we propose our models which can incorporate the effect heredity principle. We demonstrate the application of the proposed models through three examples in Sect 4, and conclude with a discussion in Sect. 5.

2 Lasso and Bayesian Lasso

Usually, experimental data can be modeled in the form of the following general linear model

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times f$ matrix of predictors and ε is a vector of n error variables $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ each of which is assumed to be independent and identically distributed as $N(0, \sigma^2)$ and $\beta = (\beta_1, \ldots, \beta_f)^T$ is a vector of parameters. Throughout this paper, we assume $\mathbf{Y}$ and $\mathbf{X}$ are centered such that the observed mean is 0.

The Lasso is proposed for variable selection and parameter estimation; moreover, it is a constrained version of ordinary least squares and, it fits a linear model by solving the optimization problem

$$\hat{\beta}^{\text{Lasso}} = \min_{\beta}(Y - X\beta)'(Y - X\beta) + \lambda \sum_{i=1}^{f} |\beta_i|, \tag{2}$$

where $\lambda > 0$ is controlling the amount of shrinkage for the fitted coefficients. A sufficiently large value of λ produces shrunken estimates of regression coefficients, often with many components equal to 0. Choosing λ can be interpreted as selecting the number of active factors that are included in a subset. In design of experiments,

this characteristic can be interpreted as the effect sparsity principle, which states that the number of active effects in an experiment is relatively small.

The Bayesian Lasso proposed by Park and Casella (2008) was based on the suggestion that estimates of the Lasso can be interpreted as Bayesian posterior modes using normal likelihood and an independent Laplace prior to each regression coefficient β. The problem with the method of Park and Casella (2008) is that its interpretation of the estimates as posterior modes makes it difficult to identify which factors are active, thereby negating the value of the Bayesian Lasso as a means of variable selection.

3 The Bayesian Lasso and the Effect Heredity Principle

We use the following linear regression model proposed by Kuo and Mallick (1998).

$$\mathbf{Y} = \mathbf{X}\mathbf{D}_\gamma \beta + \varepsilon. \tag{3}$$

Here, $\mathbf{D}_\gamma = \text{diag}(\gamma_1, \ldots, \gamma_f)$, γ_h $(h = 1, \ldots, f)$ is a binary indicator variable following Bernoulli distribution. Each γ_h takes the value 0 or 1, indicating whether the corresponding factor belongs to an inactive or an active effect, respectively. Applying the regression model Eq. (3) to the Bayesian Lasso, an inactive factor can be identified by the indicator variable even if its posterior mode wouldn't be equal to 0.

In screening experiments, the main purpose is to identify active main effects and two-factor interactions. Therefore, in our paper, we focus on a regression model with main effects and two-factor interactions.

$$f(x) = \sum_{i=1}^{p} \beta_i \gamma_i x_i + \sum_{i<j} \sum_{j=2}^{p} \beta_{ij} \gamma_{ij} x_i x_j, \tag{4}$$

where p is the number of main factors.

3.1 Strong and Weak Heredity Models

The effect heredity principle can be considered to incorporate two principles. One is the strong heredity principle which allows an interaction to be active only if both corresponding main effects are active. The other is the weak heredity principle, which allows an interaction to be active if one or more of its parent's factors are active. To identify an active subset based on either the strong or weak heredity principles, we reformulate the coefficients for the interaction terms in Eq. (4). Rather

than the single indicator parameter $\gamma_{ij} (i < j, i, j = 1, \ldots, p)$ for the interaction $x_i x_j$, we use all of the related parameters γ_{ij}, γ_i, and γ_j.

We model the strong heredity principle with main effects and linear-by-linear interactions as follows:

$$f_s(x) = \sum_{i=1}^{p} \beta_i \gamma_i x_i + \sum_{i<j} \sum_{j=2}^{p} \beta_{ij} \gamma_{ij} \gamma_i \gamma_j x_i x_j . \tag{5}$$

If either γ_i or γ_j is equal to 0, interaction $x_i x_j$ is removed. We model the weak heredity principle as shown in Eq. (6).

$$f_w(x) = \sum_{i=1}^{p} \beta_i \gamma_i x_i + \sum_{i<j} \sum_{j=2}^{p} \beta_{ij} \gamma_{ij} (\gamma_i + \gamma_j - \gamma_i \gamma_j) x_i x_j . \tag{6}$$

With the expression $(\gamma_i + \gamma_j - \gamma_i \gamma_j) = (1 - (1 - \gamma_i)(1 - \gamma_j))$, interaction $x_i x_j$ is removed when both γ_i and γ_j are equal to 0.

The prior distribution of γ is specified by choosing marginal probabilities for each effect being active. Setting the initial probability as $\Pr(\gamma_i = 1) = 1 - \Pr(\gamma_i = 0) = 0.5$ is natural, because we usually do not know whether an effect is active or not. The above reformulations indicate that the probabilities for higher-order effects can be lower than those for lower-order effects. They account for the effect hierarchy principle, which states that lower-order effects are more important than higher-order effects.

To extend the model for higher order interactions, linear-by-quadratic and quadratic-by-quadratic interactions, we adopt the convention that the parents of a term are those terms of the next lowest order. For example, consider two three-level factors: A and B. The hierarchy among all variables can be described by the diagram in Fig. 1.

Compared to our method, in Bayesian variable selection following effect heredity principle (Chipman 1996), the model hunter has to specify three different probabilities as initial tuning values. Consider two factors, A and B, and one linear-by-linear

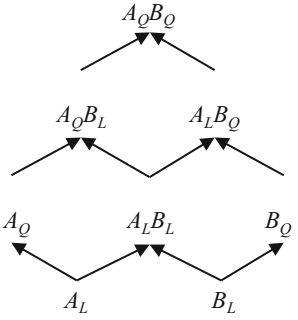


Fig. 1 Ordering and inheritance relations for two-way interaction between three-level factors

interaction, AB. The conditional probability for the interaction AB being active is as follows.

$$\text{Prob}(\gamma_{AB} = 1|\gamma_A, \gamma_B) = \begin{cases} p_0 \text{ if } (\gamma_A, \gamma_B) = (0,0) \\ p_1 \text{ if one of } \gamma_A, \gamma_B = 1 \\ p_2 \text{ if } (\gamma_A, \gamma_B) = (1,1) \end{cases}. \tag{7}$$

Chipman (1996) and Chipman et al. (1997) specified the initial probabilities under the condition of $p_0 = p_1 = 0$, $p_2 > 0$ and $p_0 = 0$, $p_1 > 0$, $p_2 > 0$ in Eq. (7) to incorporate the strong and weak heredity principle, respectively. Bergquist et al. (2011) reviewed some of the initial probabilities in an analysis of unreplicated two-level factorials. However, in any case, it is difficult to explain the propriety of the initial tuning values.

In our method, effect heredity principle is incorporated by reformulation of indicator variables for interaction terms. Therefore, we are able to avoid specifying the initial settings.

3.2 Hierarchical Model

In the Bayesian Lasso, each element of the regression coefficient parameter β has a Laplace prior which can be represented as a scale mixture of normals with an exponential density approach used by Andrews and Mallow (1974).

$$\begin{aligned} \pi(\beta) &= \prod_{h=1}^{f} \frac{\lambda}{2} \exp\{-\lambda|\beta_h|\} \\ &= \prod_{h=1}^{f} \int_0^\infty \frac{1}{\sqrt{2\pi\tau_h^2}} \exp\left\{-\frac{\beta_h^2}{2\tau_h^2}\right\} \cdot \frac{\lambda^2}{2} \exp\left\{-\frac{\lambda^2\tau_h^2}{2}\right\} d\tau. \end{aligned} \tag{8}$$

Based on the approach of Andrews and Mallow (1974), we present the following hierarchical model.

$$\begin{aligned} Y|X, \beta, \sigma^2 &\sim N(XD_\gamma\beta, \sigma^2 I) \qquad D_\gamma = \text{diag}(\gamma_1, \ldots, \gamma_f) \\ \gamma &\sim \prod_{h=1}^{f} p_h^{\gamma_h}(1-p_h)^{(1-\gamma_h)} \\ \beta|\sigma^2, \tau^2 &\sim N(0, \sigma^2 D_\tau) \qquad D_\tau = \text{diag}(\tau_1, \ldots, \tau_f) \\ \sigma^2, \tau_1^2, \ldots, \tau_f^2 &\sim \pi(\sigma^2)d\sigma^2 \prod_{h=1}^{f} \frac{\lambda^2}{2} \exp\left\{-\lambda^2\tau_h^2/2\right\} d\tau_h^2. \end{aligned}$$

We use the reformulated indicator parameters for the corresponding elements of the interaction in D_γ. Park and Casella (2008) suggest using the noninformative scale-invariant marginal prior $\pi(\sigma^2) = 1/\sigma^2$ for the residual variance. Instead of λ, we regard $\lambda^2/2$ as the parameter of interest. We consider the conjugate gamma priors $Gamma(a, b)$ on $\lambda^2/2$. To consider the prior for $\lambda^2/2$ as a noninformative, we set a and b as small value (e.g., $a = 0.1$ and $b = 0.1$).

The joint posterior distribution of the parameters given data $\pi(\beta, \gamma, \sigma^2, \tau^2, \lambda|Y)$ can be simulated using the Markov chain Monte Carlo (MCMC) algorithm. The MCMC algorithm draws from each fully conditional posterior distribution given the current values of all other unknowns and the observed data. Alternatively, β_h is updated from the normal distribution $\pi(\beta_h|Y, \beta_{-h}, \gamma, \sigma^2, \tau^2)$, where β_{-h} represents all elements of β except β_h, and γ is updated from Bernoulli distribution $\pi(\gamma_h|Y, \beta, \gamma_{-h})$. In our proposed models, the above MCMC algorithms can be performed by the Bayesian software WinBUGS14 (Spiegelhalter et al. 1994, 2003).

In our method, variable selection is based on the posterior distribution of the indicator parameter γ. Thus, the subset with the highest posterior probability, which has the most frequent combination of active factors, will be selected. Moreover, median estimates obtained from the posterior distribution of reformulated regression parameters $\gamma_h\beta_h$ might provide probable active factors.

4 Example

In this section, we apply our proposed models to the data obtained from a 12-run Plackett–Burman design and a mixed-level design. These examples are selected to demonstrate the wide applicability of our method.

4.1 12-Run Plackett–Burman Design

Table 1 presents a 12-run Plackett–Burman design and illustrates its use in a screening context that can accommodate up to 11 factors labeled A–K. In the analysis of this design, we consider main effects as well as all two-way interactions among the factors.

4.1.1 Cast Fatigue Experiment

Hunter et al. (1982) used the 12-run design to study the effects of seven factors on the fatigue life of weld-repaired castings, and these factors and levels are listed in Table 2. The seven factors (A–G) were assigned using the first seven columns of the design. The response Y_1 is the logarithm of the lifetime of the casting. This example was analyzed by Hamada and Wu (1992), and the main effect F and interaction

Table 1 Plackett–Burman 12-run design with response data

No.	A	B	C	D	E	F	G	H	I	J	K	Y_1	Y_2
1	+	+	−	+	+	+	−	−	−	+	−	6.058	1.058
2	+	−	+	+	+	−	−	−	+	−	+	4.733	1.004
3	−	+	+	+	−	−	−	+	−	+	+	4.625	−5.200
4	+	+	+	−	−	−	+	−	+	+	−	5.899	5.320
5	+	+	−	−	−	+	−	+	+	−	+	7.000	1.022
6	+	−	−	−	+	−	+	+	−	+	+	5.752	−2.471
7	−	−	−	+	−	+	+	−	+	+	+	5.682	2.809
8	−	−	+	−	+	+	−	+	+	+	−	6.607	−1.272
9	−	+	−	+	+	−	+	+	+	−	−	5.818	−0.955
10	+	−	+	+	−	+	+	+	−	−	−	5.917	0.644
11	−	+	+	−	+	+	+	−	−	−	+	5.863	−5.025
12	−	−	−	−	−	−	−	−	−	−	−	4.809	3.060

Table 2 Factors and levels for cast fatigue experiment

Factors	−	+
A:Initial structure	As received	β treat
B:Bead size	Small	Large
C:Presure treat	None	HIP
D:Heat treat	Anneal	Solution treat/age
E:Cooling rate	Slow	Rapid
F:Polish	Chemical	Mechanical
G:Final treat	None	Peen

FG were identified. With the factors F and FG, predicted fatigue lifetimes $\hat{Y} = 5.7 + 0.458F - 0.459FG$ can be obtained. To increase the predicted life, the setting of factor F and G should be "+" and "−," respectively. Therefore, our goal would be to identify the subset (F, FG), which has an adjusted R^2 (ad-R^2) of 0.871.

We applied the ordinary Bayesian Lasso and our proposed models to the data Y_1 in Table 1. In this example, we computed the posterior median using 50,000 consecutive iterations of the Gibbs sampler after 5,000 burn-in for each factor. From the result of the Bayesian Lasso, the main effect F and interactions AE and FG were identified by applying the threshold value 0.1 to the median absolute estimates. Note that this result is identical to the result obtained by applying the Lasso. Although the interaction AE was identified, its corresponding main effects A and E were not identified. The subset comprising these factors had relatively low ad-R^2 of 0.553. When we applied the strong heredity model, the most frequent subset was (F, G, FG) with posterior probability of 0.261. This subset is consistent with the strong heredity principle and had ad-R^2 of 0.910. Adding G to the desired subset (F, FG) only increased ad-R^2 to 0.039. Thus, the factor G might not be significant, whereas the weak heredity model can identify the subset (F, FG) with the highest posterior probability. Figure 2 shows boxplots for samples from each posterior distribution of the regression parameter in the Bayesian Lasso and our proposed models. When

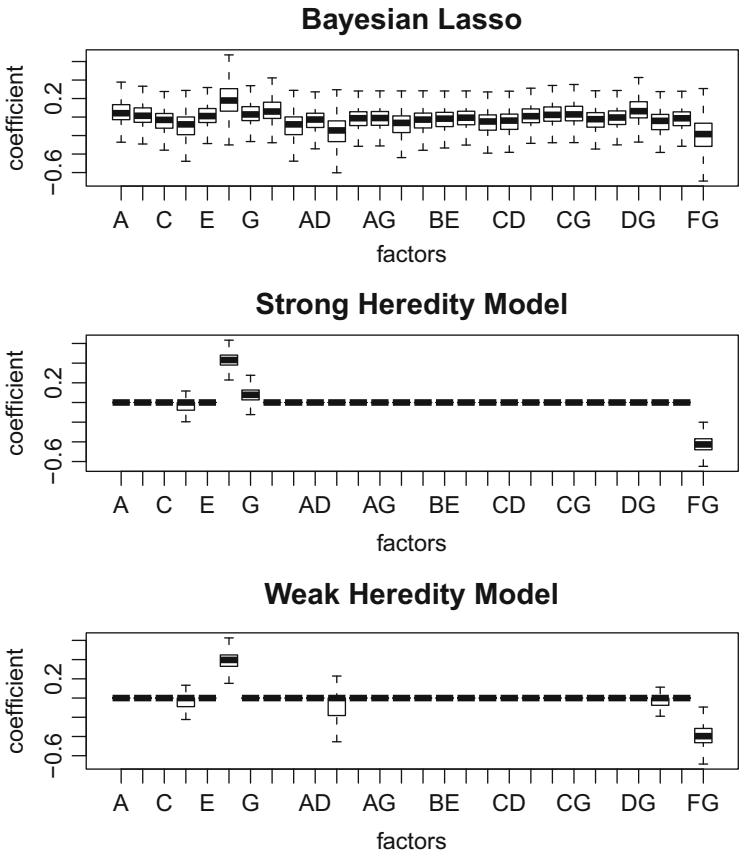


Fig. 2 Boxplots for sample from posterior distributions

using the Bayesian Lasso, identifying which factor is active is difficult; however from our models it is possible to identify which factors may be active.

4.1.2 Simulated Screening Experiment

The data Y_2 in Table 1 were constructed in Hamada and Wu (1992) based on the true model $Y = A + 2AB + 2AC + \epsilon$, here $\epsilon \sim N(0, \sigma = 0.25)$: i.e., factor A had an active main effect and there were active interactions between A and B and between A and C, although the remaining factors D–K were inactive.

The strong heredity model resulted in the selection of the subset A, B, C, AB, AC with posterior probability of 0.398 and its ad-R^2 is 0.996. The weak heredity model resulted in the selection of the subsets A, AB, AC, AJ with posterior probability of 0.298 and a high ad-R^2 of 0.997. The true subset (A, AB, AC) with the second

highest posterior probability of 0.167 was selected with the weak heredity model. However, the posterior mode estimate for the interaction AJ of the A, AB, AC, AJ subset was equal to 0. Therefore, we can say that if we use the weak heredity model, the true subset can be identified. Note that one of the methods proposed by Hamada and Wu (1992) as well as the ordinary Lasso did not identify the main effect A, which shows the advantages of our proposed models.

4.2 Blood Glucose Experiment Using Mixed-Level Design

An 18-run mixed-level design was used to study the effect of eight factors (two- and three-level factors) on blood glucose readings made by a clinical laboratory testing device (Henkin 1986). Factor A had two levels and each of the other seven factors B–H had three levels. The design of the experiment had 18 runs, and the response data are shown in Table 3. In this example, we considered linear-and-quadratic main effects and interaction effects, and denoted the linear and quadratic terms as L and Q, respectively. Each three-level factor was divided into linear and quadratic effects using orthogonal polynomial coding. Consequently, there are 15 linear-and-quadratic main effects and 98 two-factor interactions (28 linear-by-linear ($A_L B_Q, \ldots, A_L H_Q$), 49 linear-by-quadratic ($A_L B_Q, \ldots, A_L H_Q, B_L C_Q, \ldots, G_L H_Q$), and 21 quadratic-by-quadratic ($B_Q C_Q, \ldots, G_Q H_Q$) interactions: therefore, a total of 113 effects were under

Table 3 Plackett–Burman 18-run design with response data

Factor	A	G	B	C	D	E	F	H	Y
1	0	0	0	0	0	0	0	0	97.94
2	0	0	1	1	1	1	1	1	83.40
3	0	0	2	2	2	2	2	2	95.88
4	0	1	0	0	1	1	2	2	88.86
5	0	1	1	1	2	2	0	0	106.58
6	0	1	2	2	0	0	1	1	89.57
7	0	2	0	1	0	2	1	2	91.98
8	0	2	1	2	1	0	2	0	98.41
9	0	2	2	0	2	1	0	1	87.56
10	1	0	0	2	2	1	1	0	88.11
11	1	0	1	0	0	2	2	1	83.81
12	1	0	2	1	1	0	0	2	98.27
13	1	1	0	1	2	0	2	1	115.52
14	1	1	1	2	0	1	0	2	94.89
15	1	1	2	0	1	2	1	0	94.70
16	1	2	0	2	1	2	0	1	121.62
17	1	2	1	0	2	0	1	2	93.86
18	1	2	2	1	0	1	2	0	96.10

Table 4 Selected subsets and their criterion

	Active effects	ad-R^2	AIC
Stepwise	$B_L H_Q$, $B_Q H_Q$, $E_L G_L$, $A_L H_Q$, H_L, $E_L F_Q$	0.9815	68.0
Lasso	$B_L H_Q$, $B_Q H_Q$, $E_L F_L$, $A_L H_Q$, $C_Q G_Q$	0.8035	110.0
Chipman et al. (1997)	B_L, $B_L H_L$, $B_L H_Q$, $B_Q H_Q$	0.8202	107.8
Strong heredity model	B_L, D_L, F_L, H_L, B_Q, D_Q, H_Q, $B_L D_L$, $B_L H_Q$, $D_L H_L$, $F_L H_L$, $H_L B_Q$, $B_Q H_Q$	0.9990	14.6
Weak heredity model	B_L, $B_L H_L$, $B_L H_Q$, $B_Q H_Q$	0.8202	107.8

consideration. We used this example to illustrate how our proposed model can handle a complicated situation. In this example, we computed the posterior median using 50,000 consecutive iterations of the Gibbs sampler after 5,000 burn-in for each factor.

Table 4 shows selected subsets and the value of ad-R^2 and AIC obtained by the ordinary stepwise selection, the Lasso, the Bayesian variable selection with the effect heredity principle (Chipman et al. 1997), and our proposed models.

The stepwise method and the Lasso select uninterpretable subsets that consist of some interactions without corresponding main effects. Our proposed models, especially the strong heredity model, select much better subsets than the others. Both the strong and weak heredity models are able to select the well-fitted subset following the effect heredity principle. We observe that the strong heredity model could select the best subset in this example.

5 Discussion

In this paper, we extended the Bayesian Lasso to incorporate the effect heredity principle by reformulating the coefficients for the interaction terms. In the examples, the proposed models were applied to three examples of screening experiments. Our models could select subsets that were consistent with the effect heredity principle and also fit the data well. Furthermore, using our proposed models, it is possible not only to select a subset with posterior model probability but also to estimate regression coefficients from the posterior distribution. Considering the posterior distribution of regression coefficients might help our inference for variable selection. Moreover, our model can be applied in situations where the number of factors exceeds the number of observations, because it is based on the Bayesian framework.

As we see in the examples, depending on which principle we apply, the selected subset may change and it is not definite which heredity model will select a better subset. Therefore, in practice, it is better to apply both models. Note that the Bayesian approach is useful because of its flexibility in introducing new parameters;

thus, it might be possible to handle the strong and weak heredity principles in a comprehensive manner. The run-size of a screening design might also affect which principles are most appropriate for use, and this could be a topic for future study.

References

Andrews, D. F., & Mallow, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B, 36*, 99–102.

Bergquist, B., Vanhatolo, E., & Nordenvaad, M. L. (2011). A bayesian analysis of unreplicated two-level factorials using effects sparsity, hierarchy, and heredity. *Quality Engineering, 23*, 152–166.

Chipman, H. (1996). Bayesian variable selection with related predictors. *The Canadian Journal of Statistics, 24*, 17–36.

Chipman, H., Hamada, M., & Wu, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics, 39*, 372–381.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*, 881–889.

Hamada, M., & Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology, 24*, 130–137.

Henkin, E. (1986). The reduction of variability of blood glucose levels. In *Fourth Supplier Symposium on Taguchi Methods* (pp. 758–785). Dearborn, MI: American Supplier Institute.

Hunter, G. B., Hodi, F. S., & Eager, T. W. (1982). High-cycle fatigue of weld repaired cast Ti-6A1-4V. *Metallurgical Transactions, 13A*, 1589–1594.

Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhya B, 60*, 65–81.

Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression (with discussion). *Journal of the American Statistical Association, 83*, 1023–1036.

Nam, H. C., William, L., & Ji, Z. (2010). Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association, 105*, 354–364.

Noguchi, H., Ojima, Y., & Yasui, S. (2012). A practical variable selection for linear models. In H. -J. Lenz, W. Schmid, & P. -T. Wilrich (Eds.), *Frontiers in statistical quality control* (Vol. 10, pp. 349–360). Heidelberg: Physica.

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association, 103*, 681–686.

Plackett, R. L., & Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika, 33*, 305–325.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., & Lunn, D. (1994/2003). BUGS: Bayesian inference using Gibbs sampling. Cambridge, UK: MRC Biostatistics Unit. http://www.mrc-bsu.cam.ac.uk/bugs/.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, 58*, 267–288.

Part IV
Related Areas

Comparative Study of Time Scales in Optimal Time Scale Analysis of Field Reliability Data

Watalu Yamamoto and Kazuki Takeshita

Abstract The time scale of failures need not be chronological, if a failure occurs according to the accumulation of damages from its usage or exposure to some risk. The true time scale, if there could be, need not show the best fit among time scale candidates to the observed field reliability data. A semiparametric estimator of the time scale which combines multivariate lifetime data has been proposed in the literature. It does not require the specification of the lifetime distribution beforehand on the time scale also to be estimated. We first show a set of simulation results for investigation of the properties of time scale models and the sample properties of their estimates. Then we show another result on the performance under the time scale misspecifications. Finally the estimator and the time scale functions are applied to a problem of finding a suitable time scale for field reliability data. We show that the results are useful even in cases when there exists prior knowledge about the suitable time scale.

Keywords Competing risks • Counting process • Field reliability data • Optimal lifetime scale

1 Introduction

Laboratory evaluation of the reliability of products, components, parts, etc. enables local control during endurance testing and lifetime testing. The usage conditions and situation can be controlled by introducing appropriate experimental factors and levels, thereby obtaining test results reflecting the various situations each item may encounter. However, it would be impossible to define and test all possible conditions and to run all tests until an instance failed. Therefore, some items may only fail when a customer uses it under certain conditions in certain situations. To improve the reliability performance and safety of *items in the field*, it is important for a company

W. Yamamoto (✉) • K. Takeshita
University of Electro-Communications, Tokyo, Japan
e-mail: watalu@uec.ac.jp; k.takeshita@uec.ac.jp

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_22

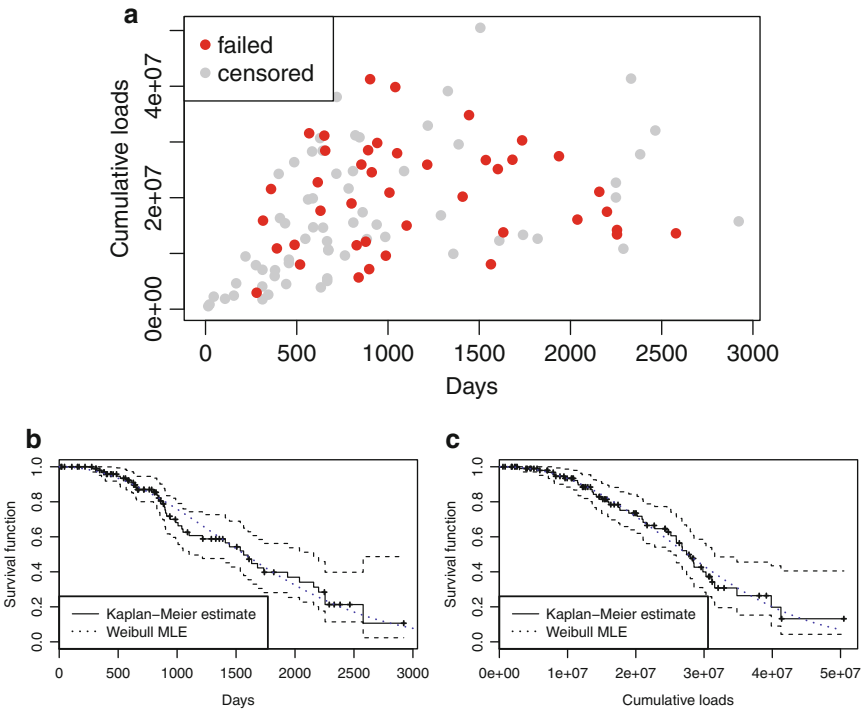


Fig. 1 Graphical presentation of Case A. (**a**) Scatterplot of all records in Case A. (**b**) Weibull probability plot for days. (**c**) Weibull probability plot for cumulative loads

or provider to have a system for collecting field performance data. A warranty claims system is a good source of such information.

This research was motivated by two cases. In Case A, there is a set of bivariate lifetime data on systems with only one failure mode. This data consists of r failure records and $n - r$ censored records. Figure 1 shows a graphical presentation of this case. Figure 1a shows a scatterplot of all records. The x-axis is days X, as a chronological time, and the y-axis is cumulative loads Y, which is the work amounts weighted by loads. Accumulated engineering knowledge supports the use of work amounts Y for the time scale of the lifetime distribution. However, the Kaplan-Meier estimate of the survival function for days X seems to fit better to Weibull distribution than a curve for cumulative loads Y, as shown in Fig. 1b, c.

In Case B, there is another set of bivariate lifetime data on systems with multiple failure modes. This data also consists of r failure records and $n-r$ censored records. A scatterplot of all records is shown in Fig. 2. The failure records are classified into six failure types, which are independent. There are competing risks since the system is considered to have failed once any failure has occurred.

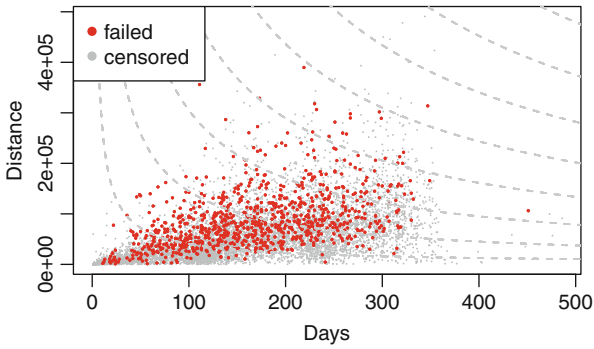


Fig. 2 Scatterplot of all records in Case B. *Dotted curves* are time scales estimated in Sect. 3

The time scale need not be chronological if a failure can occur due to cumulative damage suffered from usage or exposure to some risk. Engineering knowledge can be used to identify a more appropriate time scale, such as mileages or usage count or amounts. However, the actual time scale is often unobservable (e.g., precise rotation counts of each gear, integrals of loads by distances), in which case an observable time scale is needed as an approximation. Moreover, there is often more than one candidate time scale. There has been much researches on combining multiple lifetime scales into a unified time scale ideal for modeling the lifetimes of target items. Some authors call such a time scale ideal or intrinsic.

Farewell and Cox (1975) were possibly the first authors to investigate combining multiple time scales to obtain a more suitable time scale in the context of life testing. Oakes (1995) defined the notion of collapsibility of the time scale and proposed a parametric inference for choosing time scales and failure distributions. Kordonsky and Gertsbakh (1993, 1995a,b, 1997) used minimum coefficient of variation as a criterion to fit linear time-scale models. Duchesne and Lawless (2000, 2002) defined ideal time scales as alternative time scales for multivariate lifetime data problems and proposed a semiparametric estimator based on a counting process approach. Finkelstein (1999, 2004) investigated similar problems in slightly different settings.

In this paper, we apply the method proposed by Duchesne and Lawless (2002) to an industrial problem. We call the estimated ideal time scale the "optimal time scale." We present a series of results from simulation studies done to investigate the properties of time-scale models (linear time scale, multiplicative time scale, and log-linear time scale) and the sample properties of their estimates. We also investigate compatibility among these scales to gain some insight into choosing a time-scale function for analyzing field reliability data.

2 Optimal Lifetime Scale

2.1 Models for Lifetime Scales

Assume that we can observe k usage-amount time scales for each item. They may include the chronological time. We denote the chronological time as s and let $X_j(s)$ be the j-th usage amount at s. There are two major models for lifetime scales. *Linear time scale* is defined as the weighted mean of all scales:

$$t(s;\beta) = \sum_{j=1}^{k} \beta_j X_j(s) \quad \text{with} \quad \sum_{j=1}^{k} \beta_j = 1 \quad \text{and} \quad \beta_j \geq 0 \quad \text{for all } j.$$

The values of parameters $\beta_1, \ldots, \beta_k$ depend on the "scales" of the individual amounts, $X_1, \ldots, X_k$. Applying scale transformation to all variables by using scaling coefficients $c_1, \ldots, c_k$,

$$\tilde{X}_j(s) = X_j(s)/c_j,$$

transforms the parameters into

$$\tilde{\beta}_j = c_j \beta_j / \sum_j c_j \beta_j.$$

The denominator is needed to satisfy constraint $\sum_{j=1}^{k} \tilde{\beta}_j = 1$. This time-scale function is thus scale equivalent and requires attention when the estimated parameters are interpreted. Note that $t(s)$ is invariant under scale transformation.

Multiplicative time scale is defined as the weighted geometric mean of all scales:

$$t(s;\beta) = \prod_{j=1}^{k} X_j(s)^{\beta_j} \quad \text{with} \quad \sum_{j=1}^{k} \beta_j = 1 \quad \text{and} \quad \beta_j \geq 0 \quad \text{for all } j.$$

The values of the parameters are invariant under scale transformation of the individual usage amounts. Note that $t(s)$ is equivariant under scale transformation.

2.2 Estimation of Time Scales

Duchesne and Lawless (2002) proposed a general method for estimating the values of the parameters of time-scale functions that is applicable to cases under general censoring independent of the cause of failure.

We assume for simplicity that the usage path is linear in time unless otherwise noted:

$$X_j(s) = \Theta_j s\,,\ j = 1, \ldots, k,$$

where the Θ_j's are random variables with values equal to the slopes of the usages of X_j against a chronological time scale s. This assumption was also made by Lawless et al. (2009) and many others in the analysis of field reliability data. We also assume that Θ_j is observed for each item when it fails or it is censored.

We assume that lifetime distribution $G(t)$ on optimal time scale $t(s, \Theta; \beta)$ has a known parametric form and let $\lambda(t)$ be the hazard function corresponding to the lifetime distribution. We also introduce $Y_i(s)$ as a function of chronological time s for each item; it is 1 if the i-th item is at risk at chronological time s and 0 otherwise. The log-likelihood for β is given by

$$l(\beta) = \sum_{i=1}^{n} \int_0^{\infty} Y_i(s) \Big(\log \lambda\big(t_i(s)\big) + \log t_i'(s) \Big) \Big(dN_i(s) - \lambda\big(t_i(s)\big) t_i'(s) ds \Big),$$

where $t_i(s) = t_i(s, \theta; \beta)$ and $t_i'(s) = \partial t_i(s, \theta; \beta)/\partial s$.

If assuming a parametric form for $G(t)$ is undesired, the estimation can be done semiparametrically. By defining

$$\tilde{N}_i(t; \beta) = I\left[t_i(s_i) \le t, Y(s_i) = 1\right]$$

and

$$\tilde{Y}_i(t; \beta) = I\left[t_i(s_i) \ge t\right],$$

the function for estimating β is obtained as

$$\tilde{U}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \tilde{Y}_i(t; \beta) \left(Q_i(t; \beta) - \frac{\sum_{j=1}^{n} \tilde{Y}_j(t; \beta) Q_j(t; \beta)}{\sum_{j=1}^{n} \tilde{Y}_j(s_i; \beta)} \right) d\tilde{N}_i(t; \beta),$$

which can be alternatively written as

$$\tilde{U}(\beta) = \sum_{i=1}^{n} \tilde{Y}_i\big(t_i(s_i); \beta\big) \left(Q_i(s_i, \theta_i; \beta) - \frac{\sum_{j=1}^{n} \tilde{Y}_j\big(t_i(s_i); \beta\big) Q_j(s_i, \theta_j; \beta)}{\sum_{j=1}^{n} \tilde{Y}_j\big(t_i(s_i); \beta\big)} \right), \tag{1}$$

with

$$Q_i(s_i, \theta_i; \beta) = \frac{\partial}{\beta} \log \frac{\partial}{\partial s} t_i(s, \theta; \beta) \Big|_{s = s_i}.$$

Solving $\tilde{U}(\beta) = 0$ is not easy since the estimation function $\tilde{U}(\beta)$ is not continuous. Duchesne and Lawless (2002) defined the estimator of β as

$$\hat{\beta} = \arg\min \tilde{U}(\beta)^t \tilde{U}(\beta). \tag{2}$$

As is illustrated in Duchesne and Lawless (2002), the resulting estimator is not so complicated if the time scale is linear or multiplicative. In the case of a univariate usage path and a linear time scale, $t(s;\beta) = (1-\beta)s + \beta\theta s$, $Q_i(s_i, \theta_i; \beta) = (1-\beta)/(1-\beta+\beta\theta_i)$, and $\tilde{U}(\beta)$ becomes

$$\tilde{U}(\beta) = \sum_{i=1}^{n} \delta_i \left(\frac{1-\beta}{1-\beta+\beta\theta_i} - \frac{\sum_{j=1}^{n} \tilde{Y}_j\left(t_i(s_i);\beta\right)(1-\beta)/(1-\beta+\beta\theta_i)}{\sum_{j=1}^{n} \tilde{Y}_j\left(t_i(s_i);\beta\right)} \right),$$

where δ_i is 1 if the i-th item fails and 0 if it is censored.

The semiparametric estimation procedure is preferable in that it does not depend on distribution $G(t)$. Duchesne and Lawless (2002) demonstrated numerically that this procedure is as precise as fully parametric inference if a parametric form can be assumed for $G(t)$.

2.3 Lifetime Scale Estimations for Competing Risks Cases

The procedure proposed by Duchesne and Lawless (2002) can be applied to cases with competing risks without modification. The situation is very similar to that of the maximum likelihood estimation of parameters of competing risks models. Assume that there are k mutually independent competing risks, i.e., failure modes. Once the time scale for each failure mode has been estimated, censored items are treated as censored, and items that failed in other failure modes are also treated as censored. This enables estimation function (1) to be formulated for each failure mode. Of course, this is a marginal approach to inference in that the contribution of the censored records can be varied by constructing a set of the simultaneous estimation functions since the time scales can vary with each other.

3 Time Scale Analysis of a Field Reliability Data

3.1 Case A

In Case A, we assume a linear trace for each usage history and use a stratification approach, as was done in Duchesne (2004). First, we partition the data set into six strata in accordance with usage rate y/x. Then the parameters of the Weibull distribution are estimated for each stratum. Table 1 shows the estimated parameters.

Table 1 Estimated parameters of Weibull distributions for each stratum and for multiple time scales

Stratum	Shape parameter	Scale parameter
1	3.70	1.42×10^7
2	2.80	1.65×10^7
3	2.26	2.95×10^7
4	2.53	3.05×10^7
5	3.39	3.71×10^7
6	4.04	3.43×10^7
Time scale	3.13	3.17×10^5

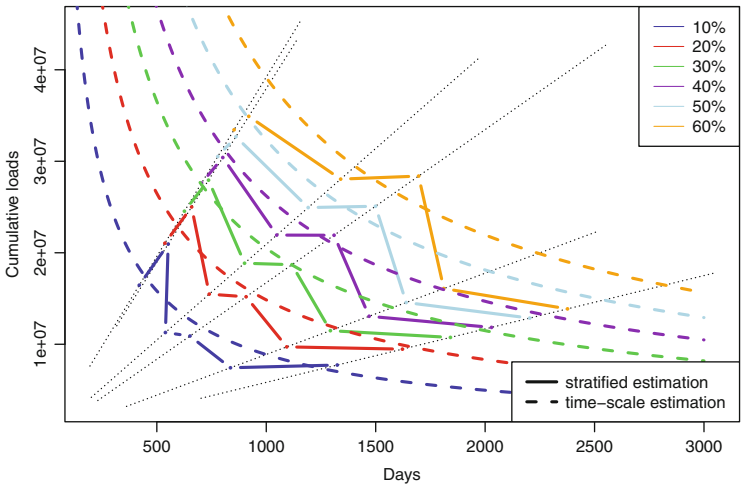


Fig. 3 *Contour lines* of estimated time scale and those of stratified analysis

From these estimates, we can estimate the conditional percentiles for each stratum. The piecewise linear lines in Fig. 3 connect the estimated conditional percentiles of the strata.

The time-scale analysis can estimate the contours of the bivariate lifetime in a much simpler way than the stratified approach. First, we fit the multiplicative time scale to the failure data and obtain the estimate $\hat{\beta} = 0.543$. Then we fit a single Weibull distribution to the transformed data $t_i = {x_i}^{\hat{\beta}} {y_i}^{1-\hat{\beta}}$ either graphically or numerically. This results in the model identification problem becoming that for the univariate case. The estimated Weibull distributions on the multiplicative time scale are shown at the bottom of Table 1. The shape parameter estimate, 3.13, is similar to the stratified estimates.

The percentiles estimated using a multiplicative time scale and the Weibull fit are shown in Fig. 3 as smooth curves. These curves are given by three parameters, much fewer than the 12 parameters used to identify Weibull distributions for each stratum.

The estimated quantiles under the Weibull assumption are shown in Fig. 3. The piecewise linear lines are stratified estimates, and the curves are estimates from the multiplicative time scale. Both sets of estimates share the same tendencies.

3.2 Case B

The set of field reliability data for a component in an electrical appliance from an anonymous industrial company is considered in Case B. When a component instance is replaced due to failure or for preventive maintenance, the number of days it was in use and the total usage amount are added to the dataset. This data set consists of the records of about nine thousand items—904 failures were observed and the rest of items were replaced before failure and are analyzed as censored. The six failure modes are labeled alphabetically, A—F.

The observed time scale for usage amount Y is roughly 100 longer than that for chronological time X. Since the linear time scale gives a poor fit even after appropriate scale transformation, we use only the multiplicative time scale.

Table 2 shows the parameter estimates for the multiplicative time scale for each failure mode, the number of records observed, and the censored ratio for the marginal inference for each failure mode. The data set contains more than 100 failure records for failure modes A, B, and C. Failure modes A and C are sensitive to changes in Y, i.e., they have a $\hat{\beta}$ larger than 0.5. Figure 4a, c show contour plots of the estimated multiplicative time scales for modes A and C, respectively. These contour plots are parallel to the x-axis when Y is not so large. Failure mode B is sensitive to changes in X, in that $\hat{\beta}$ is smaller than 0.5 for this mode. These results suggest that failure modes A and C can be modeled on a usage amount scale, and mode B can be modeled on a chronological time scale.

Note that the scatterplots are for failed and censored data. Though the estimated time scale covers the area with no dots, we do not claim that this scale can be used for extrapolation.

Table 2 Estimates of lifetime scale parameters for each failure mode

Failure mode	$\hat{\beta}$	Observed	Censored ratio (%)
All	0.543	904	89.94
A	0.695	321	93.51
B	0.332	188	97.91
C	0.618	116	98.71
D	0.329	73	99.19
E	0.397	40	99.56
F	0.528	36	99.60

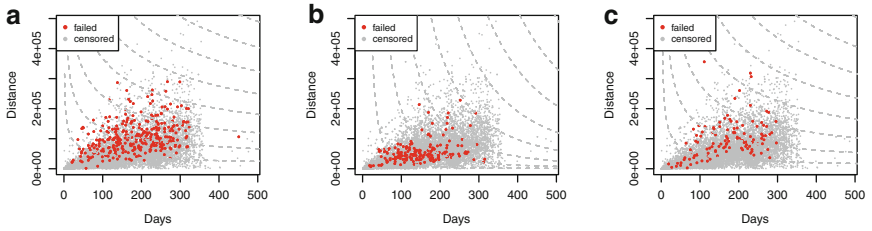


Fig. 4 Scatterplots and estimated time scales of failure and censored data. (**a**) Failure mode A, $\hat{\beta} = 0.695$. (**b**) Failure mode B, $\hat{\beta} = 0.332$. (**c**) Failure mode C, $\hat{\beta} = 0.618$

4 Simulation Study on the Properties of the Time Scale Estimator

4.1 Lifetime Scale Estimations Under Random Censoring or Competing Risks Cases

As done in Duchesne and Lawless (2002), we conducted a series of numerical experiments to investigate the properties of the estimator of the parameters of lifetime scales. We considered a two-dimensional case; one dimension was chronological time X and the other was usage amount Y. Five thousand pairs of random variables T and θ were generated numerically under four simulation settings, as described below.

A. $T \sim$ *Weibull* with shape parameter $m = 3$ and scale parameter $\eta = 1{,}000$, $\tan^{-1}\theta \sim$ *Uniform* on $[0, 1]$. (Abbreviated as `Weib-Uni`)
B. $T \sim$ *Weibull* with shape parameter $m = 3$ and scale parameter $\eta = 1{,}000$, $\theta \sim$ *Lognormal* with location parameter $\mu = 2.37$ and scale parameter $\sigma = 0.57^2$. (Abbreviated as `Weib-LogN`)
C. $T \sim$ *Lognormal* with location parameter $\mu = 6.7$ and scale parameter $\sigma = 0.83^2$, $\tan^{-1}\theta \sim$ *Uniform* on $[0, 1]$. (Abbreviated as `LogN-Uni`)
D. $T \sim$ *Lognormal* with location parameter $\mu = 6.7$ and scale parameter $\sigma = 0.83^2$, $\theta \sim$ *Lognormal* with location $\mu = 2.37$ and scale parameter $\sigma = 0.57^2$. (Abbreviated as `LogN-LogN`)

These parameter settings are chosen to be similar to those in Duchesne and Lawless (2002).

The chronological time X and the usage amount Y for each item were obtained for a linear time scale,

$$X_i = \frac{T_i}{1 - \beta + \beta\theta_i} \quad , \quad Y_i = \frac{\theta_i T_i}{1 - \beta + \beta\theta_i},$$

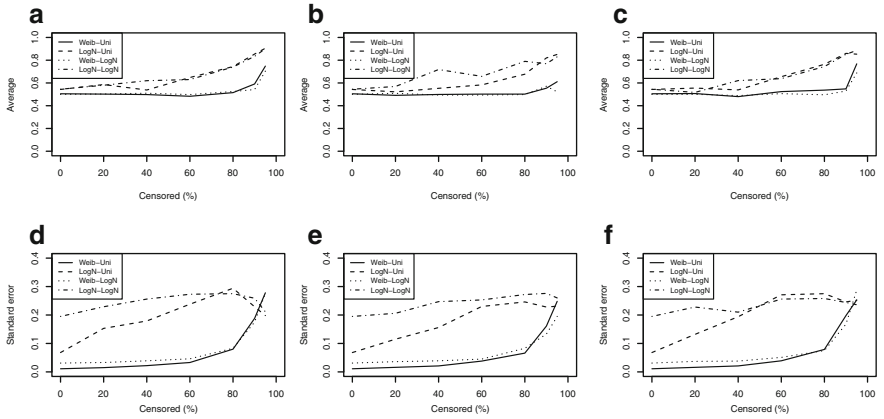


Fig. 5 Results of simulation study for linear time-scale setting with $\beta = 0.5$. (**a**) Type I cens. (Averages). (**b**) Type II cens. (Averages). (**c**) Random cens. (Averages). (**d**) Type I cens. (Std. dev.). (**e**) Type II cens. (Std. dev.). (**f**) Random cens. (Std. dev.)

and for a multiplicative time scale,

$$X_i = \frac{T_i}{\theta_i^{\beta}} \quad , \quad Y_i = \frac{T_i}{\theta_i^{\beta-1}}.$$

Type I censoring, Type II censoring, and random censoring were used for each simulated data set of 5,000 records. Though the proportions of failed items were fixed only for Type II censoring, the censoring times and numbers were set to the average so that all settings could be compared. In all the studies described here, we set $\beta = 0.5$.

Figure 5 shows the results for a linear time scale, $t = (1-\beta)X + \beta Y$. The averages of the estimates $\hat{\beta}$ were biased even when the percent censored was less than 50 for settings B and D, as shown in Fig. 5a, c, e.

The standard deviations of the estimates $\hat{\beta}$ increased with an increase in the proportion of censored items for those settings, as shown in Fig. 5a, c, e. For settings A and C, the estimator was relatively unbiased for cases in which less than 90 % of the items were censored. However, the standard deviations tended to increase rapidly when more than 50 % of the items were censored. That is, the linear time scale tended to depend on the assumptions about the underlying distributions and tended to be sensitive to censoring.

Figure 6 shows the results for a multiplicative time scale, $t = X^{1-\beta}Y^{\beta}$. This time scale was scale equivariant, and the meaning of parameter β did not depend on the scales of X and Y. The averages of the estimates $\hat{\beta}$ were less biased than those for the linear time scale for all settings, as shown in Fig. 6a, c, e.

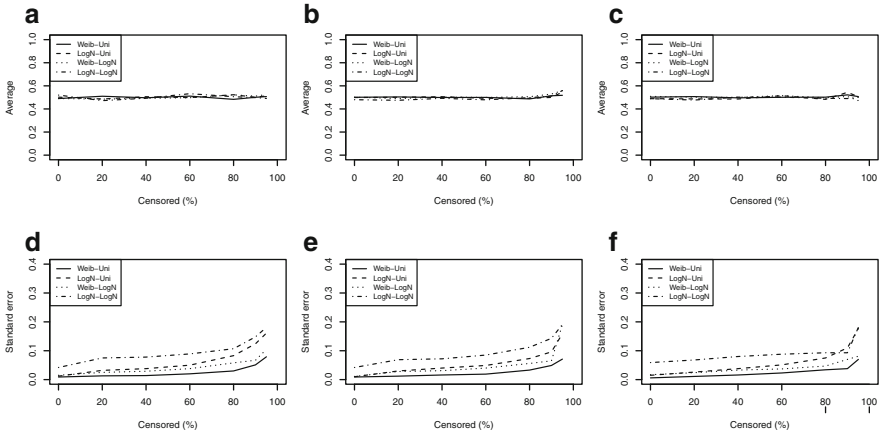


Fig. 6 Results of simulation study for multiplicative time-scale setting with $\beta = 0.5$. (**a**) Type I cens. (Averages). (**b**) Type II cens. (Averages). (**c**) Random cens. (Averages). (**d**) Type I cens. (Std. dev.). (**e**) Type II cens. (Std. dev.). (**f**) Random cens. (Std. dev.)

The standard deviations of the estimates $\hat{\beta}$ increased more slowly than those for the linear time scale with an increase in the proportion of censored items for those simulations, as shown in Fig. 6b, d, f. The parameter estimates for the multiplicative time scale tended to have a sampling distribution insensitive to the assumptions about the underlying distributions and had a certain level of precision even for the data sets with a relatively high proportion of censored items.

Recall that as mentioned in Sect. 2.2, a semiparametric estimator does not depend on the distributional assumption of the time scale, so the multiple time scale is preferable in that it is less sensitive to the distributions of T and θ.

4.2 Lifetime Scale Estimations Under Misspecified Settings

We are also interested in the performance of the time-scale estimator under misspecified model situations. We investigated three possible situations: (1) time-scale function is correct but scale is misspecified, (2) time-scale function is correct but time variables are misspecified, and (3) time-scale function is misspecified. We generated the simulation data using the "generated" time scale and fitted the data using the "fitted" time scale.

In the first situation, the time scale after some monotone transformation was true, but we do not have any technical knowledge about that transformation. Figure 7 shows the results for cases in which logarithmic and exponential transformations were combined with a linear time-scale function. The true time-scale parameter was

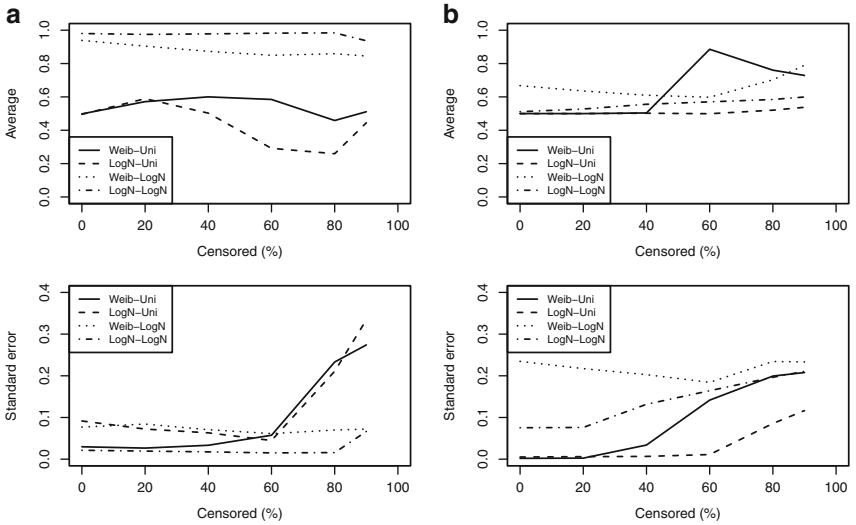


Fig. 7 Results of simulated time-scale transformation with linear time-scale function ($\beta = 0.5$). (**a**) generated: linear TS; fitted: log-linear TS. (**b**) generated: log-linear TS; fitted: linear TS

0.5 for all cases described in this section. We call a time scale that is linear after logarithmic transformation,

$$t(s;\beta) = \exp\left(\sum_{j=1}^{k} \beta_j X_j(s)\right) \quad \text{with} \quad \sum_{j=1}^{k} \beta_j = 1 \quad \text{and} \quad \beta_j \geq 0 \quad \text{for all } j.$$

a log-linear time scale. It is mathematically equivalent to a linear time scale. We see in Fig. 7 that the estimation procedure described in Sect. 2.2 worked not much worse when $\tan^{-1}\theta \sim$ *Uniform* on $[0, 1]$ (as shown by the solid lines) and censoring was not so heavy. However, it created some bias for cases with $\theta \sim$ *Lognormal*, as shown by the dashed lines. The dashed lines for standard deviations in Fig. 7a, b are rather flat for various censoring ratios, meaning that the linear time-scale function is not so flexible under exponential and logarithmic transformation.

Figure 8 presents the results of cases with time-scale misspecification, Fig. 8a,b, and with variable transformation misspecification, Fig. 8c, d.

The averages and standard deviations of the estimated time-scale parameter β tell us something about the suitability or matching among the time-scale function and the underlying model. Figure 8a shows that the log-linear time-scale function gives a $\hat{\beta}$ near 0.5 even though the underlying function is linear, for cases with $\tan^{-1}\theta \sim$ *Uniform* on $[0, 1]$. Further the standard deviation increases as the censoring becomes heavier. The log-linear time-scale function pretends as if it is the true time-scale function. We call this property by *working times-scale function*.

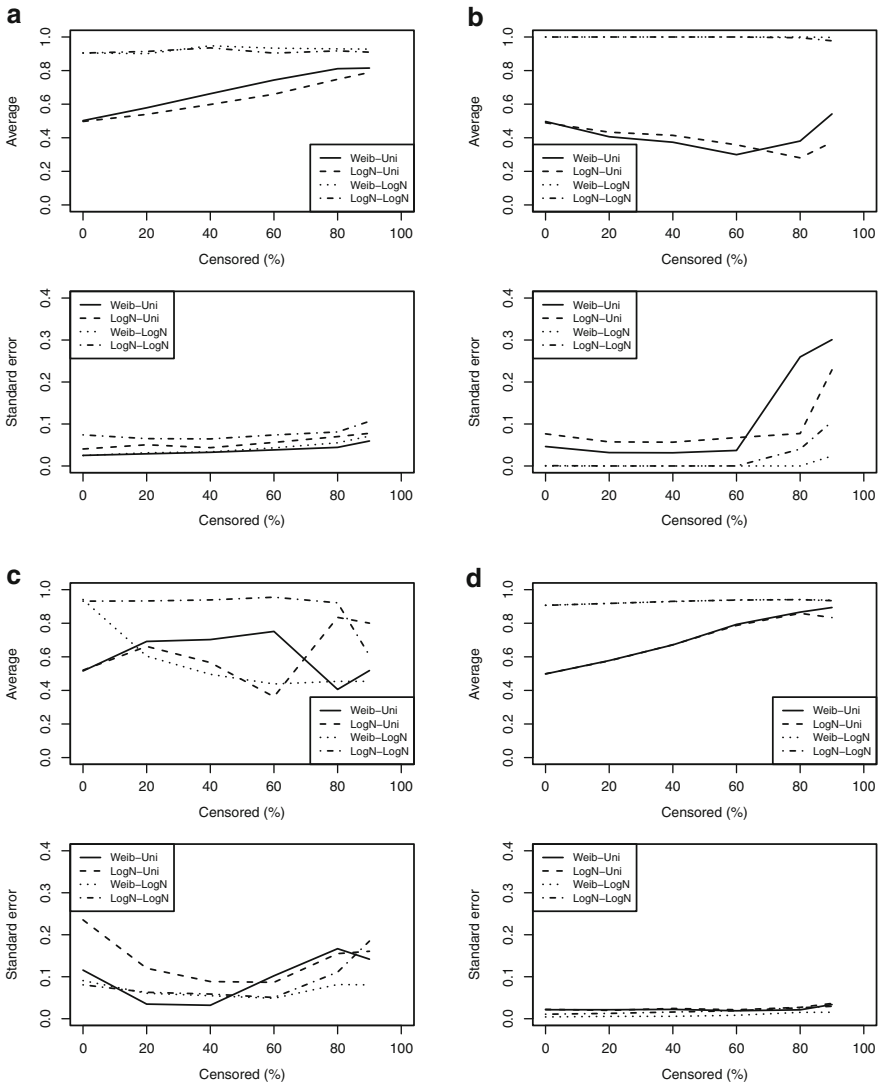


Fig. 8 Results of simulated time-scale transformation under linear time-scale setting ($\beta = 0.5$). (**a**) generated: linear TS; fitted: multiplicative TS. (**b**) generated: multiplicative TS; fitted: linear TS. (**c**) generated: multiplicative TS; fitted: log-linear TS. (**d**) generated: log-linear TS; fitted: multiplicative TS

However this time-scale function gives $\hat{\beta}$ near 1.0 for cases with $\theta \sim$ *Lognormal*, and gives small standard deviations for same cases. It looks being sticked to those values and having no effects from censoring, such as some monotone tendency in the average or increasing in the standard deviation, of the parameter estimates. So

we see that log-linear time-scale function is not suitable for these two cases, even as a working time-scale function.

Figure 8b suggests that the linear time-scale function may work as the working time-scale, in that we may not be able to identify which is the true time-scale, linear or log-linear. It gives steady averages and increasing standard deviations for cases with $T \sim$ *Lognormal*. However it works not fine for the case with $T \sim$ *Weibull* and $\theta \sim$ *Lognormal*.

The difference between the multiplicative and exponential linear time-scale functions (Fig. 8c, d) was the scale for X and Y since

$$t(s;\beta) = \prod_{j=1}^{k} X_j(s)^{\beta_j} = \exp\left(\sum_{j=1}^{k} \beta_j \log X_j(s)\right).$$

If we take the logarithm of all variables and use the log-linear time-scale function, the resulting analysis is the same as that with the multiplicative time scale. Both time-scale functions worked strangely under $\theta \sim$ *Lognormal*, as shown by the dashed lines. But both seem to have worked as a working time scale, again under $\tan^{-1}\theta \sim$ *Uniform* on $[0, 1]$, as shown by the solid lines.

We conclude that the time-scale function estimator can perform well as a working time-scale function depending on the distribution of θ. The invariance under the transformation of $t(s;\beta)$ depends on the distribution of θ.

5 Conclusion

Time-scale analysis is useful because it can be used to estimate the contour plot of a cumulative distribution function using many fewer parameters than with the stratified approach. It is also more flexible because identification of the lifetime distribution can be done after the time-scale function is estimated. Our analyses of two examples demonstrated the advantages of this approach. Simulation studies of the effects of this approach under variable transformations and time-scale misspecifications showed that the multiplicative time scale is the first candidate for application.

The scatterplots were those for failed and censored data. Though the estimated time scale covered the area with no data, we do not claim that the multiplicative time scale can be used for extrapolation. Our extension of the estimator proposed by Duchesne and Lawless (2002) to the competing risks scenario requires further investigation of the sampling properties, and our future research will include such investigation.

Acknowledgements The authors would like to thank Prof. Kazuyuki Suzuki of University of Electro-Communications, for his encouragement and his insightful advices. They are also indebted to Dr. Lu Jin and Mr. Tsutomu Ishida for their valuable comments and fruitful discussions,

Mr. Daisuke Tanada for his calculations on the analysis of Case A. They are also grateful to two anonymous companies for their corporations. This work is partly supported by a Grant-in-Aid for Scientific Research (C) No. 21500273 from the Japanese Society for the Promotion of Science.

References

Duchesne, T. (2000). *Regression Models for Lifetime Given the Usage History*, Abstracts of MMR 2004, Santa Fe, USA.

Duchesne, T., & Lawless, J. F. (2000). Alternative time scales and failure time models. *Lifetime Data Analysis, 6*, 157–179.

Duchesne, T., & Lawless, J. F. (2002). Semiprametric inference method for general time scale models. *Lifetime Data Analysis, 8*, 263–276.

Farewell, V. T., & Cox, D. R. (1975). A note on multiple time scales in life testing. *Applied Statistics, 28*, 115–124.

Finkelstein, M. S. (1999). Wearing-out of components in a variable environment. *Reliability Engineering and System Safety, 66*, 235–242.

Finkelstein, M. S. (2004). Minimum repair in heterogeneous populations. *Journal of Applied Probability, 41*, 281–286.

Kordonsky, K. B., & Gertsbakh, I. (1993). Choice of the best time scale for system reliability analysis. *European Journal of Operational Research, 65*, 235–246.

Kordonsky, K. B., & Gertsbakh, I. (1995a). System state monitoring and lifetime scales - I. *Reliability Engineering and System Safety, 47*, 1–14.

Kordonsky, K. B., & Gertsbakh, I. (1995b). System state monitoring and lifetime scales - II. *Reliability Engineering and System Safety, 49*, 145–154.

Kordonsky, K. B., & Gertsbakh, I. (1997). Multiple time scales and the lifetime coefficient of variation: Engineering applications. *Lifetime Data Analysis, 2*, 139–156.

Lawless, J. F., Crowder, M. J., & Lee, K.-A. (2009). Analysis of reliability and warranty claims in products with age and usage scales. *Technometrics, 51*, 14–24.

Oakes, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Analysis, 1*, 7–18.

Why the Naive Bayesian Classifier for Clinical Diagnostics or Monitoring Can Dominate the Proper One Even for Massive Data Sets

Hans - J. Lenz

Abstract We explain the phenomenon that the naive Bayesian classifier may dominate the proper one as happened in clinical studies, cf. Gammerman and Thatcher (*Methods of Information in Medicine*, *30*, 15–22, 1991). Today this effect may be of concern for real-time health care monitoring or surveillance. The reason for the dominance relation lies in a mix of an a-priori not fixed dimension of the state-space (symptom space) given a disease, the feature selection procedure and the parameter estimation. Estimating conditional probabilities in high dimensions when using a proper Bayesian model can lead to an "over fitting," a missing value problem, and, consequently, to a loss of classification accuracy. Due to the "Curse of dimension" the degradation may not even be compensated by big data sets.

Keywords Curse of dimension • Dependency structures • Over-fitting

1 Introduction

Consider a clinical database as a collection of big tables each of them consisting of a set of records. For example, table **P** of patients stores data like patient's name, age, gender, address, insurance company, etc., table **C** of cases relates patients, doctors, treatments, illnesses, and symptoms, cf. Fig. 1. Furthermore, the reference table (International Catalog of Diseases) **ICD** relates the set of diseases $\boldsymbol{\Theta}$ to the set of symptoms $\boldsymbol{X}$ in a many-to-many way. The last relation represents current medical knowledge as, for instance, coded by ICD-10. We assume that the tables of symptoms, diseases, and the reference table are finite, which is a reasonable assumption in this context and simplifies stochastic modeling.

Consider a new patient (case) admitted to the hospital and his recorded symptom vector $\boldsymbol{x} \in \boldsymbol{X}^d$ of fixed length d stored during anamnesis and updated while applying treatments to him. The objective of medical diagnostics during any phase of the patient's stay at a hospital or while monitoring him is to use a classifier

H.-J. Lenz (✉)
Institut für Statistik und Ökonometrie, Freie Universität Berlin, 14195 Berlin, Germany
e-mail: hans-j.lenz@fu-berlin.de

S. Knoth, W. Schmid (eds.), *Frontiers in Statistical Quality Control 11*,
Frontiers in Statistical Quality Control, DOI 10.1007/978-3-319-12355-4_23

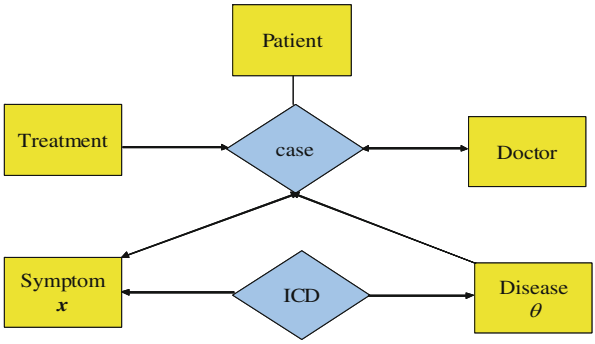


Fig. 1 Conceptual clinical database model clipping

(diagnostic function): $\phi : \boldsymbol{X}^d \rightarrow \boldsymbol{\Theta} \times \boldsymbol{R}_{[0,1]}$, $(\theta, \pi) = \phi(\boldsymbol{x})$ where $\phi \in \Psi$ identifies the disease of the current case with a degree of belief $\pi = P(\theta \mid \boldsymbol{x})$, and Ψ is the class of classifiers considered. In the study of Gammerman and Thatcher (1991) the problem is reduced to check the classification precision of computer and physicians' preliminary diagnoses during anamnesis.

As there are different classifiers ϕ available we limit ourselves to the maximum probability rule of the Bayesian framework

$$\phi(\boldsymbol{x}) = \theta^* \text{ with } \pi = P(\theta^* \mid \boldsymbol{x}) \text{ where } \theta^* = \underset{\theta \in \Theta}{\operatorname{argmax}}\, P(\theta \mid \boldsymbol{x}) \tag{1}$$

The vital question from a diagnostic point of view is how to estimate $P(\theta \mid \boldsymbol{x})$ directly or—using the Bayes theorem—indirectly by first computing the marginal $P(\theta)$ and conditional probabilities $P(\boldsymbol{x} \mid \theta)$ for each "reasonable" $\theta \in \Theta$ using the clinical database. The Bayesian approach offers two variants. The ***naïve*** or ***Idiot Bayes*** approach assumes conditional independence of the symptoms given a specific disease, which is counterintuitive and seems quite unrealistic. The advantage of this approach is that it factorizes $P(\boldsymbol{x} \mid \theta)$ by the product of the marginals $P(x_1 \mid \theta)$, $P(x_2 \mid \theta)$, …, $P(x_d \mid \theta)$. Thus the estimation problem is reduced from one d-dimensional problem to d one-dimensional ones. Accordingly, the marginals are estimated by the corresponding relative frequencies $n(x_i, \theta)/n(\theta)$ in the one-dimensional symptom subspace related to θ. Quite opposite, the proper ***Bayes*** approach assumes a conditional dependency structure between symptoms given a specific disease. The unknown conditional probabilities $P(\boldsymbol{x} \mid \theta)$ for each θ are estimated by the joint frequencies $n(x_1, x_2, \ldots, x_d, \theta)/n(\theta)$ for each θ. Note that in the Bayesian framework we usually assume that the symptom space has a fixed dimension $d \in \boldsymbol{N}$. However, in the health-care domain d varies from disease to disease, i.e. $d_\theta \leq d$, and even from case to case. So d is an upper bound. Moreover, due to computational complexity model-hunting cannot use full enumeration of all symptoms combinations, but must be based on search heuristics.

The empirical studies of Gammerman and Thatcher (1991) and Schwartz et al. (1993) surprisingly show that the proper ***Bayes*** approach leads to an inferior diagnostic rule compared to the ***naïve Bayesian*** one, i.e. $\phi_{nB} \prec \phi_B$. In the following we analyze this phenomenon cf. Lenz (1995), and give an explanation why a proper approach can produce worse results, how this happened and what should be done to avoid such artifacts. We put a stress on the fact that this effect is not limited to the Bayesian classifier but is—mutatis mutandis—also true, for example, for the perceptron model of the neural network classification approach, cf. Khanna (1990).

2 Some Well-Known Definitions and Theorems

The probability theory is a well-established theory which is capable of mapping equally well population variability, measurement errors as well as sampling and estimation errors. We shall follow the axiomatic approach originated by Kolmogorov (1933).

First, we recall some few well-known definitions about marginal and conditional probabilities. Let $(\Omega, 2^\Omega, P)$ be the underlying probability space with the finite observation set Ω, power set (sigma algebra) 2^Ω, and probability function P. Later Ω will represent the set of $0-1$ vectors of maximum length $d = 135$.

Definition 1 Conditional Probability $P(A \mid B)$

$$\text{Let } A, B \in 2^\Omega \text{ and } P(B) > 0. \text{ Then } P(A \mid B) = P(A, B)/P(B)\,. \tag{2}$$

The conditional independence of events $(A_1, A_2, \ldots, A_n)$ given an event B is defined as follows.

Definition 2 Conditional Independence

Let $\left(A_i\right)_{i \in I} \in 2^\Omega$ be a finite sequence of events and $P(B) > 0$. Then the events $(A_1, A_2, \ldots, A_n)$ are called conditional stochastically independent given $B \in 2^\Omega$ if

$$P(A_1, A_2, \ldots, A_n \mid B) = \prod_{i \in I} P(A_i \mid B). \tag{3}$$

The famous Bayes theorem combines marginal and conditional probabilities as follows, cf. Fig. 2.

Theorem 1 *Bayes Theorem*

Given $P(\boldsymbol{x} \mid \theta)$ *and* $P(\theta) > 0$ *for* $\theta \in \boldsymbol{\Theta}$ *it follows*

$$P(\theta \mid \boldsymbol{x}) = c\,P(\boldsymbol{x} \mid \theta)P(\theta) \tag{4}$$

where $c \in \boldsymbol{R_+}$ *is a normalizing constant.*

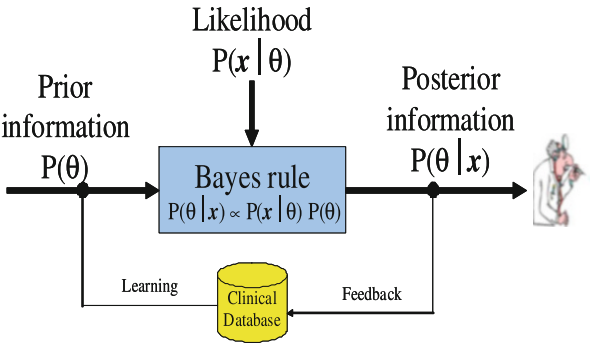


Fig. 2 Bayesian learning

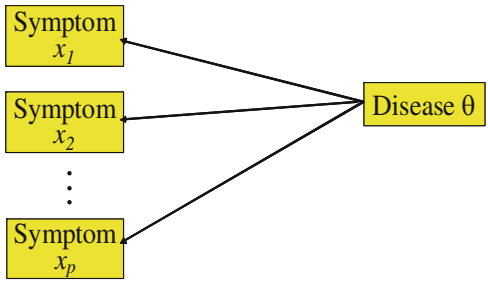


Fig. 3 Conditional independence of symptoms given disease θ

Remember that a prerequisite of applying the ***Bayes*** theorem to health care is the knowledge of the multi-dimensional distribution $P(\boldsymbol{x} \mid \theta)$, and of the marginal or prior distribution $P(\theta)$ for each disease θ.

The naïve or ***Idiot Bayes*** theorem—a term coined by Spiegelhalter (1986)—is a special case of the ***Bayes*** theorem assuming conditional independence of the d symptoms $X_1, X_2, \ldots, X_d$ given a specific disease θ, cf. Fig. 3. Note that d is disease or even case dependent. The strong assumption of conditional independence greatly simplifies determining the probabilities $P(\theta \mid \boldsymbol{x})$ for all $\theta \in \Theta$ by experts during the elicitation phase, and, of course, influences the precision of the estimates.

Theorem 2 *Idiot-Bayes Theorem*

Assuming conditional independence between the components of symptom vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ *given* θ *one can factorize the joint distribution* $P(\boldsymbol{x} \mid \theta)$ *as follows:*

$$P(\theta \mid \boldsymbol{x}) = c \prod_{i=1}^{d} P(x_i \mid \theta) P(\theta) \tag{5}$$

where $c \in \boldsymbol{R}_+$ *is a normalizing constant.*

3 Estimation of Marginal and Conditional Probabilities Using a Clinical Database System

Given a clinical database the estimation of the marginal and/or conditional probabilities can be achieved by querying the database.

3.1 Strategy 1: Marginal Probability Estimation

First, we consider the marginal (or prior) probability $P(\theta)$ for each $\theta \in \boldsymbol{\Theta}$ and estimate the probability by $\hat{P}(\theta) = n(\theta)/N$ where $n(\theta)$ counts the number of cases where patients are classified to have disease θ and N is total number of patients admitted to the hospital having an illness $\theta \in \boldsymbol{\Theta}$. Remember that $| \boldsymbol{\Theta} | < \infty$.

This simple indicator is not considered further as it neglects symptoms, and only plays a role as a prior distribution when initializing the Bayesian classifier.

3.2 Strategy 2: Direct Estimation

In principle, $P(\theta \mid \boldsymbol{x})$ for all illnesses $\theta \in \boldsymbol{\Theta}$ can be directly estimated by (at least) two table look-ups in the clinical database. First, the whole database is scanned for selecting the subset $\boldsymbol{C}_x \subseteq \boldsymbol{C}$ where the (*complete*) symptom vector $\boldsymbol{x} \in \boldsymbol{X}^d$ is the selection criterion and the counts are recorded. Let the cardinality of this set be $n(\boldsymbol{x})$. Next the smaller table $\boldsymbol{C}_x$ is scanned and the number of cases $n(\boldsymbol{x}, \theta)$ showing disease θ together with symptom vector $\boldsymbol{x}$ is counted. As an ML-estimate we get

$$\hat{P}(\theta \mid \boldsymbol{x}) = n(\boldsymbol{x}, \theta)/n(\boldsymbol{x}). \tag{6}$$

Therefore no Bayesian inference step is necessary at this point, as the conditional distribution $P(\theta \mid \boldsymbol{x})$ for each $\theta \in \Theta$ can be estimated directly.

However, the cases vary in so far as the number of observed or recorded symptoms per case varies from case to case even for the same disease θ, i.e. not all single symptoms may be observed or recorded as elements of the observation vector $\boldsymbol{x} \in \boldsymbol{X}^d$ for each patient. Evidently, d is disease and case dependent. This leads to a missing value problem. If the missing values were completely generated at random (so-called MCAR assumption), then a complete-set analysis is appropriate. However, dropping incomplete cases as suggested by Gammerman and Thatcher (1991) reduces the sample size and leads to loss of estimation accuracy.

3.3 Strategy 3: Proper Bayesian Approach

Here the definition of the (Bayesian) posterior distribution—likelihood × prior—$P(\theta \mid \boldsymbol{x}) = cP(\boldsymbol{x} \mid \theta)P(\theta)$ holds true for each $\theta \in \boldsymbol{\Theta}$ considered. Ignoring computational efficiency for a moment one can compute the estimates $\hat{P}(\theta \mid \boldsymbol{x}) = n(\boldsymbol{x}, \theta)/n(\theta)$ and $\hat{P}(\theta) = n(\theta)/N$ where $n(\theta)$ is an absolute frequency and counts the cases with disease θ, and N represents the total number of cases, i.e. $N = \mid \boldsymbol{C} \mid$. Note that not every patient shows the same symptom vector. Therefore the main problem is to select the "best" combination of d symptoms, i.e. to solve heuristically the feature selection problem.

3.4 Strategy 4: Idiot Bayes Approach

The posterior probability $P(\boldsymbol{x} \mid \theta)$ is computed as mentioned above by factorization of $P(\boldsymbol{x}, \theta) = \prod P(x_i \mid \theta)$, cf. Fig. 3. The conditional probability of each single symptom, $P(x_i \mid \theta)$, for $i = 1, 2, \ldots, d$ is estimated by

$$\hat{P}(x_i \mid \theta) = \frac{n(x_i, \theta)/N}{n(\theta)/N} \tag{7}$$

Evidently, the problem of missing values is reduced because each symptom is treated and counted separately for estimation. Given a patient suffering from disease θ the probability to be positive with one out of d symptoms is much higher than simultaneously with the bulk of all $d \gg 1$ symptoms. Or simplifying, when firing into space it is easier to hit a given rectangle in two dimensions than a hypercube in high-dimensions. For illustration of this "curse of dimension" assume independent symptoms given a disease, and a constant probability $\kappa = 5\,\%$ for a missing value (so-called "missing rate") for any of the $d \in \boldsymbol{N}$ symptoms. Then the probability $P(\textbf{at least one missing value}) = 1-(1-\kappa)^d = 0.999 \approx 1$ for $d = 135$, see Fig. 4.

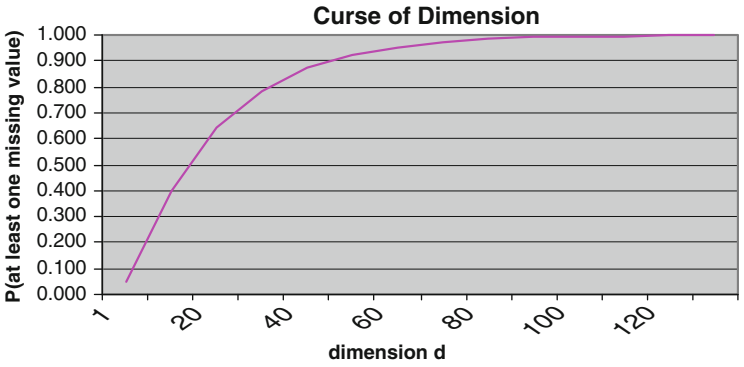


Fig. 4 Probability of at least one missing value

This effect leads to a larger sample size of complete records in lower dimensions, i.e. in a proper subspace of the symptom space $\boldsymbol{X}$.

4 The Gammerman and Thatcher Study

The empirical comparative study by Gammerman and Thatcher (1991) on the Bangour Hospital, Roxburgh, UK, is worthwhile reading, and is still representative for current machine diagnoses and health-care monitoring. It has a learning size of $n_L = 4{,}387$ cases, and left $n_T = 2{,}000$ for testing. Altogether, there are $\mid \Theta \mid = 9$ diseases (or so-called Diagnostic Groups) used for preliminary and final diagnoses like appendicitis, pancreatitis, renal colic and dyspepsia, cf. Gammerman and Thatcher (1991). If a patient cannot be classified to one of the "real" eight possible diseases, he is allocated to the residual diagnostic group labeled "non-specific abdominal pain." There are 33 main symptoms like *Pain-site*, *Anorexia*, etc. considered (some being present or absent, others multi-valued and not only binary), and expanded to all together $\mid \boldsymbol{X} \mid = 135$ binary symptoms under study. Therefore the range of the symptom space is $\boldsymbol{X} = \{0, 1\}^d$.

We condense the study to the main summary statistic and refer the reader who is interested in the details of the study to Gammerman and Thatcher (1991). It includes three diagnosis methods I-III—made by a physician, the Idiot and the proper Bayes approach, and compares them by the overall non-error rate (%) of all preliminary diagnoses related to the final (actual) ones taking the physician's decision as a yardstick, cf. Table 1. By the way, Gammerman and Thatcher (1991) present a fourth one (CART) which we ignore in our context; it performs like method III (Proper Bayesian approach) with an accuracy of 65 %.

Evidently, the Idiot Bayes dominates the proper Bayes approach, and is near to the physicians' classification. How could this happen? The study of Schwartz et al. (1993) supports this phenomenon, and the ranking is even the same for big data studies. So far the study is "representative" and by no means an artifact.

Table 1 Percentages of correct preliminary diagnoses in a comparative study

Method	Percentage of correct diagnoses
I Physician	76
II Idiot bayes	74
III Proper bayes	65 (!)

Source:Gammerman and Thatcher (1991)

5 Modeling and Estimation Accuracy

We come back to our question put forward above. The contradiction which is evident from Table 1 can be explained as a mix of factors as follows.

1. *The feature selection problem*: The probability model relating a disease θ and a corresponding symptom vector $\boldsymbol{x}$ is by no means trivial to be found-out. Even the proper dimension of the symptom space $(\boldsymbol{X}_\theta)$ given illness θ is unknown. Gammerman and Thatcher estimated these dimensional parameters and the corresponding symptoms combination—necessarily only for the proper Bayesian approach—by a forward selection procedure based on a χ^2-statistic for 2×2 tables. It includes a step-by-step procedure for up to 135 symptoms x_i and each single disease θ. For instance, during the initial phase the value of the χ^2-statistic for 2×2 table $(\theta_1, \bar{\theta}_1, x_i, \bar{x_i})$ with $i = 1, 2, \ldots, 135$ was computed. When the symptom x_i^* having the highest discrimination power measured by the χ^2-statistic is selected, the next best double symptoms combination (x_i^*, x_j) for all $j = 1, 2, \ldots, 135, i \neq j$ is searched for in a similar way. Unfortunately, it is unclear how the authors achieve the nominal value of $\alpha = 1\,\%$. During the model-hunting phase they consider a learning set of size $n_L = 4{,}387$, and a test size equal to $n_T = 2{,}000$. The best combinations lead to an overall hit-rate of diagnoses equal to 65 %. Instead of simply fixing a boundary of the dimension of $\boldsymbol{X}_\theta$ a more sophisticated approach could be to use multiple imputation, cf. Barnard and Rubin (1999). The reader interested in details of fitting a general linear model (GLM) should consult, for example, Fahrmeir et al. (1996).
2. *The over-fitting effect*: Too many symptoms per disease may be selected due to ignoring multi-collinearity effects caused by the forward selection procedure. Furthermore, the missing values effect according to "incomplete records" leads to a reduced sample size (number of cases with a d_θ-complete symptom vector per disease). Consequently, the sampling error in weakly occupied cells of a high-dimensional data cube (contingency table) is increased, i.e.

$$Var\big(\hat{P}(\boldsymbol{x}, \theta)\big) \sim 1/m_{\boldsymbol{x},\theta}. \tag{8}$$

   where $m_{\boldsymbol{x},\theta}$ is the current absolute frequency in cell $(\boldsymbol{x}, \theta)$.
3. *The effect of weakly occupied cells*: In high dimensional spaces $(\boldsymbol{X}, \Theta)$ the risk of incomplete records is seemingly reduced if the *Idiot Bayes* approach is used. From an algebraic point of view we switch to the subspace $(\boldsymbol{X}_{\boldsymbol{\theta}}, \{\theta\})$. The conditional independence assumption increases the number $n(x_i, \theta)$ of counts in each single cell (x_i, θ) for all diseases $\theta \in \Theta$ and observed symptom x_i being an element of $\boldsymbol{x} \in \boldsymbol{X}^d$. In other words, relaxing the dependency structure activates the "The law of large numbers."

Generally speaking, before applying a Bayesian diagnostic rule for clinical diagnostics or (on-line) health-care monitoring or any similar approach one should find out the trade-off between structural dependency, managing missing values (ignoring incomplete records or applying multiple imputation) and estimation efficiency. In high dimensional data spaces this effect does not vanish even with big data.

References

Barnard, J., & Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika*, *86*, 948–955.

Fahrmeir, L, Hamerle, A., & Tutz, G. (1996). *Multivariate statistische Verfahren.* Berlin: De Gruyter.

Gammerman, A., & Thatcher, A. R. (1991). Bayesian Diagnostic probabilities without assuming independence of symptoms. *Methods of Information in Medicine*, *30*, 15–22.

Khanna, T. (1990). *Foundations of neural networks.* New York: Addison-Wesley, Reading etc.

Kolmogorov, A. N. (1933). *Grundbegriffe der wahrscheinlichkeitsrechnung.* Berlin: Springer

Lenz, H.-J. (1995). On the idiot vs. proper bayes approach in clinical diagnostic systems. In A. Gammerman (Ed.), *Probabilistic Reasoning and Bayesian Belief Networks* (pp. 227–236). Alfred Waller in association with UNICOM, Henley-on-Thames.

Schwartz, S., Wiles, J., Gough, I., & Phillips, S. (1993). Connectionist, rule-based and Bayesian decision aids: an empirical comparison. In D. J. Hand (Ed.), *Artificial intelligence frontiers in statistics* (pp. 264–278). London: Chapman & Hall.

Spiegelhalter, D. J. (1986). A Statistical view of uncertainty in expert systems. In W. Gale (Ed.), *Artificial intelligence and statistics* (pp. 17–55). New York: Addison-Wesley, Reading.