ICSA Book Series in Statistics
Series Editors: Jiahua Chen · Ding-Geng (Din) Chen

Yuhlong Lio
Hon Keung Tony Ng
Tzong-Ru Tsai
Ding-Geng Chen *Editors*

Statistical Quality Technologies

Theory and Practice

EXTRAS ONLINE

Springer

ICSA Book Series in Statistics

The ICSA Book Series in Statistics showcases research from the International Chinese Statistical Association that has an international reach. It publishes books in statistical theory, applications, and statistical education. All books are associated with the ICSA or are authored by invited contributors. Books may be monographs, edited volumes, textbooks and proceedings.

More information about this series at http://www.springer.com/series/13402

Yuhlong Lio • Hon Keung Tony Ng • Tzong-Ru Tsai
Ding-Geng Chen
Editors

Statistical Quality Technologies

Theory and Practice

Springer

Editors
Yuhlong Lio
Department of Mathematical Sciences
University of South Dakota
Vermillion, SD, USA

Hon Keung Tony Ng
Department of Statistical Science
Southern Methodist University
Dallas, TX, USA

Tzong-Ru Tsai
Department of Statistics
Tamkang University
New Taipei, Taiwan

Ding-Geng Chen
Department of Statistics
University of Pretoria
Pretoria, South Africa

ISSN 2199-0980 ISSN 2199-0999 (electronic)
ICSA Book Series in Statistics
ISBN 978-3-030-20708-3 ISBN 978-3-030-20709-0 (eBook)
https://doi.org/10.1007/978-3-030-20709-0

Preface

Statistical methodologies for product quality control, acceptance sampling plans, and product reliability are essential technologies that ensure product quality to reduce both consumer and producer risks. Numerous novel statistical technologies to improve and to evaluate product quality had been developed by many scholars in the past decades. After we edited the book *Statistical Modeling for Degradation Data* (2017; Springer, Singapore), we have seen a great need to bring together experts engaged in statistical process quality control, acceptance sampling plan, and reliability testing and designs to present and discuss important issues of recent advances in product quality technologies and related applications. For this reason, we edit this book *Statistical Quality Technologies: Theory and Practice* that focuses on statistical aspects of product quality technology development.

In this book, we aim to provide theories as well as applications of statistical techniques for manufacturing quality. This book provides a venue for the timely dissemination of research on the statistical methodologies of quality improvement and assessment and to promote further research and collaborative work in this area. The authors in each chapter have made both the theoretical results and the novel statistical quality technologies publicly available, thus making it possible for readers to readily apply these new methodologies in different areas of applications and research. We believe that the topics covered in the book are timely and have high potential to impact and influence in statistics, engineering, and manufacturing.

Outline of This Book Volume

This book volume brings together 16 chapters that are categorized as follows: Statistical Process Control (Part I), Acceptance Sampling Plans (Part II), and Reliability Testing and Designs (Part III). All the chapters have undergone a thorough review process.

Part I of this book includes six papers focusing on both theoretical and applied research in statistical process control. Chapter 1 provides an overview of some

statistical process control methodologies. Qiu introduces some recent studies on nonparametric statistical process control, control charts for monitoring dynamic processes, and spatio-temporal process monitoring. In Chap. 2, Leiva, Marchant, Ruggeri, and Saulo introduce statistical quality control and reliability tools based on the Birnbaum-Saunders distribution and its generalizations, which are suitable for the situations where the distribution of product quality characteristic is asymmetric. Some possible research related to big data and business intelligence is also discussed. In Chap. 3, Koppel and Chang propose a system-wise process monitoring framework called the statistical system monitoring (SSM) for a production process equipped with thousands of process parameters and hundreds of product characteristics. The properties of the proposed SSM are studied via simulation and practical guidelines are provided. In Chap. 4, Abujiya and Lee present several location and dispersion cumulative sum (CUSUM) control charts based on the ranked set sampling (RSS) techniques. The proposed CUSUM charts are shown to be more effective compared to the standard CUSUM charts based on random sampling. In Chap. 5, Chiang, Ng, Tsai, Lio, and Chen provide a review on statistical process control for simple linear profile with independent or autocorrelated observations. Some recent developments of statistical process control on a simple linear profile model are discussed. In Chap. 6, Potgieter provides a review of some existing CUSUM procedures for monitoring location and concentration changes in circular processes. A new sequential changepoint procedure for detecting the changes in location and/or scale is proposed and the properties and performance are studied.

Part II comprises four chapters that emphasis on the statistical techniques related to acceptance sampling plans. In Chap. 7, Aslam, Rao, and Albassam present a generalized multiple dependent state sampling (GMDSS) plan for a time truncated life test to monitor product quality. Acceptable quality level and limiting quality level are used to determine the plan parameters. In Chap. 8, Prajapati, Mitra, and Kundu develop a decision theoretic sampling plan (DSP), which is an acceptance sampling plan based on Type-I and Type-I hybrid censoring via Bayes' decision theory approach with a suitable loss function. An algorithm for obtaining the optimal DSP is provided. In Chap. 9, Chiang, Ng, Tsai, Lio, and Chen provide a general structure of an economical design of acceptance sampling plan with warranty using truncated life test via Bayesian framework to tackle possible lot-to-lot variation of products. A unified algorithm to reach an optimal sample size and acceptance number for the sampling plan is established to minimize the respective expected total costs. In Chap. 10, Kumar investigates the optimal acceptance sampling plans that minimize the total expected testing cost subject to given upper bounds for the producer and consumer risks based on Type-II censored partially accelerated life test. Numerical results for the linear model and Arrhenius model are provided.

Part III includes six chapters that concentrate on reliability testing and designs. Chapter 11 deals with traditional accelerate life plan based on the c-optimality for minimizing the variance of percentile lifetime. In this chapter, Lu, Lee, and Hong propose a sequential design strategy for life tests based on the dual objectives to resolve the unknown model parameters to improve the accuracy of predicted

lifetime. In Chap. 12, Wang, Jiang, and Wang deal with the stress-strength models for reliability design of systems when both the stress and the strength variables follow the proportional hazards family or the proportional reverse hazards family. Statistical inferential methods based on the proposed model are developed. In Chap. 13, Shen, Shen, and Xu consider a Wiener-based degradation model with logistic distributed measurement errors. Efficient algorithm is provided for the estimation of parameters. In Chap. 14, Pan and Seo present a generalized linear model approach to obtain the optimal accelerated life test planning based on the proportional hazard model. The proposed approach is shown to be flexible for any failure time distribution. In Chap. 15, Ouyang, Park, Byun, and Leeds provide the background behind a dual response surface methodology that incorporates a robust design. They propose different estimation methodologies for remedying the difficulties associated with data contamination and model departure. This section concludes with Chap. 16 that deals reliability modeling with manufacturing processes of modern ultra-large-scale integrated circuits. Bae, Yuan, and Kuo discuss some latest development in modeling the non-homogeneous distributed spatial defect counts.

Vermillion, SD, USA — Yuhlong Lio
Dallas, TX, USA — Hon Keung Tony Ng
New Taipei, Taiwan — Tzong-Ru Tsai
Pretoria, South Africa — Ding-Geng Chen
March 2019

Acknowledgments

We are deeply grateful to those who have supported in the process of creating this book. We thank the contributors to this book for their enthusiastic involvements and their kindness in sharing their professional knowledge and expertise. Our sincere gratitude goes to all the chapter reviewers for their expert reviews of the book chapters, which lead to a substantial improvement in the quality of this book. We thank all the reviewers for providing thoughtful and in-depth evaluations of the papers contained in this book. We gratefully acknowledge the professional support of Nicholas Philipson and Nitza Jones-Sepulveda from Springer who made the publication of this book a reality. We would also thank the support and encouragement from the editors of ICSA Book Series in Statistics, Professors Jiahua Chen and Ding-Geng Chen.

The work of Professor Ding-Geng Chen is based upon research supported by the National Research Foundation, South Africa (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant number 114613). Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

We welcome readers' comments, including notes on typos or other errors, and look forward to receiving suggestions for improvements to future editions. Please send comments and suggestions to any of the editors.

Contents

Contributors

Mu'azu Ramat Abujiya Preparatory Year Mathematics Program, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

Mohammed Albassam Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

Muhammad Aslam Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

Suk Joo Bae Department of Industrial Engineering, Hanyang University, Seoul, South Korea

Jai-Hyun Byun Department of Industrial and Systems Engineering, Gyeongsang National University, Jinju, Gyeongnam, South Korea

Shing I Chang IMSE Department, Kansas State University, Manhattan, KS, USA

Ding-Geng Chen Department of Statistics, University of Pretoria, Pretoria, South Africa

Jyun-You Chiang School of Statistics, Southwestern University of Finance and Economics, Chengdu, China

Yili Hong Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Pei Hua Jiang School of Mathematics and Physics, Anhui Polytechnic University, Wuhu, China

Siim Koppel IMSE Department, Kansas State University, Manhattan, KS, USA

M. Kumar Department of Mathematics, National Institute of Technology Calicut, Kozhikode, Kerala, India

Debasis Kundu Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India

Yue Kuo Department of Chemical Engineering, Texas A&M University, College Station, TX, USA

I-Chen Lee Department of Statistics, National Cheng Kung University, Tainan, Taiwan

Muhammad Hisyam Lee Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Mark Leeds Statematics Consulting, New York, NY, USA

Víctor Leiva School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Yuhlong Lio Department of Mathematical Sciences, University of South Dakota, Vermillion, SD, USA

Lu Lu Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA

Carolina Marchant Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile

Sharmistha Mitra Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India

Hon Keung Tony Ng Department of Statistical Science, Southern Methodist University, Dallas, TX, USA

Linhan Ouyang Department of Management Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China

Rong Pan School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA

Chanseok Park Applied Statistics Laboratory, Department of Industrial Engineering, Pusan National University, Busan, South Korea

Cornelis J. Potgieter Department of Mathematics, Texas Christian University, Forth Worth, TX, USA
Department of Statistics, University of Johannesburg, Johannesburg, South Africa

Deepak Prajapati Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India

Peihua Qiu Department of Biostatistics, University of Florida, Gainesville, FL, USA

Gadde Srinivasa Rao Department of Statistics, The University of Dodoma, Dodoma, Tanzania

Fabrizio Ruggeri CNR-IMATI, Milano, Italy

Helton Saulo Department of Statistics, Universidade de Brasilia, Brasilia, Brazil

Kangwon Seo University of Missouri, Columbia, MO, USA

Li-Juan Shen Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, Singapore

Yan Shen Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian, China

Tzong-Ru Tsai Department of Statistics, Tamkang University, New Taipei City, Taiwan

Bing Xing Wang Department of Statistics, Zhejiang Gongshang University, Hangzhou, China

Xiaofei Wang Department of Statistics, Zhejiang Gongshang University, Hangzhou, China
School of Mathematics and Statistics, Huangshan University, Huangshan, China

Wang-Tu Xu Department of Urban Planning, School of Architecture and Civil Engineering, Xiamen University, Xiamen, Fujian, China

Tao Yuan Department of Industrial and Systems Engineering, Ohio University, Athens, OH, USA

About the Editors

Yuhlong Lio is a professor in the Department of Mathematical Sciences at the University of South Dakota, Vermillion, SD, USA. He is an associate editor of professional journals, including the *Journal of Statistical Computation and Simulation.* His research interests include reliability, quality control, censoring methodology, kernel smoothing estimate, and accelerated degradation data modeling. Dr. Lio has written more than 70 refereed publications.

Hon Keung Tony Ng is currently a professor of statistical science with the Southern Methodist University, Dallas, TX, USA. He is an associate editor of *Communications in Statistics*, *Computational Statistics*, *IEEE Transactions on Reliability, Journal of Statistical Computation and Simulation*, *Naval Research Logistics, Sequential Analysis* and *Statistics and Probability Letters*. His research interests include reliability, censoring methodology, ordered data analysis, nonparametric methods, and statistical inference. He has published more than 100 research papers in refereed journals. He is the co-author of the book *Precedence-Type Tests and Applications* and co-editor of *Ordered Data Analysis, Modeling and Health Research Methods*. Professor Ng is a fellow of the American Statistical Association, an elected senior member of IEEE, and an elected member of the International Statistical Institute.

Tzong-Ru Tsai is a professor in the Department of Statistics at Tamkang University in New Taipei City, Taiwan. His main research interests include quality control and reliability analysis. He has served as a consultant with extensive expertise in statistical quality control and experimental design for many companies in the past years. He is an associate editor of the *Journal of Statistical Computation and Simulation*. Dr. Tsai has written more than 70 refereed publications.

Ding-Geng Chen is a fellow of the American Statistical Association and currently the Wallace H. Kuralt Distinguished Professor at the University of North Carolina at Chapel Hill, USA, and an Extraordinary Professor at the University of Pretoria, South Africa. He was a Professor at the University of Rochester and the Karl E. Peace endowed eminent scholar chair in biostatistics at Georgia Southern University. He is also a senior consultant for biopharmaceuticals and government agencies with extensive expertise in clinical trial biostatistics and public health statistics. Professor Chen has written more than 150 refereed publications and co-authored/co-edited 25 books on clinical trial methodology, meta-analysis, causal inference, and public health statistics.

List of Reviewers

Jyun-You Chiang (School of Statistics, Southwestern University of Finance and Economics, Chengdu, China)

Yili Hong (Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA)

Debasis Kundu (Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, India)

Victor Leiva (School of Industrial Engineering, Pontificia University Catolica de Valparaiso, Chile)

Man Ho Ling (Department of Mathematics and Information Technology, The Education University of Hong Kong, China)

Carolina Marchant (Faculty of Basic Sciences, Universidad Catolica del Maule, Chile)

Chanseok Park (Department of Industrial Engineering, Pusan National University, Busan, Korea)

Hani Samawi (Department of Biostatistics, Georgia Southern University, Statesboro, GA, USA)

Helton Saulo (Department of Statistics, Universidad de Brasilia, Brazil)

Yan Shen (Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian, China)

Fu-Kwun Wang (Department of Industrial Management, National Taiwan University of Science and Technology)

Part I
Statistical Process Control

Some Recent Studies in Statistical Process Control

Peihua Qiu

Abstract Statistical process control (SPC) charts are widely used in manufacturing industries for quality control and management. They are used in more and more other applications, such as internet traffic monitoring, disease surveillance, and environmental protection. Traditional SPC charts designed for monitoring production lines in manufacturing industries are based on the assumptions that observed data are independent and identically distributed with a parametric in-control distribution. These assumptions, however, are rarely valid in practice. Therefore, recent SPC research focuses mainly on development of new control charts that are appropriate to use without these assumptions. In this article, we briefly introduce some recent studies on nonparametric SPC, control charts for monitoring dynamic processes, and spatio-temporal process monitoring. Control charts developed in these directions have found broad applications in practice.

1 Introduction

Since the first control chart suggested by Shewhart [46], statistical process control (SPC) charts have become a basic and powerful tool for quality control and management in manufacturing industries. Many different control charts have been developed in the past more than 80 years. These charts are mainly in the following four categories: Shewhart charts, cumulative sum (CUSUM) charts, exponentially weighted moving average (EWMA) charts, and charts based on change-point detection (CPD). For systematic descriptions about the basics of these control charts, see books Hawkins and Olwell [17], Montgomery [29], and Qiu [34].

Conventional control charts in the SPC literature are developed under the routine assumptions that process observations are independent and identically distributed (i.i.d.) with a parametric in-control (IC) distribution (e.g., normal).

P. Qiu (✉)
Department of Biostatistics, University of Florida, Gainesville, FL, USA
e-mail: pqiu@ufl.edu

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_1

These assumptions are rarely valid in practice. For instance, process observations collected at different time points could be serially correlated. Distributions of certain quality variables could be skewed and inappropriate to describe by a normal or another parametric distribution. In manufacturing industries, it might be reasonable to assume that the IC distribution of process observations does not change over time. But, in some other applications (e.g., monitoring of incidence rates of influenza over time), the process IC distribution usually changes over time and space, due to seasonality and other reasons. It has been well demonstrated in the literature that the conventional control charts designed under the above routine assumptions would not be reliable if one or more of their model assumptions are violated (e.g., [2, 4, 7, 36, 38, 42, 48]). So, much recent research effort in the SPC community has been made in developing more flexible control charts. This chapter aims to describe some of them in the research directions of nonparametric SPC, control charts for monitoring dynamic processes, and spatio-temporal process monitoring.

SPC can be roughly divided into two phases. In Phase I, we try to adjust a process under monitoring to make it run stably and satisfactorily (or IC), which usually happens when the process is first monitored (e.g., a machine for production is first installed). To know whether the process is IC, a Phase I control chart needs to be applied to a small dataset collected from the process, and adjust the process if it is not IC. This control-and-adjustment process usually needs to be repeated several times until it is certain that the process is IC. Then, a Phase II control chart can be designed properly, based on an IC dataset collected after Phase I SPC, for online process monitoring. This chapter mainly introduce methods for Phase II SPC, although many methods introduced here can be modified easily for Phase I SPC. Also, in some SPC applications, the process under monitoring cannot be adjusted easily (e.g., when monitoring incidence rates of influenza or satellite images of earth surface in a given region). In such cases, traditional Phase I SPC may not be relevant.

The remaining parts of the chapter are organized as follows. In Sect. 2, some conventional control charts are briefly introduced. Then, nonparametric SPC for cases when a parametric form is inappropriate or unavailable for describing the process distributions is discussed in Sect. 3. In Sect. 4, monitoring of processes with time-varying IC distributions (or dynamic processes) is discussed, followed by a discussion about spatio-temporal process monitoring in Sect. 5. Finally, some remarks conclude the chapter in Sect. 6.

2 Basic Control Charts

As mentioned in Sect. 1, early control charts in the SPC literature are in the framework of Shewhart charts that are described briefly in this section. Assume that X is a univariate quality variable in a specific process monitoring problem, it is continuous numerical, and its IC distribution is $N(\mu_0, \sigma^2)$. A batch of process observations obtained at the nth time point is denoted as

$$X_{n1}, X_{n2}, \ldots, X_{nm},$$

where $m \geq 2$ is the batch size. To test whether the process is IC at the nth time point, it is natural to use the following z-test: the process is declared out-of-control (OC) if

$$\overline{X}_n > \mu_0 + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{m}} \quad \text{or} \quad \overline{X}_n < \mu_0 - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{m}},$$

where $\overline{X}_n$ is the sample mean of $\{X_{n1}, X_{n2}, \ldots, X_{nm}\}$, and $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the $N(0,1)$ distribution. In practice, both μ_0 and σ could be unknown, and they need to be estimated from an IC dataset $\{(X^*_{i1}, X^*_{i2}, \ldots, X^*_{im}), i = 1, 2, \ldots, M\}$. Let $\overline{X}^*_i$ and R^*_i be the sample mean and sample range of $(X^*_{i1}, X^*_{i2}, \ldots, X^*_{im})$, for $i = 1, 2, \ldots, M$, and $\overline{\overline{X}}^*$ and $\overline{R}^*$ be the averages of $\{\overline{X}^*_i, i = 1, 2, \ldots, M\}$ and $\{R^*_i, i = 1, 2, \ldots, M\}$, respectively. Then, it can be checked that $\overline{\overline{X}}^*$ is an unbiased estimator of μ_0 and $\overline{R}^*/d_1(m)$ is an unbiased estimator of σ, where $d_1(m)$ is a constant that depends on m. When $m = 2, 3, 4, 5$, $d_1(m) = 1.128, 1.693, 2.059$ and 2.326, respectively. See Table 3.1 in [34] for more values of $d_1(m)$. After replacing μ_0 and σ by their estimates in the z-test, we obtain the Shewhart chart. So, the Shewhart chart declares a process mean shift at the nth time point if

$$\overline{X}_n > \overline{\overline{X}}^* + Z_{1-\alpha/2}\frac{\overline{R}^*}{d_1(m)\sqrt{m}} \quad \text{or} \quad \overline{X}_n < \overline{\overline{X}}^* - Z_{1-\alpha/2}\frac{\overline{R}^*}{d_1(m)\sqrt{m}}. \quad (1)$$

In manufacturing industries, we often choose $\alpha = 0.0027$. In such cases, $Z_{1-\alpha/2} = 3$. The popular terminology "six-sigma" in quality control and management is related directly to the above design of the Shewhart chart. Namely, the performance of a production process at time n can be considered IC if $\overline{X}_n$ is within an interval of six-sigma wide that is centered at μ_0, where sigma is the standard deviation of $\overline{X}_n$.

There are many different versions of the Shewhart chart (1) for detecting mean shifts. For instance, instead of using sample ranges for estimating the IC process standard deviation, we can also use sample standard deviations. The Shewhart chart (1) is constructed based on batch data with the batch size $m \geq 2$. When $m = 1$, there is only one observation at each time point. In such cases, the data are called individual observation data. There are some Shewhart charts suggested in the literature for monitoring individual observation data, some of which are based on the moving-window idea that the sample means and sample ranges used in constructing the Shewhart chart (1) are calculated from batches of observed data in different windows of observation times. There are many Shewhart charts in the literature suggested for detecting shifts in process variance, for monitoring binary or count data, and for many other purposes. See Chapter 3 in [34] for a detailed description.

The Shewhart chart (1) makes a decision whether a process is IC at a given time point based on the observed data at that time point only. It is thus ineffective for Phase II process monitoring in most cases, because the observed data in the past can also provide helpful information about the process performance at the current time point and such information is ignored completely by the Shewhart chart. To overcome this limitation, Page [32] suggested the first CUSUM chart, and then many different CUSUM charts have been suggested in the literature for different purposes (cf., [17, 34]). Next, we briefly describe the basic CUSUM chart for detecting a mean shift of a normal-distributed process. Assume that the IC process distribution is $N(\mu_0, \sigma^2)$, and the process observations for online monitoring are $\{X_n, n = 1, 2, \ldots\}$. Then, the CUSUM charting statistics for detecting a mean shift are defined by

$$\begin{aligned} C_n^+ &= \max(0, C_{n-1}^+ + (X_n - \mu_0)/\sigma - k), \\ C_n^- &= \min(0, C_{n-1}^- + (X_n - \mu_0)/\sigma + k), \qquad \text{for } n \geq 1, \end{aligned} \tag{2}$$

where $C_0^+ = C_0^- = 0$, and $k > 0$ is an allowance constant. The chart gives a signal when

$$C_n^+ > \rho_C \qquad \text{or} \qquad C_n^- < -\rho_C, \tag{3}$$

where $\rho_C > 0$ is a control limit. In the above CUSUM chart (2)–(3), the allowance constant k is usually pre-specified. Then, the control limit ρ_C is determined so that the IC average run length (ARL), denoted as ARL_0, equals a given value, where ARL_0 is defined as the average number of observation times from the beginning of process monitoring to a signal when the process is IC. From (2), it can be seen that (1) the charting statistics C_n^+ and C_n^- make use of the cumulative information in all available data by the current time point n, and (2) they re-start from 0 each time when the cumulative information suggests no significant evidence of a mean shift in the sense that $C_{n-1}^+ + (X_n - \mu_0)/\sigma < k$ and $C_{n-1}^- + (X_n - \mu_0)/\sigma > -k$. The re-starting mechanism of the CUSUM chart makes it possess a good theoretical property that it has the smallest value of OC ARL, denoted as ARL_1, among all control charts that have the same ARL_0 value (cf., [30]), where ARL_1 is defined as the average number of observation times from the occurrence of a real mean shift to a signal after the process becomes OC.

Although the CUSUM chart (2)–(3) has good properties for process monitoring, it is quite complicated to use, especially at the time when computing was expensive in the 1950s when the chart was first suggested. An alternative but simpler chart is the EWMA chart, first suggested by Roberts [45] in the first volume of *Technometrics*. In the same setup as that for the CUSUM chart (2)–(3), the EWMA charting statistic is defined as

$$E_n = \lambda X_n + (1 - \lambda) E_{n-1}, \tag{4}$$

where $E_0 = \mu_0$, and $\lambda \in (0, 1]$ is a weighting parameter. From (4), it is easy to check that

$$E_n = \lambda \sum_{i=1}^{n} (1-\lambda)^{n-i} X_i + (1-\lambda)^n \mu_0, \tag{5}$$

and when the process is IC up to the current time point n, we have

$$E_n \sim N\left(\mu_0, \ \frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2n}\right]\sigma^2\right). \tag{6}$$

Equation (5) implies that E_n is a weighted average of μ_0 and all available observations up to n, and the weight received by X_i decays exponentially fast when i moves away from n. So, it is easy to study the IC properties of E_n, including the IC distribution given in (6). Based on Expression (6), the EWMA chart gives a signal of process mean shift when

$$|E_n - \mu_0| > \rho_E \sigma \sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2n}\right]}, \tag{7}$$

where $\rho_E > 0$ is a control limit. In the EWMA chart (7), the weighting parameter λ is usually pre-specified, and the control limit ρ_E is chosen such that a given ARL_0 value is reached.

To use the Shewhart, CUSUM and EWMA charts described above, the IC parameters μ_0 and σ should be known or estimated in advance, which is inconvenient for certain applications. To overcome this limitation, Hawkins et al. [19] suggested a CPD chart described below. For process observations $X_1, X_2, \ldots, X_n$, it is assumed that they follow the following change-point model:

$$X_i = \begin{cases} \mu_0 + \varepsilon_i, & \text{if } i = 1, 2, \ldots, r, \\ \mu_1 + \varepsilon_i, & \text{if } i = r+1, r+2, \ldots, n, \end{cases}$$

where r is a change-point, and $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ is a sequence of i.i.d. random variables with the common distribution $N(0, \sigma^2)$. Then, the likelihood ratio test statistic for testing the existence of a change-point is

$$T_{\max,n} = \max_{1 \le j \le n-1} \sqrt{\frac{j(n-j)}{n}} \left|\overline{X}_j - \overline{X}'_j\right| \Big/ \widetilde{S}_j, \tag{8}$$

where $\overline{X}_j$ and $\overline{X}'_j$ are sample means of the first j and the remaining $n-j$ observations in $\{X_1, X_2, \ldots, X_n\}$, respectively, and $\widetilde{S}_j^2 = \sum_{i=1}^{j}(X_i - \overline{X}_j)^2 + \sum_{i=j+1}^{n}(X_i - \overline{X}'_j)^2$. The CPD chart gives a signal of mean shift when

$$T_{\max,n} > \rho_n, \tag{9}$$

where $\rho_n > 0$ is a control limit that may depend on n. After a signal is given, an estimate of the change-point r is given by the maximizer found in (8). Hawkins et al. [19] provided formulas for computing the values of ρ_n used in (9) for several commonly used ARL_0 values.

The description about the four types of basic control charts given above is for detecting process mean shifts when process observations are univariate. There are many different versions of each type for detecting shifts in process mean, variance and other aspects of the process distribution. There are many control charts for monitoring multivariate processes as well. See references, such as Crosier [9], Gan [11], Hawkins [14, 15], Hawkins et al. [18], Healy [20], Lowry et al. [26], Mason et al. [28], Sparks [47], Woodall and Ncube [49], Zamba and Hawkins [52], Zou and Qiu [55], and many more.

3 Nonparametric Control Charts

The basic control charts discussed in Sect. 2 are all based on the assumption that IC process observations follow a parametric (e.g., normal) distribution. In practice, this assumption is rarely valid and distributions of quality variables are often skewed and difficult to describe well by any parametric distributions. It has been well demonstrated in the literature that conventional control charts are unreliable to use in cases when their distributional assumptions are invalid (e.g., [4, 6, 7, 13, 33, 36, 38, 39]). Thus, distribution-free or nonparametric SPC is under rapid development in the past 20 years. Some fundamental nonparametric SPC charts are described below in this section.

The first type of nonparametric SPC charts makes use of the ordering or ranking information in process observations collected at different time points. Let us first discuss univariate cases when there is only one quality variable X involved in process monitoring. Assume that the batch of observed data at the nth time point is $\{X_{n1}, X_{n2}, \ldots, X_{nm}\}$, for $n \geq 1$. Let ξ_0 be the median of the IC process distribution, and R_{nj} be the rank of $|X_{nj} - \xi_0|$ in the sequence $\{|X_{n1} - \xi_0|, |X_{n2} - \xi_0|, \ldots, |X_{nm} - \xi_0|\}$. Then, the sum of the Wilcoxon signed-ranks within the nth batch of observations is defined as

$$\psi_n = \sum_{j=1}^{m} \text{sign}(X_{nj} - \xi_0) R_{nj}, \tag{10}$$

where $\text{sign}(u) = -1, 0, 1$, respectively, when $u < 0, = 0, > 0$. The absolute value of ψ_n should be small when the process is IC, because the positive and negative values in the summation of (10) will be mostly canceled out in such cases. On the other hand, the value of ψ_n will be positively (negatively) large if there is an upward (downward) mean shift. Therefore, ψ_n can be used for detecting process mean shift. Also, it can be checked that the IC distribution of ψ_n does not depend

on the original IC process distribution as long as that distribution is symmetric. In that sense, control charts constructed based on ψ_n are distribution-free. As a matter of fact, a number of distribution-free control charts based on ψ_n have been suggested in the literature. See, for instance, Bakir [3], Chakraborti and Eryilmaz [5], Graham et al. [12], Li et al. [25], and Mukherjee et al. [31]. For instance, the EWMA charting statistic suggested in Chapter 8 of Qiu [34] is defined as

$$E_n = \lambda\psi_n + (1-\lambda)E_{n-1}, \qquad \text{for } n \geq 1, \tag{11}$$

where $E_0 = 0$ and $\lambda \in (0, 1]$ is a weighting parameter. When the process distribution is symmetric, it can be checked that the IC mean and variance of E_n are 0 and $m(m+1)(2m+1)/6$, respectively. So, the chart gives a signal when

$$|E_n| > \rho_W \sqrt{\left[\frac{m(m+1)(2m+1)}{6}\right]\left[\frac{\lambda}{2-\lambda}\right]\left[1-(1-\lambda)^{2n}\right]}, \tag{12}$$

where $\rho_W > 0$ is a parameter chosen to reach a given ARL_0 value.

Besides ψ_n, there are some alternative rank-based statistics used for constructing nonparametric control charts. These include the ones based on the sign test statistic (e.g., [27]), the Cucconi test statistic (e.g., [8]), the nonparametric likelihood ratio test (e.g., [56]), the Mann-Whitney two-sample test (e.g., [16]), and more. For multivariate SPC problems, Qiu and Hawkins [38, 39] suggested CUSUM charts for detecting process mean shifts using antiranks of different quality variables, Zou and Tsung [57] suggested an EWMA chart using spatial signs, and Zou et al. [58] and Holland and Hawkins [21] suggested different nonparametric control charts using spatial ranks. See Qiu [36] for a discussion about other rank-based control charts.

The second type of nonparametric SPC charts is based on data categorization. In multivariate cases, the major difficulty in describing a process distribution when it is non-Gaussian is that the association among different quality variables can have infinitely many possibilities and it is hard to describe such association properly in general. However, if the quality variables are all categorical, then there are some mature statistical methodologies in the area of categorical data analysis (cf., [1]) for describing the association among categorical variables. Based on this observation, Qiu [33] suggested a general framework for constructing nonparametric control charts based on data categorization and categorical data analysis. Assume that process observations are

$$\mathbf{X}_n = (X_{n1}, X_{n2}, \ldots, X_{np})', \qquad \text{for } n \geq 1,$$

where p is the dimension of the quality vector $\mathbf{X}$. Let the IC median of X_{nj} be $\widetilde{\mu}_j$, for $j = 1, 2, \ldots, p$, which can be estimated from an IC data. Define

$$Y_{nj} = I(X_{nj} > \widetilde{\mu}_j), \qquad \text{for } j = 1, 2, \ldots, p, \tag{13}$$

and $\mathbf{Y}_n = (Y_{n1}, Y_{n2}, \ldots, Y_{np})'$, where $I(a)$ is an indicator function that equals 1 if a is "true" and 0 otherwise. Then, $\mathbf{Y}_n$ is the categorized version of $\mathbf{X}_n$. The IC distribution of $\mathbf{Y}_n$, can be described by a log-linear model which can be estimated from the IC data. The estimated IC distribution of $\mathbf{Y}_n$ is denoted as

$$\left\{ f^{(0)}_{j_1 j_2 \ldots j_p}, \quad j_1, j_2, \ldots, j_p = 1, 2 \right\},$$

where $j_1, j_2, \ldots, j_p$ are indices of the p-way contingency table associated with the categorized data $\mathbf{Y}_n$ and its distribution (note that each dimension of the contingency table has two categories). For $j_1, j_2, \ldots, j_p = 1, 2$, define

$$g_{nj_1 j_2 \ldots j_p} = I(Y_{n1} = j_1 - 1, Y_{n2} = j_2 - 1, \ldots, Y_{np}(i) = j_p - 1),$$

$\mathbf{g}_n$ is a vector of all $g_{nj_1 j_2 \ldots j_p}$ values, and $\mathbf{f}^{(0)}$ is the vector of all $f^{(0)}_{j_1 j_2 \ldots j_p}$. Then, $\mathbf{g}_n$ and $\mathbf{f}^{(0)}$ are vectors of the observed and expected counts of the contingency table at the time point n, respectively. Let $\mathbf{U}_0^{obs} = \mathbf{U}_0^{exp} = \mathbf{0}$ be two 2^p-dimensional vectors, and

$$\begin{cases} \mathbf{U}_n^{obs} = \mathbf{U}_n^{exp} = \mathbf{0}, & \text{if } B_n \le k \\ \mathbf{U}_n^{obs} = \left(\mathbf{U}_{n-1}^{obs} + \mathbf{g}_n\right)(1 - k/B_n), & \text{if } B_n > k \\ \mathbf{U}_n^{exp} = \left(\mathbf{U}_{n-1}^{exp} + \mathbf{f}^{(0)}\right)(1 - k/B_n), & \end{cases}$$

where

$$B_n = \left\{ \left(\mathbf{U}_{n-1}^{obs} - \mathbf{U}_{n-1}^{exp}\right) + \left(\mathbf{g}_n - \mathbf{f}^{(0)}\right) \right\}' \left(\text{diag}(\mathbf{U}_{n-1}^{exp} + \mathbf{f}^{(0)}) \right)^{-1} \left\{ \left(\mathbf{U}_{n-1}^{obs} - \mathbf{U}_{n-1}^{exp}\right) + \left(\mathbf{g}_n - \mathbf{f}^{(0)}\right) \right\},$$

$k \ge 0$ is an allowance constant, diag($\mathbf{a}$) denotes a diagonal matrix with its diagonal elements equal to the corresponding elements of the vector $\mathbf{a}$, and the superscripts "obs" and "exp" denote the observed and expected counts, respectively. Define

$$C_n = \left(\mathbf{U}_n^{obs} - \mathbf{U}_n^{exp}\right)' \left[\text{diag}(\mathbf{U}_n^{exp})\right]^{-1} \left(\mathbf{U}_n^{obs} - \mathbf{U}_n^{exp}\right). \tag{14}$$

Then, a location shift in $\mathbf{X}_n$ is signaled if

$$C_n > h, \tag{15}$$

where $h > 0$ is a control limit chosen to achieve a given ARL_0 level.

A chart similar to the one defined in (13)–(15) was suggested in univariate cases by Qiu and Li [40], where the number of categories in categorizing the quality variable can be larger than 2. From the large comparative studies in Qiu and Li [40, 41], it can be seen that the chart based on data categorization has some advantages in

terms of the ARL_1 metric over certain alternative nonparametric control charts in various cases considered, although it still has much room for possible improvement. For instance, data categorization would lose information in the original data. It is thus important to study how to make the lost information as small as possible while keeping the favorable properties of the related nonparametric control charts. In this regard, one possible improvement is to make use of the ordering information among the categories of the categorized data when constructing control charts, which has been discussed recently by Li [22].

4 Charts for Monitoring Dynamic Processes

The conventional control charts discussed in Sect. 2 usually require the assumption that process observations have an identical distribution at different time points when the related process is IC. This assumption is not valid in certain applications. One example mentioned in Sect. 1 is about monitoring of disease incidence rates. In this example, the distribution of disease incidence rates would change over time because of seasonality and other reasons, even in cases when there are no disease outbreaks. In such cases, the disease incidence rate process has a time-varying IC distribution. Such processes are called *dynamic* processes in this paper. For monitoring dynamic processes, the conventional control charts would be unreliable to use, and new monitoring charts are needed. Recently, we developed several control charts for monitoring dynamic processes, which are introduced below in this section.

Qiu and Xiang [42] suggested a so-called *dynamic screening system (DySS)* for monitoring dynamic processes with a single quality/performance variable. The DySS method consists of three main steps described below.

Step I The regular longitudinal pattern of the performance variable y is first estimated from an *IC dataset* containing longitudinal observations of a group of m well-functioning subjects.

Step II For a new subject to monitor, his/her observations are first standardized using the estimated regular longitudinal pattern obtained in Step I.

Step III The standardized observations of the new subject are then monitored by a conventional control chart and a signal is given as soon as all available data suggest a significant shift in the longitudinal pattern of the subject under monitoring from the estimated regular pattern.

Assume that the observed longitudinal data of the m well-functioning subjects in the IC dataset follow the model

$$y(t_{ij}) = \mu(t_{ij}) + \sigma(t_{ij})\varepsilon(t_{ij}), \qquad \text{for } j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, m, \quad (16)$$

where $t_{ij} \in [0, T]$ are observation times, $y(t_{ij})$ is the jth observation of the ith subject, $\mu(t_{ij})$ and $\sigma(t_{ij})$ are the mean and standard deviation of $y(t_{ij})$, and $\varepsilon(t_{ij})$ is the standardized random error with mean 0 and variance 1. Qiu and Xiang

[42] suggested a four-step procedure for estimating $\mu(t)$ and $\sigma^2(t)$ using the local linear kernel smoothing procedure. The estimators are denoted as $\widehat{\mu}(t)$ and $\widehat{\sigma}^2(t)$, respectively. For a subject to monitor, when his/her performance is IC, his/her observations should follow model (16), although the observation times, denoted as t_j^*, for $j = 1, 2, \ldots$, could be different from those in model (1). So, his/her observations $\{y(t_j^*), j \geq 1\}$ can be standardized by

$$\widehat{\varepsilon}(t_j^*) = \frac{y(t_j^*) - \widehat{\mu}(t_j^*)}{\widehat{\sigma}(t_j^*)}, \qquad \text{for } j \geq 1, \tag{17}$$

where $\widehat{\mu}(t)$ and $\widehat{\sigma}(t)$ are estimated from the IC data. Obviously, the standardized observations $\{\widehat{\varepsilon}(t_j^*), j \geq 1\}$ in (17) would have mean 0 and variance 1 when the subject under monitoring is IC. To detect an upward mean shift in the original performance variable y for the given subject, Qiu and Xiang [42] suggested using the CUSUM charting statistic defined as

$$C_j^+ = \max(0, C_{j-1}^+ + \widehat{\varepsilon}(t_j^*) - k), \qquad \text{for } j \geq 1, \tag{18}$$

where $C_0^+ = 0$ and $k > 0$ is an allowance constant. Then, the chart gives a signal when

$$C_j^+ > h_C, \tag{19}$$

where $h_C > 0$ is a control limit. For detecting a downward or arbitrary shift, a downward or two-sided CUSUM chart can be used.

As discussed in Sect. 2, the performance of a control chart is usually measured by ARL_0 and ARL_1. However, in many dynamic process monitoring applications, the observation times are often unequally spaced. In such cases, ARL_0 and ARL_1 are irrelevant any more. Instead, Qiu and Xiang [42] suggested using the average time to signal (ATS), defined below. Let ω be a basic time unit so that all observation times are its integer multiples. Then, we define $n_j^* = t_j^*/\omega$, for $j = 1, 2, \ldots$, where $n_0^* = t_0^* = 0$. For a subject whose longitudinal performance is IC, assume that a signal is given at the sth observation time. Then, $E(n_s^*)$ is called the IC ATS, denoted as ATS_0. Similarly, for a subject whose longitudinal performance starts to deviate from the regular longitudinal pattern at the time point τ, the value $E(n_s^*|n_s^* \geq \tau) - \tau$ is called OC ATS, denoted as ATS_1. For the chart (18)–(19), the value of ATS_0 can be specified beforehand, and it performs better for detecting a shift of a given size if its ATS_1 value is smaller.

The DySS method discussed above is for monitoring univariate dynamic processes only. Its multivariate version has been developed in Qiu and Xiang [43]. Li and Qiu [23, 24] suggested univariate and multivariate DySS methods that were effective in cases when process observations were serially correlated. In the chart (18)–(19), the fact that process observations are often unequally spaced is considered in the performance measures ATS_0 and ATS_1 only, and it is not taken

into account in the construction of the chart. This limitation was lifted in the EWMA chart proposed by Qiu et al. [44].

In many applications, especially those outside the manufacturing industries, longitudinal processes for monitoring often have time-varying IC distributions. Therefore, dynamic process monitoring is an important research topic. Although we have developed a number of control charts for that purpose, there are still many open research problems. For instance, the DySS method described above depends heavily on the estimated regular longitudinal pattern obtained from an IC dataset. However, it is often challenging to choose an appropriate IC dataset in practice. Proper estimation of the regular longitudinal pattern is challenging as well, especially in cases when the size of the IC dataset is quite small.

5 Charts for Monitoring Spatio-Temporal Processes

We experienced SARS and many other damaging infectious diseases (cf., [10]). To monitor their incidence rates, some global, national and regional reporting systems have been developed. For instance, Florida Department of Health (FDOH) has built the Electronic Surveillance System for the Early Notification of Community-based Epidemics at Florida (ESSENCE-FL) recently, which is a syndromic surveillance system for collecting near real-time pre-diagnostic data from participating hospitals and urgent care centers in Florida. Figure 1 presents the observed incidence rates of influenza-like illness (ILI) for all 67 counties of Florida on 06/01/2012 (a summer time) and 12/01/2012 (a winter time). One important feature of these disease incidence data is that observed data collected at different places and different times are usually correlated: with the ones closer in space or time being more correlated. This kind of spatio-temporal (ST) correlation, however, is hidden in the observed data, and cannot be observed directly. Also, the disease incidence data have the seasonality and other temporal variation, and their temporal patterns could be different at different places (i.e., the spatial variation), as seen in Fig. 1.

The conventional SPC charts described in Sect. 2 require the assumptions that process observations are independent and identically distributed when the underlying process is IC. These assumptions are all violated in the ST process monitoring problem discussed above, because of the ST data correlation and the fact that disease incidence rates would have time- and space-varying distribution even in cases when no disease outbreaks are present. Therefore, specific control charts for handling ST processes are needed.

In a case study for analyzing a foot, hand and mouth disease dataset, Zhang et al. [53] suggested a procedure consisting of three steps: detrend, decorrelation, and sequential monitoring, which are briefly described below. (1) Seasonality in the observed disease incidence data is first described by a nonparametric longitudinal model, which can be estimated from an IC dataset and then eliminated from all observed data. (2) Temporal autocorrelation in the detrended data is modeled by an ARIMA model, and then eliminated from the detrended data. (3) The detrended

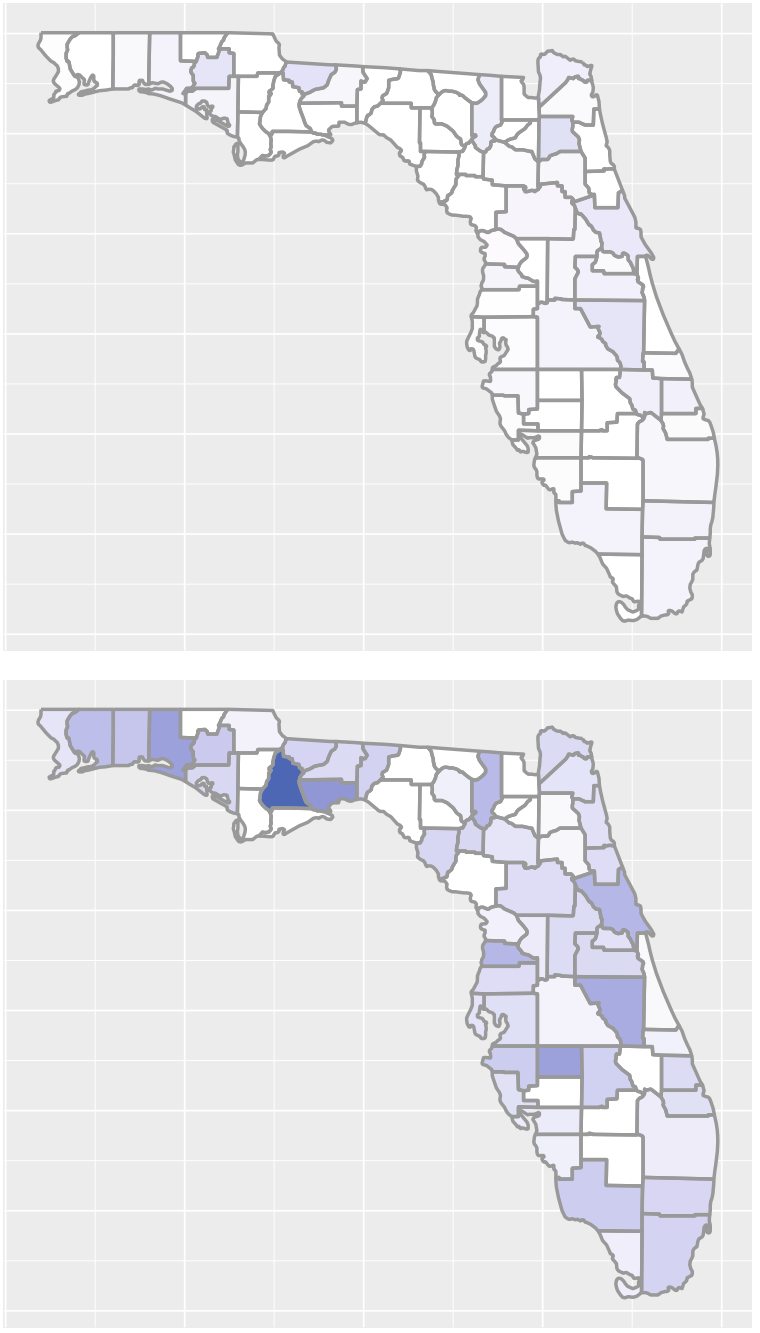


Fig. 1 Observed ILI incidence rates in Florida on 06/01/2012 (top) and 12/01/2012 (bottom). Stronger colors denote larger values

and decorrelated data obtained in step (2) are then sequentially monitored by an SPC chart. A similar procedure was used for analyzing an AIDS data in Zhang et al. [54]. This three-step method, however, can monitor the disease incidence rates at a single location only, and it cannot monitor the data at multiple locations simultaneously while accommodating ST data correlation properly.

To overcome the limitation of the three-step method by Zhang et al. [53], Yang and Qiu [50] suggested a flexible approach for spatio-temporal data modeling, briefly described below. Let Ω and $[0, T]$ be a 2-D region and a given time interval. The observed disease incidence rates in $\Omega \times [0, T]$ are assumed to follow the model

$$y(t_i, \mathbf{s}_{ij}) = \lambda(t_i, \mathbf{s}_{ij}) + \varepsilon(t_i, \mathbf{s}_{ij}), \qquad \text{for } j = 1, 2, \ldots, m_i, \ i = 1, 2, \ldots, n, \quad (20)$$

where $t_i \in [0, T]$ is the ith observation time point, $\mathbf{s}_{ij} \in \Omega$ is the jth observation location at time t_i, $\varepsilon(t_i, \mathbf{s}_{ij})$ is a zero-mean random error, m_i is the number of observation locations at time t_i, and n is the number of time points in the dataset. The correlation structure in the observed data can be described by the covariance function

$$V(\boldsymbol{u}; \boldsymbol{v}) = E\left[\varepsilon(\boldsymbol{u})\varepsilon(\boldsymbol{v})\right] = \sigma(\boldsymbol{u})\sigma(\boldsymbol{v})Cor(\varepsilon(\boldsymbol{u}), \varepsilon(\boldsymbol{v})), \qquad \text{for } \boldsymbol{u}, \boldsymbol{v} \in [0, T] \times \Omega, \quad (21)$$

where $\sigma^2(\cdot)$ is the variance function and $Cor(\cdot, \cdot)$ is the correlation function. The mean function $\lambda(t, \mathbf{s})$ is then estimated by a spatio-temporal local linear kernel smoothing procedure. The estimator is denoted as $\widehat{\lambda}(t, \mathbf{s})$. To accommodate the ST data correlation, the bandwidths used in estimating $\lambda(t, \mathbf{s})$ should be chosen carefully. To this end, a modified cross-validation procedure was proposed in Yang and Qiu [50]. After $\widehat{\lambda}(t, \mathbf{s})$ is obtained, $V(\boldsymbol{u}, \boldsymbol{v})$ can be estimated by moment estimation from the residuals. The resulting estimator is denoted as $\widehat{V}(\boldsymbol{u}, \boldsymbol{v})$.

Based on the above ST data modeling approach, Yang and Qiu [51] suggested a CUSUM chart for monitoring ST processes, which consists of several steps. First, the regular longitudinal pattern of the spatial disease incidence rates in cases when no disease outbreaks are present can be described by $\lambda(t, \mathbf{s})$ and $V(\boldsymbol{u}, \boldsymbol{v})$ in (20) and (21), which can be estimated from an IC dataset by the ST modeling procedure discussed above. Then, they can be used for online monitoring of the disease incidence rates $y(t_i^*, \mathbf{s}_{ij}^*)$ observed at locations $\{\mathbf{s}_{ij}^*, j = 1, 2, \ldots, m_i^*\}$ and times t_i^*, for $i = 1, 2, \ldots$. Define $\boldsymbol{y}(t_i^*) = (y(t_i^*, \mathbf{s}_{i1}^*), y(t_i^*, \mathbf{s}_{i2}^*), \ldots, y(t_i^*, \mathbf{s}_{im_i^*}^*))'$, for all i. When the process is IC, the observed data are assumed to follow model (20) in the sense that $y(t_i^*, \mathbf{s}_{ij}^*) = \lambda(t_i^*, \mathbf{s}_{ij}^*) + \varepsilon(t_i^*, \mathbf{s}_{ij}^*)$, for $j = 1, 2, \ldots, m_i^*$ and $i = 1, 2, \ldots$, and the mean function $\lambda(t, \mathbf{s})$ is assumed periodic in time with the period T. Namely, $\lambda(t_i^*, \mathbf{s}_{ij}^*) = \lambda(t_i^{**}, \mathbf{s}_{ij}^*)$, where $t_i^* = t_i^{**} + \ell T$ for all i, $t_i^{**} \in [0, T]$, and $\ell \geq 1$ is an integer. Second, decorrelate and standardize all observed data up to the current time point i: $\{\boldsymbol{y}(t_1^*), \boldsymbol{y}(t_2^*), \ldots, \boldsymbol{y}(t_i^*)\}$. The decorrelated and standardized data are

denoted as $\widehat{\boldsymbol{e}}(t_1^*), \widehat{\boldsymbol{e}}(t_2^*), \ldots, \widehat{\boldsymbol{e}}(t_i^*)$. Then, the suggested CUSUM charting statistic for detecting upward shifts in the disease incidence rates is

$$C_i^+ = \max\left(0, C_{i-1}^+ + \frac{\widehat{\boldsymbol{e}}(t_i^*)'\widehat{\boldsymbol{e}}(t_i^*) - m_i^*}{\sqrt{2m_i^*}} - k\right), \qquad \text{for } i \geq 1, \tag{22}$$

where $C_0^+ = 0$ and $k > 0$ is an allowance constant. The chart gives a signal when

$$C_i^+ > \gamma, \tag{23}$$

where $\gamma > 0$ is a control limit. To determine γ in (23) so that the chart (22)–(23) has a specific ARL_0 value, Yang and Qiu [51] suggested using a block bootstrap procedure.

6 Concluding Remarks

In the previous sections, we have briefly introduced some recent research in SPC, after an introduction of four types of basic SPC charts. The recent research introduced is mainly on nonparametric SPC, dynamic process control, and spatio-temporal process monitoring. These research topics aim to handle cases when the regular assumptions in SPC that IC process observations are independent and identically distributed with a specific parametric distribution are violated. In each of these research topics, there are still many open questions that need to be addressed in our future research. For instance, in the nonparametric SPC area, there have been many nonparametric control charts proposed. Systematic comparison of these charts should be important for them to be used in real-data applications. Also, both ranking and data categorization would lose information in the original observed data. It should be important to study how to minimize the lost information while keep all the favorable properties of nonparametric control charts. For dynamic process monitoring, accurate estimation of the regular longitudinal pattern from the IC data is critically important. Usually, the size of IC data is limited. In such cases, self-starting procedures to expand the initial IC dataset, by combining it with observed data during procedure monitoring after it is confirmed that the process under monitoring is IC at the current time point, might be one way to overcome the difficulty, which needs to be further studied in the future. Proper monitoring of spatio-temporal processes is important but challenging. The chart (22)–(23) represents our first research effort on that topic, and many issues, including proper accommodation of important covariates, need to be addressed in future research.

In the big data era, SPC will find more and more applications (cf., [35, 37]). In these new applications, the related process monitoring problems could become more complicated. For instance, sequential monitoring of images has broad applications

in manufacturing industries, traffic monitoring, medical diagnostics, and more. But, images often have edges and other complicated structures. Also, images obtained at different times should be geometrically aligned properly for meaning analysis of the image sequence. These features of image data, however, would make proper monitoring of an image sequence extremely challenging. So, new SPC methods are needed for handling such new applications.

References

1. Agresti, A. (2002). *Categorical data analysis*, 2nd ed. John Wiley & Sons: New York.
2. Apley, D. W., & Tsung, F. (2002). The autoregressive T^2 chart for monitoring univariate autocorrelated processes. *Journal of Quality Technology, 34*, 80–96.
3. Bakir, S. T. (2004). A distribution-free Shewhart quality control chart based on signed-ranks. *Quality Engineering, 16*, 611–621.
4. Capizzi, G. (2015). Recent advances in process monitoring: Nonparametric and variable-selection methods for phase I and phase II. *Quality Engineering, 27*, 44–67.
5. Chakraborti, S., & Eryilmaz, S. (2007). A nonparametric Shewhart-type signed-rank control chart based on runs. *Communications in Statistics - Simulation and Computation, 36*, 335–356.
6. Chakraborti, S., Qiu, P., & Mukherjee, A. (2015). Special issue on nonparametric statistical process control charts. *Quality and Reliability Engineering International, 31*, 1–152.
7. Chakraborti, S., van der Laan, P., & Bakir, S. T. (2001). Nonparametric control charts: An overview and some results. *Journal of Quality Technology, 33*, 304–315.
8. Chowdhury, S., Mukherjee, A., & Chakraborti, S. (2014). A new distribution-free control chart for joint monitoring of unknown location and scale parameters of continuous distributions. *Quality and Reliability Engineering International, 30*, 191–204.
9. Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics, 30*, 291–303.
10. Fiore, A. E., Uyeki, T. M., Broder, K., Finelli, L., Euler, G. L., et al. (2010). Prevention and control of influenza with vaccines: Recommendations of the advisory committee on immunization practices (ACIP). *MMWR Recommendations and Reports, 59*, 1–62.
11. Gan, F. F. (1993). An optimal design of CUSUMcontrol charts for binomial counts. *Journal of Applied Statistics, 20*, 445–460.
12. 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*, 2490–2503.
13. Hackl, P., & Ledolter, J. (1992). A new nonparametric quality control technique. *Communications in Statistics-Simulation and Computation, 21*, 423–443.
14. Hawkins, D. M. (1987). Self-starting CUSUM charts for location and scale. *Journal of the Royal Statistical Society. Series D (The Statistician), 36*, 299–316.
15. Hawkins, D. M. (1991). Multivariate quality control based on regression-adjusted variables. *Technometrics, 33*, 61–75.
16. Hawkins, D. M., & Deng, Q. (2010). A nonparametric change-point control chart. *Journal of Quality Technology, 42*, 165–173.
17. Hawkins, D. M., & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*, Springer, New York.
18. Hawkins, D. M., Choi, S., & Lee, S. (2007). A general multivariate exponentially weighted moving-average control chart. *Journal of Quality Technology, 39*, 118–125.
19. Hawkins, D. M., Qiu, P., & Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of Quality Technology, 35*, 355–366.

20. Healy, J. D. (1987). A note on multivariate CUSUM procedures. *Technometrics, 29*, 409–412.
21. Holland, M. D., & Hawkins, D. M. (2014). A control chart based on a nonparametric multivariate change-point model. *Journal of Quality Technology, 46*, 63–77.
22. Li, J. (2017). Nonparametric adaptive CUSUM chart for detecting arbitrary distributional changes. arXiv:1712.05072.
23. Li, J., & Qiu, P. (2016). Nonparametric dynamic screening system for monitoring correlated longitudinal data. *IIE Transactions, 48*, 772–786.
24. Li, J., & Qiu, P. (2017). Construction of an efficient multivariate dynamic screening system. *Quality and Reliability Engineering International, 33*, 1969–1981.
25. Li, S. Y., Tang, L. C., & Ng, S. H. (2010). Nonparametric CUSUM and EWMA control charts for detecting mean shifts. *Journal of Quality Technology, 42*, 209–226.
26. Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics, 34*, 46–53.
27. Lu, S. L. (2015). An extended nonparametric exponentially weighted moving average sign control chart. *Quality and Reliability Engineering International, 31*, 3–13.
28. Mason, R. L., Chou, Y. M., & Young, J. C. (2001). Applying Hotelling's T^2 statistic to batch processes. *Journal of Quality Technology, 33*, 466–479.
29. Montgomery, D. C. (2012). *Introduction to statistical quality control*. Wiley, New York.
30. Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *The Annals of Statistics, 14*, 1379–1387.
31. Mukherjee, A., Graham, M. A., & Chakraborti, S. (2013). Distribution-free exceedance CUSUM control charts for location. *Communications in Statistics - Simulation and Computation, 42*, 1153–1187.
32. Page, E. S. (1954). Continuous inspection scheme. *Biometrika, 41*, 100–115.
33. Qiu, P. (2008). Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions, 40*, 664–677.
34. Qiu, P. (2014). *Introduction to statistical process control*. Chapman Hall/CRC, Boca Raton, FL.
35. Qiu, P. (2017). Statistical process control charts as a tool for analyzing big data. In E. Ahmed (Ed.), *Big and complex data analysis: Statistical methodologies and applications* (pp. 123–138), Springer, Cham.
36. Qiu, P. (2018a). Some perspectives on nonparametric statistical process control. *Journal of Quality Technology, 50*, 49–65.
37. Qiu, P. (2018b). Jump regression, image processing and quality control (with discussions). *Quality Engineering, 30*, 137–153.
38. Qiu, P., & Hawkins, D. M. (2001). A rank based multivariate CUSUM procedure. *Technometrics, 43*, 120–132.
39. Qiu, P., & Hawkins, D. M. (2003). A nonparametric multivariate CUSUM procedure for detecting shifts in all directions. *JRSS-D (The Statistician), 52*, 151–164.
40. Qiu, P., & Li, Z. (2011a). On nonparametric statistical process control of univariate processes. *Technometrics, 53*, 390–405.
41. Qiu, P., & Li, Z. (2011b). Distribution-free monitoring of univariate processes. *Statistics and Probability Letters, 81*, 1833–1840.
42. Qiu, P., & Xiang, D. (2014). Univariate dynamic screening system: An approach for identifying individuals with irregular longitudinal behavior. *Technometrics, 56*, 248–260.
43. Qiu, P., & Xiang, D. (2015). Surveillance of cardiovascular diseases using a multivariate dynamic screening system. *Statistics in Medicine, 34*, 2204–2221.
44. Qiu, P., Zi, X., & Zou, C. (2018). Nonparametric dynamic curve monitoring. *Technometrics, 60*, 386–397.
45. Roberts, S. V. (1959). Control chart tests based on geometric moving averages. *Technometrics, 1*, 239–250.
46. Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. D. Van Nostrand Company, New York.

47. Sparks, R. S. (2000). CUSUM charts for signalling varying location shifts. *Journal of Quality Technology, 32*, 157–171.
48. Wardell, D. G., Moskowitz, H., & Plante, R. D. (1994). Run length distributions of special-cause control charts for correlated processes. *Technometrics, 36*, 3–17.
49. Woodall, W. H., & Ncube, M. M. (1985). Multivariate CUSUM quality-control procedures. *Technometrics, 27*, 285–292.
50. Yang, K., & Qiu, P. (2018). Spatio-temporal incidence rate data analysis by nonparametric regression. *Statistics in Medicine, 37*, 2094–2107.
51. Yang, K., & Qiu, P. (2019). Online sequential monitoring of disease incidence rates with an application to the Florida influenza-like illness data. submitted.
52. Zamba, K. D., & Hawkins, D. M. (2006). A multivariate change-point model for statistical process control. *Technometrics, 48*, 539–549.
53. Zhang, J., Kang, Y., Yang, Y., & Qiu, P. (2015). Statistical monitoring of the hand, foot, and mouth disease in China. *Biometrics, 71*, 841–850.
54. Zhang, J., Qiu, P., & Chen, X. (2016). Statistical monitoring-based alarming systems in modeling the AIDS epidemic in the US, 1985–2011. *Current HIV Research, 14*, 130–137.
55. Zou, C., & Qiu, P. (2009). Multivariate statistical process control using LASSO. *Journal of the American Statistical Association, 104*, 1586–1596.
56. Zou, C., & Tsung, F. (2010). Likelihood ratio-based distribution-free EWMA control charts. *Journal of Quality Technology, 42*, 1–23.
57. Zou, C., & Tsung, F. (2011). A multivariate sign EWMA control chart. *Technometrics, 53*, 84–97.
58. Zou, C., Wang, Z., & Tsung, F. (2012). A spatial rank-based multivariate EWMA control chart. *Naval Research Logistics, 59*, 91–110.

Statistical Quality Control and Reliability Analysis Using the Birnbaum-Saunders Distribution with Industrial Applications

Víctor Leiva, Carolina Marchant, Fabrizio Ruggeri, and Helton Saulo

Abstract Quality improvement has been an important aspect considered by companies since the last century. However, today it is even more relevant in business and industry, particularly for production and service companies. Statistical quality control is the quantitative tool for quality improvement. The Gaussian distribution was the main ingredient of this quantitative tool, but nowadays new distributions are being considered, some of them taking into account asymmetry. The Birnbaum-Saunders model is one of these distributions and has recently received considerable attention because of its interesting properties and its relationship with the Gaussian distribution. Since its origins and applications in material science, the Birnbaum-Saunders distribution has found widespread uses in several areas, including quality control, with now well-developed methods that allow in-depth analyses. In this work, statistical quality control and reliability tools based on the Birnbaum-Saunders distribution are introduced. Implementation of those tools is presented using the R software. For the internal quality of companies, control charts for attributes and variables, as well as their multivariate versions and capability indices, are presented, discussed and illustrated with real data. For external quality, acceptance sampling plans are also presented and discussed. The main aspects

V. Leiva (✉)
School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
e-mail: victorleivasanchez@gmail.com
URL: http://www.victorleiva.cl

C. Marchant
Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile
e-mail: carolina.marchant.fuentes@gmail.com

F. Ruggeri
CNR-IMATI, Milano, Italy
e-mail: fabrizio@mi.imati.cnr.it

H. Saulo
Department of Statistics, Universidade de Brasilia, Brasilia, Brazil
e-mail: heltonsaulo@gmail.com

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_2

of reliability, sometimes defined as ‘quality over time”, is discussed using the Birnbaum-Saunders distribution, illustrating them with data on nanotechnologies.

Keywords Acceptance sampling plans · Capability indices · Control charts for attribute · Control charts for variables · Multivariate control charts · Reliability · R software

1 Introduction

The Birnbaum-Saunders distributions are a family of probabilistic models originating from the law on cumulative damage related to fatigue and strength of materials; see [11]. Its origins allow the Birnbaum-Saunders model to be interpreted as a life distribution, because it describes the time spent until the extension of a crack exceeds a threshold leading to a failure. This family of distributions is unimodal, positively skewed and useful for modeling data that take values greater than zero; see [4]. The Birnbaum-Saunders family has two parameters, which modify shape and scale of the distribution; see [10]. Birnbaum-Saunders distributions have been extensively studied because of their good properties and relation with the Gaussian distribution. In particular, a random variable following the Birnbaum-Saunders distribution can be considered as a transformation of a random variable with (standard) Gaussian distribution; see [7, pp. 651–663] and [10, p. 18]. Although the main applications of the Birnbaum-Saunders distribution lie naturally in engineering, this model has also been successfully applied to several other areas of knowledge such as agriculture, air contamination, bioengineering, business, economics, environment, finance, food and textile industries, forestry, human and tree mortality, informatics, insurance, inventory management, medicine, nutrition, pharmacology, psychology, neurology, queueing theory, toxicology, water quality, and wind energy; see details in [10]. However, there is no compendium gathering tools for statistical quality control based on the Birnbaum-Saunders distribution.

The current work plans (i) to introduce methodologies about control charts for attributes and variables using Birnbaum-Saunders distributions; (ii) to develop multivariate control charts based on these distributions and the Hotelling statistic; (iii) to present a methodology based on capability indices for Birnbaum-Saunders processes; (iv) to derive acceptance sampling plans when the lifetimes of the test units follow the Birnbaum-Saunders distribution; and (v) to use these distributions to analyze reliability in the hardness of materials when incorporating nanotechnologies.

The remainder of the work is organized as follows. Sections 2 and 3 present a background on univariate and multivariate Birnbaum-Saunders distributions. In Sects. 4 and 5, we introduce methodologies about control charts for attribute and variables, respectively, using Birnbaum-Saunders distributions. In Sect. 6, multivariate quality control charts based on such distributions and the Hotelling statistic are developed. Section 7 presents, discusses and applies a methodology based on capability indices for Birnbaum-Saunders processes. In Sect. 8, we provide

acceptance sampling plans when the lifetimes of the test units follow a Birnbaum-Saunders distribution. In Sect. 9, the use of the Birnbaum-Saunders distribution to analyze the effect on hardness when incorporating nanotechnology is presented and illustrated with real data. Section 10 provides the main results of the work and a guidance on possible future research.

2 Univariate Birnbaum-Saunders Distributions

In this section, univariate Birnbaum-Saunders distributions are presented, along with their features, properties, logarithmic version and the sum of random variables following the Birnbaum-Saunders distribution.

2.1 Features and Properties

The cumulative distribution function (CDF) of the univariate Birnbaum-Saunders distribution with shape parameter $\alpha \in \mathbb{R}_+$ and scale parameter $\beta \in \mathbb{R}_+$ is given by

$$F_T(t;\alpha,\beta) = \Phi(A(t;\alpha,\beta)), \quad t \in \mathbb{R}_+, \tag{1}$$

where Φ is a standard Gaussian CDF and

$$A(t;\alpha,\beta) = \frac{1}{\alpha}\left[\left(\frac{t}{\beta}\right)^{\frac{1}{2}} - \left(\frac{\beta}{t}\right)^{\frac{1}{2}}\right].$$

The notation $T \sim \mathrm{BS}(\alpha,\beta)$ is used for a random variable or quality characteristic T with the CDF defined in (1). Such a random variable is a transformation of a random variable with standard Gaussian distribution. In fact, $T \sim \mathrm{BS}(\alpha,\beta)$ can be written as

$$T = T(V;\alpha,\beta) = \beta\left\{\frac{\alpha V}{2} + \left[\left(\frac{\alpha V}{2}\right)^2 + 1\right]^{\frac{1}{2}}\right\}^2, \tag{2}$$

where V is a random variable with standard Gaussian distribution. Then,

$$V = \frac{1}{\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \quad \sim \quad \mathrm{N}(0,1).$$

Thus, the probability density function (PDF) of $T \sim \mathrm{BS}(\alpha,\beta)$ is expressed as

$$f_T(t;\alpha,\beta) = \phi(A(t;\alpha,\beta))\, a(t;\alpha,\beta), \quad t \in \mathbb{R}_+, \tag{3}$$

where ϕ is the standard Gaussian PDF and $a(t;\alpha,\beta)$, the derivative of $A(t;\alpha,\beta)$ defined below (1), is given by

$$a(t;\alpha,\beta)=\frac{1}{2\alpha\beta}\left[\left(\frac{\beta}{t}\right)^{\frac{1}{2}}+\left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right]. \tag{4}$$

Let $T\sim \mathrm{BS}(\alpha,\beta)$. Then, we have the following three properties:

(P1) $k\,T\sim \mathrm{BS}(\alpha,k\,\beta)$, with $k\in\mathbb{R}_+$;
(P2) $1/T\sim \mathrm{BS}(\alpha,1/\beta)$;
(P3) $V^2=(T/\beta+\beta/T-2)/\alpha^2\sim\chi^2(1)$, that is, V^2 follows a χ^2 distribution with one degree of freedom.

Maximum likelihood (ML) estimators for the Birnbaum-Saunders model parameters are unique and can be easily obtained, solving numerically the ML equations. For details about this ML estimation, see [3, 10].

2.2 Log-Birnbaum-Saunders Distribution

If T follows a Birnbaum-Saunders distribution as defined in (2), then its logarithm, $Y=\log(T)$, follows a log-Birnbaum-Saunders distribution. Specifically, if $T\sim \mathrm{BS}(\alpha,\beta)$, then $Y=\log(T)\sim \text{log-BS}(\alpha,\mu_Y)$, where $\mu_Y=\log(\beta)$. Thus, the corresponding PDF of Y is expressed as

$$f_Y(y;\alpha,\mu_Y)=\phi(B(y;\alpha,\mu_Y))\,b(y;\alpha,\mu_Y),\quad y\in\mathbb{R}, \tag{5}$$

with shape parameter $\alpha\in\mathbb{R}_+$ and mean $\mu_Y\in\mathbb{R}$, where $b(y;\alpha,\mu_Y)$ is expressed as

$$b(y;\alpha,\mu_Y)=\frac{1}{\alpha}\cosh\left(\frac{y-\mu_Y}{2}\right),$$

which is the derivative of

$$B(y;\alpha,\mu_Y)=\frac{2}{\alpha}\sinh\left(\frac{y-\mu_Y}{2}\right).$$

Let $Y\sim \text{log-BS}(\alpha,\mu_Y)$. Then, we have the following three properties:

(P4) $Y=\mu_Y+2\,\mathrm{arcsinh}(\alpha\,W/2)\sim \text{log-BS}(\alpha,\mu_Y)$, with $W\sim \mathrm{N}(0,1)$, that is, a random variable with log-Birnbaum-Saunders distribution can be obtained directly from a random variable with standard Gaussian distribution;
(P5) $W=B(Y;\alpha,\mu_Y)=(2/\alpha)\sinh((Y-\mu_Y)/2)\sim \mathrm{N}(0,1)$;
(P6) $W^2=B^2(Y;\alpha,\mu_Y)\sim\chi^2(1)$.

2.3 The BSsum Distribution

Two approaches have been proposed to obtain the distribution of the sum of random variables with Birnbaum-Saunders distribution. The first approach is due to Raaijmakers [22, 23], whereas the second one is due to Fox et al. [5].

Approach 1

Let $T_i \sim \text{BS}(\alpha, \beta = 1)$, for $i = 1, \ldots, k$. Then, according to Raaijmakers [22, 23], the PDF and CDF of the sum of k random variables with Birnbaum-Saunders distribution, $S = \sum_{i=1}^{k} T_i$, are given respectively by

$$f_S(s; \alpha, \beta = 1, k) = \frac{a}{2^k} \exp(2ka(1-s)) \sum_{i=0}^{k} \binom{k}{i} (2ka)^{i-2} \varsigma_i \left(\frac{s}{4k^2 a}\right),$$

$$F_S(s; \alpha, \beta = 1, k) = \frac{1}{2^k} \exp(2ka(1-s)) \sum_{i=2}^{k} \frac{l_i \, \varsigma_i(s/(4k^2 a))}{(2ka)^{2-i}} + \Phi\left(\frac{\varphi(s;k)}{\alpha}\right),$$

where $s > 0$, $\alpha > 0$, $k = 1, 2, \ldots$, $a = 1/(2\alpha^2)$, $l_i = l_{i+2} - \binom{k}{i}$, with $l_{k+2} = l_{k+1} = 0$, for $i = 2, \ldots, k$, $\varphi(s;k) = \sqrt{s} - k/\sqrt{s}$, and ς_i is a function as defined in Raaijmakers [22, 23]. The notation $S \sim \text{BSsum}(\alpha, \beta, k)$ is used in this case. The quantile function of S is given by

$$Q_S(q; \alpha, \beta, k) = F_S^{-1}(q; \alpha, \beta, k), \quad 0 < q < 1, \tag{6}$$

which must be solved numerically.

Approach 2

The exact expression for the PDF of the sum of random variables with Birnbaum-Saunders distribution was proposed by Fox et al. [5] for modeling count data. Let T_i, for $i = 1, \ldots, n$, be independent random variables following Birnbaum-Saunders distributions with parameters Δ_{T_i}, μ_{T_i} and σ_{T_i}. Fox et al. [5] derived the distribution of $S = \sum_{i=1}^{n} T_i$ assuming that a positive constant v exists, such that $\sigma_{T_i}/\mu_{T_i} = v$, for all i. Then, the random variable S corresponds to a mixture distribution with PDF given by

$$f_S(s; \Delta_S, \mu_S, \sigma_S, n) = \frac{1}{2^n} f_0(s) + \sum_{j=1}^{n} \frac{1}{2^n} \binom{n}{j} f_j(s),$$

where

$$f_j(s) = \frac{\Delta_S \mu_S^j}{\sqrt{2\pi} 2^{j/2} \sigma_S^{j+1}} \exp\left(-\frac{1}{2}\left(\frac{\Delta_S - s\mu_S}{\sqrt{s}\sigma_S}\right)^2\right) s^{j/2-3/2} K_2\left(\frac{j}{2}, \frac{3}{2}, \frac{{\Delta_S}^2}{2s\sigma_S^2}\right), s > 0,$$

with $K_2(0, 3/2, \Delta_S{}^2/(2s\sigma_S^2)) = 1$ for $j = 0$, $K_2(1/2, 3/2, \Delta_S{}^2/(2s\sigma_S^2)) = \sqrt{2s}\sigma_S/\Delta_S$ for $j = 1$, $\mu_S = \sqrt{\sum_{i=1}^n \mu_{T_i}}$, $\sigma_S = \sqrt{\sum_{i=1}^n \sigma_{T_i}}$ and $\Delta_S = \sum_{i=1}^n \Delta_{T_i}/\mu_{T_i}$.

3 Multivariate Birnbaum-Saunders Distributions

In this section, multivariate Birnbaum-Saunders distributions are presented, along with their features, properties, logarithmic version, associated Mahalanobis distance and random number generator.

3.1 Features and Properties

The random vector $\underline{T} = (T_1, \ldots, T_m)^\top \in \mathbb{R}_+^m$ follows an m-variate Birnbaum-Saunders distribution with parameters $\underline{\alpha} = (\alpha_1, \ldots, \alpha_m)^\top \in \mathbb{R}_+^m$, $\underline{\beta} = (\beta_1, \ldots, \beta_m)^\top \in \mathbb{R}_+^m$ and scale matrix $\boldsymbol{\Sigma} = (\sigma_{kl}) \in \mathbb{R}^{m\times m}$, if $T_i = T(V_i; \alpha_i, \beta_i)$, for $i = 1, \ldots, m$, where T is given in (2) and $\underline{V} = (V_1, \ldots, V_m)^\top \in \mathbb{R}^m \sim \mathrm{N}_m(\underline{0}_{m\times 1}, \boldsymbol{\Gamma})$, with $\boldsymbol{\Gamma} = (\rho_{kl}) \in \mathbb{R}^{m\times m}$ being a correlation matrix. Furthermore, the Birnbaum-Saunders case implies the diagonal elements of $\boldsymbol{\Sigma}$ are equal to one, that is, $\sigma_{kk} = 1$, for all $k = 1, \ldots, m$. Then

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{12} & 1 & \cdots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1m} & \rho_{2m} & \cdots & 1 \end{pmatrix} = \boldsymbol{\Gamma}. \tag{7}$$

The m-variate Birnbaum-Saunders distribution is denoted by $\underline{T} \sim \mathrm{BS}_m(\underline{\alpha}, \underline{\beta}, \boldsymbol{\Gamma})$. The CDF and PDF of $\underline{T}$ are, respectively, defined as

$$F_{\underline{T}}(\underline{t}; \underline{\alpha}, \underline{\beta}, \boldsymbol{\Gamma}) = \Phi_m(\underline{A}; \boldsymbol{\Gamma}),$$

$$f_{\underline{T}}(\underline{t}; \underline{\alpha}, \underline{\beta}, \boldsymbol{\Gamma}) = \phi_m(\underline{A}; \boldsymbol{\Gamma})\, a(\underline{t}; \underline{\alpha}, \underline{\beta}), \quad \underline{t} = (t_1, \ldots, t_m)^\top \in \mathbb{R}_+^m,$$

where Φ_m and ϕ_m are the m-variate standard Gaussian CDF and PDF, respectively $\underline{A} = A(\underline{t}; \underline{\alpha}, \underline{\beta}) = (A_1, \ldots, A_m)^\top$, with $A_j = A(t_j; \alpha_j, \beta_j)$,

$$a(\underline{t}; \underline{\alpha}, \underline{\beta}) = \prod_{j=1}^m a(t_j; \alpha_j, \beta_j),$$

and both $A(t_j;\alpha_j,\beta_j)$ and $a(t_j;\alpha_j,\beta_j)$ are as given in (3). Let $\underline{T} \sim \mathrm{BS}_m(\underline{\alpha},\underline{\beta},\boldsymbol{\Gamma})$. Then, three properties of the m-variate Birnbaum-Saunders distribution are the following:

(P7) $k\,\underline{T} \sim \mathrm{BS}_m(\underline{\alpha}, k\,\underline{\beta}, \boldsymbol{\Gamma})$, with $k \in \mathbb{R}_+$;

(P8) $\underline{T}^* = (1/T_1,\ldots,1/T_m)^\top \sim \mathrm{BS}_m(\underline{\alpha},\underline{\beta}^*,\boldsymbol{\Gamma})$, with $\underline{\beta}^* = (1/\beta_1,\ldots,1/\beta_m)^\top$;

(P9) $A^\top(\underline{T};\underline{\alpha},\underline{\beta})\boldsymbol{\Gamma}^{-1}A(\underline{T};\underline{\alpha},\underline{\beta}) \sim \chi^2(m)$.

Unlike parameter estimation for the univariate Birnbaum-Saunders distribution, where uniqueness is guaranteed, (see [3]) in the multivariate case there is no certainty that the system of ML equations has a unique solution. Therefore, care must be taken to ensure that numerical procedures yield a global maximum.

3.2 Log-Birnbaum-Saunders Distribution

Let $\underline{T} = (T_1,\ldots,T_m)^\top \sim \mathrm{BS}_m(\underline{\alpha},\underline{\beta},\boldsymbol{\Gamma})$. Then, $\underline{Y} = (Y_1,\ldots,Y_m)^\top = (\log(T_1),\ldots,\log(T_m))^\top$ follows an m-variate log-Birnbaum-Saunders distribution with shape parameters $\underline{\alpha} = (\alpha_1,\ldots,\alpha_m)^\top$, mean vector $\underline{\mu}_Y = \mathrm{E}[\underline{Y}] = (\mathrm{E}[Y_1],\ldots,\mathrm{E}[Y_m])^\top = (\log(\beta_1),\ldots,\log(\beta_m))^\top \in \mathbb{R}^m$, and correlation matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{m\times m}$ given in (7). This is denoted by $\underline{Y} \sim \text{log-BS}_m(\underline{\alpha},\underline{\mu}_Y,\boldsymbol{\Gamma})$. The CDF of $\underline{Y}$ is defined as

$$F_{\underline{Y}}(\underline{y};\underline{\alpha},\underline{\mu}_Y,\boldsymbol{\Gamma}) = \Phi_m(\underline{B};\boldsymbol{\Gamma}), \quad \underline{y} = (y_1,\ldots,y_m)^\top \in \mathbb{R}^m,$$

where $\underline{B} = B(\underline{y};\underline{\alpha},\underline{\mu}_Y) = (B_1,\ldots,B_m)^\top$, with $B_j = B(y_j;\alpha_j,\mu_{Y_j})$, for $j = 1,\ldots,m$, as given in (5). The PDF of $\underline{Y}$ is expressed as

$$f_{\underline{Y}}(\underline{y};\underline{\alpha},\underline{\mu}_Y,\boldsymbol{\Gamma}) = \phi_m(\underline{B};\boldsymbol{\Gamma})b(\underline{y};\underline{\alpha},\underline{\mu}_Y), \quad \underline{y} \in \mathbb{R}^m,$$

where $b(\underline{y};\underline{\alpha},\underline{\mu}_Y) = \prod_{j=1}^m b(y_i;\alpha_j,\mu_{Y_j})$, with $b(y_j;\alpha_j,\mu_{Y_j})$ as given in (5), for $j = 1,\ldots,m$. If $\underline{Y} \sim \log-\mathrm{BS}_m(\underline{\alpha},\underline{\mu}_Y,\boldsymbol{\Gamma})$, then from (P5) and (7), the following two properties hold:

(P10) $\boldsymbol{D}(\underline{\alpha})\,B(\underline{Y};\underline{\alpha},\underline{\mu}_Y) \sim \mathrm{N}_m(\underline{0}_{m\times 1}, \boldsymbol{D}(\underline{\alpha})\boldsymbol{\Gamma}\boldsymbol{D}(\underline{\alpha}))$, where $\boldsymbol{D}(\underline{\alpha}) = \mathrm{diag}(\alpha_1,\ldots,\alpha_m)$
and

$$\boldsymbol{D}(\underline{\alpha})\boldsymbol{\Gamma}\boldsymbol{D}(\underline{\alpha}) = \begin{pmatrix} \alpha_1^2 & \alpha_1\alpha_2\rho_{12} & \cdots & \alpha_1\alpha_m\rho_{1m} \\ \alpha_1\alpha_2\rho_{12} & \alpha_2^2 & \cdots & \alpha_2\alpha_m\rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1\alpha_m\rho_{1m} & \alpha_2\alpha_m\rho_{2m} & \cdots & \alpha_m^2 \end{pmatrix}.$$

(P11) $(B(\underline{Y};\underline{\alpha},\underline{\mu}_Y))^\top\boldsymbol{\Gamma}^{-1}B(\underline{Y};\underline{\alpha},\underline{\mu}_Y) \sim \chi^2(m)$, that is, a χ^2 distribution with m degrees of freedom.

3.3 Mahalanobis Distance and Number Generation

The Mahalanobis distance for observation i, using Property (P10), is given by

$$\mathrm{MD}_i(\underline{\theta}) = (B(\underline{Y}_i; \underline{\alpha}, \underline{\mu}_Y))^\top \boldsymbol{\Gamma}^{-1} B(\underline{Y}_i; \underline{\alpha}, \underline{\mu}_Y), \quad i = 1, \ldots, n,$$

with $\underline{\theta} = (\underline{\alpha}^\top, \underline{\mu}_Y^\top, \mathrm{svec}(\boldsymbol{\Gamma})^\top)^\top$, where 'svec' denotes vectorization of a symmetric matrix. The Mahalanobis distance is used to identify multivariate outliers and to assess the goodness of fit in multivariate log-Birnbaum-Saunders distributions. Random vectors from multivariate log-Birnbaum-Saunders distributions can be generated using Algorithm 1; see [15] for details of the generation of numbers following Birnbaum-Saunders and log-Birnbaum-Saunders distributions.

Algorithm 1 Generator of random vectors from multivariate log-Birnbaum-Saunders distributions

1: Make a Cholesky decomposition of $\boldsymbol{\Gamma}$ as $\boldsymbol{\Gamma} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is a lower triangular matrix with real and positive diagonal entries.
2: Generate m independent standard Gaussian random numbers $\underline{W} = (W_1, \ldots, W_m)^\top$.
3: Compute $\underline{Z} = (Z_1, \ldots, Z_m)^\top = \boldsymbol{L}\underline{W}$.
4: Obtain the vector $\underline{Y}$ with components $Y_j = \mu_{Y_j} + 2\arcsin(\alpha_j Z_j/2)$ for $j = 1, \ldots, m$.
5: Repeat Steps 1 to 4 until the required vector of data is generated.

4 Control Charts for Attributes

In this section, a criterion for monitoring production processes based on an attribute control chart for the number of defective items is designed and discussed. The defective aspect to be monitored is evaluated by a quality variable which follows a Birnbaum-Saunders distribution. The control coefficient and the minimum inspection level for the designed criterion are determined to yield the specified in-control average run length (ARL), while the out-of-control case is obtained according to a shift in the target mean. This criterion is implemented in the `R` software. An application with real-world data is carried out; see details in [12].

4.1 Formulation

An np-chart is an adaptation of the control chart for non-conforming fraction when samples of equal size (n) are taken from the process. The np-chart is based on the binomial distribution as detailed below. In quality monitoring processes, one could be concerned about the random variable corresponding to the number (D) of times that the quality variable (T) exceeds a fixed value (t) established for the process,

given an exceedance probability (p). Here, p can be computed by means of the continuous distribution of the quality variable T as $p = \mathrm{P}(T > t) = 1 - F_T(t)$, where F_T is the CDF of T. Thus, D follows a binomial distribution with parameters n and p and then

$$\mathrm{P}(D = d) = \binom{n}{d} p^d (1-p)^{n-d}, \quad d = 0, 1, \ldots, n, \tag{8}$$

where $\mathrm{E}[D] = np$ and $\mathrm{Var}[D] = np(1-p)$. Based on (8), an np-chart is proposed with lower control limit (LCL), central line (CL) and upper control limit (UCL) given by

$$\mathrm{LCL} = \max\left\{0, np_0 - k\sqrt{np_0(1-p_0)}\right\}, \ \mathrm{CL} = np_0, \ \mathrm{UCL} = np_0 + k\sqrt{np_0(1-p_0)}, \tag{9}$$

where k is a control coefficient such that $k = 2$ indicates a warning level and $k = 3$ a dangerous level, p_0 is the non-conforming fraction corresponding to a target mean μ_{T_0} of the quality variable T, when the process is in-control, and n is the size of the subgroups. Note that the non-conforming fraction is the probability that the random variable exceeds a value (t_0) and, therefore, this probability is $\mathrm{P}(T > t_0) = 1 - F_T(t_0)$. The Birnbaum-Saunders distribution can be reparametrized from (α, β) to (α, μ_T), switching from the median β in the original Birnbaum-Saunders parametrization to its mean given by $\mu_T = \beta(2 + \alpha^2)/2$. Consider t_0 as proportional to μ_{T_0}, that is, $t_0 = a\mu_{T_0}$, relating them to establish the monitoring criterion, where $a > 0$ is a proportionality constant. Note that the target mean μ_{T_0} and dangerous level t_0 can be taken from process specifications. Then, the Birnbaum-Saunders CDF can be reparametrized in terms of its mean and expressed in function of t_0 and a as

$$F_T(t_0; \alpha, a, \mu_T) = \Phi\left(\frac{1}{\alpha}\xi\left(\frac{a(1+\alpha^2/2)}{\mu_T/\mu_{T_0}}\right)\right). \tag{10}$$

Note that $\xi(y) = y^{1/2} - 1/y^{1/2}$. Thus, when a monitoring process is in-control ($\mu_T = \mu_{T_0}$) for a quality variable following a Birnbaum-Saunders distribution, from (10), the non-conforming fraction is given by

$$p_0 = 1 - F_T(t_0; \alpha, a) = \Phi\left(-\frac{1}{\alpha}\xi\left(a(1+\alpha^2/2)\right)\right). \tag{11}$$

Note that the specification of the point t_0 is equivalent to specifying the inspection point $a > 0$, because $t_0 = a\mu_{T_0}$, where μ_{T_0} is the target mean, which is assumed to be known. Algorithm 2 provides a criterion for monitoring processes using an np-chart for a quality variable $T \sim \mathrm{BS}(\alpha, \mu_T)$.

Consider a shift in the process mean and assume that the new (shifted) process mean becomes $\mu_{T_1} = a\mu_{T_0}$, for a shift constant $a > 0$. Assume that the value of

Algorithm 2 np control chart based on the Birnbaum-Saunders distribution

1: Take N subgroups of size n.
2: Collect n data $t_1, \ldots, t_n$ of the random variable of interest T for each subgroup.
3: Fix the target mean μ_{T_0}, the inspection constant a and the control coefficient k.
4: Count in each subgroup of n data the number d of times that t_i exceeds $t_0 = a\mu_{T_0}$, for $i = 1, \ldots, n$.
5: Compute $\text{LCL} = \max\{0, n\widehat{p}_0 - k\sqrt{n\widehat{p}_0(1-\widehat{p}_0)}\}$ and $\text{UCL} = n\widehat{p}_0 + k\sqrt{n\widehat{p}_0(1-\widehat{p}_0)}$, which are obtained from (9), where $\widehat{p}_0 = \Phi(-(1/\widehat{\alpha})\xi(a(1+\widehat{\alpha}^2/2)))$ is given as in (11), with $\widehat{\alpha} = \sqrt{2(\sqrt{s/r}-1)}$ being the modified moment estimate of α (see [10]), for $s = (1/n)\sum_{i=1}^{n} t_i$ and $r = \left[(1/n)\sum_{i=1}^{n}(1/t_i)\right]^{-1}$.
6: Declare the process as out-of-control if $d \geq \text{UCL}$ or $d \leq \text{LCL}$, or as in-control if $\text{LCL} \leq d \leq \text{UCL}$.

the shift constant a is greater than one, because of the interest in the case where the mean of the quality variable to be monitored becomes greater than the target mean. Note that μ_{T_0} and μ_{T_1} are different means corresponding to in-control and out-of-control processes, respectively, but both of them are means of the Birnbaum-Saunders distribution. Hence, the non-conforming fraction corresponding to the new mean of the random variable T is obtained from (10) as

$$p_1 = 1 - F_T(t_0; \alpha, a) = \Phi\left(-\frac{1}{\alpha}\xi\left(\frac{a}{\alpha}\left(1+\alpha^2/2\right)\right)\right). \tag{12}$$

As mentioned, in general, a process is said in-control if $\text{LCL} \leq D \leq \text{UCL}$. Thus, when the process is actually in-control, the probability to be in-control is given by

$$P_{\text{in}}^0 = \text{P}(\text{LCL} \leq D \leq \text{UCL}|p_0) = \sum_{d=\lfloor \text{LCL} \rfloor + 1}^{\lfloor \text{UCL} \rfloor} \binom{n}{d} p_0^d (1-p_0)^{n-d}, \tag{13}$$

whereas, if the process mean has shifted to the new mean μ_{T_1}, the probability given in (13) is expressed as

$$P_{\text{in}}^1 = \text{P}(\text{LCL} \leq D \leq \text{UCL}|p_1) = \sum_{d=\lfloor \text{LCL} \rfloor + 1}^{\lfloor \text{UCL} \rfloor} \binom{n}{d} p_1^d (1-p_1)^{n-d}, \tag{14}$$

where $\lfloor x \rfloor$ denotes the integer part of the number x. Efficiency of the proposed criterion can be evaluated by using ARL_0 and ARL_1. ARL_0 is defined as the expected number of observations taken from an in-control state until the chart falsely signals an out-of-control case. ARL_0 is regarded as acceptable if it is large enough to keep the level of false alarms at a reasonable value. ARL_1 is defined as the expected number of observations taken from an out-of-control state until the chart correctly

signals an out-of-control. The ARL_1 value should be as small as possible. In-control and out-of-control ARLs are respectively given by

$$\mathrm{ARL}_0 = \frac{1}{1 - P_{\text{in}}^0}, \quad \mathrm{ARL}_1 = \frac{1}{1 - P_{\text{in}}^1},$$

where P_{in}^0 and P_{in}^1 are given in (13) and (14), respectively. The control coefficient k is selected such that ARL_0 is close to a specified ARL denoted by r_0. Then, with the selected value of k, it is possible to obtain the value of ARL_1 for the shift constant a given in (12).

4.2 Illustration

Fuchs and Kenett [6] considered measurements (in centimeters, cm) of six quantities related to aluminum pins carved from aluminum blocks using a numerically controlled machine. The variables in the data set are: diameter in three different locations on the main part of the pin (T_1, T_2 and T_3), diameter at the cap (T_4), length without cap (T_5) and length with cap (T_6). In this illustration, only the variable T_6 is considered and the target value for this variable is set at 60.05; see [6]. Table 1 provides descriptive statistics for T_6, including central tendency statistics, standard deviation (SD) and coefficients of variation (CV), skewness (CS) and kurtosis (CK). Figure 1a shows a histogram and an adjusted boxplot for asymmetric data; see [24]. Table 1 and Fig. 1a show a distribution with positive skewness, moderate kurtosis and some univariate atypical data, suggesting that the Birnbaum-Saunders distribution is appropriate for these data.

The assumption that the data follow a Birnbaum-Saunders distribution is supported by the quantile-quantile (QQ) plot with its envelope shown in Fig. 1b. The application of the Kolmogorov-Smirnov (KS) goodness-of-fit (GOF) test provides a p-value $= 0.2091$, from which is possible to conclude that the Birnbaum-Saunders distribution is a good model for these data at 5% of significance.

The criterion based on the Birnbaum-Saunders np-chart, proposed in Sect. 4.1, is now applied. Specifically, this chart is used to monitor the number of non-conforming length measurements in N subgroups of size n: in this case they are $N = 14$ and $n = 5$. The control chart is constructed following Algorithm 2. Figure 2 shows the Birnbaum-Saunders np-chart for the data under analysis. In this figure,

Table 1 Summary statistics for length with cap (T_6) data

n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
70	59.910	60.020	60.027	60.150	0.047	7.90%	0.568	0.734

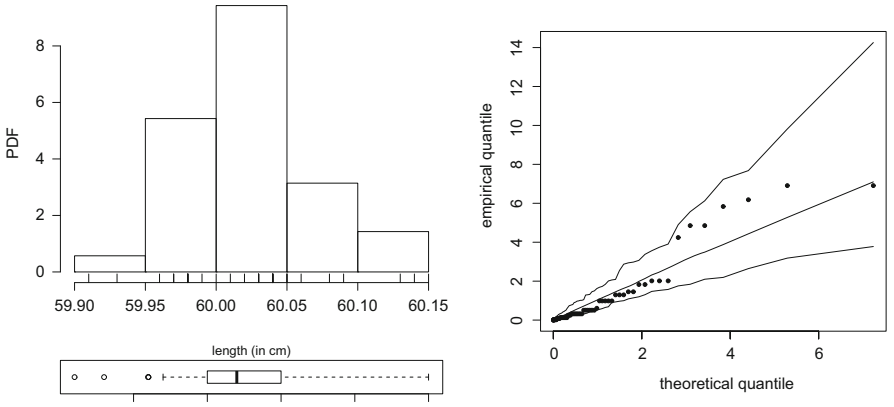


Fig. 1 Histogram with adjusted boxplot (**a**) and QQ plot with its envelope (**b**) for length with cap (T_6) data

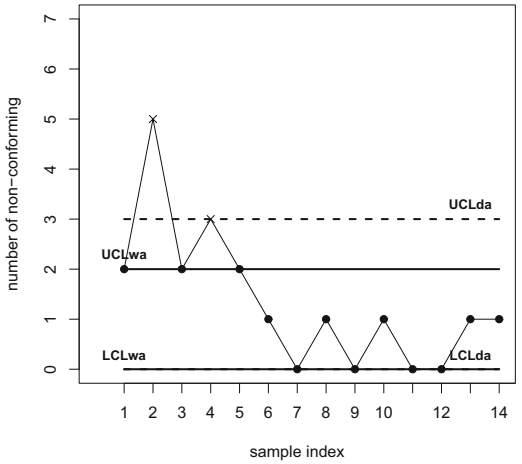


Fig. 2 Birnbaum-Saunders np chart with $t_0 = 60.05$, where index "wa" denotes warning ($k = 2$, solid line) and index "da" dangerous ($k = 3$, dashed line) situations

the limits with $k = 2$ and $k = 3$ are plotted, indicating warning ($\mathrm{UCL_{wa}}$) and dangerous ($\mathrm{UCL_{da}}$) levels, respectively. From Fig. 2, note that subgroups 2 and 4 are above the $\mathrm{UCL_{wa}}$ and also subgroup 2 is above the $\mathrm{UCL_{da}}$, indicating that the length measurement exceeds the value t_0 three and two times, respectively.

5 Control Charts for Variables

In this section, X-bar control charts are presented in the context of Birnbaum-Saunders models, that is, the quality characteristic has a Birnbaum-Saunders distribution, which in this section is denoted by X instead of T, as used in others sections, to be coherent with the X-bar concept; see details in [25]. In Sect. 2.3, two approaches are described to obtain the distribution of the sum of Birnbaum-Saunders random variables; those results are now useful to derive X-bar control charts.

5.1 Formulation

X-bar control charts based on the Birnbaum-Saunders and BSsum distributions have been proposed by Saulo et al. [25]. These charts are used to monitor the process average or mean quality level when the quality characteristic follows a Birnbaum-Saunders distribution. Particularly, LCL and UCL are obtained by using the BSsum quantile function defined in (6). The steps presented in [25] to construct a Birnbaum-Saunders X-bar control charts are summarized in Algorithm 3.

5.2 Illustration

The Birnbaum-Saunders X-bar control chart is now used to analyze a process for producing 6061-T6 aluminum coupons at the Boeing Aircraft Company; see [17]. They used simulated data from a Birnbaum-Saunders distribution with parameters fixed at the estimates based on a data set corresponding breaking strength (in hundreds of thousands of cycles) of 6061-T6 aluminum coupons. The aim is to

Algorithm 3 Birnbaum-Saunders X-bar control chart

1: Collect m samples, each containing n observations, on the quality characteristic following a Birnbaum-Saunders distribution.
2: Compute the mean of each sample $\overline{x}_1, \ldots, \overline{x}_m$, where $\overline{x}_i = (x_{i1} + \cdots + x_{in})/n$, for $i = 1, \ldots, m$.
3: Estimate the parameters of the BSsum distribution α and β based on the sample $s_1, \ldots, s_m$, where $s_i = x_{i1} + \cdots + x_{in}$, for $i = 1, \ldots, m$.
4: Set the value for γ corresponding to the desired false alarm rate (FAR) –out-of-control probability– when the process is actually under control.
5: Compute the LCL and UCL as the $(\gamma/2) \times 100$-th and $(1 - \gamma/2) \times 100$-th quantiles of the estimated BSsum quantile function divided by n, with parameters fixed at the estimates obtained in Step 2. Add a CL to the chart by using the median of the data as reference.
6: Insert the points $\overline{x}_1, \ldots, \overline{x}_m$ to the control chart and if $\overline{x}_i \geq$ UCL and/or $\overline{x}_i \leq$ LCL, for $i = 1, \ldots, m$, and/or the chart exhibits non-random pattern, then declare the process as out of control, otherwise it is under control.

Table 2 Summary statistics for breaking strength data

m	$N = m \times n$	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
1–20 (in-control)	100	0.738	1.402	1.443	2.625	0.361	25.04%	0.655	0.597
21–40 (out-of-control)	100	0.219	1.256	1.804	8.532	1.599	88.66%	1.808	3.330

mimic a process for producing these coupons. The original data set was transformed into units of hundreds of thousands of cycles to failure and its sample size is equal to $N = 100 = m \times n$; see [3]. Initially, [17] generated $m = 20$ in-control samples each containing $n = 5$ observations with Birnbaum-Saunders parameters $\alpha = 0.2795$ and $\beta = 1.358$. Then, they assumed that the process had shifted to an out-of-control state by generating additional $m = 20$ samples of size $n = 5$, setting a different shape parameter, that is, $\alpha = 0.8782$. Table 2 provides descriptive statistics for the in-control and out-of-control process data sets, including central tendency statistics, SD, CV, CS and CK. The histograms with adjusted boxplots for asymmetric data in Fig. 3 confirm the skewness noticed in Table 2. The QQ plots with simulated enveloped shown in Fig. 3 confirm the good agreement between the Birnbaum-Saunders distribution and the data.

Figure 4 displays the Birnbaum-Saunders X-bar control chart, constructed according to Algorithm 3, to monitor the average breaking strength of the aluminum material. The control limits are based on 20 sample average values from the in-control process. In addition, 20 additional sample mean values from the out-of-control process are plotted on the control chart. From Fig. 4, note that several points outside the control limits are detected, suggesting that the process is out-of-control, as expected.

6 Multivariate Control Charts

In this section, multivariate control charts for monitoring the quality of a process are presented. Multivariate monitoring is carried out taking into account correlated quality characteristics and simultaneously determining whether these characteristics are in control or out of control. A methodology using multivariate quality control charts for subgroups based on Birnbaum-Saunders distributions and an adapted Hotelling statistic is introduced and applied. The corresponding parameters are estimated with the ML method and the parametric bootstrapping is used to obtain the distribution of the adapted Hotelling statistic. Furthermore, the Mahalanobis distance is considered to detect multivariate outliers and used to assess the adequacy of the distributional assumption. The methodology is implemented in the `R` software. An illustration is given with real-world data; see details in [18, 19].

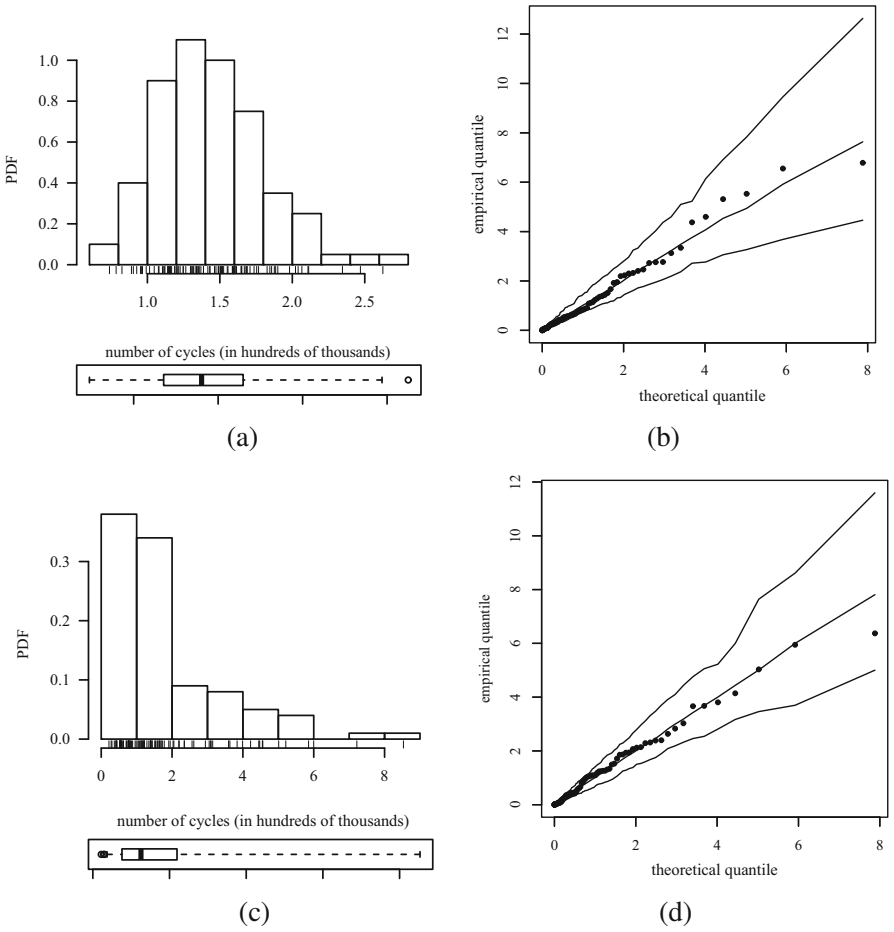


Fig. 3 Histograms with adjusted boxplots for in-control (**a**) and out-of-control (**c**) processes; and QQ plots with their envelopes for in-control (**b**) and out-of-control (**d**) processes with breaking strength data

6.1 Formulation

Suppose that one is interested in modeling a dynamic process with p quality characteristics and that for each there is a sample of n observations from the evolving process. Let $\underline{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top \in \mathbb{R}^p$ denote a random vector associated with measured log-values corresponding to subset i, for $i = 1, \ldots, n$. Assume that $\underline{Y}_i$ follows a p-variate log-Birnbaum-Saunders distribution, that is, $\underline{Y}_i \sim \text{log-BS}_p(\underline{\alpha}, \underline{\mu}_Y, \boldsymbol{\Gamma})$, with the vectors $\underline{Y}_i$ being independent over time and $\underline{\mu}_{Y_0}$ being the mean vector of the in-control process. To confirm that the process is in control it is necessary to test the hypothesis: H_0: $\underline{\mu}_Y = \underline{\mu}_{Y_0} =$

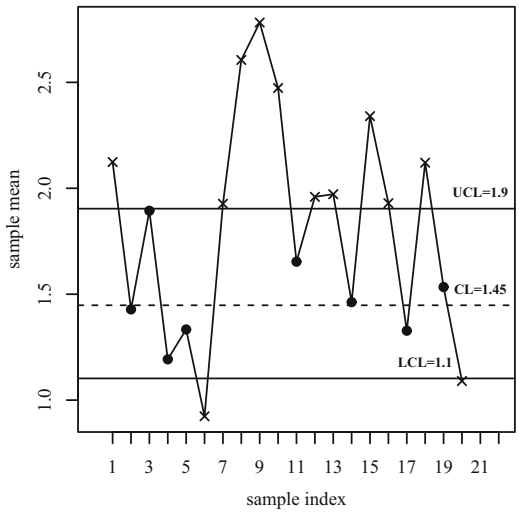


Fig. 4 Birnbaum-Saunders X-bar control chart for the mean with breaking strength data

$(\mu_{Y_{1_0}}, \ldots, \mu_{Y_{p0}})^\top$ versus H_1: $\underline{\mu}_Y \neq \underline{\mu}_{Y_0}$. These hypotheses can be contrasted by using a modified Hotelling T^2 statistic constructed as follows. Using Property (P11) and considering that

$$\underline{b}_i = \left(2\sinh\left(\frac{Y_{i1}-\mu_{1_0}}{2}\right), \ldots, 2\sinh\left(\frac{Y_{ip}-\mu_{p0}}{2}\right)\right)^\top \sim \mathrm{N}_p(\underline{0}_{p\times 1}, \boldsymbol{D}(\underline{\alpha})\boldsymbol{\Gamma}\boldsymbol{D}(\underline{\alpha})),$$

for $i = 1, \ldots, n$, a Hotelling T^2 statistic adapted for multivariate log-Birnbaum-Saunders distributions can be obtained as

$$T^2 = n(n-1)\overline{\underline{b}}^\top \boldsymbol{C}^{-1}\overline{\underline{b}}, \quad \text{with} \quad \overline{\underline{b}} = \sum_{i=1}^{n} \underline{b}_i/n, \quad \text{and} \quad \boldsymbol{C} = \sum_{i=1}^{n} \underline{b}_i \underline{b}_i^\top. \tag{15}$$

Note that, if $\underline{Y} \sim \text{log-BS}_p(\underline{\alpha}, \underline{\mu}_Y, \boldsymbol{\Gamma})$, then T^2 has a Fisher distribution with p and $n-p$ degrees of freedom, that is, $T^2 \sim \mathscr{F}(p, n-p)$; see [9]. Algorithm 1 can be used to construct the bootstrap distribution of the T^2 statistic using random vectors generated from the p-variate log-Birnbaum-Saunders distribution, and then, Algorithm 4 may be used to construct the corresponding control limits. Once constructed, the multivariate Birnbaum-Saunders control chart can be used to identify if the evolving process remains in control. Consider a new vector of values of the quality characteristics, and let T^2_{new} be the corresponding Hotelling statistic calculated using (15). As the process evolves, a sequence of values T^2_{new} is produced. Algorithm 5 details how to construct p-variate control charts based on Birnbaum-Saunders distributions for process monitoring.

Algorithm 4 Computation of multivariate Birnbaum-Saunders control limits

1: Collect k samples $(\underline{y}_{1h}, \ldots, \underline{y}_{nh})^\top$ of size n for an in-control process, with $h = 1, \ldots, k$, assuming that the p-variate vector with the logarithms of the data follows a log-BS$_p(\underline{\alpha}, \underline{\mu}_Y, \boldsymbol{\Gamma})$ distribution.
2: Compute the ML estimates of $\underline{\alpha}$, $\underline{\mu}_Y$ and $\boldsymbol{\Gamma}$ using the data of the pooled sample of size $N = k \times n$ collected in Step 1, and check the distributional assumption using GOF tools.
3: Generate a parametric bootstrap sample $(\underline{y}_1^*, \ldots, \underline{y}_n^*)^\top$ of size n from a p-variate log-BS distribution using the ML estimates obtained in Step 2 as the distribution parameters.
4: Compute T^2 defined in (15) with $(\underline{y}_1^*, \ldots, \underline{y}_n^*)^\top$, which is denoted by T^{2*}, assuming a target $\underline{\mu}_{Y_0}$.
5: Repeat Steps 3 and 4 a sufficiently large number of times (for example, B = 10,000) and obtain B bootstrap statistics of T^2, denoted by $T_1^{2*}, \ldots, T_B^{2*}$.
6: Fix γ as the FAR of the chart.
7: Use the B bootstrap statistics obtained in Step 5 to find the $100(\gamma/2)$-th and $100(1-\gamma/2)$-th quantiles of the distribution of T^2, which are the LCL and UCL for the chart of FAR γ, respectively.

Algorithm 5 Process monitoring using the multivariate Birnbaum-Saunders chart

1: Collect a sample of size n, $\underline{y}_1, \ldots, \underline{y}_n$, from the process.
2: Calculate the T^2_{new} statistic from the sample obtained in Step 1.
3: Declare the process as in control if T^2_{new} falls between LCL and UCL obtained in Algorithm 4; otherwise, the chart indicates an out-of-control state.
4: Repeat Steps 1 to 3 for each sample collected at regular time intervals.

Table 3 Summary statistics for length without cap (T_5) and with cap (T_6) data

Variable	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
T_5	70	49.810	49.910	49.911	50.070	0.044	8.90%	0.907	2.206
T_6	70	59.910	60.020	60.027	60.150	0.047	7.90%	0.568	0.734

6.2 *Illustration*

In Sect. 4.2 a data set, about measurements of aluminum pins carved from aluminum blocks using a numerically controlled machine, is presented and one variable (T_6), out of six, is considered. Now the variables T_5 and T_6 are considered and, for illustrative purposes, their target values are set at 49.91 and 60.05, respectively; see [6]. Table 3 provides descriptive statistics for these variables, including central tendency statistics, SD, CV, CS and CK. Figure 5a, b shows histograms with adjusted boxplots for asymmetric data. Table 3 and Fig. 5a, b show that distributions with positive skewness, different degrees of kurtosis and some univariate atypical data, are appropriate, suggesting the use of Birnbaum-Saunders distributions. In addition, the scatter-plot in Fig. 5c indicates a moderate to high correlation (0.626) between T_5 and T_6, which confirms the need to use a multivariate control chart for monitoring production based on this data set.

The assumption that this data set follows a Birnbaum-Saunders distribution is supported by the probability-probability (PP) plot with KS acceptance regions at 5%

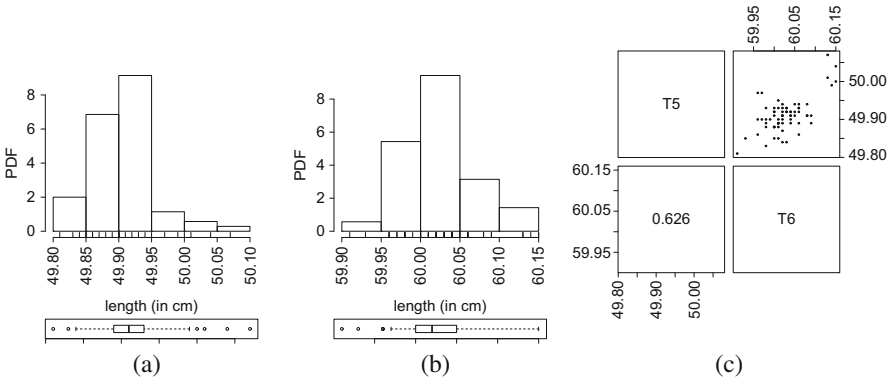


Fig. 5 Histograms with adjusted boxplots for T_5 (**a**) and T_6 (**b**) and their scatter-plot (**c**) for length data

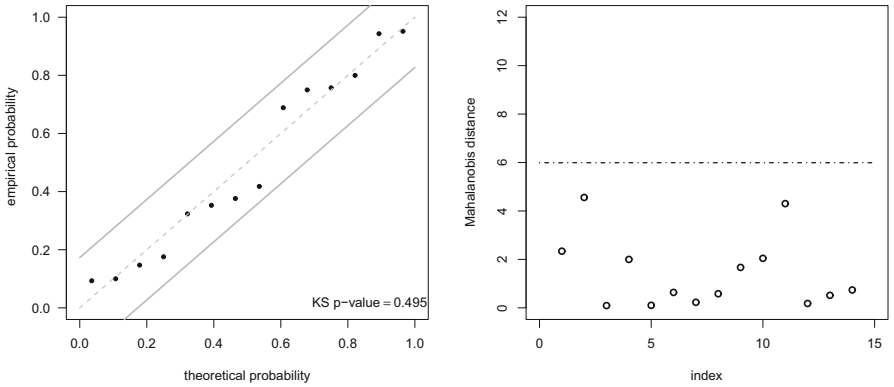


Fig. 6 PP plots with 5% KS acceptance regions for the Mahalanobis distance using the BS_2 distribution (**a**) and Mahalanobis distance index plot (**b**) with length data of T_5 and T_6

shown in Fig. 6a. The modified Mahalanobis distance is used with data transformed by the Wilson-Hilferty approximation to obtain a Gaussian distribution, which can be verified by the GOF methods as described in Step 2 of Algorithm 4. This figure indicates that the BS_2 distribution has a good fit to the data, which is supported by the p-value of 0.495 from the corresponding KS test.

The multivariate Birnbaum-Saunders control chart is used to monitor T_5 and T_6. The data set is partitioned into $m = 14$ new subgroups, each of size $n = 5$, giving 70 observations. The index plot of the modified Mahalanobis distance presented in Fig. 6b does not identify multivariate atypical data in this data set. Using Algorithm 5, a bivariate Birnbaum-Saunders control chart is constructed for monitoring this process with a FAR fixed at $\gamma = 0.027$. Figure 7 displays the

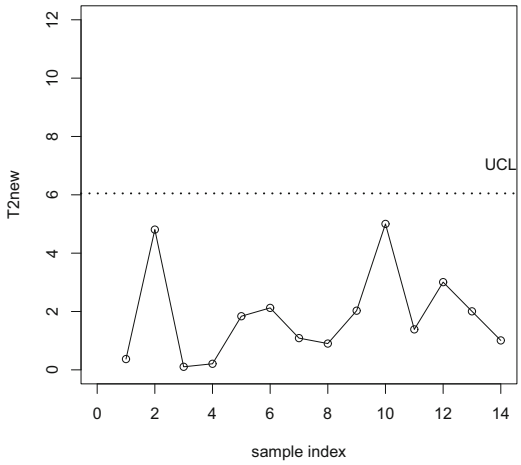


Fig. 7 Bivariate Birnbaum-Saunders control chart for without cap (T_5) and with cap (T_6) data

bivariate Birnbaum-Saunders control chart, which indicates an in-control state since all T^2_{new} of this process fall between LCL and UCL obtained with Algorithm 4.

7 Process Capability Indices

In this section, process capability indices (PCIs) for Birnbaum-Saunders distributions are introduced and a real data set is analyzed to illustrate the methodology; see details in [13]. In general, standard versions of capability indices are employed for processes whose quality characteristics have a Gaussian distribution. However, non-Gaussianity is usually present in productive processes. Therefore, misinterpretation of the process capability can be conducted if non-Gaussianity is ignored, possibly leading to inaccurate business decisions.

7.1 Formulation

A PCI represents the ratio between specification and process statistical variability ranges. This quantity is useful for analyzing the process variability related to product requirements or specifications. When a random variable X has a Gaussian distribution, denoted by $X \sim \text{N}(\mu, \sigma^2)$, the PCIs C_p, C_pl, C_pu and C_pk are defined, respectively, as

$$\text{C}_\text{p} = \frac{\text{USL} - \text{LSL}}{6\sigma}, \text{C}_\text{pl} = \frac{\mu - \text{LSL}}{3\sigma}, \text{C}_\text{pu} = \frac{\text{USL} - \mu}{3\sigma}, \text{C}_\text{pk} = \min\{\text{C}_\text{pl}, \text{C}_\text{pu}\}, \tag{16}$$

where USL is the upper specification limit and LSL is the lower specification limit. Note that 6σ is a range such that the output percentage of the process falling outside the $\mu \pm 3\sigma$ limits is 0.27%. Moreover, C_{pl}, C_{pu} are one-sided PCIs and C_{pk} is the corresponding one-sided PCI for the specification limit closer to the process mean.

Under non-Gaussianity, the PCIs established in (16) are not adequate; see [8]. There have been attempts to modify the PCIs in (16) by substituting 6σ by a range $t(0.99865) - t(0.00135)$, where $t(0.00135)$ and $t(0.99865)$ are the 0.135-th and 99.865-th quantiles of the corresponding non-Gaussian distribution, respectively. Note that the range $t(0.99865) - t(0.00135)$ covers a 99.73% of the distribution of the monitored process data. Then, a general method to calculate non-Gaussian PCIs is defined as

$$C'_p = \frac{\text{USL} - \text{LSL}}{t(q_2) - t(q_1)}, C'_{pl} = \frac{2(t(0.5) - \text{LSL})}{t(q_2) - t(q_1)},$$
$$C'_{pu} = \frac{2(\text{USL} - t(0.5))}{t(q_2) - t(q_1)}, C'_{pk} = \min\{C'_{pl}, C'_{pu}\},$$

where $t(q_i)$ is the $q_i \times 100$-th quantile of the non-Gaussian distribution, which is assumed for the quality characteristic of the process to be monitored, with $i = 1, 2$.

Based on (16) and the Birnbaum-Saunders distribution discussed in Sect. 2, [13] proposed Birnbaum-Saunders PCIs given by

$$C_p^{BS} = \frac{\text{USL} - \text{LSL}}{t(q_2) - t(q_1)} = \frac{\text{USL} - \text{LSL}}{\beta\alpha\left\{z_2\left[\frac{\alpha z_2}{2} + \sqrt{\left(\frac{\alpha z_2}{2}\right)^2 + 1}\right] - z_1\left[\frac{\alpha z_1}{2} + \sqrt{\left(\frac{\alpha z_1}{2}\right)^2 + 1}\right]\right\}},$$

$$C_{pl}^{BS} = \frac{2(\beta - \text{LSL})}{t(q_2) - t(q_1)} = \frac{2(\beta - \text{LSL})}{\beta\alpha\left\{z_2\left[\frac{\alpha z_2}{2} + \sqrt{\left(\frac{\alpha z_2}{2}\right)^2 + 1}\right] - z_1\left[\frac{\alpha z_1}{2} + \sqrt{\left(\frac{\alpha z_1}{2}\right)^2 + 1}\right]\right\}},$$

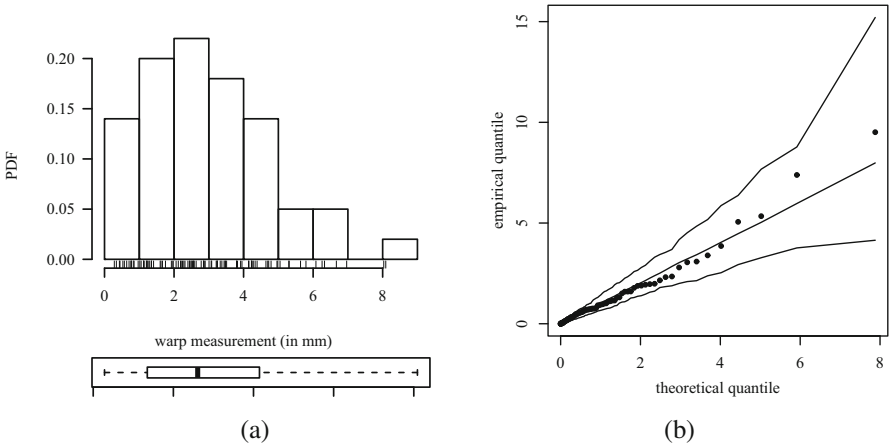
$$C_{pu}^{BS} = \frac{2(\text{USL} - \beta)}{t(q_2) - t(q_1)} = \frac{2(\text{USL} - \beta)}{\beta\alpha\left\{z_2\left[\frac{\alpha z_2}{2} + \sqrt{\left(\frac{\alpha z_2}{2}\right)^2 + 1}\right] - z_1\left[\frac{\alpha z_1}{2} + \sqrt{\left(\frac{\alpha z_1}{2}\right)^2 + 1}\right]\right\}}.$$

$$C_p^{BS} = \frac{2(\text{USL} - \text{LSL})/\beta}{\left[\alpha^2 z_2{}^2 + \alpha z_2\sqrt{(\alpha z_2)^2 + 4}\right] - \left[\alpha^2 z_1{}^2 + \alpha z_1\sqrt{(\alpha z_1)^2 + 4}\right]},$$

and $C_{pk}^{BS} = \min\{C_{pl}^{BS}, C_{pu}^{BS}\}$, where $z_1 = z(q_1)$ and $z_2 = z(q_2)$, with $z(q_i)$ being the $q_i \times 100$-th quantile of $Z \sim \text{N}(0, 1)$, for $i = 1, 2$. Leiva et al. [13] used the ML method to estimate the model parameters, and the normal and quantile bootstrap methods to obtain confidence intervals (CIs) for the PCIs. Values of q_1 and q_2 are

Table 4 Summary statistics for the warping data

n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
100	0.282	2.607	2.923	8.091	1.786	61.09%	0.687	0.008



(a) (b)

Fig. 8 Histogram with adjusted boxplot (**a**) and QQ plot with its envelope (**b**) for warping data

obtained by minimizing the variance of the corresponding Birnbaum-Saunders PCI estimator.

7.2 Illustration

A data set containing warp measurements (in millimeters, mm) in a tiles factory is analyzed. The aim is to ensure the production quality. The engineers obtained $n = 100$ observations on the warping. The USL for the warping measurement is 8 mm. This data set is available in the `qAnalyst` package of the `R` software; www.R-project.org. Table 4 provides descriptive statistics for the warping data set, including central tendency statistics, SD, CV, CS and CK. From this table, the right skewed nature of the data distribution can be noted, as confirmed by the histogram with adjusted boxplot for asymmetric data shown in Fig. 8a. The assumption that the data follow a Birnbaum-Saunders distribution is supported by the QQ plot with its envelope displayed in Fig. 8b. The application of the KS GOF test provides a $p-$value $= 0.05264$, from which is possible to conclude that the Birnbaum-Saunders distribution is a good model for these data at 5% of significance.

The ML estimates of the Birnbaum-Saunders parameters α and β are given by $\widehat{\alpha} = 0.791$ and $\widehat{\beta} = 2.208$, respectively. Table 5, about Birnbaum-Saunders PCI based on a process with only the USL, presents the estimate of C_{pu}^{BS} and its corresponding bootstrap CIs, along with sample median and mean. Note that the

Table 5 USL, sample median and mean, and point and interval estimates for PCIs C_{pu}^{BS} with warping data

USL	$\bar{t}$	$t(0.5)$	$\widehat{C}_{pu}^{BS}$	Normal bootstrap CI	Quantile bootstrap CI
8 mm	2.923	2.607	3.914	(2.964; 4.668)	(3.299; 4.941)

Table 6 Minimum values recommended for PCIs

Type of process	Sigma level	Two-sided specifications	One-sided specifications
Existing process	4.0	1.33	1.25
New process	4.5	1.50	1.45
Existing process including safety, strength, or critical parameters	4.5	1.50	1.45
New process including safety, strength, or critical parameters	5.0	1.67	1.60
Six-sigma process	6.0	2.00	2.00

Source: see [20, p.776]

estimated capability index $\widehat{C}_{pu}^{BS} = 3.914$ is greater than the generally accepted industry guideline of 1.45 (see Table 6), which indicates that the tiles manufacturing process is meeting the specification. Therefore, one can conclude that the process is capable and does meet the requirements.

8 Acceptance Sampling Plans

In this section, acceptance sampling plans are provided as an important aspect to consider when validating the external quality of a company. Acceptance sampling plans are presented and discussed when the life test is truncated at a pre-fixed time. The minimum sample size needed to ensure the specified median life is obtained assuming that the lifetimes of the test units follow a Birnbaum-Saunders distribution. The operating characteristic values of the sampling plans, as well as the producer risk, are analyzed; see details in [1].

8.1 Formulation

The lifetime of the product under study (T) is assumed to follow a Birnbaum-Saunders distribution. A common practice in life experiments is to stop the life test at a pre-fixed time (t) and record the number of failures. In these tests, one often sets a (lower) confidence limit on the mean or median life (or any other quantile of the distribution). It is therefore desired to establish a specified mean or median life with a probability of at least P^* (consumer's risk). Normally, the mean life is used if the

life distribution is symmetric, whereas the median life is preferred if the distribution is skewed. The decision to accept the specified mean or median life occurs if the number of failures at the end of the pre-fixed time t does not exceed a given number c. Thus, the test terminates at time t or at the $(c+1)$-th failure, whichever occurs first. For a truncated life test and the associated decision rule, it is important to find a sampling plan, that is, the minimum sample size needed to achieve the goal.

An acceptance sampling plan based on truncated life tests consists of: (i) the number of units on test (n); (ii) the acceptance number (c), such that the lot is accepted if at most c failures out of n occur at the end of the pre-fixed time t; and (iii) the ratio t/μ_0, where μ_0 is the specified mean or median life and t is the maximum test duration. Hence, β is the lot quality parameter for a random variable $T \sim \text{BS}(\alpha, \beta)$, since β is the median life as well as the scale parameter of the Birnbaum-Saunders distribution. Thus, an acceptance sampling plan based on truncated life tests for the Birnbaum-Saunders distribution is $(n, c, t/\beta_0)$.

The probability of accepting a bad lot (consumer's risk), that is, when the true median life β is below the specified β_0, is fixed not to exceed $1 - P^*$. The lot size (N) is assumed large enough to be considered infinite (for example, $n/N \leq 0.10$), so that the binomial distribution can be used. Thus, the acceptance and non-acceptance criteria for the lot are equivalent to the decisions of accepting or rejecting the hypothesis $\beta \geq \beta_0$. The minimum sample size (n) is sought such that

$$\sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x} \leq 1 - P^*, \tag{17}$$

where $p = F_T(t; \alpha, \beta)$, given in (1), is monotonically increasing in t/β and decreasing in β, for t fixed; this can be easily established for the Birnbaum-Saunders distribution. Thus, $p = F_T(t; \alpha, \beta) = F_T(t/\beta, \alpha, 1)$ depends only on the ratio t/β, once α is fixed. Hence, it is sufficient to specify just this ratio. Therefore, if the number of observed failures is at most c, from (17), it can be established with probability P^* that $F_T(t/\beta) \leq F_T(t/\beta_0)$, which implies that $\beta \geq \beta_0$. The minimum values of n satisfying (17) for $P^* = 0.75, 0.9, 0.95, 0.99$ and $t/\beta_0 = 0.628, 0.942, 1.257, 1.571, 2.356, 3.141, 3.927, 4.712$ are presented in Table 3 of [1], for α fixed. The operating characteristic function of the plan $(n, c, t/\beta_0)$ gives the probability that the lot can be accepted. For the plan under analysis, this probability is given by

$$L(p) = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x}, \tag{18}$$

where p, given in (17), is a monotonically decreasing function of $\beta \geq \beta_0$, for t fixed, while $L(p)$ is decreasing in p. Based on (18), the operating characteristic values, as a function of β/β_0, for α fixed, are presented in Table 4 of [1] regarding the plan $(n, c, t/\beta_0)$, with c also being fixed and assuming different values of P^*. For given P^* and t/β_0, the choice of c and n can be made on the basis of the operating characteristic function.

The producer's risk is defined as the probability of rejecting a lot when $\beta \geq \beta_0$. For the sampling plan under consideration and a given value for the producer's risk ω, one may be interested in knowing the value of β/β_0 that ensures this risk to be at most ω. Note that (4) can be written as

$$a(t, \alpha, \beta) = \frac{1}{\alpha}\left(\frac{\sqrt{t/\beta_0}}{\sqrt{\beta/\beta_0}} - \frac{\sqrt{\beta/\beta_0}}{\sqrt{t/\beta_0}}\right), \tag{19}$$

which is denoted by $a(t, \alpha, \beta/\beta_0)$, for α fixed. Based on (19), the probability $p = F_T(a(t; \alpha, \beta))$ may be obtained as function of β/β_0, that is, $p = F_T(a(t, \alpha, \beta/\beta_0))$. Then, β/β_0 is the smallest positive number for which $p = F_T(a(t, \alpha, \beta/\beta_0))$ satisfies

$$\sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x} \geq 1 - \omega.$$

For a given plan $(n, c, t/\beta_0)$, at a specified confidence level P^*, the minimum values of β/β_0 satisfying (17) are presented in Table 5 of [1], with α being fixed. For extensions and approximations of the acceptance plan presented above, see [1].

8.2 Illustration

Consider a problem associated with software reliability; see details in [1] and references therein. Failure times T (in hours) of a software are collected from its release. Use of the software might imply the development of intangible cumulative degradation, worsening its performance. A typical example is provided by the efficiency of an operating system which is worsening because of many causes, for example, malwares, badly installed programs, fragmented files. Then, it is reasonable to suppose that T follows a Birnbaum-Saunders distribution and can be used to classify units of the software as defective or non-defective.

Let the specified median life be $\beta_0 = 1000$ hours and the testing time be $t = 1257$ hours, so the ratio t/β_0 is 1.257. Thus, for an acceptance number $c = 2$ and a confidence level $P^* = 0.95$, the required n is found; in the case of truncated life tests from the Birnbaum-Saunders distribution, the acceptance sampling plan is $(n = 9, c = 2, t/\beta_0 = 1.257)$.

Consider the ordered sample t_i for $i = 1, \ldots, 9$, of size $n = 9$ from T (in hours): 519, 968, 1430, 1893, 2490, 3058, 3625, 4422, and 5218. Based on these data, a decision has to be made whether to accept or reject the lot, which is accepted if the number of failures before 1257 h is greater than two.

Since the confidence level is assured by this sampling plan only if $T \sim \mathrm{BS}(\alpha, \beta)$, then a preliminary test is needed to check if the given sample comes from such distribution. The Birnbaum-Saunders model is fit to the data and its parameters

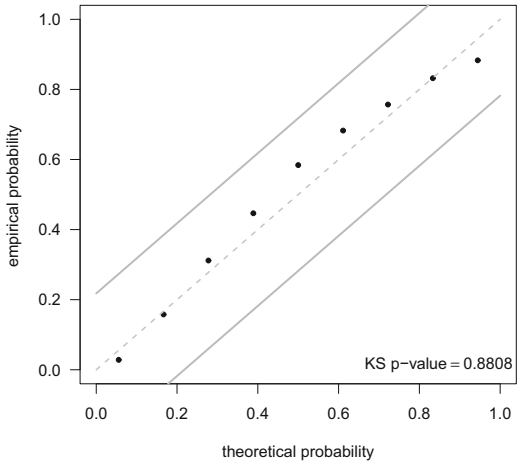


Fig. 9 PP plots with 5% KS acceptance regions using software reliability data

α and β are estimated using the ML method. Figure 9 shows the PP plots of the data. Based on this figure and in the KS p-value 0.8808, the Birnbaum-Saunders distribution fits the data very well.

Note that, in the given sample of $n = 9$ observations, there are only $c = 2$ failures at 519 and 968 h before $t = 1250$ h. Therefore, the lot is accepted, assuring a median life $\beta_0 = 1000$ h with a confidence level of $P^* = 0.95$.

9 Reliability Analysis

In this section, the Birnbaum-Saunders distribution is used to analyze the effect on hardness when incorporating nanotechnology. In biomaterials, one can study the effect of nanoparticles on the reliability of the mechanical response. Hardness is modeled by the Birnbaum-Saunders distribution and Bayesian inference is performed to derive a methodology about reliability, which allows us to evaluate the effect of the use of nanotechnology at different loadings; see details in [14].

9.1 Formulation

Materials can be compared considering the reliability function of T given by

$$\mathrm{P}(T > t|\underline{t}) \approx \frac{1}{N}\sum_{i=1}^{N} \Phi\left(\frac{1}{\alpha^{(i)}}\left(\sqrt{\beta^{(i)}/t} - \sqrt{t/\beta^{(i)}}\right)\right). \tag{20}$$

For details about the approximation given in (20), see Appendix B of [14]. Consider two materials with hardness $T_1 \sim \mathrm{BS}(\alpha_1, \beta_1)$ and $T_2 \sim \mathrm{BS}(\alpha_2, \beta_2)$, respectively, and the corresponding observed samples $\underline{t}_1$ and $\underline{t}_2$. Let T_1 and T_2 be independent random variables. The reliability function of T_1 is denoted as $R_1(t|\alpha_1, \beta_1)$, whereas $f_2(t|\alpha_2, \beta_2)$ denotes the PDF of T_2. The materials can be compared by

$$\mathrm{P}(T_1 > T_2|\underline{t}_1, \underline{t}_2) \approx \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \int R_1(t|\alpha_1^{(i)}, \beta_1^{(i)}) f_2(t|\alpha_2^{(j)}, \beta_2^{(j)}) \mathrm{d}t, \tag{21}$$

where N_1 and N_2 are the sizes of the posterior samples for the first and second material, respectively. For details about the approximation given in (21), see Appendix B of [14]. The proper Bayesian estimator of $\mathrm{P}(T_1 > T_2|\underline{t}_1, \underline{t}_2)$ is given by (21), which has the drawback of requiring $N_1 N_2$ integrations. The Monte Carlo method can be used for carrying out this integration. An alternative estimator, easier to compute, is provided by

$$\widehat{\mathrm{P}}(T_1 > T_2|\underline{t}_1, \underline{t}_2) \approx \int R_1(t|\widehat{\alpha}_1, \widehat{\beta}_1) f_2(t|\widehat{\alpha}_2, \widehat{\beta}_2) \mathrm{d}t, \tag{22}$$

where the parameter estimates are plugged in and so just one integral is computed. Although the error incurred when considering (22) instead of (21) cannot be quantified, one may consider only the former for practical reasons. This alternative estimator is also used in other situations, for example, when estimating the mean value function of a non-homogeneous Poisson process; see [21].

Consider the hardness of two materials and model their medians (β in the Birnbaum-Saunders distribution) with $\exp(\mu_T)$ and $\exp(\mu_T + \delta)$, respectively. Therefore, the effect of two treatments can be detected through δ. A negative (positive) value for δ implies a decrease (increase) of the median, denoting a smaller (larger) hardness. No significant difference between the two materials is detected when $\delta = 0$. Suppose that the materials share the same α in the Birnbaum-Saunders distribution. To compare the hardness of two materials, data are arranged as $\underline{t} = (t_1, \ldots, t_m, t_{m+1}, \ldots, t_n)^\top$, where the first m values refer to the material with median $\exp(\mu_T)$ and the remaining $n - m$ to the other material. A Gibbs sampling with Metropolis-Hastings steps is considered by obtaining the PDFs of the posterior distribution of $(\alpha, \mu_T, \delta)^\top$ through their corresponding conditional distributions; see details in Appendix B of [14]. There is interest in assessing if a positive effect exists when considering the second material. Thus, one could look at the posterior distribution of δ, and particularly at $\mathrm{P}(\delta > 0|\underline{t})$. Given a sample $\{\alpha^{(i)}, \mu_T^{(i)}, \delta^{(i)}\}_{i=1}^N$ from the posterior distribution, it is possible to estimate $\mathrm{P}(\delta > 0|\underline{t})$ simply counting the frequency of positive values of δ in the sample. Therefore, $\mathrm{P}(\delta > 0|\underline{t})$ can be compared to $\mathrm{P}(\delta \leq 0|\underline{t})$ and then one can decide that a positive effect exists if $\mathrm{P}(\delta > 0|\underline{t}) > \mathrm{P}(\delta \leq 0|\underline{t})$. The comparison corresponds to contrast the hypotheses

$$\mathrm{H}_0\colon \delta > 0 \quad \text{versus} \quad \mathrm{H}_1\colon \delta \leq 0. \tag{23}$$

One could also be interested in the test

$$\mathrm{H}_0\colon \delta = 0 \quad \text{versus} \quad \mathrm{H}_1\colon \delta \neq 0. \tag{24}$$

Then, a first approach considers a test as in (24), which relies on a prior PDF $\pi(\delta)$ as a mixture of a continuous PDF $\pi_0(\delta)$ and a Dirac ξ_0 (a point mass) at zero, that is, $\pi(\delta) = (1-\varepsilon)\pi_0(\delta) + \varepsilon\,\xi_0(\delta)$, where ε denotes the prior probability $\mathrm{P}(\delta = 0)$. The posterior PDF is given by $\pi^*(\delta) = (1-\varepsilon^*)\pi_0^*(\delta) + \varepsilon^*\xi_0(\delta)$, where ε^* is the posterior probability $\mathrm{P}(\delta = 0|\underline{t})$. Here, $\mathrm{H}_0\colon \delta = 0$ is accepted if $\mathrm{P}(\delta = 0|\underline{t}) > 0.5$. The second approach modifies the test given in (24) into $\mathrm{H}_0\colon |\delta| \leq \eta$ versus $\mathrm{H}_1\colon |\delta| > \eta$, with η sufficiently close to zero. In this case, a continuous prior PDF $\pi(\delta)$ could be considered and the null hypothesis is accepted if $\mathrm{P}(|\delta| \leq \eta|\underline{t}) > 0.5$. Unilateral tests as given in (23) could be performed similarly. Bayesian credible intervals (BCIs) may be obtained either in closed form or from a sample from the Markov chain Monte Carlo method. The BCIs provide the probability for the parameter(s) (random variables on the Bayesian framework) to be in predetermined intervals, unlike the confidence intervals (random variables in the frequentist framework) providing a procedure which should contain the parameter (constant and unknown) with a predetermined frequency, upon repeated experiments. A summary of the presented reliability methodology is provided in Algorithm 1 of [14].

9.2 Illustration

The data correspond to hardness (in megaPascal units, MPa) of a bone cement called Palacos R, treated with low-loadings of mesoporous silica nanoparticles (MSNs). Three types of MSNs are used in loading ratios of 0.1 and 0.2 related to the powder of the Palacos R bone cement, defined as: (i) plain unmodified (SiO_2); (ii) propylamine functionalized (NH_2); and (iii) propylcarboxylic acid functionalized (COOH). The plain cement (Palacos) is used as a control and six treatments containing MSNs (three particle types and two loading ratios): 0.1/SiO_2, 0.2/SiO_2, 0.1/NH_2, 0.2/NH_2, 0.1/COOH and 0.2/COOH. For the control and each treatment, hardness is measured after micro-indentation testing in the material. The goal is to verify whether an addition of MSNs increases hardness or not. For more details about this biomaterial-related problem and the data, see [14]. Table 7 provides the median, mean, SD, CV, CS and CK for hardness data. Note that the Birnbaum-Saunders distribution can be reasonably assumed to model all those data sets due to their asymmetric nature and kurtosis level; see also the histograms and boxplots in Fig. 10. The choice of the Birnbaum-Saunders distribution for modeling the data under analysis is supported by the results of the descriptive statistics.

Figure 10 provides the histogram with adjusted box-plot for asymmetric data. In addition, this figure displays the Mahalanobis distance plot and PP plot with 95% acceptance bands for hardness data in the control and treatment groups under study. From Fig. 10, it is possible to note that the PP plots, and the KS p-values, support

Table 7 Descriptive statistics for the indicated hardness data set

Data set	n	Minimum	Median	Mean	Maximum	SD	CV	CS	CK
Palacos	20	171.15	177.87	178.59	190.49	4.82	2.70%	0.79	3.02
0.1/SiO_2	20	176.92	180.80	181.00	187.55	2.67	1.48%	0.60	2.7
0.2/SiO_2	20	177.40	185.05	185.09	198.79	5.18	2.80%	0.68	3.19
0.1/NH_2	20	168.61	178.52	178.97	189.54	4.18	2.34%	0.01	4.20
0.2/NH_2	20	174.45	180.04	181.60	193.83	4.48	2.46%	0.91	3.53
0.1/COOH	20	170.63	178.69	179.08	193.71	4.79	2.67%	1.07	5.15
0.2/COOH	20	156.10	170.56	169.78	193.66	9.64	5.68%	0.38	2.70

the use of the Birnbaum-Saunders model. A notable influence of atypical data on the estimated model is not detected using the Mahalanobis distance.

Table 8 reports posterior summaries of the model parameters for each treatment. The results are obtained using a Gibbs sampling with a sample size of 10,000; see more details about this procedure in [14]. The hardness of the six treatments related to standard Palacos and percentage of MSN loading ratio is compared. Tables 8 and 9 provide the posterior summaries from which is possible to conclude that:

(i) The hardness of 0.2/SiO_2 is clearly the largest, according to both estimate of the median β provided by Table 8 and estimate of δ in the "Palacos versus 0.2/SiO_2" case of Table 9.

(ii) The estimates are very accurate as shown by the small SDs and the corresponding very narrow BCIs; see Tables 8 and 9.

(iii) Related to standard Palacos, the probability $P(\delta > 0|\underline{t})$ shows that the addition of MSNs significantly increases the hardness with 0.1/SiO_2, 0.2/SiO_2, 0.1/NH_2, 0.2/NH_2 and 0.1/COOH. A significant hardness decrease is verified with 0.2/COOH; see Table 9. However, it is worth noting that the 95% BCI does not include the value zero only in the case of 0.2/SiO_2.

(iv) The small values of the estimates of $P(T_1 > T_2|\underline{t}_1, \underline{t}_2)$, for the 0.1/$SiO_2$, 0.2/$SiO_2$, 0.1/$NH_2$, 0.2/$NH_2$ and 0.1/COOH treatments, support the results mentioned above in item (iii); see Table 9. These estimates indicate that the hardness related to the mentioned treatments are much larger than the hardness associated with Palacos.

(v) The probability $P(\delta > 0|\underline{t})$ suggests that increasing the MSN content from 0.1 to 0.2 significantly increases the hardness for SiO_2 and NH_2 treatments, but not for COOH; see Table 9. The estimates of $P(T_1 > T_2|\underline{t}_1, \underline{t}_2)$ also support this result.

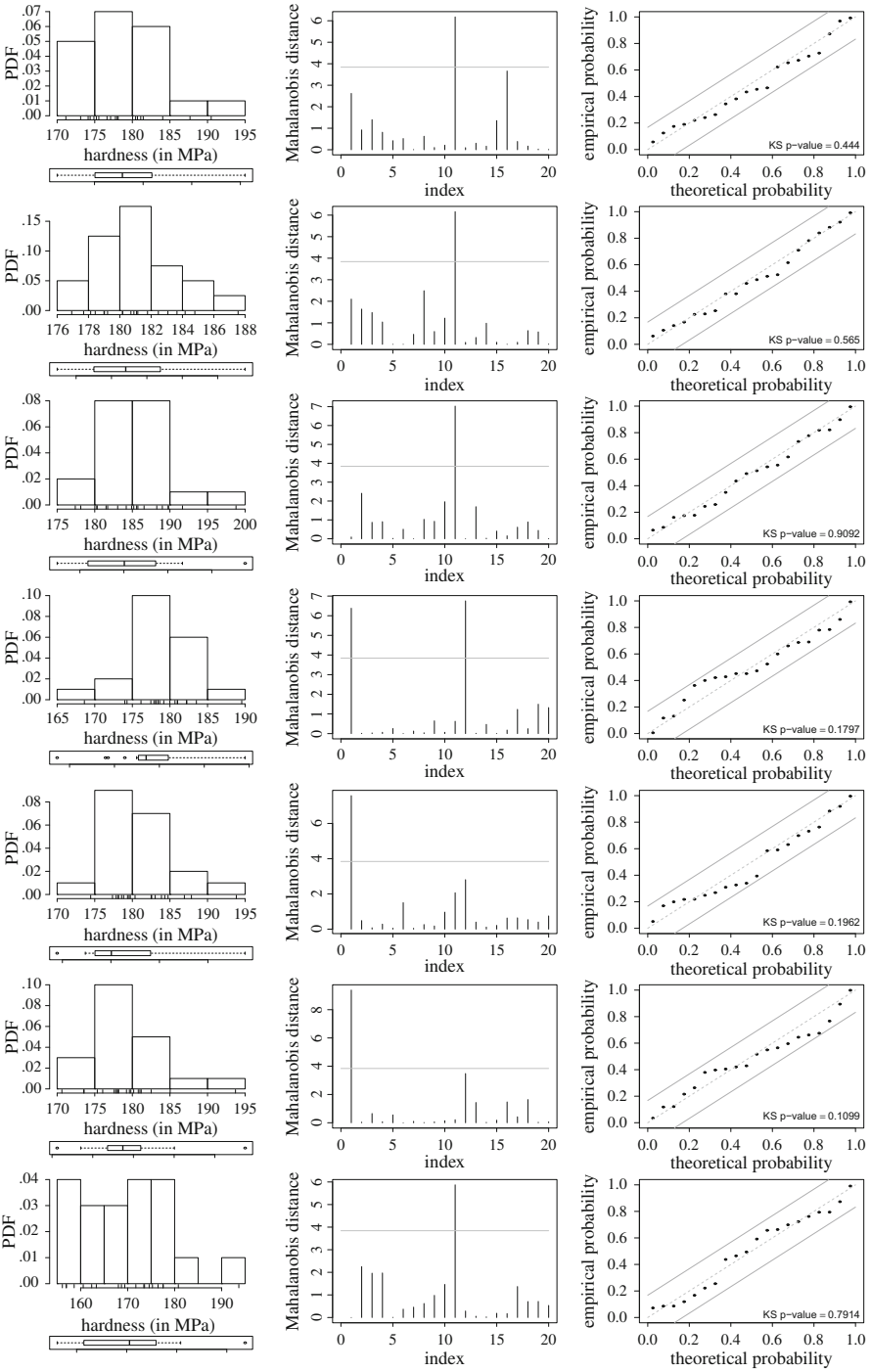


Fig. 10 Histogram with adjusted boxplot, Mahalanobis distance plot and PP plot with 95% acceptance bands for the Palacos (1st line), 0.1/SiO_2 (2nd line), 0.2/SiO_2 (3rd line), 0.1/NH_2 (4th line), 0.2/NH_2 (5th line), 0.1/COOH (6th line) and 0.2/COOH (7th line) using hardness data sets

Table 8 Bayesian estimates, SDs and 95% BCIs of the indicated Birnbaum-Saunders parameter and hardness data set

	α			β		
Data set	Mean	SD	95% BCI	Mean	SD	95% BCI
Palacos	0.101	0.034	(0.055,0.186)	178.536	0.497	(177.28,179.79)
0.1% SiO_2	0.100	0.033	(0.054,0.182)	180.984	0.492	(179.71,182.22)
0.2% SiO_2	0.101	0.034	(0.055,0.184)	185.029	0.492	(183.79,186.30)
0.1% NH_2	0.101	0.034	(0.055,0.185)	178.924	0.495	(177.68,180.19)
0.2% NH_2	0.100	0.033	(0.055,0.182)	181.554	0.490	(180.35,182.82)
0.1% COOH	0.101	0.033	(0.055,0.184)	179.027	0.492	(177.76,180.25)
0.2% COOH	0.103	0.034	(0.056,0.185)	169.521	0.497	(168.26,170.78)

10 Conclusions and Future Research

In the dawn of statistics, techniques were linear and based mainly on the Gaussian distribution. These techniques were attractive due to ease of calculation, but lacked flexibility. Then, the use of such techniques was restrictive. For practical purposes, several statistical methods, constructed on varied distributions, are needed. Each use requires a suitable model. The Birnbaum-Saunders model is one of these distributions and has recently received considerable attention due to its interesting properties and its relationship with the Gaussian distribution. It has now been 50 years since the origins of the Birnbaum-Saunders distribution, but since then, it went a long way. Its origins were as a model for the cumulative damage describing fatigue and failure of materials. As a unimodal and skewed to the right distribution, it has found widespread use as a life distribution in many applications, even outside material science. Extensions and variations conducted at multivariate Birnbaum-Saunders and logarithmic Birnbaum-Saunders distributions, allowing several other applications. In the last decades, there have been many advances and applications. Therefore, the Birnbaum-Saunders distribution provides another opportunity in modeling. This has already proven to be sharp, giving reliable and insightful results. The study of Birnbaum-Saunders distributions is now a mature topic, with a substantial literature, which has had significant impact in important and interesting fields, for example, in business and industry. The methods are rigorously defined with well understood and developed theoretical results. This work has provided a part of these theoretical results in a unified and consistent notation to enable a major appreciation.

In the quality setting, the Gaussian distribution was also the most relevant ingredient of statistical quality tools, but nowadays new models are considered, mainly in asymmetric frameworks. To this end, procedures based on the Birnbaum-Saunders distribution have been presented, discussed and applied to quality improvement through statistical quality control, acceptance sampling and reliability analysis. Such tools have been implemented in the `R` software. For the internal quality of the company, the control charts for attributes and variables, as well as their multivariate

Table 9 Estimates, SDs, 95% CIs, $P(\delta > 0|\underline{t})$ and $P(T_1 > T_2|\underline{t}_1, \underline{t}_2)$ for the indicated comparison with hardness data sets

Parameter	Mean	SD	95% CI	$P(\delta > 0\|\underline{t})$	$P(\delta \leq 0\|\underline{t})$	$P(T_1 > T_2\|\underline{t}_1, \underline{t}_2)$
Palacos versus 0.1% SiO_2						
α	0.116	0.025	(0.077,0.175)			
μ	5.185	0.010	(5.160,5.210)			
δ	0.014	0.012	(−0.018,0.044)	0.932	0.068	0.186
Palacos versus 0.2% SiO_2						
α	0.117	0.026	(0.077,0.178)			
μ	5.185	0.010	(5.158,5.210)			
δ	0.036	0.012	(0.004,0.067)	0.982	0.018	0.206
Palacos versus 0.1% NH_2						
α	0.117	0.026	(0.076,0.176)			
μ	5.185	0.011	(5.157,5.210)			
δ	0.002	0.013	(−0.031,0.034)	0.888	0.112	0.221
Palacos versus 0.2% NH_2						
α	0.117	0.025	(0.076,0.175)			
μ	5.185	0.010	(5.159,5.211)			
δ	0.017	0.012	(−0.015,0.048)	0.939	0.061	0.198
Palacos versus 0.1% COOH						
α	0.117	0.026	(0.077,0.177)			
μ	5.185	0.010	(5.160,5.211)			
δ	0.003	0.013	(−0.029,0.034)	0.886	0.046	0.205
Palacos versus 0.2% COOH						
α	0.118	0.026	(0.077,0.180)			
μ	5.185	0.010	(5.159,5.210)			
δ	−0.052	0.013	(−0.084,−0.020)	0.006	0.994	0.426
0.1% SiO_2 versus 0.2% SiO_2						
α	0.117	0.026	(0.076,0.177)			
μ	5.198	0.010	(5.173,5.223)			
δ	0.022	0.013	(−0.010,0.053)	0.954	0.046	0.162
0.1% NH_2 versus 0.2% NH_2						
α	0.117	0.026	(0.077,0.176)			
μ	5.187	0.011	(5.161,5.214)			
δ	0.015	0.012	(−0.014,0.046)	0.941	0.059	0.235
0.1% COOH versus 0.2% COOH						
α	0.118	0.026	(0.078,0.180)			
μ	5.188	0.010	(5.162,5.213)			
δ	−0.055	0.013	(−0.087,−0.023)	0.005	0.995	0.433

versions and capability indices, were presented and illustrated with real data. For external quality, acceptance sampling plans were also treated and illustrated. Finally, reliability models based on the Birnbaum-Saunders distribution have been presented and illustrated with data on nanotechnologies.

Some distributions may be expressed as a mixture of Gaussian distributions, which may be useful for performing Bayesian analysis using Markov chain Monte Carlo methods; see, for example [26]. Generalizations of Birnbaum-Saunders distributions can be represented as a mixture of Gaussians, so that these ideas on Bayesian analysis can be more explored, and this work presented a possible approach; see [2, 16].

In summary, with more sophisticated and appropriate statistical models, quality monitoring can be assessed more reliably and inaccurate decisions avoided. New tools have been presented to explore further investigations within this important application field of quality improvement. Other researchers are encouraged to begin this exploration in the belief that it can have a positive impact.

Acknowledgements This research work was partially supported by FONDECYT 1160868 grant from the Chilean government.

References

1. Balakrishnan, N., Leiva, V., & López, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum-Saunders distribution. *Communications in Statistics: Simulation and Computation, 36*, 643–656.
2. Balakrishnan, N., Leiva, V., Sanhueza, A., & Vilca, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of Gaussians and the EM-algorithm. *Statistics and Operations Research Transactions, 33*, 171–192.
3. Birnbaum, Z. W., & Saunders, S. C. (1969a). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability, 6*, 328–347.
4. Birnbaum, Z. W., & Saunders, S. C. (1969b). A new family of life distributions. *Journal of Applied Probability, 6*, 319–327.
5. Fox, E., Gavish, B., & Semple, J. (2008). *A general approximation to the distribution of count data with applications to inventory modeling*. Technical Report 07-002. Dallas, US: SMU Cox School of Business.
6. Fuchs, C., & Kenett, R. (1998). *Multivariate quality control: theory and applications*. New York, US: Marcel Dekker.
7. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (Vol. 2). New York, US: Wiley.
8. Kane, V. (1986). Process capability indices. *Journal of Quality Technology, 18*, 41–52.
9. Kundu, D. (2015). Bivariate log-Birnbaum-Saunders distribution. *Statistics, 49*, 900–917.
10. Leiva, V. (2016). *The Birnbaum-Saunders distribution*. New York, US: Academic Press.
11. Leiva, V., & Saunders, S. C. (2015). Cumulative damage models. *Wiley StatsRef: Statistics Reference Online*, 1–10.
12. Leiva, V., Marchant, C., Ruggeri, F., & Saulo, H. (2015). A criterion for environmental assessment using Birnbaum-Saunders attribute control charts. *Environmetrics, 26*, 463–476.
13. Leiva, V., Marchant, C., Saulo, H., Aslam, M., & Rojas, F. (2014a). Capability indices for Birnbaum-Saunders processes applied to electronic and food industries. *Journal of Applied Statistics, 41*, 1881–1902.
14. Leiva, V., Ruggeri, F., Saulo, H., & Vivanco, J. F. (2017). A methodology based on the Birnbaum-Saunders distribution for reliability analysis applied to nano-materials. *Reliability Engineering and System Safety, 157*, 192–201.

15. Leiva, V., Sanhueza, A., Sen, P. K., & Paula, G. A. (2008). Random number generators for the generalized Birnbaum-Saunders distribution. *Journal of Statistical Computation and Simulation, 78*, 1105–1118.
16. Leiva, V., Saulo, H., Leão, J., & Marchant, C. (2014b). A family of autoregressive conditional duration models applied to financial data. *Computational Statistics and Data Analysis, 79*, 175–191.
17. Lio, Y. L., & Park, C. (2008). A bootstrap control chart for Birnbaum-Saunders percentiles. *Quality and Reliability Engineering International, 24*, 585–600.
18. Marchant, C., Leiva, V., Cysneiros, F. J. A., & Liu, S. (2018). Robust multivariate control charts based on Birnbaum-Saunders distributions. *Journal of Statistical Computation and Simulation, 88*, 182–202.
19. Marchant, C., Leiva, V., Christakos, G., & Cavieres, M.F. Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile. *Environmetrics* (in press). https://doi.org/10.1002/env.2551
20. Montgomery, D. (2005). *Introduction to statistical quality control*. New York, US: Wiley.
21. Pievatolo, A., & Ruggeri, F. (2010). Bayesian modelling of train doors' reliability. In A. O'Hagan & M. West (Eds.), *Handbook of Applied Bayesian Analysis*. Oxford: Oxford University Press.
22. Raaijmakers, F. J. M. (1980). The lifetime of a standby system of units having the Birnbaum-Saunders distribution. *Journal of Applied Probability, 17*, 490–497.
23. Raaijmakers, F. J. M. (1981). Reliability of standby system for units with the Birnbaum-Saunders distribution. *IEEE Transactions on Reliability, 30*, 198–199.
24. Rousseeuw, P. J., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., et al. (2016). *robustbase: Basic robust statistics*. R package version 0.92-6.
25. Saulo, H., Leiva, V., & Ruggeri, F. (2015). Monitoring environmental risk by a methodology based on control charts. In C. Kitsos, T. Oliveira, A. Rigas, & S. Gulati (Eds.), *Theory and practice of risk assessment* (pp. 177–197). Switzerland: Springer.
26. Wand, M. P., Ormerod, J. T., Padoan, S. A., & Frühwirth, R. (2011). Mean field variational Bayes for elaborate distributions. *Bayesian Analysis, 6*, 847–900.

Statistical System Monitoring (SSM) for Enterprise-Level Quality Control

Siim Koppel and Shing I Chang

Abstract The rapid development and adoption of sensors and data storage solutions such as IOT (Internet of Things) has enabled the collection of a large amount of data in production facilities. The data may come from different sources such as process parameters and quality characteristics. However, traditional statistical process control (SPC) tools were not built to take the full advantages of the data provided. Traditional SPC tools such as control charts are often applied to critical quality characteristics (QCs) on a product rather than incorporating process parameters associated with the critical QCs. This chapter proposes a method that is capable of monitoring all process and quality data simultaneously. The proposed method adopts precontrol and group control chart ideas to pinpoint change location and timeframe in a production system. After the change location and timeframe have been identified, more elaborate models or data analytics methods can be used to identify potential assignable causes. Simulation studies are conducted to establish the properties of the proposed method. Guidelines are provided to help users how to implement the proposed method in any production facility including those facing big data issues.

Keywords Statistical system monitoring · Statistical process monitoring · Change detection · Big data · Continuous variable · Production

1 Introduction

Statistical process control (SPC) approaches were first introduced by Water Shewhart in 1924 [1]. The use of statistical methods such as hypothesis testing in graphic forms coupled with the Central Limit Theorem has served numerous

S. Koppel (✉) · S. I Chang
IMSE Department, Kansas State University, Manhattan, KS, USA
e-mail: Siim@ksu.edu; Changs@ksu.edu

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_3

applications well especially in manufacturing to achieve desirable product quality. The core concept is the use of a set of historical data deemed in control to set up a pair of control charts. This process is usually called Phase I SPC. Then during Phase II SPC, statistics of future observations of a quality characteristic are plotted for continuous monitoring. If any point plots outside control limits, then the process under monitoring is deemed out of control. Process engineers are then informed for fault diagnoses.

This basic framework remains unchanged till this day although multiple revisions such as CUSUM and EWMA charts have been proposed to improve the sensitivity of detecting small process shifts [2–4]. Hotelling T^2 [5] was proposed to extend univariate quality characteristic to multivariate quality characteristics. However, the combinational development of omnipresence of sensors and cloud computing often-called Industrial 4.0 or cyber-physical systems [6] has opened up opportunities to rethink implementation strategies of SPC for manufacturing. Traditional SPC methods, often restricted to product quality characteristics, cannot take advantage of big data generated from a production process equipped with thousands of process parameters and hundreds of product characteristics scattered throughout a production system.

We propose a system-wise process monitoring framework to answer this challenge. The proposed framework is called *statistical system monitoring* or SSM in that all process parameters and QCs are considered simultaneously for change detection. A system is composed of multiple processes that may be hierarchical. This chapter provides the SSM framework that adopts parts of collection of monitoring methods such as pre-control and group control chart algorithms. The core concept of SSM is to, first, quantify the performance of a system or subsystem composed of process parameters and product quality characteristics over time into three zones (green, yellow, and red zone in Fig. 1). Then, only those segments of time series that exhibits changes in terms of the statistics reflecting green, yellow, and red zones are to be analyzed. A small manufacturing example with three departments is used to demonstrate the use of the proposed method. The goal is to monitor the full system, not just the individual parts. Simulated data sets are generated to demonstrate the properties of the proposed method.

Unlike the traditional methods where measurement is restricted to physical products or work in progress, the proposed SSM framework integrates process parameters associated with products or work in process for process monitoring and defect prevention. Since the number of parameters is usually very large, high dimensional problem often confronts traditional control charts. For example, the production of semi-conductor wafers includes hundreds of processes and thousands of process parameters. The proposed SSM framework contains multiple techniques for dimension reduction and feature selection. The following section briefly outlines some of these methods in the content of statistical process control or monitoring.

Fig. 1 The basic set-up of pre-control chart

2 Background

2.1 Traditional Multivariate SPC

Much research has been generated on the topic of statistical process monitoring over the last decades. Most work focuses on a univariate quality characteristic.

2.1.1 Hotelling T^2

Hotelling T^2 is one of the oldest SPC methods for monitoring multivariate processes. This method is monitoring the mean vector of the process. The monitoring is done by plotting a chi-squared control chart [7].

The statistic plotted is calculated based on the vector of variable means over some time. Then covariance matrix is used to calculate the statistic. There also lies the biggest issue of the method—the estimation of the elements in a covariance matrix. It is possible to calculate the covariance matrix when there is about ten variables, but anything over that, the task becomes very difficult or next to impossible to implement. Also, Hotelling T^2 is often applied to several QCs. Usually process parameters associated with the QCs are not considered in the same vector. Multivariate versions of Exponentially Weighted Moving Average (EWMA) and cumulative sum (CUSUM) charts [3, 8, 9] suffer the same drawbacks. These control charts were merely used to enhance the chart performance of catching small shifts.

2.1.2 PCA

When a large number of multiple quality characteristics are encountered, Principal Component Analysis (PCA) is often used for dimension reduction. PCA uses an

orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [7]. Jiang and Yan [10] proposed a tool to monitor multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability and Bayesian inference. Tong and Yan [11] studied a modified multiblock PCA algorithm for decentralized Statistical Process Monitoring. In another study, Liu et al. [12] explored statistical process monitoring with integration of data projection and one-class classification using PCA. In a different study, Zheng et al. [13] studied a time series model coefficients monitoring approach for controlled processes. Jiang et al. [14] proposed distributed PCA process with the help of fault-relevant variable selection and Bayesian interference. In another study, Gajjar et al. [15] proposed to detect faults with the help of Sparse Principal Component Analysis. Gajjar and Palazoglu [16] also proposed a data driven multidimensional visualization technique to process fault detection and diagnosis. Yan et al. [17] proposed a robust multivariate process monitoring via stable principal component pursuit. Their goal was to increase the PCAs robustness against gross outliers. In a different study, Jiang et al. [18] investigated Gaussian mixture model and optimal principal component based Bayesian method for multimode fault diagnostics.

Most of these studies showed that their methods work well with Tennessee-Eastman process that has 52 process variables. Tennessee-Eastman process is a benchmark process that consists of five main process units: a two-phase reactor where an exothermic reaction occurs, a separator, a stripper, a compressor, and a mixer. This is a nonlinear open-loop unstable process that has been used in many studies as a case study for plant-wide control, statistical process monitoring, sensor fault detection, and identification of data-driven network models. However, it is unclear whether these methods are capable of scaling up for a production facility with more than a thousand process parameters.

2.2 Group Control Charts

The control charting methods reviewed so far often focus on a single product with one or multiple quality characteristics. Boyd [19], recognizing the need for applying process monitoring for multiple-stream processes, proposed the use of group control charts. A multiple stream is defined as multiple input sources of the same product. Any chart in a group is based on a pair of average (i.e. X-bar) and range (i.e. R) charts. All streams are sampled and each is monitored by a pair of X-bar and R charts. The group control chart framework only records the largest, smallest mean, and the maximum range of the streams with the understanding that if these are within the control limits, the other streams must be too. Specifically, this group control chart method adopts the idea of monitoring the worst-case scenarios [7].

2.3 Precontrol

Precontrol is a technique used to detect shifts or upsets in the process that may result in the production of nonconforming units [7]. The technique differs from a statistical process control in that conventional control charts are designed for real-time monitoring while precontrol is mainly used to assure process capability. Precontrol uses the normal distribution in determining changes in the process mean or standard deviation that could result in increased production of nonconforming units. Only three statistics related to green, yellow, and red zones are required to provide control information as shown in Fig. 1.

The original precontrol method assumes that the process is normally distributed and the natural tolerance limits ($\mu \pm 3\sigma$) exactly coincide with the specification limits. Therefore, this process produces 0.27% of fallout in the red zone. The control works by setting an upper and lower process control limit at ¼ and ¾ of the specification limits. Two consequent samples are drawn and compared to the UPCL and LPCL. If they fall within the green zone, the process as not changed. If two samples fall to the same side of between control limit and specification limit (i.e. the yellow zone), the mean might have changed. If the samples are on the opposite sides between control and specification limits, the standard variation might have changed. If they are outside the specification limits (i.e. the red zone), the process has produced non-conforming parts. Comparing the traditional control charts, precontrol needs to be executed at the beginning of a shift and about six times during a shift for quality assurance purposes. In addition, traditional control charting usually requires a pair of charts—one for mean shift and the other for variance changes while precontrol only requires one chart. It is not an emphasis in precontrol chart to connect the dots of consecutive samples like tradition control chart do. This proposed work uses the color scheme of precontrol to classify samples, while defining different limits for determining the classification of each sample. The upper and lower control limit on the proposed method are set at 1 sigma from the mean rather than 1.5 as in the traditional method. Users can set their desired limits when dealing with cases where more precision is required such as 6 sigma processes.

3 A Proposed Framework for System/Enterprise-Level Monitoring

The following framework is a proposed method for enterprise-level system monitoring. Note that existing SPC methods such as PCA are usually limited to approximately 50 variables at once or need to solve large variance-covariance matrix (Hotelling T^2). There are also possibilities of using machine learning algorithms, artificial neural networks or other methods, which are usually computer-time consuming. The proposed method can be used as a tool to reduce the size of the dataset before it is analyzed with machine learning algorithms.

An enterprise-level system monitoring method should possess that following properties:

1. Able to detect there is a change in the system of interest
2. Able to detect changes in both QCs and their corresponding process parameters
3. Able to detect the location of the change
4. Able to detect the timing of the change
5. Able to work with different types of data: continuous, profile and binary data
6. Able to be easily modeled (or model free) and implemented
7. Able to be scaled up for big data applications

Enterprise level system monitoring is assumed to aid different levels of managers. The department head is interested if all machines in the department are performing as expected or on the same level as previously, the factory manager is interested if all the departments are performing on needed levels, the area manager is interested if all the factories are performing on the needed level etc. Therefore, the monitoring does not focus on finding the cause of the change, but simply on detecting it. The machine operator or process engineer is also expected to correct any deviations promptly. The monitoring system allows to detect changes on the level that particular manager or investigator is interested in and gives a starting point for further analysis.

With the growth of sensors in a system, all the process parameters and quality characteristics can be monitored and treated as variables. This method takes into account all variables and monitors all of them. It is data driven. No knowledge of distribution is needed as the basis of the proposed method is to compare results of period currently under investigation to results from previous periods. Two main strategies are followed to maintain the scale of the problem formulated. First, only the variables that exhibit changes according the precontrol rules are flagged for further investigation. Second, values of all variables in different data types are transformed into a standard scale so that changes are easy to identify with familiar units.

The proposed method is a two-layer method: first layer is the bottom-level entities such as individual machine performance and the second layer is the aggregation of results over all bottom-level entities such as machines at the department level or factory level. The same application can be applied to department and factory level and so on. The machine level layer looks at all of the variables connected to the particular machine (process parameters, quality characteristics) and categorizes them based on “distance from target”. This can be used with both continuous data and profile data. Then the output of the variables is categorized. Summarization is done by using group control chart idea. The base here is to look at all the variables for each timeframe, that is, samples are taken (ten per second, every second, or every minute, etc.), and then the worst outcome is chosen to represent the status of the machine at that timepoint.

The higher level generates summary statistic values over time periods and machine groups that the user has identified. These statistic values are then compared with similar values from previous time periods and conclusions are made based on the results of the comparison.

When change is detected on a higher level, indexing is used to change the resolution of the report to identify the location of the change (machine, department) and the timing of the change. More precise analysis with much smaller dataset can then be started to identify cause of the change.

3.1 Formulation of the Proposed Method

Consider a small production system of n_j departments each containing n_k machines. The machines generate raw data on quality characteristics and process parameters. Each machine generates n_l variables. Assume that the system produces two different products that have different process paths and different target values for each quality characteristic and process parameter.

The proposed method consists of five steps (Fig. 2). The main purpose is to identify if there has been a change, when the change occurred and where the change occurred. This is achieved by using the indexing method and different resolutions of time and space. The indexing method uses the timestamps associated with each measurement to assign time and location identification (machine ID, department ID and factory ID) to all measurements.

The first step is to transform all the raw data into a distance measure from the target. The assumptions are that all the process parameters are continuous variables and quality characteristics have target values. The calculation of the target is as follows

$$d_{jklt} = \left| x_{jkltp} - Ta_{jklp} \right| \Big/ s_{jklp} \tag{1}$$

where

d—distance
x—raw continuous data measurement
Ta—target value
s—target standard deviation
j—index for departments in each factory ($j = 1,2,\ldots, n_j$)
k—index for machines in each department ($k = 1,2,\ldots, n_k$)
l—index for variables in each machine ($l = 1,2,\ldots, n_l$)
t—index for count of measurement ($t = 1,2,\ldots, n_t$)
p—product identifier

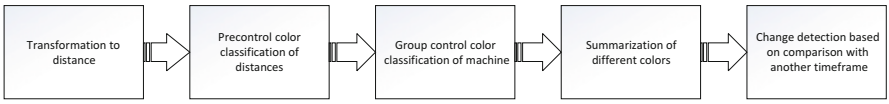


Fig. 2 The flowchart of the proposed method

The second step is to classify each distance from the measurement using precontrol idea. The classification assigns one of four colors to each distance based as follows:

$$c_{jklt} = \begin{cases} green, if the\ d_{jklt}\ is\ within\ one\ target\ standard\ deviation\ from\ 0\ for\ product \\ yellow, if\ the\ d_{jklt}\ is\ between\ 1\ and\ 3\ target\ standard\ deviations\ from\ 0\ for\ product \\ red, if\ the\ d_{jklt} is\ more\ than\ 3\ times\ the\ target\ standard\ deviation\ or\ machine\ is\ down \\ white, if\ the\ machine\ is\ scheduled\ to\ be\ down \end{cases}$$

The setup of the limits can be done using different principles. In the presented case higher value was assigned to more precision, so anything under one sigma shift was rewarded with green in the proposed method. The classification limits could also be set up based on the original precontrol charts, where the "acceptable" area is divided equally between green and yellow; and the division is at 1.5 times standard deviation.

The third step uses the group control chart idea to offset within machine variable dependency. The worst classification will be reported. The machine will be assigned into category for each sample row *ti* based on the following rules.

Green count

$$ti^{g}_{jkt} = \begin{cases} 1, when\ all\ the\ c_{jklt} = green \\ 0, otherwise \end{cases} \tag{2}$$

Yellow count

$$ti^{y}_{jkt} = \begin{cases} 1, when\ at\ least\ one\ of\ the\ c_{jklt} = yellow\ and\ none\ in\ red \\ 0, otherwise \end{cases} \tag{3}$$

Red count

$$ti^{r}_{jkt} = \begin{cases} 1, when\ at\ least\ one\ of\ the\ c_{jklt} = red\ or\ unexpexted\ stop \\ 0, otherwise \end{cases} \tag{4}$$

White count

$$ti^{w}_{jkt} = \begin{cases} 1, when\ the\ machine\ has\ a\ scheduled\ stop \\ 0, otherwise \end{cases} \tag{5}$$

The fourth step is to summarize all color counts and generate a statistic that is used for comparison and detection of changes. The summarization is completed as follows:

All the available counts:

$$T = \sum_{t=1}^{n} \sum_{j=1}^{n_j} \sum_{k=1}^{n_k} ti_{jkt} \tag{6}$$

Count of each color category:

$$T^c = \sum_{t=1}^{n} \sum_{j=1}^{n_j} \sum_{k=1}^{n_k} ti^c_{jkt} \tag{7}$$

where c is the index of the color, either *g* for green, *y* for yellow, *r* for red, or *w* for white.

The statistic is ratio between the counts of each color category divided by the overall available count of samples. Based on the calculations following table is created for each period under investigation.

Green	T^g/T
Yellow	T^y/T
Red	T^r/T
White	T^w/T

Step 4—the final step is to detect changes. The previous four steps are used on the period under investigation and also on two other periods, for example average values of the ratios over year-to-date and average values of the ratios over historical data from the same periods over the previous years. When the change is detected (i.e. the percentages of colors are different between the periods), the location and the period of change can be pin pointed by segmenting the data further and running the same method. For example, if the time frame under investigation was week, the segmentation would be to generate the same table for each day of that week and for each department of the system. Comparison would show when and where the change occurred.

Change Detection Criteria The selection of the criteria depends on many aspects, such as availability of historic data, the distributions of variables etc. In the case of historic data, the user can choose timeframe that is known to be acceptable and generate acceptable thresholds based on those timeframes and compare the results with the results under investigation. This process is very similar to Phase I operation of control charting. The examples in this chapter were generated by assuming normal distribution for each variable. The results showed that the percentage of red was the best indication of change. Depending on if there were ten or 20 variables grouped together in the third step, the threshold for change was 3.417% and 6.305% of red respectively. The process behind these suggestions can be found in the simulation study section of this chapter (Sect. 4). In practice, these thresholds should come from the computation based on a historical data set with the considerations of both Type I and Type II errors.

3.2 Examples

The following samples are presented to show how the proposed method works. First example shows the step-by-step process of how to locate the change and the second example shows how to pin-point the timeframe of the change.

3.2.1 Example 1

Assume there is a small system of three departments that are producing two products. Each department has three machines which have six critical process variables each. The goal is to detect if the system is operating on the same level 1 day as it did on the previous day. The proposed system-wide monitoring framework is implemented for the assessment.

An example of one machine from one department of raw data and target value is in Table 1. Process variables x1 . . . x6 are raw data for each of the six variables. TA1 . . . TA6 are the target values for each variable. Ten sample periods are shown in this table. The measurements are assumed to be recorded at the same time for each variable in this case for the sake of simplicity, but in real life that does not have to be the case. This will be addressed in the discussion section (Sect. 5).

The distance for each variable is calculated based on the equation in the first step of the proposed method. The results are listed in Table 2.

The distances are classified based on the logic presented in the second step of the proposed method using the pre-control methodology. In this example case, all the values for all process variables for Machine 1 are within one standard deviation from the target except for variable 1 at row 9 in which three or more standard deviations is away from the target. Table 3 show the color classifications c1, c2, . . . , c6 for corresponding process variables of Machine 1.

In the next step—the third step of the proposed method, each machine is given an overall color classification based on the group control chart idea. The worst classification across all process variables of a machine is chosen as the performance for a machine. Table 4 show the overall color classification for machine one for each sample.

Next the overall available number of samples is calculated based on the Eq. (2) in the fourth step of the proposed method as well as the count of each color classification found in the sample list. In this case for Machine 1: $\sum ti^{g}_{jkt} = 9$and $\sum ti^{r}_{jkt} = 1$. The results for Machine 1 are then summarized according to Eqs. (6) and (7) shown in Table 5.

Note that Table 4 is the result of one machine while the overall system would generate the following Table 6.

Table 1 Raw data example of one machine

Sample period no.	x1	TA1	x2	TA2	x3	TA3	x4	TA4	x5	TA5	x6	TA6
1	20.645	20.000	81.491	100.000	48.659	50.000	1079.103	1000.000	69.268	70.000	71.991	66.000
2	19.062	20.000	93.879	100.000	51.037	50.000	1039.937	1000.000	69.828	70.000	66.385	66.000
3	19.690	20.000	101.773	100.000	51.288	50.000	1021.800	1000.000	69.210	70.000	68.343	66.000
4	20.441	20.000	94.206	100.000	53.557	50.000	1060.937	1000.000	69.359	70.000	67.117	66.000
5	20.654	20.000	113.126	100.000	49.391	50.000	972.231	1000.000	68.941	70.000	65.596	66.000
6	20.403	20.000	87.734	100.000	49.528	50.000	1037.427	1000.000	71.952	70.000	62.247	66.000
7	20.475	20.000	93.849	100.000	49.716	50.000	933.001	1000.000	69.856	70.000	63.205	66.000
8	21.967	20.000	91.667	100.000	51.327	50.000	1010.222	1000.000	69.861	70.000	62.996	66.000
9	14.953	20.000	83.017	100.000	49.024	50.000	1039.758	1000.000	71.692	70.000	65.149	66.000
10	20.532	20.000	117.300	100.000	47.953	50.000	1070.887	1000.000	70.254	70.000	66.333	66.000

Table 2 The results of distance calculations based on the first step of the proposed method. d1 . . . d6 are the distance values for each variable

Sample period no.	d1	d2	d3	d4	d5	d6
1	0.645	18.509	1.341	79.103	0.732	5.991
2	0.938	6.121	1.037	39.937	0.172	0.385
3	0.310	1.773	1.288	21.800	0.790	2.343
4	0.441	5.794	3.557	60.937	0.641	1.117
5	0.654	13.126	0.609	27.769	1.059	0.404
6	0.403	12.266	0.472	37.427	1.952	3.753
7	0.475	6.151	0.284	66.999	0.144	2.795
8	1.967	8.333	1.327	10.222	0.139	3.004
9	5.047	16.983	0.976	39.758	1.692	0.851
10	0.532	17.300	2.047	70.887	0.254	0.333

Table 3 Pre-control classification for Machine 1

Sample period	c1	c2	c3	c4	c5	c6
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

Table 4 Overall color code for Machine 1

Sample period	ti
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

3.2.2 Example on Detecting the Time and Location of Change

The proposed method is also designed to identify the location and the time of changes by changing the resolution of the output. Let's assume there is a department with 10 machines that are producing different products. The goal is to determine if the department is producing on the similar level as previous day. Over an 8 h shift

Table 5 The final results in term of percentage of samples in each color zone

Level	Statistic
	10.0%
	0.0%
	90.0%
IDLE	0.0%

Table 6 The system-wide look at the performance

Sample period	Machine 1	Machine 2	Machine 3	Machine 4	Machine 5	Machine 6	Machine 7	Machine 8	Machine 9
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									

Table 7 Comparison of 2 days of production of the department overall

today	Level	Statistic	previous day	Level	Statistic
	RED	4.15%		RED	3.20%
	YELLOW	85.22%		YELLOW	86.72%
	GREEN	10.63%		GREEN	10.08%
	IDLE	0%		IDLE	0%

Table 8 Hour-to-hour production of the whole department to provide higher resolution

8:00-9:00	Level	Statistic	9:00-10:00	Level	Statistic	10:00-11:00	Level	Statistic	11:00-12:00	Level	Statistic
		3.0%			3.15%			2.95%			3.50%
		86.170%			86.770%			86.850%			86.600%
		10.63%			10.08%			10.20%			9.90%
	IDLE	0%		IDLE	0%		IDLE	0%		IDLE	0%
12:00-13:00	Level	Statistic	13:00-14:00	Level	Statistic	14:00-15:00	Level	Statistic	15:00-16:00	Level	Statistic
		3.17%			3.40%			7.20%			7.00%
		87.080%			86.040%			81.900%			82.150%
		9.75%			10.56%			10.90%			10.85%
	IDLE	0%		IDLE	0%		IDLE	0%		IDLE	0%

the data is collected, targets are deducted from the raw data, all the variables and time points are classified into color classes with the help of pre-control part of the proposed method. Then the machines are assigned to color classes based on the group control part. All the data is summarized and the output for the department is reported in Tables 7, 8, and 9.

In Table 7, the red percentage has changed from 3.2% to 4.15% which means more measurements were beyond three times the deviation from the target value.

Table 9 Time 14:00–15:00 based on individual machines to provide higher resolution

Machine 1	Level	Statistic	Machine 2	Level	Statistic	Machine 3	Level	Statistic	Machine 4	Level	Statistic	Machine 5	Level	Statistic
		2.83%			3.18%			3.40%			3.22%			45.10%
		86.400%			86.780%			86.780%			85.810%			45.130%
		10.77%			10.04%			9.82%			10.97%			9.77%
	IDLE	0%		IDLE	0%		IDLE	0%		IDLE	0%		IDLE	0%
Machine 6	Level	Statistic	Machine 7	Level	Statistic	Machine 8	Level	Statistic	Machine 9	Level	Statistic	Machine 10	Level	Statistic
		2.60%			3.17%			3.18%			2.65%			2.67%
		86.790%			87.320%			86.710%			86.430%			87.180%
		10.61%			9.51%			10.11%			10.92%			10.15%
	IDLE	0%		IDLE	0%		IDLE	0%		IDLE	0%		IDLE	0%

The recommendations for the decision criteria will be discussed under the simulation portion of this chapter. Obviously there has been a change somewhere and some time. To determine when the change happened, the results are viewed in higher resolution. Specifically, the categorized machine output is divided into hourly blocks with the help of timestamps and presented in the series of outputs.

In Table 8, it is clear that the changed happened somewhere between 14:00 and 15:00 in the department. The next step would be to identify the machine/machines that are responsible for the change. Since the time of the change is known, the time is limited only to that slot and the categorized machine output is summarized over that time slot. The machines are not summarized into department level to identify the culprit. The results are presented in series of outputs.

In Table 9, the results show that machine 5 has started to produce higher percentages of signals that are classified as red and now more thorough analysis of the reasons behind that can start with much smaller time window.

4 Simulation Studies

Three simulation studies were carried out to identify different characteristics of the proposed method. The goal of the first study is to determine what threshold should be used for making a decision if there has been a change. The second study aims to determine the maximum number of variables that can be grouped together in the third step of the method. Finally, the third study determines how sensitive this method is. All the simulation studies were carried out with an assumption that the data was normally distributed for simplicity and demonstration purposes. In real life applications the distribution does not have to be predetermined and the decisions would be made based on the comparison with historic data.

4.1 Determining Threshold for Decision Making

Let's assume there is a machine with ten variables. Each variable measures different parameter of the machine. There are 3600 data points in each variable. For

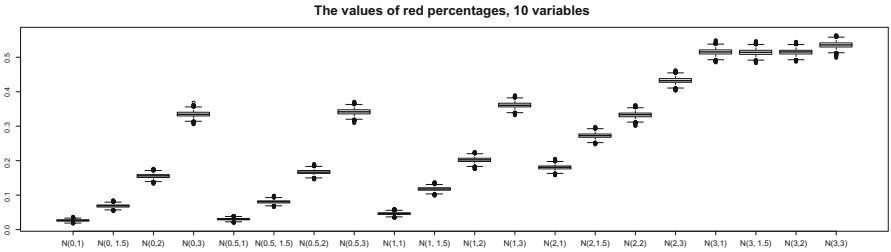


Fig. 3 Boxplots red zone percentages from various out-of-control situations at machine level

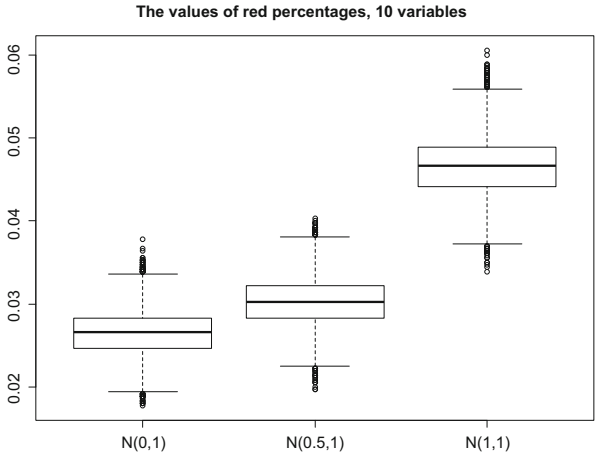


Fig. 4 Percentage of red zone percentages of processes N(0,1), N(0.5,1), and N(1,1)

simplicity, all the variables have been normalized so that the unchanged variable would be normally distributed with N(0,1). In order to determine the threshold for change detection and Type II errors for different scenarios all but one variable were left at N(0,1) and one was changed according to series of parameter changes. The changes introduced were mean shifts of 0, 0.5, 1, 2 and 3 while standard deviations' changes of 1, 1.5, 2 and 3, and the combination of both. After each iteration of the simulation, the proposed method was applied to the new dataset and the results of the color percentages were recorded. Each combination was repeated 10,000 times. One of the additional findings of this simulation was that in this case the red percentage was the best indicator of change. The green and yellow percentages were more random and together mirroring the red color percentage. The results of the red color percentages on the machine level are presented in the following box-plots.

Figure 3 shows all the scenarios on the same graph. As can be seen, most of the scenarios have not overlapping results with the unchanged scenario, which is the process is at N(0,1). The only overlapping scenarios to N(0,1) are N(0.5,1) and N(1,1). Figure 4 shows more detailed look at the results of those three scenarios.

More detailed look reveals that scenario N(1,1) has much less overlaps than N(0.5,1). The smaller change results are much closer to the unchanged variable

Table 10 Type II errors for different scenarios

	Mean				
St.Dev	0	0.5	1	2	3
1		0.914	0.0001	0	0
1.5	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0

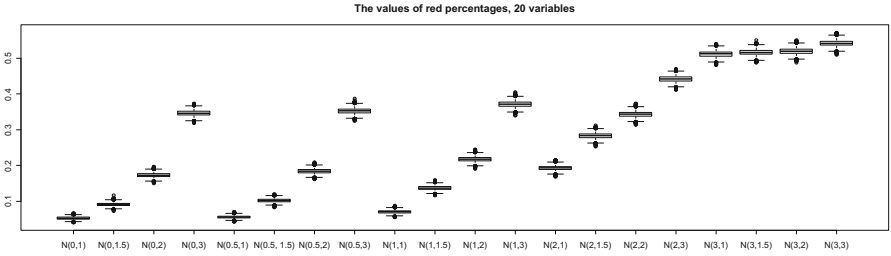


Fig. 5 Boxplots red zone percentages from various out-of-control situations with 20 variables

results. To determine what threshold should be use in the decision making, the overall Type I error will be set at 0.0027. In the case of 10 variables, the threshold of red color percentage would be 3.417%. Anything over that would be considered as changed variable.

Table 10 shows the Type II errors of different scenarios when considering the set threshold. The results are drawn from the 10,000 repetition results.

The table confirms the previous finding that only N(0.5,1) and N(1,1) have overlaps with the unchanged variable and therefore are also the only ones with Type II error. Given the overall type I error at 0.0027, the proposed monitoring method is not capable of detecting a very small process mean shift.

The second part of this simulation study was to determine if the threshold will change when the number of variables in each machine is different. The simulation steps and scenario parameters stayed the same, only this time the machine was assumed to have 20 variables with 3600 data points each and one of those is responsible for changed based on the scenario. Each scenario was repeated 10,000 times. Similarly to the first part, the red color percentage was found to be the best indicator of change. The overall results of the study are in the box-plot of Fig. 5.

Once again only the small mean shifts of N(0.5, 1) and N(1, 1) have seem to have overlapping parts with unchanged variable. More detailed boxplot of the three is provided in Fig. 6.

The threshold in this case is again based on the Type I error being 0.0027. When the machine uses 20 variables and is assumed to be normally distributed, the red percentage threshold for change is 6.305%. Additionally there were two revelations. First, the unchanged variable produces much more "red" colored signals which is explained with Bonferroni curse of dimensions. The threshold is almost two times

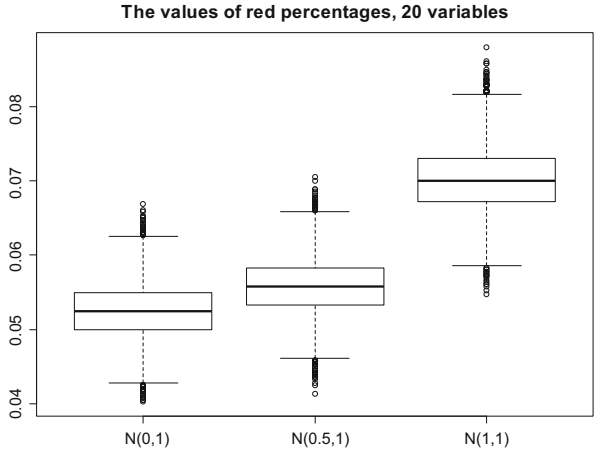


Fig. 6 Percentage of red zone percentages of processes N(0,1), N(0.5,1), and N(1,1) with 20 variables

Table 11 Type II errors of different scenarios with 20 variables

	Mean				
St.dev	0	0.5	1	2	3
1		0.9664	0.0411	0	0
1.5	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0

larger than on the ten variable case. The second is that there is less difference or more overlapping between scenarios.

Table 11 shows the Type II errors of the scenarios based on the 10,000 repetitions and the threshold set previously at 6.305%.

As can be seen, the Type II error for both N(0.5,1) and N(1,1) has increased. The larger changes are detected with 100% accuracy, while small mean shift is virtually undetectable.

4.2 Determining the Max Number of Variables in One Machine

This simulation study was inspired by the first study. The purpose of this study was to determine how many variables can be grouped together in the group control part of the proposed method without losing too much detection power.

For this simulation, the statistics monitored was the percentage of "red" classifications as the previous study indicated to be the best indicator. The scenarios were based on 30–100 variables in 10 variable step. Each number of variables was

Table 12 Type II errors of different combinations of scenarios and numbers of variables with Type I error 0.0027

No of variable	Threshold	N(0,1.5)	N(0.5,1)	N(0.5,1.5)	N(1,1)	N(1,1.5)
30	0.0905	0	0.9735	0	0.0496	0
40	0.1175	0.0002	0.9878	0	0.3711	0
50	0.1417	0	0.9821	0	0.2756	0
60	0.164	0.0085	0.992	0	0.6904	0
70	0.193	0.008	0.993	0	0.574	0
80	0.222	0.03	0.989	0	0.776	0
90	0.24	0.017	0.99	0	0.571	0
100	0.255	0.036	0.986	0.001	0.742	0

Table 13 Type II errors of different combinations of scenarios and numbers of variables with Type I error 0.05

No of var	Threshold	N(0,1.5)	N(0.5,1)	N(0.5,1.5)	N(1,1)	N(1,1.5)
30	0.0852	0	0.7796	0	0.0021	0
40	0.1113	0	0.8445	0	0.0659	0
50	0.136	0	0.8556	0	0.0568	0
60	0.1565	0.0002	0.888	0	0.2274	0
70	0.1844	0	0.866	0	0.137	0
80	0.2119	0.001	0.897	0	0.217	0
90	0.2315	0	0.863	0	0.261	0
100	0.248	0.003	0.907	0	0.353	0

simulated with variable changing N(0,1), N(0.5,1), N(1,1), N(0,1.5), N(0.5,1.5) and N(1,1.5), because the previous study showed that smaller changes are more suspect to be overlapping with the unchanged variable. For 30, 40, 50 and 60 variables, each combination of mean and standard deviation was recorded 10,000 times. For 70, 80, 90 and 100 variables, each combination was recorded 1000 times due to the high demand in computational time.

Tables 12 and 13 show the results for this simulation. Both Type I and Type II errors are considered to reach a balance performance in setting the threshold values. Table 12 shows the threshold values in the second column Type II errors of different number of variables when the Type I error is set at 0.0027 while Table 13 shows the results of Type I errors when the Type II error is set at 0.05.

As shown in Tables 12 and 13, the more variables is used during the group control part of the proposed method, the higher the threshold is for keeping the type I error at 0.0027 and 0.05 respectively. Again, this is explained by the Bonferroni curse of dimensions. What also can be seen is that in each combination there is a trend for Type II error to grow when the number of variables inclines. The more variables is used, the less obvious the difference between unchanged and changed variable is. For the recommendation of how many variables could be used in the group control part of the proposal, a decision criteria must be established. The proposal is to base the decision on the N(1,1) scenario, because in the previous simulation study that

scenario showed small Type II error and it is important change that needs to be captured. The second side of the decision criteria is where to draw the line of what is acceptable. The recommendation here would be to use 5% of Type II error, because at that level, a lot of changes are still detected. Based on the decision criteria, the recommended maximum number of variables to be used in the group control part of the algorithm would be 30 variables for the proposed method to be effective with the consideration of both Type I and Type II errors.

4.3 Determining the Sensitivity of the Proposed Method

This simulation study on the sensitivity of the proposed method aims to determine how sensitive it is by leveraging the knowledge gained from the previous two studies. Assume that there is a factory with ten departments. Each department consists of ten machines and each machine monitors ten variables. Overall, there are 1000 variables to monitor. Each variable has 3600 data points or rows. The simulation is done with a small mean shift of N(1,1), because that was used in the number of variables simulation study decision criteria. In first scenario 1% of random variables changed by N(1,1). This means ten random variables out of 1000. The selection was made with uniformly distributed random number generator. Each variable had an equal chance of being selected. Also the changed variables could have appeared in the same machine. The selected variables changed from iteration to iteration. After each dataset was generated, the proposed method was applied. First the pre-control part of the algorithm classified each data point into either red, yellow or green color. Then the group control part grabbed the worst classification for each row and assigned that classification to the machine for that row. The next steps were to sum the results up on the machine, department and factory level and report the percentages for each color. This was repeated 1000 times. The red color percentage was used for the identification statistic. The same process was carried out with 5% of variables (50 random variables out of 1000) and 10% of variables (100 random variables out of 1000) experiencing changes.

Two thresholds were used for alpha: 3.417% for 0.0027 and 3.111% for 0.05. The results can be found in Table 14. The Type I error reported in this table represents the amount of machines/departments/factory that were labelled as changed but in fact they did not have the changed variable in them. On the other hand, the Type II error are machines/departments/factory that was labelled unchanged but there was a changed component in them.

Table 14 shows that the proposed method is capable of detecting changes at the machine level well. The proposed method failed to detect changes in department and factory level when Type I is 0.0027 in the 1% case. In the case of 5% variables change case, the departmental detection is getting much better. In the 10% case, the factory level and the department level are almost always labelled as changed. The revelation is that on any managerial level the change is brought to attention if it either large change or a lot of small changes.

Table 14 The sensitivity of the proposed method

N(1,1)		Alpha = 0.05		Alpha = 0.0027	
		Type I	Type II	Type I	Type II
1% variables changed	Factory (1)	0.0000	1.0000	0.0000	1.0000
	Department (10)	0.0000	0.6613	0.0000	0.9585
	Machine (100)	0.0612	0.0000	0.0047	0.0010
5% variables changed	Factory (1)	0.00000	0.00000	0.00000	0.00000
	Department (10)	0.00000	0.02536	0.00000	0.14874
	Machine (100)	0.00897	0.00002	0.00074	0.00052
10% variables changed	Factory (1)	0.0000	0.0000	0.0000	0.0000
	Department (10)	0.0000	0.0001	0.0000	0.0009
	Machine (100)	0.0084	0.0000	0.0006	0.0004

5 Discussions

The proposed method is capable of practical implementation in many different environments. The example given in the previous chapter was when all the samples were taken at the same time point. In most real life cases that might not be possible. For example, the measurement of each part in height might happen ten times a second, but temperature is measured only in every 2–3 s. The recommendation would be to assign the less frequent measurement to each row of the sample that is between the less frequent sampling times. In cases where there is so much data that the traditional computational tool cannot handle the data, the use of big data applications might be useful. Another facet of the proposed method worth mentioning is that it is not in-time monitoring method where the sample in hand is the item under investigation. The proposed method can be used as near-time, since the focus is on the results from a period of time.

5.1 Implementation in Enterprise Environment

The proposed method is scalable and can be adapted to a system ranging from a factory to a supply chain. The recommendation of the implementation of the proposed method in a real-world enterprise would consist of following steps:

1. Determine the target value for each variable in the system for each product produced in the facility to calculate the distances from target:
   (a) If the targets can be set by requirements, those could be used
   (b) If not, the historical data must be used to calculate the targets
   (c) Not all process variables are continuous. Some may be discrete and the others may be in a form of a profile. The proposed method introduced in Sect. 3.1 is based on the fraction of time as shown in Eqs. (2), (3), (4), and (5) computed

by the number of samples falling into each color zone. Discrete variables can be easily accommodated using the same set of equation. However, profile variables require additional procedures to convert. It will be introduced in another study.

2. Determine the standard deviations to be used in the pre-control part of the algorithm
   (a) If the standard deviations be set by requirements, those should be used
   (b) If not, the historical data of "good" products should be used. This is often referred as the Phase I of control charting.
3. Group variables together for the group control part of the proposed method:
   (a) If there is more than 30 variables in the machine, then the variables should be separated and grouped together to form subgroups. For example, if the machine has 10 variables that are connected to temperature measurements, 20 connected to speeds and ten connected to the shape of the product, then 40 variables together will not produce good output. However grouping them according to function within the machine would give the desired result.
   (b) While it would be advisable to keep the variable groups similar size, it is not required
4. Run the proposed method on the full set of historic data. This step might help with solving problems in the future. The data should be analyzed for both "good" parts as well as "bad" parts. If in the future a certain set of changes has been detected, then comparing it to historical data and finding similar set of changes with the current issue might help solving the problem and also predicting the quality of the products or even predicting an eminent failure.
5. Start the real time monitoring using the target values from step 1 and deviations from step2
6. For reporting purposes, the recommendation would be to have three output charts side-by-side that show the current period, averages of year-to-date and averages of historical data from similar timeframe (for example: all the Julys in history). This might help understanding trends and also helps in the case of autocorrelated data
7. Different levels of management are assumed to be interested in different level of data. The department head is interested in the machine level report in their department. The factory manager is interested in department level report and CEO is interested in the factory level report. Each level can dig deeper if more precise analysis is needed.
8. Prior knowledge of the distributions of the variables is not required, because the decisions are based on the comparison with the historical data that can be chosen on the basis of "good" products.

   The addition of new machines is not an issue, because all the statistics are generated based on all the available time and they are percentage of that. New machines adds to the total time.

5.2 Implementation in Big Data Environment

MapReduce is a framework for executing highly parallelizable and distributable algorithms across large datasets using hundreds or thousands of commodity computers [20]. A MapReduce algorithm does parts of the calculations in the server that the data segment is stored in parallel (hence the name "parallel computing").

MapReduce consists of two procedures that the user must develop: mapping and reducing. The system manages the parallel execution, coordination of tasks that execute mapping or reducing and also deals with possible failures of some of the tasks execution. In the mapping procedure, the data segment in each server is split, sorted and filtered. If needed, other calculations are also carried out. Users must define two critical parameters that are used as the input and output of each server: key and value. The key is the identification parameter that depends on the goal of the algorithm and the value is the output of the segment in that server. All the key-value pairs are collected by master controller and divided among all Reduce tasks in a way that all the pairs with same key end up in the same Reduce task [21].

In the reducing procedure, the outputs of the mapping procedures are shuffled and sorted based on the key defined in the mapper and then reduced by combining the values defined previously in some manner defined by the user.

The proposed method could be turned into MapReduce function using following logic. The assumption is that all the targets and standard deviations are stored in the top-level computer so they can be accessed by the program at any time. The data is assumed to have timestamp and variable identification for each measurement. In the mapper function the distances are calculated, colors are assigned, the machine level is assigned. The output is timestamp, machine ID and the color assignment.

In the reducer function the summarization of the colors are made based on the level on interest of the user. This is where the resolution of the report is set. The output would be the traffic light for the timeframe and location under investigation. Details of MapReduce algorithm applied to the proposed method will be reported in another study.

6 Conclusions

An enterprise-level monitoring system is proposed. All the raw data is transformed into distances from target. The distances are then classified into colored groups with the help of the precontrol chart idea. All the variables associated with a machine are then used to classify the machine into a color class according to the group control chart idea. The resulting counts are turned into percentages of time and that statistic is used to determine if the process is changed or not. The simulation studies show that the proposed method is capable of handling 1000 variables per department and produce good results with different scenarios. A recommendation for implementation of it in real-life situation is proposed as well as a recommendation for Big Data application with the help of the MapReduce method.

The proposed method has following characteristics:

1. It is capable of detecting changes, also identify the spatial and temporal space of the change
2. It does not use covariance matrix, which makes the calculation much easier
3. It is easily scalable from a few to 1000 process variables
4. It can be extended in Big Data format
5. It does not require previous knowledge of distributions

Future studies include implementation of the proposed method on different types of data, such as profile data and binary data. In addition, the proposed method may be able to be integrated into a maintenance planning—usually a topic traditionally studied from the field of reliability. Finally, the proposed method may be applied to supply chain applications.

References

1. Shewhart, W. A. (1930). Economic quality control of manufactured product. *Bell System Technical Journal, 9*(2), 364–389.
2. Runger, G. C., & Testik, M. C. (2004). Multivariate extensions to cumulative sum control charts. *Quality and Reliability Engineering International, 20*(6), 587–606.
3. 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.
4. Hu, J., Runger, G., & Tuv, E. (2007). Tuned artificial contrasts to detect signals. *International Journal of Production Research, 45*(23), 5527–5534.
5. Hotelling, H. (1947). Multivariate quality control—Illustrated by air testing of sample bombsights. In C. Eisenhart, M. W. Hastay, & W. A. Wallis (Eds.), *Techniques of statistical analysis* (pp. 111–184). New York: McGraw Hill.
6. Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters, 3*, 18–23.
7. Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). Hoboken, NJ: Wiley.
8. Pignatiello, J. J., & Runger, G. C. (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology, 22*, 173–186.
9. Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality control schemes. *Technometrics, 30*, 291–303.
10. Jiang, Q., & Yan, X. (2014). Monitoring multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability and Bayesian interference. *Chemometrics and Intelligent Laboratory Systems, 136*, 121–137.
11. Tong, C., & Yan, X. (2017). A novel decentralized process monitoring scheme using a modified multiblock PCA algorithm. *IEEE Transactions on Automation Science and Engineering, 14*, 1129–1138.
12. Liu, Y., Pan, Y., Wang, Q., & Huang, D. (2015). Statistical process monitoring with integration of data projection and one-class classification. *Chemometrics and Intelligent Laboratory Systems, 149*, 1–11.
13. Zheng, Y., Wang, Y., Wong, D. S. H., & Wang, Y. (2015). A time series model coefficients monitoring approach for controlled processes. *Chemical Engineering Research and Design, 100*, 228–236.

14. Jiang, Q., Huang, B., & Yan, X. (2016). GMM and optimal principal components-based Bayesian method for multimode fault diagnosis. *Computers and Chemical Engineering, 84*, 338–349.
15. Gajjar, S., Kulachi, M., & Palazoglu, A. (2016). Use of sparse principal component analysis (SPCA) for fault detection. *IFAC-PapersOnLine, 49*, 693–698.
16. Gajjar, S., & Palazoglu, A. (2016). A data-driven multidimensional visualization technique for process fault detection and diagnosis. *Chemometrics and Intelligent Laboratory Systems, 154*, 122–136.
17. Yan, J., Chen, C. Y., Yao, Y., & Huang, C. C. (2016). Robust multivariate statistical process monitoring via stable principal component pursuit. *Industrial and Engineering Chemistry Research, 55*, 4011–4021.
18. Jiang, Q., Yan, X., & Huang, B. (2016). Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference. *IEEE Transactions on Industrial Electronics, 63*, 377–386.
19. Boyd, D. F. (1950). Applying the group control chart for –x and R. *Industrial Quality Control, 7*(3), 22–25.
20. Lublinsky, B., Smith, K., & Yakubovich, A. (2013). *Professional Hadoop solutions* (pp. 63–97). Indianapolis, IN: Wiley.
21. Rajaraman, A., & Ullman, J. D. (2012). *Mining of massive datasets* (pp. 21–42). Cambridge: Cambridge Printing House.

Enhanced Cumulative Sum Charts Based on Ranked Set Sampling

Mu'azu Ramat Abujiya and Muhammad Hisyam Lee

Abstract The cumulative sum (CUSUM) control charts are widely used for the monitoring of normal processes for changes in the location and dispersion parameters. This study presents several CUSUM charts designed structures based on the ranked set sampling (RSS) data for overall efficient detection of changes in the process mean and variance. The run-length properties of these charts are examined and compared to the classical CUSUM location and dispersion charts. Results show that the application of RSS technique has significantly improved upon the standard CUSUM chart. Using real RSS data set, we present a practical example of the implementation of the CUSUM schemes.

Keywords Average run length · CUSUM · Quality control chart · Ranked set sampling · Statistical process control

1 Introduction

The history of quality control goes back to the beginning of manufacturing and competition associated with the manufactured products. The scientific foundation of quality control as per how many random samples of products from a production line to inspect and what conclusions to draw from the sampled products gave birth to the statistical process control (SPC). A control chart is a well-known SPC tool used in modern industries to monitor the quality characteristics of processes effectively. It has also found applications in non-manufacturing sectors. The Shewhart control

M. R. Abujiya
Preparatory Year Mathematics Program, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
e-mail: abujiya@kfupm.edu.sa

M. H. Lee (✉)
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
e-mail: mhl@utm.my

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_4

chart proposed by Shewhart [1] is the first control chart and is still very popular among the practitioners of quality control. The chart is based on the most-recent information in the sampled products and completely ignores all the past information and as such only sensitive to large changes in a process. To detect small changes, alternative control charts that put its weight on both the past and the current information in the sampled products, such as the cumulative sum (CUSUM) chart by Page [2] were introduced. The CUSUM charts have been accepted, widely, for monitoring production processes because of their effectiveness in detecting small to moderate changes in a process. Moreover, the Shewhart chart is a special case of the CUSUM chart obtained by setting the decision interval or the control limits equal to zero [3].

To construct a classical two-sided CUSUM control chart, let $x_{1j}, x_{2j}, x_{3j}, \ldots, x_{nj}$; $j = 1, 2, 3, \ldots, m$ be independent and identically distributed random samples of the quality characteristic of interest of subgroup size n. Suppose that the in-control process follows a normal distribution with a known mean μ_0 and standard deviation σ_0. For example, if changes in the process location parameter is of interest, define the mean of the jth sample as $\overline{x}_j = (1/n)\sum_{i=1}^{n} x_{ij}$ and the standardized statistic of x_{ij} to be $z_j = \sqrt{n}\left(\overline{x}_j - \mu_0\right)/\sigma_0$. Then the upper and lower sided CUSUM statistics for monitoring the process mean are defined as:

$$CUSUM_j^+ = \max\left[0, CUSUM_{j-1}^+ + z_j - k\right] \tag{1}$$

$$CUSUM_j^- = \max\left[0, CUSUM_{j-1}^- - z_j - k\right] \tag{2}$$

where $\max[a, b]$ is the maximum of a and b. The constant k is the control chart allowance popularly called the reference value and is determined based on the anticipated shift δ in the process mean level of a quality characteristic. The initial values $CUSUM_0^+$ and $CUSUM_0^-$ are usually set equal to zero. If either $CUSUM_0^+$ or $CUSUM_0^-$ exceeds the control limit $h > 0$, then the process is assumed to have shifted from μ_0 to $\mu_0 + \delta\left(\sigma_0/\sqrt{n}\right)$. That is, the process is out-of-control. Several modifications have been made in the literature to improve the sensitivity of a standard CUSUM chart introduced by Page [2]. However, most of the recommended design strategies are based on random sampling [4].

Recently, the ranked set sampling (RSS) technique has found applications in the development of quality control charts. The scheme is gaining popularity among some investigators in monitoring the process mean and variance of a normal process. See, for example, Muttlak and Al-Sabah [5], Abujiya and Muttlak [6], Al-Nasser and Al-Rawwash [7], Al-Sabah [8], Al-Omari and Al-Nasser [9], Jafari and Mirkamali [11], Al-Omari and Haq [11], Mehmood et al. [12], Haq et al. [13], Mehmood et al. [14], Abujiya et al. [15], Abujiya et al. [16], Abujiya et al. [17], Abujiya et al. [18], Abid et al. [19], Abujiya et al. [20] among others. The RSS procedure suggested by McIntyre [21] in the estimation of population mean of pasture yields uses inexpensive judgment ranking. The samples produce

by the RSS scheme are more representative of a population than the classical random sampling. Particularly, when the exact measurement of variable of interest is expensive, difficult or time-consuming but could readily be ranked visually or by some cost-effective methods that do not require the precise measurements of the variables. The mathematical foundation for the RSS method was developed by Takahasi and Wakimoto [22]. This study presents the construction of several CUSUM location and dispersion charts based on RSS technique and its variations for monitoring normal processes.

2 Designing CUSUM Charts Based on RSS Data

To improve the classical design of some control charts, RSS has achieved some remarkable enhancements. The scheme has particularly been used to increase the sensitivity of a basic CUSUM control chart to monitor the process location parameter as can be seen in Al-Sabah [8], Haq et al. [13], Abujiya et al. [16] among others. Furthermore, Abujiya et al. [15], Abujiya et al. [17] and Abujiya et al. [18] have also used the method to improve the performance of the basic CUSUM dispersion charts. Below is brief background information on RSS technique followed by CUSUM design structure based on the RSS schemes.

2.1 RSS Technique

The RSS methodology has established itself as an efficient data collection method in major research areas, including process monitoring. For a comprehensive review on the theory and applications of RSS, see Chen et al. [23]. More recent contributions can be found in Wolfe [24] and Al-Omari and Bouza [25], and the references therein.

2.1.1 The Setup

The procedure for collecting ranked set samples as suggested by McIntyre [21] can be summarized as outlined below:

(a) Identify n^2 samples from the target population.
(b) Randomly allocate these samples into n sets, each of size n units.
(c) Rank the n units in each set by judgement or any cost-effective method, concerning a variable of interest.
(d) The unit with the lowest rank is selected from the first set, the unit with the second lowest rank is selected from the second set, and so on, until the unit with the highest rank is selected from the last nth set.
(e) The steps (a) through (d) may be repeated m times to collect a balanced RSS sampled data of nm units.

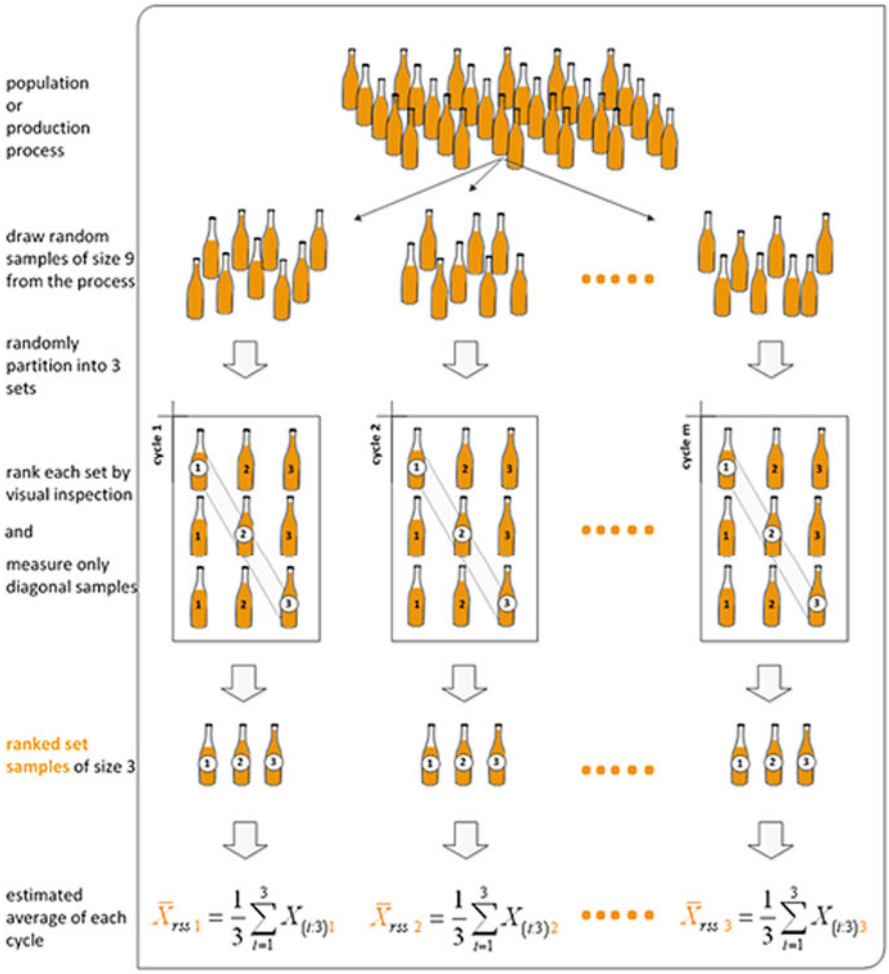


Fig. 1 Design of a ranked set sample on a bottle filling process with set size of $n = 3$

The above procedure is illustrated in Fig. 1, as we consider an example on the filling up of a bottle in a production line of a bottling company. In this example, the variable of interest is the amounts of the liquid inside a bottle, and this could be a soft drink, mineral water, vegetable oil, milk, honey, paint, etc. Considering a sample size is three, randomly collect nine bottles from the production line and divide them into three sets, each of size three. Using the level of the liquid, visually rank the bottles in each of the three sets from the least to the highest filled. Repeat the process and the main diagonal bottles, of each cycle, constitute the RSS data.

2.1.2 Location and Dispersion Parameters

Let $X_{11}, X_{12}, \ldots, X_{1n}, X_{21}, X_{22}, \ldots, X_{2n}, \ldots, X_{n1}, X_{n2}, \ldots, X_{nn}$ be independent random samples each of size n and assume that all samples have the same cumulative distributive function $F(x)$, with mean μ and variance σ^2. Suppose $X_{i(1:n)}, X_{i(2:n)}, \ldots, X_{i(n:n)}$ is the corresponding order statistic of $X_{i1}, X_{i2}, \ldots, X_{in}$ for $i = 1, 2, \ldots, n$. Then $X_{(1:1)}, X_{(2:2)}, \ldots, X_{(n:n)}$ is the RSS data for one cycle. For m repeated cycle, denote the ith order statistic of the ith sample of size n in the jth cycle by $X_{(i:n)j}$. The RSS estimate for the location parameter μ, is an unbiased estimator regardless of the judgement ranking errors [22]. It is defined by

$$\overline{X}_{rss} = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n} X_{(i:n)j}, \tag{3}$$

and its variance is given to be

$$\mathrm{Var}\left(\overline{X}_{rss}\right) = \frac{1}{nm}\left[\sigma^2 - \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(\mu_{(i:n)j} - \mu\right)^2\right]. \tag{4}$$

Here, $\mu_{(i:n)}$ is the mean of the i^{th} ranked set $X_{(i:n)}$ which is defined as

$$\mu_{(i:n)} = \int_{-\infty}^{\infty} x \frac{n!}{(i-1)!\,(n-i)!} F^{i-1}(x)[1 - F(x)]^{n-1} f(x)\, dx, \tag{5}$$

where $f(x)$ is the probability density function of the random samples $X_{i1}, X_{i2}, \ldots, X_{in}$; $i = 1, 2, \ldots, n$ [26].

For the measure of variation, the RSS estimate for the process standard deviation σ, is an unbiased estimator [27], and is given by

$$S_{rss} = \sqrt{\frac{1}{nm - 1 + \gamma_n}\sum_{j=1}^{m}\sum_{i=1}^{n}\left[X_{(i:n)j} - \overline{X}_{rss}\right]^2}, \tag{6}$$

where $\gamma_n = (1/n)\sum_{i=1}^{n}\upsilon_{(i:n)}^2$ is a correction constant and $\upsilon_{(i:n)} = [\mu_{(i:n)} - \mu]/\sigma$. The numerical values of $\upsilon_{(i:n)}$ for set sizes $n = 2, 3, 4, 5, 6$ and 10 can be found in Yu et al. [28]. The expectation and standard deviation of S_{rss} are defined as $E(S_{rss}) = c_4^*\sigma$ and $\sigma(S_{rss}) = c_5^*\sigma$, respectively, where c_4^* and c_5^*, are the S chart constants [15] which are sample size dependent.

Many researchers have shown that the use of sample range as a measure of variation is equally effective but probably not as effective as the standard deviation which is the natural measure of process dispersion. The choice of range is because of its simplicity to compute. Thus, we define

$$R_{rssj} = \left|X_{(n:n)j} - X_{(1:1)j}\right|, \tag{7}$$

to be the sample range of set size n in the jth cycle of RSS data, and the average as

$$\overline{R}_{rss} = \frac{1}{m}\sum_{j=1}^{m} R_{rssj}. \tag{8}$$

We also define the mean and the standard deviation of R_{rss} to be $E\left(R_{rssj}\right) = d_2^*\sigma$ and $\sigma\left(R_{rssj}\right) = d_3^*\sigma$, respectively, where d_2^* and d_3^* are the R chart parameters [15]. Like c_4^* and c_5^*, the constants d_2^* and d_3^* are also sample size n dependent.

2.1.3 Imperfect Ranking

In RSS setup, the ranking of the variable of interest X may not always be perfect as a result of wrong human judgments. Thus, an imperfect ranking of units or ranking based on the auxiliary variable, denoted by Y, may be unavoidable. In such case, the values of the auxiliary variable Y are used to rank the values of the main variable X. The usual practice is to use an auxiliary variable that is highly correlated with the variable of interest as the accuracy of the ranking depends on the linear relationship between X and Y. To study the effect of imperfect ranking, assume that the pair (X, Y) follows a bivariate normal distribution and that the regression of X on Y is linear. Thus, by Stokes [29] we have

$$X = \mu_X + \rho_{XY}\left(\sigma_X/\sigma_Y\right)\left(Y - \mu_Y\right) + \varepsilon, \tag{9}$$

where μ_X, μ_Y are the process means of X and Y; σ_X, σ_Y are the process standard deviations of X and Y; ρ_{XY} is the correlation between X and Y; and ε is the error term which is independent of Y. Here, ε is assumed to have a mean of zero and standard deviation $\sigma_X\sqrt{1-\rho_{XY}^2}$. For the corresponding RSS model, let $X_{[i:n]j}$ denotes the ith judgment order statistic of the study variable X based on auxiliary variable Y, then the above equation can be written as

$$X_{[i:n]j} = \mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}\left[Y_{(i:n)j} - \mu_Y\right] + \varepsilon_{ij}, \tag{10}$$

where $Y_{(i:n)j}$ are the perfectly ranked ordered statistics on Y. Thus, we can replace $X_{(i:n)j}$ in Eqs. (3), (6) and (7) by $X_{[i:n]j}$ to obtain the corresponding imperfect RSS estimators for the process mean, standard deviation and range. Note that imperfect ranking is a special case of perfect ranking when $\rho_{XY} = 1$.

2.1.4 RSS Variants

In practice, an increase in the relative efficiency of RSS scheme is achievable when the variations such as the median RSS (MRSS), and the extreme RSS (ERSS) are used instead of the balanced scheme. These variants of RSS are easier to execute and

have fewer errors associated with ranking of units. The MRSS scheme suggested by Muttlak [30] involves the measurement of the middle-most value (*for odd set size*) or two middle-most values (*for even set size*) in step *d* of Sect. 2.1.1. MRSS is an outlier-resistant scheme that gives a better and more robust location estimator than the balanced RSS, particularly for samples that contain outliers. The MRSS estimators for the population mean for odd and even set sizes are respectively given by:

$$\overline{X}_{mrss\ o} = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n} X_{(i:(n+1)/2)j} \tag{11}$$

$$\overline{X}_{mrss\ e} = \frac{1}{nm}\sum_{j=1}^{m}\left[\sum_{i=1}^{n/2} X_{(i:n/2)j} + \sum_{i=\mathrm{n}/2+1}^{n} X_{(i:(n+2)/2)j}\right], \tag{12}$$

where $X_{(i:(n+1)/2)}$ is the *i*th median for an odd set size and, $X_{(i:n/2)}$ and $X_{(i:(n+2)/2)}$ are the middle-most values for the *i*th set of an even set size. For more information, see Muttlak [30].

The ERSS is also a variant of RSS that was suggested by Samawi et al. [31]. Like the MRSS, the ERSS does not require complete ranking of units. The scheme involves the quantification of the smallest and largest ranked units in step *d* of Sect. 2.1.1 (*for even set size*). If the set size is odd, the median observation is measured from one set in addition to the extreme values from the other sets. Shaibu and Muttlak [32] reported that if the underlying distribution normal, the ERSS is more efficient in estimating the process variance than the balanced RSS scheme. Sinha and Purkayastha [27] define the unbiased estimator for the process standard deviation as

$$S_{erss} = \sqrt{\frac{1}{n-1+\gamma_n}\sum_{j=1}^{m}\sum_{i=1}^{n}\left[X_{(i:e)j} - \overline{X}_{erss}\right]^2}, \tag{13}$$

where $X_{(i:e)j}$ is the *i*th ERSS observation in the *j*th cycle and $\overline{X}_{erss} = (1/n)\sum_{j=1}^{m}\sum_{i=1}^{n} X_{(i:e)j}$ is the ERSS estimator for the process mean. The constant $\gamma_n = (1/n)\sum_{i=1}^{n}\upsilon_{(i:e)}^2$, is a known constant, with the values of $\upsilon_{(i:e)} = [\mu_{(i:e)} - \mu]/\sigma$ presented in Yu et al. [28].

2.2 CUSUM Location Charts Based on RSS

In constructing a CUSUM chart to monitor the process location parameter, it is normally assumed that the process is initially stable (in-control) with mean μ_0 and standard deviation σ_0, and becomes unstable (out-of-control) at some point in time with a shift in mean from μ_0 to $\mu_1 = \mu_0 + \delta\sigma_0/\sqrt{n}$. Here, $\delta = \sqrt{n}\,|\mu_1 - \mu_0|\,/\sigma_0$ is the amount of shift in the process mean. Define the standardized RSS location

parameter to be $Z_{rss\ j} = \left(\overline{X}_{rss\ j} - \mu_0\right)/\sigma_{\overline{X}_{rss}}$, and the two-sided CUSUM statistics for monitoring increases and decreases in the process mean level [6, 16] are respectively, given by

$$L^+_{rssj} = \max\left[0, L^+_{rssj-1} + Z_{rssj} - k\right] \tag{14}$$

$$L^-_{rssj} = \max\left[0, L^-_{rssj-1} - \mathrm{Z}_{rssj} - k\right], \tag{15}$$

where $k = (\mu_1 - \mu_0)/2$ is called the reference value. It is customary to set the initial values of the CUSUM statistics $L^+_{rss\ 0}$ and $L^-_{rss\ 0}$ equal to zero. Whenever $L^+_{rss\ j}$ or $L^-_{rss\ j}$ exceed the pre-determined control limit denoted by h, the CUSUM control chart signals. The h is often selected to give a specified in-control run length property.

2.3 CUSUM Dispersion Charts Based on RSS

In monitoring changes in the process variability, we assume that the process is initially in-control with mean μ_0 and standard deviation σ_0. After some time, the process goes out-of-control with a change in process standard deviation from σ_0 to $\sigma_1 = \tau\sigma_0$, where $\tau = \sigma_1/\sigma_0$ is the amount of shift in the process standard deviation. The measure of variation considered in this study, includes the sample range and standard deviation. Let $V_{rss\,j}$ denote the either of the RSS based variation measure R_{rssj} or S_{rssj}. We define the standardized two-sided CUSUM statistics to monitor the process dispersion [15] as

$$D^+_{rssj} = \max\left[0, D^+_{rssj-1} + \left(V_{rssj}/\sigma_0\right) - k\right] \tag{16}$$

$$D^-_{rssj} = \max\left[0, D^-_{rssj-1} - \left(V_{rssj}/\sigma_0\right) + k\right], \tag{17}$$

where $D^+_{rss\ 0} = D^-_{rss\ 0} = 0$. Here, we have $k = d_n(1+\tau)/2$ with $d_n = \overline{R}_{erss}/\sigma_0$ for the R chart and $k = c_n(1+\tau)/2$ where $c_n = (1/\sigma_0)E(S_{rss})$ for the S chart. Thus, the RSS CUSUM dispersion chart gives a signal when either $D^+_{rssj} > h$ or $D^-_{rssj} > h$. Using simulation of 10^6 replicated iterations, Abujiya et al. [15] gave the estimated values of d^*_2, d^*_3, c^*_4 and c^*_5 for different sample sizes and RSS schemes which are partly re-produced in Table 1.

Table 1 *R* and *S* chart constants for RSS and ERSS of normal order statistics

	d_2^*		d_3^*		c_4^*		c_5^*	
Sample size *n*	RSS	ERSS	RSS	ERSS	RSS	ERSS	RSS	ERSS
2	1.327	1.327	0.937	0.937	0.817	0.817	0.577	0.577
3	1.929	1.929	0.907	0.907	0.906	0.906	0.422	0.422
4	2.305	2.850	0.855	0.861	0.941	0.958	0.339	0.288
5	2.574	3.090	0.813	0.819	0.958	0.969	0.286	0.248
6	2.781	3.640	0.781	0.754	0.969	0.982	0.249	0.191
7	2.949	3.779	0.757	0.736	0.975	0.985	0.220	0.174
8	3.090	4.139	0.738	0.695	0.980	0.990	0.199	0.144
9	3.210	4.236	0.722	0.684	0.983	0.991	0.182	0.136
10	3.315	4.498	0.708	0.657	0.986	0.993	0.167	0.117

3 Performance Measure

Statistical performance of control charts is usually evaluated using the average run length (ARL). The ARL is the mean of the run-length distribution which is the average number of samples taken until an out-of-control signal is given. The standard deviation of the run length (SDRL) distribution is also often used as a supplementary measure of performance. Together, the two metrics measures how quickly a control chart to respond to changes is a process. The run-length properties of a CUSUM chart can be computed either by the approximation of an integral equation, Markov Chain or Monte Carlo simulation. Following Hawkins [33], Chang and Gan [34], Mehmood et al. [12], Haq [35] among others, this study uses the Monte Carlo simulation approach through an algorithm developed in FORTRAN.

3.1 Measuring Indices

An efficient CUSUM control chart is one that has sufficiently large in-control ARL (ARL_0) and SDRL ($SDRL_0$) values and small out-of-control ARL (ARL_1) and SDRL ($SDRL_1$) values when a shift in the process parameter occurs. The large value of ARL_0 is required to guide against unnecessary false alarms, while a chart with smaller ARL_1 and $SDRL_1$ at more shift values is considered more efficient. In addition to the computations of ARLs and SDRLs, which measure performance of a control chart at only some specific points, we have calculated the Extra Quadratic Loss (EQL) to measure the overall effectiveness of the CUSUM charts (cf. Wu et al. [36] and Wu et al. [37]). The EQL measures the performance of a control chart

over some shift intervals rather than at some specific shift values. If φ denotes either δ mean shift or τ shift in standard deviation, then

$$EQL = \frac{1}{\varphi_{\max} - \varphi_{\min}} \int_{\varphi_{\min}}^{\varphi_{\max}} \varphi^2 ARL(\varphi)\ d\varphi \tag{18}$$

where $\varphi_{\max}$ and $\varphi_{\min}$ are respectively, the upper and lower bounds of the desired shift, while $ARL(\varphi)$ is the ARL value of a particular chart at φ. The EQL of a CUSUM chart can be calculated using numerical integration or approximated by the summation $EQL = (1/r)\sum_{i=1}^{r} \varphi^2 ARL(\varphi)$, where r is the number of equally divided shifts from $\varphi_{\min}$ to $\varphi_{\max}$. A control chart with the smallest EQL value is considered the most efficient chart.

3.2 Evaluating CUSUM Location Charts

The evaluate the CUSUM location chart, we assume that for a fixed standard deviation σ_0, both the in-control and out-of-control processes are normally distributed with mean μ_0 and $\mu_1 = \mu_0 + \delta\left(\sigma_0/\sqrt{n}\right)$, respectively. Thus, the mean shift δ is equal to zero when the process is in-control and a value of $\delta > 0$ is an indication that the process is out-of-control. Without loss of generality, we assume $\mu_0 = 0$ and $\sigma_0 = 1$. Using simulation of 10^5 iterations for each mean shift δ level in a process, we compute the run length properties of the RSS CUSUM charts and its median variation. Using a set size of $n = 5$, $k = 0.25$, 0.5, 0.75 and 1.0, the control limit h is adjusted to obtain the desired in-control ARL_0 values of 100, 300 and 500.

For the other choices of ARL_0 values, the pair (h, k) can easily be worked out. Tables 2, 3, and 4 presents the ARL (cf. Al-Sabah [8], Abujiya et al. [16]) and EQL values for the classical, RSS and MRSS based CUSUM location control charts. We have presented only the SDRL values for $ARL_0 = 100$ in Table 5, since others have similar conclusion [16].

3.3 Evaluating CUSUM Dispersion Charts

To measure the performance of CUSUM dispersion chart, we assume that both the in-control and the out-of-control processes are both normally distributed with mean μ_0 but with different standard deviation σ_0 and $\sigma_1 = \tau\sigma_0$, respectively. Here, the process is said to be in-control if the shift in standard deviation $\tau = 1$, while an out-of-control condition is indicated by $\tau \neq 1$. Unlike the value of δ, here τ could either be greater or less than one, i.e., $\tau > 1$ or $\tau < 1$, which represents an increase and decease in process standard deviation, respectively. Like the location chart, we assume without loss of generality that the stable process has $\mu_0 = 0$ and $\sigma_0 = 1$.

Table 2 ARL and EQL values for the two-sided CUSUM location chart ($n = 5$, $ARL_0 = 100$)

	k	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00
δ	h	5.595	3.500	2.482	1.876	5.605	3.508	2.491	1.885	5.594	3.500	2.482	1.878
		Classical				RSS				MRSS			
0.00		100.09	100.04	100.08	100.07	100.01	100.07	100.12	100.12	100.02	100.09	100.00	100.13
0.25		44.41	52.27	60.74	67.97	24.51	28.56	35.17	42.33	21.06	24.09	29.64	36.20
0.50		19.27	21.73	26.53	32.67	10.22	9.81	11.10	13.42	8.89	8.19	8.92	10.63
0.75		11.59	11.43	13.32	16.24	6.38	5.43	5.42	5.98	5.59	4.64	4.50	4.78
1.00		8.20	7.39	7.84	9.17	4.65	3.74	3.48	3.53	4.12	3.26	2.96	2.92
1.25		6.34	5.39	5.38	5.95	3.70	2.89	2.57	2.47	3.29	2.55	2.23	2.09
1.50		5.19	4.24	4.04	4.20	3.09	2.38	2.06	1.92	2.76	2.12	1.80	1.65
1.75		4.40	3.51	3.23	3.22	2.67	2.06	1.73	1.58	2.39	1.85	1.52	1.38
2.00		3.83	3.00	2.70	2.61	2.36	1.82	1.49	1.35	2.15	1.62	1.31	1.20
2.25		3.41	2.65	2.32	2.20	2.15	1.62	1.32	1.20	2.00	1.43	1.17	1.09
2.50		3.08	2.38	2.06	1.90	2.01	1.45	1.18	1.10	1.89	1.25	1.08	1.04
2.75		2.81	2.16	1.84	1.69	1.92	1.29	1.09	1.04	1.76	1.13	1.03	1.01
3.00		2.59	1.99	1.68	1.52	1.81	1.16	1.04	1.02	1.60	1.05	1.01	1.00
3.25		2.41	1.85	1.53	1.38	1.67	1.08	1.02	1.01	1.41	1.02	1.00	1.00
3.50		2.26	1.73	1.41	1.28	1.51	1.04	1.01	1.00	1.24	1.01	1.00	1.00
3.75		2.14	1.62	1.30	1.20	1.35	1.01	1.00	1.00	1.12	1.00	1.00	1.00
4.00		2.06	1.51	1.22	1.13	1.21	1.00	1.00	1.00	1.05	1.00	1.00	1.00
EQL		15.842	12.585	11.192	10.963	10.108	7.394	6.768	6.643	8.872	6.786	6.339	6.254

Table 3 ARL and EQLL values for the two-sided CUSUM location chart ($n = 5$, $ARL_0 = 300$)

	k	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00
δ	h	7.621	4.577	3.206	2.417	7.623	4.581	3.215	2.432	7.612	4.575	3.204	2.419
		Classical				RSS				MRSS			
0.00		300.16	300.35	300.26	300.66	300.01	300.14	300.22	300.50	300.20	300.10	300.09	300.25
0.25		76.73	107.54	139.27	166.78	35.81	46.69	65.40	87.32	30.07	37.47	52.02	70.45
0.50		27.23	32.86	45.41	61.23	13.66	12.90	15.24	20.15	11.82	10.64	11.96	15.18
0.75		15.62	15.47	19.05	25.59	8.38	6.84	6.85	7.77	7.34	5.83	5.59	6.04
1.00		10.88	9.51	10.36	12.84	6.07	4.66	4.26	4.35	5.36	4.04	3.61	3.54
1.25		8.36	6.82	6.78	7.68	4.80	3.57	3.12	2.98	4.26	3.12	2.69	2.51
1.50		6.81	5.32	5.00	5.26	3.99	2.90	2.49	2.30	3.56	2.58	2.18	1.97
1.75		5.77	4.36	3.95	3.95	3.43	2.48	2.10	1.88	3.08	2.23	1.85	1.63
2.00		4.99	3.73	3.29	3.17	3.03	2.20	1.83	1.61	2.72	2.00	1.61	1.39
2.25		4.42	3.26	2.81	2.64	2.72	2.00	1.61	1.39	2.43	1.83	1.40	1.22
2.50		3.98	2.90	2.48	2.28	2.45	1.85	1.43	1.24	2.20	1.66	1.24	1.10
2.75		3.63	2.63	2.23	2.01	2.23	1.70	1.27	1.13	2.06	1.47	1.12	1.04
3.00		3.34	2.41	2.03	1.81	2.09	1.53	1.15	1.06	2.00	1.30	1.05	1.02
3.25		3.10	2.24	1.87	1.64	2.02	1.37	1.07	1.02	1.97	1.16	1.02	1.00
3.50		2.90	2.10	1.73	1.50	1.99	1.23	1.03	1.01	1.91	1.07	1.01	1.00
3.75		2.71	2.00	1.61	1.38	1.95	1.12	1.01	1.00	1.80	1.03	1.00	1.00
4.00		2.55	1.90	1.49	1.29	1.89	1.06	1.00	1.00	1.65	1.01	1.00	1.00
EQL		20.552	15.810	14.287	14.132	13.034	9.077	7.713	7.458	11.978	8.013	7.011	6.813

Table 4 ARL and EQL values for the two-sided CUSUM location chart ($n = 5$, $ARL_0 = 500$)

	k	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00
δ	h	8.591	5.077	3.544	2.667	8.593	5.082	3.554	2.686	8.585	5.075	3.542	2.668
		Classical				RSS				MRSS			
0.00		500.13	500.43	500.16	500.05	500.22	500.39	500.20	500.07	500.90	500.80	500.64	500.44
0.25		94.73	146.00	202.55	250.93	41.33	57.23	85.68	120.37	34.39	44.66	66.22	93.89
0.50		31.16	38.93	57.23	81.46	15.33	14.40	17.55	23.94	13.23	11.79	13.49	17.54
0.75		17.50	17.34	22.17	30.90	9.35	7.53	7.52	8.62	8.20	6.40	6.12	6.65
1.00		12.17	10.54	11.60	14.68	6.77	5.10	4.63	4.72	5.97	4.41	3.91	3.83
1.25		9.34	7.50	7.49	8.53	5.33	3.88	3.37	3.23	4.73	3.40	2.90	2.70
1.50		7.58	5.82	5.45	5.72	4.42	3.16	2.69	2.46	3.94	2.79	2.35	2.11
1.75		6.41	4.77	4.29	4.29	3.80	2.69	2.25	2.03	3.40	2.39	2.00	1.75
2.00		5.55	4.05	3.55	3.42	3.35	2.35	1.97	1.72	3.02	2.13	1.76	1.50
2.25		4.92	3.54	3.04	2.86	3.01	2.13	1.76	1.50	2.72	1.97	1.55	1.30
2.50		4.41	3.15	2.67	2.45	2.75	1.98	1.57	1.32	2.43	1.83	1.36	1.16
2.75		4.01	2.85	2.40	2.17	2.49	1.86	1.39	1.19	2.20	1.67	1.20	1.07
3.00		3.69	2.60	2.18	1.94	2.26	1.72	1.24	1.10	2.07	1.49	1.10	1.03
3.25		3.42	2.41	2.01	1.76	2.11	1.57	1.13	1.04	2.01	1.31	1.04	1.01
3.50		3.20	2.25	1.87	1.62	2.03	1.40	1.06	1.02	1.99	1.17	1.01	1.00
3.75		3.01	2.13	1.75	1.49	2.00	1.26	1.03	1.01	1.97	1.08	1.00	1.00
4.00		2.84	2.03	1.63	1.38	1.99	1.14	1.01	1.00	1.91	1.03	1.00	1.00
EQL		22.849	17.284	15.874	15.951	14.034	10.037	8.286	7.949	12.975	8.764	7.412	7.145

Table 5 SDRL values for the two-sided CUSUM location chart ($n = 5$, $ARL_0 = 100$)

	k	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00
δ	h	5.595	3.500	2.482	1.876	5.605	3.508	2.491	1.885	5.594	3.500	2.482	1.878
		Classical				RSS				MRSS			
0.00		91.49	95.90	97.84	98.76	90.92	96.27	98.17	98.79	91.30	96.26	98.21	98.84
0.25		36.11	48.35	58.28	66.41	16.97	24.48	32.68	40.82	13.87	20.13	27.17	34.52
0.50		12.31	17.81	24.07	30.88	4.88	6.40	8.76	11.84	3.95	4.96	6.67	9.00
0.75		5.86	7.87	10.90	14.57	2.38	2.71	3.37	4.41	1.96	2.15	2.61	3.30
1.00		3.49	4.27	5.63	7.49	1.47	1.53	1.77	2.15	1.23	1.25	1.39	1.63
1.25		2.37	2.69	3.36	4.39	1.03	1.02	1.12	1.27	0.87	0.84	0.91	1.01
1.50		1.73	1.87	2.22	2.75	0.79	0.75	0.81	0.88	0.68	0.63	0.68	0.71
1.75		1.36	1.40	1.59	1.90	0.65	0.60	0.65	0.67	0.55	0.54	0.57	0.54
2.00		1.10	1.09	1.20	1.39	0.53	0.53	0.56	0.53	0.41	0.52	0.48	0.41
2.25		0.92	0.89	0.96	1.08	0.41	0.52	0.48	0.41	0.32	0.50	0.38	0.29
2.50		0.79	0.75	0.81	0.88	0.32	0.50	0.39	0.30	0.34	0.43	0.26	0.19
2.75		0.70	0.64	0.70	0.74	0.33	0.45	0.29	0.21	0.43	0.33	0.17	0.11
3.00		0.62	0.58	0.63	0.63	0.40	0.37	0.20	0.13	0.49	0.23	0.09	0.05
3.25		0.55	0.54	0.57	0.55	0.47	0.27	0.12	0.08	0.49	0.14	0.05	0.03
3.50		0.48	0.53	0.52	0.48	0.50	0.19	0.07	0.04	0.43	0.07	0.02	0.01
3.75		0.41	0.52	0.47	0.41	0.48	0.11	0.04	0.02	0.33	0.03	0.01	0.00
4.00		0.35	0.51	0.42	0.34	0.41	0.07	0.02	0.01	0.22	0.01	0.00	0.00

Considering only the case when the in-control $ARL_0 = 200$, we have evaluated the run length properties of the CUSUM dispersion charts using RSS and its extreme variation via simulation study of 10^5 iterations for each shift τ in the process variability.

Using a set size of $n = 5$ and setting $k = d_n(1 + \tau)/2$ (CUSUM R chart) or $k = c_n(1 + \tau)/2$ (CUSUM S chart) to the required percentage of detection, the control limit h is adjusted to obtain the target ARL_0 value of 200. All the dispersion charts were designed to detect upward shifts ($\tau > 1$) and downward shifts ($\tau < 1$) of 20%, 30%, 40% and 50% in the process standard deviation. Tables 6, 7, 8, 9, 10, and 11 presents the ARL and SDRL values for the classical, RSS and ERSS based CUSUM R and S charts for different shift value τ. The approximated EQL values are also displayed in Tables 6, 8, 9 and 10.

4 Results and Discussion

4.1 Performance of CUSUM Location Charts

Based on the ARL, SDRL and EQL values for the classical, RSS and MRSS CUSUM control charts presented in Tables 2, 3, 4, and 5, we summarize our findings below. We have reported only the SDRL values for $ARL_0 = 100$ (cf. Table 5) since the resulting pattern is the same for other ARL_0 values.

1. The low values of ARL_1 shows that the use of RSS and its median variant has greatly enhanced the sensitivity of CUSUM chart in the quick detection of small, moderate and large shifts in the process mean (cf. Tables 2, 3, and 4).
2. With in-control ARL_0 fixed, and given reference value k, there is no significant difference between the control limits h of the RSS scheme and those of the classical CUSUM control charts (cf. Tables 2, 3, and 4). Observe that h is indirectly proportional to k.
3. In the out-of-control cases when $\delta > 0$, the values of ARL_1 and $SDRL_1$ for the RSS and MRSS based CUSUM charts decreases more rapidly than the classical charts, as the shift in mean value δ increases (cf. Tables 2, 3, 4, and 5).
4. The in-control $SDRL_0$ values of the three CUSUM control charts appeared to be in agreement, and relatively very close to the corresponding ARL_0 values except when $k < 0.5$. The difference is very minimal as k gets larger (cf. Table 5).
5. Similar to the classical chart, the ARL_1 and $SDRL_1$ values reveal that the RSS and MRSS CUSUM charts are very effective in detecting small changes $\delta \leq 0.5$, in the process mean when k is small. However, the charts are more sensitive to larger shifts when the value of k is large.
6. The RSS and MRSS have both maintained all the properties of the classical chart, including its advantage for early detection of small changes in the process. Moreover, the RSS schemes have improved the monitoring of large shifts in the process mean (cf. Tables 2, 3, 4, and 5).

Table 6 ARL and EQL values for the two-sided CUSUM R control chart ($\tau > 1$, $n = 5$, $\text{ARL}_0 = 200$)

	k	2.559	2.675	2.791	2.908	2.831	2.96	3.089	3.218	3.399	3.554	3.708	3.863
	h	4.910	3.884	3.220	2.746	4.188	3.267	2.686	2.271	3.749	2.853	2.300	1.916
τ	σ	1.20	1.30	1.40	1.50	1.20	1.30	1.40	1.50	1.20	1.30	1.40	1.50
		Classical				RSS				ERSS			
1.00		200.22	200.22	200.18	200.02	200.16	200.04	200.09	200.25	200.16	200.08	200.02	200.35
1.10		40.25	42.75	45.58	48.43	34.48	37.86	41.63	44.77	28.73	32.02	35.71	39.58
1.20		17.51	17.60	18.26	19.26	14.25	14.48	15.19	16.41	11.30	11.57	12.43	13.64
1.30		10.77	10.31	10.30	10.53	8.60	8.33	8.46	8.82	6.76	6.48	6.63	6.95
1.40		7.77	7.21	7.00	7.01	6.26	5.82	5.65	5.74	4.87	4.49	4.38	4.43
1.50		6.11	5.56	5.31	5.22	4.89	4.43	4.24	4.22	3.84	3.46	3.30	3.26
1.60		5.07	4.57	4.30	4.15	4.06	3.61	3.47	3.37	3.19	2.84	2.65	2.59
1.70		4.35	3.88	3.62	3.47	3.49	3.13	2.92	2.79	2.75	2.42	2.25	2.18
1.80		3.83	3.40	3.14	3.00	3.08	2.73	2.53	2.43	2.43	2.13	1.97	1.90
1.90		3.43	3.04	2.80	2.65	2.76	2.45	2.26	2.16	2.19	1.91	1.76	1.70
2.00		3.12	2.75	2.53	2.40	2.52	2.21	2.05	1.94	2.00	1.75	1.62	1.55
EQL		1.559	1.416	1.342	1.309	1.256	1.141	1.090	1.072	0.991	0.894	0.854	0.846

Table 7 SDRL values for the two-sided CUSUM R control chart ($\tau > 1$, $n = 5$, $\text{ARL}_0 = 200$)

	k	2.559	2.675	2.791	2.908	2.831	2.96	3.089	3.218	3.399	3.554	3.708	3.863
	h	4.910	3.884	3.220	2.746	4.188	3.267	2.686	2.271	3.749	2.853	2.300	1.916
τ	σ	1.20	1.30	1.40	1.50	1.20	1.30	1.40	1.50	1.20	1.30	1.40	1.50
		Classical				RSS				ERSS			
1.00		194.12	195.69	197.30	197.96	193.73	193.84	196.12	197.65	194.97	197.08	198.02	199.18
1.10		33.74	38.40	42.50	46.10	28.45	34.13	39.14	43.32	23.89	28.97	33.64	38.17
1.20		12.11	13.71	15.33	16.94	9.59	11.15	12.78	14.53	7.50	8.86	10.42	12.10
1.30		6.44	7.04	7.76	8.45	5.04	5.62	6.22	7.02	3.77	4.23	4.85	5.52
1.40		4.19	4.44	4.75	5.14	3.22	3.49	3.76	4.14	2.43	2.60	2.86	3.19
1.50		3.08	3.17	3.35	3.58	2.34	2.44	2.59	2.79	1.78	1.85	1.96	2.11
1.60		2.44	2.46	2.57	2.67	1.85	1.87	2.00	2.11	1.39	1.41	1.47	1.56
1.70		2.01	2.01	2.06	2.13	1.55	1.55	1.57	1.65	1.16	1.16	1.18	1.23
1.80		1.73	1.71	1.73	1.76	1.33	1.30	1.33	1.36	1.00	0.99	0.99	1.02
1.90		1.52	1.49	1.49	1.52	1.16	1.15	1.16	1.16	0.89	0.86	0.86	0.86
2.00		1.36	1.32	1.32	1.32	1.04	1.01	1.02	1.02	0.81	0.77	0.76	0.74

Table 8 ARL and EQL values for the two-sided CUSUM R control chart ($\tau < 1$, $n = 5$, $\text{ARL}_0 = 200$)

	k	2.093	1.977	1.861	1.745	2.317	2.188	2.059	1.931	2.781	2.627	2.472	2.318
	h	4.335	3.182	2.434	1.912	3.667	2.633	1.985	1.538	3.313	2.345	1.745	1.350
τ	σ	0.80	0.70	0.60	0.50	0.80	0.70	0.60	0.50	0.80	0.70	0.60	0.50
		Classical				RSS				ERSS			
1.00		200.08	200.12	200.02	200.28	200.26	200.21	200.09	200.57	200.04	200.35	200.29	200.08
0.95		89.86	100.94	111.39	120.77	81.24	92.85	104.55	116.05	73.44	87.52	101.07	114.87
0.90		45.44	53.34	62.59	72.12	38.09	46.12	55.75	66.32	31.88	40.37	51.01	63.97
0.85		26.32	30.02	35.76	43.17	21.10	24.67	30.54	39.02	16.94	20.35	26.36	34.81
0.80		17.42	18.71	21.77	26.28	13.75	14.75	17.99	22.22	10.80	11.83	14.52	19.22
0.75		12.70	12.80	14.14	16.67	9.99	9.98	11.17	13.22	7.79	7.83	8.89	11.21
0.70		9.95	9.54	9.94	11.17	7.79	7.46	7.79	8.77	6.06	5.72	6.03	7.04
0.65		8.14	7.49	7.45	7.95	6.35	5.79	5.74	6.19	4.95	4.48	4.45	4.85
0.60		6.89	6.16	5.89	6.04	5.38	4.78	4.55	4.60	4.19	3.68	3.50	3.61
0.55		5.96	5.21	4.86	4.78	4.66	4.04	3.75	3.68	3.65	3.13	2.90	2.86
0.50		5.26	4.52	4.12	3.95	4.14	3.51	3.18	3.05	3.25	2.72	2.48	2.38
EQL		0.762	0.739	0.771	0.850	0.604	0.587	0.621	0.696	0.477	0.467	0.504	0.586

Table 9 ARL and EQL values for the two-sided CUSUM S control chart ($\tau > 1$, $n = 5$, $\text{ARL}_0 = 200$)

	k	1.034	1.081	1.128	1.175	1.054	1.102	1.150	1.198	1.066	1.114	1.163	1.211
	h	1.901	1.494	1.230	1.044	1.392	1.071	0.867	0.723	1.085	0.816	0.646	0.532
τ	σ	1.20	1.30	1.40	1.50	1.20	1.30	1.40	1.50	1.20	1.30	1.40	1.50
		Classical				RSS				ERSS			
1.00		200.24	200.13	200.05	200.54	200.37	200.61	200.70	200.19	200.42	200.42	200.71	200.95
1.10		38.90	41.09	43.79	46.34	31.63	34.53	37.64	40.55	26.71	29.42	32.76	36.09
1.20		16.83	16.83	17.36	18.19	12.90	12.99	13.65	14.49	10.47	10.63	11.25	12.19
1.30		10.33	9.86	9.79	9.96	7.72	7.32	7.45	7.69	6.30	5.96	5.98	6.23
1.40		7.46	6.89	6.66	6.61	5.62	5.17	5.00	4.96	4.52	4.13	3.99	4.00
1.50		5.87	5.31	5.04	4.94	4.44	4.00	3.78	3.67	3.58	3.20	3.02	2.96
1.60		4.87	4.36	4.09	3.95	3.69	3.28	3.04	2.93	2.98	2.63	2.44	2.36
1.70		4.20	3.73	3.46	3.29	3.14	2.78	2.58	2.50	2.58	2.25	2.08	1.99
1.80		3.69	3.26	2.99	2.86	2.79	2.45	2.27	2.16	2.28	1.99	1.82	1.75
1.90		3.30	2.91	2.68	2.53	2.50	2.20	2.04	1.91	2.06	1.79	1.64	1.57
2.00		3.01	2.64	2.42	2.29	2.32	2.03	1.84	1.76	1.87	1.63	1.51	1.44
EQL		1.500	1.356	1.280	1.244	1.137	1.026	0.972	0.948	0.927	0.830	0.786	0.773

Table 10 ARL and EQL values for the two-sided CUSUM S control chart ($\tau < 1$, $n = 5$, $\text{ARL}_0 = 200$)

	k	0.846	0.799	0.752	0.705	0.862	0.814	0.766	0.719	0.872	0.824	0.775	0.727
	h	1.704	1.25	0.958	0.754	1.260	0.912	0.691	0.541	0.995	0.711	0.532	0.416
τ	σ	0.80	0.70	0.60	0.50	0.80	0.70	0.60	0.50	0.80	0.70	0.60	0.50
		Classical				RSS				ERSS			
1.00		200.32	200.06	200.50	200.91	200.27	200.68	200.77	200.40	200.07	200.81	201.08	201.17
0.95		89.61	101.04	111.95	122.16	80.92	94.86	107.13	118.79	73.76	88.96	103.61	117.84
0.90		44.97	53.05	63.00	72.93	37.25	46.44	57.00	67.98	31.56	40.83	52.62	66.85
0.85		25.84	29.75	35.75	43.92	20.08	24.41	30.95	39.97	16.55	20.34	27.13	36.30
0.80		17.06	18.40	21.62	26.26	13.07	14.19	17.68	22.48	10.48	11.70	14.74	20.03
0.75		12.41	12.55	13.97	16.62	9.33	9.57	10.98	13.30	7.53	7.68	8.89	11.45
0.70		9.71	9.32	9.78	11.05	7.37	7.04	7.46	8.57	5.84	5.58	5.96	7.11
0.65		7.94	7.30	7.31	7.90	5.90	5.47	5.49	5.97	4.77	4.35	4.38	4.84
0.60		6.71	6.00	5.77	5.92	5.01	4.51	4.34	4.47	4.04	3.57	3.44	3.57
0.55		5.82	5.09	4.76	4.71	4.36	3.81	3.55	3.51	3.52	3.04	2.83	2.82
0.50		5.12	4.41	4.03	3.88	3.86	3.30	3.02	2.89	3.13	2.65	2.43	2.35
EQL		0.745	0.725	0.762	0.846	0.569	0.563	0.606	0.690	0.462	0.459	0.504	0.597

7. On the overall, the smaller EQL values of RSS and MRSS show the immediate benefit of using the schemes in constructing CUSUM charts. As expected, the outlier-resistant MRSS is the most efficient with the ability to detect a wide range of shifts in the process mean than the RSS and classical charts (cf. Tables 2, 3, and 4).

4.2 Performance of CUSUM Dispersion Charts

To access the performance of the CUSUM R chart and CUSUM S chart using RSS schemes, we refer to the ARL, SDRL and EQL values presented in Tables 6, 7, 8, 9, 10, and 11. We have reported only two tables for SDRL values (cf. Tables 7 and 11) to save space. In addition to Tables 6, 7, 8, 9, 10, and 11, we have also presented the ARL and EQL values for the classical, RSS and ERSS CUSUM R and S charts when $\text{ARL}_0 = 500$ with 30% σ increase and decrease in Table 12, for a quick comparison of the dispersion charts. Hence, we summarize our findings as follows:

1. The in-control SDRL_0 values of the classical, RSS and the ERSS are very similar with no significant differences from the corresponding ARL_0 values. Interestingly, SDRL_0 values of the ERSS CUSUM charts are closer to the target value of 200 (cf. Tables 7 and 11).
2. The ARL_1 and SDRL_1 values of RSS and ERSS decreases more rapidly than the classical case when the shift in standard deviation increases ($\tau > 1$), (cf. Tables 6, 7, and 9), or decreases ($\tau < 1$), (cf. Tables 8, 10, and 11). If $\tau < 1$, the h value is directly proportional to k and indirectly proportional when $\tau > 1$.
3. There appear to be no much difference between the CUSUM R chart and CUSUM S chart when changes in the downward shifts if of interest. However, the ERSS CUSUM S chart is more effective in detecting decreases in the process standard deviation than others (cf. Tables 6, 7, 8, 9, 10, and 11).
4. Small values of EQL for the RSS and ERSS CUSUM R and CUSUM S charts over the classical chart, shows the significant improvements on the overall effectiveness of the scheme in monitoring the process variability (cf. Tables 6, 7, 8, 9, 10, and 11).
5. The ERSS CUSUM dispersion charts are more efficient, regarding ARL, SDRL and EQL values than its RSS counterpart. Furthermore, the ERSS CUSUM S chart dominates all the other scale charts in detecting increases and decreases in the process standard deviation (cf. Table 12). The charts are arranged from left to right based on their overall effectiveness.

Table 11 SDRL values for the two-sided CUSUM S control chart ($\tau < 1$, $n = 5$, $\text{ARL}_0 = 200$)

	k	0.846	0.799	0.752	0.705	0.862	0.814	0.766	0.719	0.872	0.824	0.775	0.727
	h	1.704	1.25	0.958	0.754	1.260	0.912	0.691	0.541	0.995	0.711	0.532	0.416
τ	σ	0.80	0.70	0.60	0.50	0.80	0.70	0.60	0.50	0.80	0.70	0.60	0.50
		Classical				RSS				ERSS			
1.00		192.59	195.52	196.96	201.28	193.58	194.69	201.04	202.23	194.65	198.08	199.08	199.38
0.95		80.59	95.20	108.11	117.65	74.79	89.99	103.86	116.89	67.51	85.51	101.09	116.25
0.90		36.14	47.13	58.67	69.70	30.13	41.82	53.63	66.66	25.43	36.89	50.09	65.09
0.85		17.45	23.88	31.40	40.81	13.46	19.67	27.17	37.42	10.88	16.46	24.26	34.44
0.80		9.45	12.73	17.18	22.53	7.02	9.64	14.29	19.96	5.48	7.95	11.96	17.82
0.75		5.63	7.20	9.71	13.14	4.11	5.36	7.57	10.60	3.19	4.25	6.16	9.21
0.70		3.68	4.47	5.79	7.65	2.67	3.24	4.40	5.98	2.05	2.51	3.40	4.93
0.65		2.55	2.91	3.59	4.78	1.77	2.09	2.63	3.48	1.41	1.62	2.04	2.79
0.60		1.83	2.03	2.38	3.00	1.32	1.49	1.73	2.23	1.02	1.13	1.33	1.68
0.55		1.37	1.46	1.65	1.97	0.99	1.05	1.20	1.40	0.76	0.83	0.92	1.10
0.50		1.03	1.08	1.18	1.37	0.75	0.78	0.86	0.97	0.59	0.64	0.65	0.76

Table 12 ARL comparison among CUSUM R and S charts ($\tau > 1$, $\tau < 1$, $n = 5$, $\text{ARL}_0 = 500$)

	Classical CUSUM R chart	Classical CUSUM S chart	RSS CUSUM S chart	ERSS CUSUM R chart	ERSS CUSUM S chart		Classical CUSUM R chart	Classical CUSUM S chart	RSS CUSUM R chart	RSS CUSUM S chart	ERSS CUSUM R chart	ERSS CUSUM S chart
k	2.675	1.081	1.102	3.554	1.114	k	1.977	0.799	2.188	0.814	2.627	0.824
h	4.929	1.894	1.350	3.576	1.020	h	4.002	1.570	3.331	1.135	2.918	0.884
σ	1.30	1.30	1.30	1.30	1.30	σ	0.70	0.70	0.70	0.70	0.70	0.70
τ						τ						
1.00	500.36	500.26	500.24	500.43	501.04	1.00	500.21	499.90	500.08	500.48	500.71	500.82
1.10	69.46	66.43	53.91	49.85	45.36	0.95	205.95	205.82	186.91	189.10	175.60	178.61
1.20	24.13	22.94	17.33	15.18	13.76	0.90	88.59	88.17	75.80	75.89	65.51	66.67
1.30	13.25	12.65	9.37	8.03	7.33	0.85	43.45	43.00	35.00	34.31	28.43	28.41
1.40	9.03	8.59	6.37	5.43	4.97	0.80	24.81	24.38	19.29	18.51	15.17	14.93
1.50	6.87	6.56	4.83	4.13	3.80	0.75	16.39	16.00	12.56	11.91	9.67	9.45
1.60	5.56	5.33	3.94	3.36	3.11	0.70	11.83	11.56	9.11	8.57	6.96	6.78
1.70	4.70	4.49	3.34	2.86	2.64	0.65	9.22	8.98	7.06	6.63	5.40	5.25
1.80	4.09	3.92	2.91	2.50	2.31	0.60	7.57	7.37	5.79	5.43	4.42	4.29
1.90	3.63	3.48	2.60	2.24	2.07	0.55	6.39	6.22	4.89	4.58	3.75	3.63
2.00	3.29	3.15	2.35	2.03	1.89	0.50	5.53	5.38	4.25	3.98	3.27	3.17
2.10	3.00	2.88	2.16	1.86	1.73	0.45	4.88	4.75	3.75	3.52	2.92	2.83
2.20	2.77	2.66	2.00	1.73	1.61	0.40	4.38	4.27	3.37	3.18	2.62	2.53
2.30	2.58	2.48	1.87	1.62	1.52	0.35	3.99	3.88	3.10	2.95	2.32	2.24
2.40	2.41	2.32	1.76	1.53	1.43	0.30	3.65	3.54	2.93	2.75	2.09	2.05
2.50	2.28	2.19	1.67	1.45	1.36	0.25	3.30	3.20	2.75	2.43	2.01	2.00
EQL	2.562	2.457	1.844	1.593	1.479		1.195	1.167	0.940	0.890	0.728	0.713

4.3 Effect of Imperfect Ranking

The ranking of units based on the auxiliary variable discussed in Sect. 2.1.3, may affect the efficiency of RSS estimators and could have a negative impact on the performance of control charts based on RSS techniques. To access the performance of the imperfect RSS schemes, namely, imperfect RSS (IRSS), imperfect MRSS (IMRSS) and imperfect ERSS (IERSS), we follow the corresponding perfect ranking procedure where bivariate random samples are generated from a standard normal distribution. Using Monte Carlo simulation, we set $n = 5$, in-control $\text{ARL}_0 = 740$, correlation coefficients $\rho_{XY} = 0.25, 0.50, 0.75$ and 0.90 and compute the run length properties for the two-sided CUSUM control charts based on IRSS, IMRSS and IERSS when $k = 0.5$ for different shift values. In RSS technicality, both the traditional random sampling and perfect ranking are special cases of imperfect ranking with of $\rho_{XY} = 0$ and $\rho_{XY} = 1$, respectively. Tables 13 and 14 presents the ARL, SDRL and EQL values for the IRSS and IMRSS based location charts. Graphical displays of ARL curves for the IERSS CUSUM dispersion charts are also given in Fig. 2.

Observations from the obtained results shows that ranking errors does have negative impact on the performance RSS schemes based control charts but very negligible particularly as ρ_{XY} gets larger. Both the IRSS and IMRSS CUSUM location charts have smaller ARL_1 and SRDL_1 values than their corresponding classical chart irrespective of the errors associated with ranking of the variables. That is, regardless of the imperfectness in the ranking of units, the RSS schemes are more effective in detecting changes in the process mean level. This point is also supported by the EQL as its value decreases from left to right, with IMRSS dominating others. Interestingly, the ARL_0 and SRDL_0 values are also in agreement and approximately close to the target value.

The log ARL curves designed to detect 20% upward shift in standard deviation and presented in Fig. 2, shows that both the R and S CUSUM control charts based on IERSS procedure is doing great despite the presence of ranking errors. All the IERSS curves are lower in terms of the log ARL values than the classical charts.

4.4 An Application Example on the Filling of Bottles

This application example is based on an industrial data on filling up of the bottle problem, obtained from a production line of the Pepsi-Cola production company, Al-Khobar, Saudi Arabia (cf. Muttlak and Al-Sabah [5], Abujiya et al. [16], Abujiya et al. [15]). Following Takahasi and Wakimoto [22], we re-sampled the original data set to obtain thirty MRSS and ERSS samples each of set size $n = 5$, of which the last ten data points were contaminated. The statistics of CUSUM $\overline{X}$ and S charts were computed and using the in-control ARL_0 of 200. We used $k = 0.5$ for the location chart and the dispersion chart was designed to detect a 20% increase in the process

Table 13 ARL and EQL values for the two-sided CUSUM $\overline{X}$ chart under imperfect ranking ($k = 0.5$, $n = 5$, $\text{ARL}_0 = 740$)

	Classical	Imperfect Ranking								Perfect Ranking	
	$\rho_{XY} = 0$	$\rho_{XY} = 0.25$		$\rho_{XY} = 0.50$		$\rho_{XY} = 0.75$		$\rho_{XY} = 0.90$		$\rho_{XY} = 1.0$	
δ	$h = 5.465$	5.467	5.461	5.467	5.461	5.467	5.461	5.467	5.461	5.467	5.461
		IRSS	IMRSS	IRSS	IMRSS	IRSS	IMRSS	IRSS	IMRSS	RSS	MRSS
0.00	740.38	747.67	736.86	748.97	739.87	748.71	743.47	746.74	740.47	740.41	740.41
0.25	184.37	176.92	176.07	157.43	153.37	120.75	112.62	90.24	78.00	65.85	51.18
0.50	43.74	42.07	41.66	36.24	35.41	27.15	25.25	20.47	17.90	15.53	12.69
0.75	18.83	18.14	18.05	16.02	15.61	12.63	11.86	10.03	9.03	8.04	6.80
1.00	11.30	10.97	10.92	9.83	9.65	8.02	7.65	6.55	6.00	5.42	4.69
1.25	8.00	7.79	7.75	7.07	6.95	5.89	5.62	4.92	4.52	4.12	3.62
1.50	6.20	6.04	6.03	5.55	5.46	4.67	4.49	3.95	3.66	3.35	2.96
1.75	5.09	4.97	4.93	4.57	4.51	3.89	3.74	3.32	3.09	2.84	2.52
2.00	4.32	4.22	4.20	3.90	3.85	3.35	3.22	2.88	2.69	2.48	2.23
2.25	3.76	3.69	3.67	3.42	3.37	2.95	2.85	2.55	2.39	2.23	2.05
2.50	3.35	3.28	3.26	3.05	3.01	2.65	2.56	2.31	2.18	2.06	1.92
2.75	3.02	2.96	2.94	2.76	2.72	2.41	2.33	2.13	2.04	1.95	1.80
3.00	2.76	2.70	2.69	2.52	2.49	2.23	2.17	2.01	1.93	1.84	1.64
3.25	2.54	2.49	2.49	2.34	2.31	2.09	2.04	1.92	1.84	1.71	1.46
3.50	2.37	2.32	2.31	2.19	2.17	2.00	1.96	1.83	1.71	1.56	1.28
3.75	2.23	2.19	2.18	2.08	2.07	1.92	1.88	1.71	1.58	1.39	1.15
4.00	2.12	2.09	2.08	2.00	1.99	1.84	1.78	1.58	1.42	1.25	1.07
EQL	18.461	18.047	17.959	16.735	16.508	14.535	14.044	12.642	11.780	10.797	9.405

Table 14 SDRL values for the two-sided CUSUM $\overline{X}$chart under imperfect ranking ($k = 0.5$, $n = 5$, $\text{ARL}_0 = 740$)

	Classical	Imperfect Ranking								Perfect Ranking	
	$\rho_{XY} = 0$	$\rho_{XY} = 0.25$		$\rho_{XY} = 0.50$		$\rho_{XY} = 0.75$		$\rho_{XY} = 0.90$		$\rho_{XY} = 1.0$	
δ	$h = 5.465$	5.467	5.461	5.467	5.461	5.467	5.461	5.467	5.461	5.467	5.461
		IRSS	IMRSS	IRSS	IMRSS	IRSS	IMRSS	IRSS	IMRSS	RSS	MRSS
0.00	736.19	741.48	734.03	741.59	731.31	740.07	738.30	740.09	738.52	734.16	732.91
0.25	176.31	168.59	169.52	149.16	144.66	112.27	103.91	82.87	82.36	57.50	42.99
0.50	35.56	34.05	33.67	28.26	27.41	19.68	17.88	13.43	13.46	9.20	6.88
0.75	12.00	11.46	11.30	9.65	9.20	6.82	6.18	4.84	4.85	3.47	2.69
1.00	5.77	5.51	5.50	4.70	4.54	3.45	3.23	2.54	2.54	1.90	1.53
1.25	3.41	3.30	3.28	2.85	2.78	2.16	2.01	1.63	1.63	1.24	1.02
1.50	2.32	2.25	2.24	1.97	1.94	1.51	1.42	1.17	1.16	0.91	0.76
1.75	1.74	1.66	1.63	1.46	1.43	1.14	1.08	0.90	0.90	0.72	0.60
2.00	1.34	1.29	1.28	1.14	1.13	0.91	0.86	0.73	0.73	0.59	0.47
2.25	1.08	1.05	1.04	0.94	0.92	0.76	0.72	0.61	0.61	0.47	0.35
2.50	0.91	0.88	0.87	0.79	0.78	0.65	0.62	0.51	0.51	0.36	0.34
2.75	0.78	0.76	0.75	0.69	0.67	0.56	0.52	0.41	0.40	0.33	0.41
3.00	0.69	0.67	0.66	0.60	0.59	0.46	0.42	0.34	0.34	0.38	0.48
3.25	0.61	0.59	0.59	0.52	0.51	0.38	0.35	0.33	0.34	0.45	0.50
3.50	0.54	0.52	0.51	0.44	0.43	0.33	0.33	0.39	0.39	0.50	0.45
3.75	0.46	0.44	0.44	0.37	0.37	0.34	0.36	0.45	0.45	0.49	0.36
4.00	0.39	0.37	0.37	0.33	0.33	0.38	0.42	0.49	0.49	0.43	0.25

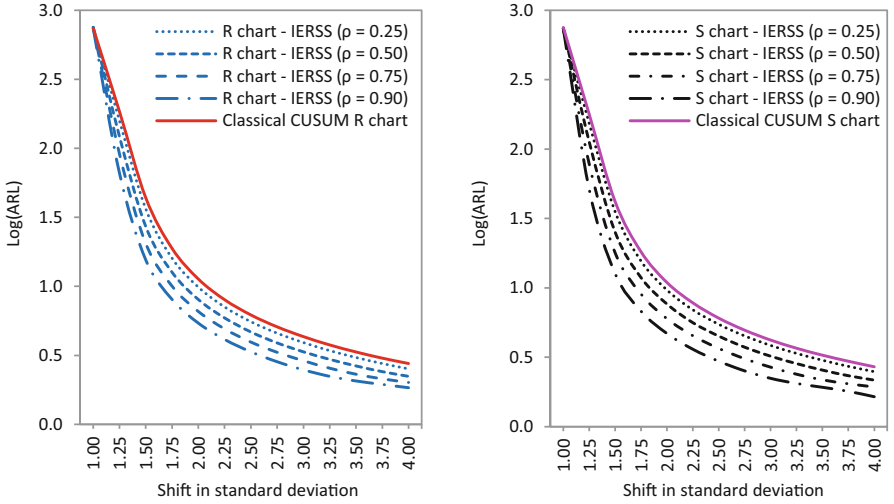


Fig. 2 ARL curves for IERSS vs. classical CUSUM R and CUSUM S charts for detecting 1.2σ ($n = 5$, $\text{ARL}_0 = 740$)

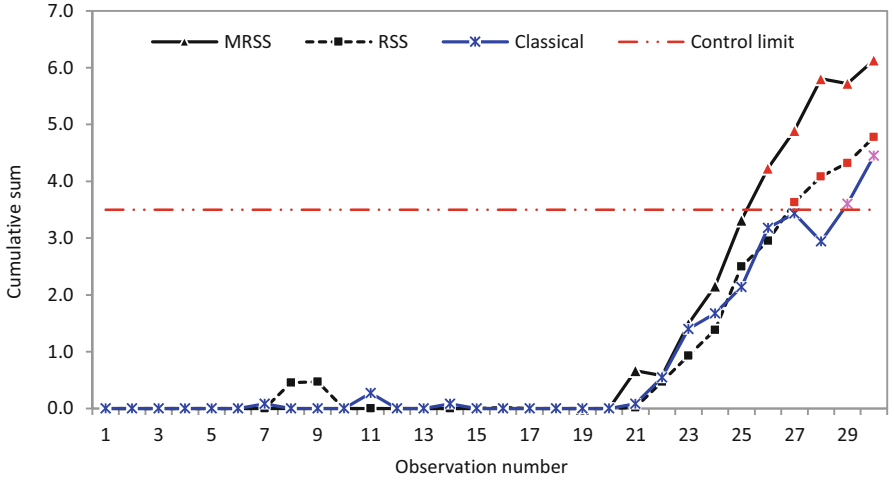


Fig. 3 Classical CUSUM $\overline{X}$ chart vs. RSS and MRSS CUSUM $\overline{X}$ chart for monitoring the process mean

variability following Hawkins and Zamba [38]. For more details, see Abujiya et al. [15] and Abujiya et al. [17]. Figures 3 and 4 present graphic control charts of the classical versus RSS based CUSUM chart.

From the location chart in Fig. 3, it is observed that the RSS and MRSS CUSUM $\overline{X}$ charts gives earlier signals than the classical chart based on random sampling. Observe that even though the contamination was introduced after the 20th data

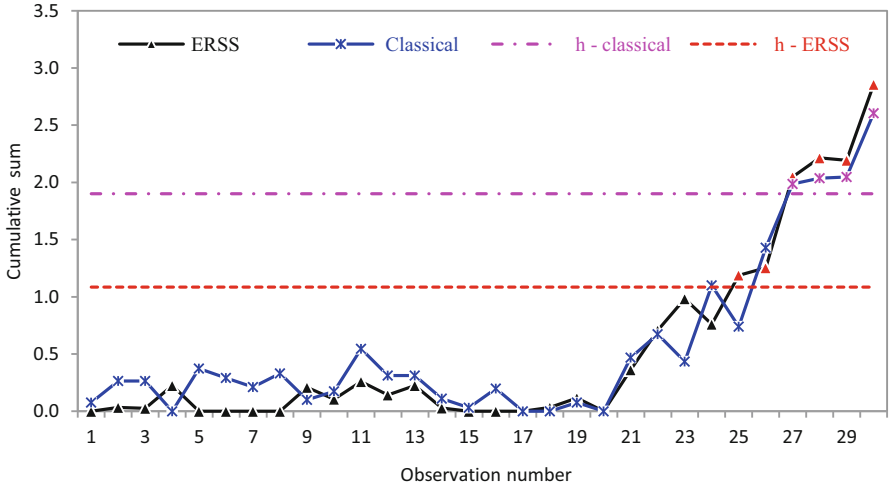


Fig. 4 Classical CUSUM *S* chart vs. ERSS CUSUM *S* chart for monitoring the process variability

point, the classical chart gives only two off-target point as compared to the four and five out-of-control signals of RSS and MRSS schemes, respectively. The use of RSS scheme in the construction of CUSUM *S* chart is also not doing badly as can be seen in Fig. 4. Based on the used data, the classical chart gives four out-of-control signals while the corresponding ERSS gave six off-target points. In both the figures, all control charts signify adrift in the process right from the 21st data point, but the classical chart appears to be dragging feet to signal in Fig. 3, while in Fig. 4, the tighter control limit of the ERSS scheme puts it at an advantage over the classical chart with a wider limit.

5 Conclusions

This study examines the performance of several CUSUM location and dispersion charts based on RSS and its variations for effective monitoring of changes in a normal process. From the run length properties obtained via simulation, we observed a great deal of improvements in the sensitivity of the CUSUM charts using RSS schemes over the standard CUSUM charts based on random sampling. For location chart, the outlier-resistant MRSS CUSUM $\overline{X}$ chart is more effective in monitoring the process mean than its corresponding balanced RSS and classical counterparts. Hence, the use of the median variation of RSS technique to enhance the sensitivity of basic location charts structure is recommended to quality control engineers.

The run length performance of the CUSUM *R* chart and CUSUM *S* chart for monitoring the process standard deviation were also evaluated. The structures were

based on the classical, RSS and ERSS data collection techniques. We found the ERSS CUSUM *S* chart more efficient than other dispersion control charts. This point is equally supported by the real data application example in the problem of filling the bottles. Thus, we recommend the use of the extreme variation of RSS to improve the performance of classical charts for quick detection of changes in the process variability. The scope of study be extended to the design strategies of multivariate control charts.

References

1. Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Milwaukee: ASQC.
2. Page, E. S. (1954). Continuous inspection schemes. *Biometrika, 41*(1-2), 100–114.
3. Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). New York: Wiley.
4. Hawkins, D. M., & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. New York: Springer.
5. Muttlak, H. A., & Al-Sabah, W. S. (2003). Statistical quality control based on ranked set sampling. *Journal of Applied Statistics, 30*(9), 1055–1078.
6. Abujiya, M. R., & Muttlak, H. A. (2004). Quality control chart for the mean using double ranked set sampling. *Journal of Applied Statistics, 31*(10), 1185–1201.
7. Al-Nasser, A. D., & Al-Rawwash, M. (2007). A control chart based on ranked data. *Journal of Applied Sciences, 7*(14), 1936–1941.
8. Al-Sabah, W. S. (2010). Cumulative sum statistical control charts using ranked set sampling data. *Pakistan Journal of Statistics, 26*(2), 365–378.
9. Al-Omari, A. I., & Al-Nasser, A. D. (2011). Statistical quality control limits for the sample mean chart using robust extreme ranked set sampling. *Economic Quality Control, 26*, 73–89.
10. Jafari, J. M., & Mirkamali, S. J. (2011). Control charts for attributes with maxima nominated samples. *Journal of Statistical Planning and Inference, 141*(7), 2386–2398.
11. Al-Omari, A. I., & Haq, A. (2012). Improved quality control charts for monitoring the process mean, using double-ranked set sampling methods. *Journal of Applied Statistics, 39*(4), 745–763.
12. Mehmood, R., Riaz, M., & Does, R. J. M. M. (2012). Control charts for location based on different sampling schemes. *Journal of Applied Statistics, 40*(3), 483–494.
13. Haq, A., Brown, J., & Moltchanova, E. (2014). A new cumulative sum quality control scheme for monitoring the process mean. *Quality and Reliability Engineering International, 30*(8), 1165–1177.
14. Mehmood, R., Riaz, M., & Does, R. J. M. M. (2014). Quality quandaries: On the application of different ranked set sampling schemes. *Quality Engineering, 26*(3), 370–378.
15. Abujiya, M. a. R., Riaz, M., & Lee, M. H. (2015). Enhanced cumulative sum charts for monitoring process dispersion. *PLoS One, 10*(4), e0124520.
16. Abujiya, M. R., Lee, M. H., & Riaz, M. (2015). Increasing the sensitivity of cumulative sum charts for location. *Quality and Reliability Engineering International., 31*(6), 1035–1051.
17. Abujiya, M. a. R., Riaz, M., & Lee, M. H. (2016). A new combined Shewhart–cumulative sum S chart for monitoring process standard deviation. *Quality and Reliability Engineering International, 32*(3), 1149–1165.
18. Abujiya, M. R., Lee, M. H., & Riaz, M. (2016). Combined application of Shewhart and cumulative sum R chart for monitoring process dispersion. *Quality and Reliability Engineering International, 32*(1), 51–67.

19. Abid, M., Nazir, H. Z., Riaz, M., & Lin, Z. (2017). Investigating the impact of ranked set sampling in nonparametric CUSUM control charts. *Quality and Reliability Engineering International, 33*(1), 203–214.
20. Abujiya, M. R., Riaz, M., & Lee, M. H. (2017). New EWMA S2 control charts for monitoring of process dispersion. *Scientia Iranica. Transaction E, Industrial Engineering, 24*(1), 378–389.
21. McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research, 3*(4), 385–390.
22. Takahasi, K., & Wakimoto, K. (1968). On unbiased estimates of population mean based on sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics, 20*(1), 1–31.
23. Chen, Z., Bai, Z., & K Sinha, B. (2004). *Ranked set sampling. Theory and applications*. New York: Springer.
24. Wolfe, D. A. (2012). Ranked set sampling: Its relevance and impact on statistical inference. *ISRN Probability and Statistics, 2012*, 32.
25. Al-Omari, A. I., & Bouza, C. N. (2014). Review of ranked set sampling: Modifications and applications. *Revista Investigación Operacional, 35*(3), 215–240.
26. Wolfe, D. A. (2004). Ranked set sampling: An approach to more efficient data collection. *Statistical Science, 19*(4), 636–643.
27. Sinha, B. K., & Purkayastha, S. (1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistics and Decisions, 14*(3), 223–240.
28. Yu, P. H., Lam, K., & Sinha, B. (1999). Estimation of normal variance based on balanced and unbalanced ranked set samples. *Environmental and Ecological Statistics, 6*(1), 23–46.
29. Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics Part a-Theory and Methods, 6*(12), 1207–1211.
30. Muttlak, H. A. (1998). Median ranked set sampling with concomitant variables and a comparison with ranked set sampling and regression estimators. *Environmetrics, 9*(3), 255–267.
31. Samawi, H. M., Ahmed, M. S., & Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal, 38*(5), 577–586.
32. Shaibu, A. B., & Muttlak, H. A. (2004). Estimating the parameters of the normal, exponential and gamma distributions using median and extreme ranked set samples. *STATISTICA, 64*(1), 75–98.
33. Hawkins, D. M. (1981). A CUSUM for a scale parameter. *Journal of Quality Technology, 13*(4), 228–231.
34. Chang, T. C., & Gan, F. F. (1995). A cumulative sum control chart for monitoring process variance. *Journal of Quality Technology, 27*(2), 109–119.
35. Haq, A. (2014). An improved mean deviation exponentially weighted moving average control chart to monitor process dispersion under ranked set sampling. *Journal of Statistical Computation and Simulation, 84*(9), 2011–2024.
36. Wu, Z., Yang, M., Jiang, W., & Khoo, M. B. C. (2008). Optimization designs of the combined Shewhart-CUSUM control charts. *Computational Statistics & Data Analysis, 53*(2), 496–506.
37. Wu, Z., Jiao, J., Yang, M., Liu, Y., & Wang, Z. (2009). An enhanced adaptive CUSUM control chart. *IIE Transactions, 41*(7), 642–653.
38. Hawkins, D. M., & Zamba, K. D. (2003). On small shifts in quality control. *Quality Engineering, 16*(1), 143–149.

A Survey of Control Charts for Simple Linear Profile Processes with Autocorrelation

Jyun-You Chiang, Hon Keung Tony Ng, Tzong-Ru Tsai, Yuhlong Lio, and Ding-Geng Chen

Abstract In quality control, the quality of process or product can be characterized by a profile that defines as a functional relationship between a quality response variable and one or more explanatory variables. Many research works have been accomplished on statistical process control for simple linear profile with independent or autocorrelated observations. This chapter will serve as a review of some recent works on statistical quality control on autocorrelated simple linear profiles.

1 Introduction

In many statistical process control (SPC) applications, the quality of a product can be characterized by a functional relationship between a response variable and one or more independent variables. This functional relationship is often called profile. The profile that links a response variable to the explanation variables can be linear and nonlinear in nature. In statistical quality control, Jensen et al. [6] considered a general linear profile model with m profiles. We assume that there are n observations

J.-Y. Chiang (✉)
School of Statistics, Southwestern University of Finance and Economics, Chengdu, China
e-mail: jiangjy@swufe.edu.cn

H. K. T. Ng
Department of Statistical Science, Southern Methodist University, Dallas, TX, USA
e-mail: ngh@mail.smu.edu

T.-R. Tsai
Department of Statistics, Tamkang University, New Taipei City, Taiwan
e-mail: tzongru@gms.tku.edu.tw

Y. Lio
Department of Mathematical Sciences, University of South Dakota, Vermillion, SD, USA
e-mail: Yuhlong.Lio@usd.edu

D.-G. Chen
Department of Statistics, University of Pretoria, Pretoria, South Africa
e-mail: chend05@gmail.com

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_5

in each of the j-th profile, $j = 1, 2, \ldots, m$. The general linear profile model is defined as

$$\mathbf{y}_j = \boldsymbol{\gamma}_{0,j} + \mathbf{X}_j \boldsymbol{\gamma}_j + \mathbf{Z}_j \mathbf{b}_j + \boldsymbol{\epsilon}_j, \tag{1}$$

where $\mathbf{y}_j$ is a $n \times 1$ vector of responses, $\mathbf{X}_j$ is a $n \times p$ matrix of the regressor variables associated with the fixed effects, $\boldsymbol{\gamma}_j$ is a $p \times 1$ parameter vector of coefficients for the fixed effects, $\mathbf{Z}_j$ is a $n \times q$ matrix of predictors associated with random effects, $\mathbf{b}_j \sim MN(\mathbf{0}, \mathbf{D})$ is a $q \times 1$ vector of random effect coefficients, $MN(\mathbf{0}, \mathbf{D})$ denotes a multivariate normal distribution with zero mean vector $\mathbf{0}$ and positive definite variance-covariance matrix $\mathbf{D}$, $\boldsymbol{\epsilon}_j \sim MN(\mathbf{0}, \mathbf{R}_j)$ is a $n \times 1$ vector of errors, and $MN(\mathbf{0}, \mathbf{R}_j)$ is a multivariate normal distribution with zero mean vector $\mathbf{0}$ and positive definite variance-covariance matrix $\mathbf{R}_j$ for $j = 1, 2, \cdots, m$. If the errors are assumed to be independent, then $\mathbf{R}_j = \sigma^2 \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. If the errors are correlated, $\mathbf{R}_j$ is often assumed as a simple form such as the autoregressive (AR) model in order to reduce the number of covariance parameters needed to be estimated. When Eq. (1) is reduced to the model that has only one fixed effect regressor variable and no random effect terms, the profile model is called a simple linear profile (SLP) model. The currently developed SPC on SLP models have been focused on the same levels of fixed effect regressor for all the profiles. Therefore, the SLP model can be defined as

$$y_{i,j} = \gamma_0 + x_i \gamma_1 + \epsilon_{i,j}, \tag{2}$$

where $y_{i,j}$ is the response in the j-th profile at the i-th level of predictor, x_i, γ_0 is the intercept, γ_1 is the model parameter for the predictor and $\epsilon_{i,j}$ are $N(0, \sigma^2)$ distributed for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$. SPC methods that involve monitoring a SLP process have drawn considerable attention over the past two decades because a SLP model is easy to handle and can be applied to many production processes. Process monitoring using control charts is a two-stage process, which has Phase I and Phase II. Phase I is to evaluate the stability of the process and estimate the in-control values of the process parameters, and Phase II is to monitor the future online data obtained after Phase I and detect shifts in the process parameters. In phase I, it is important to confirm the process stability under a given false alarm rate, i.e., a type-I error probability. In Phase II, the emphasis is on detecting process change as soon as possible. Both stages are usually measured by parameters, mean and standard deviation, of the run length distribution, where run length is the number of samples taken before an out-of-control signal is occurred. Therefore, the average run length (ARL_1) when the process out-of-control during the Phase II monitoring is often used to compare the performance of competing control chart methods under a given average run length ARL_0 for in-control Phase I. Numerous studies have been conducted on the use of profile monitoring methods, for example, Mestek et al. [11], Stover and Brill [17], Kang and Albin [7], Kim et al. [8], and Mahmoud et al. [10] studied some Phase I monitoring methods for SLP processes in order to set and evaluate the stability of a process and

to estimate process parameters. Kang and Albin [7], Kim et al. [8], Noorossana [13], Gupta et al. [5], Zou et al. [19], and Saghaei et al. [14] studied Phase II monitoring methods for SLP processes to detect shifts in the process parameters as soon as possible. Woodall et al. [18] reviewed the research papers related to SPC using profiles, provided examples of profile monitoring methods, identified some weaknesses in existing methods and proposed some new research directions. All the aforementioned studies have assumed that all the m profile models have the same n values of one fixed effect regressor variable, and the error terms in the models are independent and follow a normal distribution. When the error terms are not independent, Noorossana et al. [12] studied a SLP model with error terms that have a first-order autocorrelation structure between profiles and showed the impact on the ARL performance of the T^2 control chart proposed by Kang and Albin [7]. In Noorossana et al. [12], three methods based on exponentially weighted moving average/range (EWMA/R) and T^2 [7] and EWMA-3 [8] were provided to eliminate the effect of autocorrelation between profiles and the ARL. When the response variables are produced at set of points over time, the response of a profile are very often autocorrelated. In this case, Soleimani et al. [16] considered a SLP model with error terms that follow a first-order autoregressive model (AR(1)) within profile and suggested the use of Hotelling T^2 control charts and EWMA-type control charts, namely EWMA-3 charts, for monitoring SLP processes in the presence of within-profile autocorrelation. These charts are simple to detect process shifts in SLP processes. The simulation results in Soleimani et al. [16] showed that among many competitive methods, EWMA-type control charts are the most effective in detecting shifts in the intercept or slope of a SLP model. Three EWMA control charts are obtained for EWMA-3 charts by re-parameterizing the SLP model in terms of its intercept and slope. The intercept of the transformed SLP model being monitored is a linear combination of the intercept and slope in the original SLP model. The EWMA-3 charts of Soleimani et al. [16] are less sensitive if the original intercept and slope shift simultaneously in opposite directions. Chiang et al. [3] investigated an operational and effective Phase II monitoring method for a SLP model with error terms follow an AR(1) within-profile. In Chiang et al. [3], a new multivariate MEWMA control chart, namely MEWMA-SLP chart, was developed on the basis of the design of Lowry et al. [9] for quickly detecting process shifts associated with the original intercept or slope in the presence of within-profile autocorrelation.

In this chapter, we survey the work of SLP model with AR(1) autocorrelation for the error terms. The rest of this chapter is organized as follows. A SLP model with between-profile autocorrelation and EWMA-3 charts are reviewed in Sect. 2. In Sect. 3, a SLP model with within-profile autocorrelation is introduced along with the Hotelling T^2, EWMA-3 and MEWMA-SLP control charts for this SLP model. In Sect. 4, we present the construction of the two process capability indices studied by Chiang et al. [3]. In Sect. 5, Monte Carlo simulations are conducted to explore the performance of the MEWMA-SLP chart are discussed, and the applications of the proposed process capability indices are illustrated. Finally, conclusions are provided in Sect. 6.

2 SLP Model with Between-Profile Autocorrelated Error Terms

When the error terms in a SLP model satisfy the following autocorrelation structure

$$\epsilon_{i,j} = \rho\epsilon_{i,j-1} + a_{i,j},\ i = 1, 2, \cdots, n,\ \ j = 1, 2, \cdots, m, \tag{3}$$

where $a_{i,j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m$ are independent and identically distributed normal random variables with mean 0 and variance σ^2 and the random error terms $\epsilon_{i,j}$ for $i = 1, 2, \cdots, n$ within the j-th profile are independent, then the SLP model is a SLP model with between-profile autocorrelation [12]. Based on the autocorrelation structure between errors defined in Eq. (3), Noorossana et al. [12] provided the following autocorrelation structure between two different profiles at the same level of x_i:

$$y_{i,j} - (\gamma_0 + x_i\gamma_1) = \rho(y_{i,j-1} - (\gamma_0 + x_i\gamma_1)) + a_{i,j}, \tag{4}$$

and the prediction equation, $\hat{y}_{i,j} = \rho y_{i,j-1} + (1 - \rho)(\gamma_0 + x_i\gamma_1)$. Although there exists an AR(1) structure between errors corresponding to each level of predictor in different profiles in Eq. (4), Noorossana et al. [12] showed that the residual $e_{i,j} = y_{i,j} - \hat{y}_{i,j}$ equates to $a_{i,j}$, for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$ and the residuals are independent random variables with the expected value $E(e_{i,j}) = 0$ and variance $var(e_{i,j}) = \sigma^2$. Hence, $e_{i,j}$ can be monitored using traditional control charts. Noorossana et al. [12] proposed three methods to monitor the residuals $e_{i,j}$ for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$. Let $\bar{e}_j = \sum_{i=1}^{n} e_{i,j}/n$ for the j-th profile. The three methods are reviewed as follows. When the model parameters are unknown, the corresponding maximum likelihood estimates can be used to replace the values of the model parameters to establish the control charts.

Method 1: EWMA Chart Combining with R-Chart for Monitoring the Error Variance

The first method is EWMA in combining with R-chart, which is the same as the control charts used by Kang and Albin [7]. The EWMA chart procedure is constructed as follows:

Step 1: Define the EWMA sequence using residuals, $e_{i,j}$ for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$, as $z_j = \theta\bar{e}_j + (1 - \theta)z_{j-1}$ with $0 < \theta < 1$ as a smoothing constant and $z_0 = 0$

Step 2: Define the lower control limit (LCL) and the upper control limit (UCL) for EWMA chart as $LCL_{EWMA} = -L\sigma\sqrt{\frac{\theta}{(2-\theta)n}}$ and $UCL_{EWMA} = L\sigma\sqrt{\frac{\theta}{(2-\theta)n}}$, where L is a positive constant selected to give a specified in-control ARL, ARL_0.

The R-chart to detect shifts in the process variance is constructed as follows:

Step 1: Define the R sequence as $R_j = max_i(e_{i,j}) - min_i(e_{i,j})$ for the j-th profile.

Step 2: Define the lower control limit (LCL) and the upper control limit (UCL) for the R chart as $LCL_R = \sigma(d_2 - Ld_3)$ and $UCL_R = \sigma(d_2 + Ld_3)$, where L is a positive constant chosen to give a specified in-control ARL_0 and the values of d_2 and d_3 are constants depending on the sample size n.

Method 2: Modified Hotelling T^2 Control Chart

The second method proposed by Noorossana et al. [12] is a modified Hotelling T^2 control chart studied by Kang and Albin [7]. The modified Hotelling T^2 control chart is described as follows:

Step 1: Define $T_j^2 = (\mathbf{e}_j - \mathbf{0})\Sigma^{-1}(\mathbf{e}_j - \mathbf{0})^T$ where $\mathbf{e}_j = (e_{1,j}, e_{2,j}, \cdots, e_{n,j})$, $\mathbf{0}$ is a $n \times 1$ zero vector, $\Sigma = \sigma^2\mathbf{I}$ and $\mathbf{I}$ is a $n \times n$ identity matrix.

Step 2: Define the upper control limit as $UCL_T = \chi^2_{n,\alpha}$, which is the $100(1-\alpha)$ percentile of the chi-square distribution with n degree of freedom.

Method 3: Modified EWMA-3

The third method proposed by Noorossana et al. [12] is a modified EWMA-3 studied by Kim et al. [8] which was designed to deal with the autocorrelation between profiles. Noorossana et al. [12] proposed the scaling of the x-variable such that the average x-value is zero and the transformed SLP model of Eq. (2) with autocorrelation in Eq. (3) is

$$y_{i,j} = b_0 + x_i^* b_1 + \epsilon_{i,j}, \epsilon_{i,j} = \rho\epsilon_{i,j-1} + a_{i,j}, \ i = 1, 2, \cdots, n, \ j = 1, 2, \cdots, m, \tag{5}$$

where $x_i^* = (x_i - \bar{x})$, $b_0 = \gamma_0 + \gamma_1\bar{x}$ and $\bar{x} = \sum_{i=1}^n x_i/n$. Noorossana et al. [12] showed that the AR(1) structure between observations can be transformed into the intercept and the slope estimators in the successive profiles. Let $\hat{b}_{0,j}$ be the least squared estimate for b_0 and $\hat{b}_{1,j}$ be the least squared estimate for b_1 using the sample from the j-th profile. Noorossana et al. [12] calculated the residual for the intercept as $e_0(j) = \hat{b}_{0,j} - \rho\hat{b}_{0,j-1} - (1-\rho)b_0$, the residual for the slope as $e_1(j) = \hat{b}_{1,j} - \rho\hat{b}_{1,j-1} - (1-\rho)b_1$, and the residual as $e_{i,j} = y_{i,j} - \hat{y}_{i,j} = y_{i,j} - \rho y_{i,j-1} - (1-\rho)(b_0 + b_1 x_i^*)$, then the mean squared error (MSE) is defined as $MSE_j = \sum_{i=1}^n e_{i,j}^2/n$ for the j-th profile. The process of EWMA-3 that contains three control charts is described as follows:

Step 1: EWMA-3 for monitoring the intercept b_0 is established as follows:

1.1 Define $EWMA_0(j) = \theta e_0(j) + (1-\theta)EWMA_0(j-1)$, where $0 < \theta \leq 1$ is a smoothing constant and $EWMA_0(0) = 0$.

1.2 Define the lower control limit (LCL) and upper control limit (UCL) respectively as $LCL_1 = -L_0\sigma\sqrt{\theta/[(2-\theta)n]}$ and $UCL_1 = L_0\sigma\sqrt{\theta/[(2-\theta)n]}$, where $L_0 > 0$ is chosen to give a specified in-control ARL_0.

Step 2: EWMA-3 for monitoring the slop b_1 is established as follows:

2.1 Define $EWMA_1(j) = \theta e_1(j) + (1-\theta)EWMA_1(j-1)$, where $0 < \theta \leq 1$ is a smoothing constant and $EWMA_1(0) = 0$.

2.2 Define the lower control limit (LCL) and upper control limit (UCL) respectively as $LCL_2 = -L_1\sigma\sqrt{\theta/[(2-\theta)\sum_{i=1}^{n} x_i^{*2}]}$ and $UCL_2 = L_1\sigma\sqrt{\theta/[(2-\theta)\sum_{i=1}^{n} x_i^{*2}]}$, where $L_1 > 0$ is chosen to give a specified in-control ARL_0.

Step 3: EWMA-3 for monitoring the error variance is established as follows:

3.1 Define $EWMA_E(j) = \max\{\theta(MSE_j - 1) + (1-\theta)EWMA_E(j-1), 0\}$ with $0 < \theta \leq 1$ as a smoothing constant and $EWMA_E(0) = 0$

3.2 Define the upper control limit as $UCL_3 = L_E\sqrt{\theta Var(MSE_j)/(2-\theta)}$ where $Var(MSE_j) = Var(\sum_{i=1}^{n} e_{i,j}^2/n) = 2\sigma^4/n$ and $L_E > 0$ is chosen to give a specified in-control ARL_0.

3 SLP Model with Within-Profile Autocorrelated Error Terms

When the error terms in a SLP model satisfy the following autocorrelation structure

$$\epsilon_{i,j} = \rho\epsilon_{i-1,j} + a_{i,j},\ i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m, \tag{6}$$

where $a_{i,j}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m$ are independent and identically distributed normal random variables with mean 0 and variance σ^2 and the random error terms $\epsilon_{i,j}$ for $j = 1, 2, \cdots, m$ between any two different profiles are independent, then the SLP model is a SLP model with within-profile autocorrelation. Let $y'_{i,j} = y_{i,j} - \rho y_{i-1,j}$, then the model based on Eqs. (2) and (6) can be expressed as

$$y'_{i,j} = \gamma'_0 + \gamma'_1 x'_i + a_{i,j},\ i = 2, 3, \cdots, n,\ j = 1, 2, \cdots, m, \tag{7}$$

where $\gamma'_0 = \gamma_0(1-\rho)$, $\gamma'_1 = \gamma_1$, and $x'_i = x_i - \rho x_{i-1}$. If the model parameters are unknown, they are replaced with their corresponding maximum likelihood estimates. In the Phase II monitoring, the model parameters γ_0, γ_1, σ^2, and ρ are treated as known constants. We define the residuals for the j-th profile as $e_{i,j} = y'_{i,j} - \gamma'_0 - \gamma'_1 x'_i$, $i = 1, 2, \cdots, n$. Following the approach proposed by Kang and Albin [7] for monitoring independent SLP processes, Soleimani et al. [16] proposed the four control charts, T^2 method, T^2 based on the residuals, EWMA/R and EWMA-3, based on the residuals from the transformed model in Eq. (7) to

monitor SLP processes with within-profile autocorrelation. The first two Hotelling T^2 charts are stated as follows:

Hotelling T^2 Method for the Transformed Model

Step 1: Let $\hat{\gamma}'_{0,j}$ and $\hat{\gamma}'_{1,j}$ be the respective estimators of γ'_0 and γ'_1 based on the jth profile sample.

Step 2: Define $T_j^2 = [\hat{\gamma}'_{0,j}, \hat{\gamma}'_{1,j}] - [\gamma'_{0,j}, \gamma'_{1,j}]^T S^{-1} [\hat{\gamma}'_{0,j}, \hat{\gamma}'_{1,j}] - [\gamma'_{0,j}, \gamma'_{1,j}]$, where

$$S = \begin{bmatrix} \sigma^2(1/(n-1) + \bar{x}'^2/s_{x'x'}) & -\sigma_2 \bar{x}'/s_{x'x'} \\ -\sigma_2 \bar{x}'/s_{x'x'} & \sigma_2/s_{x'x'} \end{bmatrix}.$$

Step 3: Under in-control process, T^2 has a central chi-square distribution of two degree freedom. Therefore, the upper control limit (UCL) for the chart is $UCL = \chi^2_{2,\alpha}$, where $\chi^2_{2,\alpha}$ is the $100(1-\alpha)$-th percentile of the chi-square distribution with two degree of freedom.

Hotelling T^2 Method Based on the Residuals from the Transformed Model

Step 1: Define $T_j^2 = (\mathbf{e}_j - \mathbf{0})\Sigma^{-1}(\mathbf{e}_j - \mathbf{0})^T$ where $\mathbf{e}_j = (e_{1,j}, e_{2,j}, \cdots, e_{n,j})$, $\mathbf{0}$ is a $n \times 1$ vector of zeros, $\Sigma = \sigma^2\mathbf{I}$ and $\mathbf{I}$ is a $n \times n$ identity matrix.

Step 2: Define upper control limit as $UCL = \chi^2_{n-1,\alpha}$, which is the $100(1-\alpha)$-th percentile of the chi-square distribution with $n-1$ degree of freedom.

The EWMA-3 charts for the intercept, slope, and the variance of error terms [16] can be respectively constructed using the following three charting processes:

Control Chart 1: EWMA-3 Chart for Monitoring the Intercept

The first EWMA-3 chart, namely the EWMA-3I chart, is constructed using the following steps:

Step 1: Let $B_0 = \gamma'_0 + \gamma'_1\bar{x}'$, $B_1 = \gamma'_1$, $x''_i = (x'_i - \bar{x}')$, and express the model in Eq. (7) as

$$y'_{i,j} = B_0 + B_1 x''_i + a_{i,j},\ i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m. \tag{8}$$

Step 2: Derive the EWMA sequence of the B_0 estimates from the equation

$$\omega_I(j) = \theta b_{0,j} + (1-\theta)\omega_I(j-1),\ j = 1, 2, \cdots, m,$$

where $b_{0,j}$ is the least squares estimate obtained based on the observations in the j-th subgroup. Furthermore, $\omega_I(0) = B_0$, and θ $(0 < \theta \leq 1)$ is a smoothing constant. The control limits can be represented as

$$LCL_I = B_0 - L_I\sigma\sqrt{\frac{\theta}{(2-\theta)(n-1)}}$$

and

$$UCL_I = B_0 + L_I\sigma\sqrt{\frac{\theta}{(2-\theta)(n-1)}},$$

where $L_I(>0)$ is a control chart parameter.

Control Chart 2: EWMA-3 Chart for Monitoring the Slope
The second EWMA-3 chart, namely the EWMA-3S chart, is constructed using the following steps:

Step 1: On the basis of the transformations of B_0, B_1, and x_i'' in Eq. (8), the EWMA sequence of the B_1 estimates can be derived through

$$\omega_S(j) = \theta b_{1,j} + (1-\theta)\omega_S(j-1),\ \ j = 1, 2, \cdots, m,$$

where $\omega_S(0) = B_1$, and $b_{1,j}$ is the least squares estimate obtained based on the observations in the j-th subgroup.

Step 2: The control limits can be obtained as

$$LCL_S = B_1 - L_S\sigma\sqrt{\frac{\theta}{(2-\theta)S_{xx}}}$$

and

$$UCL_S = B_1 + L_S\sigma\sqrt{\frac{\theta}{(2-\theta)S_{xx}}},$$

where $S_{xx} = \sum_{i=2}^{n} x_i''^2$ and $L_S(>0)$ is a control chart parameter.

Control Chart 3: EWMA-3 Chart for Monitoring the Error Variance
The third chart, namely the EWMA-3e chart, is constructed using the following steps:

Step 1: Evaluate the MSE for each profile as

$$MSE_j = \frac{\sum_{i=2}^{n} e_{i,j}^2}{n-1},\ \ j = 1, 2, \cdots, m.$$

Step 2: The EWMA sequence of MSE_j can be derived through

$$\omega_E(j) = \max\{\theta(MSE_j - 1) + (1-\theta)\omega_E(j-1), 0\},\ \ j = 1, 2, \cdots, m,$$

where $\omega_E(0) = 0$. The upper control limit can be obtained as

$$UCL = L_E\sqrt{\frac{\theta Var(MSE_j)}{2-\theta}},$$

where

$$Var(MSE_j) = \frac{2\sigma^4}{n-1}$$

and $L_E(>0)$ is a control chart parameter.

The control chart parameters L_I, L_S, and L_E in the three EWMA-3 charts can be determined numerically such that a specified in-control average run length ARL_0 is reached.

3.1 MEWMA-SLP Chart

For the SLP model defined in Eqs. (2) and (6), the variance-covariance matrix of the error terms can be represented as $\sigma^2\Psi$, where

$$\Psi = \frac{1}{1-\rho^2}[\psi_{i,j}], \tag{9}$$

$\psi_{k,l} = 1$ for $k = l$, and $\psi_{k,l} = \rho^{(l-k)}$ for $k < l$. For each subgroup of size n, the SLP model in Eqs. (2) and (6) can be represented by $\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, where $\mathbf{Y}^T = (y_1, y_2, \cdots, y_n)$, $\mathbf{X} = (\mathbf{1}, \mathbf{x})$, $\mathbf{1}$ is a $n \times 1$ vector with entries 1, $\mathbf{x}^T = (x_1, x_2, \cdots, x_n)$, $\boldsymbol{\gamma}^T = (\gamma_0, \gamma_1)$, and $\boldsymbol{\epsilon}^T = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$. Because of Ψ is a positive definite matrix, the linear model can be expressed as $\mathbf{Y}_* = \mathbf{X}_*\boldsymbol{\gamma} + \boldsymbol{\epsilon}_*$, where $\mathbf{Y}_* = \Psi^{-1/2}\mathbf{Y}$, $\mathbf{X}_* = \Psi^{-1/2}\mathbf{X}$ and $\boldsymbol{\epsilon}_* = \Psi^{-1/2}\boldsymbol{\epsilon}$. Chiang et al. [3] mentioned that $E(\boldsymbol{\epsilon}_*) = \mathbf{0}$ and $Cov(\boldsymbol{\epsilon}_*) = \sigma^2 I$, where $\mathbf{0}$ is an order n column vector with entries 0 and I is an order n identity matrix. The generalized least squares (GLS) estimator of the model parameters γ_0 and γ_1, $\boldsymbol{\gamma}_G^T = (\hat{\gamma}_{G0}, \hat{\gamma}_{G1})$, of $\boldsymbol{\gamma}^T$, can be obtained as

$$\hat{\boldsymbol{\gamma}}_G = (\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T\mathbf{Y} = (\mathbf{X}^T\Psi^{-1}\mathbf{X})^{-1}\mathbf{X}^T\Psi^{-1}\mathbf{Y}. \tag{10}$$

Chiang et al. [3] constructed the MEWMA-SLP chart based on the design of Lowry et al. [9] using the GLS estimates obtained from the Phase I samples of SLP process. The control chart procedure can be described as follows:

MEWMA-SLP Chart

Step 1: Obtain the GLS estimates $\hat{\boldsymbol{\gamma}}_{G,j}$ for $j = 1, 2, \cdots, m$ using Eq. (10) and the Phase I samples of SLP process.

Step 2: Compute $\bar{\hat{\boldsymbol{\gamma}}}_G = \sum_{j=1}^{m} \hat{\boldsymbol{\gamma}}_{G,j}/m = (\bar{\hat{\gamma}}_{G0}, \bar{\hat{\gamma}}_{G1})^T$. Let S_0 and S_1 denote the sample standard deviation of $\hat{\gamma}_{G0,j}$ and $\hat{\gamma}_{G1,j}$, $j = 1, 2, \cdots, m$, respectively, and let $\mathbf{U}_j$ be the rescaled vector of $\hat{\boldsymbol{\gamma}}_{G,j}$ for $j = 1, 2, \cdots, m$, where the rescaling involves the use of $\bar{\hat{\boldsymbol{\gamma}}}_G$, S_0, and S_1. The sample variance-covariance matrix of $\mathbf{U}_j$'s, denoted by $\mathbf{S}_\mathbf{U}$, can be obtained as

$$\mathbf{S}_\mathbf{U} = \frac{1}{m-1} \sum_{j=1}^{m} (\mathbf{U}_j - \bar{\mathbf{U}})(\mathbf{U}_j - \bar{\mathbf{U}})^T,$$

where $\bar{\mathbf{U}}$ is the sample mean of $\mathbf{U}_j$, $j = 1, 2, \cdots, m$.

Step 3: Derive the chart parameters κ ($0 < \kappa \leq 1$), and $UCL(= H)$ from the values suggested by Lowry et al. [9] for an in-control $ARL_0 = 200$. Useful parameter combinations are outlined as (κ, H) = (0.2, 9.65), (0.4, 10.29), (0.6, 10.53) and (0.8, 10.58). Other chart parameters can be obtained from the study of Lowry et al. [9] or through Monte Carlo simulations.

Step 4: Compute the EWMA series of $\mathbf{U}_j$ from

$$\mathbf{Z}_j = \kappa \mathbf{U}_j + (1-\kappa)\mathbf{Z}_{j-1}, \ \ j = 1, 2, \cdots, m,$$

and the test statistics from

$$T_i^2 = \mathbf{Z}_i^T \mathbf{S}_{\mathbf{Z}_i}^{-1} \mathbf{Z}_i, \ \ j = 1, 2, \cdots, m,$$

where

$$\mathbf{S}_{\mathbf{Z}_j} = \frac{\kappa}{2-\kappa}[1-(1-\kappa)^{2j}]\mathbf{S}_\mathbf{U} \rightarrow \frac{\kappa}{2-\kappa}\mathbf{S}_\mathbf{U} \text{ as } j \text{ goesto } \infty.$$

4 Process Capability Indices

The process capability indices C_P and C_{PK} are widely used to evaluate the capability of a univariate process. These indices are defined as $C_P = (U-L)/6\sigma$ and $C_{PK} = \min\{U-\mu, \mu-L\}/3\sigma$, where L and U are the lower and upper specification limits of the univariate process characteristic quality and μ and σ are the mean and standard deviation of the process, respectively. For two-dimensional processes, C_P and C_{PK} are denoted by BC_P and BC_{PK}, respectively. The evaluations of BC_P and BC_{PK} for SLP processes are described as follows.

Let $\hat{\boldsymbol{\gamma}}_{G,j}$, $j = 1, 2, \cdots, m$ be the GLS estimators of the of the model parameters $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)$ obtained by using the j-th in-control subgroups of the SLP process and the specification limits of γ_0 and γ_1 be labeled as (L_1', U_1') and (L_2', U_2'), respectively. Chiang et al. [3] improved the method proposed by Castagliola and Garcia Castellanos [2] through the following two steps:

Step 1: Let $\mathbf{U}_j$ be the bivariate vectors defined in the Step 2 of MEWMA-SLP, and let $L_1 = (L_1' - \bar{\hat{\gamma}}_{G0})/S_1$, $U_1 = (U_1' - \bar{\hat{\gamma}}_{G0})/S_1$, $L_2 = (L_2' - \bar{\hat{\gamma}}_{G1})/S_2$, and $U_2 = (U_2' - \bar{\hat{\gamma}}_{G1})/S_2$. Derive the bivariate process capability indices from $\mathbf{U}_j$, $j = 1, 2, \cdots, m$, L_1, U_1, L_2, and U_2.

Step 2: Use an orthogonal decomposition method to obtain the variance-covariance matrix of $\mathbf{U}_j$, $j = 1, 2, \cdots, m$. The variance-covariance matrix is given by

$$\mathbf{S}_{\mathbf{U}} = \mathbf{R}\Lambda\mathbf{R}^T, \tag{11}$$

where Λ is a diagonal matrix of rank two with eigenvalues $\lambda_1 < \lambda_2$ of $\mathbf{S}_{\mathbf{U}}$ as the diagonal elements. Furthermore, $\mathbf{R} = [\mathbf{r}_1^T, \mathbf{r}_2^T]$, where $\mathbf{r}_1^T = (r_{1,1}, r_{2,1})$ and $\mathbf{r}_2^T = (r_{1,2}, r_{2,2})$ are the eigenvectors corresponding to λ_1 and λ_2, respectively.

Let D_1 and D_2 be two lines passing through the point $G(\hat{\gamma}_{G0}, \hat{\gamma}_{G1})$ in the directions of the eigenvectors $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. The lines D_1 and D_2 split the $(\hat{\gamma}_{G0}, \hat{\gamma}_{G1})$-plane into four disjoint areas, which are denoted by A_1, A_2, A_3, and A_4 (see, Fig. 1). Because of $\hat{\boldsymbol{\gamma}}_G$ follows a bivariate normal (BVN) distribution which is symmetric about its mean $\boldsymbol{\beta}$, we can show that $P(\hat{\boldsymbol{\gamma}}_G \in A_i) = 1/4$, $i = 1, 2, 3$, and 4. Let

$$A = \{(\hat{\gamma}_{G0}, \hat{\gamma}_{G1}) | L_1 \le \hat{\gamma}_{G0} \le U_1,\ L_2 \le \hat{\gamma}_{G1} \le U_2\} = Q_1 \cup Q_2 \cup Q_3 \cup Q_4,$$

where $Q_i = A \cap A_i$ for $i = 1, 2, 3$, and 4. Let $q = P(\hat{\boldsymbol{\gamma}}_G \in A)$ and $p = 1 - q$, where p is the proportion of non-conformity. Furthermore, let $q_i = P(\hat{\boldsymbol{\gamma}}_G \in Q_i)$ and $p_i = P(\hat{\boldsymbol{\gamma}}_G \in A_i) - q_i = 1/4 - q_i$ for $i = 1, 2, 3$, and 4. Hence, $q =$

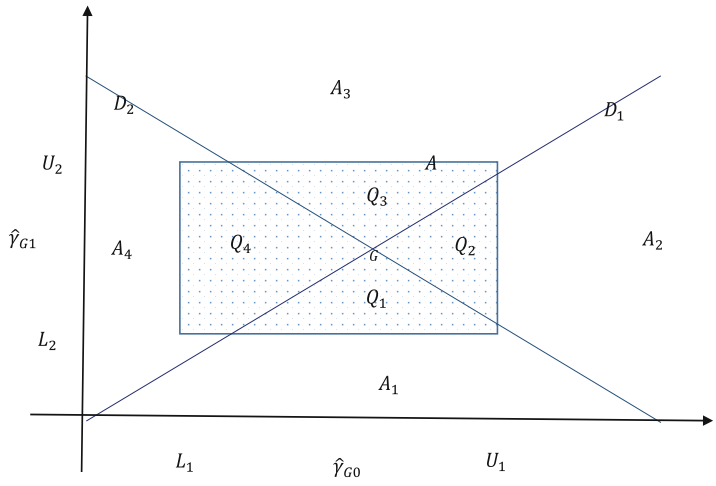


Fig. 1 The lines D_1 and D_2, and polygons Q_i for $i = 1, 2, 3, 4$

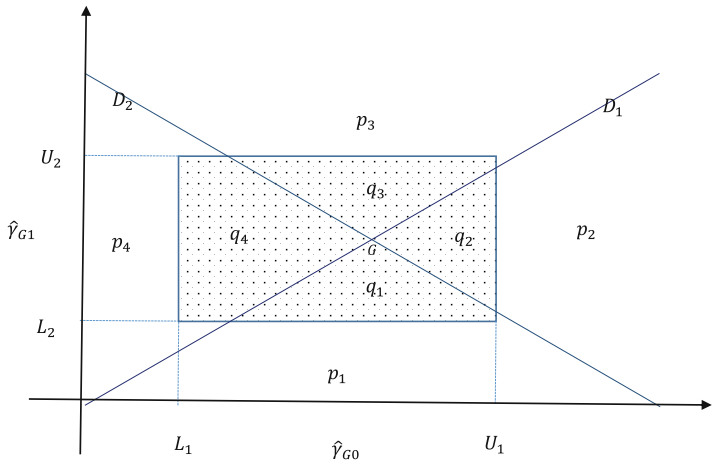


Fig. 2 The probabilities of p_i and q_i for $i = 1, 2, 3, 4$

$q_1 + q_2 + q_3 + q_4$ and $p = p_1 + p_2 + p_3 + p_4$ (Fig. 2). When the means of $\hat{\gamma}_{G0}$ and $\hat{\gamma}_{G1}$ are considered as the midpoints of their respective specification limits, then $p_1 = p_2 = p_3 = p_4 = p/4$ and $q_1 = q_2 = q_3 = q_4 = q/4$. According to the method presented in [2], the process capability indices for the GLS estimator $\hat{\boldsymbol{\gamma}}_G$ can be defined as

$$BC_P = -\frac{1}{3}\Phi^{-1}(p/2)$$

and

$$BC_{PK} = \min\{-\Phi^{-1}(2p_1), -\Phi^{-1}(2p_2), -\Phi^{-1}(2p_3), -\Phi^{-1}(2p_4)\},$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Chiang et al. [3] suggested that the probabilities p_i and q_i, $i = 1, 2, 3$, and 4, can be obtained using the R packages `mvtnorm:pmvnorm` [4] and `pracma:triquad` [1] and they proposed the following two algorithms to obtain BC_P and BC_{PK}.

Algorithm A: Evaluation of BC_{PK}

Step 1: Obtain $\mathbf{U}_j$, L_1, U_1, L_2, and U_2, then evaluate the sample variance-covariance matrix $\mathbf{S}_\mathbf{U}$ from $\mathbf{U}_j$, $j = 1, 2, \cdots, m$.

Step 2: Find the diagonal matrix Λ of eigenvalues and the matrix of eigenvectors $\mathbf{R}$ from the orthogonal decomposition $\mathbf{S}_\mathbf{U} = \mathbf{R}\Lambda\mathbf{R}^T$.

Step 3: Let $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_\ell$ be the vertices of the polygon Q_i, then obtain the transformed vertices of Q_i through $\mathbf{c}_l' = \mathbf{R}\Lambda^{-1/2}\mathbf{R}^T\mathbf{c}_l$ for $l = 1, 2, \cdots, \ell$.

Step 4: Determine the probabilities $\hat{q}_i$ and $p_i (= 1/4 - q_i)$ for $i = 1, 2, 3$, and 4.

Step 5: Estimate BC_{PK} as

$$\hat{BC}_{PK} = \frac{1}{3}\min\{-\Phi^{-1}(2\hat{p}_1), -\Phi^{-1}(2\hat{p}_2), -\Phi^{-1}(2\hat{p}_3), -\Phi^{-1}(2\hat{p}_4)\}.$$

Algorithm B: Evaluation of BC_P

Step 1: Let $\hat{\boldsymbol{\gamma}}_{G,j}$ for $j = 1, 2, \cdots, m$ follow a $BVN(\boldsymbol{\mu}_B, \mathbf{S}_{\boldsymbol{\gamma}})$ distribution with mean $\boldsymbol{\mu}_B = ((L_1' + U_1')/2, (L_2' + U_2')/2)$ and variance-covariance matrix

$$\mathbf{S}_{\boldsymbol{\gamma}} = \frac{1}{m-1}\sum_{j=1}^{m}(\hat{\boldsymbol{\gamma}}_{G,j} - \boldsymbol{\mu}_B)(\hat{\boldsymbol{\gamma}}_{G,j} - \boldsymbol{\mu}_B)^T,$$

and denote the probability density function of $BVN(\boldsymbol{\mu}_B, \mathbf{S}_{\boldsymbol{\gamma}})$ distribution as $f(\hat{\gamma}_0, \hat{\gamma}_1)$.

Step 2: Evaluate the probability

$$\hat{q} = \int_{L_1'}^{U_1'}\int_{L_2'}^{U_2'} f(t_1, t_2)dt_1dt_2,$$

and $\hat{p} = 1 - \hat{q}$.

Step 3: Estimate BC_P as

$$\hat{BC}_P = -\frac{1}{3}\Phi^{-1}(\hat{p}/2).$$

5 Monte Carlo Simulation Study and Numerical Example

5.1 Monte Carlo Simulation Study

To evaluate the performance of the MEWMA-SLP chart, Chiang et al. [3] used the model settings as described in [15]. Let $\gamma_0 = 3$ and $\gamma_1 = 2$ for the models in Eqs. (2) and (6). The points set for Eq. (2) are $x_i = 2, 4, 6$, and 8. Because Soleimani et al. [16] revealed that the EWMA-3 charts outperform other competitive methods, we only compare the performance of the EWMA-3 charts with the MEWMA-SLP chart.

The EWMA-3 charts use EWMA-3I, EWMA-3S and EWMA-3e charts to simultaneously monitor the intercept, slope, and error variance of a SLP model. The overall false alarm rate of EWMA-3 charts can be determined as

$$\alpha_{EWMA} = 1 - (1 - \alpha_1')(1 - \alpha_2')(1 - \alpha_3'),$$

where α_1', α_2' and α_3' are the false alarm rates for EWMA-3I, EWMA-3S and EWMA-3e, respectively. Chiang et al. [3] used the MEWMA-SLP chart and the EWMA-3e chart simultaneously to monitor the intercept, slope, and error variance of a SLP model. The overall false alarm rate of can be obtained as

$$\alpha_{MEWMA} = 1 - (1 - \alpha_1)(1 - \alpha_2),$$

where α_1 and α_2 are the false alarm rates of the MEWMA-SLP and EWMA-3e charts, respectively. Because the EWMA-3e chart is used to monitor the error variance for both approaches, the performance of the EWMA-3e chart is omitted in the comparison. Specifically, the error variance was assumed to be in a statistical control state in the simulation study. Hence, Chiang et al. [3] only compared the performance of monitoring SLP process based on the EWMA-3I and EWMA-3S charts with the performance of monitoring SLP process based on the MEWMA-SLP chart.

In the simulation study, the correlation coefficient between the adjacent error terms, ρ, is set to be 0.1, 0.7 or 0.9, and the overall ARL_0 is set to be 200 (i.e., the overall false alarm rate $\alpha = 0.005$). Without loss of generality, let $\alpha_1' = \alpha_2' = 1 - \sqrt{1-\alpha} \cong 0.0025$ for the EWMA-3I and EWMA-3S charts, and let $\alpha_1 = \alpha_2 = 0.005$ for the MEWMA-SLP chart. The chart parameters of the MEWMA-SLP chart are set as $\kappa = 0.2$ and $H = 9.65$.

In Phase I monitoring for the model parameter, estimation of the parameters is involved in using the EWMA-3I and EWMA-3S charts and 10,000 in-control SLP processes were prepared. A simulation with 10,000 iterations was conducted to obtain the chart parameters, which are $L_I = 2.889$ and $L_S = 2.895$, for using the EWMA-3I and EWMA-3S simultaneously to achieve an approximated overall ARL_0 of 200. The EWMA-3 charts and the MEWMA-SLP chart were constructed in Phase I. The out-of-control average run length, denoted by ARL_1, was evaluated for different parameter shifts in the Phase II monitoring.

Two scenarios were considered for the parameter shift. Scenario I (SI) was set to be same as the simulation setting of Soleimani et al. [16]. In SI, the shift in either the intercept or the slope of a SLP model was considered, and the other parameter kept unchanged. In scenario II (SII), the shifts in both the intercept and the slope of a SLP model were considered simultaneously. Assume that the intercept shifts from γ_0 to $\gamma_0' = \gamma_0 + \lambda\sigma$ and the slope shifts from γ_1 to $\gamma_1' = \gamma_1 + \beta\sigma$ when an assignable cause is introduced. Here, λ and β are two constants. The parameter combinations for SI are (1) $\lambda = 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8$ and 2, and $\beta = 0$; and (2) $\beta = 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225$ and 0.25, and $\lambda = 0$.

In many practical applications, an intercept shift often accompanies with a slope shift for a SLP process. Figure 3 presents an example, in which the nominal regression line is characterized by the conditional expected value of y, given x, as follows:

$$\mu(y|x) = \gamma_0 + \gamma_1 x. \tag{12}$$

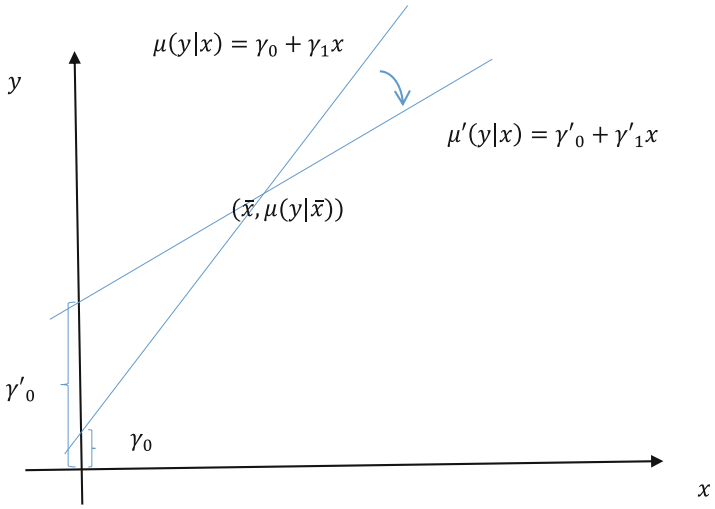


Fig. 3 An example of an intercept shift accompanies a slope shift for a SLP

When the profile parameters shift because of an assignable cause, the nominal regression line shifts to

$$\mu'(y|x) = \gamma_0' + \gamma_1' x. \tag{13}$$

In Fig. 3, both $\mu(y|x)$ and $\mu'(y|x)$ pass through the point $(P_x, P_y), = (\bar{x}, \mu(y|\bar{x}))$. Using $(P_x, P_y) = (\bar{x}, \mu(y|\bar{x}))$ in Eqs. (12) and (13), we can obtain

$$P_y = \gamma_0 + \gamma_1 \bar{x}$$

and

$$\gamma_0' = (\gamma_0 + \gamma_1 \bar{x}) - \gamma_1' \bar{x}. \tag{14}$$

Because $\gamma_0 = 3$, $\gamma_1 = 2$, and $x_i = 2, 4, 6, 8$, Eq. (14) reduces to

$$\gamma_0' = 13 - 5\gamma_1'. \tag{15}$$

To present the process shifts on the σ scale, Chiang et al. [3] obtained $\gamma_0' = \gamma_0 + \lambda\sigma = 3 + \lambda$ and $\gamma_1' = \gamma_1 + \beta\sigma = 2 + \beta$. Subsequently substituting $\gamma_0' = 3 + \lambda$ and $\gamma_1' = 2 + \beta$ into Eq. (15) yields the relationship between λ and β as $\lambda = -5\beta$. In the study conducted by Chiang et al. [3], $\beta = -0.50, -0.46, -0.44, -0.42, -0.40, -0.38, -0.36, -0.34, -0.32, -0.30, -0.28, -0.26, -0.24, -0.22, -0.20, -0.18, -0.16, -0.14, -0.12, -0.10, -0.08, -0.06, -0.04$, and -0.02 in the simulation for determining the speed of control charts

for detecting process shifts. The values of λ can be obtained using the relation $\lambda = -5\beta$. The out-of-control ARL, denoted by ARL_1, was evaluated through a simulation with 10,000 realizations. A single alarm of the EWMA-3I or the EWMA-3S indicates that the process is out of control for EWMA-3. The value of ARL_1 obtained from this charting procedure is labeled as ARL_1-$E3$, and the value of ARL_1 determined from MEWMA-SLP is labeled as ARL_1-ME.

The Monte Carlo simulation study presented in [3] clearly showed that the MEWMA-SLP chart outperforms EWMA-3I and EWMA-3S simultaneous charting. The MEWMA-SLP chart is faster than the combined EWMA-3I and EWMA-3S charts in detecting the process shift for the two scenarios considered. In particular, the simulation results show that the MEWMA-SLP chart improves the performance of the combined EWMA-3I and EWMA-3S charts considerably when ρ is close to 1. When $\rho = 0.9$, the sensitivity of the combined EWMA-3I and EWMA-3S charts in detecting a process shift engendered by an intercept shift was low. More detail information, interested readers are suggested to read Chiang et al. [3].

5.2 Numerical Example

When a SLP process is identified as in-control, the process capability should be evaluated. To illustrate the application of the two new process capability indices, Chiang et al. [3] generated data sets of size $m = 200$ with ρ =0.1, 0.7, and 0.9 based on the model in Eqs. (2) and (6) with $\gamma_0 = 3$ and $\gamma_1 = 2$. Let the lower and upper specification limits of the response variable y be LSL_y and USL_y, respectively. If the mean of y, given $x = \bar{x}$, is the midpoint of the specification limits, then LSL_y and USL_y can be expressed as

$$LSL_y = E(y|x = \bar{x}) - k\sigma = \gamma_0 + \gamma_1\bar{x} - k_1\sigma \tag{16}$$

and

$$USL_y = E(y|x = \bar{x}) + k\sigma = \gamma_0 + \gamma_1\bar{x} + k_2\sigma, \tag{17}$$

respectively. Using Eqs. (16) and (17), Chiang et al. [3] obtained

$$\gamma_0 = \frac{LSL_y + USL_y - (k_2 - k_1)\sigma}{2} - \gamma_1\bar{x},$$

$$\gamma_1 = \frac{\frac{LSL_y + USL_y - (k_2 - k_1)\sigma}{2} - \gamma_0}{\bar{x}},$$

$LSL_y - (k_2 - k_1)\sigma/2 - \gamma_1\bar{x} \le \gamma_0 \le USL_y - (k_2 - k_1)\sigma/2 - \gamma_1\bar{x}$ and $(LSL_y - (k_2 - k_1)\sigma/2 - \gamma_0)/\bar{x} \le \gamma_1 \le (USL_y - (k_2 - k_1)\sigma/2 - \gamma_0)/\bar{x}$. On the basis of the specification limits of the response variable, the specifications limits of $\hat{\gamma}_{G0}$ and $\hat{\gamma}_{G1}$ can be expressed as

$$LSL_{G0} = LSL_y - \gamma_1\bar{x},$$

$$USL_{G0} = USL_y - \gamma_1\bar{x},$$

and

$$LSL_{G1} = \frac{LSL_y - \gamma_0}{\bar{x}},$$

$$USL_{G1} = \frac{USL_y - \gamma_0}{\bar{x}}.$$

When the parameters γ_0 and γ_1 are unknown, the parameters γ_0 and γ_1 can be replaced with the sample means of $\hat{\gamma}_{G0,j}$ and $\hat{\gamma}_{G1,j}$, respectively.

In the numerical examples, let $k_1 = 4$, $k_2 = 3.5$, $LSL_y = 9$, $USL_y = 16.5$, $LSL_{G0} = LSL_y - \gamma_1\bar{x} = -0.75$, $USL_{G0} = USL_y - \gamma_1\bar{x} = 6.75$, $LSL_{G1} = LSL_y - \gamma_0/\bar{x} = 1.25$, and $USL_{G1} = USL_y - \gamma_0/\bar{x} = 2.75$. Chiang et al. [3] studied the relationship between ρ and the process capability indices BC_p and BC_{pk}. The study showed that BC_p and BC_{pk} were influenced by the with-profile correlation and they are overestimated if the with-profile correlation was ignored. Furthermore, the maximal values of BC_p and BC_{pk} were attained for $\rho = 0$. The value of BC_{pk} decreases as ρ increases and the values of BC_p and BC_{pk} were closed because the mean of $\hat{\boldsymbol{\gamma}}_{G,j}$ for $j = 1, 2, \cdots, m$ were closed to the midpoint of the specification limits. For more information, interested readers are suggested to read Chiang et al. [3].

6 Conclusion

The main purpose of this chapter is to review the recent developments on the SPC to monitor SLP processes with correlated error terms; specially, SLP processes with within-profile autocorrelation. The common skill is to transform the SLP model with autocorrelated error terms to a SLP model with independent error terms and independent residuals so that the control charts for independent errors can be applied. Interested readers are suggested to read Chiang et al. [3] for potential research directions.

References

1. Borchers, H. W. (2018). *pracma: Practical numerical math functions*. R package version 2.1.8. https://CRAN.R-project.org/package=pracma
2. Castagliola, P., & Garcia Castellanos, J. V. (2005). Capability indices dedicated to the two quality characteristics case. *Quality Technology and Quantitative Management, 2*, 201–220.
3. Chiang, J.-Y., Lio, Y. L., & Tsai, T.-R. (2017). MEWMA control chart and process capability indices for simple linear profiles with within-profile autocorrelation. *Journal of Quality and Reliability Engineering International, 33*, 1083–1094.
4. Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., et al. (2018). *mvtnorm: Multivariate normal and t distributions*. R package version 1.0-8. http://CRAN.R-project.org/package=mvtnorm
5. Gupta, S., Montgomery, D., & Woodall, W. (2006). Performance evaluation of two methods for online monitoring of linear calibration profiles. *International Journal of Production Research, 44*, 1927–1942.
6. Jensen, W. A., Brich, J. B., & Wood, W. H. (2008). Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology, 40*, 167–183.
7. Kang, L., & Albin, S. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology, 32*, 418–426.
8. Kim, K., Mahmoud, M. A., & Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology, 35*, 317–328.
9. Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics, 34*, 46–53.
10. Mahmoud, M. A., Parker, P. A., Woodall, W. H., & Hawkins, D. M. (2007). A change point method for linear profile data. *Quality and Reliability Engineering International, 23*, 247–268.
11. Mestek, O., Pavlik, J., & Suchánek, M. (1994). Multivariate control charts: Control charts for calibration curves. *Fresenius' Journal of Analytical Chemistry, 350*, 344–351.
12. Noorossana, R., Amiri, A., & Soleimani, P. (2008). On the monitoring of autocorrelated linear profiles. *Communications in Statistics–Theory and Methods, 37*, 425–442.
13. Noorossana, R., Amiri, A., Vaghefi, S. A., & Roghanian, E. (2004). Monitoring quality characteristics using linear profile. In *Proceedings of the 3rd International Industrial Engineering Conference, Tehran, Iran*.
14. Saghaei, A., Mehrjoo, M., & Amiri, A. (2009). A CUSUM-based method for monitoring simple linear profiles. *The International Journal of Advanced Manufacturing Technology, 5*, 1252–1260.
15. Shahriari, H., Huebele, N. F., & Lawrence, F. P. (1995). A multivariate process capability vector. In *Proceeding of the 4th Industrial Engineering Research Conference* (pp. 304–309). Norcross: Institute of Industrial Engineers.
16. Soleimani, P., Noorossana, R., & Amiri, A. (2009). Simple linear profiles monitoring in the presence of within profile autocorrelation. *Computers & Industrial Engineering, 57*, 1015–1021.
17. Stover, F. S., & Brill, R. V. (1998). Statistical quality control applied to ion chromatography calibrations. *Journal of Chromatography A, 804*, 37–43.
18. 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*, 309–320.
19. Zou, C., Zhang, Y., & Wang, Z. (2006). A control chart based on a change-point model for monitoring linear profiles. *IIE Transactions, 38*, 1093–1103.

Sequential Monitoring of Circular Processes Related to the von Mises Distribution

Cornelis J. Potgieter

Abstract The sequential monitoring of circular processes related to the von Mises distribution are considered. More specifically, methods for detecting changes in location and/or scale are considered when a process has in-control and out-of-control behavior following a von Mises distribution. Results on existing cumulative sum (cusum) charts are reviewed, and new sequential changepoint methods are developed. These are compared using Monte Carlo simulations. Finally, the sequential monitoring of a process with in-control distribution that is circular uniform is considered. An existing nonparametric cusum is reviewed and is compared to a new sequential changepoint method designed for a von Mises alternative.

1 Introduction

Circular data arise in a variety of settings. One of the obvious examples of data that can be expressed as a point on the circumference of circle is angular direction from a reference point, say the direction of the wind measured at a weather station, or the direction an animal is moving in. In astrophysics, the arrival times of radio pulses from neutron stars are used to study the interior physics of these stars. The pulsar data are "folded" relative to the instantaneous pulse period [10, Chapter 7]. The resulting data are points on a circle. Another example is in the study of biological systems where many processes follow a circadian rythm. For example, the time of

Electronic Supplementary Material The online version of this chapter (https://doi.org/10.1007/978-3-030-20709-0_6) contains supplementary material, which is available to authorized users.

C. J. Potgieter (✉)
Department of Mathematics, Texas Christian University, Forth Worth, TX, USA

Department of Statistics, University of Johannesburg, Johannesburg, South Africa
e-mail: c.potgieter@tcu.edu

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_6

day that systolic blood pressure attains a maximum on a given day is called the acrophase. This acrophase data are also circular.

The geometry of the circle poses some unique challenges when analyzing data. Methods for linear data can fail. Consider the hypothetical example of observing angles 5° and 355° on a circle [0°, 360°). The linear average of these two points is 180°, while the circular average is 0°. For circular data $X_1, \ldots, X_n \in [-\pi, \pi)$, define

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n}\cos(X_i) \qquad \text{and} \qquad \bar{S} = \frac{1}{n}\sum_{i=1}^{n}\sin(X_i),$$

then the mean direction is given by $\bar{X} = \text{atan2}(\bar{S}, \bar{C})$ where atan2 is the two-argument arctan function with range $(-\pi, \pi]$. Similarly, the circular variance is $V = 1 - \bar{R}$ with $\bar{R} = \left(\bar{S}^2 + \bar{C}^2\right)^{1/2}$ the mean resultant length. For comprehensive overviews of the analysis of circular data along with many commonly encountered parametric models, see the monographs by Fisher [1] and Mardia & Jupp [11].

The analysis of sequential circular data has received limited attention in the literature. The retrospective analysis of circular data to identify changepoints was considered by Fisher [1, Chapter 7] and Ghosh et al. [3], with Lombard [6] taking a nonparametric approach, and Sengupta & Laha [14] deriving a Bayesian solution.

However, with data arriving sequentially, and having as objective the detect of a change soon after it has occurred, the changepoint methodology referenced above is not adequate. Solutions for sequential data for an underlying von Mises distribution were proposed by Gadsen & Kanji [2], who developed a sequential probability ratio test for a location parameter, and by Hawkins & Lombard [4], who developed cusum procedures for location and concentration changes. Lombard & Maxwell [7] developed a nonparametric cusum to detect a deviation from angular uniformity. A rank-based cusum procedure was developed by Lombard et al. [9], and Lombard et al. [8] developed self-starting cusums for location and concentration changes.

In the present setting, a parametric approach will be considered. Specifically, methodology will be developed for an underlying von Mises distribution. However, extensions to other distributions such as the cardioid and wrapped Cauchy will also be discussed. In Sect. 2, sequential monitoring for a location change is considered when both the in-control and out-of-control distributions are von Mises. The existing cusum of Hawkins & Lombard [4] is reviewed and a new sequential changepoint (SCP) procedure is also developed. Section 3 extends the SCP methodology to monitoring for both a location and concentration change. The power and robustness properties of the procedures are also considered. Section 4 considers sequential monitoring of a process that follows a circular uniform distribution when in control. The nonparametric cusum of Lombard & Maxwell [7] is reviewed, and a parametric SCP is designed for out-of-control behavior following a von Mises distribution. Some concluding remarks are contained in Sect. 5.

2 Monitoring for a Location Change

Assume sequential observations $X_1, X_2, X_3, \ldots$ arise from a von Mises distribution which has density function

$$f(x|\mu,\kappa) = \frac{\exp[\kappa\cos(x-\mu)]}{2\pi I_0(\kappa)},\ -\pi \le x < \pi \tag{1}$$

where μ and κ denote the location and concentration parameters, and $I_p(\kappa)$ denotes the modified Bessel function of the first kind of order p,

$$I_p(t) = \sum_{k=0}^{\infty} \frac{(t/2)^{p+2k}}{k!\Gamma(p+k+1)}.$$

Small values of κ indicate that the distribution is very diffuse around the mean direction μ. On the other hand, large values of κ indicate that the distribution is concentrated around the mean direction. The special case $\kappa = 0$ corresponds to the uniform distribution on the circle for which the mean direction is undefined.

In this section, methods for monitoring the sequence to detect a change in the location parameter μ will be considered. Specifically, an existing cusum procedure will be reviewed, and a new sequential changepoint procedure will be developed. The two procedures will also be compared for various out-of-control states.

2.1 A Cusum Procedure

Cumulative sum (cusum) procedures are popular for sequential process monitoring. This popularity stems in part from their computational ease. Whenever a new data point X_n is collected, the new value of the cusum at time n can be expressed as the value of the cusum at time $n-1$ plus a function of the data point X_n.

Consider now the task of monitoring a process with sequential observations arising from a von Mises distribution. When monitoring this process to determine if/when it goes out of control, the in-control parameters, in this case (μ, κ), are generally known. It can be assumed without loss of generality that the in-control mean is $\mu = 0$. The cumulative sum (cusum) control chart proposed by Hawkins & Lombard [4] is then defined by $C_0 = 0$ and

$$\begin{aligned} C_n &= \max\left[0, C_{n-1} + \log\frac{f(X_n|\delta,\kappa)}{f(X_n|0,\kappa)}\right] \\ &= \max\left[0, C_{n-1} + \kappa\cos(X_n-\delta) - \kappa\cos(X_n)\right], \quad n \ge 1. \end{aligned} \tag{2}$$

Here, δ is a pre-specified minimal change in location that needs to occur before the process is considered to be out of control. The stopping time is defined as $N = \min\{n|C_n \geq h\}$, the first time the process exceeds the control limit h. The in-control average run-length is given by $\text{ICARL} = \text{E}_0[N]$ where the subscript 0 indicates that the expectation is evaluated using in-control parameter values. In practice, the cusum is calibrated in such a way that the control limit h corresponds to a desired in-control run-length. The resulting control limit also depends on the minimal change δ and the concentration parameter κ.

Note that the defined cusum C_n is one-sided, meaning that it is designed to detect either only an upward or downward shift in location, depending on the sign of δ. Now assume $\delta > 0$ and that it is desirable to detect a shift of size δ in either direction, a two-sided cusum should be used. To this end, define upward and downward cusums with $C_0^+ = C_0^- = 0$, and for $n \geq 1$,

$$C_n^+ = \max\left[0, C_{n-1}^+ + \kappa\cos(X_n + \delta) - \kappa\cos(X_n)\right],$$
$$C_n^- = \max\left[0, C_{n-1}^- + \kappa\cos(X_n - \delta) - \kappa\cos(X_n)\right].$$

The stopping time for the two-sided procedure is defined as the first time n such that one of the two cusums signals, $N = \min\left\{n|\max(C_n^+, C_n^-) \geq h\right\}$.

Hawkins & Lombard [4] calibrated the one-sided cusum for a range of δ and κ values and tabulate the resulting control limits. To calibrate the two-sided cusum, they recommended using the one-sided control limits along with the conventional equation

$$\text{ARL}_{\text{two-sided}} = \left(\frac{1}{\text{ARL}_{\text{up}}} + \frac{1}{\text{ARL}_{\text{down}}}\right)^{-1} \tag{3}$$

where ARL_{proc} denotes the average run-length for the "proc" cusum. However, application of (3) requires that the upward and downward cusums satisfy the necessary and sufficient non-interference condition of [16]. Specifically, if one of the two cusums signals that the process is out of control, it is required that the companion cusum has value 0. This does not hold for the von Mises distribution. Specifically, interference is present for circular data as an upward shift of δ can also be characterized as a downward shift of $2\pi - \delta$. Provided here in Table 1 are control limits for the two-sided cusum procedure for $\kappa \in \{0.5, 1, 2\}$, $\text{ICARL} \in \{100, 500, 1000\}$ and a variety of δ-values.

The values in Table 1 were found using 100,000 simulated realizations of C_n^+ and C_n^- for each set of parameter configurations. It should also be noted that the two-sided control limits in Table 1 are generally very close to the control limits found using (3). This suggests that the von Mises location cusum is robust against departure from the von Dobben de Bruyn non-interference conditions.

Table 1 Monte Carlo control limits for the two-sided von Mises location cusum

κ	δ	ICARL		
		100	500	1000
0.5	$\pi/8$	1.325	2.413	2.979
	$\pi/4$	2.081	3.428	4.072
	$3\pi/8$	2.546	3.998	4.662
	$\pi/2$	2.825	4.337	5.016
1	$\pi/8$	2.050	3.393	4.031
	$\pi/2$	2.905	4.403	5.075
	$3\pi/8$	3.343	4.896	5.580
	$\pi/2$	3.592	5.162	5.850
2	$\pi/8$	2.759	4.238	4.907
	$\pi/4$	3.528	5.089	5.774
	$3\pi/8$	3.846	5.408	6.093
	$\pi/2$	3.895	5.488	6.179

2.2 A Sequential Changepoint Procedure

Cusum procedures such as those discussed in Sect. 2.1 are generally computationally expedient, as they are based on single-step updating as each new data point is observed. However, with increased computational abilities, alternative monitoring procedures should also be considered. Let $f_0(x)$ and $f_1(x)$ denote, respectively, the density functions of data under in-control and out-of-control conditions and define stopping rule

$$N = \min\left\{ n \,\middle|\, \max_{0\le k<n} \sum_{i=k+1}^{n} \log\left[f_1(X_i)/f_0(X_i)\right] \ge h_n \right\} \tag{4}$$

where h_n is some sequence of control limits. When the density function f_1 is fully specified and h_n is constant for all n, this stopping rule simplifies to the traditional cusum procedure. On the other hand, one might try to use this stopping rule in a setting where the density f_1 is not fully known. For example, Siegmund & Venkatraman [15] considered (4) in the case of normal data with f_1 representing a change in location of *unknown* magnitude. Their developed test is equivalent to the sequential application of a fixed-sample changepoint detection procedure. This sequential changepoint approach is explored in this section for a location change of unknown size in the case of a von Mises distribution.

As the in-control parameters are assumed known, let $\mu = 0$ without loss of generality, and let $f_0(x)$ denote the corresponding von Mises density as in (1) with known concentration κ. Furthermore, let $f_1(x) = f_0(x - \delta)$ where $\delta \in (-\pi, \pi)$ is unknown. The stopping rule (4) is based on the logarithm of density-ratios, for which the individual components in this setting can be written as

$$\theta_i(\delta) = \log \frac{f_0(X_i - \delta)}{f_0(X_i)} = \kappa \left[\cos(X_i - \delta) - \cos(X_i)\right].$$

These terms have cumulative sum $T_n(\delta) = \sum_{i=1}^{n} \theta_i(\delta)$. When δ is unknown, the stopping rule can be evaluated by maximizing over δ, giving

$$N = \min \left\{ n \,\middle|\, \max_{0 \le k < n} \sup_{\delta} \left[T_n(\delta) - T_k(\delta)\right] \ge h_n \right\}. \tag{5}$$

It is easily shown that the value of δ that maximizes the difference $T_n(\delta) - T_k(\delta)$ is the maximum likelihood estimator of the von Mises location parameter based on the sample $X_{k+1}, \ldots, X_n$. Specifically, letting $C_i = \cos(X_i)$ and $S_i = \sin(X_i)$ for all i, the estimator has closed-form expression

$$\hat{\delta}_{k,n} = \text{atan2}\left(\sum_{i=k+1}^{n} S_i, \sum_{i=k+1}^{n} C_i \right)$$

where atan2 is the two-argument arctangent function. Some algebraic simplification gives

$$\begin{aligned} D_{k,n} &:= \sup_{\delta} \left[T_n(\delta) - T_k(\delta)\right] \\ &= \kappa \left[\left(\cos(\hat{\delta}_{k,n}) - 1\right) \sum_{i=k+1}^{n} C_i + \sin(\hat{\delta}_{k,n}) \sum_{i=k+1}^{n} S_i \right]. \end{aligned} \tag{6}$$

Evaluating the stopping rule can be done by tracking the statistic

$$D_{\max,n} = \max_{0 \le k < n} D_{k,n}.$$

While $D_{\max,n}$ is easily evaluated, it is associated with an increasing computational burden as the sample size grows. However, it does have an advantage over the cusum procedure in that no minimum location shift δ needs to be specified a priori for the process to be considered out of control.

The sequential changepoint (SCP) formulation described here is inherently two-sided. It eliminates the need to run two simultaneous cusums monitoring for upward and downward shifts. However, if the out-of-control state is inherently one-sided (for example, only an upward shift would case the process to be out of control), the SCP can be adjusted accordingly. Assume that an upward shift needs to be detected. The associated stopping rule looks a lot like (5), but with the supremum evaluated over the region $\delta > 0$. Here, the difference $T_n(\delta) - T_k(\delta)$ has range-restricted maximizer $\delta^{+}_{k,n} = \max\{0, \delta_{k,n}\}$. No further evaluation of this one-sided SCP is considered here, and any further reference to SCP is assumed to be the unrestricted (two-sided) version thereof.

The SCP procedure as formulated allows for the control limit h_n to depend on the sample size. While one could choose to calibrate the procedure to a fixed control limit h, both Margavio et al. [12] and Hawkins et al. [5] advocate for using a sequence of control limits $h_{n,\alpha}$ such that the probability of a false alarm is constant for all n. Specifically, when the sequence of control limits $\{h_{1,\alpha}, h_{2,\alpha,\ldots}\}$ satisfy

$$P\left[D_{\max,n} > h_{n,\alpha} | D_{\max,j} \leq h_{j,\alpha}\ \forall j < n\right] = \alpha, \qquad n \geq 1, \tag{7}$$

the in-control ARL is given by $\text{ARL} = 1/\alpha$. Theoretical evaluation of the sequence of control limits is problematic. Specifically, even if one can evaluate the marginal distribution of $D_{\max,n}$ for a given n, the required control limits depends on sequential values of $D_{\max,n}$ that are highly correlated. On the other hand, Hawkins et al. [5] noted that the sequence of control limits can easily be obtained using Monte Carlo methods. In the present setting, sequences of control limits were obtained using Monte Carlo sampling using 1,000,000 simulated realizations of $\left[D_{\max,n} \geq h_{n,\alpha} | D_{\max,j} < h_{j,\alpha}\ \forall j < n\right]$ for each $n = 1, \ldots, 500$. This was done for von Mises concentration parameters $\kappa \in \{0.5, 1, 2\}$ and exceedance probabilities $\alpha = \{0.01, 0.002, 0.001\}$. The values of α correspond to in-control ARL values of 100, 500 and 1000. The Monte Carlo control limits were then "smoothed" by fitting a rational quadratic function to the resulting sequence. The simulated and smoothed control limit sequences are illustrated for the case $\kappa = 1$ in Fig. 1. A summary of some control limits are given in Table 2.

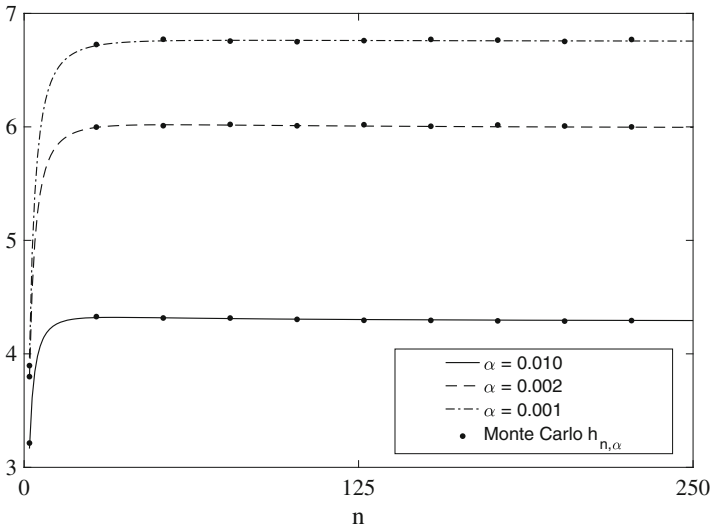


Fig. 1 Control limits for sequential changepoint method for a von Mises distribution with $\kappa = 1$. The markers indicate the Monte Carlo estimates for a subset of n-values, and the lines correspond to the rational quadratic fits

Table 2 Control limits for the von Mises sequential changepoint model for detecting location changes

	$\kappa = 0.5$			$\kappa = 1$			$\kappa = 2$		
n, α	$\alpha =$ 0.01	$\alpha =$ 0.002	$\alpha =$ 0.001	$\alpha =$ 0.01	$\alpha =$ 0.002	$\alpha =$ 0.001	$\alpha =$ 0.01	$\alpha =$ 0.002	$\alpha =$ 0.001
2	1.787	1.988	2.040	3.165	3.796	3.965	3.934	5.359	6.007
5	2.773	3.432	3.667	3.979	5.239	5.735	4.345	6.112	6.836
10	3.281	4.331	4.745	4.228	5.751	6.393	4.539	6.344	7.068
15	3.453	4.702	5.210	4.289	5.899	6.586	4.579	6.415	7.139
20	3.528	4.893	5.458	4.310	5.959	6.666	4.591	6.447	7.172
25	3.564	5.003	5.605	4.318	5.988	6.706	4.594	6.464	7.191
50	3.600	5.180	5.859	4.318	6.018	6.757	4.594	6.488	7.224
100	3.585	5.204	5.919	4.306	6.011	6.762	4.590	6.489	7.233
150	3.572	5.189	5.915	4.300	6.004	6.760	4.588	6.486	7.233
200	3.564	5.175	5.905	4.296	5.999	6.758	4.587	6.484	7.233
300	3.555	5.155	5.888	4.292	5.994	6.755	4.586	6.480	7.231
400	3.549	5.143	5.877	4.290	5.990	6.753	4.585	6.478	7.230
500	3.546	5.135	5.870	4.288	5.988	6.752	4.585	6.476	7.229

The choice of a rational quadratic fit was motivated by noting that (for the cases considered) the sequence of control limits increases rapidly for n small, has a maximum in the range $50 \leq n \leq 100$, and then has a very slights decrease over the remaining range of n considered. Full tables of both the Monte Carlo and smoothed control limit are available from the author. The code to generate control limit sequences is also available for calibrating control sequences for values of κ and α different from those included here.

Note that the computational time required to simulate control sequences grows with the sample size. At time n, it is necessary to simulate sequences of random variables from the conditional distribution $\left[X_1, \ldots, X_n | D_{\max,j} \leq h_{j,\alpha}, j \leq n-1\right]$. This is not particularly difficult, but the probability that a random sample $X_1, \ldots, X_n$ satisfies the required constraints is $(1-\alpha)^{n-1}$, which goes to 0 quite rapidly. Fortunately, the observed asymptotes in the control sequence suggest that it is possible to evaluate control limits $h_{n,\alpha}$ for all n up to some moderate large sample size denoted N. In practice, one can then carry forward this last value, i.e. use $h_{n,\alpha} = h_{K,\alpha}$ for $n > K$. In the present application, the value $K = 500$ was used.

2.3 *Comparison of Methods*

Given the two-sided location cusum in Sect. 2.1 and the sequential changepoint procedure developed in Sect. 2.2, a question of interest is how these compare when a process goes out of control. In general, it is desirable that a monitoring procedure

has a small out-of-control run length. In other words, a good monitoring procedure is one that signals fast once the underlying process has gone out of control. One measure of performance is the out-of-control ARL (OOCARL). Assume that observations $X_1, \ldots, X_{\tau-1}$ occur under in-control conditions, and that observations $X_\tau, X_{\tau+1}, \ldots$ occur under out-of-control conditions. For a process that goes out of control at time τ, the out-of-control average run-length is defined as

$$\text{OOCARL}_\tau = E_\tau[N - \tau | N > \tau] \tag{8}$$

where N denotes the stopping time.

A simulation study was done to compare the cusum and SCP procedures using OOCARL as criterion. The out-of-control time was taken to be $\tau = 0$ and location shifts $\delta \in \{j\pi/8,\ j = 1, \ldots, 8\}$ were considered. Four different cusums were evaluated. Each cusum was calibrated to optimally detect a differently sized deviation from the in-control conditions. These deviation values were $\Delta \in \{\pi/8, \pi/4, \pi/2, \pi\}$. The SCP was also evaluated and did not require an equivalent calibration. Both the cusums and the SCP were evaluated for control limits corresponding to in-control average run length ICARL $\in \{100, 500, 1000\}$. This was done for von Mises distributions with concentration parameters $\kappa \in \{0.5, 1, 2\}$. For each configuration of simulation parameters (δ, κ, ICARL), a total of 100,000 sample paths were simulated. For each simulated path, the run length was evaluated using the four differently sized cusums described above as well as the SCP. These results corresponding to concentrations $\kappa = 0.5$ and $\kappa = 1$ are reported, respectively, in Tables 3 and 4. In these tables, ICARL denotes the in-control average run length, cusum_Δ indicates the cusum has been calibrated to detect a shift of size Δ, SCP is the sequential changepoint procedure, and Max SE denotes the maximum standard error of the estimated ARLs for a specific shift δ. The results corresponding to concentration $\kappa = 2$ are plotted in Fig. 2.

Consider now the results in Tables 3 and 4. When the true location shift is small, here corresponding to values $\delta = \pi/8$ and $\pi/4$, the cusum optimally calibrated to detect that shift has the best performance by far. However, when the cusum is not optimally calibrated, it can perform much worse than the SCP. As the magnitude of the change δ increases, the SCP starts outperforming all the cusums considered, including the optimally calibrated cusum. The cusums also have the curious behavior in that the OOCARL initially decreases as a function of δ, but then starts increasing again. This same behavior was noted by [4] in the one-sided cusum case. On the other hand, the OOCARL of the SCP is decreasing as a function of δ. Generally, when monitoring for a location change and the expected change is small, it is possible to select a cusum that outperforms the SCP procedure. However, this is in part dependent on optimally calibrating the cusum for the true change. On the other hand, the SCP does not require any calibration and has superior performance when the true change is large. In fact, the SCP has out-of-control performance that also holds up favorably when compared the optimal cusum even when the size of the change is small. Inspection of Fig. 2 further illustrates this point. For δ small, the

Table 3 Monte Carlo ARL for $\kappa = 0.5$

ICARL	δ	cusum$_{\pi/8}$	cusum$_{\pi/4}$	cusum$_{\pi/2}$	cusum$_{\pi}$	SCP	max SE
100	$\pi/8$	68	69	74	88	85	0.26
	$\pi/4$	42	39	42	61	45	0.18
	$3\pi/8$	31	27	26	37	27	0.10
	$\pi/2$	27	22	19	23	18	0.06
	$5\pi/8$	27	21	15	16	14	0.04
	$3\pi/2$	30	21	14	12	11	0.04
	$7\pi/8$	36	22	13	10	10	0.05
	π	39	23	13	10	10	0.05
500	$\pi/8$	175	188	257	399	233	1.22
	$\pi/4$	84	76	94	217	91	0.65
	$3\pi/8$	59	48	47	99	48	0.28
	$\pi/2$	51	38	31	48	31	0.13
	$5\pi/8$	51	36	25	28	22	0.06
	$3\pi/4$	59	38	23	20	18	0.07
	$7\pi/8$	79	43	22	17	16	0.10
	π	97	46	22	16	15	0.12
1000	$\pi/8$	175	263	417	763	324	1.26
	$\pi/4$	84	94	124	368	112	0.32
	$3\pi/8$	59	58	57	144	58	0.12
	$\pi/2$	51	46	37	63	36	0.07
	$5\pi/8$	51	43	30	35	26	0.07
	$3\pi/4$	59	45	27	24	21	0.09
	$7\pi/8$	79	53	27	20	19	0.13
	π	97	58	27	19	18	0.15

SCP performance closely tracks the best-performing cusum. Then, as δ increases, the SCP becomes the best-performing monitoring procedure.

3 Monitoring for Concentration and Location-Concentration Changes

The focus of Chap. 2 is the detection of a change in location in the underlying process. Of course, the out-of-control behavior can take many forms different from a simple change in location. In this chapter, two other scenarios are considered. Firstly, procedures to monitor for a change in concentration are discussed. Thereafter, the more interesting problem of monitoring for a change in both location and concentration is considered. The methods discussed are developed specifically for an underlying von Mises distribution, but robustness against departure from this parametric model are considered at the end of the chapter.

Table 4 Monte Carlo ARL for $\kappa = 1$

ICARL	δ	$\text{cusum}_{\pi/8}$	$\text{cusum}_{\pi/4}$	$\text{cusum}_{\pi/2}$	cusum_{π}	SCP	max SE
100	$\pi/8$	41	43	56	82	51	0.25
	$\pi/4$	20	19	23	47	21	0.14
	$3\pi/8$	14	12	12	23	11	0.07
	$\pi/2$	12	10	8	12	7	0.03
	$5\pi/8$	12	9	6	7	5	0.02
	$3\pi/4$	14	9	6	5	4	0.02
	$7\pi/8$	19	10	6	4	4	0.02
	π	23	11	6	4	4	0.03
500	$\pi/8$	79	92	169	367	106	1.15
	$\pi/4$	34	30	42	158	34	0.49
	$3\pi/8$	24	18	18	54	18	0.16
	$\pi/2$	20	14	11	21	11	0.05
	$5\pi/8$	20	13	9	11	8	0.02
	$3\pi/4$	24	14	8	8	6	0.02
	$7\pi/8$	33	17	8	6	6	0.04
	π	46	19	8	6	5	0.05
1000	$\pi/8$	79	119	260	698	133	0.79
	$\pi/4$	34	35	52	265	40	0.13
	$3\pi/8$	24	21	21	76	20	0.04
	$\pi/2$	20	16	13	26	13	0.02
	$5\pi/8$	20	15	10	13	9	0.02
	$3\pi/4$	24	16	9	9	7	0.03
	$7\pi/8$	33	20	9	7	6	0.05
	π	46	23	10	7	6	0.06

3.1 Detecting a Change in Concentration

Assume again that the data $X_1, X_2, \ldots$ are from a von Mises distribution which (without loss of generality) has mean direction 0. Assume the in-control concentration is given by κ_0. The one-sided cusum for detecting a change from in-control concentration κ_0 to a specified out-of-control concentration κ_1 as developed by Hawkins & Lombard [4] is given by $C_0 = 0$ and

$$\begin{aligned} C_n &= \max\left[0, C_{n-1} + \log\frac{f(X_n|0,\kappa_1)}{f(X_n|0,\kappa_0)}\right] \\ &= \max\left[0, C_{n-1} - \log\frac{I_0(\kappa_1)}{I_0(\kappa_0)} + (\kappa_1 - \kappa_0)\cos(X_n)\right], \quad n \geq 1. \end{aligned}$$

These authors also tabulate control limits for the one-sided concentration cusum for select values of κ_0 and κ_1.

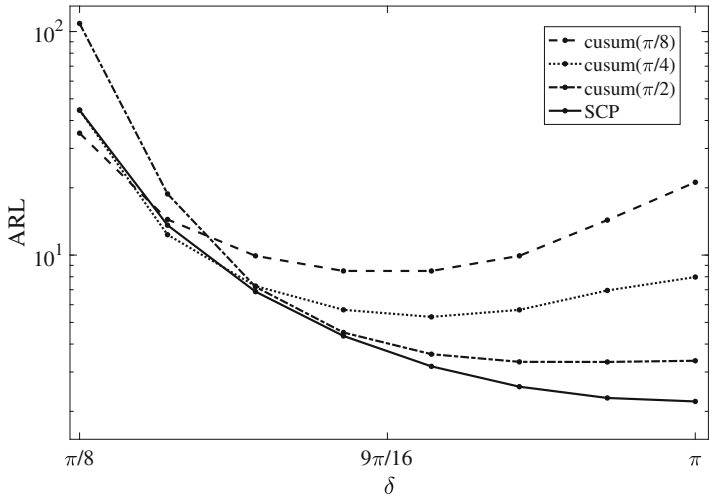


Fig. 2 Monte Carlo ARL for a von Mises distribution with $\kappa = 2$ and location shift δ. Cusums calibrated to three different shifts are shown, along with the SCP method

Now, if it is desired to detect either an increase or a decrease in concentration, say from κ_0 to $\kappa_0 e^{\pm\beta}$ with $\beta > 0$, two one-sided cusums can be used to monitor the process, say

$$C_n^+ = \max\left[0, C_{n-1}^+ - \log\frac{I_0(\kappa_0 e^{\beta})}{I_0(\kappa_0)} + \kappa_0(e^{\beta} - 1)\cos(X_n)\right],$$

$$C_n^- = \max\left[0, C_{n-1}^- - \log\frac{I_0(\kappa_0 e^{-\beta})}{I_0(\kappa_0)} - \kappa_0(1 - e^{-\beta})\cos(X_n)\right].$$

These two cusums are run simultaneously until one of them exceeds the control limit. Tabulated control limits are not included here, as these depend on the in-control concentration κ_0, as well as the values of β and the in-control run length. These are easily calculated for a specific set of parameters. Software for the calculation of control limits is available at the author's github page (link provided at the end of this chapter).

It is again possible to develop a sequential changepoint procedure for detecting a change in concentration. Let $f(x|\kappa)$ denote a von Mises density function with mean direction 0 and concentration parameter κ. Based on stopping rule (4), define

$$\begin{aligned}\theta_i(\kappa_0, \kappa_1) &= \log\left[f(X_i|\kappa_1)/f(X_i|\kappa_0)\right] \\ &= \log I_0(\kappa_1) - \log I_0(\kappa_0) + (\kappa_1 - \kappa_0)\cos(X_i)\end{aligned}$$

with associated cumulative sum $T_n(\kappa_0, \kappa_1) = \sum_{i=1}^{n} \theta_i(\kappa_0, \kappa_1)$. When the out-of-control concentration κ_1 is unknown, an appropriate stopping rule is given by

$$N = \min\left\{ n \left| \max_{0 \le k < n} \sup_{\kappa_1} [T_n(\kappa_0, \kappa_1) - T_k(\kappa_0, \kappa_1)] \ge h_n \right. \right\}. \tag{9}$$

The value of κ_1 that maximizes the difference $T_n(\kappa_0, \kappa_1) - T_k(\kappa_0, \kappa_1)$ is the maximum likelihood estimator of the von Mises concentration parameter (with known location) based on the sample $X_{k+1}, \ldots, X_n$. Specifically, it is the solution of the equation

$$\frac{I_1(\hat{\kappa}_{k,n})}{I_0(\hat{\kappa}_{k,n})} = \frac{1}{n-k} \sum_{i=k+1}^{n} \cos(X_i)$$

where I_j denote the modified Bessel function of the first kind of order j. Note that the value of this maximizer does not depend on the in-control concentration κ_0. However, the distribution of N in (9) does depend on the in-control value. Now, given the solution $\hat{\kappa}_{k,n}$, define

$$\begin{aligned} K_{k,n} &:= \sup_{\kappa_1} [T_n(\kappa_0, \kappa_1) - T_k(\kappa_0, \kappa_1)] \\ &= (n-k) \log\left[I_0(\hat{\kappa}_{k,n}) / I_0(\kappa_0) \right] + (\hat{\kappa}_{k,n} - \kappa_0) \sum_{i=k+1}^{n} \cos(X_i), \end{aligned}$$

and let $K_{\max,n} = \max_{0 \le k < n} K_{k,n}$. The stopping rule can now be written as

$$N = \min\{n | K_{\max,n} \ge h_n\}.$$

The sequence of control limits can be evaluated using (7) as described in Sect. 2.2 to give a constant hazard rate α which has associated ICARL$= 1/\alpha$.

3.2 Detecting a Change in Location and Concentration

If a process would be considered out of control if either a change in location or a change in concentration takes place, one could monitor separately for these two types of changes. For example, it is fairly standard in practice to run simultaneously two cusums, one to detect a change in location and the other to detect a change in concentration. Of course, one would not expect this approach to work well if there were a simultaneous change in location and concentration of small magnitude. In this instance, a procedure that simultaneously monitors for a location and concentration change is preferred. A cusum can be constructed for the latter situation. Let the in-control location and concentration parameters be 0 and κ_0, and assume that the out of control parameters to be detected are location δ and concentration κ_1. The associated cusum is of the form $C_0 = 0$ and

$$C_n = \max\left[0, C_{n-1} + \log\frac{f(X_n|\delta,\kappa_1)}{f(X_n|0,\kappa_0)}\right]$$
$$= \max\left[0, C_{n-1} - \log\frac{I_0(\kappa_1)}{I_0(\kappa_0)} + \kappa_1 cos(X_n - \delta) - \kappa_0\cos(X_n)\right], \quad n \geq 1.$$

Note that if the direction of changes in location and concentration are not specified, the cusum procedure will require monitoring *four* one-directional cusums. This seems rather unwieldy. The sequential changepoint procedure is a very attractive alternative here, as it only requires running one monitoring scheme. Specifically, the log-ratio of density functions is given by

$$\theta_i(\delta,\kappa_0,\kappa_1) = \log I_0(\kappa_1) - \log I_0(\kappa_0) + \kappa_1\cos(X_i - \delta) - \kappa_0\cos(X_i)$$

with cumulative sum $T_n(\delta,\kappa_0,\kappa_1) = \sum_{i=1}^{n}\theta_i(\delta,\kappa_0,\kappa_1)$. With both δ and κ_1 unknown, an appropriate stopping rule is

$$N = \min\left\{n \left| \max_{0\leq k<n-1} \sup_{\delta,\kappa_1}[T_n(\delta,\kappa_0,\kappa_1) - T_k(\delta,\kappa_0,\kappa_1)] \geq h_n\right.\right\}. \tag{10}$$

Note that stopping rule (10) has a maximum evaluated over indices $0 \leq k < n-1$, whereas the stopping rules associated with only a change in location or only a change in concentration was evaluated over indices $0 \leq k < n$. The reason for this difference in definition is that at least two observations are required to estimate both the location and concentration parameters from data. Now, define

$$L_{k,n} := \sup_{\delta,\kappa_1}[T_n(\delta,\kappa_0,\kappa_1) - T_k(\delta,\kappa_0,\kappa_1)]$$
$$= (n-k)\log\left[I_0(\hat{\kappa}_{k,n})/I_0(\kappa_0)\right]$$
$$+\hat{\kappa}_{k,n}\sum_{i=k+1}^{n}\cos(X_i - \hat{\delta}_{k,n}) - \kappa_0\sum_{i=k+1}^{n}\cos(X_i). \tag{11}$$

where $(\hat{\delta}_{k,n}, \hat{\kappa}_{k,n})$ denote the maximum likelihood estimators of the von Mises location and concentration parameters based on sample $X_{k+1},\ldots,X_n$. The stopping rule can now be expressed as

$$N = \min\{n|L_{\max,n} \geq h_n\}$$

with $L_{\max,n} = \max_{0\leq k<n-1} L_{k,n}$.

The sequence of control limits using (7) for hazard rates $\alpha \in \{0.01, 0.002, 0.001\}$ with associated ICARL= $\{100, 500, 1000\}$ are given in Table 5 for a selection of sample sizes n and for in-control concentration parameter $\kappa \in \{0.5, 1, 2\}$. As in

Table 5 Control limits for the von Mises sequential changepoint model for simultaneous detection of location and concentration changes assuming in-control parameters $\mu = 0$ and κ

	$\kappa = 0.5$			$\kappa = 1$			$\kappa = 2$		
n	$\alpha =$ 0.01	$\alpha =$ 0.002	$\alpha =$ 0.001	$\alpha =$ 0.01	$\alpha =$ 0.002	$\alpha =$ 0.001	$\alpha =$ 0.01	$\alpha =$ 0.002	$\alpha =$ 0.001
2	7.752	8.795	9.053	7.435	8.887	9.464	6.887	8.453	9.114
5	8.217	9.690	10.616	8.071	9.920	10.471	7.804	9.579	10.374
10	8.288	9.904	10.792	8.168	10.021	10.590	7.928	9.769	10.576
15	8.306	9.954	10.829	8.195	10.046	10.621	7.960	9.814	10.628
20	8.314	9.974	10.845	8.207	10.057	10.636	7.974	9.833	10.651
25	8.319	9.986	10.853	8.215	10.064	10.644	7.983	9.844	10.664
50	8.327	10.005	10.869	8.229	10.076	10.659	7.998	9.865	10.690
100	8.331	10.013	10.875	8.236	10.082	10.664	8.005	9.874	10.702
150	8.333	10.015	10.878	8.238	10.084	10.664	8.008	9.877	10.706
200	8.333	10.016	10.879	8.239	10.085	10.663	8.009	9.879	10.708
250	8.334	10.017	10.879	8.240	10.085	10.662	8.010	9.879	10.709

Table 6 OOCARL for SCP with in-control concentration $\kappa_0 = 0.5$, out-of-control location shift δ and concentration change $\kappa_1 = \gamma\kappa_0$

		γ			
ICARL	δ	1.25	1.50	1.75	2.00
100	$\pi/8$	87	78	66	54
	$\pi/4$	58	50	43	37
	$3\pi/8$	38	33	28	25
	$\pi/2$	27	23	20	18
500	$\pi/8$	273	189	131	95
	$\pi/4$	125	97	75	60
	$3\pi/8$	70	56	46	38
	$\pi/2$	46	38	32	27
1000	$\pi/8$	387	250	165	116
	$\pi/4$	154	117	90	70
	$3\pi/8$	82	66	54	44
	$\pi/2$	53	44	36	31

Sect. 2.2, the Monte Carlo control limits obtained using 1,000,000 samples for each n were "smoothed" by fitting a rational quadratic function to the sequence.

The out-of-control behavior of the SCP procedure is illustrated in Tables 6 and 7. Here, the run length was evaluated for 100,000 sequences simulated with a selection of out-of-control parameters (δ, κ_1) where $\kappa_1 = \gamma\kappa_0$. The simulated OOCARLs are reported for in-control run lengths ICARL$\in \{100, 500, 1000\}$.

When comparing the results in Tables 6 and 7, it is clear that it takes longer to detect a change for smaller concentration (κ) values. Intuitively, it is more difficult to detect a change in location when the distribution has large spread. Even so, in both instances the SCP signals fairly quickly, even for the minimal change in location and scale. Consider the setting with an ICARL of 500. Even when the location change is $\pi/8$ and the concentration is increased by 25%, the OOCARL is 273 for in-control

Table 7 OOCARL for SCP with in-control concentration $\kappa_0 = 1$, out-of-control location shift δ and concentration change $\kappa_1 = \gamma\kappa_0$

		γ			
ICARL	δ	1.25	1.50	1.75	2.00
100	$\pi/8$	58	46	36	29
	$\pi/4$	26	22	19	16
	$3\pi/8$	15	12	11	10
	$\pi/2$	9	8	7	7
500	$\pi/8$	122	83	59	45
	$\pi/4$	43	35	29	24
	$3\pi/8$	22	19	16	14
	$\pi/2$	14	12	11	9
1000	$\pi/8$	144	95	66	50
	$\pi/4$	48	39	32	26
	$3\pi/8$	24	21	18	16
	$\pi/2$	15	13	12	10

concentration $\kappa = 0.5$, and 122 for $\kappa = 1$. For an ICARL of 1000, the SCP signals even faster (relative to the ICARL). Overall, the location-concentration SCP appears to have good performance.

3.3 Power and Robustness Considerations

In Sects. 2.2 and 3.2, SCP procedures for, respectively, only a change in location and a change in location and concentration are considered. Denote by SCP_L and SCP_{LC} the location-only and location-concentration sequential changepoint procedures. Intuitively, if there is only a location change, SCP_{LC} can still be used to monitor the process. However, it may not work as well as SCP_L. Here, the question is how much detection power is sacrificed when only a location change occurs, but SCP_{LC} is used. The reduction in detection power manifests in an increase in OOCARL compared to SCP_L. A Monte Carlo study was performed to compare the average run length of the two SCP procedures. A single run of the simulation study consisted of generating sequential observations from a von Mises distribution with location $\delta \in [\pi/32, \pi/2]$ uniformly generated (and fixed for the run) and $\kappa = 1$. Both (5) and (10) calibrated to an in-control run length of 500 were used to monitor the process until both signaled that a change had occurred. The run lengths for both SCP procedures were then recorded. This was done for 100,000 simulated process paths. In Fig. 3, the estimated OOCARL curves are shown as a function of the location change δ. $\text{SCP}(\delta)$ denotes the procedure monitoring for a location change only, while $\text{SCP}(\delta, \kappa)$ denotes the procedure monitoring for a location and concentration change. The concentration was fixed at $\kappa = 1$ and the in-control ARL was 500.

Visual inspection of Fig. 3 reveals that SCP_L on average detects the change in location faster than SCP_{LC}. This is as anticipated. When considering the difference

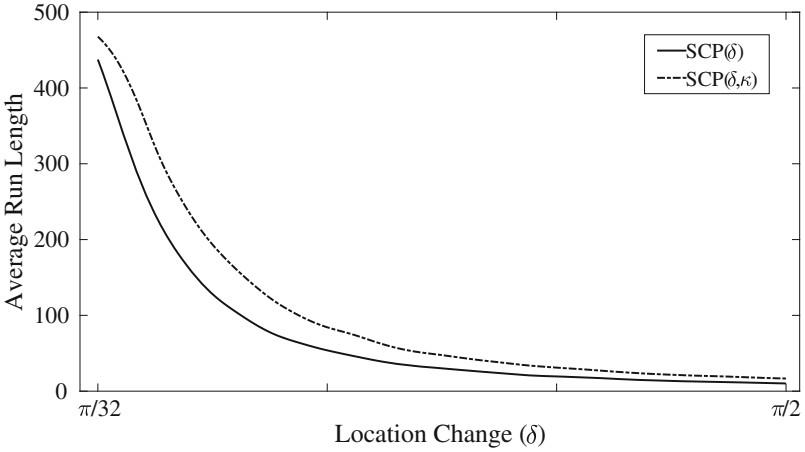


Fig. 3 ARLs as a function of δ for the two SCP procedures. ICARL = 500 and $\kappa = 1$

between the two OOCARL curves, it becomes evident that detection power is dramatically impacted by also monitoring for a concentration change which does not occur. For example, when a location change of $\pi/16$ occurs, the SCP_L has average run length 267, while SCP_{LC} has average run length 361, an increase of around 35%. Similarly, for a location change of $\pi/4$, the respective average run lengths are 33 and 51, an increase of nearly 54%. Across the range of δ values considered, SCP_{LC} generally has run length more than 50% larger than SCP_L. This conclusion carries over to scenarios with other concentration parameter values, and for SCP procedures calibrated to different ICARLs.

Another question of interest is whether the SCP procedures designed for von Mises data are robust against departures from this assumption. While it is possible to set up and calibrate monitoring procedures for other circular distributions, no distribution is likely to perfectly describe the true data generating mechanism in practice. For this reason, the behavior of the location and location-concentration SCPs were investigated with data generated from two other symmetric distributions on a circle, namely a cardioid distribution and a wrapped Cauchy distribution. The cardioid distribution has density

$$f(x) = \frac{1 + 2\rho\cos(x - \mu)}{2\pi}, \ -\pi < x < \pi$$

with $|\rho| < 0.5$. Here, μ denotes the mean direction and the concentration is given by $\kappa = A^{-1}(p)$ where A^{-1} is the inverse of the function $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. The wrapped Cauchy distribution has density

$$f(x) = \frac{1}{2\pi}\frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos(x - \mu)}, \ -\pi < x < \pi$$

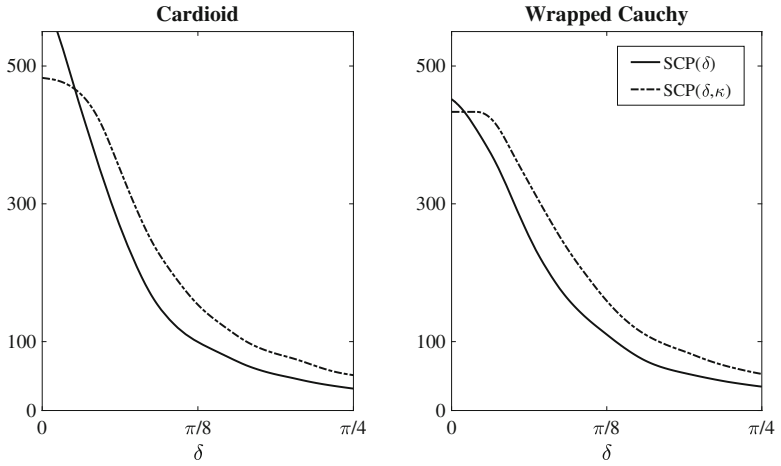


Fig. 4 ARLs of the two SCP procedures as a function of δ for cardioid distribution (left) and wrapped Cauchy distribution (right) both with $\kappa = 1$ and using a von Mises distribution calibration

with mean direction μ and concentration $\kappa = A^{-1}(\rho)$. When the von Mises, cardioid, and wrapped Cauchy distributions have common mean direction and concentration, the cardioid distribution has lighter tails and the wrapped Cauchy has heavier tails than the von Mises distribution.

Illustrated in Fig. 4 is the performance of the SCP_L and SCP_{LC} designed for an underlying von Mises distribution, but when the data are in fact from either a cardioid distribution (left figure) or a wrapped Cauchy distribution (right figure). The ARLs in these figures were evaluated using Monte Carlo methods. For the figure on the left, observations were generated from a cardioid distribution with location parameter δ fixed at 0 with probability 0.15 or uniformly generated on the interval $[0, \pi/4]$ with probability 0.85. This method of choosing δ for a given run of the simulation was used to ensure accurate estimation of ARL values close to the boundary of the δ-space. The concentration parameter was fixed at $\kappa = 1$. The SCP_L and SCP_{LC} procedures calibrated to an ICARL of 500 (for a von Mises distribution) were used to monitor the process until both signaled a change had occurred. The run lengths were then recorded. This was done for 100,000 simulated process paths. A cubic spline model was then fit to the (δ, ARL) pairs to obtain the figure. An analogous process was used for the figure on the right by sampling from a wrapped Cauchy distribution.

The performance of the SCP procedures for underlying cardioid and wrapped Cauchy distributions can be described as "fair" under in-control conditions. For the cardioid distribution, the average ICARL for SCP_L and SCP_{LC} are approximately 612 and 482, respectively. For the wrapped Cauchy distribution these same values are around 446 and 433. However, even though their ICARLs are not all that close to

the nominal value, the procedures do perform well under out-of-control condition, signaling when a location change occurs fairly quickly.

The conclusions above also held for other symmetric circular distributions considered. However, when asymmetric distributions are considered, for example the wrapped skew-normal distribution of Pewsey [13], neither the location nor location-concentration SCP procedures performed well under in-control or out-of-control conditions. Generally, it would seem that the methods developed for a von Mises distribution may perform well enough to satisfy many practitioners when the underlying distribution is symmetric but not von Mises. On the other hand, these SCP procedures do not perform well when the underlying distribution deviates from symmetry.

4 Monitoring Uniformity

In Sects. 2 and 3, the general problem of interest considered is detecting a change in location and/or concentration when the underlying distribution is von Mises. In this section, the problem of monitoring for deviation from a circular uniform distribution is considered. Even though the uniform distribution can be considered a special case of the von Mises distribution with $\kappa = 0$, the problem of monitoring for deviations from uniformity is worth exploring in its right. The circular uniform distribution has density $f(x) = (2\pi)^{-1},\ -\pi < x < \pi$, and is interesting in that the mean direction does not exist. There are many examples of processes that are uniformly distributed, see Lombard & Maxwell [7] for examples. These authors also developed a cusum procedure for detecting deviations from the uniform distribution. An overview of their methodology, which is nonparametric in nature, will be provided. A sequential changepoint method against a parametric alternative will also be developed. These two procedures will finally be compared in a simulation study.

4.1 A Nonparametric Cusum for Uniformity

Consider *iid* random variables $X_1, X_2, \ldots$ having uniform distribution on the circle and define

$$c_j = \sqrt{2}\cos X_j \quad \text{and} \quad s_j = \sqrt{2}\sin X_j,$$

for $j = 1, 2, \ldots$. Under uniformity, these random variables have $\mathrm{E}(c_j) = \mathrm{E}(s_j) = 0$, $\mathrm{Var}(c_j) = \mathrm{Var}(s_j) = 1$, and are uncorrelated with one another. The Raleigh statistic is commonly used to test uniformity of a sample of circular random variables $X_1, \ldots, X_n$. Specifically, defining

$$C_k = \sum_{j=1}^{k} c_j \quad \text{and} \quad S_k = \sum_{j=1}^{k} S_j,$$

the Raleigh statistic is proportional to

$$R_k^2 = C_k^2 + S_k^2.$$

Towards defining the cusum proposed by Lombard & Maxwell [7], set

$$\xi_k = \frac{R_k^2 - R_{k-1}^2 - 2}{2R_{k-1}}, \quad k \geq 2, \tag{12}$$

with ξ_k equal to 0 if $R_{k-1} = 0$. Some algebra shows that (12) can equivalently be expressed as

$$\xi_k = \frac{C_{k-1}c_k + S_{k-1}s_k}{C_{k-1}^2 + S_{k-1}^2}. \tag{13}$$

Both Eqs. (12) and (13) show how ξ_k can be updated sequentially. Under the in-control hypothesis of uniformity,

$$\mathrm{E}[\xi_k | X_1, \ldots, X_{k-1}] = 0 \quad \text{and} \quad \mathrm{Var}[\xi_k | X_1, \ldots, X_{k-1}] = 1.$$

One can also show that the ξ_k are uncorrelated with one another. These facts, together with noting that the $|\xi_k| \leq \sqrt{2}$ for all k, were used by Lombard & Maxwell [7] to prove that for n_0 large, the joint distribution of the sums $\sum_{k=1}^{n} \xi_k$, $n \geq n_0$ will approximate those of the sums $\sum_{k=1}^{n} Z_k$, $n \geq n_0$ for Z_k *iid* standard normal random variables. The cusum procedure then follows by setting $V_0 = 0$ and

$$V_k = \max(0, V_{k-1} + \xi_k - \zeta)$$

with stopping rule

$$N = \min\{n | V_n \geq h\}.$$

Here, ζ denotes the reference value of the cusum, and can be thought of as a minimum detection limit. If the out-of-control distribution of X_i is known to have mean μ, a good choice for the reference value is $\zeta = \mu/2$. However, this is not always known and one can use $\zeta = 0$ for the cusum.

Assume now that random variable X is not uniform. Defining

$$\alpha = \mathrm{E}[\cos X] \quad \text{and} \quad \beta = \mathrm{E}[\sin X]$$

Table 8 Control limits h for the uniformity cusum corresponding to an reference value ζ

	ICARL		
ζ	100	500	1000
0	8.879	21.201	30.476
0.25	4.418	7.267	8.585

and $\theta = \sqrt{2\left(\alpha^2 + \beta^2\right)}$. By the law of large numbers, $\xi_k \to \theta > 0$ as $k \to \infty$ when the process goes out of control. When this happens, one will expect the cusum V_k to start increasing and eventually signal as out of control.

Reported in Table 8 are the control limits h as determined by Lombard & Maxwell [7]. These values were obtained using results for cusums following a normal distribution and hold for small values of ζ. If a cusum with a larger reference value is desired, Monte Carlo methods can be employed to calibrate the cusum.

4.2 Monitoring Against a von Mises Alternative

The cusum considered in the previous section is nonparametric, in that no parametric distribution is specified for the out-of-control behavior of the process. It is possible to construct a cusum for deviation from uniformity against a parametric alternative. The increments for this cusum can be based on the log-ratio $\log[f_1(x)/f_0(x)]$ where f_0 is the uniform density and f_1 is the parametric alternative. When f_1 is the von Mises distribution, the log-ratio becomes

$$\theta_i(\mu, \kappa) = \kappa \cos(X_i - \mu) - \log I_0(\kappa) \tag{14}$$

with cusum given by $C_0 = 0$ and $C_n = \max[0, C_{n-1} + \theta_n(\mu, \kappa)]$. While this cusum is easy to calculate, it has the undesirable property that the out-of-control mean μ needs to be specified. Even so, the control limits only depend on the out-of-control concentration parameter κ. If the value for μ is mis-specified, the ability of the cusum to signal fast once the process is out of control is dramatically impacted. For this reason, the parametric cusum for deviation from uniformity is not further explored here.

On the other hand, it is possible to use (14) to specify a sequential changepoint procedure. The control limits of this SCP depend only on the null distribution, and does not require the specification of out-of-control parameters. Defining $T_n(\mu, \kappa) = \sum_{i=1}^{n} \theta_i(\mu, \kappa)$, the stopping rule for the SCP procedure is given by

$$N = \min\left\{ n \,\middle|\, \max_{0 \le k < n-1} \sup_{\mu,\kappa} [T_n(\mu, \kappa) - T_k(\mu, \kappa)] \ge h_n \right\}. \tag{15}$$

The difference $T_n(\mu, \kappa) - T_k(\mu, \kappa)$ is maximized by $\hat{\mu}_{k,n}$ and $\hat{\kappa}_{k,n}$ the maximum likelihood estimators of the von Mises mean and concentration parameters based

on the sample $X_{k+1}, \ldots, X_n$. The control limits h_n were calculated using (7). A selection of these control limits smoothed in a manner similar to that described in Sect. 2.2 are given in Table 9 for hazard rates $\alpha \in \{0.01, 0.002, 0.001\}$. The largest value in the table corresponds to $n = 150$. For an SCP running longer than this, the last value can be carried forward.

A small simulation study was done to compare the above SCP procedure against the nonparametric cusum defined in Sect. 4.1. Data were simulated from a von Mises distribution with location $\mu = 0$ and concentration parameter $\kappa \in \{0.125, 0.25, 0.5, 1\}$. For each simulated dataset, the cusums with reference values $\zeta = 0$ and 0.25 as well as the SCP procedure were using to find the run length until the process signals. For each value of κ and each ICARL$\in \{100, 500, 1000\}$, a total of 100,000 datasets were simulated. The average run-lengths across these simulations are reported in Table 10.

When inspecting Table 10, it is surprising to note that the nonparametric cusum generally signals much faster than the SCP procedure. The only instances in which the SCP is seen to outperform a cusum is for the case $\kappa = 1$, a rather strong deviation from uniformity. Even then, the SCP only performs better than the cusum with reference value $\zeta = 0$. When comparing the two cusums, their performance is very comparable for ICARL$= 100$. For larger ICARL, the cusum with reference

Table 9 Control limits $h_{\alpha,n}$ for the SCP procedure to detect deviation from uniformity

n	$\alpha = 0.01$	$\alpha = 0.002$	$\alpha = 0.001$
2	8.084	8.228	8.233
5	8.192	9.711	10.804
10	8.203	9.931	10.980
15	8.206	9.975	11.014
20	8.207	9.993	11.028
25	8.208	10.002	11.036
50	8.209	10.016	11.050
100	8.210	10.022	11.056
150	8.210	10.024	11.057

Table 10 OOCARL for cusums SCP procedure monitoring for deviation from angular uniformity

		κ			
ICARL	Method	0.125	0.25	0.50	1.00
100	$\text{cusum}_{\zeta=0}$	82	58	33	18
	$\text{cusum}_{\zeta=0.25}$	81	58	31	15
	SCP	97	83	51	23
500	$\text{cusum}_{\zeta=0}$	265	144	72	38
	$\text{cusum}_{\zeta=0.25}$	295	151	59	23
	SCP	425	263	103	35
1000	$\text{cusum}_{\zeta=0}$	397	203	101	54
	$\text{cusum}_{\zeta=0.25}$	482	215	73	27
	SCP	759	385	128	41

value $\zeta = 0$ is seen to perform better for smaller values of κ, while the cusum with reference value $\zeta = 0.25$ is seen to perform better for larger values of κ.

5 Concluding Remarks

Sequential changepoint (SCP) procedures are develop for the monitoring of circular process data with the data assumed to follow a von Mises distribution. In the context of monitoring for a location change, the SCP method is seen to compare very favorably to a cusum procedure developed by Hawkins & Lombard [4]. To implement the cusum, it is necessary to specify a location shift at which point the process would be considered out of control. No such specification is needed to implement the SCP method.

The SCP methodology is also extended to the simultaneous monitoring for a location and concentration change. If only a location change occurs, it is seen that the use of a joint location-concentration procedure can result in a substantial loss of power. Finally, an SCP procedure to monitor for departures from uniformity is developed. Here, a nonparametric cusum developed by Lombard & Maxwell [7] is demonstrated to be superior to the SCP procedure.

References

1. Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.
2. Gadsen, R. & Kanji, G. (1983). Sequential analysis applied to circular data. *Communications in Statistics. Part C: Sequential Analysis, 1*, 305–314.
3. Ghosh, K., Jammalamadaka, S. R. & Vasudaven, M. (1999). Change-point problems for the von Mises distribution. *Journal of Applied Statistics, 26*, 423–434.
4. Hawkins, D. M. & Lombard, F. (2017). Cusum control for data following the von Mises distribution. *Journal of Applied Statistics, 44*, 1319–1332.
5. Hawkins, D. M., Qiu, P. & Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of Quality Technology, 35*, 355–366.
6. Lombard, F. (1986). The change–point problem for angular data: A nonparametric approach. *Technometrics, 28*, 391–397.
7. Lombard, F. & Maxwell, R. (2012). A cusum procedure to detect deviations from uniformity in angular data. *Journal of Applied Statistics, 39*, 1871–1880.
8. Lombard, F., Hawkins, D. M. & Potgieter, C. (2018). Nonparametric cusum charts for angular data with applications in health science and astrophysics. *REVstat* .
9. Lombard, F., Hawkins, D. M. & Potgieter, C. J. (2017). Sequential rank cusum charts for angular data. *Computational Statistics & Data Analysis, 105*, 268–279.
10. Lorimer, D. R. & Kramer, M. (2005). *Handbook of pulsar astronomy* (vol. 4). Cambridge University Press, Cambridge.
11. Mardia, K. V. & Jupp, P. E. (2009). *Directional statistics* (vol. 494). Wiley, London.
12. Margavio, T. M., Conerly, M. D., Woodall, W. H. & Drake, L. G. (1995). Alarm rates for quality control charts. *Statistics & Probability Letters, 24*, 219–224.

13. Pewsey, A. (2000). The wrapped skew-normal distribution on the circle. *Communications in Statistics-Theory and Methods, 29*, 2459–2472.
14. Sengupta, A. & Laha, A. K. (2008). A Bayesian analysis of the change-point problem for directional data. *Journal of Applied Statistics, 35*, 693–700.
15. Siegmund, D. & Venkatraman, E. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics, 23*, 255–271.
16. van Dobben de Bruyn, C. S. (1968). *Cumulative sum tests: theory and practice*. Griffin, London.

Part II
Acceptance Sampling Plans

Time Truncated Life Tests Using the Generalized Multiple Dependent State Sampling Plans for Various Life Distributions

Muhammad Aslam, Gadde Srinivasa Rao, and Mohammed Albassam

Abstract This chapter presents the designing of the generalized multiple dependent state sampling (GMDSS) plans for the various statistical distributions. We will present the design of GMDSS sampling when the failure time follows the gamma distribution, Burr type XII distribution and the Birnbaum-Saunders (BS) distribution. The necessary measures including the operating characteristics (OC) function are derived. The plan parameters of the proposed test plans are determined through the non-linear optimization solution. The proposed sampling plan is studied for a minimal sample size subject to specified requirements of the consumer and producer' risks. The efficiency of the proposed plans in terms of sample sizes are discussed over the existing sampling plans using the same level of all parameters. The advantages of the proposed plans are discussed through simulated data and real data from the industry.

Keywords Acceptable quality level · Limiting quality level · Multiple deferred state sampling plan · Producer's risk · Consumer's risk · Mean lifetime ratio

1 Introduction

The high quality is the main target of all companies in the world. There is a race among industries in producing a high-quality product. To manufacture the high quality of the product, the producers are very careful in purchasing the raw material and testing or the inspection of the product at each stage of the manufacturing of the product. Therefore, the inspection of the raw material to the finished product is an important activity to maintain the high quality of the product. For inspecting the

M. Aslam (✉) · M. Albassam
Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia
e-mail: magmuhammad@kau.edu.sa; malbassam@kau.edu.sa

G. S. Rao
Department of Statistics, The University of Dodoma, Dodoma, Tanzania
e-mail: rao.gadde@udom.ac.tz

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_7

product, it is not possible to check each item at every stage of the production process. Therefore, inspection in the industry is done through a well-defined acceptance-sampling plan. A sampling plan plays a role like a guide to select a suitable random sample and acceptance number at specified parameters. As, for the lot sentencing purpose, a random sample is selected from a lot of the product, there is a chance that a good lot is rejected (producer's risk) and a bad lot is accepted (consumer's risk). The sampling schemes are designed which minimize both risks during the inspection of a lot of the product.

Mostly, the sampling plans are designed using the single sampling scheme, which is simple to inspect a lot of the product. However, the use of single sampling needs a larger sample size for the inspection of the product. Several authors designed the sampling plans using the single sampling in various circumstances, for example, Rosaiah and Kantam [1] proposed a sampling plan for the Rayleigh distribution under the time truncated life test. They determined the minimum sample size and discussed the OC curve for various specified parameters. Chen et al. [2] discussed the single and double sampling using the Bayesian approach under the type-II censoring. They showed the efficacy of the double sampling plan over the single sampling plan in the cost required for the testing of the product. Vijayaraghavan et al. [3] designed a single acceptance sampling plan under the conditions of gamma prior and the Poisson distribution. They proved the efficiency gamma-Poisson sampling plan over the traditional sampling plan in terms of sample size. Lio et al. [4] proposed the sampling plan under the assumption that the failure time of the product follows the BS distribution. They determined the minimum sample size using the percentile of the BS distribution. Rao [5] designed the group sampling plan when the failure time follows the Marshall-Olkin extended Lomax distribution with known or unknown shape parameters. He determined the minimum group size through the non-linear optimization solution and discussed the results with the help of some examples. Aslam et al. [6] designed a double sampling plan and group sapling plan for the BS and determined the parameters by satisfying the producer's risk and the consumer's risk. They discussed the efficiency of the plans with the help of some examples. Al-Nasser and Al-Omari [7] worked on the sampling plan using the exponentiated Fréchet distribution. They derived the OC function for this distribution and presented an example to illustrate the plan. Aslam et al. [8] designed the attribute sampling plan using the exponentially weighted moving average (EWMA) statistics for the Weibull distribution and Burr type X distribution. They compared the plan using EWMA statistics with the traditional sampling plans for the same distribution. Al-Omari [9] proposed time truncated the double sampling plan when the lifetime follows the transmuted generalized inverse Weibull distribution. He compared the double sampling plan with the single sampling plan at the same levels of all parameters.

The inspection cost is directly affected by the change in the sample size. Larger the sample size means the more inspection cost for the lot sentencing. The various sampling schemes such as the double sampling, repetitive sampling and multiple dependent state sampling (MDSS) are more economical than the single sampling scheme. In MDSS, in case of indecision at the results of the first

sample, the previously accepted lots are considered to make a final decision about the lot. Therefore, several authors worked on the double sampling plan. Aslam et al. [10, 11] designed the double sampling using the Burr type XII distribution percentiles. They determined the smaller sample sizes for the double sampling from the optimization. Ramasamy and Sutharani [12] proposed the double sampling plan using the minimum angle method when the lifetime of the product follows the Rayleigh distribution. They concluded that double sampling using the minimum angle method is helpful in reducing the inspection cost. The half exponential power distribution in the area of the sampling plan is considered by Gui and Xu [13]. They determined smaller samples for zero and one failure scheme. Al-Omari and Zamanzade [14] worked on the designing of the double sampling plan under transmuted generalized inverse Weibull distribution and Gui and Lu [15] designed a double sampling plan for the Burr type X distribution. For the repetitive sampling plans, Aslam et al. [10, 11] developed the decision-making framework for the repetitive sampling under the Weibull distribution and the generalized exponential distribution median life. Yen et al. [16] worked on the sampling plan based on the one-sided process capability index for the lot having the one-sided specification. Balamurali et al. [17] designed repetitive sampling plans for the gamma distribution and the Weibull distribution. They discussed the application and compared these distributions with the BS distribution. Saminathan and Mahalingam [18] proposed the mixed repetitive sampling plan using the process capability index. For the MDSS plans, Balamurali and Jun [19] designed the MDSS plan for the normal distribution. Yan et al. [20] worked for this sampling using the coefficient of variation (CV) under the normal distribution. Balamurali et al. [21, 22] proposed MDSS plan for the Weibull distribution. Balamurali et al. [21, 22] used the generalized inverted exponential distribution to proposed MDSS plan. Wu et al. [23] worked for the MDSS plan for the one-sided process capability index and Balamurali et al. [21, 22] designed MDSS to reduce the inspection cost.

In this chapter, our objective is to propose a GMDSS plan for various statistical distributions. According to the best of our knowledge, there is no work on GMDSS for the gamma, Burr type XII and BS distributions. We expect that the GMDSS plan will improve the efficacy of the sampling plans by reducing the sample size is required for the lot sentencing. We will present the design of the proposed plan under these distributions and compare the performance with the existing plans using the MDSS. The rest of the book chapter is organized as: a brief introduction about the distributions is given in Sect. 2, the perforce measures and algorithm are given in Sect. 3. The illustration using the real data is given in Sect. 4. A simulation study is given in Sect. 5 and some concluding remarks are given in the last Section.

2 Some Continuous Distributions

2.1 Gamma Distribution

Let T be a lifetime random variable of the product follows a gamma distribution with shape parameter η and scale parameter λ. The cumulative distribution function (cdf) of the gamma distribution is given by

$$F(t) = \frac{1}{\Gamma\eta}\gamma\left(\eta, \frac{t}{\lambda}\right); t \geq 0, \eta > 0, \lambda > 0, \tag{1}$$

where $\gamma\left(\eta, \frac{t}{\lambda}\right)$ is the lower incomplete gamma function.

It is assumed that the shape parameter η is known. The mean lifetime of the product using gamma distribution is given as $\mu = \eta\lambda$. The failure probability is a probability of a product failure before the time t_0, that is $p = F(T \leq t_0)$ and is expressed as given below:

$$p = \frac{1}{\Gamma\eta}\gamma\left(\eta, \frac{t_0}{\lambda}\right). \tag{2}$$

It is convenient to determine the termination time t_0 as a multiple of the specified mean life μ_0, that is $t_0 = a\mu_0$. The scale parameter λ can be written in terms of mean μ.

Hence, the failure probability given in Eq. (2) can be expressed as follows:

$$p = \frac{1}{\Gamma\eta}\gamma\left(\eta, a\eta/\frac{\mu}{\mu_0}\right). \tag{3}$$

2.2 Burr Type XII Distribution

According to Lio et al. [4] and Lio et al. [24], Burr Type XII distribution is widely used in the areas of quality control and reliability analysis. Burr [25] was initially introduced the Burr Type XII distribution and subsequently, many researchers applied to different environments. Assume that the lifetime of items follows a Burr type XII distribution whose probability density function (pdf) and cdf are given below:

$$f(t) = \frac{bc}{\sigma}\left(\frac{t}{\sigma}\right)^b\left[1 + \left(\frac{t}{\sigma}\right)^b\right]^{-c-1}; t \geq 0, \sigma > 0, b > 0, c > 0, \tag{4}$$

$$F(t) = 1 - \left[1 + \left(\frac{t}{\sigma}\right)^b\right]^{-c}; t \geq 0, \sigma > 0, b > 0, c > 0, \tag{5}$$

where σ is scale parameter, b and c are shape parameters of Burr type XII distribution, assume that shape parameters are known. The median lifetime of the product using Burr type XII distribution is given as $\mu = \sigma(2^{1/c} - 1)^{1/b}$. The failure probability of the product when experiment time t_0 is denoted by p and as follows:

$$p = 1 - \left[1 + \left(\frac{t_0}{\sigma}\right)^b\right]^{-c} \tag{6}$$

It is convenient to determine the termination time t_0 as a multiple of the specified mean life μ_0, that is $t_0 = a\mu_0$. The scale parameter σ can be written in terms of mean μ.

Hence, the failure probability given in Eq. (6) can be expressed as follows:

$$p = 1 - \left[1 + \left(\frac{a\left(2^{1/c} - 1\right)^{1/b}}{\mu/\mu_0}\right)^b\right]^{-c} \tag{7}$$

2.3 Birnbaum: Saunders Distribution

The cdf of the two-parameter BS distribution is given by

$$F(t) = \Phi\left[\frac{1}{\nu}\left\{\left(\frac{t}{\delta}\right)^{1/2} - \left(\frac{\delta}{t}\right)^{1/2}\right\}\right]; t \geq 0, \delta > 0, \nu > 0, \tag{8}$$

where ν is the shape parameter, δ is a scale parameter of BS distribution and $\Phi(.)$ is cdf of the standard normal distribution. Birnbaum and Saunders [26] were proposed a life distribution for fatigue failures under cyclic loading distribution known as BS distribution. The density function of the BS distribution is unimodal and it is a family member of the normal distribution. Lemonte et al. [27] has derived the mean life of the BS distribution and is given by $\mu = \delta\left(1 + \frac{\nu^2}{2}\right)$. The probability of the failure of the product when experiment time t_0 is denoted by p and as follows:

$$p = \Phi\left[\frac{1}{\nu}\left\{\left(\frac{t_0}{\delta}\right)^{1/2} - \left(\frac{\delta}{t_0}\right)^{1/2}\right\}\right] \tag{9}$$

It is convenient to determine the termination time t_0 as a multiple of the specified mean life μ_0, that is $t_0 = a\mu_0$. The scale parameter δ can be written in terms of mean μ. Hence, the failure probability given in Eq. (9) can be written as follows:

$$p = \Phi\left[\frac{1}{\nu}\left\{\left(\frac{a\left(1+\frac{\nu^2}{2}\right)}{\mu/\mu_0}\right)^{1/2} - \left(\frac{\mu/\mu_0}{a\left(1+\frac{\nu^2}{2}\right)}\right)^{1/2}\right\}\right] \tag{10}$$

3 Designing of Proposed Generalized Multiple Dependent State Sampling Plan

3.1 Designing Methodology

The functioning procedure of the GMDSS plan for a time truncated life test is given as follows:

- *Step 1*: Select a random sample of n items from a lot. Then put them on life test for specified time t_0.
- *Step 2*: Observe the number of failures before the experiment time t_0 and denote it as d.
- *Step 3*: If the number of failures d is smaller than or equal to c_a, then accept the lot.
- *Step 4*: Reject the lot if the number of failures is larger than c_r (*i.e.* $d > c_r$)
- *Step 5*: If $c_a < d \leq c_r$, then accept the lot if at least k out of m preceding lots were all accepted with the number of failures before the experiment time t_0 is less than or equal to c_a (*i.e.* $d \leq c_a$). Otherwise, reject the lot.

The proposed plan is described by five parameters; namely, sample size (n), the maximum number of allowable failure items for unconditional acceptance ($c_a \geq 0$), the maximum number of additional failure items for conditional acceptance ($c_r > c_a$), k and m are the number of successive (preceding) lots needed for making a decision. Our proposed GMDSS plan is the general for MDSS and it becomes MDSS Plan when $m = k$, furthermore, GMDSS plan become the single sampling plan when either $(m, k) \to \infty$ and/or $c_a = c_r = c$.

3.2 Performance Measures

The performance of any sampling plan can be disclosed by its probability of acceptance, known as OC function. The OC function of the GMDSS plan for a lifetime-truncated test is given by the following.

$$P_A(p) = P_a + \left(\sum_{d=c_a+1}^{c_r} \binom{n}{d} p^d (1-p)^{n-d} \right) \left[\sum_{j=k}^{m} \binom{m}{j} P_a^j (1-P_a)^{m-j} \right] \tag{11}$$

where $P_a = p\,(d \le c_a | p) = \sum_{d=0}^{c_a} \binom{n}{d} p^d (1-p)^{n-d}$.

The plan parameters of the GMDSS plan can be determined by the optimization the following non-linear problem.

$$\text{Minimize } n \tag{12}$$

Subject to

$$P_A\,(p_1) = P_a\,(p_1) + \left(\sum_{d=c_a+1}^{c_r} \binom{n}{d} p_1^d (1-p_1)^{n-d} \right) \left[\sum_{j=k}^{m} \binom{m}{j} P_a^j\,(p_1)\,(1-P_a\,(p_1))^{m-j} \right] \ge 1-\alpha \tag{13}$$

$$P_A\,(p_2) = P_a\,(p_2) + \left(\sum_{d=c_a+1}^{c_r} \binom{n}{d} p_2^d (1-p_2)^{n-d} \right) \left[\sum_{j=k}^{m} \binom{m}{j} P_a^j\,(p_2)\,(1-P_a\,(p_2))^{m-j} \right] \le \beta \tag{14}$$

$$c_r > c_a \ge 0$$

In order to find the design parameters of the proposed GMDSS plan, the quality is represented in terms of the ratio of true mean life and specified mean life. In this approach, the quality level is measured through the ratio of its mean lifetime to the true mean lifetime, μ/μ_0. To ensure and improve the quality of the products, the producer may use the mean ratios. The probability (α) of rejecting a good lot is called the producer's risk and the probability (β) of accepting a bad lot is known as the consumer's risk. At the interest of both producer and consumer, the inference is drawn from a GMDSS plan so developed may fulfil their specified risks. Producer requires the lot acceptance probability of at least $1-\alpha$ at the acceptable reliability level (ARL), p_1, *i.e.*, the producer wants the lot to be accepted at different values of mean ratios say for the values of $\mu/\mu_0 = 2, 2.5, 3, 3.5, 4$, another hand, the consumer wants the lot tolerance reliability level (LTRL), p_2, *i.e.,* consumer may reject the lot if $\mu/\mu_0 = 1$.

3.3 Algorithm

A grid search method is used to obtain a non-linear optimization solution. This methodology requires setting up grids in the decision space. The non-linear objective function values are evaluated at the grid points. The point which corresponding to the best objective function value is considered as an optimum solution for the non-linear function. The below algorithm can be applied to obtain the optimal parameters of the proposed plan for specified values of α, β and a.

Step 1: For a given values of α, β, a and the known shape parameter of gamma, Burr XII and BS distributions.
Step 2: Initiate the values as n, c_a, c_r as $n = 2$, $c_a = 0$, $c_r = c_a + 1$.
Step 3: For fixed values α, β, $a$and the known shape parameter of gamma, Burr XII and BS distributions with the specified mean ratio, obtain the failure probability by using Eq. (3) for gamma distribution or using Eq. (7) for Burr XII distribution and using (10) for BS distribution at p_1 and p_2.
Step 4: Determine the acceptance probabilities of the lot at p_1 and p_2 using Eqs. (13) and (14), respectively which are denote as $P_A(p_1)$ and $P_A(p_2)$, respectively.
Step 5: Obtain the parameters for both the conditions in Eqs. (13) and (14) are satisfied such that the minimum value of n.
Step 6: If there exists a value of n satisfies step 5 then the desired optimal parameters are n, c_a, c_r for fixed values of m and k. Otherwise, repeat the Steps 4–5 till to get desired optimum values of parameters.

Using the above algorithm, results are computed for $m = 3$ and the values of k considered as 3, 2 and 1. Tables 1, 2, 3, and 4 are given for gamma distribution with shape parameter $\eta = 2$ and 3 at $a = 0.5$ and $a = 1.0$ respectively. Tables 5, 6, 7, 8, 9, and 10 are provided for Burr type XII distribution with shape parameters $b = 0.85$, 2.0 and $c = 5.49$, 2.0, 1.0 at $a = 0.5$ and $a = 1.0$ respectively. Tables 11, 12, 13, 14, 15, and 16 are displayed for BS distribution with shape parameter $\nu = 1.0$, 1.5, 2.0 when $a = 0.5$ and $a = 1.0$ respectively.

From Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16, we conclude the following significant facts in the proposed Sampling Plans:

1. Result shows that the ratio of its mean lifetime to the true mean lifetime, μ/μ_0 increases from 2.0 to 4.0 values the sample size (n), decreases as expected.
2. It is observed from tables that c_a and $c_r$values shows decreasing tendency while μ/μ_0 increases from 2.0 to 4.0 when the remaining combinations are fixed.
3. Further, pointed out that as k increases the values of sample size also decreases. Similar observed for various parametric combinations considered in this work.
4. It is important to note that sample size value is small when $k = m - 2$ as compared with other values of k for fixed value of m. Furthermore, noticed that when $k = m$ the value of sample size is higher than at $k = m - 1$ and $k = m - 2$ (since when $k = m$, proposed scheme reduced to MDS scheme). It show that the GMDS sampling plans is more effective than MDS sampling plans to obtain the optimum plan parameters.

Table 1 GMSDS plan parameters for gamma distribution when $a = 0.5$, $\eta = 2$, m = 3

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	14	2	4	0.9541	0.2485	11	1	11	0.9560	0.2327	13	1	3	0.9646	0.2271
0.25	2.5	10	1	10	0.9601	0.2211	10	1	2	0.9756	0.2451	8	0	2	0.9651	0.2174
0.25	3.0	10	1	10	0.9861	0.2211	6	0	6	0.9653	0.2154	7	0	1	0.9591	0.2079
0.25	3.5	10	1	10	0.9947	0.2211	6	0	6	0.9830	0.2154	7	0	1	0.9767	0.2079
0.25	4.0	5	0	5	0.9583	0.2235	6	0	6	0.9911	0.2154	7	0	1	0.9857	0.2079
0.10	2.0	29	4	14	0.9646	0.0864	19	2	5	0.9597	0.0992	21	2	4	0.9578	0.0978
0.10	2.5	19	2	12	0.9686	0.0884	14	1	11	0.9755	0.0997	16	1	3	0.9814	0.0931
0.10	3.0	14	1	11	0.9574	0.0826	14	1	11	0.9944	0.0997	11	0	3	0.9748	0.0976
0.10	3.5	14	1	11	0.9825	0.0826	8	0	2	0.9633	0.0975	9	0	1	0.9605	0.0995
0.10	4.0	14	1	11	0.9924	0.0826	8	0	2	0.9803	0.0975	9	0	1	0.9758	0.0995
0.05	2.0	37	5	15	0.9665	0.0489	28	3	13	0.9640	0.0436	25	2	5	0.9557	0.0423
0.05	2.5	27	3	13	0.9817	0.0484	22	2	7	0.9908	0.0498	18	1	3	0.9701	0.0497
0.05	3.0	22	2	12	0.9853	0.0454	17	1	11	0.9859	0.0427	12	0	2	0.9550	0.0489
0.05	3.5	16	1	11	0.9727	0.0499	17	1	11	0.9962	0.0427	12	0	2	0.9804	0.0489
0.05	4.0	16	1	11	0.9878	0.0499	10	0	2	0.9653	0.0492	11	0	1	0.9633	0.0476
0.01	2.0	56	7	17	0.9564	0.0096	46	5	15	0.9684	0.0089	37	3	7	0.9510	0.0085
0.01	2.5	40	4	14	0.9743	0.0098	29	2	12	0.9548	0.0089	25	1	5	0.9518	0.0090
0.01	3.0	29	2	12	0.9506	0.0087	29	2	12	0.9929	0.0089	24	1	4	0.9892	0.0096
0.01	3.5	29	2	12	0.9849	0.0087	23	1	11	0.9842	0.0082	18	0	4	0.9544	0.0095
0.01	4.0	23	1	11	0.9598	0.0080	23	1	11	0.9948	0.0082	17	0	2	0.9728	0.0075

Table 2 GMSDS plan parameters for gamma distribution when $a = 1.0$, $\eta = 2$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	8	3	8	0.9527	0.1881	6	2	4	0.9759	0.2420	5	1	3	0.9571	0.2376
0.25	2.5	6	2	6	0.9786	0.1930	4	1	3	0.9863	0.2490	3	0	3	0.9508	0.2420
0.25	3.0	4	1	4	0.9718	0.1914	2	0	2	0.9527	0.2254	3	0	3	0.9806	0.2420
0.25	3.5	4	1	4	0.9881	0.1914	2	0	2	0.9753	0.2254	3	0	3	0.9918	0.2420
0.25	4.0	4	1	4	0.9947	0.1914	2	0	2	0.9864	0.2254	3	0	3	0.9963	0.2420
0.10	2.0	14	5	14	0.9501	0.0643	10	3	10	0.9572	0.0689	9	2	7	0.9626	0.0997
0.10	2.5	10	3	10	0.9698	0.0596	8	2	8	0.9836	0.0615	6	1	3	0.9752	0.0971
0.10	3.0	8	2	8	0.9750	0.0538	5	1	2	0.9677	0.0973	4	0	3	0.9533	0.0945
0.10	3.5	5	1	5	0.9726	0.0924	5	1	2	0.9855	0.0973	4	0	3	0.9790	0.0945
0.10	4.0	5	1	5	0.9872	0.0924	3	0	3	0.9623	0.0789	4	0	3	0.9901	0.0945
0.05	2.0	17	6	16	0.9519	0.0391	13	4	13	0.9639	0.0391	12	3	7	0.9717	0.0452
0.05	2.5	13	4	13	0.9803	0.0355	9	2	9	0.9656	0.0295	8	1	8	0.9510	0.0370
0.05	3.0	9	2	9	0.9563	0.0274	6	1	6	0.9760	0.0491	7	1	3	0.9846	0.0373
0.05	3.5	9	2	9	0.9858	0.0274	6	1	6	0.9926	0.0491	5	0	5	0.9591	0.0434
0.05	4.0	6	1	6	0.9752	0.0439	6	1	6	0.9976	0.0491	5	0	5	0.9799	0.0434
0.01	2.0	26	9	19	0.9601	0.0094	20	6	16	0.9559	0.0077	18	5	8	0.9612	0.0095
0.01	2.5	18	5	15	0.9662	0.0067	13	3	13	0.9705	0.0090	12	2	7	0.9713	0.0087
0.01	3.0	13	3	13	0.9699	0.0088	11	2	11	0.9780	0.0068	9	1	4	0.9745	0.0077
0.01	3.5	11	2	11	0.9629	0.0067	8	1	8	0.9709	0.0096	9	1	4	0.9933	0.0077
0.01	4.0	11	2	11	0.9864	0.0067	8	1	8	0.9894	0.0096	6	0	2	0.9539	0.0069

Table 3 GMSDS plan parameters for gamma distribution when $a = 0.5$, $\eta = 3$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	14	1	11	0.9685	0.2294	15	1	3	0.9926	0.2345	11	0	2	0.9740	0.2419
0.25	2.5	14	1	11	0.9953	0.2294	8	0	2	0.9862	0.2391	10	0	1	0.9770	0.2100
0.25	3.0	7	0	7	0.9747	0.2355	8	0	2	0.9962	0.2391	10	0	1	0.9911	0.2100
0.25	3.5	7	0	7	0.9882	0.2355	8	0	2	0.9988	0.2391	10	0	1	0.9960	0.2100
0.25	4.0	7	0	7	0.9940	0.2355	8	0	2	0.9996	0.2391	10	0	1	0.9981	0.2100
0.10	2.0	27	2	12	0.9757	0.0884	20	1	11	0.9811	0.0999	23	1	3	0.9841	0.0888
0.10	2.5	19	1	11	0.9861	0.0983	12	0	10	0.9632	0.0945	13	0	1	0.9607	0.0982
0.10	3.0	19	1	11	0.9974	0.0983	12	0	10	0.9891	0.0945	13	0	1	0.9847	0.0982
0.10	3.5	11	0	10	0.9728	0.0978	12	0	10	0.9964	0.0945	13	0	1	0.9932	0.0982
0.10	4.0	11	0	10	0.9860	0.0978	12	0	10	0.9987	0.0945	13	0	1	0.9967	0.0982
0.05	2.0	31	2	12	0.9561	0.0479	24	1	11	0.9593	0.0457	26	1	3	0.9742	0.0464
0.05	2.5	23	1	11	0.9738	0.0490	23	1	2	0.9823	0.0497	18	0	2	0.9786	0.0401
0.05	3.0	23	1	11	0.9947	0.0490	15	0	10	0.9805	0.0463	16	0	1	0.9766	0.0459
0.05	3.5	15	0	10	0.9530	0.0416	15	0	10	0.9934	0.0463	16	0	1	0.9896	0.0459
0.05	4.0	15	0	10	0.9753	0.0416	15	0	10	0.9976	0.0463	16	0	1	0.9949	0.0459
0.01	2.0	49	3	13	0.9515	0.0098	41	2	12	0.9689	0.0097	35	1	4	0.9578	0.0083
0.01	2.5	41	2	12	0.9868	0.0094	32	1	11	0.9872	0.0099	25	0	3	0.9594	0.0089
0.01	3.0	32	1	11	0.9834	0.0096	22	0	10	0.9501	0.0097	23	0	1	0.9516	0.0085
0.01	3.5	32	1	11	0.9960	0.0096	22	0	10	0.9820	0.0097	23	0	1	0.9784	0.0085
0.01	4.0	22	0	10	0.9513	0.0094	22	0	10	0.9931	0.0097	23	0	1	0.9894	0.0085

Table 4 GMSDS plan parameters for gamma distribution when $a = 1.0$, $\eta = 3$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	14	1	11	0.9685	0.2294	15	1	3	0.9926	0.2345	11	0	2	0.9740	0.2419
0.25	2.5	14	1	11	0.9953	0.2294	8	0	2	0.9862	0.2391	10	0	1	0.9770	0.2100
0.25	3.0	7	0	7	0.9747	0.2355	8	0	2	0.9962	0.2391	10	0	1	0.9911	0.2100
0.25	3.5	7	0	7	0.9882	0.2355	8	0	2	0.9988	0.2391	10	0	1	0.9960	0.2100
0.25	4.0	7	0	7	0.9940	0.2355	8	0	2	0.9996	0.2391	10	0	1	0.9981	0.2100
0.10	2.0	27	2	12	0.9757	0.0884	20	1	11	0.9811	0.0999	23	1	3	0.9841	0.0888
0.10	2.5	19	1	11	0.9861	0.0983	12	0	10	0.9632	0.0945	13	0	1	0.9607	0.0982
0.10	3.0	19	1	11	0.9974	0.0983	12	0	10	0.9891	0.0945	13	0	1	0.9847	0.0982
0.10	3.5	11	0	10	0.9728	0.0978	12	0	10	0.9964	0.0945	13	0	1	0.9932	0.0982
0.10	4.0	11	0	10	0.9860	0.0978	12	0	10	0.9987	0.0945	13	0	1	0.9967	0.0982
0.05	2.0	31	2	12	0.9561	0.0479	24	1	11	0.9593	0.0457	26	1	3	0.9742	0.0464
0.05	2.5	23	1	11	0.9738	0.0490	23	1	2	0.9823	0.0497	18	0	2	0.9786	0.0401
0.05	3.0	23	1	11	0.9947	0.0490	15	0	10	0.9805	0.0463	16	0	1	0.9766	0.0459
0.05	3.5	15	0	10	0.9530	0.0416	15	0	10	0.9934	0.0463	16	0	1	0.9896	0.0459
0.05	4.0	15	0	10	0.9753	0.0416	15	0	10	0.9976	0.0463	16	0	1	0.9949	0.0459
0.01	2.0	49	3	13	0.9515	0.0098	41	2	12	0.9689	0.0097	35	1	4	0.9578	0.0083
0.01	2.5	41	2	12	0.9868	0.0094	32	1	11	0.9872	0.0099	25	0	3	0.9594	0.0089
0.01	3.0	32	1	11	0.9834	0.0096	22	0	10	0.9501	0.0097	23	0	1	0.9516	0.0085
0.01	3.5	32	1	11	0.9960	0.0096	22	0	10	0.9820	0.0097	23	0	1	0.9784	0.0085
0.01	4.0	22	0	10	0.9513	0.0094	22	0	10	0.9931	0.0097	23	0	1	0.9894	0.0085

Table 5 GMSDS plan parameters for Burr XII distribution when $a = 0.5$, $b = 0.85$, $c = 5.49$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	42	11	21	0.9634	0.2462	30	7	12	0.9633	0.2483	30	6	10	0.9526	0.2326
0.25	2.5	25	6	8	0.9528	0.2481	19	4	6	0.9505	0.2448	19	3	7	0.9656	0.2460
0.25	3.0	19	4	14	0.9524	0.2121	16	3	9	0.9795	0.2498	15	2	5	0.9644	0.2206
0.25	3.5	15	3	13	0.9612	0.2345	12	2	4	0.9676	0.2395	11	1	4	0.9579	0.2279
0.25	4.0	15	3	13	0.9780	0.2345	12	2	4	0.9813	0.2395	11	1	4	0.9748	0.2279
0.10	2.0	69	17	27	0.9539	0.0970	49	11	21	0.9549	0.0989	46	9	15	0.9502	0.0941
0.10	2.5	45	10	20	0.9581	0.0889	31	6	16	0.9559	0.0946	30	5	9	0.9552	0.0925
0.10	3.0	34	7	17	0.9645	0.0912	23	4	8	0.9614	0.0982	23	3	8	0.9582	0.0887
0.10	3.5	27	5	15	0.9555	0.0830	19	3	6	0.9630	0.0985	19	2	8	0.9557	0.0953
0.10	4.0	23	4	14	0.9593	0.0864	15	2	5	0.9550	0.0996	18	2	5	0.9710	0.0800
0.05	2.0	–	–	–	–	–	67	15	25	0.9593	0.0497	61	12	20	0.9547	0.0479
0.05	2.5	60	13	23	0.9589	0.0434	42	8	18	0.9511	0.0427	38	6	13	0.9504	0.0470
0.05	3.0	41	8	18	0.9505	0.0475	30	5	15	0.9505	0.0466	29	4	9	0.9615	0.0475
0.05	3.5	37	7	17	0.9728	0.0496	26	4	14	0.9632	0.0465	24	3	6	0.9510	0.0414
0.05	4.0	30	5	15	0.9543	0.0419	22	3	13	0.9594	0.0449	20	2	6	0.9630	0.0495
0.01	2.0	–	–	–	–	–	–	–	–	–	–	90	18	26	0.9502	0.0100
0.01	2.5	91	19	29	0.9585	0.0093	68	13	22	0.9604	0.0100	59	10	16	0.9567	0.0090
0.01	3.0	68	13	23	0.9637	0.0098	53	9	19	0.9634	0.0086	43	6	12	0.9534	0.0090
0.01	3.5	57	10	20	0.9642	0.0085	45	7	17	0.9707	0.0083	38	5	10	0.9733	0.0098
0.01	4.0	49	8	18	0.9652	0.0084	36	5	15	0.9644	0.0097	33	4	7	0.9538	0.0085

Table 6 GMSDS plan parameters for Burr XII distribution when $a = 1.0$, $b = 0.85$, $c = 5.49$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	27	11	21	0.9517	0.2294	19	7	17	0.9574	0.2495	21	7	11	0.9636	0.2425
0.25	2.5	16	6	16	0.9541	0.2363	15	5	15	0.9717	0.2030	12	3	8	0.9594	0.2467
0.25	3.0	14	5	14	0.9700	0.2195	10	3	10	0.9742	0.2369	10	2	10	0.9538	0.2015
0.25	3.5	12	4	12	0.9714	0.1997	8	2	8	0.9611	0.1930	9	2	4	0.9615	0.1908
0.25	4.0	10	3	10	0.9615	0.1761	8	2	8	0.9785	0.1930	7	1	7	0.9681	0.2275
0.10	2.0	46	18	28	0.9507	0.0926	33	12	22	0.9577	0.0983	31	10	16	0.9564	0.0973
0.10	2.5	28	10	20	0.9534	0.0932	19	6	12	0.9529	0.1000	19	5	10	0.9535	0.0913
0.10	3.0	21	7	17	0.9608	0.0954	17	5	15	0.9713	0.0854	14	3	8	0.9510	0.0922
0.10	3.5	19	6	16	0.9707	0.0841	12	3	12	0.9554	0.0871	11	2	6	0.9557	0.0985
0.10	4.0	14	4	14	0.9561	0.0904	12	3	12	0.9778	0.0871	11	2	6	0.9769	0.0985
0.05	2.0	42	14	20	0.9628	0.0484	47	17	27	0.9615	0.0433	42	14	20	0.9628	0.0484
0.05	2.5	26	7	15	0.9584	0.0497	28	9	19	0.9656	0.0488	37	13	23	0.9572	0.0495
0.05	3.0	18	4	10	0.9514	0.0494	21	6	16	0.9635	0.0435	28	9	19	0.9565	0.0437
0.05	3.5	15	3	7	0.9589	0.0426	16	4	14	0.9533	0.0426	23	7	17	0.9655	0.0467
0.05	4.0	15	3	7	0.9808	0.0426	13	3	6	0.9543	0.0490	18	5	15	0.9542	0.0482
0.01	2.0	–	–	–	–	–	–	–	–	–	–	61	20	28	0.9554	0.0093
0.01	2.5	59	20	30	0.9577	0.0092	42	13	23	0.9508	0.0100	39	11	18	0.9671	0.0100
0.01	3.0	42	13	23	0.9550	0.0098	35	10	20	0.9704	0.0085	28	7	11	0.9535	0.0094
0.01	3.5	35	10	20	0.9559	0.0083	27	7	17	0.9692	0.0098	23	5	9	0.9586	0.0084
0.01	4.0	30	8	18	0.9570	0.0081	22	5	15	0.9534	0.0087	20	4	7	0.9511	0.0081

Table 7 GMSDS plan parameters for Burr XII distribution when $a = 0.5$, $b = 2$, $c = 2$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	N	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	22	2	12	0.9759	0.2276	16	1	3	0.9762	0.2387	20	1	3	0.9798	0.2174
0.25	2.5	15	1	11	0.9810	0.2305	9	0	9	0.9582	0.2333	10	0	1	0.9551	0.2494
0.25	3.0	15	1	11	0.9944	0.2305	8	0	1	0.9754	0.2465	10	0	1	0.9780	0.2494
0.25	3.5	8	0	8	0.9581	0.2137	8	0	1	0.9875	0.2465	10	0	1	0.9880	0.2494
0.25	4.0	8	0	8	0.9739	0.2137	8	0	1	0.9931	0.2465	10	0	1	0.9929	0.2494
0.10	2.0	36	3	13	0.9715	0.0941	29	2	5	0.9836	0.0983	25	1	3	0.9517	0.0836
0.10	2.5	29	2	12	0.9877	0.0882	21	1	3	0.9875	0.0973	16	0	2	0.9592	0.0909
0.10	3.0	21	1	11	0.9819	0.0895	13	0	10	0.9577	0.0928	14	0	1	0.9563	0.0978
0.10	3.5	21	1	11	0.9936	0.0895	13	0	10	0.9802	0.0928	14	0	1	0.9761	0.0978
0.10	4.0	21	1	11	0.9975	0.0895	13	0	10	0.9901	0.0928	14	0	1	0.9858	0.0978
0.05	2.0	49	4	14	0.9749	0.0471	34	2	12	0.9655	0.0483	29	1	4	0.9542	0.0498
0.05	2.5	33	2	12	0.9778	0.0499	25	1	4	0.9802	0.0498	28	1	3	0.9854	0.0458
0.05	3.0	25	1	11	0.9681	0.0468	25	1	4	0.9962	0.0498	19	0	2	0.9764	0.0437
0.05	3.5	25	1	11	0.9882	0.0468	16	0	10	0.9667	0.0478	17	0	1	0.9646	0.0483
0.05	4.0	25	1	11	0.9953	0.0468	16	0	10	0.9830	0.0478	17	0	1	0.9790	0.0483
0.01	2.0	70	5	15	0.9519	0.0089	53	3	13	0.9511	0.0096	48	2	6	0.9570	0.0083
0.01	2.5	53	3	13	0.9769	0.0093	44	2	12	0.9870	0.0098	36	1	3	0.9615	0.0092
0.01	3.0	44	2	12	0.9834	0.0095	35	1	11	0.9827	0.0089	27	0	3	0.9522	0.0087
0.01	3.5	35	1	11	0.9644	0.0087	35	1	11	0.9957	0.0089	26	0	2	0.9755	0.0082
0.01	4.0	35	1	11	0.9847	0.0087	24	0	10	0.9534	0.0091	25	0	1	0.9545	0.0081

Table 8 GMSDS plan parameters for Burr XII distribution when $a = 1.0$, $b = 2$, $c = 2$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	7	2	7	0.9655	0.2356	5	1	3	0.9689	0.2452	7	1	7	0.9824	0.2275
0.25	2.5	5	1	5	0.9658	0.1929	5	1	3	0.9951	0.2452	3	0	1	0.9509	0.2488
0.25	3.0	5	1	5	0.9891	0.1929	3	0	3	0.9665	0.1626	3	0	1	0.9762	0.2488
0.25	3.5	5	1	5	0.9962	0.1929	3	0	3	0.9844	0.1626	3	0	1	0.9871	0.2488
0.25	4.0	5	1	5	0.9985	0.1929	3	0	3	0.9922	0.1626	3	0	1	0.9924	0.2488
0.10	2.0	14	4	14	0.9784	0.0904	9	2	4	0.9661	0.0992	8	1	4	0.9590	0.0964
0.10	2.5	9	2	9	0.9806	0.0905	7	1	7	0.9785	0.0730	7	1	2	0.9566	0.0914
0.10	3.0	7	1	7	0.9640	0.0627	7	1	7	0.9956	0.0730	5	0	2	0.9774	0.0738
0.10	3.5	7	1	7	0.9863	0.0627	4	0	4	0.9677	0.0730	5	0	2	0.9912	0.0738
0.10	4.0	7	1	7	0.9944	0.0627	4	0	4	0.9834	0.0730	5	0	2	0.9963	0.0738
0.05	2.0	18	5	15	0.9793	0.0482	13	3	6	0.9785	0.0490	12	2	5	0.9713	0.0402
0.05	2.5	11	2	11	0.9505	0.0327	8	1	8	0.9612	0.0387	9	1	4	0.9902	0.0471
0.05	3.0	11	2	11	0.9891	0.0327	8	1	8	0.9915	0.0387	6	0	3	0.9690	0.0452
0.05	3.5	8	1	8	0.9781	0.0352	8	1	8	0.9980	0.0387	5	0	1	0.9581	0.0454
0.05	4.0	8	1	8	0.9908	0.0352	5	0	5	0.9708	0.0340	5	0	1	0.9752	0.0454
0.01	2.0	27	7	17	0.9755	0.0096	19	4	14	0.9638	0.0099	17	3	6	0.9625	0.0094
0.01	2.5	19	4	14	0.9865	0.0096	14	2	12	0.9679	0.0066	12	1	6	0.9646	0.0090
0.01	3.0	14	2	12	0.9656	0.0065	11	1	11	0.9637	0.0060	11	1	3	0.9856	0.0077
0.01	3.5	14	2	12	0.9909	0.0065	11	1	11	0.9900	0.0060	8	0	3	0.9701	0.0081
0.01	4.0	11	1	11	0.9722	0.0059	11	1	11	0.9971	0.0060	8	0	3	0.9870	0.0081

Table 9 GMSDS plan parameters for Burr XII distribution when $a = 0.5$, $b = 2$, $c = 1$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	19	2	12	0.9742	0.2470	14	1	3	0.9732	0.2489	18	1	3	0.9749	0.2071
0.25	2.5	13	1	11	0.9792	0.2434	8	0	8	0.9519	0.2302	11	0	2	0.9778	0.2114
0.25	3.0	13	1	11	0.9937	0.2434	8	0	8	0.9801	0.2302	9	0	1	0.9748	0.2402
0.25	3.5	7	0	7	0.9544	0.2170	8	0	8	0.9910	0.2302	9	0	1	0.9862	0.2402
0.25	4.0	7	0	7	0.9714	0.2170	8	0	8	0.9956	0.2302	9	0	1	0.9918	0.2402
0.10	2.0	32	3	13	0.9639	0.0938	26	2	7	0.9806	0.0998	23	1	4	0.9664	0.0928
0.10	2.5	25	2	12	0.9864	0.0991	19	1	6	0.9881	0.0994	14	0	2	0.9542	0.0950
0.10	3.0	18	1	11	0.9806	0.1000	11	0	2	0.9540	0.0970	14	0	2	0.9841	0.0950
0.10	3.5	18	1	11	0.9930	0.1000	11	0	2	0.9784	0.0970	13	0	1	0.9705	0.0829
0.10	4.0	18	1	11	0.9972	0.1000	11	0	2	0.9891	0.0970	13	0	1	0.9825	0.0829
0.05	2.0	44	4	14	0.9648	0.0441	30	2	12	0.9574	0.0496	32	2	4	0.9562	0.0476
0.05	2.5	30	2	12	0.9691	0.0443	22	1	3	0.9690	0.0499	25	1	3	0.9818	0.0439
0.05	3.0	22	1	11	0.9628	0.0481	22	1	3	0.9929	0.0499	17	0	2	0.9712	0.0415
0.05	3.5	22	1	11	0.9859	0.0481	14	0	10	0.9625	0.0494	15	0	1	0.9606	0.0486
0.05	4.0	22	1	11	0.9943	0.0481	14	0	10	0.9806	0.0494	15	0	1	0.9766	0.0486
0.01	2.0	69	6	16	0.9673	0.0092	55	4	14	0.9722	0.0088	49	3	6	0.9658	0.0089
0.01	2.5	47	3	13	0.9697	0.0093	39	2	12	0.9824	0.0097	32	1	3	0.9531	0.0090
0.01	3.0	39	2	12	0.9788	0.0095	31	1	11	0.9780	0.0089	32	1	3	0.9875	0.0090
0.01	3.5	31	1	11	0.9574	0.0087	31	1	11	0.9943	0.0089	23	0	2	0.9710	0.0081
0.01	4.0	31	1	11	0.9814	0.0087	31	1	11	0.9984	0.0089	23	0	2	0.9869	0.0081

Table 10 GMSDS plan parameters for Burr XII distribution when $a = 1.0$, $b = 2$, $c = 1$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	10	3	10	0.9613	0.1761	8	2	8	0.9783	0.1930	7	1	7	0.9679	0.2275
0.25	2.5	7	2	7	0.9899	0.2356	5	1	3	0.9905	0.2452	4	0	4	0.9598	0.2275
0.25	3.0	5	1	5	0.9817	0.1929	3	0	3	0.9511	0.1626	3	0	1	0.9672	0.2488
0.25	3.5	5	1	5	0.9932	0.1929	3	0	3	0.9762	0.1626	3	0	1	0.9820	0.2488
0.25	4.0	5	1	5	0.9973	0.1929	3	0	3	0.9877	0.1626	3	0	1	0.9893	0.2488
0.10	2.0	14	4	14	0.9557	0.0904	12	3	12	0.9776	0.0871	11	2	6	0.9767	0.0985
0.10	2.5	9	2	9	0.9641	0.0905	7	1	7	0.9610	0.0730	8	1	4	0.9895	0.0964
0.10	3.0	9	2	9	0.9920	0.0905	7	1	7	0.9909	0.0730	5	0	2	0.9639	0.0738
0.10	3.5	7	1	7	0.9765	0.0627	4	0	4	0.9519	0.0730	5	0	2	0.9852	0.0738
0.10	4.0	7	1	7	0.9900	0.0627	4	0	4	0.9743	0.0730	5	0	2	0.9935	0.0738
0.05	2.0	18	5	15	0.9537	0.0482	13	3	6	0.9539	0.0490	15	3	7	0.9806	0.0426
0.05	2.5	13	3	13	0.9766	0.0462	11	2	11	0.9832	0.0358	9	1	4	0.9798	0.0471
0.05	3.0	11	2	11	0.9780	0.0327	8	1	8	0.9829	0.0387	6	0	3	0.9507	0.0452
0.05	3.5	8	1	8	0.9632	0.0352	8	1	8	0.9955	0.0387	6	0	3	0.9797	0.0452
0.05	4.0	8	1	8	0.9838	0.0352	5	0	5	0.9557	0.0340	5	0	1	0.9652	0.0454
0.01	2.0	30	8	18	0.9565	0.0081	22	5	15	0.9529	0.0087	20	4	7	0.9507	0.0081
0.01	2.5	19	4	14	0.9674	0.0096	17	3	13	0.9795	0.0065	14	2	4	0.9586	0.0081
0.01	3.0	17	3	13	0.9812	0.0064	14	2	12	0.9893	0.0066	11	1	3	0.9736	0.0077
0.01	3.5	14	2	12	0.9808	0.0065	11	1	11	0.9794	0.0060	8	0	3	0.9512	0.0081
0.01	4.0	11	1	11	0.9532	0.0059	11	1	11	0.9935	0.0060	8	0	3	0.9775	0.0081

Table 11 GMSDS plan parameters for BS distribution when $a = 0.5$, $\nu = 1$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	13	3	13	0.9578	0.2009	10	2	4	0.9722	0.2354	9	1	4	0.9650	0.2362
0.25	2.5	10	2	10	0.9858	0.1958	7	1	7	0.9907	0.2454	5	0	2	0.9632	0.2348
0.25	3.0	7	1	7	0.9844	0.1817	4	0	4	0.9645	0.1886	5	0	2	0.9901	0.2348
0.25	3.5	3	0	3	0.9565	0.2405	4	0	4	0.9875	0.1886	5	0	2	0.9974	0.2348
0.25	4.0	3	0	3	0.9787	0.2405	4	0	4	0.9957	0.1886	5	0	2	0.9993	0.2348
0.10	2.0	22	5	15	0.9678	0.0922	16	3	13	0.9716	0.0969	15	2	6	0.9673	0.0956
0.10	2.5	12	2	12	0.9662	0.0998	9	1	9	0.9701	0.0998	11	1	4	0.9886	0.0935
0.10	3.0	9	1	9	0.9633	0.0827	9	1	9	0.9955	0.0998	7	0	3	0.9778	0.0996
0.10	3.5	9	1	9	0.9903	0.0827	5	0	1	0.9614	0.0928	6	0	1	0.9703	0.0840
0.10	4.0	9	1	9	0.9976	0.0827	5	0	1	0.9840	0.0928	6	0	1	0.9866	0.0840
0.05	2.0	28	6	16	0.9574	0.0431	22	4	14	0.9644	0.0389	20	3	7	0.9746	0.0493
0.05	2.5	18	3	13	0.9729	0.0421	15	2	12	0.9830	0.0376	12	1	3	0.9609	0.0408
0.05	3.0	14	2	12	0.9891	0.0493	11	1	11	0.9881	0.0406	8	0	2	0.9555	0.0392
0.05	3.5	11	1	11	0.9804	0.0368	11	1	11	0.9983	0.0406	7	0	1	0.9588	0.0465
0.05	4.0	11	1	11	0.9949	0.0368	7	0	7	0.9810	0.0358	7	0	1	0.9813	0.0465
0.01	2.0	44	9	19	0.9567	0.0081	34	6	16	0.9559	0.0075	28	4	9	0.9580	0.0092
0.01	2.5	26	4	14	0.9652	0.0096	23	3	13	0.9804	0.0075	19	2	4	0.9571	0.0093
0.01	3.0	19	2	12	0.9556	0.0075	15	1	11	0.9541	0.0070	15	1	3	0.9804	0.0090
0.01	3.5	19	2	12	0.9926	0.0075	15	1	11	0.9920	0.0070	11	0	3	0.9744	0.0091
0.01	4.0	15	1	11	0.9847	0.0069	10	0	10	0.9541	0.0077	10	0	1	0.9606	0.0086

Table 12 GMSDS plan parameters for BS distribution when $a = 1.0$, $\nu = 1$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	12	6	12	0.9522	0.2005	9	4	9	0.9656	0.2141	10	4	7	0.9784	0.2342
0.25	2.5	7	3	7	0.9558	0.1914	6	2	6	0.9560	0.1376	7	2	7	0.9833	0.1861
0.25	3.0	5	2	5	0.9764	0.2307	4	1	4	0.9633	0.1528	4	1	2	0.9519	0.2142
0.25	3.5	5	2	5	0.9924	0.2307	4	1	4	0.9874	0.1528	3	0	3	0.9507	0.1501
0.25	4.0	4	1	4	0.9651	0.1200	4	1	4	0.9959	0.1528	3	0	3	0.9759	0.1501
0.10	2.0	22	11	21	0.9750	0.0926	14	6	14	0.9511	0.0776	13	5	8	0.9540	0.0900
0.10	2.5	12	5	12	0.9604	0.0754	10	4	7	0.9884	0.0983	8	2	8	0.9570	0.0863
0.10	3.0	8	3	8	0.9728	0.0974	7	2	7	0.9718	0.0571	6	1	6	0.9620	0.0774
0.10	3.5	7	2	7	0.9534	0.0503	5	1	5	0.9630	0.0562	6	1	6	0.9891	0.0774
0.10	4.0	7	2	7	0.9830	0.0503	5	1	5	0.9866	0.0562	5	1	2	0.9708	0.0738
0.05	2.0	27	13	23	0.9655	0.0441	20	9	15	0.9770	0.0496	16	6	10	0.9532	0.0451
0.05	2.5	15	6	15	0.9514	0.0361	11	4	11	0.9740	0.0494	11	3	11	0.9630	0.0409
0.05	3.0	11	4	11	0.9735	0.0441	9	3	5	0.9838	0.0498	8	2	4	0.9701	0.0389
0.05	3.5	9	3	9	0.9845	0.0478	8	2	8	0.9835	0.0238	6	1	3	0.9751	0.0385
0.05	4.0	8	2	8	0.9667	0.0223	6	1	6	0.9682	0.0211	6	1	3	0.9907	0.0385
0.01	2.0	39	18	28	0.9555	0.0090	30	13	23	0.9670	0.0097	24	9	15	0.9522	0.0097
0.01	2.5	23	9	19	0.9580	0.0079	17	6	11	0.9675	0.0100	14	4	8	0.9592	0.0098
0.01	3.0	17	6	16	0.9783	0.0098	14	4	14	0.9695	0.0050	10	2	6	0.9589	0.0096
0.01	3.5	14	4	14	0.9664	0.0049	9	2	9	0.9655	0.0099	10	2	6	0.9917	0.0096
0.01	4.0	12	3	12	0.9734	0.0047	9	2	9	0.9911	0.0099	8	1	5	0.9839	0.0080

Table 13 GMSDS plan parameters for BS distribution when $a = 0.5$, $\nu = 1.5$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	26	11	21	0.9605	0.2347	21	8	18	0.9643	0.2074	19	6	16	0.9554	0.2319
0.25	2.5	16	6	16	0.9567	0.1953	12	4	12	0.9723	0.2244	11	3	6	0.9685	0.2459
0.25	3.0	9	3	9	0.9502	0.2318	10	3	10	0.9822	0.1964	9	2	6	0.9830	0.2456
0.25	3.5	9	3	9	0.9800	0.2318	5	1	5	0.9554	0.2306	6	1	3	0.9698	0.2323
0.25	4.0	7	2	7	0.9718	0.2075	5	1	5	0.9778	0.2306	6	1	3	0.9850	0.2323
0.10	2.0	42	17	27	0.9563	0.0993	32	12	22	0.9629	0.0929	30	10	16	0.9625	0.0935
0.10	2.5	25	9	19	0.9559	0.0869	16	5	10	0.9506	0.0987	18	5	9	0.9667	0.0881
0.10	3.0	18	6	16	0.9695	0.0945	14	4	14	0.9745	0.0852	13	3	6	0.9659	0.0793
0.10	3.5	14	4	14	0.9561	0.0720	9	2	9	0.9559	0.0898	10	2	4	0.9578	0.0815
0.10	4.0	11	3	11	0.9712	0.0946	9	2	9	0.9824	0.0898	8	1	5	0.9715	0.0958
0.05	2.0	58	23	33	0.9579	0.0453	41	15	25	0.9508	0.0418	36	12	18	0.9552	0.0494
0.05	2.5	34	12	22	0.9594	0.0412	25	8	18	0.9651	0.0425	22	6	11	0.9634	0.0455
0.05	3.0	22	7	17	0.9605	0.0495	15	4	8	0.9523	0.0495	17	4	8	0.9741	0.0403
0.05	3.5	18	5	15	0.9542	0.0359	13	3	13	0.9602	0.0396	12	2	7	0.9618	0.0475
0.05	4.0	15	4	14	0.9738	0.0459	10	2	10	0.9663	0.0497	11	2	4	0.9666	0.0417
0.01	2.0	86	33	43	0.9500	0.0093	66	24	34	0.9565	0.0092	57	19	28	0.9572	0.0090
0.01	2.5	50	17	27	0.9557	0.0091	39	12	22	0.9503	0.0070	33	9	16	0.9595	0.0090
0.01	3.0	36	11	21	0.9634	0.0087	26	7	17	0.9603	0.0097	22	5	9	0.9522	0.0090
0.01	3.5	26	7	17	0.9553	0.0094	21	5	15	0.9692	0.0093	19	4	7	0.9665	0.0093
0.01	4.0	21	5	15	0.9519	0.0090	19	4	14	0.9746	0.0067	14	2	6	0.9587	0.0093

Table 14 GMSDS plan parameters for BS distribution when $a = 1.0$, $\nu = 1.5$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	23	14	18	0.9503	0.2494	19	11	14	0.9609	0.2425	18	9	18	0.9545	0.2314
0.25	2.5	14	8	14	0.9559	0.2375	13	7	13	0.9814	0.2380	11	5	8	0.9625	0.2215
0.25	3.0	11	6	11	0.9674	0.2261	10	5	10	0.9843	0.2116	8	3	8	0.9671	0.2216
0.25	3.5	8	4	8	0.9565	0.2069	5	2	5	0.9543	0.2300	6	2	5	0.9744	0.2390
0.25	4.0	8	4	8	0.9784	0.2069	5	2	5	0.9758	0.2300	4	1	3	0.9605	0.2473
0.10	2.0	44	26	36	0.9580	0.0898	32	18	28	0.9643	0.0905	29	15	21	0.9620	0.0940
0.10	2.5	27	15	25	0.9671	0.0866	18	9	18	0.9506	0.0746	16	7	13	0.9550	0.0985
0.10	3.0	19	10	19	0.9720	0.0900	11	5	11	0.9536	0.0998	12	5	8	0.9666	0.0973
0.10	3.5	14	7	14	0.9734	0.0992	11	5	11	0.9842	0.0998	9	3	8	0.9633	0.0999
0.10	4.0	11	5	11	0.9599	0.0827	8	3	8	0.9564	0.0707	7	2	5	0.9591	0.0873
0.05	2.0	55	32	42	0.9534	0.0471	38	21	31	0.9512	0.0480	37	19	27	0.9575	0.0459
0.05	2.5	33	18	28	0.9642	0.0482	22	11	21	0.9532	0.0469	22	10	16	0.9660	0.0466
0.05	3.0	22	11	21	0.9500	0.0420	17	8	17	0.9726	0.0484	15	6	10	0.9593	0.0398
0.05	3.5	17	8	17	0.9549	0.0433	12	5	12	0.9587	0.0457	12	4	12	0.9510	0.0403
0.05	4.0	16	7	16	0.9570	0.0277	12	5	12	0.9848	0.0457	10	3	10	0.9616	0.0446
0.01	2.0	86	49	59	0.9522	0.0086	62	34	44	0.9545	0.0100	54	28	36	0.9569	0.0094
0.01	2.5	51	27	37	0.9624	0.0089	38	19	29	0.9642	0.0089	31	14	21	0.9556	0.0093
0.01	3.0	35	17	27	0.9516	0.0073	28	13	23	0.9750	0.0089	22	9	14	0.9640	0.0097
0.01	3.5	28	13	23	0.9690	0.0087	21	9	19	0.9759	0.0099	17	6	11	0.9567	0.0080
0.01	4.0	21	9	19	0.9579	0.0096	16	6	16	0.9542	0.0080	13	4	8	0.9508	0.0090

Table 15 GMSDS plan parameters for BS distribution when $a = 0.5$, $\nu = 2.0$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	48	25	30	0.9658	0.2495	31	15	23	0.9567	0.2490	32	14	22	0.9577	0.2438
0.25	2.5	29	14	24	0.9545	0.1936	18	8	18	0.9574	0.2416	18	7	12	0.9607	0.2469
0.25	3.0	19	9	19	0.9694	0.2461	14	6	9	0.9690	0.2414	12	4	8	0.9520	0.2292
0.25	3.5	16	7	16	0.9601	0.1861	10	4	6	0.9596	0.2483	10	3	8	0.9639	0.2461
0.25	4.0	12	5	12	0.9582	0.2000	10	4	6	0.9784	0.2483	8	2	8	0.9550	0.2283
0.10	2.0	74	37	47	0.9532	0.0994	56	27	33	0.9599	0.0984	50	22	31	0.9532	0.0970
0.10	2.5	45	21	31	0.9505	0.0818	32	14	24	0.9562	0.0861	28	11	17	0.9552	0.0999
0.10	3.0	33	15	25	0.9729	0.0993	22	9	19	0.9634	0.0956	20	7	12	0.9539	0.0890
0.10	3.5	24	10	20	0.9523	0.0786	16	6	16	0.9580	0.0959	16	5	11	0.9598	0.0965
0.10	4.0	20	8	18	0.9598	0.0803	14	5	14	0.9683	0.0938	13	4	7	0.9639	0.0992
0.05	2.0	103	51	61	0.9608	0.0489	72	34	44	0.9527	0.0443	63	28	36	0.9503	0.0475
0.05	2.5	60	28	38	0.9661	0.0492	41	18	23	0.9542	0.0498	37	15	20	0.9500	0.0475
0.05	3.0	39	17	27	0.9556	0.0482	30	12	22	0.9551	0.0384	28	10	18	0.9634	0.0471
0.05	3.5	34	14	24	0.9616	0.0349	24	9	19	0.9623	0.0372	19	6	11	0.9529	0.0477
0.05	4.0	26	10	20	0.9508	0.0346	20	7	17	0.9639	0.0347	17	5	10	0.9652	0.0447
0.01	2.0	153	74	84	0.9503	0.0096	119	56	66	0.9607	0.0100	98	44	53	0.9535	0.0098
0.01	2.5	89	40	50	0.9509	0.0084	70	30	40	0.9560	0.0073	57	23	30	0.9547	0.0089
0.01	3.0	64	27	37	0.9556	0.0075	48	19	29	0.9569	0.0076	39	14	21	0.9544	0.0088
0.01	3.5	48	19	29	0.9521	0.0075	38	14	24	0.9602	0.0068	30	10	15	0.9564	0.0092
0.01	4.0	40	15	25	0.9561	0.0069	29	10	20	0.9620	0.0090	24	7	13	0.9527	0.0090

Table 16 GMSDS plan parameters for BS distribution when $a = 1.0$, $\nu = 2.0$, $m = 3$

		(m, k) = (3, 3)					(m, k) = (3, 2)					(m, k) = (3, 1)				
β	μ/μ_0	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$	n	c_a	c_r	$P_A(p_1)$	$P_A(p_2)$
0.25	2.0	43	28	38	0.9506	0.2145	29	18	28	0.9520	0.2300	28	16	26	0.9519	0.2473
0.25	2.5	27	17	27	0.9609	0.2133	17	10	17	0.9598	0.2447	17	9	13	0.9532	0.2192
0.25	3.0	18	11	18	0.9624	0.2311	14	8	14	0.9755	0.2435	12	6	9	0.9594	0.2433
0.25	3.5	15	9	15	0.9722	0.2360	11	6	11	0.9763	0.2373	9	4	7	0.9561	0.2278
0.25	4.0	12	7	12	0.9718	0.2393	8	4	8	0.9625	0.2216	9	4	7	0.9753	0.2278
0.10	2.0	72	46	56	0.9545	0.0890	47	29	35	0.9513	0.0998	45	26	34	0.9511	0.0946
0.10	2.5	41	25	35	0.9520	0.0888	29	17	21	0.9568	0.0992	28	15	21	0.9606	0.0906
0.10	3.0	29	17	27	0.9540	0.0886	20	11	20	0.9602	0.0985	20	10	15	0.9657	0.0941
0.10	3.5	23	13	23	0.9562	0.0851	17	9	17	0.9693	0.0918	15	7	11	0.9644	0.0975
0.10	4.0	20	11	20	0.9636	0.0817	14	7	14	0.9649	0.0823	12	5	10	0.9574	0.0974
0.05	2.0	90	57	67	0.9541	0.0500	65	40	46	0.9517	0.0499	57	33	42	0.9504	0.0490
0.05	2.5	53	32	42	0.9551	0.0481	38	22	28	0.9596	0.0495	34	18	26	0.9515	0.0484
0.05	3.0	38	22	32	0.9569	0.0459	29	16	26	0.9694	0.0458	24	12	17	0.9586	0.0486
0.05	3.5	32	18	28	0.9683	0.0431	21	11	15	0.9615	0.0489	19	9	13	0.9603	0.0482
0.05	4.0	24	13	23	0.9625	0.0495	18	9	18	0.9681	0.0458	16	7	11	0.9589	0.0405
0.01	2.0	141	88	98	0.9544	0.0098	107	65	75	0.9545	0.0089	89	52	62	0.9561	0.0095
0.01	2.5	83	49	59	0.9513	0.0083	63	36	46	0.9616	0.0090	52	28	36	0.9575	0.0094
0.01	3.0	60	34	44	0.9602	0.0085	43	23	33	0.9509	0.0082	36	18	24	0.9513	0.0086
0.01	3.5	46	25	35	0.9595	0.0086	35	18	28	0.9649	0.0085	28	13	19	0.9541	0.0088
0.01	4.0	38	20	30	0.9647	0.0092	27	13	23	0.9501	0.0082	23	10	16	0.9577	0.0098

4 Industrial Application of the GMDSS Plan

In this section, to illustrate the proposed sampling plan, consider a dataset contains the 19 times in minutes to oil breakdown of an insulating fluid under high-test voltage (34 kV). The reliability application of Burr XII was studied by Zimmer et al. [28] for this data whereas this dataset was originally given by Nelson [29]. Lio et al. [24] discussed the application of this data for the sampling plan when the lifetime of the product follows the Burr XII. They have also tested Kolmogorov–Smirnov test to examine the goodness-of- fit of this data for Burr XII and showed that reasonable well fitted to Burr XII with maximum likelihood estimates of shape parameters are $b = 0.85$ and $c = 5.49$ respectively. We have also constructed tables for our proposed GMDSS plan at these estimated shape parametric values and are reported in Tables 5 and 6. The breakdown times of insulating fluids at 34 kV, which are as follows.

0.19, 0.78, 0.96, 0.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71 and 72.89.

To demonstrate the proposed sampling plan, assume that the specified mean time to breakdown of an insulating fluid is 2 min (i.e., $\mu_0 = 2$). The experiment time t_0 is 1 min. Thus we get the termination ratio of life test is $a = 0.5$. Also, assume that producer's risk α is 5% and consumer's risk β is 25% and the mean ratio is used as $\mu/\mu_0 = 4$. From Table 5, the optimal parameters of the GMDSS plan are given as $n = 12$, $c_a = 2$, $c_r = 4$, $m = 3$ and $k = 2$ for the aforementioned specified values. Thus the proposed GMDSS plan can be implemented as follows:

From the current lot (L_3), 12 electrical insulating fluids are selected randomly and put on life test. The breakdown times of each fluid are recorded. Suppose that the breakdown times of 12 insulating fluids are as follows:

CL : 0.07, 0.23, 0.32, 1.19, 1.54, 1.88, 2.12, 2.19, 5.47, 5.80, 5.96 and 8.39.

Let d denoted as the number of failed insulating fluids. From simulated data, it is noticed that three insulating fluids broke down before 1 min of time (i.e. d = 3). Hence, the current lot (CL) is accepted in a condition that the preceding tow lots out of three lots were accepted with at most two failed item (since $m = 3$, $k = 2$). Since $m = 3$, consider the preceding three lot and accept the current lot (CL) if two lots out of three lots were at most two fluids will fail before 1 min of time. Suppose that the breakdown times of 12 insulating fluids which are selected from preceding lot (L_3) are as follows.

L_3 : 0.36, 0.62, 1.09, 1.47, 1.70, 1.96, 2.02, 2.49, 2.93, 3.45, 8.38 and 13.48.

From the above data, it is observed that there are two electrical insulating fluids that broke down before 1 min and it is equal to c_a (i.e. $d = c_a$). Therefore, the preceding one lot satisfied the condition out of three lots. Now at this stage, at

least one more preceding lot accepts then the CL will be accepted. Suppose that the breakdown times of 12 insulating fluids which are selected from preceding lot (L_2) are as follows.

L_2 : 0.25, 0.75, 1.29, 1.36, 1.51, 1.59, 1.76, 1.81, 2.96, 3.88, 4.45 and 6.12.

From this data, it is noticed that there are two electrical insulating fluids that broke down before 1 min and it is equal to c_a (i.e. $d = c_a$). Therefore, one of the preceding lots satisfied the condition of acceptance out of three lots. Hence we stop checking the lots and decided that the CL is accepted since out of the preceding three lots two lots were accepted. This is an advantage of our proposed GMDSS plan, instead of testing preceding three lots we are testing only preceding two lots (Table 7).

5 Simulation

In this subsection, a simulated data is used to demonstrate the execution of the proposed plan. We explain the implementation of the proposed GMDSS plan using the simulated data for assuring that the mean life of the product when the lifetime of the product follows Burr XII distribution. In this work, we generate a data set which follows Burr XII distribution. To execute this plan, we assume the following circumstances. The shape parameters of Burr XII distribution are assumed to be $b = 2$ and $c = 2$. Consider the experiment termination time t_0 is 1000 min and specified mean life μ_0 is 1000 min. Thus, the experiment mean ratio becomes $a = 1.0$. The producer's risk and consumer's risk are respectively considered to be $\alpha = 5\%$ and $\beta = 25\%$ and the ratio of true mean life and specified mean life is $\mu/\mu_0 = 2$.

For the aforesaid particular values, from Table 8 we get the optimal parameters of the GMDSS plan as $n = 8$, $c_a = 1$, $c_r = 4$, $m = 3$ and $k = 1$. The GMDSS plan can be applied as follows:

A random sample of size 8 is selected from the CL. Conduct life test experiment for the sample items for 1000 min. Suppose that the failure times of sample items of current lot recorded are as follows:

CL : 106, 566, 982, 1037, 1568, 1651, 1941 and 2156.

From these data, we can observe that there are three sample items that failed before the experiment time, 1000 min. Hence, the current lot is accepted in a condition that the preceding one lots out of three lots were accepted with at most one failed item (since $m = 3, k = 1$). Select a random sample of eight items from the preceding lot (say L_3) and put for life testing these samples for 1000 min. Suppose that the recorded failure times of the preceding lot sample times are as follows:

L_3 : 626, 1008, 1290, 1541, 1792, 2176, 2919 and 3102.

Accept the lot L_3 since it contains one failed items (i.e. $d = c_a$). Hence our conditions satisfied, so stop checking and conclude that the current lot is accepted since k = 1 means out of three preceding lot at least one lot accepted then the current lot is accepted in this sampling plan. Therefore, our proposed GMDSS plan saves checking of two preceding lots, that is an advantage of saving cost and time(Tables 9, 10, 11, 12, 13, 14, 15, and 16).

6 Comparative Study

In this section, the proposed sampling plan is compared with the existing MDSS plan. A plan is considered as better if it has a smaller sample size. The proposed plan has a smaller sample size than the existing sampling plans for various plan parameters for three distributions as shown in Table 17. We compared the different distributions for proposed and existing sampling plan for shape parameter is 2 and the termination ratio of the experiment is considered as $a = 1.0$. It is observed that from Table 17, the proposed plan has a smaller sample size as compared to the existing sampling plans. As an illustration, if $\alpha = 0.05$, $\beta = 0.25$ and $\mu/\mu_0 = 2$ for m = 3, the sample size from the proposed sampling plan for Burr XII is 5 whereas

Table 17 Sample size for proposed and existing MDSS plans when m = 3 and shape parameter is 2 for three distributions

		Burr XII		Gamma		BS	
β	μ/μ_0	MDSS	Proposed	MDSS	Proposed	MDSS	Proposed
0.25	2.0	7	5	8	6	43	29
0.25	2.5	5	5	6	4	27	17
0.25	3.0	5	3	4	2	18	14
0.25	3.5	5	3	4	2	15	11
0.25	4.0	5	3	4	2	12	8
0.10	2.0	14	9	14	10	72	47
0.10	2.5	9	7	10	8	41	29
0.10	3.0	7	7	8	5	29	20
0.10	3.5	7	4	5	5	23	17
0.10	4.0	7	4	5	3	20	14
0.05	2.0	18	13	17	13	90	65
0.05	2.5	11	8	13	9	53	38
0.05	3.0	11	8	9	6	38	29
0.05	3.5	8	8	9	6	32	21
0.05	4.0	8	5	6	6	24	18
0.01	2.0	27	19	26	20	141	107
0.01	2.5	19	14	18	13	83	63
0.01	3.0	14	11	13	11	60	43
0.01	3.5	14	11	11	8	46	35
0.01	4.0	11	11	11	8	38	27

Table 18 Sample size of the GMDSS plan for three lifetime distributions with shape parameter 2 and $a = 1.0$

β	μ/μ_0	BS	Burr XII	Gamma
0.25	2.0	29	5	6
0.25	2.5	17	5	4
0.25	3.0	14	3	2
0.25	3.5	11	3	2
0.25	4.0	8	3	2
0.10	2.0	47	9	10
0.10	2.5	29	7	8
0.10	3.0	20	7	5
0.10	3.5	17	4	5
0.10	4.0	14	4	3
0.05	2.0	65	13	13
0.05	2.5	38	8	9
0.05	3.0	29	8	6
0.05	3.5	21	8	6
0.05	4.0	18	5	6
0.01	2.0	107	19	20
0.01	2.5	63	14	13
0.01	3.0	43	11	11
0.01	3.5	35	11	8
0.01	4.0	27	11	8

it is 7 for the MDSS plan, proposed sampling plan for gamma distribution is 6 whereas it is 8 for the MDSS plan and proposed sampling plan for BS distribution is 29 whereas it is 43 for the MDSS plan. We know that if the sample size is small then that sampling plan shows better performance. Thus, the proposed plan is performed better than existing plans in terms of sample size.

In addition, we made a comparison of the sample size of the proposed GMSDS plan for different distributions considered in our study, namely gamma distribution, Burr XII distribution and BS distribution. To compare the different distributions, we obtain the optimal parameters of the proposed sampling plan for the specified parametric values as follows. The shape parameter is 2 and the termination ratio of the experiment is considered as $a = 1.0$. The optimal parameters are determined so that both the producer and consumer risks are satisfied with a minimum average of ASN at AQL and ASN at LQL. We consider the ratio of true mean life to the specified mean life, μ/μ_0 values from 2 to 4. The sample size of the GMDSS plan for three- lifetime distributions are presented in Table 18. From this table, we can notice that the sample size for the proposed plan for BS distribution shows a larger sample size than compared with the other two distribution considered in this study. Furthermore, gamma and Burr XII distributions have a small sample size and performing well as compared with BS distribution. Also, the sample size obtained for gamma distribution is more cases smaller than the corresponding sample size obtained for the Burr XII distribution.

7 Conclusions

We presented the designing of the GMDSS for the three popular distributions in this chapter. The performance measures, algorithm, real example and the simulation study is added. The comparative study showed that the proposed plan significantly reduced the sample for the inspection of the product as compared to the existing sampling plans. The inspection of the product using the proposed sampling plan will be helpful in reducing the inspection cost in the industry. As we presented sampling plans for three different lifetime distributions that are the Gamma, Burr type XII and BS distributions. The selection of the three distributions can be done on the nature of the data. The practitioners need to fit the data for the three distributions first and choose that distribution which is the best fit to the data. After the selection of the distribution, they can apply the proposed sampling plan for the inspection of the product. If the data is well fitted to three distribution, the practitioners can choose the gamma distribution among the three distribution as sample size under the gamma distribution are smaller than the Burr type XII distribution and the BS distribution. The proposed plan can be applied in any electronic industry for the testing of the product. The proposed sampling plan using some cost model can be extended for future research.

References

1. Rosaiah, K., & Kantam, R. (2005). Acceptance sampling based on the inverse Rayleigh distribution. *Economic Quality Control, 20*(2), 277–286.
2. Chen, J., Li, K.-H., & Lam, Y. (2007). Bayesian single and double variable sampling plans for the Weibull distribution with censoring. *European Journal of Operational Research, 177*(2), 1062–1073.
3. Vijayaraghavan, R., Rajagopal, K., & Loganathan, A. (2008). A procedure for selection of a gamma-Poisson single sampling plan by attributes. *Journal of Applied Statistics, 35*(2), 149–160.
4. Lio, Y., Tsai, T.-R., & Wu, S.-J. (2009). Acceptance sampling plans from truncated life tests based on the Birnbaum–Saunders distribution for percentiles. *Communications in Statistics-Simulation and Computation, 39*(1), 119–136.
5. Rao, G. S. (2009). A group acceptance sampling plans based on truncated life tests for Marshall-Olkin extended Lomax distribution. *Electronic Journal of Applied Statistical Analysis, 3*(1), 18–27.
6. Aslam, M., Jun, C.-H., & Ahmad, M. (2011). New acceptance sampling plans based on life tests for Birnbaum–Saunders distributions. *Journal of Statistical Computation and Simulation, 81*(4), 461–470.
7. Al-Nasser, A. D., & Al-Omari, A. I. (2013). Acceptance sampling plan based on truncated life tests for exponentiated Frechet distribution. *Journal of Statistics and Management Systems, 16*(1), 13–24.
8. Aslam, M., Balamurali, S., Jun, C.-H., & Meer, A. (2017). Time-truncated attribute sampling plans using EWMA for Weibull and Burr type X distributions. *Communications in Statistics-Simulation and Computation, 46*(6), 4173–4184.

9. Al-Omari, A. I. (2018). The transmuted generalized inverse Weibull distribution in acceptance sampling plans based on life tests. *Transactions of the Institute of Measurement and Control, 40*(16), 4432–4443.
10. Aslam, M., Mahmood, Y., Lio, Y., Tsai, T.-R., & Khan, M. A. (2012). Double acceptance sampling plans for Burr type XII distribution percentiles under the truncated life test. *Journal of the Operational Research Society, 63*(7), 1010–1017.
11. Aslam, M., Niaki, S., Rasool, M., & Fallahnezhad, M. (2012). Decision rule of repetitive acceptance sampling plans assuring percentile life. *Scientia Iranica, 19*(3), 879–884.
12. Ramasamy, A. S., & Sutharani, R. (2013). Designing double acceptance sampling plans based on truncated life tests in rayleigh distribution using minimum angle method. *American Journal of Mathematics and Statistics, 3*(4), 227–236.
13. Gui, W., & Xu, M. (2015). Double acceptance sampling plan based on truncated life tests for half exponential power distribution. *Statistical Methodology, 27*, 123–131.
14. Al-Omari, A. I., & Zamanzade, E. (2017). Double acceptance sampling plan for time truncated life tests based on transmuted generalized inverse Weibull distribution. *Journal of Statistics Applications and Probability, 6*, 1–6.
15. Gui, W., & Lu, X. (2018). Double acceptance sampling plan based on the Burr type X distribution under truncated life tests. *International Journal of Industrial and Systems Engineering, 28*(3), 319–330.
16. Yen, C.-H., Chang, C.-H., & Aslam, M. (2015). Repetitive variable acceptance sampling plan for one-sided specification. *Journal of Statistical Computation and Simulation, 85*(6), 1102–1116.
17. Balamurali, S., Jeyadurga, P., & Usha, M. (2018). Optimal design of repetitive group sampling plans for Weibull and gamma distributions with applications and comparison to the Birnbaum–Saunders distribution. *Journal of Applied Statistics, 45*(14), 1–22.
18. Saminathan, B., & Mahalingam, U. (2018). A new mixed repetitive group sampling plan based on the process capability index for product acceptance. *International Journal of Quality & Reliability Management, 35*(2), 463–480.
19. Balamurali, S., & Jun, C.-H. (2007). Multiple dependent state sampling plans for lot acceptance based on measurement data. *European Journal of Operational Research, 180*(3), 1221–1230.
20. Yan, A., Liu, S., & Dong, X. (2016). Designing a multiple dependent state sampling plan based on the coefficient of variation. *Springerplus, 5*(1), 1447.
21. Balamurali, S., Jeyadurga, P., & Usha, M. (2017). Designing of multiple deferred state sampling plan for generalized inverted exponential distribution. *Sequential Analysis, 36*(1), 76–86.
22. Balamurali, S., Jeyadurga, P., & Usha, M. (2017). Optimal designing of multiple deferred state sampling plan for assuring percentile life under Weibull distribution. *The International Journal of Advanced Manufacturing Technology, 93*(9–12), 3095–3109.
23. Wu, C.-W., Lee, A. H., & Chang Chien, C.-C. (2017). A variables multiple dependent state sampling plan based on a one-sided capability index. *Quality Engineering, 29*(4), 719–729.
24. Lio, Y., Tsai, T.-R., & Wu, S.-J. (2010). Acceptance sampling plans from truncated life tests based on the Burr type XII percentiles. *Journal of the Chinese Institute of Industrial Engineers, 27*(4), 270–280.
25. Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics, 13*(2), 215–232.
26. Birnbaum, Z. W., & Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability, 6*(2), 319–327.
27. Lemonte, A. J., Cribari-Neto, F., & Vasconcellos, K. L. (2007). Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis, 51*(9), 4656–4681.
28. Zimmer, W. J., Keats, J. B., & Wang, F. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology, 30*(4), 386–394.
29. Nelson, W. B. (2005). *Applied life data analysis* (Vol. 577). Hoboken: Wiley.

Decision Theoretic Sampling Plan for One-Parameter Exponential Distribution Under Type-I and Type-I Hybrid Censoring Schemes

Deepak Prajapati, Sharmistha Mitra, and Debasis Kundu

Abstract In this paper, we design a decision theoretic sampling plan (DSP) based on Type-I and Type-I hybrid censored lifetime data from a one-parameter exponential distribution. The Bayes estimator of the mean lifetime is used to define a decision function. A suitable loss function is considered to derive the Bayes risk of this DSP. A finite algorithm is provided to obtain the optimum DSP and the corresponding Bayes risk. It has been observed numerically that the optimum DSP is better than the sampling plan proposed by Lam (Ann Stat 22:696–711, 1994) and Lin et al. (Commun Stat Simul Comput 37:1101–1116, 2008; Commun Stat Simul Comput 39:1499–1505, 2010) and it is as good as the Bayesian sampling plan (BSP) of Lin et al. (Ann Inst Stat Math 54:100–113, 2002) and Liang and Yang (J Stat Comput Simul 83: 922–940, 2013). It is observed that the Bayes risk of the optimum DSP is approximately equal to the Bayes risk of the BSP. In case of higher degree polynomials and for a non-polynomial loss function the DSP can be obtained without any additional effort as compared to BSP.

1 Introduction

In reliability life testing experiments, manufacturers usually choose a suitable acceptance sampling plan and do inspection to make decision on the reliability of the experimental items in the batch. If we can find the optimal acceptance sampling plan then we make a better decision on batch of items. So acceptance sampling plan plays an important role in reliability life testing or quality control. In the vast literature of acceptance sampling plans, various approaches have been discussed. The approach based on decision theory has become more popular because here the sampling plan is obtained by making an optimal decision on some economic considerations such as maximizing the return or minimizing the loss. Many researchers have widely

D. Prajapati · S. Mitra · D. Kundu (✉)
Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, India
e-mail: deepakdp@iitk.ac.in; smitra@iitk.ac.in; kundu@iitk.ac.in

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_8

used this approach to obtain the optimal sampling plan. See, for example, Hald [6], Fertig and Mann [5], Lam [10–12], Lam and Choy [13], Lin et al. [18], Huang and Lin [8, 9], Chen et al. [3], Lin et al. [16, 17], Liang and Yang [14] and the references cited therein.

In life testing experiments, usually we collect a censored or incomplete data. The experimenter does not observe the failure of all items because inspection cost increases with time, items may be expensive, etc. So, to save time or resources we try to get the optimal result based on censored or incomplete data. If the inspection cost increases with time, then we put n items on test and terminate the test at a preassigned time τ. This type of censoring is called Type-I censoring. In Type-I hybrid censoring, the experiment is terminated at time of rth failure or at time τ, which ever occurs first. It is useful when items are expensive and the inspection cost also increases with time. In this paper it is assumed that the lifetimes of the experimental units follow one parameter exponential distribution. If M is the number of failures in the experiment, then in both the censoring schemes for $M = 0$, the maximum likelihood (ML) estimator of the mean lifetime does not exist. Lam [12] has obtained the Bayesian sampling plan for Type-I censoring where decision function is based on an estimator which is equal to the ML estimator of the mean lifetime when $M \geq 1$ and at $M = 0$, the estimator is equal to $n\tau$. Lin et al. [18] has observed that the loss function used by Lam [12] does not involve any cost on duration of the experiment. So if we extend the duration of the experiment, we can observe complete sample to get a better decision on the batch of items. Therefore, the sampling plan of Lam [12] is " neither optimal, nor Bayes ". Lin et al.[16, 17] proposed optimal sampling plans for Type-I and Type-I hybrid censoring schemes using a decision function based on the ML estimator of the mean lifetime conditioning on the event that $M \geq 1$.

In this study, we consider the case where life tests are Type-I and Type-I hybrid censored. To avoid the non-existence of the ML estimator of the mean lifetime, we use its Bayes estimator which exists for all values of M. We find the decision theoretic sampling plan (DSP) using the Bayes estimator of the mean lifetime in the decision function. A loss function which includes the sampling cost, the cost on per unit duration of the experiment, the salvage value and the decision loss are used to determine the optimal sampling plan by minimizing the Bayes risk. Numerical results for quadratic loss function show that the optimum DSP is a better plan than the sampling plans proposed by Lam [12] and Lin et al. [16, 17], and as good as the Bayesian sampling plan (BSP) given by Lin et al. [18] and Liang and Yang [14]. Theoretically, for fifth or higher degree polynomial and for a non-polynomial loss function, the optimum DSP is quite easy to derive but BSP cannot be obtained very conveniently.

The rest of the paper is organized as follows. In Sect. 1.1 we introduce the decision function based on the Bayes estimator of the mean lifetime. All necessary theoretical results and algorithm to obtain the optimum DSP are provided in Sect. 2. In Sect. 3 we compare the proposed DSP and the BSP for higher degree polynomial and for a non-polynomial loss function. Numerical comparisons and results on

DSP are given in Sect. 4. Finally we conclude the paper in Sect. 5. All necessary derivations are provided in the Appendix.

1.1 Model Formulation, Assumptions and DSP

Let $X_1, X_2, \ldots, X_n$ denote the lifetime of n items put to test in an experiment. It is assumed that these lifetimes of components are mutually independent and follow the exponential distribution with probability density function (PDF)

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0, \ \lambda > 0 \\ 0, & otherwise. \end{cases} \tag{1}$$

Let θ denote the mean lifetime of an item and for the above distribution, the mean lifetime of an item is $\theta = \frac{1}{\lambda}$. So use of λ is equivalent to the use of θ. Henceforth, we will use whichever is more convenient. Let $X_{(1)} < \ldots < X_{(n)}$ be the ordered observations from the given sample of n lifetimes $X_1, \ldots, X_n$. For Type-I and Type-I hybrid censoring schemes, if τ^* denotes the duration of the experiment, then $\tau^* = \tau$ in a Type-I censoring and $\tau^* = min\{X_{(r)}, \tau\}$ in a Type-I hybrid censoring schemes. Here τ denotes the pre-fixed time and r is a pre-fixed integer, $1 \leq r \leq n$. We observe that τ^* is fixed in Type-I censoring and random for Type-I hybrid censoring. Define $M = max\{i : X_{(i)} \leq \tau^*\}$, i.e., M is the number of failures among the n items put on the life test, before time τ^*. For Type-I censoring the joint PDF of $\mathbf{X} = (X_{(1)}, X_{(2)}, \ldots, X_{(M)})$ is as follows:

$$f_{X_{(1)},\ldots,X_{(M)}}(x_1, \ldots, x_m, M = m|\lambda) = \frac{n!}{(n-m)!} \ \lambda^m e^{-\lambda\left[\sum_{i=1}^{m} x_i + (n-m)\tau\right]}; \\ 0 < x_1 < \ldots < x_m < \infty, \tag{2}$$

and for Type-I hybrid censoring the joint PDF is given as follows:

$$f_{X_{(1)},\ldots,X_{(M)}}(x_1, \ldots, x_m, M = m|\lambda) = \begin{cases} \frac{n!}{(n-m)!}\lambda^m e^{-\lambda\left[\sum_{i=1}^{m} x_i + (n-m)\tau\right]}, & \text{if } x_r \geq \tau \\ \frac{n!}{(n-r)!}\lambda^r e^{-\lambda\left[\sum_{i=1}^{r} x_i + (n-r)x_r\right]}, & \text{if } x_r < \tau. \end{cases} \tag{3}$$

Lam [12] proposed the Bayesian sampling plan for Type-I censoring using a decision function

$$\delta(\mathbf{X}) = \begin{cases} 1, & \text{if } \widehat{\theta} \geq \xi \\ 0, & \text{if } \widehat{\theta} < \xi, \end{cases} \tag{4}$$

where $\xi(>0)$ denotes the minimum acceptable surviving time to take a decision on a batch, i.e., whether to accept it with an action 1 or reject it with an action 0. Estimator $\widehat{\theta}$ is given by

$$\widehat{\theta} = \begin{cases} \widehat{\theta}_M, & \text{if } M \geq 1 \\ n\tau, & \text{if } M = 0, \end{cases}$$

where $\widehat{\theta}_M$ is the ML estimator of the mean lifetime θ and is defined as:

$$\widehat{\theta}_M = \begin{cases} \frac{\sum_{i=1}^{M} X_{(i)} + (n-M)\tau}{M}, & \text{if } M \geq 1 \\ \text{does not exist}, & \text{if } M = 0. \end{cases}$$

Similarly, for Type-I hybrid censoring also the ML estimator of the mean lifetime does not exist for $M = 0$. Therefore, in place of ML estimator, we propose to use the Bayes estimator of θ in decision function because it exists for all values of M. The Bayes estimator with respect to the squared error loss function (see Berger[2]) when λ has a prior distribution $G(a, b)$, with PDF

$$\pi(\lambda; a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b}; \quad \lambda > 0, \quad a, b > 0, \tag{5}$$

is as follows:

1. *Type-I Censoring:*

$$\widehat{\theta}_B = \frac{b + \sum_{i=1}^{M} X_{(i)} + (n-M)\tau}{M + a},$$

2. *Type-I Hybrid Censoring:*

$$\widehat{\theta}_B = \begin{cases} \frac{b + \sum_{i=1}^{M} X_{(i)} + (n-M)\tau}{M+a}, & \text{if } X_{(r)} \geq \tau \\ \frac{b + \sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}}{r+a}, & \text{if } X_{(r)} < \tau. \end{cases}$$

Based on the observed sample, we define our decision function as:

$$\delta(\mathbf{X}) = \begin{cases} 1, & \text{if } \hat{\theta}_B \geq \xi \\ 0, & \text{if } \hat{\theta}_B < \xi, \end{cases} \tag{6}$$

where $\hat{\theta}_B$ is the Bayes estimator of θ, $\xi(>0)$ is the minimum acceptable surviving time to take a decision on a batch, i.e., whether to accept it with an action 1 or reject it with an action 0. We consider the loss function:

$$L(\delta(\mathbf{X}), \lambda) = \begin{cases} nC_s - (n-M)r_s + \tau^* C_\tau + g(\lambda), & \text{if } \delta(\mathbf{x}) = 1 \\ nC_s - (n-M)r_s + \tau^* C_\tau + C_r, & \text{if } \delta(\mathbf{x}) = 0, \end{cases} \tag{7}$$

which depends on various costs. C_r is the cost due to rejecting the batch, C_s is the cost due to inspection of each item and C_τ is the cost per unit duration of the experiment. Cost due to accepting the batch is denoted by $g(\lambda)$, so that $g(\lambda)$ has to be positive and increasing in λ. In general, $g(\lambda)$ will be determined by the inspection requirement or experience before designing a sampling plan. Hence, the form of $g(\lambda)$ can vary and accordingly, form of the loss function also varies. If an item does not fail, then this item can be reused with a salvage value r_s such that $C_s > r_s \geq 0$, see, Liang and Yang [14] or Liang et al. [15].

Any sampling plan consists of the sampling parameters and the decision parameter ξ. Therefore, (n, τ, ξ) and (n, r, τ, ξ) are the sampling plans for Type-I censoring and Type-I hybrid censoring, respectively, and we denote these by DSP. To determine the optimum DSP namely, (n_0, τ_0, ξ_0) and $(n_0, r_0, \tau_0, \xi_0)$ for Type-I censoring and Type-I hybrid censoring, respectively, we determine that decision function in (6) which minimizes the Bayes risk of the DSP under the given loss function in (7) among all possible DSPs.

2 Computation of Bayes Risk and Optimum DSP

In this section we compute the Bayes risk of the DSP and an algorithm is presented for obtaining the optimum DSP. First, we derive the general expression of the Bayes risk for any given sampling plan and the distribution of $\hat{\theta}_B$. In Sects. 2.1 and 2.3 we derive the exact expressions of the Bayes risks for Type-I and Type-I hybrid censoring schemes.

2.1 Computation of Bayes Risk and Distribution of $\hat{\theta}_B$

A number of authors used a quadratic loss function to obtain the Bayesian sampling plans (for example see, Lam [11, 12], Lam and Choy [13], Lin et al. [18], Huang and Lin [8, 9], Lin et al. [16, 17], Liang and Yang [14], Liang et al. [15], etc). They use this functional form because computation becomes easier and $g(\lambda) = a_0 + a_1\lambda + a_2\lambda^2$ is a reasonable approximation of the true acceptance cost. However, a higher degree polynomial may be a better approximation of the true acceptance cost. So we consider the functional form of the loss function defined as

$$L(\delta(\mathbf{X}), \lambda) = \begin{cases} nC_s - (n-M)r_s + \tau^* C_\tau + a_0 + a_1\lambda + \ldots + a_k\lambda^k, & \text{if } \delta(\mathbf{X}) = 1 \\ nC_s - (n-M)r_s + \tau^* C_\tau + C_r, & \text{if } \delta(\mathbf{X}) = 0, \end{cases} \tag{8}$$

where $a_0, a_1, \ldots$ and a_k are fixed positive constants. Based on the decision rule in (6) and the loss function in (8), the Bayes risk takes the form

$$\begin{aligned} R_B(\mathbf{S}) &= E\big\{L(\delta(\mathbf{X}), \lambda)\big\} = E_\lambda E_{X/\lambda}\big\{L(\delta(\mathbf{X}), \lambda)\big\} \\ &= E_\lambda E_{X/\lambda}\big\{nC_s - (n-M)r_s + \tau^* C_\tau + a_0 + a_1\lambda + \ldots + a_k\lambda^k \\ &\quad +(1-\delta(\mathbf{X}))(C_r - a_1\lambda - \ldots - a_k\lambda^k)\big\} \\ &= n(C_s - r_s) + r_s E(M) + C_\tau E(\tau^*) + a_0 + a_1\mu_1 + \ldots + a_k\mu_k \\ &\quad +E_\lambda\big\{(C_r - a_0 - a_1\lambda - \ldots - a_k\lambda^k)P(\hat{\theta}_B < \xi \mid \lambda)\big\}, \end{aligned} \tag{9}$$

where $\mathbf{S}$ is the sampling plan under Type-I or Type-I hybrid censoring and $\mu_i = E(\lambda^i)$ for $i = 1, 2, \ldots, k$. Therefore, in order to derive the Bayes risk of the DSP from (9), it is clear that we need the distribution of $\hat{\theta}_B$ given λ for Type-I and Type-I hybrid censoring schemes.

So, first we compute the distribution of $\hat{\theta}_B$ for a given λ, which is

$$\begin{aligned} P(\widehat{\theta}_B \le x|\lambda) &= P(M=0|\lambda)P(\widehat{\theta}_B \le x|M=0,\lambda) \\ &\quad +P(M \ge 1|\lambda)P(\widehat{\theta}_B \le x|M \ge 1,\lambda) \\ &= pS_\lambda(x) + (1-p)H_\lambda(x), \end{aligned} \tag{10}$$

where $p = P(M=0|\lambda) = e^{-n\lambda\tau}$ and

$$S_\lambda(x) = P(\widehat{\theta}_B \le x|M=0,\lambda) = \begin{cases} 1, & \text{if } x \ge \frac{b+n\tau}{a} \\ 0, & otherwise. \end{cases}$$

$$H_\lambda(x) = P(\widehat{\theta}_B \le x|M \ge 1,\lambda) = \begin{cases} \int_0^x h(u|\lambda)du, & \text{if } 0 < x < \frac{b+n\tau}{a} \\ 0, & otherwise, \end{cases}$$

where $h(u|\lambda)$ is the PDF of the absolutely continuous part of $\widehat{\theta}_B$ given λ.

2.2 Type- I Censoring

When the samples are Type-I censored, then as mentioned above, evaluation of the Bayes risk of DSP requires the PDF $h(.|\lambda)$. The following lemma provides the PDF of the absolutely continuous part of the distribution.

Lemma 1 *The PDF $h(.|\lambda)$ of $\widehat{\theta}_B$ given λ under a Type-I censoring scheme is given by*

$$h(x|\lambda) = \frac{1}{1-p}\sum_{m=1}^{n}\sum_{j=0}^{m}\binom{n}{m}\binom{m}{j}(-1)^j e^{-\lambda(n-m+j)\tau}\pi\big(x - \tau_{j,m,a,b}; m, (m+a)\lambda\big),$$

for $\tau_{j,m,a,b} < x < \frac{b+n\tau}{a}$, where $\tau_{j,m,a,b} = \frac{b+(n-m+j)\tau}{m+a}$ and $\pi(\cdot)$ is as defined in (5).

Proof The PDF can be easily obtained using the result of Bartholomew [1]. □

Hence, using the distribution function of $\hat{\theta}_B$ and the prior density in (5) of λ, the following theorem provides the Bayes risk of the DSP (n, τ, ξ) for the loss function in (8).

Theorem 1 *The Bayes risk of the DSP* (n, τ, ξ) *w.r.t the loss function in (8) is given as*

$$R_B(n,\tau,\xi) = n(C_s - r_s) + r_s E(M) + \tau C_\tau + a_0 + a_1\mu_1 + \ldots + a_k\mu_k + \sum_{l=0}^{k} C_l \frac{b^a}{\Gamma(a)}$$
$$\times\left[\frac{\Gamma(a+l)I_{(\frac{b+n\tau}{a}<\xi)}}{(b+n\tau)^{(a+l)}} + \sum_{m=1}^{n}\sum_{j=0}^{m}(-1)^j\binom{n}{m}\binom{m}{j}\frac{\Gamma(a+l)I_{S^*_{j,m,a,b}}(m,a+l)}{((m+a)\tau_{j,m,a,b})^{a+l}}\right],$$

where $I_{(\frac{b+n\tau}{a}<\xi)}$ *is an indicator function and the exact expressions for* $E(M)$, C_l *and* $I_{S^*_{j,m,a,b}}(m, a+l)$ *are provided in the proof of the theorem.*

Proof See Appendix. □

Based on the explicit expression of the Bayes risk $R_B(n, \tau, \xi)$, an optimum DSP (n_0, τ_0, ξ_0) can be determined by

$$R_B(n_0, \tau_0, \xi_0) = \min_{n}\{\min_{\tau,\xi}[R_B(n, \tau, \xi)]\}. \tag{11}$$

The explicit expression of the Bayes risk is very complicated and hence, it is not possible to obtain optimal values of n, τ and ξ analytically. Lam [12] has given a grid search algorithm for obtaining an approximate optimal sampling plan. Following that approach, we present a similar algorithm to find the optimum DSP:

Algorithm A

1. Fix n and τ, minimize $R_B(n, \tau, \xi)$ with respect to ξ using grid search method and denote the minimum Bayes risk by $R_B(n, \tau, \xi_0(n, \tau))$.
2. For fixed n, minimize $R_B(n, \tau, \xi_0(n, \tau))$ with respect to τ, using grid search method and denote the minimum Bayes risk by $R_B(n, \tau_0(n), \xi_0(n, \tau_0(n)))$.
3. Choose sample size n_0 such that

$$R_B(n_0, \tau_0(n_0), \xi_0(n_0, \tau_0(n_0))) \leq R_B(n, \tau_0(n), \xi_0(n, \tau_0(n))) \quad \forall \quad n \geq 0.$$

The choice of the number of grid points for τ and ξ depends on the Bayes risk function $R_B(n, \tau, \xi)$. If $R_B(n, \tau, \xi)$ has a unique minimum then the algorithm works very well to find the optimum DSP. In relation to Algorithm A, the following

theorem can significantly reduce the effort in the search for the optimum DSP and also ensures that it will be obtained in a finite number of search steps.

Theorem 2 *Assuming* $0 < \xi \leq \xi^*$, *let us denote* $R_B(n, \tau, \xi') = min_{0<\xi\leq\xi^*} R_B(n, \tau, \xi)$ *for some fixed* n (≥ 1) *and* τ. *Further, let* n_0 *and* τ_0 *be the optimal sample size and censoring time. Then,*

$$n_0 \leq \min\left\{\frac{C_r}{C_s - r_s}, \frac{a_0 + a_1\mu_1 + \ldots + a_k\mu_k}{C_s - r_s}, \frac{R_B(n,\tau,\xi')}{C_s - r_s}\right\},$$

$$\tau_0 \leq \min\left\{\frac{C_r}{C_\tau}, \frac{a_0 + a_1\mu_1 + \ldots + a_k\mu_k}{C_\tau}, \frac{R_B(n,\tau,\xi')}{C_\tau}\right\}.$$

Proof See Appendix. □

2.3 Type-I Hybrid Censoring

When the random sample is coming from a Type-I hybrid censoring scheme, the PDF of the absolutely continuous part of the Bayes estimator $\hat{\theta}_B$ given λ is $h(.|\lambda)$, which is given in the following lemma.

Lemma 2 *The PDF* $h(.|\lambda)$ *of* $\widehat{\theta}_B$ *given* λ *under a Type-I hybrid censoring scheme is given by*

$$h(x|\lambda) = \frac{1}{1-p}\Bigg[\sum_{m=1}^{r-1}\sum_{j=0}^{m}\binom{n}{m}\binom{m}{j}(-1)^j e^{-\lambda\tau(n-m+j)}\pi\left(x - \tau_{j,m,a,b}; m, (m+a)\lambda\right)$$

$$+\pi\left(x - \frac{b}{r+a}; r, (r+a)\lambda\right) + r\binom{n}{r}\sum_{j=1}^{r}\binom{r-1}{j-1}\frac{(-1)^j e^{-\lambda\tau(n-r+j)}}{n-r+j}$$

$$\times\pi\left(x - \tau_{j,r,a,b}; r, (r+a)\lambda\right)\Bigg],$$

for $0 < x < \frac{b}{a} + \frac{n\tau}{a}$, *where* $\tau_{j,m,a,b} = \frac{b}{m+a} + \frac{(n-m+j)\tau}{m+a}$ *and* $\pi(\cdot)$ *is as defined in (5).*

Proof It can be easily obtained using the result of Childs et al. [4]. □

Then in the expression (9), the distribution of $\hat{\theta}_B$ given λ and prior density (5) of λ is used to obtain the Bayes risk of DSP (n, r, τ, ξ). The following theorem provides the Bayes risk of DSP.

Theorem 3 *The Bayes risk of DSP* (n, r, τ, ξ) *w.r.t the loss function (8) is given as follows*

$$R_B(n,r,\tau,\xi) = n(C_s - r_s) + E(M)r_s + E(\tau^*)C_\tau + a_0 + a_1\mu_1 + \ldots + a_k\mu_k + \sum_{l=0}^{k} C_l \frac{b^a}{\Gamma(a)} \left\{ \frac{\Gamma(a+l)}{(b+n\tau)^{(a+l)}} I_{(\frac{b+n\tau}{a} < \xi)} + \sum_{m=1}^{r-1} \sum_{j=0}^{m} \binom{n}{m} \binom{m}{j} (-1)^j R_{l,j,m} + R_{l,r-n,r} + \sum_{j=1}^{r} \binom{n}{r} \binom{r-1}{j-1} (-1)^j \frac{r}{(n-r+j)} R_{l,j,r} \right\},$$

where $I_{(\frac{b+n\tau}{a} < \xi)}$ is an indicator function as defined before, and the expressions for $E(M)$, $E(\tau^*)$, C_l and $R_{l,j,m}$ are provided in the proof of the theorem.

Proof See Appendix. □

Based on the explicit expression of the Bayes risk, an optimum DSP $(n_0, r_0, \tau_0, \xi_0)$ can be determined by

$$R_B(n_0, r_0, \tau_0, \xi_0) = \min_n \{\min_{r \le n} \{\min_{\tau,\xi} [R_B(n, r, \tau, \xi)]\}\}. \tag{12}$$

In this case also the Bayes risk expression is very complicated. So we present a similar algorithm as given in Sect. 2.2 to obtain the optimum DSP $(n_0, r_0, \tau_0, \xi_0)$.

Algorithm B

1. Fix n, r and τ, minimize $R_B(n, r, \tau, \xi)$ with respect to ξ using a grid search method. Denote the minimum Bayes risk by $R_B(n, r, \tau, \xi_0(n, r, \tau))$.
2. For fixed n and r, minimize $R_B(n, r, \tau, \xi_0(n, r, \tau))$ with respect to τ using the grid search method and denote the minimum Bayes risk by

$$R_B(n, r, \tau_0(n, r), \xi_0(n, r, \tau_0(n, r))).$$

3. For fixed n, choose $r \le n$ for which $R_B(n, r, \tau_0(n, r), \xi_0(n, r, \tau_0(n, r)))$ is minimum and denote it by

$$R_B(n, r_0(n), \tau_0(n, r_0(n)), \xi_0(n, r_0(n), \tau_0(n, r_0(n)))).$$

4. Choose sample size n_0 such that

$$R_B(n_0, r_0(n_0), \tau_0(n_0, r_0(n_0)), \xi_0(n_0, r_0(n_0), \tau_0(n_0, r_0(n_0)))) \le R_B(n, r_0(n), \tau_0(n, r_0(n)), \xi_0(n, r_0(n), \tau_0(n, r_0(n)))) \quad \forall \quad n \ge 0.$$

For each sample size n and for given values of r and ξ, Bayes risk $R_B(n, r, \tau, \xi)$ is a function of τ. If we have to find an optimum DSP, we need an upper bound of τ.

Tsai et al. [19] suggested to choose suitable range of τ, say $[0, \tau_\alpha]$, where τ_α is such that

$$P(0 < X < \tau_\alpha) = 1 - \alpha$$
$$\int_0^\infty \int_0^{\tau_\alpha} \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b} \lambda e^{-\lambda x} dx d\lambda = 1 - \alpha$$
$$1 - \left(1 + \frac{\tau_\alpha}{b}\right)^{-a} = 1 - \alpha,$$

hence

$$\tau_\alpha = b\left(\alpha^{-\frac{1}{a}} - 1\right), \tag{13}$$

where α is a preassigned number satisfying $0 < \alpha < 1$. The choice of α depends on the prescribed precision. Higher the precision required, smaller the value of α should be chosen. In the range $[0, \tau_\alpha]$ we use grid search to find the optimal value of τ. The next theorem shows that Algorithm B is finite, i.e., we can find an optimal sampling plan in a finite number of search steps.

Theorem 4 *Assuming* $0 < \xi \leq \xi^*$, *let us denote* $R_B(n, r, \tau, \xi') = min_{\xi \leq \xi^*} R_B(n, r, \tau, \xi)$ *for some fixed* n (≥ 1) *and* τ. *Further, let* n_0 *be the optimal sample size. Then,*

$$n_0 \leq \min\left\{\frac{C_r}{C_s - r_s}, \frac{a_0 + a_1\mu_1 + \ldots + a_k\mu_k}{C_s - r_s}, \frac{R_B(n, r, \tau, \xi')}{C_s - r_s}\right\}$$

and $r_0 \leq n_0$.

Proof Proof is similar to Theorem 2. □

3 Comparisons with Existing Bayesian Sampling Plan

Lin et al. [18] proposed the BSP (n_B, τ_B, δ_B) for Type-I censoring and Liang and Yang [14] proposed BSP $(n_B, r_B, \tau_B, \delta_B)$ for Type-I hybrid censoring using a quadratic loss function. In this section we show that for more general loss functions, the optimum DSP can be obtained without any additional effort as compared to BSP.

3.1 Higher Degree Polynomial Loss Function

In Sect. 3 we have shown that for a higher degree polynomial loss function, Bayes risk of the DSP can be easily computed for Type-I and Type-I hybrid censoring. In

this section we show that the Bayes risk of BSP for higher degree polynomial loss function cannot be expressed in explicit form. The optimal sampling plans obtained are "approximate optimal" plans if $g(\lambda)$, cost due to acceptance, is an approximation of the true acceptance cost. So it is necessary to have a better approximation of the true cost for better results. Therefore, it is meaningful to study what happens when $g(\lambda)$ is a higher degree polynomial loss function. Bayes decision function (see, Lin et al. [18] and Liang and Yang [14]) for BSP is given as

$$\delta_B(\mathbf{x}) = \begin{cases} 1, & \text{if } \phi_\pi(m, z) \leq C_r \\ 0, & otherwise, \end{cases}$$

where for Type-I censoring

$$z = \sum_{i=1}^{m} x_i + (n-m)\tau,$$

and for Type-I hybrid censoring

$$z = \begin{cases} \sum_{i=1}^{m} x_i + (n-m)\tau, & \text{if } m = 1, 2, \ldots, r-1 \\ \sum_{i=1}^{r} x_i + (n-r)x_r, & \text{if } m = r, \end{cases}$$

with

$$\phi_\pi(m, z) = \int_0^\infty g(\lambda)\pi(\lambda|m, z)d\lambda.$$

Since the prior distribution of λ is $G(a, b)$, it is well known that the posterior distribution of λ is also gamma, viz.,

$$\pi(\lambda|m, z) \sim G(m + a, z + b).$$

Now when $g(\lambda) = a_0 + a_1\lambda + \ldots + a_k\lambda^k$ in the loss function (7), then,

$$\phi_\pi(m, z) = \int_0^\infty g(\lambda)\pi(\lambda|m, z)d\lambda = a_0 + \sum_{j=1}^{k} a_j \frac{(m+a)\ldots(m+a+j-1)}{(z+b)^j}.$$

Thus to find the closed form of the decision function we need to obtain the set

$$A = \{z;\ z \geq 0, \phi_\pi(m, z) \leq C_r\},$$

and to construct A, we need to obtain the set of $z \geq 0$, such that

$$h_1(z) = a_0 + \sum_{j=1}^{k} a_j \frac{(m+a)\ldots(m+a+j-1)}{(z+b)^j} \leq C_r, \tag{14}$$

which is equivalent to finding $z \geq 0$, such that,

$$h_2(z) = (C_r - a_0)\big(z+b\big)^k - \sum_{j=1}^{k} a_j (m+a)\ldots(m+a+j-1)\big(z+b\big)^{k-j} \geq 0. \tag{15}$$

Note that for $k = 2$,

$$h_2(z) = (C_r - a_0)\big(z+b\big)^2 - a_1(m+a)\big(z+b\big) - a_2(m+a)(m+a+1) \geq 0.$$

It can easily be shown that if $D_n(m)$ is the only real root or $D_n(m)$ is the maximum real root of $h_2(z) = 0$, then the Bayes decision function will take the following form:

$$\delta_B(\mathbf{x}) = \begin{cases} 1, & \text{if } z \geq D_n(m) - b \\ 0, & otherwise. \end{cases} \tag{16}$$

It is well known that there is no algebraic solution to polynomial equations of degree five or higher (see chapter 5, Herstein [7]). So closed form of the Bayes risk of BSP cannot be determined for five or higher degree polynomial loss functions. This difficulty does not arise in computation of the proposed optimum DSP, because the decision function does not depend on the form of loss function. So for higher degree polynomial loss functions, applying the DSP is easier than the BSP.

3.2 Non-polynomial Loss Function

In Sect. 3.1 we discussed the limitations of BSP for polynomial loss function where DSP can be obtained quite easily. In this section, we discuss that if we have a non-polynomial loss function, then obtaining DSP is easier as compared to BSP. To illustrate this, we consider a very simple non-polynomial loss function

$$L(\delta(\mathbf{X}), \lambda) = \begin{cases} nC_s - (n-M)r_s + \tau * C_\tau + a_0 + a_1\lambda + a_2\lambda^{5/2}, & \text{if } \delta(\mathbf{X}) = 1 \\ nC_s - (n-M)r_s + \tau * C_\tau + C_r, & \text{if } \delta(\mathbf{X}) = 0, \end{cases} \tag{17}$$

where $g(\lambda) = a_0 + a_1\lambda + a_2\lambda^{5/2}$, which is an increasing function in λ. The Bayes risk of DSP for the non-polynomial loss function (17) is computed by a similar approach as in Sect. 2 and is given by:

For Type -I censoring

$$R_B(n,\tau,\xi) = n(C_s - r_s) + r_s E(M) + \tau C_\tau + a_0 + a_1\mu_1 + a_2\frac{\Gamma(a+\frac{5}{2})}{\Gamma(a)b^{\frac{5}{2}}} + \sum_{l=0}^{2} C_l \frac{b^a}{\Gamma(a)}$$
$$\times\left[\frac{\Gamma(a+p_l)I_{(\frac{b+n\tau}{a}<\xi)}}{(b+n\tau)^{(a+p_l)}} + \sum_{m=1}^{n}\sum_{j=0}^{m}(-1)^j\binom{n}{m}\binom{m}{j}\frac{\Gamma(a+p_l)I_{S^*_{j,m,a,b}}(m,a+p_l)}{((m+a)\tau_{j,m,a,b})^{a+p_l}}\right],$$

where the expression of $E(M)$ and $I_{S^*_{j,m,a,b}}(m,a+p_l)$ are given in the proof of the Theorem 1 in the Appendix. For Type -I hybrid censoring,

$$R_B(n,r,\tau,\xi) = n(C_s - r_s) + E(M)r_s + E(\tau^*)C_\tau + a_0 + a_1\mu_1 + a_2\frac{\Gamma(a+\frac{5}{2})}{\Gamma(a)b^{\frac{5}{2}}}$$
$$+\sum_{l=0}^{2} C_l\frac{b^a}{\Gamma(a)}\left[\frac{\Gamma(a+p_l)}{(b+n\tau)^{(a+p_l)}}I_{(\frac{b+n\tau}{a}<\xi)} + \sum_{m=1}^{r-1}\sum_{j=0}^{m}\binom{m}{n}\binom{m}{j}(-1)^j R_{p_l,j,m}\right.$$
$$\left.+R_{p_l,r-n,r} + \sum_{k=1}^{r}\binom{n}{r}\binom{r-1}{k-1}(-1)^k\frac{r}{(n-r+k)}R_{p_l,k,r}\right],$$

where the expressions for $E(M)$, $E(\tau^*)$, C_l and $R_{p_l,j,m}$ are defined in the proof of the Theorem 3 in the Appendix and

$$p_l = \begin{cases} 0, & \text{if } l = 0 \\ 1, & \text{if } l = 1 \\ \frac{5}{2}, & \text{if } l = 2. \end{cases}$$

Now, in case of BSP, when $g(\lambda) = a_0 + a_1\lambda + a_2\lambda^{5/2}$, then by Sect. 3.1

$$\phi_\pi(m,z) = \int_0^\infty g(\lambda)\pi(\lambda|m,z)d\lambda = a_0 + \frac{a_1(m+a)}{(z+b)} + \frac{a_2\Gamma(m+a+\frac{5}{2})}{\Gamma(m+a)(z+b)^{\frac{5}{2}}}.$$

So to find a closed form of the decision function we need to obtain the set A as defined in Sect. 3.1. Note that to construct A, we need to obtain the set of $z \geq 0$, such that

$$h_1(z) = a_0 + \frac{a_1(m+a)}{(z+b)} + \frac{a_2\Gamma(m+a+\frac{5}{2})}{\Gamma(m+a)(z+b)^{\frac{5}{2}}} \leq C_r,$$

which is equivalent to finding $z \geq 0$, such that,

$$h_2(z) = (C_r - a_0)\Gamma(m+a)\big(z+b\big)^{\frac{5}{2}} - a_1(m+a)\Gamma(m+a)\big(z+b\big)^{\frac{3}{2}} - a_2\Gamma(m+a+\frac{5}{2}) \geq 0.$$

It is obvious that finding the closed form solution of the non-polynomial equation $h_2(z) = 0$ is not possible. So, for the given very simple non-polynomial loss function the Bayes risk of BSP cannot be obtained analytically but the Bayes risk of DSP is obtained quite easily.

Hence, we see that in general BSP cannot be obtained easily for all type of functional forms of the loss function but we can obtain the DSP for such cases quite easily because we have to compute just the expected value of the loss function.

4 Numerical Results

In this section, we focus on comparing the optimum DSP with the sampling plans of Lam [12], Lin et al.[16, 17] and with the BSP proposed by Lin et al. [18] and Liang and Yang [14]. We also obtain the optimum DSP for fifth degree polynomial loss function and for non-polynomial loss function proposed in Sect. 3.2. Let n_1^*, τ_b, ξ^* from Theorem 2, denote the upper bound for optimal values of n, τ and ξ, under Type-I censoring, i.e.,

$$0 \leq n_0 \leq n_1^*, \quad 0 \leq \tau_0 \leq \tau_b \quad and \quad 0 < \xi_0 \leq \xi^*.$$

Since τ and ξ are continuous and n is an integer, therefore, to obtain numerical results we take a grid size of 0.0125 for τ and 0.0015 for ξ. Then by applying the Algorithm A, we obtain the optimum DSP (n_0, τ_0, ξ_0) for Type-I censoring.

Similarly, let n_2^* and ξ^* from Theorem 4 denote the upper bound for optimal value of n and ξ under Type-I hybrid censoring, i.e.,

$$0 \leq n_0 \leq n_2^*, \quad r_0 \leq n_0, \quad 0 \leq \xi_0 \leq \xi^*$$

and optimal values of $\tau \in [0, \tau_\alpha]$ where τ_α is determined by (13) for given value of a and b. Here we use $\alpha = 0.01$. To obtain numerical results we take a grid size of 0.0015 for ξ and for τ we take the grid size of 0.0125. Then we apply Algorithm B to obtain the optimum DSP for Type-I hybrid censoring. The results presented in Tables 1, 2, 3, 4, 5, 6, and 7 are obtained by using a program written in *R*.

Table 1 Numerical comparison with sampling plan of Lam [12] and Lin et al. [16, 17] for different values of a and b

Scheme	a	b	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0	a	b	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0
DSP	2.5	0.8	24.8419	4	1.3125	0.3255	1.5	0.8	16.5825	3	0.7000	0.3330
LAM			24.9367	3	0.7077	0.3539			16.6233	3	0.5262	0.2631
Lin et al.[16, 17]			24.9893	4	0.6808	0.3404			16.7533	3	0.5262	0.2631
DSP	2.5	1.0	21.7081	4	1.1125	0.3255	2.0	0.8	21.1398	4	1.1500	0.3270
LAM			21.7640	3	0.5483	0.2742			21.2153	3	0.6051	0.3026
Lin et al.[16, 17]			21.8515	4	0.5819	0.2910			21.2875	4	0.6051	0.3026
DSP	3.0	0.8	27.5581	3	1.1625	0.3270	2.5	0.6	27.7267	3	1.2125	0.3285
LAM			27.6136	3	0.8170	0.4085			27.7834	3	0.8537	0.4268
Lin et al.[16, 17]			27.6521	3	0.8170	0.4085			29.8193	3	0.8537	0.4268
DSP	3.5	0.8	29.2789	2	1.0125	0.3285	10.0	3.0	29.5166	2	0.7875	0.3165
LAM			29.2789	2	1.0037	0.5019			29.5166	2	0.7928	0.3964
Lin et al.[16, 17]			29.3642	2	1.0037	0.5019			29.5959	2	0.8194	0.4097

Table 2 Numerical comparison between DSP and BSP for Type-I censoring

		BSP	DSP					BSP	DSP			
a	b	$R_B(n_B, \tau_B, \delta_B)$	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0	C_s	$R_B(n_B, \tau_B, \delta_B)$	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0
2.5	0.8	25.2777	25.2777	3	0.7250	0.3285	0.3	24.4279	24.4282	5	0.7750	0.3240
2.5	1.0	22.0361	22.0361	3	0.5625	0.3285	0.7	25.8777	25.8777	3	0.7250	0.3285
3.5	0.8	29.7131	29.7131	2	0.8125	0.3285	2.0	27.9535	27.9535	1	0.3750	0.4095
C_t							C_r					
0.1		24.9446	24.9446	4	0.9625	0.3255	20	19.3292	19.3293	2	0.8750	0.4275
0.7		25.4197	25.4197	3	0.7000	0.3285	50	32.2081	32.2085	5	0.5625	0.2400
2.0		26.1439	26.1439	4	0.3875	0.3285	100	35.5937	35.5937	0	0.0000	0.0000

Table 3 Numerical comparison between DSP and BSP for Type-I hybrid censoring

		BSP[a]	DSP						BSP[a]	DSP				
a	b	$R_B(n_B, r_B, \tau_B, \delta_B)$	$R_B(n_0, r_0, \tau_0, \xi_0)$	n_0	r_0	τ_0	ξ_0	C_s	$R_B(n_B, r_B, \tau_B, \delta_B)$	$R_B(n_0, r_0, \tau_0, \xi_0)$	n_0	r_0	τ_0	ξ_0
2.5	1.0	21.6762	21.6772	4	4	0.5625	0.3270	0.5	24.9678	24.9625	4	4	0.7500	0.3255
2.5	0.8	24.9678	24.9625	4	4	0.7500	0.3255	0.7	25.5932	25.5953	3	3	0.7250	0.3285
3.0	0.8	27.7339	27.7370	3	3	0.8750	0.3270	2.0	27.7681	27.7681	1	1	0.3750	0.3975
C_t								C_r						
0.1		24.7344	24.7312	4	4	0.8375	0.3255	20	19.0392	19.0383	2	2	0.9375	0.4275
0.7		25.0531	25.0533	5	4	0.4500	0.3255	30	24.9678	24.9625	4	4	0.7500	0.3255
2.0		25.4595	25.4552	4	3	0.3750	0.3285	50	31.7324	31.7326	6	6	0.3625	0.2385

[a]Bayes risk of BSP is obtained by simulation

Table 4 Minimum Bayes risk and corresponding optimum DSP for fifth degree polynomial loss function for Type-I censoring

a	b	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0	C_s	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0
1.5	0.8	12.7301	3	1.1875	0.9750	0.4	24.2273	7	1.6750	0.8525
3.0	2.5	24.8613	6	1.7125	0.8650	0.7	25.8920	4	1.9875	0.9000
3.0	3.0	22.4152	6	1.5750	0.8650	1.0	26.9562	3	2.1125	0.9125
C_τ						C_r				
0.3		24.4827	6	2.1125	0.8625	30	24.8613	6	1.7125	0.8650
1.0		25.6103	6	1.3500	0.8750	50	33.0079	8	1.7375	0.7300
1.5		26.2238	6	1.0750	0.8850	100	45.7463	12	1.6000	0.5900

4.1 Numerical Results for Quadratic Loss Function

In Sect. 2 we have obtained the Bayes risk of DSP for a kth order polynomial loss function. For comparison with Lam [12], Lin et al. [16, 17], Lin et al. [18] and Liang and Yang [14] we assume $k = 2$. They choose this value of k because a quadratic polynomial makes computation easy and straightforward.

For comparison with Lam [12] and Lin et al. [16, 17] we fix the coefficients as: $a_0 = 2,\ a_1 = 2,\ a_2 = 2,\ C_s = 0.5,\ C_\tau = 0, r_s = 0$ and $C_r = 30$, which they have used to obtain the sampling plan and we take $\xi^* = 3$. In Table 1, comparison is shown by varying the hyper-parameters a and b and keeping others fixed. From Table 1 it is clear that Bayes risk of the optimum DSP is less than or equal to the Bayes risk of the sampling plan of Lam [12] and Lin et al.[16, 17].

For comparison with the BSP proposed by Lin et al. [18] and Liang and Yang [14] we choose the same set of parameter values and coefficients used by Lin et al. [18] and Liang and Yang [14], which are as follows, for Type-I censoring, $a = 2.5,\ b = 0.8,\ a_0 = 2,\ a_1 = 2,\ a_2 = 2,\ C_s = 0.5,\ r_s = 0,\ C_\tau = 0.5,\ C_r = 30,\ \xi^* = 3$ and for Type-I hybrid censoring, $a = 2.5,\ b = 0.8,\ a_0 = 2,\ a_1 = 2,\ a_2 = 2,\ C_s = 0.5,\ r_s = 0.3,\ C_\tau = 0.5,\ C_r = 30,\ \xi^* = 3$. Comparison of the Bayes risk of the BSP and the optimum DSP is given in Tables 2 and 3 by varying a, b, C_s, C_τ and C_r one at a time and keeping others fixed.

In Table 2, $R_B(n_B, \tau_B, \delta_B)$ denotes the Bayes risk of the BSP and $R_B(n_0, \tau_0, \xi_0)$ denotes the Bayes risk of the optimum DSP (n_0, τ_0, ξ_0). Similarly, in Table 3 the Bayes risk of the BSP is denoted by $R_B(n_B, r_B, \tau_B, \delta_B)$ and $R_B(n_0, r_0, \tau_0, \xi_0)$ denotes the Bayes risk of the optimum DSP $(n_0, r_0, \tau_0, \xi_0)$. In case of Type-I hybrid censoring, the Bayes risk of BSP includes a complicated integral which is computed by simulation techniques. So the Bayes risk for BSP in case of Type-I hybrid censoring is an approximation of the exact Bayes risk. From Tables 2 and 3 it is clear that performance of the optimum DSP is as good as the BSP.

In Type-I censoring, the Bayes risk of the DSP is a function of the sampling plan (n, τ, ξ) in which n takes discrete values and others are continuous. Theorem 2 states that the optimal value of n is bounded above, so it is sufficient to provide the

Table 5 Minimum Bayes risk and corresponding optimum DSP for fifth degree polynomial loss function for Type-I hybrid censoring

a	b	$R_B(n_0, r_0, \tau_0, \xi_0)$	n_0	r_0	τ_0	ξ_0	C_s	$R_B(n_0, r_0, \tau_0, \xi_0)$	n_0	r_0	τ_0	ξ_0
1.5	0.8	26.2979	5	4	1.6375	0.9300	0.4	23.4093	10	7	0.9500	0.8450
3.0	2.5	24.2369	7	6	1.4125	0.8600	0.7	25.4201	5	4	1.3625	0.8900
3.0	3.0	20.7241	7	6	1.1875	0.8625	1.5	27.6478	2	2	1.9500	0.9475
C_τ							C_r					
0.1		23.7217	6	6	2.4125	0.8575	20	18.6112	4	3	1.4625	1.0275
1.0		24.6729	7	5	1.0250	0.8725	50	32.2362	10	9	1.2125	0.7225
1.5		25.0037	8	5	0.8125	0.8725	100	44.6304	14	13	1.0125	0.5875

Table 6 Minimum Bayes risk and corresponding optimum DSP for non-polynomial loss function for Type-I censoring

a	b	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0	C_s	$R_B(n_0, \tau_0, \xi_0)$	n_0	τ_0	ξ_0
2.5	0.8	27.5536	4	1.0625	0.4185	0.3	26.6463	6	0.9250	0.4110
2.5	1.0	24.9524	4	1.0000	0.4185	0.7	28.1649	3	1.0875	0.4245
3.0	0.8	29.6926	2	1.0750	0.4275	2.0	29.9411	1	0.7250	0.6000
C_t						C_r				
0.1		27.0722	4	1.3875	0.4170	30	27.5536	4	1.0625	0.4185
0.7		27.7546	4	0.9500	0.4200	50	38.4987	6	0.9250	0.3225
2.0		28.6323	4	0.5625	0.4230	100	54.6498	9	0.7250	0.2325

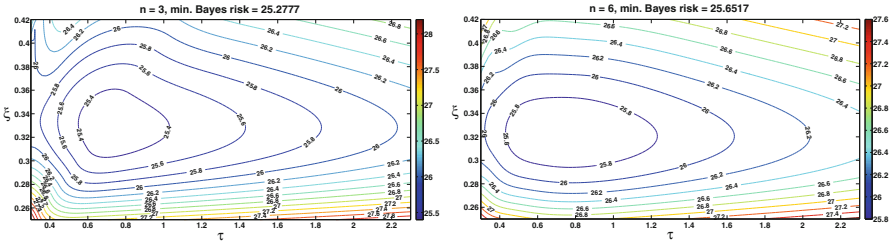


Fig. 1 Contour plot of the Bayes risk with set of coefficients $a_0 = 2, a_1 = 2, a_2 = 2, C_s = 0.5, r_s = 0, C_\tau = 0.5, C_r = 30$, and $a = 2.5, b = 0.8$ for Type-I censoring

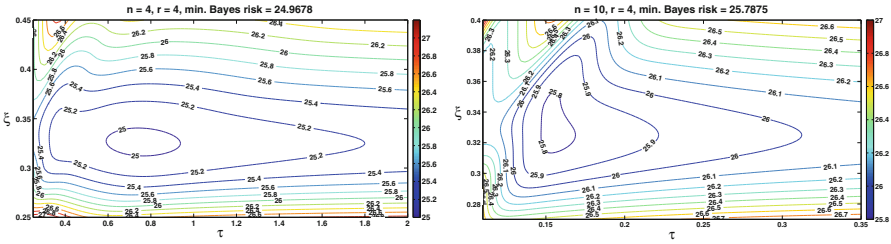


Fig. 2 Contour plot of the Bayes risk with set of coefficients $a_0 = 2, a_1 = 2, a_2 = 2, C_s = 0.5, r_s = 0.3, C_\tau = 0.5, C_r = 30$, and $a = 2.5, b = 0.8$ for Type-I hybrid censoring

contour plot of the Bayes risk with respect to τ and ξ in Fig. 1 for different values of n to show that it has a unique minimum. We provide the plots for which the minimum Bayes risk is $R_B(3, 0.7250, 0.3285) = 25.2777$. It is clear from the plots that Bayes risk first decreases then increases as n increases which ensures that the Bayes risk has a unique minimum w.r.t n, τ and ξ.

For Type-I hybrid censoring, the Bayes risk is a function of the sampling plan (n, r, τ, ξ) in which n and r take discrete values and others are continuous. Since optimal values of n and r are bounded above (Theorem 4), so for different values of n and r, we provide the contour plot of Bayes risk with respect to τ and ξ in Fig. 2. We provide the plots for which the minimum Bayes risk is

Table 7 Minimum Bayes risk and corresponding optimum DSP for non-polynomial loss function for Type-I hybrid censoring

a	b	$R_B(n_0, r_0, \tau_0, \xi_0)$	n_0	r_0	τ_0	ξ_0	C_s	$R_B(n_0, r_0, \tau_0, \xi_0)$	n_0	r_0	τ_0	ξ_0
2.5	1.0	24.5869	5	4	0.6000	0.4170	0.5	27.2090	4	4	1.2125	0.4170
2.5	0.8	27.2090	4	4	1.2125	0.4170	0.7	27.8214	3	3	1.1750	0.4230
3.0	0.8	29.3249	2	2	1.2000	0.4275	2.0	29.6794	1	1	0.7500	0.4530
C_t							C_r					
0.1		26.9124	4	4	1.3375	0.4170	20	19.7253	1	1	1.0750	0.5685
0.7		27.3031	5	4	0.6875	0.4170	30	27.2090	4	4	1.2125	0.4170
2.0		27.7698	5	3	0.4125	0.4230	50	38.1061	7	6	0.5750	0.3225

$R_B(4, 4, 0.7500, 0.3255) = 24.9678$. In this case also the Bayes risk has unique minimum and as n increases the Bayes risk first decreases then increases. The contour plot can also be used for predicting the range which includes the optimal values of τ and ξ by which we can reduce significantly the effort in the search for the optimum plan.

4.2 Numerical Results for Fifth Degree Polynomial and Non-polynomial Loss Function

For a better approximation of the optimal sampling plan we assume that the loss function is of order $k = 5$ in (8). To obtain the optimum DSP, the following parameter and coefficient values are used: $a = 3.0$, $b = 2.5$, $a_0 = 2$, $a_1 = 2$, $a_2 = 2$, $a_3 = 2$, $a_4 = 2$, $a_5 = 2$, $C_s = 0.5$, $r_s = 0$, $C_\tau = 0.5$, $C_r = 30$, $\xi^* = 3$ for Type-I censoring. For Type-I hybrid censoring, we use following sets of coefficients: $a = 3.0$, $b = 2.5$, $a_0 = 2$, $a_1 = 2$, $a_2 = 2$, $a_3 = 2$, $a_4 = 2$, $a_5 = 2$, $C_s = 0.5$, $r_s = 0.3$, $C_\tau = 0.5$, $C_r = 30$, $\xi^* = 3$. Then varying the parameters a and b or one coefficient out of C_s, C_τ and C_r at a time, and setting others fixed, we obtain the minimum Bayes risk and the optimum DSP which is given in Tables 4 and 5. We observe in Tables 4 and 5 that as costs C_s, C_τ and C_r increase, the minimum Bayes risk increases. It is also observed that as cost per unit inspection C_s increases the optimal value of n_0 and r_0 decreases and the optimal values of τ_0 and ξ_0 increases. If the rejection cost C_r increases then n_0 and r_0 increases and the optimal values of τ_0 and ξ_0 decreases. Therefore, for fifth degree polynomial loss function results presented in Tables 4 and 5 have similar behaviour with those obtained from the quadratic loss function and they are quite acceptable in terms of n_0, r_0, τ_0 and ξ_0.

To obtain the optimum DSP for non-polynomial loss function given in Sect. 3.2, we use the following set of parameter values and coefficients for Type-I censoring: $a = 2.5$, $b = 0.8$, $a_0 = 2$, $a_1 = 2$, $a_2 = 2$, $C_r = 30$, $C_s = 0.5$, $r_s = 0$, $C_\tau = 0.5$, $\xi^* = 3$. Similarly, for Type-I hybrid censoring: $a = 2.5$, $b = 0.8$, $a_0 = 2$, $a_1 = 2$, $a_2 = 2$, $C_r = 30$, $C_s = 0.5$, $r_s = 0.3$, $C_\tau = 0.5$, $\xi^* = 3$, i.e., the set of parameter values and coefficients are used. The minimum Bayes risk and optimum DSP are obtained in Tables 6 and 7 by varying parameters a and b or one coefficient out of C_s, C_τ and C_r at a time and setting others fixed. It is observed from Tables 6 and 7 that as costs C_s, C_τ and C_r increase, the minimum Bayes risk increases. The optimal values of n_0 and r_0 decrease when C_s increases and when C_r increases, then n_0 and r_0 increase. The value of ξ_0 increases when cost C_s increases and decreases when cost C_r increases. So, for given non-polynomial loss function also, the results in Tables 6 and 7 are quite reasonable and acceptable.

5 Conclusion

In this paper, we have shown that a sampling plan can be obtained using the Bayes estimator of the mean lifetime θ. This estimator always exists for both Type-I and Type-I hybrid censoring. We have developed a methodology for finding the DSP using a decision function which is based on the Bayes estimator of θ. We propose an algorithm to find the optimum DSP. It is shown numerically that the optimum DSP is better than the sampling plans of Lam [12] and Lin et al. [16, 17] and as good as the BSP proposed by Lin et al. [18] and Liang and Yang [14] for Type-I censoring and Type-I hybrid censoring. Further, we have generalized the existing work for higher degree polynomial loss function and for a specific choice of the non-polynomial loss function, which cannot be handled easily by the BSP. Thus, we see that the DSP is applicable to a wider class of loss functions than the BSP.

Acknowledgements The authors would like to thank the reviewers for their careful review and valuable suggestions which led to the improvement of the presentation of the paper.

Appendix

Proof of Theorem 1

To derive the Bayes risk we consider the decision function (6) and loss function (8) with $\tau^* = \tau$. Then from (9) Bayes risk under Type-I censoring is,

$$R_B(n,\tau,\xi) = n(C_s - r_s) + r_s E(M) + \tau C_\tau + a_0 + a_1\mu_1 + \ldots + a_k\mu_k + \sum_{l=0}^{k} C_l \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} P(\hat{\theta}_B < \xi|\lambda)\, d\lambda, \tag{18}$$

where C_l is defined as

$$C_l = \begin{cases} C_r - a_l, & \text{if} \quad l = 0 \\ -a_l, & \text{if} \quad l = 1, 2, \ldots, k. \end{cases} \tag{19}$$

Using (10) and Lemma 1 we get

$$\int_0^\infty \lambda^{a+l-1} e^{-\lambda b} P(\hat{\theta}_B < \xi|\lambda)\, d\lambda$$
$$= \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} p S_\lambda(\xi)\, d\lambda + \int_0^\infty \lambda^{a+l-1} e^{-\lambda b}(1-p) H_\lambda(\xi)\, d\lambda$$

$$= \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} e^{-n\lambda\tau}\, I_{(\frac{b+n\tau}{a}<\xi)}\, d\lambda + \sum_{m=1}^{n}\sum_{j=0}^{m} \binom{n}{m}\binom{m}{j}(-1)^j$$

$$\times \frac{(m+a)^m}{\Gamma(m)} \int_0^\infty \int_{\tau_{j,m,a,b}}^{\xi} \lambda^{a+l+m-1} e^{-\lambda\{(m+a)x\}} \big(x-\tau_{j,m,a,b}\big)^{m-1} dx\, d\lambda$$

$$= \frac{\Gamma(a+l)}{(b+n\tau)^{(a+l)}} I_{(\frac{b+n\tau}{a}<\xi)} + \sum_{m=1}^{n}\sum_{j=0}^{m} \binom{n}{m}\binom{m}{j}(-1)^j$$

$$\times \frac{\Gamma(a+l+m)}{\Gamma(m)(m+a)^{(a+l)}} \int_0^{\xi-\tau_{j,m,a,b}} \frac{y^{m-1}}{\{\tau_{j,m,a,b}+y\}^{a+l+m}} dy.$$

$$= \frac{\Gamma(a+l)}{(b+n\tau)^{(a+l)}} I_{(\frac{b+n\tau}{a}<\xi)} + \sum_{m=1}^{n}\sum_{j=0}^{m} \binom{n}{m}\binom{m}{j}(-1)^j$$

$$\times \frac{\Gamma(a+l)}{((m+a)\tau_{j,m,a,b})^{(a+l)}} \frac{\Gamma(a+l+m)}{\Gamma(m)\Gamma(a+l)} \int_0^{\frac{\xi-\tau_{j,m,a,b}}{\tau_{j,m,a,b}}} \frac{z^{m-1}}{(1+z)^{a+l+m}} dz.$$

Now taking a transformation $z = u/(1-u)$, we have

$$\int_0^{C^*_{j,m,a,b}} \frac{z^{m-1}}{(1+z)^{a+l+m}} dz = \int_0^{S^*_{j,m,a,b}} u^{m-1}(1-u)^{a+l-1} du = B_{S^*_{j,m,a,b}}(m, a+l),$$

where $C^*_{j,m,a,b} = \frac{\xi-\tau_{j,m,a,b}}{\tau_{j,m,a,b}}$, $S^*_{j,m,a,b} = \frac{C^*_{j,m,a,b}}{1+C^*_{j,m,a,b}}$ and

$$B_x(\alpha,\beta) = \int_0^x u^{\alpha-1}(1-u)^{\beta-1} du, \quad 0 \le x \le 1, \tag{20}$$

is the incomplete beta function. Let us denote the cumulative distribution function of beta as $I_x(\alpha,\beta) = B_x(\alpha,\beta)/B(\alpha,\beta)$. Then Bayes risk is finally obtained as

$$R_B(n,\tau,\xi) = n(C_s - r_s) + r_s E(M) + \tau C_\tau + a_0 + a_1\mu_1 + \ldots + a_k\mu_k + \sum_{l=0}^{k} C_l \frac{b^a}{\Gamma(a)}$$

$$\times \Bigg[\frac{\Gamma(a+l) I_{(\frac{b+n\tau}{a}<\xi)}}{(b+n\tau)^{(a+l)}} + \sum_{m=1}^{n}\sum_{j=0}^{m} (-1)^j \binom{n}{m}\binom{m}{j} \frac{\Gamma(a+l) I_{S^*_{j,m,a,b}}(m, a+l)}{((m+a)\tau_{j,m,a,b})^{a+l}} \Bigg], \tag{21}$$

where

$$E(M) = E_\lambda\{E(M|\lambda)\}.$$

Since for each $m = 1, 2, \ldots, n$, the probability mass function of M given λ is

$$P(M=m|\lambda)=\binom{n}{m}(1-e^{-\lambda\tau})^m(e^{-\lambda\tau})^{n-m}=\sum_{j=0}^{m}\binom{n}{m}\binom{m}{j}(-1)^j e^{-(n-m+j)\lambda\tau},$$

therefore,

$$E(M)=\sum_{m=1}^{n}\sum_{j=0}^{m} m\binom{n}{m}\binom{m}{j}(-1)^j\frac{b^a}{(b+(n-m+j)\tau)^a}.$$

Proof of Theorem 2

Note that the Bayes risk can be written as,

$$R_B(n,\tau,\xi)=n(C_s-r_s)+\tau C_\tau+E(M)r_s \\ +E_\lambda\left\{(a_0+a_1\lambda+\ldots+a_k\lambda^k)P(\hat{\theta}_B\geq\xi|\lambda)+C_rP(\hat{\theta}_B<\xi|\lambda)\right\}.$$

Now we know that $a_0+a_1\lambda+\ldots+a_k\lambda^k\geq 0$ and C_r the rejection cost is non negative. Therefore, if (n_0,τ_0,ξ_0) is the optimal sampling plan then the corresponding Bayes risk is

$$R_B(n_0,\tau_0,\xi_0)\geq n_0(C_s-r_s)+\tau_0C_\tau. \tag{22}$$

Now when $\xi=\infty$ we reject the batch without sampling and the corresponding Bayes risk is given by $R_B(0,0,\infty)=C_r$. When $\xi=0$ we accept the batch without sampling and corresponding Bayes risk is given by $R_B(0,0,0)=a_0+a_1\mu_1+\ldots+a_k\mu_k$. Then the optimal Bayes risk is

$$R_B(n_0,\tau_0,\xi_0)\leq min\left\{R_B(0,0,0),R_B(0,0,\infty),R_B(n,\tau,\xi')\right\}. \tag{23}$$

Hence, from Eqs. (22) and (23) we have

$$n_0(C_s-r_s)+\tau_0C_\tau\leq min\left\{R_B(0,0,0),R_B(0,0,\infty),R_B(n,\tau,\xi')\right\},$$

from where it follows that

$$n_0\leq min\left\{\frac{C_r}{C_s-r_s},\frac{a_0+a_1\mu_1+\ldots+a_k\mu_k}{C_s-r_s},\frac{R_B(n,\tau,\xi')}{C_s-r_s}\right\},$$

$$\tau_0\leq min\left\{\frac{C_r}{C_\tau},\frac{a_0+a_1\mu_1+\ldots+a_k\mu_k}{C_\tau},\frac{R_B(n,\tau,\xi')}{C_\tau}\right\}.$$

Proof of Theorem 3

To derive the Bayes risk for Type-I hybrid censoring, we consider the decision function (6) and loss function (8) with $\tau^* = min\{X_{(r)}, \tau\}$. Then from (9)

$$R_B(n,r,\tau,\xi) = n(C_s - r_s) + E(M)r_s + E(\tau^*)C_\tau + a_0 + a_1\mu_1 + \ldots + a_k\mu_k + \sum_{l=0}^{k} C_l \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} P(\widehat{\theta}_B \le \xi|\lambda) d\lambda, \tag{24}$$

where μ_i and C_l are defined earlier. Let $\xi_* = min\{\frac{b}{a} + \frac{n\tau}{a}, \xi\}$, then define

$$\begin{aligned}
R_{l,j,m} &= \int_0^\infty \int_0^{\xi_*} \lambda^{a+l-1} e^{-\lambda(b+\tau(n-m+j))} \pi\Big(y - \tau_{j,m,a,b}; m, (m+a)\lambda\Big) dy\, d\lambda \\
&= \frac{(m+a)^m}{\Gamma(m)} \int_0^\infty \int_{\tau_{j,m,a,b}}^{\xi_*} \lambda^{a+l+m-1} e^{-\lambda\{(m+a)y\}} \big(y - \tau_{j,m,a,b}\big)^{m-1} dy\, d\lambda \\
&= \frac{(m+a)^m}{\Gamma(m)} \int_0^{\xi_* - \tau_{j,m,a,b}} \frac{v^{m-1}\Gamma(a+l+m)}{\big((m+a)\tau_{j,m,a,b} + (m+a)v\big)^{a+l+m}} dv \\
&= \frac{\Gamma(a+l+m)}{\Gamma(m)(m+a)^{a+l}\tau_{j,m,a,b}^{a+l+m}} \int_0^{\xi_* - \tau_{j,m,a,b}} \frac{v^{m-1}}{\Big(1 + \frac{v}{\tau_{j,m,a,b}}\Big)^{a+l+m}} dv \\
&= \frac{\Gamma(a+l)}{((m+a)\tau_{j,m,a,b})^{a+l}} \frac{\Gamma(a+l+m)}{\Gamma(m)\Gamma(a+l)} \int_0^{\frac{\xi_* - \tau_{j,m,a,b}}{\tau_{j,m,a,b}}} \frac{z^{m-1}}{(1+z)^{a+l+m}} dz.
\end{aligned}$$

Taking similar transformation as in Type-I censoring and using the incomplete beta function, we obtain

$$R_{l,j,m} = \frac{\Gamma(a+l)}{((m+a)\tau_{j,m,a,b})^{a+l}} I_{S^*_{j,m,a,b}}(m, a+l). \tag{25}$$

Using (10), Lemma 2 in (24) and relation (25) we get

$$\begin{aligned}
&\int_0^\infty \lambda^{a+l-1} e^{-\lambda b} P(\widehat{\theta}_B \le \xi|\lambda) d\lambda \\
&= \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} p S_\lambda(\xi)\, d\lambda + \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} (1-p) H_\lambda(\xi)\, d\lambda \\
&= \int_0^\infty \lambda^{a+l-1} e^{-\lambda b} e^{-n\lambda\tau} I_{(\frac{b+n\tau}{a} < \xi)}\, d\lambda + \sum_{m=1}^{r-1} \sum_{j=0}^{m} \binom{n}{m} \binom{m}{j} (-1)^j
\end{aligned}$$

$$\times \int_0^{\infty} \int_0^{\xi_*} \lambda^{a+l-1} e^{-\lambda\{b+\tau(n-m+j)\}} \pi\Big(y - \tau_{j,m,a,b}; m, (m+a)\lambda\Big) dy\, d\lambda$$

$$+ \int_0^{\infty} \int_0^{\xi_*} \lambda^{a+l-1} e^{-\lambda b} \pi\Big(y - \frac{b}{r+a}; r, (r+a)\lambda\Big) dy\, d\lambda + \sum_{j=1}^{r} \binom{n}{r}\binom{r-1}{j-1}$$

$$\times \frac{(-1)^j r}{(n-r+j)} \int_0^{\infty} \int_0^{\xi_*} \lambda^{a+l-1} e^{-\lambda\{b+\tau(n-r+j)\}} \pi\Big(y - \tau_{j,r,a,b}; r, (r+a)\lambda\Big) dy\, d\lambda$$

$$= \frac{\Gamma(a+l)}{(b+n\tau)^{(a+l)}} I_{(\frac{b+n\tau}{a} < \xi)} + \sum_{m=1}^{r-1} \sum_{j=0}^{m} \binom{n}{m}\binom{m}{j} (-1)^j R_{l,j,m} + R_{l,r-n,r}$$

$$+ \sum_{j=1}^{r} \binom{n}{r}\binom{r-1}{j-1} (-1)^j \frac{r}{(n-r+j)} R_{l,j,r}.$$

Thus Bayes risk of DSP (n, r, τ, ξ) under Type-I hybrid censoring is given by

$$R_B(n, r, \tau, \xi) = n(C_s - r_s) + E(M) r_s + E(\tau^*) C_\tau + a_0 + a_1\mu_1 + \ldots + a_k\mu_k$$

$$+ \sum_{l=0}^{k} C_l \frac{b^a}{\Gamma(a)} \Bigg\{ \frac{\Gamma(a+l)}{(b+n\tau)^{(a+l)}} I_{(\frac{b+n\tau}{a} < \xi)} + \sum_{m=1}^{r-1} \sum_{j=0}^{m} \binom{n}{m}\binom{m}{j} (-1)^j R_{l,j,m} + R_{l,r-n,r}$$

$$+ \sum_{j=1}^{r} \binom{n}{r}\binom{r-1}{j-1} (-1)^j \frac{r}{(n-r+j)} R_{l,j,r} \Bigg\}, \tag{26}$$

where

$$E(M) = \sum_{m=1}^{r-1} \sum_{j=0}^{m} m \binom{n}{m}\binom{m}{j} \frac{(-1)^j b^a}{(b+(n-m+j)\tau)^a}$$

$$+ \sum_{i=r}^{n} \sum_{j=0}^{i} r \binom{n}{i}\binom{i}{j} \frac{(-1)^j b^a}{(b+(n-i+j)\tau)^a},$$

$$E(\tau^*) = r \binom{n}{r} \sum_{j=0}^{r-1} \binom{r-1}{j} (-1)^{r-1-j} \Bigg\{ \frac{b}{(n-j)^2(a-1)} - \frac{t b^a}{(n-j)((n-j)\tau+b)^a}$$

$$- \frac{b^a}{(n-j)^2(a-1)((n-j)\tau+b)^{a-1}} \Bigg\} + \sum_{i=r}^{n} \sum_{j=0}^{i} \tau \binom{n}{i}\binom{i}{j} (-1)^j \frac{b^a}{(b+(n-i+j)\tau)^a}.$$

For computation of $E(M)$ and $E(\tau^*)$ see Liang and Yang [14].

References

1. Bartholomew, D. J. (1963). The sampling distribution of an estimate arising in life testing. *Technometrics, 5*, 361–374.
2. Berger, J. O. (1985). *Statistical decision theory and bayesian analysis*. New York: Springer.
3. Chen, J., Chou, W., Wu, H., & Zhou, H. (2004). Designing acceptance sampling schemes for life testing with mixed censoring. *Naval Research Logistics, 51*, 597–612.
4. Childs, A., Chandrasekar, B., Balakrishnan, N., & Kundu, D. (2003). Exact likelihood inference based on Type-I and Type-II hybrid censored samples from the exponential distribution. *Annals of the Institute of Statistical Mathematics, 55*, 319–330.
5. Fertig, K. W., & Mann, N. R. (1974). A decision-theoretic approach to defining variables samplingplans for finite lots: Single sampling for exponential and Gaussian processes. *Journal of the American Statistical Association, 69*, 665–671.
6. Hald, A. (1967). Asymptotic properties of Bayesian single sampling plans. *Journal of the Royal Statistical Society Series B, 29*, 162–173.
7. Herstein, I. N. (1975). *Topics in algebra*. New York: John Wiley & Sons.
8. Huang, W. T., & Lin, Y. P. (2002). An improved Bayesian sampling plan for exponential population with type I censoring. *Communications in Statistics: Theory and Methods, 31*, 2003–2025.
9. Huang, W. T., & Lin, Y. P. (2004). Bayesian sampling plans for exponential distribution based on uniform random censored data. *Computational Statistics and Data Analysis, 44*, 669–691.
10. Lam, Y. (1988). Bayesian approach to single variable sampling plans. *Biometrika, 75*, 387–391.
11. Lam, Y. (1990). An optimal single variable sampling plan with censoring. *The Statistician, 39*, 53–66.
12. Lam, Y. (1994). Bayesian variable sampling plans for the exponential distribution with Type-I censoring. *The Annals of Statistics, 22*, 696–711.
13. Lam, Y., & Choy, S. (1995). Bayesian variable sampling plans for the exponential distribution with uniformly distributed random censoring. *Journal of Statistical Planning and Inference, 47*, 277–293.
14. Liang, T., & Yang, M. C. (2013). Optimal Bayesian sampling plans for exponential distributions based on hybrid censored samples. *Journal of Statistical Computation and Simulation, 83*, 922–940.
15. Liang, T., Yang, M. C., & Chen, L. S. (2017). Optimal Bayesian variable sampling plans for exponential distributions based on modified type-II hybrid censored samples. *Communications in Statistics-Simulation and Computation, 46*, 4722–4744.
16. Lin, C., Huang, Y., & Balakrishnan, N. (2008). Exact Bayesian variable sampling plans for the exponential distribution based on Type-I and Type-II hybrid censored samples. *Communications in Statistics: Simulation and Computation, 37*, 1101–1116.
17. Lin, C., Huang, Y., & Balakrishnan, N. (2010). Corrections on "Exact Bayesian variable sampling plans for the exponential distribution based on type-I and type-II hybrid censored samples". *Communications in Statistics: Simulation and Computation, 39*, 1499–1505.
18. Lin, Y., Liang, T., & Huang, W. (2002). Bayesian sampling plans for exponential distribution based on type-I censoring data. *Annals of the Institute of Statistical Mathematics, 54*, 100–113.
19. Tsai, T. R., Chiang, J. Y., Liang, T., & Yang, M. C. (2014). Efficient Bayesian sampling plans for exponential distributions with type-I censored samples. *Journal of Statistical Computation and Simulation, 84*, 964–981.

Economical Sampling Plans with Warranty

Jyun-You Chiang, Hon Keung Tony Ng, Tzong-Ru Tsai, Yuhlong Lio, and Ding-Geng Chen

Abstract Designing a proper life test plan to evaluate the quality of a lot of products in order to decide on accepting or rejecting the lot between the manufacturer and customers is an important objective in quality control studies. Most existing life test plans are developed based on the mean time to failure (MTTF) of the products in which a lot is acceptable if the MTTF of products is higher than a given threshold and is rejected otherwise. To save the time and cost of a life test, truncated life test with a prefixed upper limit of the test time can be used in acceptance sampling plan. Instead of life test plans based on MTTF, we consider here life test plans simply based on the number of product failures. Nowadays, to make products more competitive in the market, providing product warranty is a common strategy for manufacturers. Therefore, the development of acceptance sampling plans with warranty considerations is desired. In this chapter, the general structure of an economical design of acceptance sampling plan with warranty using truncated life test is studied. To take into account the uncertainty of the underlying model of the product lifetimes, Bayesian approach using prior information and/or preliminary samples is used to design acceptance sampling plan. Methodologies

J.-Y. Chiang (✉)
School of Statistics, Southwestern University of Finance and Economics, Chengdu, China
e-mail: jiangjy@swufe.edu.cn

H. K. T. Ng
Department of Statistical Science, Southern Methodist University, Dallas, TX, USA
e-mail: ngh@mail.smu.edu

T.-R. Tsai
Department of Statistics, Tamkang University, New Taipei City, Taiwan
e-mail: tzongru@gms.tku.edu.tw

Y. Lio
Department of Mathematical Sciences, University of South Dakota, Vermillion, SD, USA
e-mail: Yuhlong.Lio@usd.edu

D.-G. Chen
Department of Statistics, University of Pretoria, Pretoria, South Africa
e-mail: chend05@gmail.com

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_9

and algorithms to obtain the optimal sampling plan that minimizes the expected total cost with warranty considerations are discussed. The proposed methodologies are illustrated with two flexible lifetime distributions, the Burr type XII and the generalized exponential distributions, along with real data examples.

1 Introduction

In quality control, designing a proper life test plan to evaluate the quality of a lot of products in order to decide on accepting or rejecting the lot between the manufacturer and customers is an important issue. Nowadays, due to the development of quality technology, products are highly reliable and it could be time consuming to collect product lifetime information. Truncated life testing scheme, a scheme that prefixed the upper limit of the experimental time and evaluate the product quality based on number of failed items, is one of the competitive schemes to collect lifetime information that can save the test time and cost. On the other hand, to make products more competitive in the market, it is common for manufacturer to provide product warranty when selling the products. Therefore, it is desire to design acceptance sampling plans that takes the cost of warranty into account.

Numerous research papers on acceptance sampling plans have been written in the past few decades, for example, Sobel and Tischendrof [32], Goode and Kao [10], Gupta and Groll [11], Kantam and Rosaiah [17], Vintr [42], Kantam et al. [18], Wu and Tsai [45], Wu and Yu [46], Tsai and Wu [40], Tsai et al. [37], Lio et al. [24, 25], Aslam and Jun [4] and Aslam et al. [5]. Among those, Aslam and Jun [4], Aslam et al. [5], Goode and Kao [10], Gupta and Groll [11], Kantam and Rosaiah [17], Kantam et al. [18], Lio et al. [24, 25], Sobel and Tischendrof [32], Tsai and Wu [40], Wu and Tsai [45] and Wu and Yu [46] constructed acceptance sampling plans via using truncated censoring schemes. Acceptance sampling plans developed under the Bayesian framework can be found in AL-Hussaini and Jaheen [3], AliMousa [1], Kwon [20], Tsai [36] and Tsai et al. [37]. Acceptance sampling plans with warranty considerations have been developed by, for example, Menke [27], Blischke and Scheuer [6], Kwon [20], Thomas and Rao [35], Murthy and Blischke [29], Tsai et al. [37], Huang et al. [16] and Tsai et al. [38, 39].

Following the developments of the work on economical sampling plan, the main objective of this chapter is to provide a general structure of an economical design of acceptance sampling plan with warranty using truncated life test. Bayesian approach using prior information and/or preliminary samples is used to incorporate the uncertainty of the underlying model of the product lifetimes when designing the acceptance sampling plan. The rest of this chapter is organized as follows. In Sect. 2, the general economical design of sampling plan is discussed under the truncated life test. Methodologies and algorithms to obtain the optimal sampling plan that minimizes the expected total cost with warranty considerations are presented. In Sects. 3 and 4, the proposed methodologies are illustrated with the Burr type XII and the generalized exponential distributions, respectively, along with real data examples. Some concluding remarks and information related to the computer programs for the proposed methodologies are given in Sect. 5.

2 Economical Sampling Plan

A truncated acceptance sampling plan is defined by the sample size, n, the acceptance number r_b and the preassigned upper bound of the termination time, τ, and it is denoted by (n, r_b, τ). Suppose a sample of size n is randomly drawn from a lot with N items and put on a life test, the life test is terminated when either more than r_b failed items are observed or the test time reaches the preassigned upper bound τ. If there are more than r_b items failed before time τ, then the lot is rejected; otherwise, the lot is accepted. Because of the destructive nature for the life testing procedure, engineers are often interested in using a truncated acceptance sampling plan to ensure the quality of the product in terms of the mean time to failure (MTTF) or a specified lifetime percentile of the product. In this case, an acceptance sampling plan is determined based on the customer risk and (or) the producer risk. When only the customer risk is under protection, for a given value of r_b and the upper bound of the probability of accepting a bad lot (i.e., the MTTF or a specified lifetime percentile of products is lower than the respective thresholds), the minimum sample size n can be determined such that the customer risk is controlled. When both the customer and producer risks are under protection, the sample size n and acceptance number r_b are determined simultaneously based on controlling the probability of accepting a bad lot and the probability of rejecting a good lot. However, for economical sampling plan, it is determined through a different principle. Given a preassigned upper bound of termination time, τ, the optimal economical sampling plan with sample size n^* and acceptance number r_b^* are determined such that the expected total cost of the life test is minimized. Therefore, to develop an economic sampling plan, the total cost for the life test must be established. Let the cost of a unit on the life test be c_s, the operating cost per unit time for the life test be c_τ, the cost per unit of a rejecting lot be c_r and the cost per unit of an accepted lot be c_a, which is associated with the failure of a unit in an accepted lot during the warranty period.

Suppose the lifetime of an item in a lot, T, is a random variable that follows a two-parameter probability distribution with cumulative distribution function (cdf) $F_T(x|\theta, \delta)$ and probability density function (pdf) $f_T(x|\theta, \delta)$, where θ and δ are the model parameters. It is assumed that all the units in an accepted lot are sold under a general rebate warranty policy, and all the units in a rejected lot are scrapped or reprocessed after the rejection. The general rebate warranty policy to be a mixture of free replacement policy and prorated warranty policy. Specifically, suppose w_1 is the warranty period of free replacement warranty and from w_1 to $w_2 (> w_1)$ is the warranty period of prorated warranty for item sold. Therefore, the overall warranty cost of each item in an accepted lot can be defined by

$$c_a^*(T) = \begin{cases} c_a & \text{if } T < w_1, \\ \frac{c_a(w_2 - T)}{w_2 - w_1} & \text{if } w_1 \le T < w_2, \\ 0 & \text{if } T \ge w_2, \end{cases} \tag{1}$$

and the expected warranty cost per item in an accepted lot is defined as

$$\varpi(\theta, \delta) = c_a \int_0^{w_1} f_T(t|\theta, \delta)dt + \frac{c_a}{w_2 - w_1} \int_{w_1}^{w_2} (w_2 - t) f_T(t|\theta, \delta)dt \quad (2)$$

The warranty policy in Eq. (1) has been studied by many authors including Kwon [20], Huang et al. [16], Tsai [36], and Tsai et al. [37–39]. For $N > n$, the warranty cost of selling an accepted lot can be expressed as

$$W(\theta, \delta) = (N - n)\varpi(\theta, \delta),$$

and the cost of rejecting a lot is $(N - n)c_r$. Let the probability of accepting a lot based on the economical sampling plan be P_A,

$$P_A = \sum_{i=0}^{r_b} \binom{n}{i} p_\tau^i (1 - p_\tau)^{n-i},$$

where $p_\tau = F_T(\tau|\theta, \delta)$, then the total cost for the sampling plan (n, r_b, τ), given the model parameters θ and δ, can be expressed as

$$\begin{aligned} \Psi(n, r_b, \tau|\theta, \delta) &= nc_s + c_\tau T^*(\theta, \delta, \tau) + W(\theta, \delta)P_A + (N - n)c_r(1 - P_A), \\ &= nc_s + c_\tau T^*(\theta, \delta, \tau) + (N - n)[(\varpi(\theta, \delta) - c_r)P_A + c_r], \quad (3) \end{aligned}$$

where $T^*(\theta, \delta, \tau) = E[\min\{\tau, T_{r_b+1}\}] = \int_0^\tau t f_{T_{r_b+1}}(t|\theta, \delta)dt + \tau \int_\tau^\infty f_{T_{r_b+1}}(t|\theta, \delta)dt$, T_{r_b+1} is the $(r_b + 1)$th order failure time, and

$$\begin{aligned} f_{T_{r_b+1}}(t|\theta, \delta) = (r_{b+1})\binom{n}{r_{b+1}} \Big[F_T(t|\theta, \delta)\Big]^{r_b} \Big[1 - F_T(t|\theta, \delta)\Big]^{n-r_b-1} \\ f_T(t|\theta, \delta), \ t > 0. \quad (4) \end{aligned}$$

If $T_{r_b+1} \le \tau$, then the life test is terminated at T_{r_b+1}; otherwise, the life test is terminated at τ. When the lifetime distribution and its parameters θ and δ are known, Eq. (3) can be used to find the optimal economical sampling plan, n^* and r_b^*, that minimizes the expected total cost, i.e.,

$$\Psi(n^*, r_b^*, \tau|\theta, \delta) = \min_{n \in \mathscr{A},\ 0 \le r_b \le n-1} \Psi(n, r_b, \tau|\theta, \delta),$$

where $\mathscr{A}$ is the space of sample sizes and a subset of positive integers. For example, the producer may want to set the smallest and largest sample sizes to be n_l and n_u, respectively, then $\mathscr{A} = \{n_l, n_{l+1}, \ldots, n_u\}$.

In practice, however, the parameters of the lifetime distribution are usually unknown. For a given two-parameter distribution with unknown parameters θ and δ, if prior information on θ and δ is available, then the Bayesian decision method

can be implemented by using Eq. (3) as a loss function in order to reduce the subjectivity of specifying the model parameters. Suppose that the joint prior pdf, denoted by $\pi(\theta, \delta|\Theta)$, is available from prior experience where Θ is the vector of hyper-parameters. The expected total cost to conduct the economical sampling plan (n, r_b, τ), using the priori, is given by

$$\begin{aligned}\Psi_B(n, r_b, \tau) &= E_{\pi(\theta,\delta|\Theta)}[\Psi(n, r_b, \tau|\theta, \delta)] \\ &= nc_s + \int_0^\infty \left[c_\tau T^*(\theta, \delta, \tau) + (N-n)[(\varpi(\theta, \delta) - c_r)P_A + c_r]\right] \\ &\quad \pi(\theta, \delta|\Theta)\, d\theta d\delta. \end{aligned} \tag{5}$$

When the hyper-parameter, Θ, is specified based on prior experience or expert opinion, Eq. (5) can be used as the expected total cost of the sampling plan. Hence, the optimal sampling plan, (n^*, r_b^*, τ), can be obtained by minimizing $\Psi_B(n, r_b, \tau)$, i.e.,

$$\Psi_B(n^*, r_b^*, \tau|\theta, \delta) = \min_{n\in\mathscr{A},\ 0\le r_b\le n-1} \Psi_B(n, r_b, \tau|\theta, \delta),$$

where $\mathscr{A}$ is the space of sample sizes and a subset of positive integers. When the hyper-parameter, Θ, is unknown, an empirical estimate, $\hat{\Theta}$, of the hyper-parameter based on a preliminary sample is needed. Then, we can consider

$$\begin{aligned}\Psi_M(n, r_b, \tau) &= E_{\pi(\theta,\delta|\hat{\Theta})}[\Psi(n, r_b, \tau|\theta, \delta)] \\ &= nc_s + \int_0^\infty \left[c_\tau T^*(\theta, \delta, \tau) + (N-n)[(\varpi(\theta, \delta) - c_r)P_A + c_r]\right] \\ &\quad \pi(\theta, \delta|\Theta)\, d\theta d\delta, \end{aligned} \tag{6}$$

as the expected total cost. In this situation, the optimal sampling plan, (n^*, r_b^*, τ), can be obtained by minimizing $\Psi_M(n, r_b, \tau)$, i.e.,

$$\Psi_M(n^*, r_b^*, \tau|\theta, \delta) = \min_{n\in\mathscr{A},\ 0\le r_b\le n-1} \Psi_M(n, r_b, \tau|\theta, \delta).$$

Let $\Psi(n, r_b, \tau)$ be $\Psi(n, r_b, \tau|\theta, \delta)$, $\Psi_B(n, r_b, \tau)$ or $\Psi_M(n, r_b, \tau)$, then the aforementioned optimal sampling plan, (n^*, r_b^*, τ), for each corresponding loss function can be obtained by the following searching algorithm:

Step 1: Provide the preassigned upper bound of the termination time τ;
Step 2: For each (n, τ), the optimal acceptance number, $r_b^*(n)$, can be determined by searching over the space $\mathscr{B} = \{0, 1, \ldots, n-1\}$ such that

$$\Psi(n, r_b^*(n), \tau) = \min_{r_b(n)\in\mathscr{B}} \Psi(n, r_b(n), \tau);$$

Step 3: The economical sampling plan (n^*, r_b^*, τ) can be obtained by searching over the space $\mathscr{A}$ such that

$$\Psi(n^*, r_b^*, \tau) = \min_{n \in \mathscr{A}} \Psi(n, r_b^*(n), \tau).$$

3 Burr Type XII Distribution

The Burr type XII distribution, denoted by Burr(θ, δ), was originally introduced by Burr [8]. Tadikamalla [34] discovered the connections of the Burr type XII distribution with many other useful lifetime distributions and concluded that the Burr type XII distribution could be used to model different kinds of data sets due to its flexibility with two shape parameters. Since then, the Burr type XII distribution has gained much attention for its potential applications in the reliability studies by many authors, such as Wingo [43, 44], Aslam et al. [5], AliMousa [1], AliMousa and Jaheen [2], Gupta et al. [12], Lio and Tsai [23], Upadhyay et al. [41], Wu and Yu [46], Wu et al. [47], Zimmer et al. [49], Wu et al. [48] and Soliman et al. [33]. The pdf and cdf of the Burr(θ, δ) distribution are given by

$$f_T(t) = f(t|\theta, \delta) = \theta\delta t^{\theta-1}(1+t^\theta)^{-\delta-1}, \ t > 0,$$

and

$$F_T(t) = F(t|\theta, \delta) = 1 - (1+t^\theta)^{-\delta}, \ t > 0,$$

respectively, where $\theta > 0$ and $\delta > 0$ are shape parameters. Hence, the expected total time for the life test is

$$T^*(\theta, \delta, \tau) = (r_b+1)\binom{n}{r_b+1}\left\{\int_0^\tau t^\theta \theta\delta(1-(1+t^\theta)^{-\delta})^{r_b}(1+t^\theta)^{-\delta(n-r_b)-1}dt + \tau\int_{p_\tau}^1 y^{r_b}(1-y)^{n-r_b-1}dy\right\}. \quad (7)$$

3.1 Bayesian Approach

Since there is no conjugate priors for the two shape parameters of the Burr XII distribution, using the procedure suggested by Tsai et al. [38], we let θ be a known constant θ^* and consider the conjugate prior denoted by $\pi_2(\delta|\Theta)$ for the shape parameter δ, where $\Theta = (\alpha, \beta)$ is the vector of hyper-parameters. Here, the prior distribution of δ is given by

$$\pi_2(\delta|\alpha,\beta) = \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}\delta^{\alpha}e^{-\beta\delta}, \quad \delta > 0, \tag{8}$$

where $\alpha > -1$ and $\beta > 0$. Then, the expected total cost presented in Eq. (5) can be expressed as

$$\begin{aligned}\Psi_B(n, r_b, \tau) &= E_{g(k|\Theta)}[\Psi(n, r_b, \tau|\theta, \delta)] \\ &= nc_s + \int_0^{\infty} \{c_\tau T^*(\theta, \delta, \tau) + (N-n)\left[(\varpi(\theta, \delta) - c_r)P_A + c_r\right]\} \\ &\quad \pi_2(\delta|\Theta)\, d\delta,\end{aligned} \tag{9}$$

and the optimal sampling plan, (n^*, r_b^*, τ), can be obtained as

$$\Psi_B(n^*, r_b^*, \tau) = \min_{n\in\mathscr{A},\ 0\le r_b\le n-1} \Psi_B(n, r_b, \tau). \tag{10}$$

To specify the hyper-parameters of the prior distribution of δ, an empirical Bayesian-moment method proposed by Ali Mousa [1] can be applied. Suppose $t_1 < t_2, \cdots, < t_m$ is a preliminary type II censored sample of items that follow the Burr type XII distribution and we let

$$X = \sum_{i=1}^{m} \log(1 + t_i^{\theta^*}) + (n-m)\log(1 + t_m^{\theta^*}).$$

It can be shown that the conditional distribution of the statistic X with a specific value of θ^* is a gamma distribution with pdf

$$f_{X|\delta}(x|\delta) = \frac{\delta^m}{\Gamma(m)}x^{m-1}e^{-\delta x}, \quad x > 0.$$

The posterior pdf of δ, given the value of the statistic X based on the data, i.e., $X = x$, is also a gamma distribution with pdf

$$g(\delta|x, \alpha, \beta) = \frac{(\beta+x)^{m+\alpha+1}}{\Gamma(m+\alpha+1)}\delta^{m+\alpha}e^{-(\beta+x)\delta}, \quad \delta > 0. \tag{11}$$

The marginal pdf of X can be obtained as

$$f_X(x) = \frac{\Gamma(m+\alpha+1)}{\Gamma(m)\Gamma(\alpha+1)}\frac{\beta^{\alpha+1}x^{m-1}}{(\beta+x)^{m+\alpha+1}}, \quad x > 0, \tag{12}$$

which is the pdf of an inverted beta distribution with parameters m and $\alpha + 1$. Let $L(\alpha, \beta|x) = \log(f_X(x))$, then

$$L(\alpha, \beta|x) \propto m \log(\Gamma(m+\alpha+1)) - \log(\Gamma(\alpha+1)) + (\alpha+1)\log(\beta) - (m+\alpha+1)\log(\beta+x). \tag{13}$$

Based on a simulation study, Tsai et al. [38] mentioned that it could be difficult to estimate the parameters α and β precisely with a truncated sample using the maximum likelihood estimation method which maximizes Eq. (13) with respect to the parameters. For the hyper-parameters α and β, based on the priori in Eq. (8), the mean of the prior distribution can be obtained as

$$\eta_0 = \frac{\alpha+1}{\beta}. \tag{14}$$

Let $\partial L(\alpha, \beta|x)/\partial\beta = 0$, we can obtain

$$\beta = \frac{x(\alpha+1)}{m}. \tag{15}$$

From Eqs. (14) and (15), an estimate of η_0 can be obtained as $\hat{\eta}_0 = m/x$.

Suppose we place an upper bound, γ_U, as the variance of the prior distribution, $\gamma = \eta_0/\beta$, then based on Eq. (14) and setting $\gamma_U = \hat{\eta}_0/\hat{\beta}_M$, the estimates of α and β can be obtained, respectively, as

$$\hat{\alpha}_M = \frac{\hat{\eta}_0^2}{\gamma_U} - 1 \tag{16}$$

and

$$\hat{\beta}_M = \frac{\hat{\eta}_0}{\gamma_U}. \tag{17}$$

Then, the posterior pdf of Eq. (11) can be replaced by $g_M(\delta) = g(\delta|x, \hat{\alpha}_M, \hat{\beta}_M)$, and the posterior expected total cost of the sampling plan can be expressed as

$$\begin{aligned}\Psi_M(n, r_b, \tau) &= E_{g_M(\delta)}[\Psi(n, r_b, \tau|\theta^*, \delta)] \\ &= nc_s + \int_0^\infty \left\{c_\tau T^*(\theta^*, \delta, \tau) + (N-n)[(\varpi(\theta^*, \delta) - c_r)P_A + c_r]\right\} \\ &\quad g_M(\delta)\, d\delta.\end{aligned} \tag{18}$$

The optimal sampling plan, (n^*, r_b^*, τ), can be obtained by searching over the space of sample sizes, $\mathscr{A}$, such that

$$(n^*, r_b^*, \tau) = \arg \min_{n \in \mathscr{A},\ 0 \le r_b \le n-1} \Psi_M(n, r_b, \tau) \tag{19}$$

If the value of θ^* is unknown, then it can be replaced by the maximum likelihood estimate of θ, denoted by $\hat{\theta}_{ML}$, based on the preliminary type II censored sample $t_1 < t_2, \cdots, < t_m$.

3.2 *Illustrative Example*

In this subsection, one of the numerical examples in Tsai et al. [39] is used to illustrate the methodologies. The first failure times (in months) from 20 small electric carts, which are used for internal transportation and delivery in a large manufacturing facility, are used to develop the proposed sampling plan. It has been shown that the Burr type XII distribution can be used to model this data set [25, 49]. Lio et al. [25] demonstrated that the Burr type XII distribution is a suitable model for this data set using a Kolmogorov-Smirnov goodness-of-fit test with a p-value 0.1008 at the maximum likelihood estimates $\hat{\theta}_{ML} = 5.47$ and $\hat{\delta}_{ML} = 0.08$.

In order to save the cost of the life test, the producer set the smallest sample size and the largest sample size for the sampling plan as $n_l = 5$ and $n_u = 20$, respectively and the life test is terminated after 3 months (i.e., $\tau = 3$). For the warranty policy after sale, the values of w_1 and w_2 are set as the 20th and 50th percentiles of the Burr type XII distribution with some pre-specified parameter values, respectively. To investigate the optimal economic sampling plan and the sensitivity of the cost components on the optimal sampling plan, we consider $c_s = 1$, $c_r = p_1 \times c_s$, $c_a = p_2 \times c_s$ and $c_\tau = p_3 \times c_s$, where $(p_1, p_2, p_3) = (1, 1, 1)$, (0.1, 0.1, 1), (0.1, 0.5, 1), (0.1, 0.5, 0.5).

First, suppose the lifetime distribution is known as the Burr type XII distribution with $\theta = 5.47$ and $\delta = 0.08$, then the values of w_1 and w_2 are set as the 20th and 50th percentiles of Burr($\theta = 5.47$,$\delta = 0.08$), i.e., $w_1 = 1.65$ and $w_2 = 4.87$. The optimal economical sampling plans can (n^*, r_b^*, τ) can be obtained by searching over the space $n \in \mathscr{A} = \{5, 6, \ldots, 20\}$ and $r_b \in 0, 1, \ldots, n-1$ through the algorithm described in Sect. 2 such that the expected total cost is being minimized, i.e.,

$$\Psi(n^*, r_b^*, \tau|\theta, \delta) = \min_{n\in\mathscr{A}, 0\le r_b\le n-1} \Big\{ nc_s + c_\tau T^*(\theta, \delta, \tau) + (N-n)[(\varpi(\theta, \delta)P_A + (1-c_r)P_A] \Big\}. \tag{20}$$

Table 1 shows the smallest expected total cost for economical sampling plan $(n, r_b^*(n))$ given $n \in \mathscr{A}$ and Table 2 shows the expected total costs under the optimal sampling plans (n^*, r_b^*, τ) from different settings of (p_1, p_2, p_3). From Table 1, we observe that for setting (p_1, p_2, p_3), the smallest expected total cost $\Psi(n, r_b^*(n), \tau)$ decreases monotonically to the minimum with n increases and then

Table 1 Economic sampling plans for Burr($\theta = 5.47, \delta = 0.08$), $\tau = 3$ and $N = 500$

	$(p_1, p_2, p_3) = (1, 1, 1)$		$(p_1, p_2, p_3) = (0.1, 0.1, 1)$		$(p_1, p_2, p_3) = (0.1, 0.5, 1)$		$(p_1, p_2, p_3) = (0.1, 0.5, 0.5)$	
n	$r_b^*(n)$	$\Psi(n, r_b^*(n), \tau \mid \theta, \delta)$	$r_b^*(n)$	$\Psi(n, r_b^*(n), \tau \mid \theta, \delta)$	$r_b^*(n)$	$\Psi(n, r_b^*(n), \tau \mid \theta, \delta)$	$r_b^*(n)$	$\Psi(n, r_b^*(n), \tau \mid \theta, \delta)$
5	4	202.04	4	27.40	0	60.20	0	59.44
6	5	201.13	5	28.21	0	59.39	0	58.69
7	6	201.17	6	29.12	0	59.21	0	58.55
8	7	201.56	7	30.06	0	59.43	0	58.81
9	8	202.09	8	31.01	0	59.90	0	59.30
10	9	202.67	9	31.97	0	60.52	0	59.95
11	10	203.27	10	32.93	0	61.23	0	60.68
12	11	203.88	11	33.89	0	62.01	0	61.48
13	12	204.49	12	34.85	0	62.83	0	62.31
14	13	205.10	13	35.81	0	63.67	0	63.17
15	14	205.71	14	36.77	0	64.53	0	64.03
16	15	206.33	15	37.73	0	65.40	0	64.91
17	16	206.94	16	38.69	0	66.27	0	65.79
18	17	207.55	17	39.66	0	67.15	0	66.68
19	18	208.16	18	40.62	0	68.03	0	67.57
20	19	208.78	19	41.58	0	68.92	0	68.46

Table 2 Optimal test plans for $\tau = 3$ and $N = 500$

(p_1, p_2, p_3)	(n^*, r_b^*)	$\Psi(n^*, r_b^*, \tau \mid \theta, \delta)$
(1, 1, 1)	(6, 5)	201.13
(0.1, 0.1, 1)	(5, 4)	27.4
(0.1, 0.5, 1)	(7, 0)	59.21
(0.1, 0.5, 0.5)	(7, 0)	58.55

Table 3 Simulated type II censored sample from Burr($\theta = 5.47, \delta = 0.08$), censored rate 1/7

0.460	0.627	0.645	0.673	0.685	0.756	0.759	0.768	0.789	0.789
0.865	0.885	0.890	0.897	0.947	0.952	0.972	0.994	1.005	1.026
1.030	1.032	1.033	1.059	1.063	1.112	1.119	1.130	1.159	1.171
1.171	1.184	1.191	1.193	1.232	1.233	1.240	1.268	1.277	1.299
1.301	1.315	1.344	1.389	1.400	1.401	1.411	1.413	1.423	1.439

increases monotonically with n afterward. It means that the smallest expected total cost, $\Psi(n, r_b^*(n), \tau \mid \theta, \delta)$, is a convex function of n given a setting of (p_1, p_2, p_3). Therefore, the expected total cost for the optimal sampling plan (n^*, r_b^*) over $5 \le n \le 20$ can be obtained and they are presented in Table 2. In view of Table 2, when the cost of accepting the lot is higher than the cost for rejecting a lot, the optimal acceptance number r_b^* would be forced to be close to or equal to zero; otherwise, r_b^* would be close to the sample size n^*.

When the parameters of the two-parameter lifetime population distribution are unknown, we consider obtaining a preliminary type II censored sample in order to get the information related to the parameters. Following the simulation study conducted by Tsai et al. [39], we consider a type II censored sample with $r = 50$ observed failures of small electric carts from a sample of size 350 presented in Table 3. The parameter θ and the hyper-parameters α and β for the prior distribution of δ can be estimated based on the preliminary type II censored sample. Based on the type II censored sample presented in Table 3, the maximum likelihood estimate of θ and δ are $\hat{\theta}_{ML} = 5.03$ and $\hat{\delta}_{ML} = 0.078$, respectively, and the upper variation bound of δ is set to be $\gamma_U = 3 \times 0.078 = 0.234$. Suppose that the values of w_1 and w_2 for the warranty setting are set as the 20th and 50th percentiles of Burr($\theta = 5.03, \delta = 0.078$), respectively, i.e., $w_1 = 1.72$ and $w_2 = 5.60$. The optimal economic sampling plan n^*, r_b^*, τ can be obtained by searching over the space $n \in \mathscr{A} = 5, 6, \cdots, 20$ and $r_b \in 0, 1, \cdots, n-1$ through the algorithm described in Sect. 2 such that the posterior expected total cost in Eq. (18) is being minimized, i.e.,

$$\Psi_M(n^*, r_b^*, \tau) = \min_{n \in \mathscr{A}, 0 \le r_b \le n-1} \Psi_M(n, r_b, \tau), \tag{21}$$

Table 4 shows the smallest expected total costs for economical sampling plan, $(n, r_b^*(n))$, where $n \in \mathscr{A}$ under different settings of (p_1, p_2, p_3). Table 5 displays the corresponding smallest expected total cost for the optimal sampling plan, (n^*, r_b^*), for different settings of (p_1, p_2, p_3). From Table 4, the behavior of the smallest expected total cost, $\Psi_M(n, r_b^*, \tau)$, is similar to the behavior of the smallest

Table 4 Life test economic sampling plans for Burr($\theta = 5.47, \delta = 0.08$), $\tau = 3$ and $N = 500$ when the model parameters are unknown

	$(p_1, p_2, p_3) = (1, 1, 1)$		$(p_1, p_2, p_3) = (0.1, 0.1, 1)$		$(p_1, p_2, p_3) = (0.1, 0.5, 1)$		$(p_1, p_2, p_3) = (0.1, 0.5, 0.5)$	
n	$r_b^*(n)$	$\Psi_M(n, r_b^*(n), \tau)$	$r_b^*(n)$	$\Psi_M(n, r_b^*(n), \tau)$	$r_b^*(n)$	$\Psi_M(n, r_b^*(n), \tau)$	$r_b^*(n)$	$\Psi_M(n, r_b^*(n), \tau)$
5	4	202.00	4	27.40	0	61.39	0	60.60
6	5	201.53	5	28.25	0	60.33	0	59.59
7	6	201.75	6	29.18	0	59.94	0	59.25
8	7	202.22	7	30.12	0	59.99	0	59.34
9	8	202.78	8	31.08	0	60.31	0	59.70
10	9	203.37	9	32.04	0	60.83	0	60.24
11	10	203.97	10	33.00	0	61.47	0	60.90
12	11	204.58	11	33.96	0	62.19	0	61.64
13	12	205.19	12	34.92	0	62.96	0	62.43
14	13	205.80	13	35.88	0	63.77	0	63.25
15	14	206.42	14	36.84	0	64.60	0	64.09
16	15	207.03	15	37.80	0	65.45	0	64.95
17	16	207.64	16	38.76	0	66.31	0	65.83
18	17	208.25	17	39.73	0	67.18	0	66.70
19	18	208.86	18	40.69	0	68.06	0	67.59
20	19	209.47	19	41.65	0	68.94	0	68.47

Table 5 Optimal test plans for $\tau = 3$ and $N = 500$ when the model parameters are unknown

(p_1, p_2, p_3)	(n^*, r_b^*)	$\Psi_M(n^*, r_b^*, \tau)$
(1, 1, 1)	(6, 5)	201.53
(0.5,0.1,0.5)	(5, 4)	27.05
(0.1, 0.1, 1)	(5, 4)	27.40
(0.1, 0.5, 1)	(7, 0)	59.94
(0.1, 0.5, 0.5)	(7, 0)	59.25

expected total cost, $\Psi(n, r_b^*, \tau)$ in Table 1. Furthermore, the optimal sampling plans in Table 5 are the same as the optimal sampling plans in Table 2. If the accept cost is larger than the reject cost, it needs a large sample size with a small acceptance number; otherwise, the optimal sampling plan requires a small sample size and a large acceptance number for the truncated life test.

4 Generalized Exponential Distribution

The generalized exponential distribution was originally introduced by Gupta and Kundu [13], which has the pdf and cdf, respectively,

$$f_T(t) = \delta\theta(1 - e^{-\theta t})^{\delta-1}e^{-\theta t},\ t > 0, \tag{22}$$

and

$$F_T(t) = (1 - e^{-\theta t})^{\delta},\ \delta > 0, \theta > 0,\ t > 0, \tag{23}$$

where δ is the shape parameter and θ is the rate parameter. When $\delta = 1$, the generalized exponential distribution reduces to the conventional exponential distribution. The pdf in (22) is a decreasing function if $\delta < 1$, and it is a unimodal function if $\delta > 1$. Similar to the Weibull distribution, the hazard function of generalized exponential distribution can be increasing, decreasing or constant depending on the value of δ. Mudholkar and Srivastava [28] commented that the generalized exponential distribution can be an alternative to the commonly used lifetime distributions such as gamma and Weibull distributions. Gupta and Kundu [14, 15] showed that the generalized exponential distribution can be used to model many types of lifetime data effectively.

When the lifetime distribution of the items is generalized exponential distribution, the expected total time for the life test is

$$T^*(\theta, \delta, \tau) = \sum_{k=0}^{n-r_b-1} \frac{(-1)^k n!}{k! r_b! (n - r_b - 1 - k)!} \left\{ \delta \int_0^{\tau} \theta t (1 - e^{-\theta t})^{\delta(k+r_b+1)-1} e^{-\theta t} dt \right.$$
$$\left. + \frac{\tau}{k + r_b + 1} [1 - (1 - e^{-\theta\tau})^{\delta(k+r_b+1)}] \right\}. \tag{24}$$

4.1 Bayesian Approach

When both θ and δ are unknown, Kundu and Gupta [19] suggested independent gamma distributions for the generalized exponential distribution parameters, θ and δ to be used for Bayesian inference. Following the suggestion in Kundu and Gupta [19], we consider the prior distributions of θ and δ as

$$\pi_1(\theta) \propto \theta^{b-1} e^{-a\theta}, \quad \theta > 0, \tag{25}$$

and

$$\pi_2(\delta) \propto \delta^{d-1} e^{-c\delta}, \quad \delta > 0. \tag{26}$$

Therefore, the joint prior distribution, $\pi(\theta, \delta|\Theta)$, is given by

$$\pi(\theta, \delta|\Theta) = \pi_1(\theta)\pi_2(\delta), \quad \theta > 0, \quad \delta > 0, \tag{27}$$

where $\Theta = (a, b, c, d)$. Then Eq. (5) can be represented as

$$\begin{aligned} \Psi_1(n, r_b, \tau) &= E_{\pi(\theta,\delta|\Theta)}[\Psi(n, r_b, \tau|\theta, \delta)] \\ &= nc_s + \int_0^\infty \int_0^\infty \Big[c_\tau T^*(\theta, \delta, \tau) + W(\theta, \delta) + R(\theta, \delta)\Big] \\ &\quad \pi(\theta, \delta) d\theta d\delta, \end{aligned} \tag{28}$$

and the optimal sampling plan, (n^*, r_b^*, τ), can be obtained by using the algorithm described in Sect. 2 with $\Psi(n, r_b, \tau)$ replaced by $\Psi_1(n, r_b, \tau)$, i.e.,

$$\Psi_1(n^*, r_b^*, \tau) = \min_{n \in \mathscr{A},\ 0 \le r_b \le (n-1)} \Psi_1(n, r_b, \tau), \tag{29}$$

where $\mathscr{A}$ denotes the space of sample size.

If additional information about the hyper-parameters is available, for example, a preliminary type II censored sample, $\mathbf{t} = (t_1, t_2, \cdots, t_r)$, from the generalized exponential distribution is observed, then the likelihood can be expressed as

$$L(\theta, \delta|\mathbf{t}) = \delta^r \theta^r \exp\Big\{ -\theta \sum_{i=1}^{r} t_i + (\delta - 1) \sum_{i=1}^{r} \ln(1 - e^{-\theta t_i}) \Big\}. \tag{30}$$

Using the joint prior distribution in Eq. (27), the joint posterior pdf is

$$\begin{aligned} \pi(\theta, \delta|\boldsymbol{t}) \propto\ & \delta^{r+d-1} \theta^{r+b-1} e^{-\theta(a + \sum_{i=1}^{r} t_i)} \\ & \times e^{-\delta(c - \sum_{i=1}^{r} \ln(1 - e^{-\theta t_i}))} e^{-\sum_{i=1}^{r} \ln(1 - e^{-\theta t_i})}, \end{aligned} \tag{31}$$

and it can be represented as

$$\pi(\theta,\delta|\boldsymbol{t}) \propto g_{\theta}\Big(r+b, a+\sum_{i=1}^{r} t_i\Big) \times g_{\delta|\theta}\Big(r+d, c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})\Big) e^{-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})}, \quad (32)$$

where $g_{\theta}(r+b, a+\sum_{i=1}^{r} t_i)$ and $g_{\delta|\theta}(r+d, c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i}))$ are posterior gamma pdfs for θ and δ, respectively. The posterior model is essentially an updated version of the prior knowledge on θ and δ in the light of the preliminary sample data. Then, the marginal posterior pdf of θ is

$$\pi(\theta|\boldsymbol{t}) \propto g_{\theta}\Big(r+b, a+\sum_{i=1}^{r} t_i\Big) \times \Big(c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})\Big)^{-(r+d)} e^{-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})}, \quad (33)$$

and the conditional posterior pdf of δ, given θ, is

$$g_{\delta|\theta}\Big(r+d, c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})\Big) \propto \delta^{r+d-1} e^{-\delta(c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i}))}, \quad (34)$$

which is a gamma pdf with shape parameter $r+d$ and rate parameter $(c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i}))$. Replacing $\pi(\theta,\delta)$ in Eq. (28) by the joint pdf in Eq. (32), the posterior expected total cost for determining the sampling plan (n, r_c, τ) is given by

$$\begin{aligned}\Psi_2(n, r_b, \tau) &= E_{\pi(\theta,\delta|\boldsymbol{t})}[\Psi(n, r_b, \tau|\theta,\delta)] \\ &= nc_s + \int_0^{\infty}\int_0^{\infty}\Big[c_{\tau}T^*(\theta,\delta,\tau) + W(\theta,\delta) + R(\theta,\delta)\Big]\pi(\theta,\delta|\boldsymbol{t})d\theta d\delta. \quad (35)\end{aligned}$$

Then, Eq. (35) can be formulated as

$$\Psi_2(n, r_b, \tau) = nc_s + \frac{E_{g_{\delta|\theta}g_{\theta}}\Big[(c_{\tau}T^*(\theta,\delta,\tau) + W(\theta,\delta) + R(\theta,\delta))e^{-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})}\Big]}{E_{g_{\theta}}\Big[e^{-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})}\Big]}, \quad (36)$$

where the expected value $E_{g_{\delta|\theta}g_\theta}[\cdot]$ in the numerator is taken with respect to the joint pdf

$$g_{\delta|\theta}\Big(r+d, c-\sum_{i=1}^{r}\ln(1-e^{-\theta t_i})\Big)g_\theta\Big(r+b, a+\sum_{i=1}^{r}t_i\Big), \tag{37}$$

and the expected value $E_{g_\theta}[\cdot]$ in the denominator is taken with respect to the distribution $g_\theta(r+b, a+\sum_{i=1}^{r}t_i)$. The optimal sampling plan (n^*, r_b^*, τ) can be determined by using the algorithm described in Sect. 2 with $\Psi(n, r_b, \tau)$ replaced by $\Psi_2(n, r_b, \tau)$, i.e.,

$$\Psi_2(n^*, r_b^*, \tau) = \min_{n\in\mathscr{A},\ 0\le r_b\le(n-1)} \Psi_2(n, r_b, \tau). \tag{38}$$

4.2 Illustrative Example

The numerical example related to integrated circuit (IC) manufacturing provided by Huang et al. [16] and revisited by Tsai et al. [38] is used here for illustration. In this example, the IC manufacturer needs establishing a cost-saving life test procedure to make a decision of accepting or rejecting the submitted lots. Huang et al. [16] modeled the IC lifetime data with the Weibull distribution and Tsai et al. [38] used the generalized exponential distribution in Eqs. (22) and (23) to model the IC lifetime. In this case, each submitted lot has $N = 1000$ IC products and the sampling cost is $c_s = \$25$ per unit. The total rebate warranty is within 300 h (i.e., $w_1 = 0.3$, in 1000 h), and the time limit of prorated warranty is 750 h (i.e., $w_2 = 0.75$, in 1000 h). The cost of rejecting a unit is $c_r = 5c_s$, the cost associated with an external failure in an accepted lot is $c_a = 10c_s$, and the cost per unit of test time is $c_\tau = 2c_s$. The life test will last at most τ 1000 h or it will be terminated when $r_b + 1$ failure items are observed within τ 1000 h.

The optimal sampling plan in the range of $\mathscr{A} = \{n|6 \le n \le 20\}$ using Bayesian approach need to establish with the joint prior pdf and joint posterior pdf. If there is no prior information available, then the optimal sampling plan can be established by minimizing $\Psi_1(n, r_b, \tau)$ in Eq. (29). The hyper-parameters are conjectured to be $a = 5$, $b = 3$, $c = 4$ and $d = 2$ for the priori. Note that this selection of hyper-parameters could be subjective in the initial stage, however, the prior assumption can be modified after additional lifetime information is collected from the life test.

Using the proposed algorithm described in Sect. 2, the optimal sampling plan, $r_b^*(n)$, for different sample sizes in $\mathscr{A}$ and different values of τ, are obtained and the results for $\tau = 0.35$ 1000 h are displayed in Table 6. It can be seen that the expected total cost, $\Psi_1(n, r_b^*(n), \tau)$, for economic sampling plan is again a convex function of n for different settings. The patterns of the smallest expected total cost $\Psi_1(n, r_b^*(n), \tau)$ for the rest cases for different τ are similar to the results shown in Table 6. The optimal sampling plans (n_1^*, r_{b1}^*, τ) for different τ and the

Table 6 Economic sampling plans for generalized exponential distribution with $\tau = 350$ 1000 h and $N = 1000$

	Optimal sampling plan using *priors*		Optimal sampling plan using *posteriors*	
n	$r_b^*(n)$	$\Psi_1(n, r_b^*(n), \tau)$	$r_b^*(n)$	$\Psi_2(n, r_b^*(n), \tau)$
6	2	109816.72	2	117342.54
7	2	109405.19	2	117029.29
8	3	108990.86	3	116563.32
9	3	108589.97	3	116216.25
10	4	108552.10	4	116058.87
11	4	108373.42	4	115825.26
12	5	108701.57	5	115909.86
13	5	109278.62	5	116086.82
14	6	110415.41	6	116502.23
15	8	112013.91	7	117269.24
16	9	113558.75	8	118254.75
17	10	115239.79	9	119356.43
18	11	117035.24	10	120500.75
19	13	118147.29	12	121585.41
20	14	119205.81	13	122281.06

Table 7 Optimal sampling plans for generalized exponential distribution with different values of τ (in 1000 h)

τ	(n_1^*, r_{b1}^*)	$\Psi_1(n_1^*, r_{b1}^*, \tau)$	(n_2^*, r_{b2}^*)	$\Psi_2(n_2^*, r_{b2}^*, \tau)$
200	(11, 3)	108079.15	(12, 3)	112820.89
250	(11, 3)	108265.99	(12, 4)	115131.51
300	(11, 4)	108325.84	(12, 4)	115892.23
350	(11, 4)	108373.42	(11, 4)	115825.26
400	(10, 4)	108466.18	(12, 5)	115133.76

corresponding expected total cost values, $\Psi_1(n^*1, r_{b1}^*, \tau)$, are shown on Table 7. Table 7 indicates $\Psi_1(n_1^*, r_{b1}^*, \tau)$ increases with respect to τ.

Huang et al. [16] reported a type II censored life test for 10 IC products that can provide additional information for adjusting the prior distribution. The first 6 lifetimes reported by Huang et al. [16] were 187.56, 212.88, 266.55, 326.55, 395.19, 420.82 in 1000 h. To implement the economical sampling plan, the products that have lifetimes shorter than τ 1000 h will be used to adjust the prior distribution in this case study. Based on those IC lifetimes smaller than τ 1000 h, the joint posterior pdf in Eq. (36) can be established and it is applied in $\Psi_2(n, r_b, \tau)$ to obtain the revised optimal life test plan. Table 6 also provides the expected total cost, $\Psi_2(n, r_b^*(n), \tau)$, under $\tau = 0.35$ 1000 h for all the sampling plans considered here. Comparing $\Psi_1(n, r_b^*(n), \tau)$ with $\Psi_2(n, r_b^*(n), \tau)$, it can be seen that both functions are convex and $\Psi_2(n, r_b^*(n), \tau) \geq \Psi_1(n, r_b^*(n), \tau)$ in this example. Therefore, the optimal sampling plan (n_2^*, r_{b2}^*, τ) that minimizes $\Psi_2(n, r_b^*(n), \tau)$ can be obtained and the optimal sample plans for different values of τ are presented in Table 7.

Based on the aforementioned type II censoring sample with 6 failure times, the termination time for $\tau = 200, 250, 300, 350$ and 400 h are selected to assess the sensitivity of the optimal test plan on τ. Table 7 shows the optimal life test plans for different values of τ via the prior and posterior distributions. It can be noticed that the optimal sample size n^* for the life test and optimal acceptance number, r_b^* are insensitive to the value of τ. The optimal sample size is around 10–12 and the acceptance number is about 3–5.

5 Concluding Remarks

In this chapter, a review of the essential techniques for constructing economical sampling plans for acceptance or rejection of a product lot is provided. Bayesian approach is used to incorporate the uncertainty of the underlying lifetime distribution of the product. The proposed methodologies and algorithms are illustrated with the Burr type XII and the generalized exponential distributions.

Note that the proposed methodologies can be extended to other lifetime distributions with suitable modifications. Moreover, since the maximum likelihood estimators do not possess a closed form and numerical methods are needed to obtain the estimates, instead of using maximum likelihood estimation method in determining the hyper-parameters of the prior distributions in the Bayesian approach, other estimation methods can be adopted to simplify the algorithm to obtain information from a preliminary sample. Another issue on the methodologies proposed here is that there are several multi-dimensional integrals need to be evaluated in order to compute the expected total cost and most of these multi-dimensional integrals cannot be obtained explicitly mathematically. In this study, we applied a suitable transformation to convert the infinite integral interval, $[0, \infty]$ to $[0, 1]$ and then Riemann sum approximation with larger number of subintervals over $[0, 1]$ is used to evaluate the integral numerically. It will be interesting to apply other numerical or Monte Carlo integration methods and compare the accuracy as well as the effect to the optimal sampling plan for these methods. For the detailed computations related to the economical sampling plans presented this chapter, computer programs written in R [30] can be obtained at http://faculty.smu.edu/ngh/R_program_ESP.zip.

References

1. Ali Mousa, M. A. M. (1995). Empirical Bayes estimators for the Burr type XII accelerated life testing model based on type-2 censored data. *Journal of Statistical Computation and Simulation, 52*, 95–103.
2. Ali Mousa, M. A. M., & Jaheen, Z. F. (2002). Statistical inference for the Burr model based on progressively censored data. *Computers and Mathematics with Applications, 43*, 1441–1449.

3. AL-Hussaini, E. K., & Jaheen, Z. F. (1992). Bayesian estimation of the parameters. *Journal of Statistical Computation and Simulation, 41*, 31–40.
4. Aslam, M., & Jun, C.-H. (2009). A group acceptance sampling plan for truncated life test having Weibull distribution. *Journal of Applied Statistics, 36*, 1021–1027.
5. Aslam, M., Lio, Y. L., & Jun, C.-H. (2013). Repetitive acceptance sampling plans for Burr type XII distribution. *International Journal of Advanced Manufacturing Technology, 68*, 495–507.
6. Blischke, W. R., & Scheuer, E. M. (1975). Calculation of the cost of warranty policies as a function of estimated life distributions. *Naval Research Logistics, 22*, 681–696.
7. Brush, G. G. (1986). A comparison of classical and Bayes producer's risk. *Technometrics, 28*, 69–72.
8. Burr, I. W. (1946). Cumulative frequency functions. *Annals of Mathematical Statistics, 13*, 215–232.
9. Gerontidis, I., & Smith, R. L. (1982). Monte carlo generation of order statistics from general distributions. *Journal of the Royal Statistical Society Series C Applied Statistics, 31*, 238–243.
10. Goode, H. P., & Kao, J. H. K. (1961). Sampling plans based on the Weibull distribution. In *Proceedings of seventh national symposium on reliability and quality control, Philadelphia* (pp. 24–40).
11. Gupta, S. S., & Groll, P. A. (1961). Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association, 56*, 942–970.
12. Gupta, P. L., Gupta, R. C., & Lvin, S. J. (1996). Analysis of failure time data by Burr distribution. *Communications in Statistics: Theory and Methods, 25*, 2013–2024.
13. Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics, 41*, 173–188.
14. Gupta, R. D., & Kundu, D. (2001). Exponentiated exponential distribution: An alternative to gamma and Weibull distributions. *Biometrical Journal, 43*, 117–130.
15. Gupta, R. D., & Kundu, D. (2003). Discriminating between Weibull and generalized exponential distributions. *Computational Statistics and Data Analysis, 43*, 179–196.
16. Huang, Y. S., Hsieh, C.-H., & Ho, J.-W. (2008). Decisions on an optimal life test sampling plan with warranty considerations. *IEEE Transactions on Reliability, 57*, 643–649.
17. Kantam, R. R. L., & Rosaiah, K. (1998). Half logistic distribution in acceptance sampling based on life tests. *IAPQR Transactions, 23*, 117–125.
18. Kantam, R. R. L., Rosaiah, K., & Rao, G. S. (2001). Acceptance sampling based on life tests: Log-logistic model. *Journal of Applied Statistics, 28*, 121–128.
19. Kundu, D., & Gupta, R. D. (2008). Generalized exponential distribution: Bayesian estimations. *Computational Statistics and Data Analysis, 52*, 1873–1883.
20. Kwon, Y. I. (1996). A Bayesian life test sampling plan for products with Weibull lifetime distribution sold under warranty. *Reliability Engineering and System Safety, 53*, 61–66.
21. Liang, T. C., Chen, L.-S., & Yang, M.-C. (2012). Efficient Bayesian sampling plans for exponential distributions with random censoring. *Journal of Statistical Planning and Inference, 142*, 537–551.
22. Lin, C.-T., Huang, Y.-L., & Balakrishnan, N. (2008). Exact Bayesian variable sampling plans for the exponential distribution based on type-I and type-II censored samples. *Communications in Statistics – Simulation and Computation, 37*, 1101–1116.
23. Lio, Y. L., & Tsai, T.-R. (2012). Estimation of $\delta = P(X < Y)$ for Burr XII distribution based on the progressively first failure-censored samples. *Journal of Applied Statistics, 39*, 309–322.
24. Lio, Y. L., Tsai, T.-R., & Wu, S.-J. (2009). Acceptance sampling plans from truncated life tests based on the Birnbaum-Saunders distribution for percentiles. *Communications in Statistics – Simulation and Computation, 39*, 119–136.
25. Lio, Y. L., Tsai, T.-R., & Wu, S.-J. (2010). Acceptance sampling plans from truncated life tests based on the Burr type XII percentiles. *Journal of the Chinese Institute of Industrial Engineers, 27*, 270–280.
26. Marshall, A. W., Meza, J. C., & Olkin, I. (2001). Can data recognize its parcent distribution. *Journal of Computational and Graphical Statistics, 10*, 555–580.
27. Menke, W. W. (1969). Determination of warranty reserves. *Management Science, 15*, 542–549.

28. Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure data. *IEEE Transactions on Reliability, 42*, 299–302.
29. Murthy, D. N. P., & Blischke, W. R. (2009). Strategic warranty management: A life-cycle approach. *IEEE Transactions on Engineering Management, 47*, 40–54.
30. R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/
31. Singpurwalla, N. D. (2002). Some cracks in the empire of chance flaws in the foundations of reliability. *International Statistical Review, 70*, 53–78.
32. Sobel, M., & Tischendrof, J. A. (1959). Acceptance sampling with sew life test objective. In *Proceedings of fifth national symposium on reliability and quality control, Philadelphia* (pp. 108–118).
33. Soliman, A. A., Abd Ellah, A. H., Abou-Elheggag, N. A., & Modhesh, A. A. (2014). *Estiatmtion from Burr type XII distribution using progressive first-failure censored data* (Vol. 83, pp. 2270–2290).
34. Tadikamalla, P. -R. (1980). A look at the Burr and related distributions. *International Statistical Review, 48*, 373–344.
35. Thomas, M. U., & Rao, S. S. (1999). Warranty economic decision models: A summary and some suggested directions for future research. *Operation Research, 47*, 807–820.
36. Tsai, T.-R. (2008). Bayesian sampling plans with progressive censoring and warranty policy. *International Journal of Reliability, Quality and Safety Engineering, 15*, 329–340.
37. Tsai, T.-R., Lu, Y.-T., & Wu, S.-J. (2008). Reliability sampling plan for Weibull distribution with limited capacity of test facility. *Computers and Industrial Engineering, 55*, 721–728.
38. Tsai, T.-R., Jiang, N., & Lio, Y. L. (2015). Econmonic design of the life test with a warranty policy. *Journal of Industrial and Production Engineering, 32*, 225–231.
39. Tsai, T.-R., Lio, Y. L., Jiang, N., Lin, Y.-J., & Fan, Y.-Y. (2015). Econmonic sampling plans with warranty based on truncated data from Burr type XII distribution. *Journal of the Operational Research Society, 66*, 1511–1518.
40. Tsai, T.-R., & Wu, S.-J. (2006). Acceptance sampling based on truncated life tests for generalized Rayleigh distribution. *Journal of Applied Statistics, 33*, 595–600.
41. Upadhyay, S. K., Javed, I. A., & Peshwani, M. (2004). Bayesian analysis of generalized four-parameter Burr distribution via Gibbs sampler. *METRON, 62*, 115–135.
42. Vintr, Z. (1999). Optimization of the reliability requirements from manufacturer's point of view. In *IEEE 1999 proceeding annual reliability and maintainability symposium* (pp. 183–189).
43. Wingo, D. R. (1983). Maximum likelihood methods for fitting the Burr type XII distribution to life test data. *Biometrical Journal, 25*, 77–84.
44. Wingo, D. R. (1993). Maximum likelihood estimation of Burr XII distribution parameters under type II censoring. *Microelectronis and Reliability, 33*, 1251–1257.
45. Wu, C.-J., & Tsai, T.-R. (2005). Acceptance sampling plans for Birnbaum-Saunders distribution under truncated life tests. *International Journal of Reliability, Quality and Safety Engineering, 12*, 507–519.
46. Wu, J.-W., & Yu, H.-Y. (2005). Statistical inference about the shape parameter of the Burr type XII distribution under the failure-censored sampling plan. *Applied Mathematics and Computation, 163*, 443–482.
47. Wu, S.-J., Chen, Y.-J., & Chang, C.-T. (2007). Statistical inference based on progressively censored samples with random removals from the Burr Type XII distribution. *Journal of Statistical Computation and Simulation, 77*, 19–27.
48. Wu, S. F., Wu, C. C., Chen, Y. L., Yu, Y. R., & Lin, Y. P. (2010). Interval estimation of a two-parameter Burr-XII distribution under progressive censoring. *Statistics, 44*, 77–88.
49. Zimmer, W. J., Keats, J. B., & Wang, F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology, 30*, 386–394.

Design of Reliability Acceptance Sampling Plans Under Partially Accelerated Life Test

M. Kumar

Abstract In this chapter, constant-stress Partially Accelerated Life Tests (PALT) are considered for products with the assumption that the lifetimes of products follow Weibull distribution with known shape parameter and unknown scale parameter. Based on data obtained using Type-II censoring, the maximum likelihood estimates (MLEs) of the Weibull parameters and acceleration factor are obtained assuming linear and Arrhenius relationships with the lifetime characteristics and stress. Exact distributions of the MLEs of the parameters of Weibull distribution are also obtained. Optimal acceptance sampling plans are developed using both linear and Arrhenius relationships. Some numerical results are also presented to illustrate the resulted test plans.

Keywords Reliability · Partially accelerated life test · Linear life-stress model · Arrhenius life-stress model · Optimal test plan

1 Introduction

In life testing, acquiring a test data at a specified normal use condition requires long period of time. This problem makes life testing a difficult, time consuming and costly procedure. Under such circumstances accelerated life tests (ALTs) or partially accelerated life tests (PALTs), which can shorten the lives of test units are used. ALT and PALT differ on the conditions at which they are applied. The test units are run only at accelerated conditions in an ALT, whereas test units are run both at accelerated and normal use conditions in a PALT.

Usage of ALT can often be seen in reliability prediction. Here, in order to induce early failures, specimens are tested at high stress levels. Then through an existing stress dependent model, the failure information is related to specimens at an

M. Kumar (✉)
Department of Mathematics, National Institute of Technology Calicut, Kozhikode, Kerala, India
e-mail: mahesh@nitc.ac.in

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_10

operational stress level. In the absence of such a model, the ALT can't be conducted. In such conditions PALT, which is a combination of both ordinary and accelerated life tests, becomes suitable option. Reliability analysis using PALT helps to save time and money needed in ordinary or traditional life tests. The time at which a test unit can be switched from the standard stress conditions to higher stresses is under the control of the experimenter. This assumption was made by authors in [5]. They also assumed that, in accelerated life testing, it will be possible to choose various levels of higher stresses. Therefore, they restricted themselves to problems in which higher stress levels were fixed in advance.

Bai and Chung [3] studied optimal designs for two partially accelerated life tests (PALTs), in which units were run at both use and accelerated conditions for a predetermined time. In constant PALT, each test unit runs only at either use or accelerated condition and in the step PALT, at a specified time, testing can be changed from use to accelerated condition. For units having exponentially distributed lives, maximum likelihood estimators (MLEs) of failure rates at use condition were obtained. Also, they have obtained the ratios of the failure rates at accelerated condition to that at use condition. The change-time for the step PALT is determined to minimize the generalized asymptotic variance of MLEs of the acceleration factor and the failure rate at use condition. The sample proportion allocated to accelerated condition for the constant PALT is determined to minimize the asymptotic variance of MLE of the acceleration factor. Bai et al. [4] extended the work of Bai and Chung [3] to units having log normally distributed lives. Ismail [8] considers the estimation of parameters of Weibull distribution based on hybrid censored data. The parameters are estimated by the maximum likelihood method under step-stress partially accelerated test. The maximum likelihood estimates (MLEs) of the unknown parameters are obtained by Newton-Raphson algorithm. Also, an approximate Fisher information matrix is obtained for constructing asymptotic confidence bounds for the model parameters. The biases and mean square errors of the maximum likelihood estimators are computed to assess their performances through a Monte Carlo simulation study.

Abd-Elfattah et al. [1] consider step-stress partially accelerated life tests, when the lifetime of a product follows a Burr type XII distribution. Based on Type I censoring, the maximum likelihood estimates (MLEs) are obtained for the distribution parameters and acceleration factor. In addition, asymptotic variance and covariance matrix of the estimators are given. An iterative procedure is used to obtain the estimators numerically. Furthermore, confidence intervals of the estimators are presented. Simulations are carried out to study the precision of MLEs of the parameters involved.

Assuming Weibull distribution as a lifetime model, Ismail and Aly [10] consider optimum plans for failure step-stress partially accelerated life tests with two stress levels under Type-II censoring. The optimum proportion of test units failing at each stress, according to certain optimality criterion, is determined by the optimum test plans. Here, the D-optimality criterion is considered and some numerical results are provided for illustrating the proposed procedure. Ismail [9] discussed maximum likelihood estimators of parameters of Weibull distribution and the acceleration

factor, based on two different types of progressively hybrid censoring schemes under step-stress partially accelerated life test. Using two progressively hybrid censoring schemes, the performances of estimators of the model parameters are evaluated and compared in terms of biases and mean squared errors through a Monte Carlo simulation study.

Aly and Ismail [2] discuss step-stress partially accelerated life tests (PALT) under Type-I censoring, and obtained MLEs of the parameters of a simple (only two stresses) time step-stress model. Further, the confidence intervals for the estimators are constructed. Also, optimum time step-stress test plans are obtained. The optimal stress switching point is obtained from the optimum test plan. These plans minimize the generalized asymptotic variance of the maximum likelihood estimators of the model parameters.

Ismail [7] applied the Bayesian approach to the estimation problem in the case of step stress partially accelerated life tests with two stress levels under Type-I censoring. Here, Gompertz distribution is considered as a lifetime model. The authors have not obtained Bayes estimates in closed forms. Using the method of Lindley [11], approximate Bayes estimates are computed. In addition to this, using Monte Carlo simulation, approximate Bayes estimates are obtained under the assumption of non-informative priors. The results are compared with corresponding maximum likelihood estimates. The optimal design of step-stress partially accelerated life tests (PALTs) under censoring is described by Srivastava and Mittal [12]. It is assumed that the lifetimes of units follow truncated logistic distribution truncated at point zero. This is because, when sample selection is not possible in some sub-region of the sample space, truncated distribution becomes a natural choice. Also, the support of logistic distribution contains negative values (which could conceivably results in modeling negative times-to-failure), it becomes inappropriate for modeling lifetime data. This demands the use of truncated logistic distribution truncated at point zero for modeling lifetime data. The optimal change-time for the step PALT is resolved by minimizing either the generalized asymptotic variance of maximum likelihood estimates (MLEs) of the acceleration factor and the hazard rate at use condition, or the asymptotic variance of MLE of the acceleration factor. Inferential procedure involving model parameters and acceleration factor are studied, and sensitivity analysis is performed as well.

Even though there exists a number of research work related with PALT, the optimal test plan for a product with Weibull lifetime under PALT is not addressed, satisfying the requirements of Type I and Type II errors constraints. The purpose of this chapter is to explore the optimum expected total testing costs incurred in conducting a life test plan for Weibull distribution using linear and Arrhenius life-stress relationships. The test plans are obtained minimizing the expected total testing cost satisfying the constraints of Type I and Type II errors. The exact distribution of the MLE of scale parameter of Weibull distribution is obtained for constructing the reliability acceptance sampling plan.

The rest of the chapter is organized in section wise as follows: Sect. 2 discusses the estimation of Weibull parameters under linear stress-lifetime relationship. An optimal test plan which minimizes the total expected testing cost satisfying the

constraints of Type I and Type II errors, is designed as well. In Sect. 3, estimation of Weibull parameter and optimal test plan are discussed by assuming an Arrhenius life-stress model. Some numerical results are presented in Sect. 4 and concluding remarks are made in Sect. 5.

2 Acceptance Sampling Plan for Weibull Distribution Using Linear Model in PALT

In this section, we derive acceptance sampling plan for Weibull distribution using data from PALT that uses a linear life-stress relation namely, $X = \frac{T}{\lambda}$, where X denotes the lifetime of units under accelerated stress condition, and T is the lifetime under normal stress condition.

2.1 Procedure of PALT

Consider a sample of n independent and identically distributed (iid) units from a lot. Let p be the proportion of units allocated to accelerated stress condition and $1 - p$ be the proportion of units allocated to normal stress condition. Then np is the number of randomly chosen units from n iid units to be allocated to accelerated stress condition and $n(1 - p)$ is the number of units randomly chosen from n units to be allocated to normal stress condition. Each unit under normal stress condition is run until the occurrence of r_1 number of failures and each unit under accelerated stress condition is run until the occurrence of r_2 number of failures.

2.2 Distribution of Lifetime Under Normal Stress Condition

Let T be the lifetime of units under normal stress condition, having Weibull distribution with known shape parameter α and unknown scale parameter θ. Then the probability density function of T is given by

$$f_1(t, \theta, \alpha) = \frac{\alpha}{\theta}\left(\frac{t}{\theta}\right)^{\alpha-1} e^{-(\frac{t}{\theta})^{\alpha}},\ t \geq 0,\ \alpha > 0,\ \theta > 0, \tag{1}$$

where t is the value of the random variable T.

2.3 Distribution of Lifetime Under Accelerated Stress Condition

Let X be the lifetime of units under accelerated stress condition with acceleration factor λ. Assume that, at accelerated stress condition, the lifetime of an unit is $X = \frac{T}{\lambda}$ with cumulative distribution function

$$F_2(X) = P(X \leq x) = P\left(\frac{T}{\lambda} \leq x\right) = P\left(T \leq \lambda x\right) = 1 - e^{-\left(\frac{\lambda x}{\theta}\right)^{\alpha}}, \tag{2}$$

and probability density function

$$f_2(x, \alpha, \lambda, \theta) = \frac{d}{dx}\left(F_2(x)\right) = \frac{\alpha\lambda}{\theta}\left(\frac{\lambda x}{\theta}\right)^{\alpha-1} e^{-\left(\frac{\lambda x}{\theta}\right)^{\alpha}},\ \alpha > 0,\ \theta > 0,\ \lambda > 0. \tag{3}$$

From (2) and (3), it is clear that X follows Weibull distribution with shape parameter α and scale parameter $\frac{\theta}{\lambda}$.

2.4 An MLE of Weibull Parameter Using Transformed Data

In this section, to derive the MLE of Weibull parameter, we use the transformation $Z = T^{\alpha}$. Since the shape parameter α is constant and known in the probability density function of T, the probability density function of Z is given by

$$f_3(z, \theta, \alpha) = \frac{1}{\theta^{\alpha}} e^{-\frac{z}{\theta^{\alpha}}} = \frac{1}{\delta} e^{-\frac{z}{\delta}},\ \delta > 0, \tag{4}$$

where $\delta = \theta^{\alpha}$, and the cumulative density function is

$$F_3(z) = 1 - e^{-\frac{z}{\delta}}. \tag{5}$$

Also, $X = \frac{T}{\lambda}$ implies $X^{\alpha} = \left(\frac{T}{\lambda}\right)^{\alpha} = \frac{T^{\alpha}}{\lambda^{\alpha}}$. Let $Z' = X^{\alpha}$, then the cumulative distribution function of Z' is given by

$$F_4(z') = P\left(Z' \leq z'\right) = P\left(X^{\alpha} \leq z'\right) = P\left(\frac{T^{\alpha}}{\lambda^{\alpha}} \leq z'\right) \tag{6}$$

$$= P\left(T^{\alpha} \leq z'\lambda^{\alpha}\right) = F_3(x\lambda^{\alpha}) = 1 - e^{-\frac{x\lambda^{\alpha}}{\delta}}, \tag{7}$$

and the probability density function of Z' is

$$f_4(z', \theta, \alpha, \lambda) = \frac{\lambda^{\alpha}}{\delta} e^{-\frac{x\lambda^{\alpha}}{\delta}},\ \lambda > 0,\ \delta > 0. \tag{8}$$

That is, $Z' = X^{\alpha}$ follows Exponential distribution with parameter $\frac{\lambda^{\alpha}}{\delta}$ having mean lifetime $\frac{\delta}{\lambda^{\alpha}}$.

Let $T_j,\ j = 1, 2, \cdots, n(1-p)$ be the lifetime of j-th unit under normal stress condition and $X_j,\ j = 1, 2, \cdots, np$ be the lifetime of j-th unit under accelerated stress condition. Then $Z_j = T_j^{\alpha},\ j = 1, 2, \cdots, n(1-p)$ follows Exponential distribution with parameter δ and $Z'_j = X_j^{\alpha},\ j = 1, 2, \cdots, np$ follows exponential distribution with parameter $\frac{\lambda^{\alpha}}{\delta}$. Now, consider the likelihood function for the joint probability density function of $Z'_j,\ j = 1, 2, \cdots, np$ and $Z_j,\ j = 1, 2, \cdots, n(1-p)$ as

$$L = \frac{1}{\delta} e^{-\frac{z_1}{\delta}} \frac{1}{\delta} e^{-\frac{z_2}{\delta}} \cdots \frac{1}{\delta} e^{-\frac{z_{r_1}}{\delta}} \left(e^{-\frac{z_{r_1}}{\delta}}\right)^{n(1-p)-r_1} \frac{\lambda^{\alpha}}{\delta} e^{-\frac{z'_1\lambda^{\alpha}}{\delta}} \frac{\lambda^{\alpha}}{\delta} e^{-\frac{z'_2\lambda^{\alpha}}{\delta}} \cdots$$

$$\frac{\lambda^{\alpha}}{\delta} e^{-\frac{z'_{r_2}\lambda^{\alpha}}{\delta}} \left(\frac{\lambda^{\alpha}}{\delta} e^{-\frac{z'_{r_2}\lambda^{\alpha}}{\delta}}\right)^{np-r_2}.$$

That is,

$$L = \left(\frac{1}{\delta}\right)^{r_1} \left(\frac{\lambda^{\alpha}}{\delta}\right)^{r_2} e^{-\frac{1}{\delta}\sum_{j=1}^{r_1} z_j} e^{-(n(1-p)-r_1)\frac{z_{r_1}}{\delta}} e^{-\frac{\lambda^{\alpha}}{\delta}\sum_{j=1}^{r_2} z'_j} e^{-(np-r_2)\frac{z'_{r_2}\lambda^{\alpha}}{\delta}}.$$

Then the log likelihood function $ln(L)$ is given by

$$ln(L) = -r_1\, ln(\delta) + r_2\, ln(\lambda^{\alpha}) - r_2\, ln(\delta) - \frac{1}{\delta}\sum_{j=1}^{r_1} z_j - (n(1-p)-r_1)\frac{z_{r_1}}{\delta} -$$

$$\frac{\lambda^{\alpha}}{\delta}\sum_{j=1}^{r_2} z'_j - (np-r_2)\frac{z'_{r_2}\lambda^{\alpha}}{\delta}.$$

To find the maximum likelihood estimator of δ and λ^{α}, equate $\frac{\partial ln(L)}{\partial \delta}$ and $\frac{\partial ln(L)}{\partial \lambda}$ to zero. Thus,

$$\frac{\partial ln(L)}{\partial \delta} = 0 \Rightarrow -\frac{r_1}{\delta} - \frac{r_2}{\delta} + \frac{\sum_{j=1}^{r_1} z_j}{\delta^2} + \frac{(n(1-p)-r_1)z_{r_1}}{\delta^2} + \frac{\lambda^{\alpha}\sum_{j=1}^{r_2} z'_j}{\delta^2} + \frac{(np-r_2)z'_{r_2}}{\delta^2} = 0.$$

This implies

$$-(r_1+r_2) + \frac{\sum_{j=1}^{r_1} z_j}{\delta} + \frac{(n(1-p)-r_1)z_{r_1}}{\delta} + \frac{\lambda^{\alpha}\sum_{j=1}^{r_2} z'_j}{\delta} + \frac{(np-r_2)z'_{r_2}}{\delta} = 0.$$

Then, after simplification, we get

$$\hat{\delta} = \frac{\sum_{j=1}^{r_1} z_j + (n(1-p) - r_1) z_{r_1} + \lambda^{\alpha} \left[\sum_{j=1}^{r_2} z'_j + (np - r_2) z'_{r_2} \right]}{r_1 + r_2}.$$

Let $P_1 = \sum_{j=1}^{r_1} z_j + (n(1-p) - r_1) z_{r_1}$ and $P_2 = \sum_{j=1}^{r_2} z'_j + (np - r_2) z'_{r_2}$.
Then, we see that

$$\hat{\delta} = \frac{P_1 + \lambda^{\alpha} P_2}{r_1 + r_2}. \tag{9}$$

Now

$$\frac{\partial ln(L)}{\partial \lambda} = 0 \Rightarrow \frac{\alpha\, r_2}{\lambda} - \frac{\alpha\, \lambda^{\alpha-1} \sum_{j=1}^{r_2} z'_j}{\delta} - \frac{(np - r_2) z'_{r_2} \alpha \lambda^{\alpha-1}}{\delta} = 0.$$

This implies

$$\frac{r_2}{\lambda} - \lambda^{\alpha-1} \left(\frac{\sum_{j=1}^{r_2} z'_j}{\delta} + \frac{(np - r_2) z'_{r_2}}{\delta} \right) = 0 \Rightarrow \frac{r_2}{\lambda} - \lambda^{\alpha-1} \left(\frac{P_2}{\delta} \right) = 0,$$

where

$$P_2 = \sum_{j=1}^{r_2} z'_j + (np - r_2) z'_{r_2}.$$

Then from Eq. (8), we have

$$\lambda^{\alpha} = \frac{r_2\, \delta}{P_2} = \frac{r_2\, P_1 + r_2\, \lambda^{\alpha}\, P_2}{P_2 (r_1 + r_2)} \Rightarrow \lambda^{\alpha} \left(1 - \frac{r_2}{r_1 + r_2} \right) = \frac{r_2\, P_1}{P_2 (r_1 + r_2)}.$$

Thus

$$\hat{\lambda}^{\alpha} = \frac{r_2\, P_1}{r_1\, P_2}. \tag{10}$$

Using the above equation, the maximum likelihood estimator of δ can be rewritten as

$$\hat{\delta} = \frac{P_1 + \lambda^{\alpha} P_2}{r_1 + r_2} = \frac{P_1 + \left(\frac{r_2}{r_1}\frac{P_1}{P_2}\right) P_2}{r_1 + r_2} = \frac{P_1}{r_1}. \tag{11}$$

Since $\theta^{\alpha} = \delta$, we have $\hat{\theta}^{\alpha} = \frac{P_1}{r_1}$.

2.5 The Distribution of $\hat{\delta}$ and $\hat{\lambda}^{\alpha}$

From the above discussions, we have

$$\hat{\delta} = \frac{\sum_{j=1}^{r_1} z_j + (n(1-p) - r_1) z_{r_1}}{r_1} \Rightarrow \hat{\delta}^{\frac{1}{\alpha}} = \left(\frac{\sum_{j=1}^{r_1} z_j + (n(1-p) - r_1) z_{r_1}}{r_1} \right)^{\frac{1}{\alpha}}.$$

Since $\delta = \theta^{\alpha}$ and $\hat{\delta}$ is a maximum likelihood estimator of δ,

$$\hat{\theta} = \hat{\delta}^{\frac{1}{\alpha}} = \left(\frac{\sum_{j=1}^{r_1} z_j + (n(1-p) - r_1) z_{r_1}}{r_1} \right)^{\frac{1}{\alpha}}$$

is a maximum likelihood estimator of $\theta = \delta^{\frac{1}{\alpha}}$. The probability density function of the random variable $Y_1 = \hat{\delta}$ is given by (see [6] for proof)

$$f_5(y_1) = \frac{1}{\Gamma(r_1)} \left(\frac{1}{\delta}\right)^{r_1} y_1^{r_1 - 1} e^{-\frac{r_1 y_1}{\delta}}, \; y_1 > 0. \tag{12}$$

We have $\hat{\theta} = \hat{\delta}^{\frac{1}{\alpha}}$. Let $Y_2 = Y_1^{\frac{1}{\alpha}}$, then the probability density function of $\hat{\theta}$ is given by

$$f_6(y_2) = \frac{\alpha}{\Gamma(r_1)} \left(\frac{r_1}{\delta}\right)^{r_1} y_2^{r_1 \alpha - 1} e^{-\frac{r_1 y_2^{\alpha}}{\delta}}, \; y_2 > 0. \tag{13}$$

Note that one can rewrite

$$\hat{\lambda}^{\alpha} = \frac{r_2}{r_1}\frac{P_1}{P_2} = \frac{\left(\frac{P_1}{r_1}\right)}{\left(\frac{P_2}{r_2}\right)}.$$

From Eq. (11), we have $\frac{P_1}{r_1}$ follows $f_5(y_1)$ with parameter $(\delta,\ r_1)$ and $\frac{P_2}{r_2}$ follows $f_5(y_1)$ with parameter $\left(\frac{\delta}{\lambda^{\alpha}}, r_2\right)$.

A simple straight forward calculation shows that $\frac{2r_1\hat{\delta}}{\delta} \sim \chi^2(2r_1)$ and $\frac{2r_2\left(\frac{\hat{\delta}}{\lambda^{\alpha}}\right)}{\frac{\delta}{\lambda^{\alpha}}} \sim \chi^2(2r_2)$ (see, [6]).

To find the distribution of $\hat{\lambda}^{\alpha} = \frac{\left(\frac{P_1}{r_1}\right)}{\left(\frac{P_2}{r_2}\right)}$, let $W = \hat{\lambda}^{\alpha} = \frac{W_1}{W_2}$, where $W_1 = \hat{\delta} = \frac{P_1}{r_1}$ and $W_2 = \frac{P_2}{r_2}$. The probability density function of W_1 is given by Eq. (11) and the probability density function of W_2 is given by

$$f_7(w_2) = \frac{1}{\Gamma(r_2)}\left(\frac{r_2\lambda^{\alpha}}{\delta}\right)^{r_2} w_2^{r_2-1} e^{-\frac{r_2\lambda^{\alpha}w_2}{\delta}},\ w_2 > 0. \tag{14}$$

Let $f_8(w)$ be the probability density function of W and $F(w)$ be the cumulative distribution function of W. Assume that W_1 and W_2 are independent, then

$$F(w) = P\left(\frac{W_1}{W_2} \le w\right) = P\left(W_1 \le W_2 w\right) = F_{w_1}(w_2 w) = \int_0^{\infty} F_{w_1}(w w_2) f_7(w_2) dw_2,$$

where

$$F_{w_1}(w w_2) = \int_0^{w w_2} \frac{1}{\Gamma(r_1)}\left(\frac{r_1}{\delta}\right)^{r1} w_1^{r_1-1} e^{-\frac{r_1 w_1}{\delta}} dw_1$$

$$= \frac{1}{\Gamma(r_1)} e^{-\frac{r_1 w w_2}{\delta}} \left(\frac{r_1 w w_2}{\delta}\right)^{r_1} \sum_{n=0}^{\infty} \frac{\left(\frac{r_1 w w_2}{\delta}\right)^n}{r_1(r_1+1)\cdots(r_1+n)}.$$

That is,

$$F(w) = \sum_{n=0}^{\infty} \frac{\frac{1}{\Gamma(r_1)}\left(\frac{r_1 w}{\delta}\right)^{r_1} \frac{1}{\Gamma(r_2)}\left(\frac{r_2\lambda^{\alpha}}{\delta}\right)^{r_2}\left(\frac{r_1 w}{\delta}\right)^n}{r_1(r_1+1)\cdots(r_1+n)} \int_0^{\infty} e^{-\frac{r_1 w_2 w}{\delta}} w_2^{r_1+n} w_2^{r_2-1} e^{-\frac{r_2\lambda^{\alpha}w_2}{\delta}} dw_2.$$

This implies

$$F(w) = \sum_{n=0}^{\infty} \frac{\frac{1}{\Gamma(r_1)}\frac{1}{\Gamma(r_2)}\left(\frac{r_1 w}{\delta}\right)^{r_1+n}\left(\frac{r_2\lambda^\alpha}{\delta}\right)^{r_2}}{r_1(r_1+1)\cdots(r_1+n)} \int_0^{\infty} w_2^{r_1+r_2+n-1} e^{-w_2\left(\frac{r_1 w}{\delta}+\frac{r_2\lambda^\alpha}{\delta}\right)} dw_2.$$

Hence

$$F(w) = \sum_{n=0}^{\infty} \frac{\frac{1}{\Gamma(r_1)}\frac{1}{\Gamma(r_2)}\left(\frac{r_1 w}{\delta}\right)^{r_1+n}\left(\frac{r_2\lambda^\alpha}{\delta}\right)^{r_2}}{r_1(r_1+1)\cdots(r_1+n)} \frac{\Gamma(r_1+r_2+n)}{\left(\frac{r_1 w}{\delta}+\frac{r_2\lambda^\alpha}{\delta}\right)^2}. \tag{15}$$

As $w \to \infty$, $F(w) \to \infty$. That is, $F(w)$ is not a cumulative distribution function. Hence our assumption that W_1 and W_2 are independent, is not valid and hence W_1 and W_2 must be dependent. Since the joint density function of W_1 and W_2 is unknown, one cannot get a closed-form expression for the probability density function of W, that is, closed-form expression for probability density function of $\hat{\lambda}^\alpha$ cannot be obtained.

2.6 Optimal Sampling Plan Using Linear Life-Stress Relation

Consider a lot of units having Weibull failure time with probability density function given in Eq. (1). We construct a statistical testing procedure to assess whether the lifetime characteristics $\delta = \theta^\alpha$ adheres to the required level. The proposed acceptance sampling plan is given in following steps.

1. Take a random sample of size n and test np units under accelerated stress condition and $n(1-p)$ units are put under normal stress condition.
2. Under Type-II censoring, observe r_1 failures from $n(1-p)$ units and r_2 failures from np units.
3. From observed data under Type-II censoring, calculate the MLE $\hat{\delta}$ of δ.
4. If $\hat{\delta} \geq k$, where k is a constant to be determined, accept the units in the lot, otherwise reject the lot. Observe that our acceptance rule is based on the fact that, the lot will be accepted only when the mean lifetime under normal stress level exceeds some constant, say k, which is to be determined.

Let δ^* denote the acceptable quality level (AQL) and δ^{**} denote the unacceptable quality level (UQL) of a unit in the lot. The decision on the lot as to accept or reject will be based upon the following probability requirements:

$$P(\text{Reject the lot} \mid \delta \geq \delta^*) \leq \alpha_1, \tag{16}$$

$$P(\text{Accept the lot} \mid \delta \leq \delta^{**}) \leq \alpha_2, \tag{17}$$

where α_1 is the producer's risk and α_2 is the consumer's risk. The unknown quantities of the test plan (p, r_1, r_2, k) are determined by formulating an optimization problem, which minimizes the total expected testing cost (ETC) subjecting to the conditions in (16) and (17). The total cost of testing involves cost associated with the testing time and cost of failed units. Since the testing time is random, we consider expected testing time and which results into the computation of total expected testing cost. Next, we consider the following theorem to proceed further.

Theorem 1 *Let $G(t)$ be the CDF of chi-square distribution with $2r$ degrees of freedom then $G(t)$ can be written as*

$$G(t) = 1 - e^{-t/2} \sum_{j=0}^{r-1} \frac{(t/2)^j}{j!},\ t > 0.$$

Now, we define the acceptance rule for accepting the entire lot as $P_a = P(\hat{\delta} \geq k)$, where k is a constant to be determined. Using $\frac{2r_1\hat{\delta}}{\delta} \sim \chi^2(2r_1)$ and $t = \frac{2r_1 k}{\delta}$ in Theorem 1, we have

$$P_a = P\left(\frac{2r_1\hat{\delta}}{\delta} \geq k\frac{2r_1}{\delta}\right) = 1 - P\left(\frac{2r_1\hat{\delta}}{\delta} \leq k\frac{2r_1}{\delta}\right) = e^{-t/2} \sum_{j=0}^{r-1} \frac{(t/2)^j}{j!}. \quad (18)$$

Let $S_1 = \sum_{i=1}^{r_1} T_i^{\alpha} + (n(1-p) - r_1)\, T_{r_1}^{\alpha}$ be the total time of testing under normal stress condition after transformation, and $S_2 = \sum_{i=1}^{r_2} X_i^{\alpha} + (np - r_2) X_{r_2}^{\alpha}$ be the total time of testing under accelerated stress condition after transformation. Then by Epstein and Sobel [6], one can see that

$$E\,(S_1) = \delta \sum_{i=1}^{r_1} \frac{1}{n(1-p) - i + 1}, \quad (19)$$

$$E\,(S_2) = \frac{\delta}{\lambda^{\alpha}} \sum_{i=1}^{r_2} \frac{1}{np - i + 1}. \quad (20)$$

Hence the total expected testing time using transformed data is

$$E(S_1) + E(S_2) = \delta \left(\sum_{i=1}^{r_1} \frac{1}{n(1-p) - i + 1} + \frac{1}{\lambda^{\alpha}} \sum_{i=1}^{r_2} \frac{1}{np - i + 1} \right). \quad (21)$$

Let C_1 be the cost of testing a unit for unit time and C_2 be the cost of a failed unit, then the total expected testing cost (ETC) involved in conducting the experiment (after transformation) is

$$ETC = (E(S_1) + E(S_2))\, C_1 + (r_1 + r_2)\, C_2.$$

That is (by using above expressions in (21)), the ETC can be written as

$$ETC = \delta \left(\sum_{i=1}^{r_1} \frac{1}{n(1-p) - i + 1} + \frac{1}{\lambda^{\alpha}} \sum_{i=1}^{r_2} \frac{1}{np - i + 1} \right) C_1 + (r_1 + r_2) C_2. \quad (22)$$

Now we formulate an optimization problem which minimizes the total expected testing cost at acceptable quality level δ^*. Hence using (16), (17), (18) and (21), the optimization problem to find (p, r_1, r_2, k) can be written as

$$\min_{p, r_1, r_2, k} \delta^* \left(\sum_{i=1}^{r_1} \frac{1}{n(1-p) - i + 1} + \frac{1}{\lambda^{\alpha}} \sum_{i=1}^{r_2} \frac{1}{np - i + 1} \right) C_1 + (r_1 + r_2)\, C_2$$

such that

$$e^{\frac{-r_1 k}{\delta}} \sum_{j=0}^{r_1 - 1} \frac{\left(\frac{r_1 k}{\delta}\right)^j}{j!} \geq 1 - \alpha_1,\ \delta \geq \delta^*,$$

$$e^{\frac{-r_1 k}{\delta}} \sum_{j=0}^{r_1 - 1} \frac{\left(\frac{r_1 k}{\delta}\right)^j}{j!} \leq \alpha_2,\ \delta \leq \delta^{**}.$$

Since δ is an unknown parameter, we rewrite the above optimization problem as

$$\min_{p, r_1, r_2, k} \delta^* \left(\sum_{i=1}^{r_1} \frac{1}{n(1-p) - i + 1} + \frac{1}{\lambda^{\alpha}} \sum_{i=1}^{r_2} \frac{1}{np - i + 1} \right) C_1 + (r_1 + r_2)\, C_2$$

such that

$$\min_{\delta}\ e^{\frac{-r_1 k}{\delta}} \sum_{j=0}^{r_1 - 1} \frac{\left(\frac{r_1 k}{\delta}\right)^j}{j!} \geq 1 - \alpha_1,\ \delta \geq \delta^*, \quad (23)$$

$$\max_{\delta}\ e^{\frac{-r_1 k}{\delta}} \sum_{j=0}^{r_1 - 1} \frac{\left(\frac{r_1 k}{\delta}\right)^j}{j!} \leq \alpha_2,\ \delta \leq \delta^{**}. \quad (24)$$

Note that as δ increases $e^{\frac{-r_1 k}{\delta}}$ increases. Hence the minimum with respect to δ in (23) occurs at $\delta = \delta^*$ and the maximum with respect to δ in (24) occurs at $\delta = \delta^{**}$. Hence the above optimization problem becomes

$$\min_{p,r_1,r_2,k} \delta^* \left(\sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1} + \frac{1}{\lambda^\alpha} \sum_{i=1}^{r_2} \frac{1}{np-i+1} \right) C_1 + (r_1 + r_2) C_2$$

such that

$$e^{\frac{-r_1 k}{\delta^*}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta^*}\right)^j}{j!} \geq 1 - \alpha_1,$$

$$e^{\frac{-r_1 k}{\delta^{**}}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta^{**}}\right)^j}{j!} \leq \alpha_2.$$

This optimization problem can be solved using genetic algorithm solver in MATLAB.

3 Acceptance Sampling Plan for Weibull Distribution Using Arrhenius Model in PALT

Consider a sample of n independent and identically distributed units from a lot. Let p be the proportion of units allocated to accelerated stress condition and $1 - p$ be the proportion of units allocated to normal stress condition. Then np is the number of randomly chosen units from n iid units to be allocated to accelerated stress condition and $n(1 - p)$ is the number of units randomly chosen from n units to be allocated to normal stress condition. Each unit under normal stress condition is run until occurrence of r_1 number of failures and each unit under accelerated stress condition is run until the occurrence of r_2 number of failures. In this section, we consider the Arrhenius life-stress relationship and which is given by

$$A(\zeta) = a_0 e^{\frac{a_1}{\zeta}}, \tag{25}$$

where A is a quantifiable life measure, ζ is the stress level and $a_0 > 0$, a_1 are the model parameter to be determined.

3.1 Maximum Likelihood Estimation of a_0 and a_1

Let T be the lifetime of a unit under normal stress condition, where the probability density function of T follows Weibull distribution with parameters α and θ_1. Then the probability density function of T is given by

$$f_8(t, \theta_1, \alpha) = \frac{\alpha}{\theta_1}\left(\frac{t}{\theta_1}\right)^{\alpha-1} e^{-\left(\frac{t}{\theta_1}\right)^{\alpha}}, \; t \geq 0, \; \alpha > 0, \; \theta_1 > 0. \tag{26}$$

Let X be the lifetime of a unit under accelerated stress condition and let X follow Weibull probability density function with parameters α and θ_2. Then the probability density function of X is given by

$$f_9(x, \theta_2, \alpha) = \frac{\alpha}{\theta_2}\left(\frac{x}{\theta_2}\right)^{\alpha-1} e^{-\left(\frac{x}{\theta_2}\right)^{\alpha}}, \; x \geq 0, \; \alpha > 0, \; \theta_2 > 0. \tag{27}$$

As considered in Sect. 2.4, use the transformation $Z_i = T_i^{\alpha}$ and $Z_i' = X_i^{\alpha}$, then each Z_i follows $exp(\delta_1)$ and each Z_i' follows $exp(\delta_2)$, where $\delta_1 = \theta_1^{\alpha}$ and $\delta_2 = \theta_2^{\alpha}$.

Let ζ_1 be the normal stress level and ζ_2 be the accelerated stress level. Under Arrhenius life-stress model, assume that

$$\theta_1^{\alpha} = a_0 e^{\frac{a_1}{\zeta_1}} \text{ and } \theta_2^{\alpha} = a_0 e^{\frac{a_1}{\zeta_2}}. \tag{28}$$

The likelihood function obtained from the observed data under normal stress level ζ_1 is given by

$$L_1(z_1, z_2, \cdots, z_{r_1}, \delta_1) \approx \left(\frac{1}{\delta_1}\right)^{r_1} e^{-\frac{S_1}{\delta_1}} = \left(\frac{1}{a_0 e^{\frac{a_1}{\zeta_1}}}\right)^{r_1} e^{-\frac{S_1}{\delta_1}},$$

where $S_1 = \sum_{i=1}^{r_1} z_i + (n(1-p) - r_1)\, z_{r_1}$.

The likelihood function obtained from the observed data under accelerated stress level ζ_2 is given by

$$L_2(z_1', z_2', \cdots, z_{r_2}', \delta_2) \approx \left(\frac{1}{\delta_2}\right)^{r_2} e^{-\frac{S_2}{\delta_2}} = \left(\frac{1}{a_0 e^{\frac{a_1}{\zeta_2}}}\right)^{r_2} e^{-\frac{S_2}{a_0 e^{\frac{a_1}{\zeta_2}}}},$$

where $S_2 = \sum_{i=1}^{r_2} z_i' + (np - r_2) z_{r_2}'$.

Now the joint likelihood function obtained using normal and accelerated stress levels is given by

$$L \approx L_1 L_2 = \left(\frac{1}{\delta_1}\right)^{r_1} e^{-\frac{S_1}{\delta_1}} \left(\frac{1}{\delta_2}\right)^{r_2} e^{-\frac{S_2}{\delta_2}}.$$

Then the log likelihood function is given by

$$\ln L = -r_1 (\ln \delta_1) - \frac{S_1}{\delta_1} - r_2 (\ln \delta_2) - \frac{S_2}{\delta_2}.$$

The normal equations, which are obtained by differentiating L partially with respect to δ_1 and δ_2 respectively, are given by

$$\frac{\partial \ln L}{\partial \delta_1} = 0 \text{ and } \frac{\partial \ln L}{\partial \delta_2} = 0.$$

This implies

$$\frac{-r_1}{\delta_1} + \frac{S_1}{\delta_1^2} = 0 \text{ and } \frac{-r_2}{\delta_2} + \frac{S_2}{\delta_2^2} = 0.$$

Now MLE of δ_1 and δ_2 are given by the following equations:

$$\hat{\delta_1} = \frac{S_1}{r_1} \text{ and } \hat{\delta_2} = \frac{S_2}{r_2}. \tag{29}$$

From (28) we have

$$\hat{a_0} e^{\frac{\hat{a_1}}{\zeta_1}} = \frac{S_1}{r_1} \text{ and } \hat{a_0} e^{\frac{\hat{a_1}}{\zeta_2}} = \frac{S_2}{r_2}.$$

By taking logarithm, we have

$$\ln \hat{a_0} + \frac{\hat{a_1}}{\zeta_1} = \ln\left(\frac{S_1}{r_1}\right), \tag{30}$$

$$\ln \hat{a_0} + \frac{\hat{a_1}}{\zeta_2} = \ln\left(\frac{S_2}{r_2}\right). \tag{31}$$

By subtracting (30) from (31), we have

$$\hat{a_1} = \frac{\zeta_1 \zeta_2}{\zeta_2 - \zeta_1}\left[\ln\left(\frac{S_1}{r_1}\right) - \ln\left(\frac{S_2}{r_2}\right)\right] = \frac{\zeta_1 \zeta_2}{\zeta_2 - \zeta_1} \ln\left(\frac{\hat{\delta_1}}{\hat{\delta_2}}\right). \tag{32}$$

Adding (30) and (31), we get

$$\hat{a_1}\left[\frac{1}{\zeta_1}+\frac{1}{\zeta_2}\right]+2\ln\hat{a_0}=\ln\left(\frac{S_1S_2}{r_1r_2}\right).$$

That is,

$$\frac{\zeta_1\zeta_2}{\zeta_2-\zeta_1}\frac{\zeta_1+\zeta_2}{\zeta_1\zeta_2}\left[\ln\left(\frac{S_1}{r_1}\right)-\ln\left(\frac{S_2}{r_2}\right)\right]+2\ln\hat{a_0}=\ln\left(\frac{S_1S_2}{r_1r_2}\right).$$

This implies

$$\frac{2}{\zeta_1-\zeta_2}\left[\zeta_1\, ln\left(\frac{S_1}{r_1}\right)-\zeta_2\, ln\left(\frac{S_2}{r_2}\right)\right]=2\ln\hat{a_0}.$$

This implies

$$ln\left[\frac{\left(\frac{S_1}{r_1}\right)^{\frac{\zeta_1}{\zeta_1-\zeta_2}}}{\left(\frac{S_2}{r_2}\right)^{\frac{\zeta_2}{\zeta_1-\zeta_2}}}\right]=\ln\hat{a_0}.$$

Hence,

$$\hat{a_0}=\frac{\left(\frac{S_1}{r_1}\right)^{\frac{\zeta_1}{\zeta_1-\zeta_2}}}{\left(\frac{S_2}{r_2}\right)^{\frac{\zeta_2}{\zeta_1-\zeta_2}}}=\frac{\left(\hat{\delta_1}\right)^{\frac{\zeta_1}{\zeta_1-\zeta_2}}}{\left(\hat{\delta_2}\right)^{\frac{\zeta_2}{\zeta_1-\zeta_2}}}. \tag{33}$$

Thus $\hat{a_0}$ and $\hat{a_1}$ represent MLE of a_0 and a_1 respectively, by using invariance property of MLE. Then MLE's of Weibull scale parameters θ_1 and θ_2 can be obtained by using the same invariance property of MLE, by using equations in (28).

Let $h_0=\frac{\zeta_1}{\zeta_1-\zeta_2}$ and $h_1=\frac{\zeta_2}{\zeta_1-\zeta_2}$.

Define $U_1=e^{\frac{\zeta_2-\zeta_1}{\zeta_1\zeta_2}}\left(\hat{a_1}-a_1\right)$ and $U_2=(2r_1)^{h_0}(2r_2)^{-h_1}\frac{\hat{a_0}}{a_0}$. Using (30) and (31), U_1 and U_2 can be rewritten as

$$U_1=\frac{\left(\frac{2S_1}{\delta_1}\right)}{\left(\frac{2S_2}{\delta_2}\right)}\left(\frac{2r_2}{2r_1}\right)\text{ and }U_2=\frac{\left(\frac{2S_1}{\delta_1}\right)^{h_0}}{\left(\frac{2S_2}{\delta_2}\right)^{h_1}},$$

where U_1 and U_2 are pivotal quantities. We state the following theorem which will be useful for obtaining intended result.

Theorem 2 *If S_1 and S_2 are as defined in Sect. 3, then S_1 and S_2 are independent, and the distribution of $\frac{2S_i}{\delta_i} \sim \chi^2(2r_i),\ i = 1, 2.$*

Proof For proof one can refer [6]. □

Next, we give following theorem for obtaining the distribution of U_1 and U_2.

Theorem 3 *The cumulative density functions of the pivotal quantities U_1 and U_2 are given by*

1. $F(u_1, r_1, r_2) = I_{\frac{r_1 u_1}{r_1 u_1 + r_2}}\left(\frac{r_1}{2}, \frac{r_2}{2}\right)$, *where I is the regular incomplete beta function.*
2. $F(u_2, r_1.r_2) = 1 - \int_0^\infty g_1(t) e^{-w/2} \sum_{j=0}^{r-1} \frac{(w/2)^j}{j!} dt.$

Proof Let $V_1 = \frac{2S_1}{\delta_1}$ and $V_2 = \frac{2S_2}{\delta_2}$.

1. The cumulative distribution function of U_1 is obtained as follows.
   Since U_1 is the ratio of two independent Chi-square random variables with degrees of freedom $2r_1$ and $2r_2$, $U_1 \sim F(2r_1, 2r_2)$.
2. The cumulative distribution function of U_2 is obtained as follows.

$$F_{U_2}(u) = P\left(U_2 \le u\right) = P\left(\frac{V_1^{h_0}}{V_2^{h_1}} \le u\right) = P\left(V_1^{h_0} \le u V_2^{h_1}\right) = P\left(V_1 \le (u V_2^{h_1})^{\frac{1}{h_0}}\right)$$

$$= E\left(P\left(V_1 \le (u V_2^{h_1})^{\frac{1}{h_0}}\right) \middle| V_2\right) = \int_0^\infty P\left(V_1 \le (u V_2^{h_1})^{\frac{1}{h_0}} \middle| V_2 = t\right) g_1(t) dt$$

$$= \int_0^\infty G(y) g_1(t) dt = 1 - \int_0^\infty g_1(t) e^{-w/2} \sum_{j=0}^{r-1} \frac{(w/2)^j}{j!} dt,$$

where $G(y)$ is the cumulative distribution function of $\chi^2(2r_1)$, $g_1(t)$ is the probability density function of $\chi^2(2r_2)$ and $w = (u V_2^{h_1})^{\frac{1}{h_0}}$.

□

3.2 Optimal Sampling Plan Using Arrhenius Life-Stress Relation

Consider a lot of units having Weibull failure time with probability density function given in Eq. (1). We construct a statistical testing procedure to assess whether the lifetime characteristics $\delta_1 = {\theta_1}^\alpha$ adheres to the required level. The proposed acceptance sampling plan is stated as follows:

1. Take a random sample of size n and test np units under accelerated stress condition and $n(1-p)$ units are put under normal stress condition.
2. Under Type-II censoring, observe r_1 failures from $n(1-p)$ units and r_2 failures from np units.
3. From observed data under Type-II censoring, calculate the MLE $\hat{\delta_1}$ of δ_1.
4. If $\hat{\delta_1} \geq k$, where k is a constant to be determined, accept the units in the lot, otherwise reject the lot. Observe that our acceptance rule is based on the fact that, the lot will be accepted only when the mean lifetime under normal stress level exceeds some constant, say k, which is to be determined.

Let ${\delta_1}^*$ denote the acceptable quality level (AQL) and ${\delta_1}^{**}$ denote the unacceptable quality level (UQL) of a unit in the lot. The decision on the lot as to accept or reject will be based upon the following probability requirements:

$$P(\text{Reject the lot} \mid \delta_1 \geq {\delta_1}^*) \leq \alpha_1, \tag{34}$$

$$P(\text{Accept the lot} \mid \delta_1 \leq {\delta_1}^{**}) \leq \alpha_2, \tag{35}$$

where α_1 is the producer's risk and α_2 is the consumer's risk. The unknown quantities of the test plan (p, r_1, r_2, k) are determined using an optimization problem, which minimizes the total expected testing cost (ETC) subjecting to conditions in (34) and (35). The ETC that we have considered here is the same as which is defined in Sect. 3.

Hence, the acceptance rule for accepting the entire lot is defined as $P_a = P(\hat{\delta_1} \geq k)$, where k is a constant to be determined later. Using $\frac{2r_1\hat{\delta_1}}{\delta_1} \sim \chi^2(2r_1)$ and $t = \frac{2r_1 k}{\delta_1}$ in Theorem 1, we have

$$P_a = P\left(\frac{2r_1\hat{\delta_1}}{\delta_1} \geq k\frac{2r_1}{\delta_1}\right) = 1 - P\left(\frac{2r_1\hat{\delta_1}}{\delta_1} \leq k\frac{2r_1}{\delta_1}\right) = e^{-t/2}\sum_{j=0}^{r-1}\frac{(t/2)^j}{j!}. \tag{36}$$

Also we have,

$$E\,(S_1) = \delta_1 \sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1}, \tag{37}$$

$$E\,(S_2) = \delta_2 \sum_{i=1}^{r_2} \frac{1}{np-i+1}. \tag{38}$$

Hence the total expected testing time using transformed data is

$$E(S_1) + E(S_2) = \delta_1 \sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1} + \delta_2 \sum_{i=1}^{r_2} \frac{1}{np-i+1}.$$

Let C_1 be the cost of testing a unit for unit time and C_2 be the cost of a failed unit, then the total expected testing cost (ETC) involved in conducting the experiment (after transformation) is

$$ETC = \left(\delta_1 \sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1} + \delta_2 \sum_{i=1}^{r_2} \frac{1}{np-i+1}\right) C_1 + (r_1 + r_2)\, C_2. \quad (39)$$

Since δ_1 and δ_2 are there in the expression for ETC and δ_1 and δ_2 are unknown, we minimize ETC at δ_1^* and δ_2^* (that is, for fixed values of δ_1 and δ_2 at normal and accelerated stress level respectively). Hence using Eqs. (36)–(39), the optimization problem to find (p, r_1, r_2, k) can be written as

$$\min_{p,r_1,r_2,k} \left(\delta_1^* \sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1} + \delta_2^* \sum_{i=1}^{r_2} \frac{1}{np-i+1}\right) C_1 + (r_1 + r_2)\, C_2$$

such that

$$e^{\frac{-r_1 k}{\delta_1}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta_1}\right)^j}{j!} \geq 1-\alpha_1,\ \delta_1 \geq {\delta_1}^*,$$

$$e^{\frac{-r_1 k}{\delta_1}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta_1}\right)^j}{j!} \leq \alpha_2,\ \delta_1 \leq {\delta_1}^{**}.$$

Since δ_1 and δ_2 are unknown parameters, we rewrite the above optimization problem as

$$\min_{p,r_1,r_2,k} \left(\delta_1^* \sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1} + \delta_2^* \sum_{i=1}^{r_2} \frac{1}{np-i+1}\right) C_1 + (r_1 + r_2)\, C_2$$

such that

$$\min_{\delta_1} e^{\frac{-r_1 k}{\delta_1}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta_1}\right)^j}{j!} \geq 1-\alpha_1,\ \delta_1 \geq {\delta_1}^*, \quad (40)$$

$$\max_{\delta_1} e^{\frac{-r_1 k}{\delta_1}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta_1}\right)^j}{j!} \leq \alpha_2,\ \delta_1 \leq {\delta_1}^{**}. \quad (41)$$

Note that as δ_1 increases $e^{\frac{-r_1 k}{\delta_1}}$ increases. Hence the minimum with respect to δ_1 given in (38) occurs at $\delta_1 = \delta_1{}^*$ and the maximum with respect to δ_1 given in (39) occurs at $\delta_1 = \delta_1{}^{**}$. Hence the above optimization problem becomes

$$\min_{p,r_1,r_2,k} \left(\delta_1^* \sum_{i=1}^{r_1} \frac{1}{n(1-p)-i+1} + \delta_2^* \sum_{i=1}^{r_2} \frac{1}{np-i+1} \right) C_1 + (r_1 + r_2)\, C_2$$

such that

$$e^{\frac{-r_1 k}{\delta_1{}^*}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta_1{}^*}\right)^j}{j!} \geq 1 - \alpha_1,$$

$$e^{\frac{-r_1 k}{\delta_1{}^{**}}} \sum_{j=0}^{r_1-1} \frac{\left(\frac{r_1 k}{\delta_1{}^{**}}\right)^j}{j!} \leq \alpha_2.$$

This optimization problem can be solved using genetic algorithm solver in MATLAB.

4 Numerical Results and Discussions

In this section, we present the numerical results obtained for the optimal test plan discussed in Sects. 2 and 3.

First, we consider the case of acceptance sampling plan for Weibull distribution using linear model discussed in Sect. 2. Consider an example, which is included in Table 1 given below. For a set of values, the cost of testing a unit for unit time ($C_1 = 1$), the cost of a failed unit ($C_2 = 1$) and the value of the shape parameter of the Weibull distribution $\alpha = 2$, for the choices of producer's risk $\alpha_1 = 0.1$, consumer's risk $\alpha_2 = 0.1$, acceptable quality level $\delta^* = 900$, unacceptable quality level $\delta^{**} = 200$, number of samples $n = 40$ and for the value of acceleration factor $\lambda = 2$, the optimal values of total expected testing cost (ETC) is 1079.5, the number of failures $r_1 = 5$, $r_2 = 2$ and the lot acceptance constant $k = 329.2345$ are obtained. Thus the test plan is to accept the lot whenever $\hat{\delta}$ exceeds 329.2345, otherwise reject the lot.

Next, we discuss some examples in Table 2 to illustrate acceptance test plan using Arrhenius model. Consider an example, which is included in Table 2. For a set of values, $C_1 = 1$, $C_2 = 1$ and $\alpha = 1$, for $\alpha_1 = 0.05$, $\alpha_2 = 0.05$, acceptable quality level $\delta_1^* = 270$, unacceptable quality level $\delta_1^{**} = 50$, $n = 50$ and the acceleration factors $\zeta_1 = 1$, $\zeta_2 = 2$, the optimal values of total expected testing cost (ETC) is 43.7748, the number of failures $r_1 = 6$, $r_2 = 1$ and the constant $k = 80.1196$ are obtained. Thus the lot will be rejected whenever $\hat{\delta}_1 < 0.801196$.

Table 1 Optimal acceptance sampling plans using linear life-stress model

$(\alpha_1,\ \alpha_2)$	$(\delta^*,\ \delta^{**})$	n	λ	p	$(r_1,\ r_2)$	k	ETC
The values of $k,\ r_1$ and r_2 for various choices of $(\alpha_1,\ \alpha_2),\ (\delta^*,\ \delta^{**}),\ n,\ \lambda, \alpha = 2$ and $C_1 = 1,\ C_2 = 1$							
(0.1, 0.1)	(900, 200)	40	2	0.6554	(5, 2)	329.2345	1079.5
(0.1, 0.05)	(900, 200)	40	2	0.6554	(5, 2)	428.4553	1079.5
(0.05, 0.05)	(900, 200)	40	2	0.6492	(6, 2)	348.8147	1334.3
(0.1, 0.1)	(1600, 400)	50	3	0.7301	(6, 2)	729.1776	3299.3
(0.1, 0.05)	(1600, 400)	50	3	0.7158	(9, 2)	591.6600	4625.9
(0.05, 0.05)	(1600, 400)	50	3	0.7159	(9, 2)	587.5489	4266.0
(0.05, 0.05)	(1800, 400)	30	3	0.6821	(9, 2)	587.1894	9752.2
(0.01, 0.05)	(1800, 400)	30	3	0.7240	(5, 2)	728.7510	5220.8
(0.1, 0.1)	(1800, 400)	30	3	0.7360	(4, 2)	700.3257	4010.1

Table 2 Optimal acceptance sampling plans using Arrhenius life-stress model

$(\alpha_1,\ \alpha_2)$	$(\delta_1^*,\ \delta_1^{**})$	n	δ_2^*	p	$(r_1,\ r_2)$	k	ETC
The values of $k,\ r_1$ and r_2 for various choices of $(\alpha_1,\ \alpha_2),\ (\delta_1^*,\ \delta_1^{**}),\ n,\ \delta_2,\ \alpha = 1,\ \zeta_1 = 1,\ \zeta_2 = 2$ and $C_1 = 1,\ C_2 = 1$							
(0.05, 0.05)	(270, 50)	50	37	0.1217	(6, 1)	80.1196	43.7748
(0.05, 0.05)	(504, 60)	30	42	0.0929	(5, 1)	101.6579	92.3034
(0.1, 0.1)	(1200, 160)	30	60	0.0687	(4, 1)	510.9784	153.5640
(0.05, 0.1)	(1200, 160)	30	60	0.1221	(4, 2)	362.7429	221.3770
(0.05, 0.05)	(1200, 160)	30	60	0.1221	(4, 2)	321.8990	221.3770
(0.05, 0.05)	(1500, 200)	20	67	0.1002	(5, 2)	521.5325	507.1096
(0.1, 0.05)	(1500, 200)	20	67	0.1306	(4, 2)	324.6586	309.4506
(0.1, 0.1)	(1500, 200)	20	67	0.1621	(3, 2)	277.6770	215.2409
(0.1, 0.1)	(2000, 300)	10	90	0.1	(4, 1)	402.5021	722.3354
(0.05, 0.1)	(2000, 300)	10	90	0.2	(5, 2)	403.6784	1172.7

Table 3 Optimal ASP using linear and Arrhenius model with fixed $\alpha_1 = 0.1$

Linear model					Arrhenius model					
α_2	p	r_1, r_2	k	ETC	α_2	p	r_1, r_2	k	ETC	%R
0.1	0.7360	4,1	670.2815	4009.1	0.1	0.0687	4,1	711.891	297.2392	92.5
0.01	0.711	6,2	930.8076	6547.7	0.01	0.0494	6,1	866.5972	444.6027	93.21
0.001	0.6818	8,2	987.26	9661	0.001	0.0463	8,1	982.5218	604.2052	93.74
0.0005	0.7	8,3	1033.4	9699	0.0005	0.0333	9,1	918.0819	684.3313	92.94
0.0001	0.6	10,5	1048.5	13,693	0.0001	0.0334	11,1	959.8846	771.0734	94.36

Table 4 Optimal ASP using linear and Arrhenius model with fixed $\alpha_2 = 0.1$

Linear model					Arrhenius model					
α_1	p	r_1, r_2	k	ETC	α_1	p	r_1, r_2	k	ETC	%R
0.1	0.7364	4,1	671.5725	4009.1	0.1	0.0687	4,1	698.7491	297.2392	92.5
0.01	0.6818	8,3	592.7106	9662	0.01	0.0361	8,1	598.0348	601.471	93.77
0.001	0.6291	11,3	561.2797	16,247	0.001	0.0425	11,1	537.2504	865.8627	94.67
0.0005	0.6	12,3	555.6039	19,408	0.0005	0.0337	12,1	542.9913	957.9362	95.06
0.0001	0.5	15,5	537.7993	42,873	0.0001	0.0346	15,1	540.2972	1281.8	97.01

4.1 Comparative Study and Discussion on Linear and Arrhenius Life-Stress Models

Let $C_1 = 1$, $C_2 = 1$, Weibull shape parameter $\alpha = 2$, acceptable quality level $\delta_1^* = 1800$, unacceptable quality level $\delta_1^{**} = 400$, $n = 40$ (sample size), $\lambda = 3$ (acceleration factor). Table 3 given below presents optimal test plans and associated expected testing costs in linear model and Arrhenius model, for fixed producer's risk (α_1) and various values of consumer's risk (α_2). The last column in the Table (%R) denote the percentage of reduction in testing costs in Arrhenius model as compared to that in linear model. Similar results are presented in Table 4, for fixed consumer's risk (α_2) and various values of producer's risk (α_1).

5 Conclusions

In this chapter, optimal acceptance sampling plans based on data obtained from partially accelerated life test are developed using linear and Arrhenius stress-life relations. Type-II censoring scheme is used to obtain required data. The Maximum likelihood estimates of unknown parameter of Weibull distribution and acceleration factor are obtained for linear model. Similarly, MLEs of model parameters are obtained in case of Arrhenius model as well. Several examples are presented in Tables 1 and 2 to illustrate our optimal acceptance test plans. It is observed that test cost involved in constructing acceptance sampling plan is random in nature. Hence an expression for expected testing cost is given and the same is illustrated through

several examples. It is observed that when the values of α_1 decreases for fixed α_2, expected testing cost increases. The same is true for decreasing values of α_2 with α_1 kept fixed. Also the total expected testing cost is less in Arrhenius model as compared to that in linear model and it is observed that Arrhenius model is more cost effective than linear model. However the actual cost involved in testing may be less than that reported in this work. However, we believe that when testing is carried out, the actual cost may be less than the one (ETC) reported in this work.

Acknowledgements The author would like to express his gratitude to the editor for his constructive comments which improved the presentation of the chapter. The author would also like to thank Mr. Bajeel P N for his computational assistance.

References

1. Abd-Elfattah, A. M., Hassan, A. S., & Nassr, S. G. (2008). Estimation in step-stress partially accelerated life tests for the Burr type XII distribution using type I censoring. *Statistical Methodology, 5*(6), 502–514.
2. Aly, H. M., & Ismail, A. A. (2008). Optimum simple time-step stress plans for partially accelerated life testing with censoring. *Far East Journal of Theoretical Statistics, 24*(2), 175–200.
3. Bai, D. S., & Chung, S. W. (1992). Optimal design of partially accelerated life tests for the exponential distribution under type-I censoring. *IEEE Transactions on Reliability, 41*(3), 400–406.
4. Bai, D. S., Chung, S. W., & Chun, Y. R. (1993). Optimal design of partially accelerated life tests for the lognormal distribution under type I censoring. *Reliability Engineering & System Safety, 40*(1), 85–92.
5. DeGroot, M. H., & Goel, P. K. (1979). Bayesian estimation and optimal designs in partially accelerated life testing. *Navel Research Logistics Quarterly, 26*(2), 223–235.
6. Epstein, B., & Sobel, M. (1953). Life testing. *Journal of the American Statistical Association, 48*(263), 486–502.
7. Ismail, A. A. (2010). Bayes estimation of Gompertz distribution parameters and acceleration factor under partially accelerated life tests with type-I censoring. *Journal of Statistical Computation and Simulation, 80*(11), 1253–1264.
8. Ismail, A. A. (2012). Estimating the parameters of Weibull distribution and the acceleration factor from hybrid partially accelerated life test. *Applied Mathematical Modelling, 36*(7), 2920–2925.
9. Ismail, A. A. (2014). Inference for a step-stress partially accelerated life test model with an adaptive type-II progressively hybrid censored data from Weibull distribution. *Journal of Computational and Applied Mathematics, 260*, 533–542.
10. Ismail, A. A., & Aly, H. M. (2014). Optimal planning of failure-step stress partially accelerated life tests under type-II censoring (Retraction of vol. 80, pg. 1335, 2010). *Journal of Statistical Computation and Simulation, 84*(12), 2759–2759.
11. Lindley, D. V. (1980). Approximate Bayesian methods. *Trabajos Estadistica, 31*, 223–237.
12. Srivastava, P. W., & Mittal, N. (2010). Optimum step-stress partially accelerated life tests for the truncated logistic distribution with censoring. *Applied Mathematical Modelling, 34*(10), 3166–3178.

Part III
Reliability Testing and Designs

Bayesian Sequential Design Based on Dual Objectives for Accelerated Life Tests

Lu Lu, I-Chen Lee, and Yili Hong

Abstract Traditional accelerated life test plans are typically based on optimizing the C-optimality for minimizing the variance of an interested quantile of the lifetime distribution. These methods often rely on some specified planning values for the model parameters, which are usually unknown prior to the actual tests. The ambiguity of the specified parameters can lead to suboptimal designs for optimizing the reliability performance of interest. In this paper, we propose a sequential design strategy for life test plans based on considering dual objectives. In the early stage of the sequential experiment, we suggest allocating more design locations based on optimizing the D-optimality to quickly gain precision in the estimated model parameters. In the later stage of the experiment, we can allocate more observations based on optimizing the C-optimality to maximize the precision of the estimated quantile of the lifetime distribution. We compare the proposed sequential design strategy with existing test plans considering only a single criterion and illustrate the new method with an example on the fatigue testing of polymer composites.

Keywords Bayesian sequential design · C-optimality · D-optimality · Dual objectives · Fatigue test · Polymer composites

L. Lu
Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA
e-mail: lulu1@usf.edu

I.-C. Lee
Department of Statistics, National Cheng Kung University, Tainan, Taiwan
e-mail: iclee@mail.ncku.edu.tw

Y. Hong (✉)
Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
e-mail: yilihong@vt.edu

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_11

1 Introduction

1.1 Background

For many long life products or systems, accelerated lifetime tests (ALTs) [1, 2] are broadly used to accelerate the failure process by exposing the units under harsher conditions than usual and to collect timely information for effectively predicting the lifetime under the normal operating conditions. An ALT plan is often chosen to minimize the anticipated variance of an estimated reliability metric under the normal use condition of interest given an assumed ALT model.

This paper considers an example on planning fatigue tests for polymer composites. Polymer composites are broadly used in many fields of industry such as aircraft, wind turbine, transportation, construction and manufacture, because of their many desirable features such as light weight, high strength, and long-term durability. However, the performance of polymer composites can change after long periods of use due to the fatigue of the materials resulted from being exposed to varied stress levels. Hence fatigue tests of polymer composites aim to assess the material's reliability at some specified normal stress levels. Due to the extreme long life property of polymer composites, the ALTs are desired to acquire failure information in a more timely fashion at much higher stress levels than the normal use conditions.

In a typical ALT setting, n units are tested under some elevated stress conditions. Let x_i represent the stress level at which unit i is tested, and t_i represents the recorded failure time (e.g., the number of cycles for the polymer composite fatigue test) if a failure was observed during the test or the censoring time if unit i had not failed by the end of the test. Then an ALT model is fitted to relate the failure time with the stress level, which is used to predict the product reliability at other stress levels under the normal use conditions. A particular problem of interest for planning an ALT is to determine the stress levels at which individual units will be tested given an affordable sample size. In this case, the number of test units is already determined based on the available resources and time or budget constraints. The goal of the test plan is to choose the stress levels, x_i for $i = 1, \ldots, n$, that offer the best precision of the estimated reliability of interest.

ALT plans have been studied extensively in the past few decades. Most of the methods focus on finding the optimal test plans that maximize some specified utility functions based on the available information. Since the ALT failure models are typically non-linear, the associated information matrices usually depend on the model parameters, and hence the selected optimal designs are dependent on the values of the model parameters used at the planning stage. Given the true values of the model parameters are usually unknown precisely at the planning stage, the resulted ALT plans could be suboptimal for assessing the reliability performance of interest. Bayesian methods have been utilized to leverage prior knowledge on the model parameters through carefully specified prior distributions, which allow the

uncertainty of the prior information to be properly propagated through the statistical inference and offer more flexibility in combining information from various sources.

In addition, the special challenges for polymer composites tests also involve the use of expensive test equipment and the lack of prior information on the performance of new composites materials. Due to the limited test equipment (it is not uncommon that each test laboratory may only have a couple of testing machines available), simultaneously testing multiple samples are usually impractical. As a result, the tests need to be performed sequentially, which offers an opportunity to improve our limited prior knowledge on the model parameters as more data are collected. In this case, it is natural to employ a Bayesian sequential design to select the next design point based on optimizing the expected utility function over the prior distribution based on the current information of the model parameters updated by the observed data, which is captured by the posterior distribution conditioned on the observed failure times from the tested units.

Given the different choices of the utility functions for measuring different aspects of test performance, the sequential Bayesian design strategy can result in different test plans. For example, the commonly used D-optimality criterion focuses on achieving the most precise estimation of the model parameters of an ALT model by maximizing the determinant of the information matrix. While the C-optimality criterion aims to maximize the precision of a linear function of the model parameters by minimizing the its asymptotic variance. In the ALT setting, the C-optimal designs are often used for obtaining the most precise prediction of some interesting quantile of the lifetime distribution under the normal use conditions, which involves extrapolating the stress variable outside the range of observed values (at the elevated stress levels). The optimal designs based on different objectives will push in different directions for allocating test units at different stress levels. Hence, an optimal test plan based on a single criterion may select suboptimal design when other criteria are also of primary interest. This has motivated us to consider multiple objectives in the design selection process.

In this paper, we consider dual objectives in the sequential Bayesian design setting. More specifically at the early stage of the sequential experiment, the sequential runs will be selected based on considering the D-optimality to quickly gain most precision in the estimated model parameters. Then in the later stage, more sequential runs will be selected based on the C-optimality to gain most precision of the predicted quantile of the lifetime distribution at the normal use conditions. It is expected that by considering the dual objectives at different stages of the sequential experiment, we can seek for more balanced performance of design on both the estimation and prediction. In addition, improving the estimation of the model parameters at the early stage of the sequential experiment when little prior information is available can improve our estimation of the anticipated variance of the quantile of the lifetime distribution, and hence results in selecting a more effective test plan for optimizing product reliability at the normal use conditions.

1.2 Related Literature

Traditional non-Bayesian methods for designing an ALT are based on properties of maximum likelihood (ML) estimators. Meeker and Escobar [1] provided a general guideline for planning life tests to obtain precise predictions at the normal use conditions. Several work such as Chernoff [3] and Meeker and Hahn [4, 5] studied optimum designs and compromise plans and outlined practical guidelines for planning an efficient ALT. Some recent developments on ALT include Ye et al. [6], Pan and Yang [7], Tsai et al. [8], and Ng et al. [9]. Escobar and Meeker [2] provided a review on the accelerated test models. Limon et al. [10] provided a review on planning and analysis of the accelerated tests for assessing reliability.

Bayesian techniques have been broadly used in design of experiments. Bayesian methods assume prior distributions for the unknown parameters and the inference of unknown quantities are based on the posterior distribution of the parameters given the observed data. Chaloner and Verdinelli [11] provided a comprehensive review of Bayesian experimental design techniques for linear and non-linear models, among which Bayesian D-optimal designs and Bayesian C-optimal designs are popular choices that have been broadly used in reliability test plans. Zhang and Meeker [12] developed a Bayesian test plan based on minimizing the pre-posterior expectation of the posterior variance over the marginal distribution of all possible test data. Hong et al. [13] proposed new Bayesian criteria based on the large-sample approximations, which offer simplified solutions to the Bayesian test plans.

Sequential test plans have been popular for reliability tests that are either very expensive or very time-consuming. Instead of determining the design locations (i.e., the stress levels) for all test units prior to the experiment, the test in a sequential design is determined and implemented one at a time given the current information gained from the previously observed data. At each step, the next optimal design point is determined by optimizing a design criterion summarized over the posterior distribution of the parameters given all the data observed prior to the current step. It was first introduced in the non-Bayesian framework using the ML estimator with the D- or C-optimality criterion for designing nonlinear experiments [14, 15]. In the Bayesian framework, sequential designs based on considering the D-optimality criterion were used often. For example, Dror and Steinberg [16] developed the Bayesian sequential D-optimal designs for generalized linear models. Roy et al. [17] and Zhu et al. [18] proved convergence properties of the Bayesian sequential D-optimal designs for different forms of models.

In planning for the fatigue tests for polymer composites, King et al. [19] proposed an optimum test plan for the constant amplitude cyclic fatigue testing of polymer composites material. Lee et al. [20] proposed a sequential Bayesian C-optimal test plan for the polymer composites fatigue testing, which selected the sequential design points based on optimizing the posterior asymptotic variance of an quantile life of interest across a range of normal use conditions.

1.3 Overview

The remaining of the paper is organized as follows. Section 2 provides the basic background on the ALT plans. Section 3 discusses the Bayesian sequential design which helps improve the efficiency of the fatigue testing plan when imprecise prior knowledge on the planning parameter values is available. Section 4 proposes a new method based on considering dual objectives in a Bayesian sequential design, where D-optimality is recommended initially to quickly improve the precision of the estimated model parameters in the early runs and followed with C-optimal sequential runs to further improve the precision of the estimated reliability quantile at the specified normal use conditions. The new method will be illustrated in Sect. 5 using the polymer composites fatigue testing example, and compared with the Bayesian sequential designs that consider only a single criterion. At the end, Sect. 6 offers additional discussion and conclusions.

2 Accelerated Life Test Plans

In this section, we give an introduction to the general ALT plans. We focus on the ALT plan with a single accelerating factor such as in the fatigue testing example for polymer composites. The ALTs involve testing units at different stress levels of the accelerating factor for quickly obtaining failure information for the product. Hence, an ALT plan requires the determination of the levels of the accelerating factor to be implemented in the tests (i.e., the stress levels) and the sample size at each level (how many units to be tested at each stress level). Similar to the regular design of experiments, where optimal designs (e.g., D-, A-, G-, or I-optimality [21]) are often chosen for achieving the best precision of quantities of interest (e.g., model parameters or predictions throughout the design space) under an assumed response model, the ALT plans are often based on optimizing the precision of some reliability metric, such as a quantile of the lifetime, given an assumed ALT model. Because the data are collected under the elevated stress levels to make inference on reliability at the normal use conditions, extrapolation on the accelerating factors is naturally involved in ALT plans and hence requires strong assumptions on the specified ALT models.

For modeling, the log-location-scale family of distributions (e.g., the Weibull and Lognormal) is often used to model the accelerated lifetime distribution. The models often assume a common scale parameter but allow the location parameter to change at the different stress levels through a parametric model. An optimal ALT plan is often determined based on minimizing the asymptotic variance of a reliability metric under the assumed model, which is dependent on the model parameters (e.g., the scale and location parameters) that are unknown prior to the data collection. Hence, the efficiency of the ALT plan is largely dependent on the choice of the parameter values at the planning stage.

For the polymer composites fatigue test example, let T denote the cycles to failure which is assumed to follow a log-location-scale family of distribution with the cumulative distribution function (cdf) and the probability density function (pdf) given in the forms below

$$F(t;\boldsymbol{\theta}) = \Phi\left[\frac{\log(t)-\mu}{\nu}\right], \quad \text{and} \quad f(t;\boldsymbol{\theta}) = \frac{1}{\nu t}\phi\left[\frac{\log(t)-\mu}{\nu}\right].$$

In the equations above, μ and ν are the location and scale parameters, respectively. The $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution. Let $\boldsymbol{\theta} = (\mu, \nu)$ denote the vector of the unknown parameters included in the model. The scale parameter ν is often assumed to be constant, while the location parameter $\mu = \mu_{\boldsymbol{\beta}}(x)$ is assumed to be dependent on the stress level x through the model parameters $\boldsymbol{\beta}$. A physically motivated nonlinear model [22] is used to model the relationship between the cycles-to-failure and the stress level as given in the form

$$\mu_{\boldsymbol{\beta}}(x) = \frac{1}{B}\log\left\{\left(\frac{B}{A}\right)h^B\left(\frac{\sigma_{ult}}{x}-1\right)\left(\frac{\sigma_{ult}}{x}\right)^{\gamma(\alpha)-1}[1-\psi(R)]^{-\gamma(\alpha)}+1\right\}. \quad (1)$$

In Eq. (1), the unknown parameters are $\boldsymbol{\beta} = (A, B)$, where A is a model parameter representing the environmental effect on the material fatigue, and B is the material specific effect. The remaining involved variables that are known for the test planning include: the stress ratio, $R = \sigma_m/\sigma_M$, where σ_M and σ_m are the maximum and minimum stresses, the frequency of the cyclic stress testing, h, the ultimate stress of the material, σ_{ult}, and the smallest angle between the testing direction and the fiber direction, α. In addition, functions of known parameters $\psi(R)$ and $\gamma(\alpha)$ are defined as

$$\psi(R) = \begin{cases} R \text{ for } \infty < R < 1 \\ \frac{1}{R} \text{ for } 1 < R < \infty \end{cases},$$

and $\gamma(\alpha) = 1.6 - \psi|\sin(\alpha)|$, respectively. The empirical model in Eq. (1) is flexible to be applied to a wide variety of materials with different settings of stress levels, angles and frequencies. Then, in the assumed ALT model, the unknown parameters are $\boldsymbol{\theta} = (A, B, \nu)$.

In reliability analysis, the quantile of the lifetime in the lower tail of the distribution ($p \leq 0.5$) is often of interest for capturing important reliability characteristics. Let $\zeta_{p,x}$ denote the pth quantile at the stress level x, which is related to the ALT model parameters through the following form

$$\log(\zeta_{p,x}) = \mu_{\boldsymbol{\beta}}(x) + z_p\nu, \quad (2)$$

where z_p is the pth quantile of the standard log-location-scale distribution, and x is a specified stress level under the normal use condition. The pth quantile of the lifetime can be estimated by substituting the model parameters in Eq. (2) by its estimators,

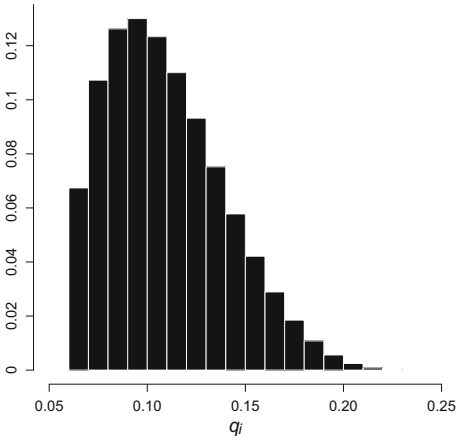


Fig. 1 An illustration of the use stress profile

$\hat{\beta}$ and $\hat{\nu}$. The ALT plan can be chosen based on minimizing the asymptotic variance of the estimated pth quantile life, denoted by AVar $\left[\log\left(\hat{\zeta}_p, x\right)\right]$. Considering in many real applications the stress level can vary across a range of use conditions, a use stress profile shown in Fig. 1 is considered for the polymer composites fatigue test plan. The use condition stress levels range between $x_L = 0.05\sigma_{ult}$ and $x_U = 0.25\sigma_{ult}$. Let $\{x_1, \ldots, x_k\}$ denote all the use stress levels with $x_i = q_i\sigma_{ult}$ for $q_i \in [0.05, 0.25]$ and $\{w_1, \ldots, w_k\}$ denote their relative frequencies in the use profile with $\sum_{k=1}^{K} w_k = 1$. The total weighted asymptotic variance at all use stress levels is defined as

$$\sum_{k=1}^{K} w_k \text{AVar}\left[\log\left(\hat{\zeta}_{p,x_k}\right)\right], \tag{3}$$

for capturing the overall asymptotic variance throughout all possible use conditions.

The ML estimation is often used to calculate the weighted total asymptotic variance given in Eq. (3). The ML approach estimates the parameters based on maximizing the likelihood function of the parameters given the observed data. For the polymer composites fatigue test, the failure time data were collected within a predetermined test duration. Some test units had not failed by the end of the test period, which were censored observations. Let (x_i, t_i, δ_i) represent the observed data for the ith test unit, where δ_i is the censoring indicator and $\delta_i = 1$ if the ith unit is censored and $\delta_i = 0$ otherwise, and t_i is the failure time when $\delta_i = 0$ and the censoring time when $\delta_i = 1$. Given the observed test data $(\mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n)$, where $\mathbf{x}_n = (x_1, \ldots, x_n)'$, $\mathbf{t}_n = (t_1, \ldots, t_n)'$, and $\boldsymbol{\delta}_n = (\delta_1, \ldots, \delta_n)'$, the likelihood function is given by

$$L\left(\boldsymbol{\theta}|\mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n\right) = \prod_{i=1}^{n} \left\{\frac{1}{\nu t_i}\phi\left[\frac{\log(t_i) - \mu_{\boldsymbol{\beta}}(x_i)}{\nu}\right]\right\}^{(1-\delta_i)} \left\{1 - \Phi\left[\frac{\log(t_i) - \mu_{\boldsymbol{\beta}}(x_i)}{\nu}\right]\right\}^{\delta_i}.$$

Let $z_i = \left[\log(t_i) - \mu_{\boldsymbol{\beta}}(x_i)\right]/\nu$, then the log-likelihood function is expressed as

$$l(\boldsymbol{\theta}|\mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n) = \sum_{i=1}^{n} (1-\delta_i)\left[\log\phi(z_i) - \log(\nu) - \log(t_i)\right] + \delta_i \log\left[1 - \Phi(z_i)\right]. \tag{4}$$

The ML estimate of $\boldsymbol{\theta}$ is the solution to the equation $\partial l(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$ and the asymptotic variance of the ML estimator, $\hat{\boldsymbol{\theta}}$ is given by

$$\Sigma_{\boldsymbol{\theta}} = I_n^{-1}(\boldsymbol{\theta}, \mathbf{x_n}) = \left\{E\left[-\frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right\}^{-1}. \tag{5}$$

Then the total weighted asymptotic variance in Eq. (3) can be calculated as

$$\sum_{k=1}^{K} w_k \mathrm{AVar}\left[\log\left(\hat{\zeta}_{p,x_k}\right)\right] = \sum_{k=1}^{K} w_k \mathbf{c}_k' \Sigma_{\boldsymbol{\theta}} \mathbf{c}_k, \tag{6}$$

where $\mathbf{c}_k = \left[\partial\mu_{\boldsymbol{\beta}}(x_k)/\partial A, \partial\mu_{\boldsymbol{\beta}}(x_k)/\partial B, z_p\right]'$. Note that using the ML estimation when there is a small number of observed failures with a high censoring rate can result in unstable estimates. If the censoring time is random, then the expectation-maximization approach can be used to obtain more stable estimates of the model parameters (e.g., Park and Lee [23]).

The C-optimal design [11], which minimizes the asymptotic variance of the estimated quantile of the lifetime distribution at a range of normal use conditions measured by Eq. (6) can be selected as the best ALT plan for determining the number of test units at different stress levels. Note the test plan based on minimizing Eq. (6) should be performed prior to the data collection, and hence heavily depends on the values of the parameters, which are often unknown or at least not known precisely at the planning stage. Therefore, when little prior information was available on the true parameter values, an alternative to the ML approach for estimating the model parameters was proposed by Lee et al. [20] based on using a Bayesian sequential design, which selects each single design point based on optimizing Eq. (6) over the posterior distribution of the parameter $\boldsymbol{\theta}$ given the previously observed data.

3 Sequential Bayesian Design

The sequential design framework used in this paper is similar to Lee at al. [20]. Here we provide a brief description of the general framework within the context of the fatigue test plan for the polymer composites. Given a use stress profile, the $(n+1)$th design point was selected based on minimizing the average posterior asymptotic

variance over the posterior distribution of the model parameters given the first n observations, which is calculated by:

$$\phi\left(x_{n+1}\right)=\int_{\Theta}\left[\sum_{k=1}^{K} w_k \mathbf{c}_k^{\prime} \Sigma_{\boldsymbol{\theta}}\left(x_{n+1}\right) \mathbf{c}_k\right] \pi\left(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n\right) d \boldsymbol{\theta}, \tag{7}$$

where $\Sigma_{\boldsymbol{\theta}}\left(x_{n+1}\right)=\left[I_{n+1}\left(\theta, x_{n+1}\right)\right]^{-1}=\left[I_n\left(\theta, \mathbf{x}_n\right)+I_1\left(\theta, x_{n+1}\right)\right]^{-1}$. Note in Eq. (7), the posterior distribution updated by the first n observations, $\pi\left(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n\right)$, can be considered a prior distribution of parameters prior to selecting the $(n+1)$th design point. A Markov chain Monte Carlo (MCMC) method was developed in Lee et al. [20] to approximate the posterior distribution $\pi\left(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n\right)$, where $\boldsymbol{\theta}=(A, B, \nu)$, for the selection of sequential optimal design points. Uniform distributions were assumed for both the environmental and physical parameters A and B, as in $A \sim \text{Uniform}\left(a_1, a_2\right)$ and $B \sim \text{Uniform}\left(b_1, b_2\right)$, and a_1, a_2, b_1 and b_2 are constants specified based on the anticipated ranges of the parameters A and B. An inverse gamma distribution with the shape parameter κ and the scale parameter γ was used for ν^2 as it is the conjugate prior for the lognormal distribution, which is one of the commonly used log-location-scale distributions in reliability analysis.

In Lee et al. [20], two algorithms were provided to evaluate the anticipated asymptotic variance in Eq. (7) and to select the next optimal design location in the sequential Bayesian design. At each iteration of the sequential Bayesian design, the two algorithms are used (1) to draw samples from the posterior distribution $\pi\left(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{t}_n, \boldsymbol{\delta}_n\right)$ and use the Monte Carlo integration for approximating Eq. (7), and (2) to determine the optimal design location that minimizes Eq. (7) over all possible design points.

To encourage broad applications, an R package named "SeqBayesDesign" [24] was developed to implement the sequential Bayesian design developed in Lee et al. [20]. The package can be applied to determine the sequential Bayesian design for traditional ALTs and the constant amplitude fatigue test based on either the log-normal distribution or the Weibull distribution. Users can easily implement the method by providing the inputs on the historical test data, anticipated use condition levels, the prior information on the model parameters which can be specified in the form of a uniform or normal distribution, and the candidate design points. After specifying the input settings, the optimal design point for the next step will be calculated by the two algorithms.

4 Sequential Design Under Dual Objectives

In this section, we describe a new sequential design strategy based on dual objectives. We first describe the motivation for considering dual objectives. Typically, the optimal design based on a single criterion can have suboptimal performance on aspects that are not measured by the chosen criterion. For example, the D-optimal designs maximize the information gain measured by the determinant

of the information matrix, $|I_n(\theta, \boldsymbol{x}_n)|$. These designs allocate the design points to obtain the most precise estimates of the model parameters based on the data collected at the accelerated stress levels, and may not offer sufficient efficiency when predicting reliability at the normal use conditions which requires extrapolation beyond the range of the observed data. On the other hand, the C-optimal design, which focuses on obtaining the most precise prediction at the specified normal use condition(s), may be suboptimal for precisely estimating the model parameters. In recent decades, it has become more desirable to consider multiple aspects for design selection. The desirability function approach by Derringer and Suich [25] has been broadly used to choose an optimal design based on combining multiple criteria through a user-defined desirability function that heavily relies on subjective user choices on weighting, scaling and metric form prior to understanding the design performance and potential trade-offs. Lu et al. [26] introduced the Pareto front approach to design selection and optimization by intentionally separating the objective and subjective selection stages and allows to make informed decisions based on quantitatively understanding the trade-offs and robustness to different user priorities.

We consider dual objectives in the sequential Bayesian design context. Given the sequential designs are often used in scenarios with small test units and limited prior knowledge on the model parameters, we propose to use the D-optimality at the early stage of the sequential experiment for a quick improvement of the precision of estimated model parameters, and then followed with the C-optimality criterion for a more effective improvement on the precision of the estimated quantile of the lifetime at the normal use conditions. Considering the selection of sequential runs are dependent on the current knowledge of the model parameters (either based on the assumed planning values or from the observed test data), when the planning values or the estimated parameters from limited observations are far off the true values, using the D-optimality for choosing sequential runs is expected to quickly improve the precision of the estimated parameter values. Then with sufficiently precise estimates of model parameters, we have more precise estimate of asymptotic variance of the reliability quantity of interest, and hence can more efficiently allocate additional runs for directly improving the precision of predictions at the normal use conditions. When there are essentially sufficient number of runs collected, the Bayesian sequential design based on considering dual objectives may offer similar performance as the regular optimal designs based on considering a single criterion. However, for situations when only limited number of tests could be afforded, the dual objective sequential test is expected to offer more robustness and balanced performance between the two objectives on improving both the estimation and prediction outside the test region.

More specifically, suppose we want to design an N-run sequential Bayesian test plan and we want to choose the first N_1 runs based on the Bayesian D-optimality and the next $N - N_1$ runs based on the Bayesian C-optimality. Then, among the first N_1 runs, the $(n + 1)$th run at x_{n+1} is selected based on maximizing the expected D-optimality over the posterior distribution of the model parameters based on the first n observed test units as given by

$$\psi\left(x_{n+1}\right)=\int_{\Theta}\log\left\{\left|I_{n+1}\left(\theta,x_{n+1}\right)\right|\right\}\pi\left(\boldsymbol{\theta}|\mathbf{x}_n,\mathbf{t}_n,\boldsymbol{\delta}_n\right)d\boldsymbol{\theta}, \tag{8}$$

where $|I_{n+1}(\theta, x_{n+1})|$ is the determinant of the information matrix based on the first $n+1$ runs which can be updated from the information matrix for the first n runs, $I_n(\theta, x_n)$, by the information gained from the additional observation at x_{n+1} as in

$$I_{n+1}\left(\theta,x_{n+1}\right)=I_n\left(\theta,\mathbf{x}_n\right)+I_1\left(\theta,x_{n+1}\right). \tag{9}$$

Then the optimal design point for the $(n+1)$th run is selected by

$$x_{n+1}^{*}=\text{argmax}_{x_{n+1}\in[x_L,\,x_U]}\psi\left(x_{n+1}\right). \tag{10}$$

Then from the N_1+1 run on, the remaining design locations will be selected based on minimizing the total weighted asymptotic variance of the interested quantile of the lifetime distribution given in Eq. (6) by seeking the optimal design location at $(n+1)$th observation as

$$x_{n+1}^{*}=\text{argmin}_{x_{n+1}\in[x_L,\,x_U]}\phi\left(x_{n+1}\right), \tag{11}$$

until all N runs are selected and executed.

The implementation of the dual objective Bayesian sequential design using the R package "SeqBayesDesign" [24] is convenient. This package allows the users to generate the Bayesian sequential optimal design based on a user choice on the optimization criterion between the D-optimality and C-optimality. By changing the optimization criterion among the sequential runs, we can easily implement the dual objective Bayesian sequential design for any desired fraction of the D-optimal and C-optimal runs. In the next section, we will illustrate the proposed method using the polymer composites fatigue testing example and compare a few Bayesian sequential test plans with different fractions of D-optimal runs to the Bayesian sequential designs considering only a single criterion. The R code for implementing the proposed method for the polymer composites fatigue test example is available from the authors upon request.

5 Application in Polymer Composites Fatigue Testing

In this section, we illustrate the sequential Bayesian design considering dual objectives using the polymer composites fatigue testing example. The original data consisting of 14 observations from a fatigue testing experiment for glass fibers were summarized in Lee et al. [20]. Among the 14 existing observations, there

were 11 failures and 3 right-censored observations. The ML estimates of the model parameters based on the 14 observations are $\hat{A} = 0.00157$, $\hat{B} = 0.3188$ and $\hat{\nu} = 0.7259$, which are used as the best available estimates of model parameters when evaluating the performance of the sequential dual objective Bayesian designs. Other variables that were set as constants for the test plan include $h = 2$, $R = 0.1$, $\alpha = 0$, and $\sigma_{ult} = 1339.67$.

Considering in many polymer composites fatigue testing, the available historical data are often limited, we select only three observations (the minimum number required to estimate all three model parameters) from the 14 observations to represent the limited existing data. First, we chose to evaluate the same subset of three observations that were evaluated in Lee et al. [20], namely the Data Set 1, based on which the ML estimates of model parameters are $\hat{\boldsymbol{\theta}}_1 = (0.0005, 0.7429, 0.1658)$. Figure 2 shows the stress-life relationship for all the 14 observations and the subset of three that are considered as limited existing data in our example. We adapted the Bayesian sequential algorithms developed in Lee et al. [20] for considering the dual objectives and implemented using the R package "SeqBayesDesign" [24].

Given the three selected observations shown in Fig. 2 (solid diamonds), we plan to select 12 additional observations using the Bayesian dual objective sequential design. Five different design strategies are compared which are summarized in Table 1. Due to the sampling variation, we simulated 100 test plans based on each strategy and summarize the average performance across the 100 simulations to compare the design performance. Figure 3 shows the AVar from Eq. (6) of the 12 sequential runs for the five design strategies. We can see for the first sequential run, all five designs substantially reduce the AVar by almost the same amount. Starting from the 2nd run, the sequential C-optimal design (shown in the red line) achieves the minimum AVar consistently for all remaining runs. The sequential D-optimal design (shown in the blue line) increases the AVar at the 2nd run and then starts to reduce the AVar steadily afterwards. However, it has consistently higher AVar than the sequential C-optimal design since it focuses primarily on obtaining

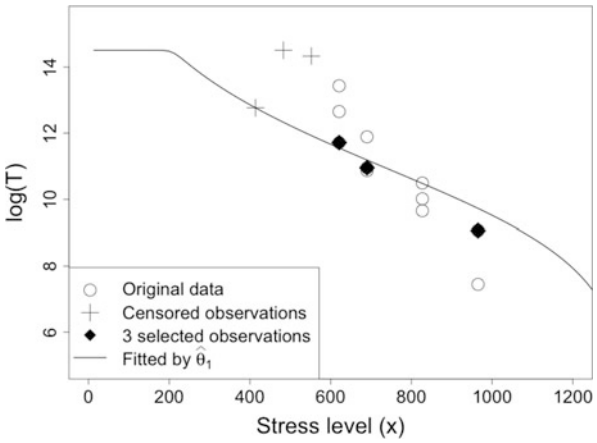


Fig. 2 Plot shows the three selected observations from the original data for Data Set 1 with its fitted stress-life relationship by $\hat{\boldsymbol{\theta}}_1$

Table 1 Five sequential Bayesian designs for comparison

Design scenarios	Description
a: 12 C-opt	All 12 sequential runs are generated based on the Bayesian C-optimality criterion
b: 12 D-opt	All 12 sequential runs are generated based on the Bayesian D-optimality criterion
c: 6 D-opt + 6 C-opt	First 6 runs are based on D-optimality and last 6 runs are based on C-optimality
d: 4 D-opt + 8 C-opt	First 4 runs are based on D-optimality and last 8 runs are based on C-optimality
e: 2 D-opt + 10 C-opt	First 2 runs are based on D-optimality and last 10 runs are based on C-optimality

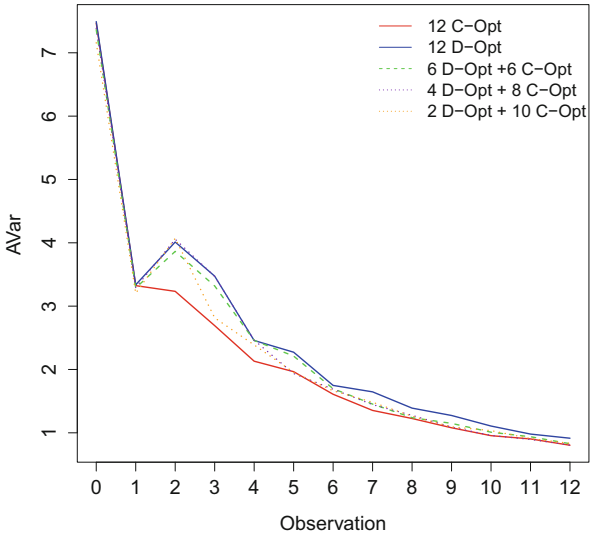


Fig. 3 Plot of AVar, the estimated asymptotic variance of the estimated quantile life averaged over a range of specified use conditions as in Eq. (6), for the five designs from Table 1

the most precise estimates of the model parameters instead of making the most precise prediction of reliability at the normal use conditions. However, the difference between the two designs becomes noticeably smaller as more sequential runs are observed. The other three designs (c–e) from Table 1 achieve AVar values between those offered by the D-optimal and C-optimal designs. Generally, the more D-optimal runs generated at the early stage of the sequential experiment, the more closely the design performs compared to the 12 run sequential D-optimal design for those early runs that are generated based on the D-optimal criterion. But for the later runs generated based on the C-optimal criterion, the design usually offers slightly better AVar than the D-optimal design. On the other hand, the fewer D-optimal runs the design has in the early runs, the faster the AVar value is improved and approaching the best AVar offered by the C-optimal design for all the sequential runs.

In addition, we compare the overall precision of the estimated model parameters among the five design strategies. Particularly, we adopt the M measure for quantifying the relative error of the estimated parameters from Lee et al. [20]. Particularly, we define

$$m\left(\theta_j\right) = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{|\hat{\theta}_{j,k} - \theta_j|}{\theta_j}\right)^2,$$

where θ_j represents the jth parameter in the assumed ALT model, and $\hat{\theta}_{j,k}$ represents the estimated value of θ_j from the kth simulation trial for $k = 1, \ldots, K = 100$. Hence, $m(\theta_j)$ measures the relative mean squared error of the estimate of the jth parameter. Then the M measure, defined as $M = \sum_{j=1}^{3} m\left(\theta_j\right)$, measures the total relative mean squared error of all the parameters, which quantifies the overall precision of the estimated model parameters from the ALT model. The smaller M value indicates more precision of the estimated parameters. Figure 4 shows the M measure for the five designs shown in Table 1. The sequential C-optimal design has the largest M and hence the least precision of all the estimated parameters for all the 12 sequential runs. This is expected as the C-optimal design focuses more on obtaining the most precise estimation of the quantile lifetime of interest at the normal use conditions and less on obtaining the most precise estimation of model parameters. It is interesting that the D-optimal design does not offer the most precise estimation of parameters at the early runs but does catch up later as more runs are implemented. Also, some of the sequential dual objective designs (e.g., design c with 6 D-opt runs followed by 6 C-opt runs) slightly outperform the D-optimal design at early runs (between the 2nd and the 7th sequential runs). This could be resulted from sampling variation and the fact of D-optimality being a better

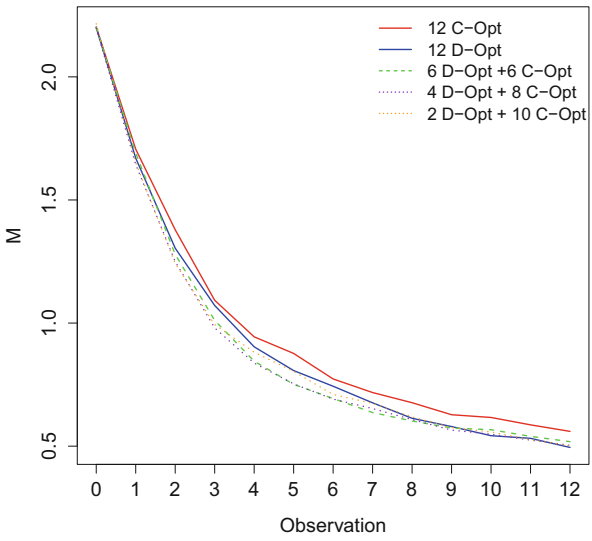


Fig. 4 Comparison of the M measure, the total relative mean squared error for all three model parameters, for the five designs from Table 1

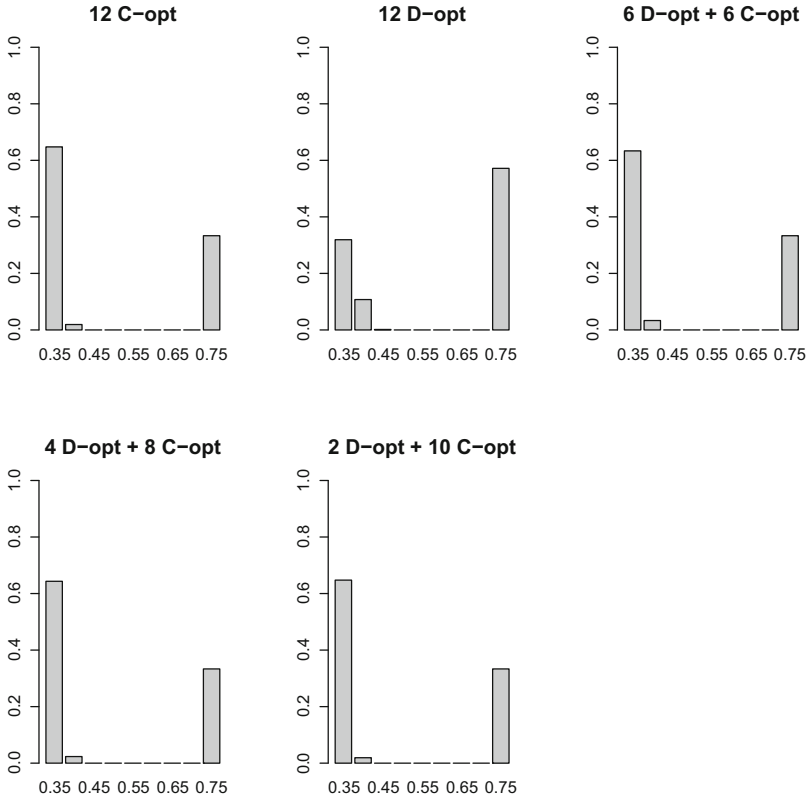


Fig. 5 Plot of sample size allocation for the five designs from Table 1

measure of the precision of estimated model parameters for the larger sample cases. As more sequential runs are obtained (8 or more runs), the D-optimal design offers the best precision of estimated model parameters. For our case study, when less than 8 sequential runs are allowed, the dual objective sequential designs offer slightly better precision for both the estimation of model parameters and the prediction at the normal use conditions.

Figure 5 shows the fraction of sample allocated to different stress levels for the five designs averaged over the 100 simulation trials. A few patterns can be observed prominently. First of all, all the test plans allocate majority of the runs to the extreme stress levels within the range of consideration for $q \in [0.35, 0.75]$. The sequential C-optimal design places about $2/3$ of the total sequential runs to the lowest stress level at $q = 0.35$. In contrast, the sequential D-optimal design places about 60% of the total runs to the highest stress level at $q = 0.75$. This is intuitive as the C-optimal aims for improving the estimation at the normal use conditions at much lower stress levels than the design region, and hence tends to allocation more runs closer to the region of prediction. On the other hand, the D-optimal design aims to improve

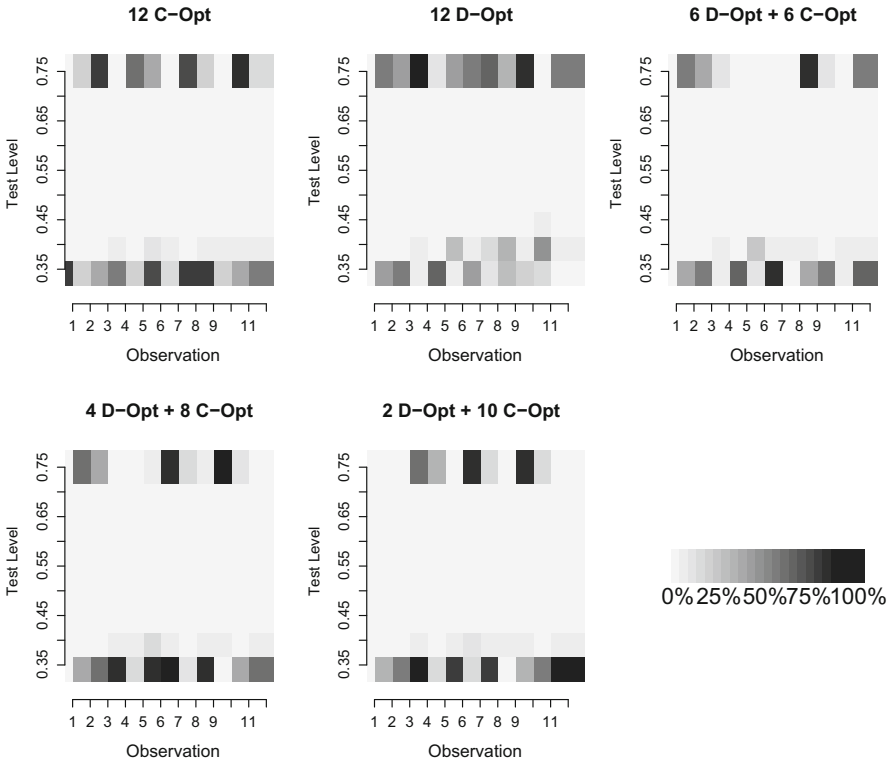


Fig. 6 Plot of the sample size allocation of the 12 sequential runs for the five designs

the estimated model parameters, which are estimated more precisely when more failures are observed at the higher stress levels assuming the failure mechanism does not change under more stressed conditions. Other sequential plans based on considering dual objectives show slightly more balanced performance between the two optimal designs considering only a single criterion. But they are consistently much closer to the C-optimal design by allocating more runs closer to the normal use considerations. For all the test plans, there are only a small number of runs ($\leq 1\%$) located at the second lowest stress level at $q = 0.4$ among the 100 simulation trials.

Figure 6 displays the faction of sample allocation at each of the sequential runs across the 100 simulation trials for the five designs. The dark-to-light gray shades indicate large to small frequency of sample allocation at the different stress levels. We can see there is a large middle region in each panel (for each design) shown in the lightest gray indicating a rare sample allocation to the middle stress levels between $q = 0.4$ and $q = 0.7$. By contrasting the C-optimal and D-optimal designs, we can see while the earlier runs are more evenly split between the high and low stress levels, the D-optimal design seems to allocate more runs at the higher stress levels for the later runs, while the C-optimal design tends to have more runs at the

lower stress levels at the later stage of the sequential experiment. And this pattern matches our understanding that more tests at the higher stress levels can improve the estimation more efficiently while tests at the lower levels are more helpful for improving the prediction at the normal use conditions. We also did the comparison between the several strategies based on other data sets, including the Data Set 2 in Lee et al. [20], and observed similar patterns. Hence, the results are not shown here.

6 Discussion and Conclusions

In this paper, we extend the Bayesian sequential test plans by considering dual objectives when selecting the sequential design points. Bayesian sequential tests are particularly useful when there is little prior information on the lifetime distribution and also tight constraints on the total number of test units or the number of units that can be tested simultaneously. In this case, it is helpful to timely update our understanding as more valuable information is collected during the sequential experiment. We recommend using D-optimality to guide the early design selection for quickly improving the precision of the estimated ALT models to help the planning strategy focus in the right region and then switch to C-optimality to seek precise prediction when extrapolating into the normal use conditions.

We applied the dual objective Bayesian sequential design to the polymer composites fatigue test problem, in which case there is limited prior knowledge on the new composites materials and the testing requires using an expensive equipment with limited availability and hence there is little chance of testing multiple units at the same time. Our comparison between the dual objective and single objective Bayesian sequential designs showed a more robust and balanced performance of the dual objective design. Especially when very limited test units are possibly to be considered, the dual objective designs have shown to offer noticeably more precise estimation of model parameters than the C-optimal design and better prediction across a range of possible normal use conditions than the D-optimal design. As more test units are affordable, the dual objective designs have shown to catch up with the single criterion optimal design more quickly than the optimal design considering only the other criterion and eventually offer near-optimal performance on both the estimation and prediction.

The implementation of the method is convenient and straightforward with the "SeqBayesDesign" R package [24] for the polymer composites fatigue test example and other applications using similar ALT models. However, the proposed method is very general to be applied to broad applications with different lifetime distributions, ALT models, or even different sequential testing strategies involving blocking or other constraints. The method should also adapt for higher dimensional problems with more than one accelerating factors.

Acknowledgements The authors thank the editors and one referee who provided comments that helped us improve this paper. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources. The work by Hong was partially supported by the National Science Foundation under Grant CNS-1565314 to Virginia Tech.

References

1. Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. Wiley, New York.
2. Escobar, L. A., & Meeker, W. Q. (2006). A review of accelerated test models. *Statistical Science, 21*, 552–577.
3. Chernoff, H. (1962). Optimal accelerated life designs for estimation. *Technometrics, 4*, 381–408.
4. Meeker, W. Q. & Hahn, G. J. (1977). Asymptotically optimum over-stress tests to estimate the survival probability at a condition with a low expected failure probability. *Technometrics, 19*, 381–399.
5. Meeker, W. Q., & Hahn, G. J. (1985). *How to plan an accelerated life test: Some practical guidelines*. ASQC Statistics Division, Ames, IA.
6. Ye, Z. S., Hong, Y., & Xie, Y. (2013). How do heterogeneities in operating environments affect field failure predictions and test planning? *The Annals of Applied Statistics, 7*, 2249–2271.
7. Pan, R., & Yang, T. (2014). Design and evaluation of accelerated life testing plans with dual objectives. *Journal of Quality Technology, 46*, 114–126.
8. Tsai, T. R., Jiang, N., & Lio, Y. L. (2015). Economic design of the life test with a warranty policy. *Journal of Industrial and Production Engineering, 32*, 225–231.
9. Ng, H. K. T., Kinaci, I., Kus, C., & Chan, P. S. (2017). Optimal experimental plan for multi-level stress testing with Weibull regression under progressive Type-II extremal censoring. *Communications in Statistics-Simulation and Computation, 46*, 2611–2637.
10. Limon, S., Yadav, O. P., & Liao, H. (2017). A literature review on planning and analysis of accelerated testing for reliability assessment. *Quality and Reliability Engineering International, 33*, 2361–2383.
11. Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science, 10*, 273–304.
12. Zhang, Y., & Meeker, W. Q. (2006). Bayesian methods for planning accelerated life tests. *Technometrics, 48*, 49–60.
13. Hong, Y., King, C., Zhang, Y., & Meeker, W. Q. (2015). Bayesian life test planning for the log-location-scale family of distributions. *Journal of Quality Technology, 47*, 336–350.
14. Wu, J. C. F. (1985). Efficient sequential data with binary data. *Journal of the American Statistical Association, 80*, 974–984.
15. McLeish, D. L., & Tosh, D. (1990). Sequential designs in bioassay. *Biometrics, 46*, 103–116.
16. Dror, H. A., & Steinberg, D. M. (2008). Sequential experimetal designs for generalized linear models. *Journal of the American Statistical Association, 103*, 288–298.
17. Roy, A., Ghosal, S., & Rosenberger, W. R. (2009). Convergence properties of sequential Bayesian D-optimal designs. *Journal of Statistical Planning and Inference, 139*, 425–440.
18. Zhu, L., Dasgupta, T., & Huang, Q. (2014). A D-optimal design for estimation of parameters of an exponential-linear growth curve of nanostructures. *Technometrics, 56*, 432–442.
19. King, C., Hong, Y., DeHart, S. P., DeFeo, P. A., & Pan, R. (2016). Planning fatigue tests for polymer composites. *Journal of Quality Technology, 48*, 227–245.
20. Lee, I.-C., Hong, Y., Tseng, S.-T., & Dasgupta, T. (2018). Sequential Bayesian design for accelerated life tests. *Technometrics, 60*, 472–483. https://doi.org/10.1080/00401706.2018.1437475.
21. Montgemory, D. C. (2017). *Deisng and analysis of experiments*, 9th ed. Wiley, New York.

22. Epparachchi, J. A., & Clausen, P. D. (2003). An empirical model for fatigue behavior prediction of glass fibre-reinforced plastic composites for various stress ratios and test frequencies. *Composites Part A: Applied science and manufacturing, 34*, 313–326.
23. Park, C., & Lee, S. B. (2012). Parameter estimation from censored samples using the Expectation-Maximization algorithm. *Technical Report: arXiv:*1203.3880.
24. SeqBayesDesign: Sequential Bayesian design, 2018, R package version 1.0.
25. Derringer, G., & Suich, R. (1980). Simultaneous optimization of several response variable. *Journal of Quality Technology, 12*, 214–219.
26. Lu, L., Anderson-Cook, C. M., & Robinson, T. (2011). Optimization of designed experiments based on multiple criteria using Pareto frontier. *Technometrics, 53*, 353–365.

The Stress-Strength Models for the Proportional Hazards Family and Proportional Reverse Hazards Family

Bing Xing Wang, Pei Hua Jiang, and Xiaofei Wang

Abstract The stress-strength model has been widely used for reliability design of systems. The reliability of the model is defined as the probability that the strength is larger than the stress. This chapter considers the stress-strength model when both the stress and the strength variables follow the two-parameter proportional hazards family or the proportional reverse hazards family. These two distribution families include many commonly-used distributions, such as the Weibull distribution, the Gompertz distribution, the Kumaraswamy distribution and the generalized exponential distribution, etc. Based on complete samples and record values, we derive the maximum likelihood estimation for the these stress-strength reliability. We also present the generalized confidence intervals for these stress-strength reliability. The simulation results show that the proposed generalized confidence intervals work well.

1 Introduction

The proportional hazards family (PHF) and proportional reverse hazards family (PRHF) are two classes important distribution families. The cumulative distribution functions (CDFs) of the two-parameter PHF and PRHF are given by

$$F(x;\alpha,\beta) = 1 - [1 - G(x;\alpha)]^{\beta},\ x > 0 \tag{1}$$

B. X. Wang (✉)
Department of Statistics, Zhejiang Gongshang University, Hangzhou, China
e-mail: wangbingxing@163.com

P. H. Jiang
School of Mathematics and Physics, Anhui Polytechnic University, Wuhu, China
e-mail: jiangph2017@163.com

X. F. Wang
Department of Statistics, Zhejiang Gongshang University, Hangzhou, China

School of Mathematics and Statistics, Huangshan University, Huangshan, China
e-mail: xiaofei@hsu.edu.cn

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_12

and

$$F(x;\alpha,\beta) = [G(x;\alpha)]^{\beta},\ x > 0, \tag{2}$$

respectively. Here, $G(x;\alpha)$ is a cumulative distribution function dependent only on the parameter α. The two distribution families will be denoted by PHF(α,β) and PRHF(α,β), respectively. The two distribution families include many commonly-used lifetime distributions. Examples of the two-parameter proportional hazards family (1) include the Weibull distribution, the Gompertz distribution, the Kumaraswamy distribution, etc. The generalized exponential distribution and the inverse Weibull distribution are two special cases of the two-parameter proportional reverse hazards family (2).

The stress-strength model is an important reliability model. It was first introduced by Birnbaum [6] and developed by Birnbaum and McCarty [7]. Since then, it has been widely used in many fields [9, 10, 28, etc.]. In the classical stress-strength model, both the stress X_1 and the strength X_2 of the unit are assumed to be random, and the stress-strength reliability is defined as $\delta = P(X_1 < X_2)$.

Inference for δ under different stress and strength distributions has been considered in the literature. When the stress and the strength have the Weibull distributions with the same but unknown shape parameter, McCool [17], Kundu and Gupta [15] considered interval estimation for δ. Krishnamoorthy and Lin [13] proposed the generalized confidence intervals (GCIs) for δ under the Weibull distributions with the same or different shape parameters. Wang and Ye [22] derived the GCIs for the Weibull stress-strength model based on the upper record values. Reiser and Guttman [19] derived the two approximate confidence intervals and an approximate Bayesian probability interval for δ under normal case. Weerahandi and Johnson [26] considered hypotheses testing and interval estimation of δ when the stress and strength variables are independent normally distributed. Guo and Krishnamoorthy [8] gave some new approximate inferential methods for δ when the stress and strength are independent normal variables. Guttman et al. [9] and Aminzadeh [1] considered the estimation for stress-strength models with explanatory variables under normal and exponential distributions, respectively. Kundu and Gupta [14] discussed the estimation for δ under the generalized exponential distributions with the same scale parameter. Baklizi [3] obtained the exact confidence interval and the Bayesian credible interval for δ based on the lower record values under the generalized exponential distributions with the known scale parameters. Wang et al. [25] proposed the GCIs for the generalized exponential stress-strength model with the same or different scale parameters. Surles and Padgett [20] derived the maximum likelihood estimation and the approximate confidence interval of δ for a scaled Burr type X Distribution. Raqab and Kundu [18] compared different estimators of δ for a scaled Burr type X distribution. Lio and Tsai [16] discussed the estimation of δ for Burr XII distribution based on the progressively first failure-censored samples. Wang et al. [24] obtained the GCIs for the stress-strength reliability under the

Kumaraswamy distributions. When $G(x;\alpha)$ is known, Basirat et al. [4, 5] studied the estimation for the PHF stress-strength model based on progressive type II censored samples and record values, respectively. Johnson [11] provided a review of many results for stress-strength model using parametric as well as nonparametric approach. Kotz et al. [12] provided some excellent information on past and current developments in the area.

The reminder of this chapter is organized as follows. Section 2 presents the estimation for the PHF and PRHF stress-strength models based on complete samples. Section 3 derived the maximum likelihood estimation and GCIs for δ based on record values. Finally, Sect. 4 concludes this chapter.

2 Inference for the Stress-Strength Model Based on Complete Samples

Suppose that the stress X_1 and strength X_2 are two independent but not identically distributed random variables with CDF defined in (1). In particular, we assume that $X_i \sim \text{PHF}(\alpha_i, \beta_i)$, $i = 1, 2$. Then the reliability of the stress-strength model is given by

$$\delta = P(X_1 < X_2) = \int_0^\infty F(x;\alpha_1,\beta_1)dF(x;\alpha_2,\beta_2). \tag{3}$$

In particular, when $\alpha_1 = \alpha_2 \hat{=} \alpha$, δ is simplified as

$$\begin{aligned}\delta &= \int_0^\infty \{1-[1-G(x;\alpha)]^{\beta_1}\}d\{1-[1-G(x;\alpha)]^{\beta_2}\} \\ &= \int_0^1 \left[1-(1-u)^{\beta_1/\beta_2}\right]du = \frac{\beta_1}{\beta_1+\beta_2}.\end{aligned}$$

Remark 1 For the PRHF stress-strength model, the expression of the reliability δ are the same as one in (3) of the PHF stress-strength model. When $\alpha_1 = \alpha_2 \hat{=} \alpha$, the reliability δ is given by $\beta_2/(\beta_1+\beta_2)$.

2.1 Maximum Likelihood Estimation

Let $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,n_i})$ be the sample from $\text{PHF}(\alpha_i, \beta_i)$, $i = 1, 2$, and $X_{i,(1)} < X_{i,(2)} < \cdots < X_{i,(n_i)}, i = 1, 2$, be the corresponding order statistics. Then the stress-strength reliability δ can be estimated by the maximum likelihood method.

2.1.1 $\alpha_1 = \alpha_2 \hat{=} \alpha$ Case

In this case, the likelihood function is

$$L(\alpha, \beta_1, \beta_2) = \prod_{i=1}^{2} \prod_{j=1}^{n_i} \beta_i g(x_{i,j}; \alpha)[1 - G(x_{i,j}; \alpha)]^{\beta_i - 1},$$

where $g(x; \alpha)$ is the corresponding density function of $G(x; \alpha)$. Then the maximum likelihood estimates (MLEs) $\widehat{\alpha}, \widehat{\beta}_1, \widehat{\beta}_2$ for the parameters α, β_1, β_2 can be obtained from the following equations.

$$\sum_{i=1}^{2} \sum_{j=1}^{n_i} \frac{1}{g_{ij}} \cdot \frac{\partial g_{ij}}{\partial \alpha} + \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[\frac{n_i}{\sum_{j=1}^{n_i} \log(1 - G_{ij})} + 1 \right] \frac{1}{1 - G_{ij}} \cdot \frac{\partial G_{ij}}{\partial \alpha} = 0, \quad (4)$$

$$\beta_1 = -\frac{n_1}{\sum_{j=1}^{n_1} \log(1 - G_{1j})}, \ \beta_2 = -\frac{n_2}{\sum_{j=1}^{n_2} \log(1 - G_{2j})}, \quad (5)$$

where $g_{ij} = g(x_{i,j}; \alpha)$, $G_{ij} = G(x_{i,j}; \alpha)$. Therefore, the MLE of δ is given by

$$\widehat{\delta} = \frac{\widehat{\beta}_1}{\widehat{\beta}_1 + \widehat{\beta}_2}.$$

Example 1 Weibull distributions with the common shape parameter.

Notice that for the Weibull distribution, $G(x; \alpha) = 1 - \exp(-x^{\alpha})$, $x > 0$, thus the Eqs. (4) and (5) can be written as

$$\frac{1}{\alpha} + \frac{1}{n_1 + n_2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \log(x_{i,j}) - \sum_{i=1}^{2} \frac{n_i}{n_1 + n_2} \cdot \frac{\sum_{j=1}^{n_i} x_{i,j}^{\alpha} \log(x_{i,j})}{\sum_{j=1}^{n_i} x_{i,j}^{\alpha}} = 0, \quad (6)$$

$$\beta_1 = \frac{n_1}{\sum_{j=1}^{n_1} x_{1,j}^{\alpha}}, \ \beta_2 = \frac{n_2}{\sum_{j=1}^{n_2} x_{2,j}^{\alpha}}.$$

Therefore, the MLE of δ is

$$\widehat{\delta} = \frac{n_1 / \sum_{j=1}^{n_1} x_{1,j}^{\widehat{\alpha}}}{n_1 / \sum_{j=1}^{n_1} x_{1,j}^{\widehat{\alpha}} + n_2 / \sum_{j=1}^{n_2} x_{2,j}^{\widehat{\alpha}}},$$

where $\widehat{\alpha}$ is the solution of the Eq. (6).

Example 2 Kumaraswamy distributions with the common shape parameter α.

Suppose that X_i follows the Kumaraswamy distribution Kum(α_i, β_i). Its CDF is given by

$$F(x; \alpha_i, \beta_i) = 1 - (1 - x^{\alpha_i})^{\beta_i},\ 0 < x < 1,$$

where $\alpha_i (> 0), \beta_i (> 0)$ are the shape parameters.

Notice that for the Kumaraswamy distribution, $G(x; \alpha) = x^{\alpha}, 0 < x < 1$, thus when $\alpha_1 = \alpha_2 \hat{=} \alpha$, the Eqs. (4) and (5) can be written as

$$\frac{1}{\alpha} + \frac{1}{n_1 + n_2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left\{ \log x_{i,j} + \frac{n_i\ x_{i,j}^{\alpha} \log x_{i,j}}{\sum_{j=1}^{n_i} \log(1 - x_{i,j}^{\alpha})} \right\} \frac{1}{1 - x_{i,j}^{\alpha}} = 0, \quad (7)$$

$$\beta_1 = -\frac{n_1}{\sum_{j=1}^{n_1} \log(1 - x_{1,j}^{\alpha})},\ \beta_2 = -\frac{n_2}{\sum_{j=1}^{n_2} \log(1 - x_{2,j}^{\alpha})}.$$

Therefore, the MLE of δ is

$$\widehat{\delta} = \frac{n_1 / \sum_{j=1}^{n_1} \log(1 - x_{1,j}^{\widehat{\alpha}})}{n_1 / \sum_{j=1}^{n_1} \log(1 - x_{1,j}^{\widehat{\alpha}}) + n_2 / \sum_{j=1}^{n_2} \log(1 - x_{2,j}^{\widehat{\alpha}})},$$

where $\widehat{\alpha}$ is the solution of the Eq. (7).

2.1.2 $\alpha_1 \neq \alpha_2$ Case

In this case, the likelihood function is

$$L(\alpha_i, \beta_i) = \prod_{j=1}^{n_i} \beta_i\, g(x_{i,j}; \alpha_i)[1 - G(x_{i,j}; \alpha_i)]^{\beta_i - 1},\ i = 1, 2.$$

Then the MLEs $\widetilde{\alpha}_i, \widetilde{\beta}_i$ for the parameters α_i, β_i can be obtained from the following equations.

$$\sum_{j=1}^{n_i} \frac{1}{g_{ij}} \cdot \frac{\partial g_{ij}}{\partial \alpha_i} + \sum_{j=1}^{n_i} \left[\frac{n_i}{\sum_{j=1}^{n_i} \log(1 - G_{ij})} + 1 \right] \frac{1}{1 - G_{ij}} \cdot \frac{\partial G_{ij}}{\partial \alpha_i} = 0, \quad (8)$$

$$\beta_i = -\frac{n_i}{\sum_{j=1}^{n_i} \log(1 - G_{ij})},\ i = 1, 2, \quad (9)$$

where $g_{ij} = g(x_{i,j}; \alpha_i)$, $G_{ij} = G(x_{i,j}; \alpha_i)$. Therefore, the MLE for δ is given by

$$\widetilde{\delta} = \int_0^{\infty} F(x; \widetilde{\alpha}_1, \widetilde{\beta}_1) dF(x; \widetilde{\alpha}_2, \widetilde{\beta}_2).$$

Example 3 Weibull distributions with the unequal shape parameters.

In this case, the Eqs. (8) and (9) can be written as

$$\frac{1}{\alpha_i} + \frac{1}{n_i}\sum_{j=1}^{n_i}\log(x_{i,j}) - \frac{\sum_{j=1}^{n_i} x_{i,j}^{\alpha}\log(x_{i,j})}{\sum_{j=1}^{n_i} x_{i,j}^{\alpha}} = 0, \tag{10}$$

$$\beta_i = \frac{n_i}{\sum_{j=1}^{n_i} x_{i,j}^{\alpha}},\ i = 1, 2. \tag{11}$$

Notice that when the two Weibull shape parameters are unequal, δ is given by

$$\delta = 1 - \beta_2 \int_0^{\infty} \exp(-\beta_1 x^{\alpha_1/\alpha_2} - \beta_2 x)dx,$$

then the MLE of δ is

$$\widetilde{\delta} = 1 - \widetilde{\beta}_2 \int_0^{\infty} \exp(-\widetilde{\beta}_1 x^{\widetilde{\alpha}_1/\widetilde{\alpha}_2} - \widetilde{\beta}_2 x)dx,$$

where $\widetilde{\alpha}_i$, $\widetilde{\beta}_i$ are the solutions of the Eqs. (10) and (11), respectively.

Example 4 Kumaraswamy distributions with $\alpha_1 \neq \alpha_2$.

In this case, the Eqs. (8) and (9) can be written as

$$\frac{1}{\alpha_i} + \frac{1}{n_i}\sum_{j=1}^{n_i}\left\{\log x_{i,j} + \frac{n_i\, x_{i,j}^{\alpha_i}\log x_{i,j}}{\sum_{j=1}^{n_i}\log(1 - x_{i,j}^{\alpha_i})}\right\}\frac{1}{1 - x_{i,j}^{\alpha_i}} = 0, \tag{12}$$

$$\beta_i = -\frac{n_i}{\sum_{j=1}^{n_i}\log(1 - x_{i,j}^{\alpha_i})},\ i = 1, 2. \tag{13}$$

Notice that when $\alpha_1 \neq \alpha_2$, δ is given by

$$\delta = 1 - \int_0^1 [1 - (1 - x^{1/\beta_2})^{\alpha_1/\alpha_2}]^{\beta_1} dx,$$

then the MLE of δ is

$$\widetilde{\delta} = 1 - \int_0^1 [1 - (1 - x^{1/\widetilde{\beta}_2})^{\widetilde{\alpha}_1/\widetilde{\alpha}_2}]^{\widetilde{\beta}_1} dx,$$

where $\widetilde{\alpha}_i$, $\widetilde{\beta}_i$ are the solutions of the Eqs. (12) and (13), respectively.

2.2 Interval Estimation

In this subsection, we derive the GCIs for δ under both $\alpha_1 = \alpha_2$ and $\alpha_1 \neq \alpha_2$ scenarios. In order to derive the GCIs of δ, we need the following lemma (see Wang et al. [21]).

Lemma 1 *Suppose that $Z_1, Z_2, \ldots, Z_n$ is a random sample from the exponential distribution with mean θ. Let $S_i = \sum_{j=1}^{i} Z_j$, $i = 1, \ldots, n$, and $T = 2\sum_{i=1}^{n-1} \log(S_n/S_i)$. Then (1) T and S_n are independent; (2) $T \sim \chi^2(2n-2)$, $2S_n/\theta \sim \chi^2(2n)$.*

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be the sample from PHF(α, β) and $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the corresponding order statistics. Then

$$-\log(1 - F(X_{(j)}; \alpha, \beta)) = -\beta \log[1 - G(X_{(j)}; \alpha)],\ j = 1, 2, \ldots, n$$

are the order statistics from the standard exponential distribution.

Let

$$S_j(\alpha, \mathbf{X}) = -\sum_{k=1}^{j} \log[1 - G(X_{(k)}; \alpha_i)] - (n_i - j)\log[1 - G(X_{(j)}; \alpha_i)],$$

$$U(\alpha, \mathbf{X}) = 2\sum_{j=1}^{n-1} \log \frac{S_n(\alpha, \mathbf{X})}{S_j(\alpha, \mathbf{X})},\ \ V(\alpha, \beta, \mathbf{X}) = 2\beta S_n(\alpha, \mathbf{X}).$$

Then we have from Lemma 1 that $U(\alpha, \mathbf{X})$, $V(\alpha, \beta, \mathbf{X})$ are mutually independent, and that $U(\alpha, \mathbf{X}) \sim \chi^2(2n-2)$, $V(\alpha, \beta, \mathbf{X}) \sim \chi^2(2n)$.

2.2.1 $\alpha_1 = \alpha_2 \hat{=} \alpha$ Case

In this case, we first construct the pivotal quantities of α, β_1 and β_2 in order to derive the generalized pivotal quantities (GPQ) of δ.

Let $W(\alpha) = U(\alpha, \mathbf{X}_1) + U(\alpha, \mathbf{X}_2)$. Then $W(\alpha) \sim \chi^2(2n_1 + 2n_2 - 4)$. Notice that $W(\alpha)$ depends only on the parameter α, thus $W(\alpha)$ is a pivotal quantity. When $W(\alpha)$ is the strictly monotonic function of α, the equation $W(\alpha) = t$ has the unique solution for $t > 0$. This solution is denoted by $h_1(t, \mathbf{X}_1, \mathbf{X}_2)$. The monotonicity of $W(\alpha)$ can be verified on a case by case basis. In addition, from $V_i = 2\beta_i S_{n_i}(\alpha, \mathbf{X}_i)$, we have $\beta_i = V_i/(2S_{n_i}(\alpha, \mathbf{X}_i))$. According to the substitution method given by Weerahandi [27], the GPQ for δ is given by

$$Y_1 = \frac{V_1/S_{n_1}(h_1(W_0, \mathbf{X}_1, \mathbf{X}_2), \mathbf{X}_1)}{V_1/S_{n_1}(h_1(W_0, \mathbf{X}_1, \mathbf{X}_2), \mathbf{X}_1) + V_2/S_{n_2}(h_1(W_0, \mathbf{X}_1, \mathbf{X}_2), \mathbf{X}_2)}, \tag{14}$$

where $W_0 \sim \chi^2(2n_1 + 2n_2 - 4)$, $V_i \sim \chi^2(2n_i)$, $i = 1, 2$.

If $Y_{1,\gamma}$ denotes the γ percentile of Y_1, then $[Y_{1,\gamma/2}, Y_{1,1-\gamma/2}]$ is a $1-\gamma$ GCI for δ. The percentiles of Y_1 can be obtained by the following Monte Carlo algorithm.

Algorithm 1 (The Percentiles for Y_1)

(1) For a given data set $(n_1, n_2, \mathbf{X}_1, \mathbf{X}_2)$, generate $W_0 \sim \chi^2(2n_1+2n_2-4)$ and $V_i \sim \chi^2(2n_i)$, $i = 1, 2$, independently. Using these values, compute $h_1(W_0, \mathbf{X}_1, \mathbf{X}_2)$ from the equation $W(\alpha) = W_0$.
(2) Compute value of Y_1 using the Eq. (14).
(3) Repeat the steps 1-2 $B(\geq 10000)$ times. Then there are the B values of Y_1.
(4) Arrange all Y_1 values in ascending order: $Y_{1,(1)} < Y_{1,(2)} < \ldots < Y_{1,(B)}$. Then $Y_{1,\gamma}$ can be estimated by $Y_{1,(B\gamma)}$.

2.2.2 $\alpha_1 \neq \alpha_2$ Case

In this case, let $h_2(t, \mathbf{X})$ be the solution of the equation $U(\alpha, \mathbf{X}) = t$. Using the substitution method, the GPQ for δ is given by

$$Y_2 = \int_0^\infty F(x; h_2(W_{0,1}, \mathbf{X}_1), Y_{0,1}) dF(x; h_2(W_{0,2}, \mathbf{X}_2), Y_{0,2}), \tag{15}$$

where $Y_{0,i} = V_i/(2S_{n_i}(h_2(W_{0,i}, \mathbf{X}_i), \mathbf{X}_i)$, $W_{0,i} \sim \chi^2(2n_i - 2)$ and $V_i \sim \chi^2(2n_i)$, $i = 1, 2$.

If $Y_{2,\gamma}$ denotes the γ percentile of Y_2, then $[Y_{2,\gamma/2}, Y_{2,1-\gamma/2}]$ is a $1-\gamma$ GCI for δ. Similarly, the value $Y_{2,\gamma}$ can be obtained by Monte Carlo simulation. Some simulation results show that under some distributions, the coverage probabilities of the proposed generalized confidence lower limit (GCLL) based on the GPQ (15) are larger than the nominal coverage probabilities. The following modified GPQ (MGPQ) is proposed to improve the performance of the GCI for these distributions. The MGPQ is given by

$$Y_3 = |\log[(1 + Y_2)/(1 - Y_2)] - RZ|, \tag{16}$$

where $RZ = \log[(1 + \widetilde{\delta})/(1 - \widetilde{\delta})]$.

If $Y_{3,\gamma}$ denotes the γ percentile of Y_3, then a $1-\gamma$ modified GCI for the reliability δ is given by

$$\left[\frac{e^{RZ-Y_{3,1-\gamma}} - 1}{e^{RZ-Y_{3,1-\gamma}} + 1}, \frac{e^{RZ+Y_{3,1-\gamma}} - 1}{e^{RZ+Y_{3,1-\gamma}} + 1}\right],$$

and $(e^{RZ-Y_{3,1-2\gamma}} - 1)/(e^{RZ-Y_{3,1-2\gamma}} + 1)$ is a $1-\gamma$ modified GCLL for the reliability δ. Similarly, the value $Y_{3,\gamma}$ can be obtained by Monte Carlo simulation.

Algorithm 2 (The Percentiles for Y_2 and Y_3)

(1) For a given data set $(n_1, n_2, \mathbf{X}_1, \mathbf{X}_2)$, compute the MLEs $\widetilde{\alpha}_i$ and $\widetilde{\beta}_i$, $i = 1, 2$. Then compute the value of $\widetilde{R}$.
(2) Generate $W_{0,i} \sim \chi^2(2n_i - 2)$. Then compute $h_2(W_{0,i}, \mathbf{X}_i)$ from the equation $U(\alpha_i) = W_{0,i}$, $i = 1, 2$.
(3) Generate $V_i \sim \chi^2(2n_i)$, $i = 1, 2$. Then compute Y_2 and Y_3 on the basis of the Eqs. (15) and (16).
(4) Repeat the steps (1) and (2) B times. Then there are the B values of Y_2 and Y_3, respectively.
(5) Arrange all Y_i values in ascending order: $Y_{i,(1)} < Y_{i,(2)} < \ldots < Y_{i,(B)}$, $i = 2, 3$. Then the γ percentile of Y_i can be estimated by $Y_{i,(B\gamma)}$.

Remark 2 The proposed procedures can been extended to the proportional reverse hazards family situation. The details are given as follows.

Let $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,n_i})$ be the sample from PRHF(α_i, β_i), $i = 1, 2$, and $X_{i,(1)} < X_{i,(2)} < \cdots < X_{i,(n_i)}$, $i = 1, 2$, be the corresponding order statistics. Further, let

$$S_j(\alpha_i, \mathbf{X}_i) = -\sum_{k=1}^{j} \log G(X_{i,(n_i-k+1)}; \alpha_i) - (n_i - j) \log G(X_{i,(n_i-j+1)}; \alpha_i),$$

$$U(\alpha_i, \mathbf{X}_i) = 2 \sum_{j=1}^{n_i-1} \log[S_{n_i}(\alpha_i, \mathbf{X}_i)/S_j(\alpha_i, \mathbf{X}_i)], \; V_i = 2\beta_i S_{n_i}(\alpha_i, \mathbf{X}_i), i = 1, 2.$$

Notice that

$$Z_i = -\log F(X_{i,(n_i-j+1)}; \alpha_i, \beta_i) = -\beta_i \log G(X_{i,(n_i-j+1)}; \alpha_i), \; j = 1, 2, \ldots, n_i$$

are the order statistics from the standard exponential distribution, thus $U(\alpha_i, \mathbf{X}_i)$ and V_i are mutually independent, and $U(\alpha_i, \mathbf{X}_i) \sim \chi^2(2n_i - 2)$ and $V_i \sim \chi^2(2n_i)$. The remaining steps are the same as those of the PHF stress-strength model.

Remark 3 Similar to the discussion in Wang et al. [21], the proposed procedures can be extended to the progressively censored samples situation.

Remark 4 The proposed MGPQ has been successfully used in the generalized exponential distribution situation ([25]).

Example 5 Weibull distributions with the equal or unequal shape parameters.

In the Weibull distribution case, $S_j(\alpha, \mathbf{X}) = \sum_{k=1}^{j} X_{(k)}^{\alpha} + (n - j)X_{(j)}^{\alpha}$. It is obvious that $U(\alpha, \mathbf{X})$ is the strictly increasing function of α and that $\lim_{\alpha \to 0+} U(\alpha, \mathbf{X}) = 0$, $\lim_{\alpha \to \infty} U(\alpha, \mathbf{X}) = \infty$. Thus the equations $W(\alpha) = t$ and $U(\alpha, \mathbf{X}_i) = t$ have the unique solutions $h_1(t, \mathbf{X}_1, \mathbf{X}_2)$ and $h_2(t, \mathbf{X}_i)$, respectively. Therefore, the GPQs Y_1, Y_2 and Y_3 can be used to obtain the GCIs for δ.

Tables 1 and 2 report the coverage probabilities and the average lower limits of the proposed GCLLs for the Weibull stress-strength model with the same or different

Table 1 The coverage probabilities and the average lower limits (in parentheses) of GCLLs for the Weibull stress-strength model with the common shape parameter based on the complete samples

$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	δ	(n_1, n_2)	$1-\gamma$	GPQ Y_1
(2, 2, 4, 1)	0.8000	(10, 10)	0.90	0.8978 (0.6710)
			0.95	0.9476 (0.6295)
		(15, 15)	0.90	0.8984 (0.6963)
			0.95	0.9498 (0.6638)
(1, 1, 6, 1)	0.8571	(10, 10)	0.90	0.8982 (0.7418)
			0.95	0.9460 (0.7030)
		(15, 15)	0.90	0.9006 (0.7660)
			0.95	0.9492 (0.7364)
(3, 3, 15, 1)	0.9375	(10, 10)	0.90	0.8974 (0.8575)
			0.95	0.9466 (0.8279)
		(15, 15)	0.90	0.9012 (0.8769)
			0.95	0.9506 (0.8555)

Table 2 The coverage probabilities and the average lower limits (in parentheses) of GCLLs for the Weibull stress-strength model with the unequal shape parameters based on the complete samples

$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	δ	(n_1, n_2)	$1-\gamma$	MGPQ Y_3	GPQ Y_2
(2, 3, 3, 2)	0.7091	(10, 10)	0.90	0.9014 (0.5455)	0.9336 (0.5367)
			0.95	0.9578 (0.4891)	0.9716 (0.4900)
		(15, 15)	0.90	0.9006 (0.5773)	0.9232 (0.5713)
			0.95	0.9550 (0.5329)	0.9648 (0.5333)
(1, 3, 3, 1)	0.8935	(10, 10)	0.90	0.8914 (0.7698)	0.9378 (0.7456)
			0.95	0.9510 (0.7184)	0.9722 (0.6996)
		(15, 15)	0.95	0.8894 (0.7978)	0.9266 (0.7820)
			0.95	0.9480 (0.7595)	0.9648 (0.7472)
(1, 3, 4, 1)	0.9402	(10, 10)	0.90	0.8982 (0.8382)	0.9420 (0.8130)
			0.95	0.9546 (0.7923)	0.9754 (0.7705)
		(15, 15)	0.90	0.8902 (0.8636)	0.9334 (0.8476)
			0.95	0.9502 (0.8308)	0.9672 (0.8169)

shape parameters when the nominal levels are 0.9, 0.95. All simulation results are based on 5000 replications with $B = 10,000$.

It is observed from Table 1 that the coverage percentages of the proposed GCLL based on the GPQ Y_1 are quite close to the nominal coverage probabilities, even for small sample sizes in the common parameter case. The simulation results in Table 2 show that when the shape parameters are unequal, the coverage percentages of the proposed GCLL based on the MGPQ Y_3 are quite close to the nominal coverage probabilities, even for small sample sizes, but the coverage percentages of the proposed GCLL based on the GPQ Y_2 are larger than the nominal coverage probabilities. These findings show that the proposed GPQs Y_1 and Y_3 work well for the Weibull stress-strength model with the equal and unequal parameters, respectively.

Table 3 The coverage probabilities and the average lower limits (in parentheses) of GCLLs for the Kumaraswamy stress-strength model with the equal α_1 and α_2 based on the complete samples

$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	δ	(n_1, n_2)	$1-\gamma$	GPQ Y_1
(1, 1, 3, 1)	0.7500	(10, 10)	0.90	0.9034 (0.6165)
			0.95	0.9490 (0.5749)
		(15, 15)	0.90	0.9030 (0.6416)
			0.95	0.9522 (0.6087)
(1, 1, 5, 1)	0.8333	(10, 10)	0.90	0.9010 (0.7161)
			0.95	0.9510 (0.6772)
		(15, 15)	0.90	0.9034 (0.7400)
			0.95	0.9510 (0.7101)
(1, 1, 10, 1)	0.9091	(10, 10)	0.90	0.9004 (0.8186)
			0.95	0.9518 (0.7862)
		(15, 15)	0.90	0.9028 (0.8391)
			0.95	0.9502 (0.8152)

Table 4 The coverage probabilities and the average lower limits (in parentheses) of GCLLs for the Kumaraswamy stress-strength model with the unequal α_1 and α_2 based on the complete samples

$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	δ	(n_1, n_2)	$1-\gamma$	MGPQ Y_3	GPQ Y_2
(3, 2, 5, 1)	0.7215	(10, 10)	0.90	0.8928 (0.5697)	0.9206 (0.5616)
			0.95	0.9496 (0.5134)	0.9616 (0.5145)
		(15, 15)	0.90	0.8938 (0.5961)	0.9134 (0.5912)
			0.95	0.9544 (0.5520)	0.9622 (0.5533)
(3, 4, 3, 1)	0.8220	(10, 10)	0.90	0.8996 (0.6837)	0.9326 (0.6653)
			0.95	0.9560 (0.6294)	0.9710 (0.6181)
		(15, 15)	0.95	0.9008 (0.7104)	0.9304 (0.6984)
			0.95	0.9552 (0.6689)	0.9680 (0.6617)
(2, 5, 4, 2)	0.9227	(10, 10)	0.90	0.9012 (0.8171)	0.9418 (0.7941)
			0.95	0.9592 (0.7706)	0.9776 (0.7518)
		(15, 15)	0.90	0.9002 (0.8414)	0.9338 (0.8267)
			0.95	0.9566 (0.8077)	0.9732 (0.7958)

Example 6 Kumaraswamy distributions with the equal or unequal α_1, α_2.

When the stress and strength variables follow the Kumaraswamy distributions, $S_j(\alpha, \mathbf{X}) = \sum_{k=1}^{j} \log(1 - X_{(k)}^{\alpha}) + (n-j)\log(1 - X_{(j)}^{\alpha})$. It is obvious that $U(\alpha, \mathbf{X})$ is the strictly increasing function of α and that $\lim_{\alpha \to 0+} U(\alpha, \mathbf{X}) = 0$, $\lim_{\alpha \to \infty} U(\alpha, \mathbf{X}) = \infty$. Thus the equations $W(\alpha) = t$ and $U(\alpha, \mathbf{X}_i) = t$ have the unique solutions $h_1(t, \mathbf{X}_1, \mathbf{X}_2)$ and $h_2(t, \mathbf{X}_i)$, respectively. Therefore, the GPQs Y_1, Y_2 and Y_3 can be used to obtain the GCIs for δ.

Tables 3 and 4 report the coverage probabilities and the average lower limits of the proposed GCLLs for the Kumaraswamy stress-strength model with the equal or

unequal α_1 and α_2 when the nominal levels are 0.9, 0.95. All simulation results are based on 5000 replications with $B = 10,000$.

The simulation results show that similar to the Weibull stress-strength model case, the GPQ Y_1 works well when $\alpha_1 = \alpha_2$, while the MGPQ Y_3 works well under $\alpha_1 \neq \alpha_2$ for the Kumaraswamy stress-strength model.

3 Inference for the Stress-Strength Model Based on Record Values

Let $\{Z_n, n = 1, 2, \dots\}$ be a sequence of independent and identical copies of Z. An observation Z_j is called an upper (lower) record value if its value exceeds (is less than) all previous observations. This means Z_j is an upper (lower) record value if $Z_j > (<) Z_i$ for all $i < j$. Based on this fact, the record time sequence $\{T_n, n \geq 1\}$ is defined as: $T_1 = 1$ and $T_n = min\{j : Z_j > (<) Z_{T_{n-1}}\}$ for $n \geq 2$. The subsequence $\{R_n = Z_{T_n}, n = 1, 2, \dots\}$ is then the sequence of upper (lower) record values from the original sequence $\{Z_n, n = 1, 2, \dots\}$.

Let $\mathbf{R}_i = (R_{i,1}, R_{i,2}, \dots, R_{i,n_i})$ be the upper record values from the PHF(α_i, β_i). From Arnold et al. [2], the likelihood function based on the upper record values is given by

$$L_i(\alpha_i, \beta_i) = \beta_i g(r_{i,n_i}; \alpha_i)[1 - G(r_{i,n_i}; \alpha_i)]^{\beta_i - 1} \prod_{j=1}^{n_i - 1} \frac{\beta_i g(r_{i,j}; \alpha_i)[1 - G(r_{i,j}; \alpha_i)]^{\beta_i - 1}}{[1 - G(r_{i,j}; \alpha_i)]^{\beta_i}}$$

$$= \beta_i^{n_i} [1 - G(r_{i,n_i}; \alpha_i)]^{\beta_i} \prod_{j=1}^{n_i} \lambda(r_{i,j}; \alpha_i),$$

where $\lambda(x; \alpha) = g(x; \alpha)/[1 - G(x; \alpha)]$ is the failure rate function of the distribution $G(x; \alpha)$.

3.1 Maximum Likelihood Estimation

3.1.1 $\alpha_1 = \alpha_2 \hat{=} \alpha$

Notice that in this case, the likelihood function is given by

$$L(\alpha, \beta_1, \beta_2) = L_1(\alpha, \beta_1) \cdot L_2(\alpha, \beta_2),$$

thus the MLEs $\widehat{\alpha}$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ of α, β_1, β_2 can be obtained from the following equations:

$$\sum_{i=1}^{2}\left\{\sum_{j=1}^{n_i}\frac{1}{\lambda_{i,j}}\frac{\partial\lambda_{i,j}}{\partial\alpha}+\frac{n_i}{(1-G_{i,n_i})\log(1-G_{i,n_i})}\frac{\partial G_{i,n_i}}{\partial\alpha}\right\}=0, \tag{17}$$

$$\beta_1=-\frac{n_1}{\log(1-G_{1,n_1})},\ \beta_2=-\frac{n_2}{\log(1-G_{2,n_2})},$$

where $\lambda_{i,j}=\lambda(r_{i,j};\alpha)$, $G_{i,n_i}=G(r_{i,n_i};\alpha)$. Hence, the MLE of δ is given by

$$\widehat{\delta}=\frac{n_1/\log[1-G(r_{1,n_1};\widehat{\alpha})]}{n_1/\log[1-G(r_{1,n_1};\widehat{\alpha})]+n_2/\log[1-G(r_{2,n_2};\widehat{\alpha})]},$$

where $\widehat{\alpha}$ is the solution of the Eq. (17).

3.1.2 $\alpha_1 \neq \alpha_2$ Case

In this case, the MLEs $\widetilde{\alpha}_i$, $\widetilde{\beta}_i$ of α_i, β_i are given by

$$\sum_{j=1}^{n_i}\frac{1}{\lambda_{i,j}}\frac{\partial\lambda_{i,j}}{\partial\alpha_i}+\frac{n_i}{(1-G_{i,n_i})\log(1-G_{i,n_i})}\frac{\partial G_{i,n_i}}{\partial\alpha_i}=0, \tag{18}$$

$$\beta_i=-\frac{n_i}{\log[1-G(r_{i,n_i};\alpha_i)]},\ i=1,2. \tag{19}$$

Thus the MLE of δ is given by

$$\widetilde{\delta}=\int_0^\infty F(x;\widetilde{\alpha}_1,\widetilde{\beta}_1)dF(x;\widetilde{\alpha}_2,\widetilde{\beta}_2).$$

3.2 *Interval Estimation*

In order to derive the GCIs for δ, the following lemma is needed.

Lemma 2 *Let $R_1, R_2, \ldots, R_n$ be n upper record values observed from the standard exponential distribution, then $R_1, R_2 - R_1, \ldots, R_n - R_{n-1}$ are i.i.d. the standard exponential variables.*

Notice that $R_{i,1}, R_{i,2}, \ldots, R_{i,n_i}$ are the upper record values from the PHF(α_i, β_i), then we have that $-\log[1 - F(R_{i,j};\alpha_i,\beta_i)]$, $j = 1, 2, \ldots, n_i$ are the upper record values from the standard exponential distribution. Using Lemma 2,

we have that $Z_{i,1} = -\log[1-G(R_{i,1};\alpha_i)]$, $Z_{i,j} = \log[1-G(R_{i,j-1};\alpha_i)]-\log[1-G(R_{i,j};\alpha_i)]$, $j = 2, 3, \ldots, n_i$ are i.i.d. the exponential variables with mean $1/\beta_i$.

Since $S_j(\alpha_i, \mathbf{R}_i) = Z_{i,1} + Z_{i,2} + \cdots + Z_{i,j} = -\log[1 - G(R_{i,j};\alpha_i)]$, we have from Lemma 1 that $U(\alpha_i, \mathbf{R}_i) = 2\sum_{j=1}^{n_i-1} \log(S_{n_i}(\alpha_i, \mathbf{R}_i)/S_j(\alpha_i, \mathbf{R}_i))$ and $V_i = 2\beta_i S_{n_i}(\alpha_i, \mathbf{R}_i)$ are mutually independent, and that $U(\alpha_i, \mathbf{R}_i) \sim \chi^2(2n_i - 2)$, $V_i \sim \chi^2(2n_i)$. Here, we assume that $U(\alpha_i; \mathbf{R}_i)$ is the strictly monotonic function of α_i.

Let $h_3(t, \mathbf{R}_1, \mathbf{R}_2)$ is the solution of the equation $U(\alpha; \mathbf{R}_1)+U(\alpha; \mathbf{R}_2) = t$. When $\alpha_1 = \alpha_2 \hat{=} \alpha$, along the same line as the derivation of Y_1, the GPQ for δ is given by

$$Y_4 = \frac{P_1}{P_1 + P_2},$$

where $W_0 \sim \chi^2(2n_1 + 2n_2 - 4)$, $P_i = V_i / \log[1 - G(R_{i,n_i}; h_3(W_0, \mathbf{R}_1, \mathbf{R}_2))]$.

When $\alpha_1 \neq \alpha_2$, the GPQ for δ is given by

$$Y_5 = \int_0^\infty F(x; h_4(W_{0,1}, \mathbf{R}_1), Q_1/2) dF(x; h_4(W_{0,2}, \mathbf{R}_2), Q_2/2),$$

where $W_{0,i} \sim \chi^2(2n_i - 1)$, $Q_i = -V_i / \log[1 - G(R_{i,n_i}; h_4(W_{0,i}, \mathbf{R}_i))]$ and $h_4(W_{0,i}, \mathbf{R}_i)$ is the solution of the equation $U(\alpha_i; \mathbf{R}_i) = W_{0,i}$. Similarly, if the performance of Y_5 is not satisfactory under some distributions, the MGPQ can be used to improve the performance of the GCI for these distributions. The percentiles of the GPQs Y_4 and Y_5 can also be obtained by the Monte Carlo algorithms. These algorithms are similar to those in the complete samples situation.

Remark 5 Similar to the complete samples case, the proposed procedures in this section can also be extended to the PRHF situation, as follows.

Let $\mathbf{R}_i = (R_{i,1}, R_{i,2}, \ldots, R_{i,n_i})$ be the lower record values from the PRHF(α_i, β_i). Notice that $F(R_{i,1}; \alpha_i, \beta_i), F(R_{i,2}; \alpha_i, \beta_i), \ldots, F(R_{i,n_i}; \alpha_i, \beta_i)$ be the lower record values from the standard uniform distribution, we have from Lemma 1 in Wang et al. [23] that $Z_{i,1} = -\log G(R_{i,1}; \alpha_i)$, $Z_{i,j} = \log G(R_{i,j-1}; \alpha_i) - \log G(R_{i,j}; \alpha_i)$, $j = 2, 3, \ldots, n_i$ are i.i.d. the exponential random variables with mean $1/\beta_i$.

Let

$$S_j(\alpha_i, \mathbf{R}_i) = Z_{i,1} + Z_{i,1} + \cdots + Z_{i,j} = -\log G(R_{i,j}; \alpha_i),$$

$$U(\alpha_i, \mathbf{R}_i) = 2\sum_{j=1}^{n_i-1} \log(S_{n_i}(\alpha_i, \mathbf{R}_i)/S_j(\alpha_i, \mathbf{R}_i)).$$

Then $U(\alpha_i, \mathbf{R}_i)$ and $V_i = 2\beta_i S_{n_i}(\alpha_i, \mathbf{R}_i)$ are mutually independent, and $U(\alpha_i, \mathbf{R}_i) \sim \chi^2(2n_i - 2)$, $V_i \sim \chi^2(2n_i)$. The rest of steps is the same as the PHF stress-strength model based on upper record values.

Example 7 Weibull distributions with the equal or unequal shape parameters.

Notice that for the Weibull distribution, $G(x;\alpha) = 1 - e^{-x^{\alpha}}$ and $\lambda(x;\alpha) = \alpha x^{\alpha-1}$, then when $\alpha_1 = \alpha_2 \hat{=} \alpha$, the MLEs $\widehat{\alpha}$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ of α, β_1, β_2 are given by

$$\widehat{\alpha} = \frac{n_1 + n_2}{\sum_{i=1}^{2}\sum_{j=1}^{n_i-1}\log(R_{i,n_i}/R_{i,j})},\ \widehat{\beta}_1 = \frac{n_1}{R_{1,n_1}^{\widehat{\alpha}}},\ \widehat{\beta}_2 = \frac{n_2}{R_{2,n_2}^{\widehat{\alpha}}},$$

respectively. Hence, the MLE of δ is given by

$$\widehat{\delta} = \frac{n_1/R_{1,n_1}^{\widehat{\alpha}}}{n_1/R_{1,n_1}^{\widehat{\alpha}} + n_2/R_{2,n_2}^{\widehat{\alpha}}}.$$

When $\alpha_1 \neq \alpha_2$, the MLEs $\widetilde{\alpha}_i$, $\widetilde{\beta}_i$ of α_i, β_i are given by

$$\widetilde{\alpha}_i = \frac{n_i}{\sum_{j=1}^{n_i-1}\log(R_{i,n_i}/R_{i,j})},\ \widetilde{\beta}_i = \frac{n_i}{R_{i,n_i}^{\widetilde{\alpha}_i}},\ i = 1, 2,$$

respectively. Therefore, the MLE of δ is given by

$$\widetilde{\delta} = 1 - \widetilde{\beta}_2 \int_0^{\infty} \exp(-\widetilde{\beta}_1 x^{\widetilde{\alpha}_1/\widetilde{\alpha}_2} - \widetilde{\beta}_2 x)dx.$$

We now discuss the GCIs for the reliability δ. For the Weibull distribution, $U(\alpha_i, \mathbf{R}_i) = 2\alpha_i \sum_{j=1}^{n_i-1}\log(R_{i,n_i}/R_{i,j})$, $V_i = 2\beta_i R_{i,n_i}^{\alpha_i}$. Hence when $\alpha_1 = \alpha_2 \hat{=} \alpha$, the GPQ for δ is given by

$$Y_4 = \frac{V_1/R_{1,n_1}^{T}}{V_1/R_{1,n_1}^{T} + V_2/R_{2,n_2}^{T}},$$

where $T = W_0/[2\sum_{i=1}^{2}\sum_{j=1}^{n_i-1}\log(R_{i,n_i}/R_{i,j})]$.

When $\alpha_1 \neq \alpha_2$, the GPQ for δ is given by

$$Y_5 = 1 - \frac{V_2}{R_{2,n_2}^{T_2}} \int_0^{\infty} \exp\left(-\frac{V_1}{R_{1,n_1}^{T_1}} x^{T_1/T_2} - \frac{V_2}{R_{2,n_2}^{T_2}} x\right) dx,$$

where $T_i = W_{0,i}/[2\sum_{j=1}^{n_i-1}\log(R_{i,n_i}/R_{i,j})]$. Hence, the GPQs Y_4 and Y_5 can be used to obtain the GCIs for δ. The more details were shown in Wang and Ye [22].

Example 8 Generalized exponential distributions with the equal or unequal scale parameters.

Suppose that X_i follows the generalized exponential distribution GED(α_i, β_i). Its CDF is given by

$$F(x; \alpha_i, \beta_i) = (1 - e^{-\alpha_i x})^{\beta_i},\ x > 0,$$

where $\alpha_i (> 0)$, $\beta_i (> 0)$ are the scale and shape parameters, respectively.

Let $R_{i,1}, R_{i,2}, \ldots, R_{i,n_i}$ be the lower record values from the generalized exponential distribution GED(α_i, β_i). Then the likelihood function is given by

$$\begin{aligned} L(\alpha_1, \alpha_2, \beta_1, \beta_2) &= \prod_{i=1}^{2} \left[f(R_{i,n_i}) \prod_{j=1}^{n_i - 1} \frac{f(R_{i,j})}{F(R_{i,j})} \right] \\ &= \prod_{i=1}^{2} \alpha_i^{n_i} \beta_i^{n_i} e^{-\alpha_i \sum_{j=1}^{n_i} R_{i,j}} (1 - e^{-\alpha_i R_{i,n_i}}) / \prod_{j=1}^{n_i} (1 - e^{-\alpha_i R_{i,j}}). \end{aligned}$$

Therefore, when $\alpha_1 = \alpha_2 \hat{=} \alpha$, the MLEs $\widehat{\alpha}, \widehat{\beta}_1, \widehat{\beta}_2$ of the parameters α, β_1, β_2 are given by the following equations:

$$\frac{1}{\alpha} - \frac{1}{n_1 + n_2} \sum_{i=1}^{2} \left[\frac{R_{i,n_i} e^{-\alpha R_{i,n_i}}}{(1 - e^{-\alpha R_{i,n_i}}) \log(1 - e^{-\alpha R_{i,n_i}})} + \sum_{j=1}^{n_i} \frac{R_{i,j}}{1 - e^{-\alpha R_{i,j}}} \right] = 0, \quad (20)$$

$$\beta_1 = -\frac{n_1}{\log(1 - e^{-\alpha R_{1,n_1}})},\ \beta_2 = -\frac{n_2}{\log(1 - e^{-\alpha R_{2,n_2}})},$$

respectively. Thus the MLE of δ is given by

$$\widehat{\delta} = \frac{n_2 / \log(1 - e^{-\widehat{\alpha} R_{2,n_2}})}{n_1 / \log(1 - e^{-\widehat{\alpha} R_{1,n_1}}) + n_2 / \log(1 - e^{-\widehat{\alpha} R_{2,n_2}})},$$

where $\widehat{\delta}$ is the solution of the Eq. (20).

When $\alpha_1 \neq \alpha_2$, the MLEs $\widetilde{\alpha}_i, \widetilde{\beta}_i$ of the parameters α_i, β_i are given by the following equations:

$$\frac{1}{\alpha_i} - \frac{1}{n_i} \left[\frac{R_{i,n_i} e^{-\alpha_i R_{i,n_i}}}{(1 - e^{-\alpha R_{i,n_i}}) \log(1 - e^{-\alpha_i R_{i,n_i}})} + \sum_{j=1}^{n_i} \frac{R_{i,j}}{1 - e^{-\alpha_i R_{i,j}}} \right] = 0,$$

$$\beta_i = -\frac{n_i}{\log(1 - e^{-\alpha_i R_{i,n_i}})},\ i = 1, 2,$$

respectively. Therefore the MLE of δ is given by

$$\widetilde{\delta} = \widetilde{\alpha}_2 \widetilde{\beta}_2 \int_0^\infty e^{-\widetilde{\alpha}_2 x}(1 - e^{-\widetilde{\alpha}_2 x})^{\widetilde{\beta}_2 - 1}(1 - e^{-\widetilde{\alpha}_1 x})^{\widetilde{\beta}_1} dx.$$

Notice that for the generalized exponential distribution, $G(x; \alpha) = 1 - e^{-\alpha x}$, $x > 0$, then $S_j(\alpha_i, \mathbf{R}_i) = -\log[1 - \exp(-\alpha_i R_{i,n_i})]$. We have from Lemma 2 in Wang et al. [25] that

$$U(\alpha_i, \mathbf{R}_i) = 2 \sum_{j=1}^{n_i - 1} \log \frac{\log[1 - \exp(-\alpha_i R_{i,n_i})]}{\log[1 - \exp(-\alpha_i R_{i,j})]}$$

is the strictly increasing function of α_i. In addition, It is easy to verify the following results:

$$\lim_{\alpha_i \to 0+} U(\alpha_i, \mathbf{R}_i) = 0, \quad \lim_{\alpha_i \to \infty} U(\alpha_i, \mathbf{R}_i) = \infty.$$

Therefore, the equations $U(\alpha_1, \mathbf{R}_1) + U(\alpha_2, \mathbf{R}_2) = W_0$ and $U(\alpha_i, \mathbf{R}_i) = W_{0,i}$ have the unique solutions, respectively. Hence, the GPQs Y_4 and Y_5 can be used to obtain the GCIs for δ.

Table 5 reports the coverage probabilities and the average lower limits of the proposed GCLLs for the generalized exponential stress-strength model with the equal or unequal α_1 and α_2 when the nominal levels are 0.9, 0.95. All simulation results are based on 5000 replications with $B = 10,000$. The simulation results show that the GPQs Y_4 and Y_5 work well for $\alpha_1 = \alpha_2$ and $\alpha_1 \neq \alpha_2$, respectively.

Table 5 The coverage probabilities and the average lower limits (in parentheses) of GCLLs for the generalized exponential stress-strength model based on the lower record values

$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	δ	(n_1, n_2)	$1 - \gamma$	GPQ Y_4 or Y_5
(1, 1, 1, 2)	0.6667	(10, 10)	0.90	0.9014 (0.5265)
			0.95	0.9500 (0.4853)
(1, 1, 1, 4)	0.8000	(10, 10)	0.90	0.9020 (0.6749)
			0.95	0.9488 (0.6357)
(1, 1, 1, 10)	0.9091	(10, 10)	0.90	0.8972 (0.8203)
			0.95	0.9486 (0.7889)
(3, 2, 2, 3)	0.7452	(10, 10)	0.90	0.9114 (0.3041)
			0.95	0.9548 (0.2067)
(2, 1, 2, 3)	0.8286	(10, 10)	0.90	0.9116 (0.3690)
			0.95	0.9560 (0.2585)
(2, 1, 2, 5)	0.9127	(10, 10)	0.90	0.9110 (0.4526)
			0.95	0.9558 (0.3246)

4 Conclusion

The proportional hazards family and the proportional reverse hazards family include many commonly-used distributions, such as the Weibull distribution, the Gompertz distribution, the Kumaraswamy distribution and the generalized exponential distribution, etc. In this chapter, we have systematically explored statistical inference procedures for the proportional hazards family and proportional reverse hazards family stress-strength models based the complete samples or the record values. In complete samples case, the maximum likelihood estimators for the stress-strength reliability δ with the common or different parameters α_1 and α_2 were derived. To construct confidence intervals for δ, the method of generalized pivotal quantities was used. The simulation results validated the satisfactory performance of the proposed GPQs Y_1 and Y_3.

We then developed the inference procedures for the stress-strength model when the samples are the record values. we derived the maximum likelihood estimators and GCIs for δ with the common or different parameters α_1 and α_2. The good performance of the generalized confidence intervals based on the GPQs Y_4 and Y_5 was verified.

Acknowledgements This work is supported by the National Natural Science Foundation of China under the contract numbers 11871431, 11671303, and First Class Discipline of Zhejiang - A (Zhejiang Gongshang University—Statistics).

References

1. Aminzadeh, M. S. (1997). Estimation of reliability for exponential stress-strength models with explanatory variables. *Applied Mathematics and Computation, 84*, 269–274.
2. Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1998). *Records*. Wiley, New York.
3. Baklizi, A. (2008). Likelihood and Bayesian estimation of $Pr(X < Y)$ using lower record values from generalized exponential distribution. *Computational Statistics and Data Analysis, 52*, 3468–3473.
4. Basirat, M., Baratpour, S., & Ahmadi, J. (2015). Statistical inferences for stress-strength in the proportional hazard models based on progressive Type-II censored samples. *Journal of Statistical Computation and Simulation, 85*(3), 431–449.
5. Basirat, M., Baratpour, S., & Ahmadi, J. (2016). On estimation of stress-strength parameter using record values from proportional hazard rate models. *Communications in Statistics – Theory and Methods, 45*(19), 5787–5801.
6. Birnbaum, Z. M. (1956). On a use of the Mann-Whitney statistics. *Proceedings of the Third Berkeley Symposium in Mathematics, Statistics and Probability, 1*, 13–17.
7. Birnbaum, Z. W., & McCarty, R. C. (1958). A distribution-free upper confidence bound for $Pr(Y < X)$, based on independent samples of X and Y. *The Annals of Mathematical Statistics, 29*, 558–562.
8. Guo, H., & Krishnamoorthy, K. (2004). New approximate inferential methods for the reliability parameter in a stress-strength model: The normal case. *Communications in Statistics–Theory and Methods, 33*, 1715–1731.

9. Guttman, I., Johnson, R. A., Bhattacharyya, G. K., & Reiser, B. (1988). Confidence limits for stress-strength models with explanatory variables. *Technometrics, 30*, 161–168.
10. Hall, I. J. (1984). Approximate one-sided tolerance limits for the difference or sum of two independent normal variates. *Journal of Quality Technology, 16*, 15–19.
11. Johnson, R. A. (1988). Stress-strength models for reliability. In P. R. Krishnaiah, C. R. Rao (Eds.), *Handbook of statistics* (vol. 7, pp. 27–54). Elsevier, Amsterdam.
12. Kotz, S., Lumelskii, Y., & Pensky, M. (2003). *The stress-strength model and its generalizations*. World Scientific Press, Singapore.
13. Krishnamoorthy, K., & Lin, Y. (2010). Confidence limits for stress-strength reliability involving Weibull models. *Journal of Statistical Planning and Inference, 140*, 1754–1764.
14. Kundu, D., & Gupta, R. D. (2005). Estimation of $R = P(Y < X)$ for the generalized exponential distribution. *Metrika, 61*, 291–308.
15. Kundu, D., & Gupta, R. D. (2006). Estimation of $R = P(Y < X)$ for Weibull distribution. *IEEE Transaction on Reliability, 55*, 270–280.
16. Lio, Y. L., & Tsai, T. R. (2012). Estimation of $\delta = P(X < Y)$ for Burr XII distribution based on the progressively first failure-censored samples. *Journal of Applied Statistics, 39*, 309–322.
17. McCool, J. I. (1991). Inference on $P(Y < X)$ in the Weibull case. *Communications in Statistics–Simulation and Computations, 20*, 129–148.
18. Raqab, M. Z., & Kundu, D. (2005). Comparison of different estimators of $P(Y < X)$ for a scaled Burr type X distribution. *Communications in Statistics–Simulation and Computation, 34*, 465–483.
19. Reiser, B., & Guttman, I. (1986). Statistical inference for $Pr(Y < X)$: The normal case. *Technometrics, 28*, 253–257.
20. Surles, J. G., & Padgett, W. J. (2001). Inference for reliability and stress-strength for a scaled Burr Type X distribution. *Lifetime Data Analysis, 7*, 187–200.
21. Wang, B. X., Yu, K., & Jones, M. C. (2010). Inference under progressively Type II right censored sampling for certain lifetime distributions. *Technometrics, 52*, 453–460.
22. Wang, B. X., & Ye, Z. S. (2015). Inference on the Weibull distribution based on record values. *Computational Statistics and Data Analysis, 83*, 26–36.
23. Wang, B. X., Yu, K., & Coolen, F. P. A. (2015). Interval estimation for proportional reversed hazard family based on lower record values. *Statistics and Probability Letters, 98*, 115–122.
24. Wang, B. X., Wang, X. K., & Yu, K. (2017). Inference on the Kumaraswamy distribution. *Communications in Statistics - Theory and Methods, 46*(5), 2079–2090.
25. Wang, B. X., Geng, Y. P., & Zhou, J. X. (2018). Inference for the generalized exponential stress-strength model. *Applied Mathematical Modelling, 53*, 267–275.
26. Weerahandi, S., & Johnson, R. A. (1992). Testing reliability in a stress-strength model when X and Y are normally distributed. *Technometrics, 34*, 83–91.
27. Weerahandi, S. (2004). *Generalized inference in repeated measures: Exact methods in MANOVA and mixed models*. Wiley, New York.
28. Wolfe, D. A., & Hogg, R. V. (1971). On constructing statistics and reporting data. *The American Statistician 25*, 27–30.

A Degradation Model Based on the Wiener Process Assuming Non-Normal Distributed Measurement Errors

Yan Shen, Li-Juan Shen, and Wang-Tu Xu

Abstract For highly reliable products whose failure times are scarce, traditional methods for lifetime time analysis are no longer effective and efficient. Instead, degradation data analysis that investigates degradation processes of products becomes a useful tool in evaluating reliability. It focuses on the inherent randomness of products, and investigates the lifetime properties by developing degradation models and extrapolating to lifetime variables. But degradation data are often subject to measurement errors, which may have tails and better be described by non-normal distribution. In this chapter, we consider a Wiener-based model and assume logistic distributed measurement errors. For parameter estimation of the model, the Monte-Carlo expectation-maximization method is adopted together with the Gibbs sampling. Also an efficient algorithm is proposed for a quick approximation of maximum likelihood value. Moreover the remaining useful lifetime is estimated and discussed. From the simulation results, we find that the proposed model is more robust than the model based on the Wiener process assuming normal-distributed errors. Finally, an example is given to illustrate the application of the proposed model.

Y. Shen (✉)
Department of Statistics, School of Economics, Xiamen University, Xiamen, Fujian, China
e-mail: sheny@xmu.edu.cn

L.-J. Shen
Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore, Singapore
e-mail: isesl@nus.edu.sg

W.-T. Xu
Department of Urban Planning, School of Architecture and Civil Engineering, Xiamen University, Xiamen, Fujian, China
e-mail: ato1981@163.com

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_13

1 Introduction

For highly reliable products, sufficient failure data are usually difficult to obtain during a relatively short time period of lifetime testing even if an accelerated lifetime test is adopted. The scarcity of failure data challenges the traditional methods for lifetime time analysis, and at the same time booms the development of degradation data analysis. It is found that most products degrade over time, and their measurable physical deterioration almost always precedes failure, for example, the loss of strength for bridge beams, the increase of vibration amplitude for bearings, the outdoor weathering of coating systems [1–4], etc. Therefore, instead of the traditional lifetime data analysis that studies the exact failure times, the degradation data analysis provides a microscopic point of view in understanding the inherent degradation processes of products and thus their failure mechanisms [5].

Degradation is usually modeled by a stochastic process [6]. Among the degradation models, the Weiner process with a positive drift rate is the most favorite one because of its mathematical properties and physical interpretations. Suppose that the degradation of a product $\{X(t), t \geq 0\}$ is governed by the following Wiener process [7]

$$X(t) = \mu\Lambda(t) + \sigma B(\Lambda(t)), \quad t \geq 0, \tag{1}$$

where μ is the drift rate with $\mu > 0$, σ is the diffusion coefficient, $B(t)$ is a standard Brownian motion, and $\Lambda(t)$ is a monotonic increasing function used to transform time scale. For continence, we often assume $\Lambda(0) = 0$. This process has extensive applications in modeling degradation data of various products. Whitmore and Schenkelberg [7] used it to describe resistance increase of a self-regulating heating cable. Hu et al. [8] considered the Wiener process as a suitable model for light intensity of a light-emitting diode (LED). Zhai and Ye [9] proposed an adaptive Wiener process model to predict residual useful life of lithium-ion batteries. Wang et al. [10] proposed a general degradation model based on the Wiener process, which has several Wiener-based models as its limiting cases.

In practice, degradation data are often noisy as the measurements of degradation can be affected by imperfect inspections, which include imperfect instruments, procedures and environments. These imperfect inspections bring extra variation to the real degradation process, which is referred to 'measurement error'. Thus it is necessary to take the measurements errors into account when developing degradation models. Whitmore [11] studied the Weiner process subject to measurement errors and described the related inference procedures. By allowing the drift rate to be a random variable, Ye et al. [12] proposed a random effects model with measurement errors based on the Wiener process. Tang et al. [13] also considered the Wiener process assuming the existence of measurement errors and applied the model in the prediction of residual useful life for lithium-ion battery. Recently, Li et al. [14] concerned a more general Wiener-based degradation model with two transformed time scales and measurement errors and studied its performances in detail. Zhai and

Ye [15] proposed a robust model for degradation analysis also based on the Wiener process but assuming t-distributed measurement errors.

In the above works except [15], the measurement errors are almost all assumed to follow a normal or a truncated-normal distribution. The normality assumption is common in most statistical models, as the measurement errors are usually expected to be the sum of many independent processes and often considered to have distributions that are nearly normal according to the large sample theory. However, some fatal errors may be also introduced during the observation process, which would significantly alter the behavior of degradation observations. Figure 1 depicts the capacity loss of lithium-ion Battery #6 with repeated charge-discharge cycles, based on the data set provided by [16]. From the figure, several large jumps are observed, and the data set shows that there were obvious time delays when measuring those dropped data points. These extraordinary measurements can be considered to have brought in large errors to degradation data, which may not be captured by the normal distribution that has low probabilities in large values. In this case, a distribution with relatively heavy tails may be expected for describing the measurement errors.

In this work, we consider the Weiner process with logistic distributed measurement errors in the modelling of degradation data. The error assumption is enlightened by Friedman [17], which claimed that the logistic distribution could approximate a normal distribution for small values while an exponential distribution for large values in each of two tails. The excess kurtosis of the logistic distribution is 1.2, which implies its relatively heavier tails than the normal distribution. Due to the complicated likelihood function of the proposed model, the Monte Carlo expectation-maximization (MCEM) algorithm together with the Gibbs sampling is used to obtain the maximum likelihood estimates (MLE) of the model parameters. To quickly approximate the likelihood value at the MLE, an algorithm is developed at the same time. Moreover, the estimation of remaining useful lifetime (RUL)

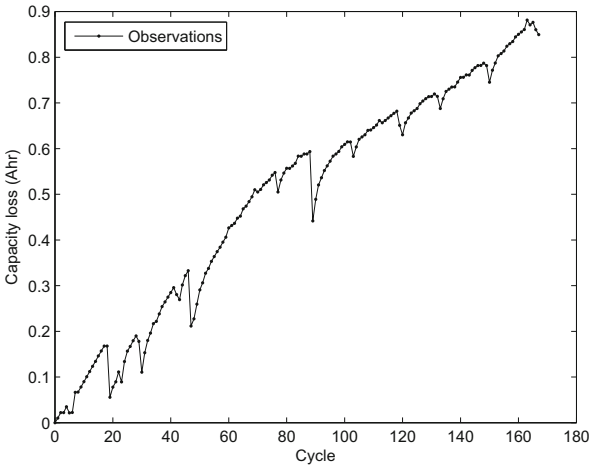


Fig. 1 The capacity loss of Battery #6 with operational cycle

is discussed. Monte Carlo experiments show that with non-normal measurement errors, the proposed model could give better performances in estimating both the real degradation process and the RUL.

This chapter is organized as follows. Section 2 describes the degradation model. Section 3 presents the detailed MCEM algorithms for parameter estimation as well as the developed algorithm for the approximation of maximum likelihood value. The sequent estimation of the RUL is discussed in Sect. 4. Section 5 presents Monte Carlo simulations and the results. Section 6 presents an illustrative example, and Sect. 7 concludes the chapter.

2 The Degradation Model

Suppose that the true degradation for a product can be modelled by the Wiener process $\{X(t), t \geq 0\}$ given by (1), which is however not directly observable. When a measurement is taken at time t, only the observation $Y(t)$ is obtained, which involves a measurement error ϵ_t subject to a logistic distribution with zero mean. Then the observed degradation process we concern is

$$Y(t) = X(t) + \epsilon_t, \quad t \geq 0, \tag{2}$$

where the measurement errors ϵ_t's are independent and identically distributed (i.i.d.) and independent of $X(t)$. The ϵ_t's are assumed to follow a logistic distribution with location parameter 0 and scale parameter s, $Logistic(0, s)$. The probability density function (PDF) and some useful properties are, respectively,

$$f_{\epsilon_t}(\epsilon) = \frac{e^{-\epsilon/s}}{s(1+e^{-\epsilon/s})^2}, \quad -\infty < \epsilon < \infty,$$

$\mathrm{E}(\epsilon_t) = 0$, $\mathrm{Var}(\epsilon_t) = s^2\pi^2/3$, and the mode of the PDF is 0.

Suppose that the process is measured at n ordered inspection times $\mathsf{t} = (t_1, \ldots, t_n)'$. The observed degradations are $\mathsf{Y} = (Y_1, \ldots, Y_n)'$, and the associated unobservable true degradations are $\mathsf{X} = (X_1, \ldots, X_n)'$, where $Y_i = X_i + \epsilon_i$. Since the Wiener process has independent and stationary increments that are normally distributed, we have for complete data $\{\mathsf{Y}, \mathsf{X}\}$,

$$\begin{aligned} &Y_i - X_i \sim Logistic(0, s), \\ &\Delta X_i = X_i - X_{i-1} \sim N(\mu\lambda_i, \sigma^2\lambda_i), \\ &\mathrm{Cov}(\Delta X_i, \Delta X_j) = 0, \quad i \neq j, \end{aligned} \tag{3}$$

where $\Lambda_i = \Lambda(t_i)$, $\Lambda_0 = 0$ and $\lambda_i = \Lambda_i - \Lambda_{i-1}$, $i = 1, \ldots, n$, $j = 1, \ldots, n$. Note that the distribution of $Y(t)$ involves integration over either $X(t)$ or ϵ_t, thus direct maximum likelihood estimation based on the observations $\mathsf{Y} = (Y_1, \ldots, Y_n)'$

is analytically infeasible and the maximum likelihood estimator cannot be obtained directly. As an alternative method, the expectation-maximization (EM) algorithm can be used for parameter estimation. To do this, some density functions between Y and X are needed.

Denote by $\mathsf{x} = (x_1, \ldots, x_n)'$ and $\mathsf{y} = (y_1, \ldots, y_n)'$ realizations of X and Y, and by $(\lambda_1, \ldots, \lambda_n)'$ the scaled time intervals obtained from the observed times $\mathsf{t} = (t_1, \ldots, t_n)'$. Also let $\theta = (\mu, \sigma, s)'$. The joint density function for (X, Y) given the parameters θ is

$$
\begin{aligned}
&p_{\mathsf{X},\mathsf{Y}|\theta}(\mathsf{x}, \mathsf{y}|\theta) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \prod_{i=1}^{n} \lambda_i^{-\frac{1}{2}} \exp\left\{-\frac{(x_i - x_{i-1} - \mu\lambda_i)^2}{2\sigma^2\lambda_i}\right\} \frac{e^{-(y_i - x_i)/s}}{s\left(1 + e^{-(y_i - x_i)/s}\right)^2},
\end{aligned} \tag{4}
$$

where $-\infty < x_i < \infty$, $-\infty < y_i < \infty$, $i = 1, \ldots, n$.

According to the Bayesian theorem, we can obtain the conditional density function of the true degradations $\mathsf{X} = (X_1, \ldots, X_n)'$ given the observations $\mathsf{Y} = (Y_1, \ldots, Y_n)'$ based on (4),

$$
p_{\mathsf{X}|\mathsf{Y},\theta}(\mathsf{x}|\mathsf{y}, \theta) = C \prod_{i}^{n} \exp\left\{-\frac{(x_i - x_{i-1} - \mu\lambda_i)^2}{2\sigma^2\lambda_i}\right\} \frac{e^{-(y_i - x_i)/s}}{\left(1 + e^{-(y_i - x_i)/s}\right)^2}, \tag{5}
$$

where C is the normalisation constant that enables $p_{\mathsf{X}|\mathsf{Y},\theta}(\mathsf{x}|\mathsf{y}, \theta)$ to be a density function. The value of C can be computed by

$$
1/C = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i}^{n} \exp\left\{-\frac{(x_i - x_{i-1} - \mu\lambda_i)^2}{2\sigma^2\lambda_i}\right\} \frac{e^{-(y_i - x_i)/s}}{\left(1 + e^{-(y_i - x_i)/s}\right)^2} dx_1 \cdots dx_n.
$$

Furthermore, the marginal conditional density function for each X_i of X given the other X_j's, denoted by $\mathsf{X}_{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, is expressed as

- for $i = 1, 2, \ldots, n-1$,

$$
\begin{aligned}
&p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}(x_i|\mathsf{x}_{(i)}, \mathsf{y}, \theta) \\
&= C_i \exp\left\{-\frac{(x_i - x_{i-1} - \mu\lambda_i)^2}{2\sigma^2\lambda_i} - \frac{(x_{i+1} - x_i - \mu\lambda_{i+1})^2}{2\sigma^2\lambda_{i+1}}\right\} \frac{e^{-(y_i - x_i)/s}}{\left(1 + e^{-(y_i - x_i)/s}\right)^2},
\end{aligned} \tag{6}
$$

- for $i = n$,

$$
p_{X_n|\mathsf{X}_{(n)},\mathsf{Y},\theta}(x_n|\mathsf{x}_{(n)}, \mathsf{y}, \theta) = C_n \exp\left\{-\frac{(x_n - x_{n-1} - \mu\lambda_n)^2}{2\sigma^2\lambda_n}\right\} \frac{e^{-(y_n - x_n)/s}}{\left(1 + e^{-(y_n - x_n)/s}\right)^2}, \tag{7}
$$

where $x_0 = 0$, and all C_i's are normalisation constants for the marginal densities.

3 Algorithms for Parameter Estimation

3.1 Parameter Estimation Based on the EM Algorithm

Suppose that there are observed degradation data from m units, and for the lth unit, we have observations $\mathsf{y}_l = (y_{l,1}, \ldots, y_{l,n_l})'$ measured at times $\mathsf{t}_l = (t_{l,1}, \ldots, t_{l,n_l})'$, $l = 1, \ldots, m$. Further assume that the unobservable true degradations at times t_l are $\mathsf{x}_l = (x_{l,1}, \ldots, x_{l,n_l})'$, $l = 1, \ldots, m$. Also for the lth unit, the transformed time is $\Lambda_{l,i} = \Lambda_l(t_i)$ and $\lambda_{l,i} = \Lambda_{l,i} - \Lambda_{l,i-1}$. For convenience, the drift rate parameters for different units are assumed homogeneous.

Based on the complete data $(\mathsf{y}_l, \mathsf{x}_l)$'s of all m units, the complete-data log-likelihood function of $\theta = (\mu, \sigma, s)'$ is given by

$$\begin{aligned}
&\ell(\theta|\mathsf{y}_1, \ldots, \mathsf{y}_m, \mathsf{x}_1, \ldots, \mathsf{x}_m) = \sum_{l=1}^{m} p_{\mathsf{X},\mathsf{Y}|\theta}(\mathsf{x}_l, \mathsf{y}_l|\theta) \\
&= -\frac{\sum_{l=1}^{m} n_l}{2} \ln(2\pi) - \frac{\sum_{l=1}^{m} n_l}{2} \ln \sigma^2 - \frac{1}{2} \sum_{l=1}^{m} \sum_{i=1}^{n_l} \ln \lambda_{l,i} - \sum_{l=1}^{m} \sum_{i=1}^{n_l} \frac{(x_{l,i} - x_{l,i-1} - \mu\lambda_{l,i})^2}{2\sigma^2 \lambda_{l,i}} \\
&\quad -\frac{1}{s} \sum_{l=1}^{m} \sum_{i=1}^{n_l} (y_{l,i} - x_{l,i}) - \left(\sum_{l=1}^{m} n_l \right) \ln s - 2 \sum_{l=1}^{m} \sum_{i=1}^{n_l} \ln(1 + e^{-(y_{l,i} - x_{l,i})/s}). \qquad (8)
\end{aligned}$$

Because the observation data include only y_l and t_l but not x_l, the complete-data log-likelihood cannot be used directly for parameter estimation. Thus instead we consider the expected value under the posterior distribution of the unobservable variable X_l, which corresponds to the E step of the EM algorithm. Given the current parameter values θ^{old}, the expectation of $\sum_{l=1}^{m} p_{\mathsf{X},\mathsf{Y}|\theta}(\mathsf{x}_l, \mathsf{y}_l|\theta)$, denoted by $\mathcal{Q}(\theta, \theta^{\text{old}})$, is given by

$$\begin{aligned}
\mathcal{Q}(\theta, \theta^{\text{old}}) = &-\frac{\sum_{l=1}^{m} n_l}{2} \ln(2\pi) - \frac{\sum_{l=1}^{m} n_l}{2} \ln \sigma^2 - \frac{1}{2} \sum_{l=1}^{m} \sum_{i=1}^{n_l} \ln \lambda_{l,i} \\
&- \sum_{l=1}^{m} \sum_{i=1}^{n_l} \frac{\mathbb{E}_{\mathsf{X}_l}[(X_{l,i} - X_{l,i-1} - \mu\lambda_{l,i})^2 | \mathsf{y}_l, \theta^{\text{old}}]}{2\sigma^2 \lambda_{l,i}} \\
&- \frac{1}{s} \sum_{l=1}^{m} \sum_{i=1}^{n_l} (y_{l,i} - \mathbb{E}_{\mathsf{X}_l}[X_{l,i}|\mathsf{y}_l, \theta^{\text{old}}]) - \left(\sum_{l=1}^{m} n_l \right) \ln s \\
&- 2 \sum_{l=1}^{m} \sum_{i=1}^{n_l} \mathbb{E}_{\mathsf{X}_l} \left[\ln(1 + e^{-(y_{l,i} - X_{l,i})/s}) | \mathsf{y}_l, \theta^{\text{old}} \right]. \qquad (9)
\end{aligned}$$

In the M step, we determine the revised parameter estimate θ^{new} by maximizing the above function

$$\theta^{\text{new}} = \arg\max_{\theta} \mathscr{Q}(\theta, \theta^{\text{old}}),$$

provided that those expectations in (9) are given.

In our model, the standard EM algorithm cannot be applied, because the conditional PDF for X_l given by (5) implies that the explicit expressions of the expectation quantities in (9) are difficult if not impossible to obtain. Alternative methods are thus desired. In this chapter, we adopt Gibbs sampling together with rejection method for estimating the expectations, which is described in detail in next subsection.

Provided that the expectations in (9) can be given in some way, the estimation procedure based on the EM algorithm is summarized as follows.

1. Choose an initial setting for the unknown parameters θ^{old};
2. **E step.** With θ^{old}, estimate the conditional expectations related to X_l's;
3. **M step.** Re-estimate the parameters using

$$\mu^{\text{new}} = \frac{\sum_{l=1}^{m} \hat{\mathbb{E}}_{\mathsf{X}_l}[X_{l,n_l}|\mathsf{y}_l, \theta^{\text{old}}]}{\sum_{l=1}^{m} \Lambda_{l,n_l}}, \tag{10}$$

$$\sigma^{2^{\text{new}}} = \frac{\sum_{l=1}^{m}\sum_{i=1}^{n_l} \hat{\mathbb{E}}_{\mathsf{X}_l}[(X_{l,i} - X_{l,i-1} - \mu^{\text{new}}\lambda_{l,i})^2|\mathsf{y}_l, \theta^{\text{old}}]/\lambda_{l,i}}{\sum_{l=1}^{m} n_l}, \tag{11}$$

and s^{new} solves the following equation

$$\sum_{l=1}^{m}\sum_{i=1}^{n_l}(y_{l,i} - \mathbb{E}_{\mathsf{X}_l}[X_{l,i}|\mathsf{y}_l, \theta^{\text{old}}]) - s\sum_{l=1}^{m} n_l - 2\sum_{l=1}^{m}\sum_{i=1}^{n_l}\mathbb{E}_{\mathsf{X}_l}\left[\frac{y_{l,i} - X_{l,i}}{1 + e^{(y_{l,i} - X_{l,i})/s}}|\mathsf{y}_l, \theta^{\text{old}}\right] = 0; \tag{12}$$

4. Check for convergence of either the log-likelihood or the parameter values. If the convergence criterion is not satisfied, then let $\theta^{\text{old}} \leftarrow \theta^{\text{new}}$, and return to step 2.

Based on the above procedure, the obtained estimate for θ, denoted by $\hat{\theta}$, is the MLE.

If the transformed time scale $\Lambda(t)$ also contains parameters, the vector θ can be rewritten to include them. And the M step in the EM algorithm should be modified so as to involve the estimation of the parameters in $\Lambda(t)$, which will further affect μ^{new} and $\sigma^{2^{\text{new}}}$. For example, if we consider $\Lambda(t) = t^b$, then the parameter vector is reset as $\theta = (\mu, \sigma, b, s)'$. The M step can be revised in the following way.

3. **M step.** Denote

$$\mu(b) = \frac{\sum_{l=1}^{m} \hat{\mathbb{E}}_{\mathsf{X}_l}[X_{l,n_l}|\mathsf{y}_l, \theta^{\text{old}}]}{\sum_{l=1}^{m} \Lambda_{l,n_l}},$$

$$\sigma^2(b) = \frac{\sum_{l=1}^{m} \sum_{i=1}^{n_l} \hat{\mathbb{E}}_{\mathsf{X}_l}[(X_{l,i} - X_{l,i-1} - \mu(b)\lambda_{l,i})^2|\mathsf{y}_l, \theta^{\text{old}}]/\lambda_{l,i}}{\sum_{l=1}^{m} n_l}.$$

Firstly calculate $b^{\text{new}} = \arg\max_b\{-\sum_{l=1}^{m} n_l \ln \sigma^2(b) - \sum_{l=1}^{m} \sum_{i=1}^{n_l} \ln \lambda_{l,i}\}$. Secondly re-estimate the parameters using $\mu^{\text{new}} = \mu(b^{\text{new}})$, $\sigma^{2\text{new}} = \sigma^2(b^{\text{new}})$. And s^{new} solves the same equation as (12).

Note that (i) both Λ_{l,n_l} and λ_{l,n_l} are functions of the parameter b; (ii) they are in fact the second and third terms of $\mathcal{Q}(\theta, \theta^{\text{old}})$ used in the M step for getting b^{new}.

3.2 The Gibbs Sampling Algorithm and the MCEM Method

Gibbs sampling, which is a special Monte Carlo Markov Chain (MCMC) algorithm, generates random vectors from the marginal conditional PDF of each element. This means that enough sample data for the unobservable variable X_l could be generated according to $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$ $(i = 1, \ldots, n_l)$ given by (6) and (7), which are then used for the estimation of the conditional expectations required in the EM algorithm. This type of method combining the Gibbs sampling and the EM algorithm is the so-called MCEM method [18].

Next we present the Gibbs sampling procedure in detail. For convenience, the case of only one unit is considered, while the case of m units can be dealt with similarly and separately. The subscript l is omitted for abbreviation. Denote the generated vector at the qth step for $\mathsf{X} = (X_1, \ldots, X_n)'$ by $\mathsf{Z}^q = (Z_1^q, \ldots, Z_n^q)'$, the generated values for Z^q by $\mathsf{z}^q = (z_1^q, \ldots, z_n^q)'$, and the observations of $Y(t)$ by $\mathsf{y} = (y_1, \ldots, y_n)'$. The procedure of the Gibbs sampling given the parameter θ is as follows,

Step 0 Set an initial vector $\mathsf{z}^0 = (z_1^0, \ldots, z_n^0)'$;
Step 1 Given z^0, generate a new vector $\mathsf{z}^1 = (z_1^1, \ldots, z_n^1)'$ by

1. Generate the value z_1^1 based on its marginal PDF $p_{X_1|\mathsf{X}_{(1)},\mathsf{Y},\theta}(z_1|z_2^0, \ldots, z_n^0, \mathsf{y}, \theta)$;
2. Generate the value z_i^1 based on its marginal PDF $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}(z_i|z_1^1, \ldots, z_{i-1}^1, z_{i+1}^0, \ldots, z_n^0, \mathsf{y}, \theta)$, $2 \le i \le n-1$;
3. Generate the value z_n^1 based on its PDF $p_{X_n|\mathsf{X}_{(n)},\mathsf{Y},\theta}(z_n|z_1^1, \ldots, z_{n-1}^1, \mathsf{y}, \theta)$;

Step q Given z^{q-1}, generate a new vector $\mathsf{z}^q = (z_1^q, \ldots, z_n^q)'$ by

1. Generate the value z_1^q based on $p_{X_1|\mathsf{X}_{(1)},\mathsf{Y},\theta}(z_1|z_2^{q-1}, \ldots, z_n^{q-1}, \mathsf{y}, \theta)$;
2. Generate the value z_i^q based on $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}(z_i|z_1^q, \ldots, z_{i-1}^q, z_{i+1}^{q-1}, \ldots, z_n^{q-1}, \mathsf{y}, \theta)$, $2 \le i \le n-1$;
3. Generate the value z_n^q based on $p_{X_n|\mathsf{X}_{(n)},\mathsf{Y},\theta}(z_n|z_1^q, \ldots, z_{n-1}^q, \mathsf{y}, \theta)$.

At each step q, the Gibbs algorithm produces the elements of the vector $\mathsf{Z}^q = (Z_1^q, \ldots, Z_n^q)'$ successively by utilizing Z^{q-1} obtained in previous step and the Z_i^q's already updated in step q. Since $\{\mathsf{Z}^q, q \in \mathsf{N}\}$ is actually a Markov chain, the stationary property of an ergodic Markov chain ensures that Z^q could be a good approximation of $\mathsf{X} = (X_1, \ldots, X_n)$ for enough large value of q, i.e. after enough number of iterations in the Gibbs sampling procedure.

Rejection Method

The main difficulty in the above sampling procedure is to generate each z_i^q according to the marginal conditional PDF $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$ in (6) or (7), which includes a normalization constant C_i that is difficult to work out. A feasible approach that exempts from calculating the constant is the Rejection method [19], which does not require the exact expression of the PDF $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$, only some $\tilde{p} \propto p$ and a bound on $\tilde{p}/g$,where g is a known PDF.

Fortunately, the marginal conditional PDF $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$ $(i = 1, \ldots, n)$ has two upper bounds given by the product of a constant and a normal PDF. We firstly consider the more general case for X_i, $i = 1, \ldots, n-1$. Since $1 + \exp\{-(y_i - x_i)/s\} > 1$, the PDF for X_i $(i = 1, \ldots, n-1)$ in (6) can be bounded by

$$\begin{aligned} & p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}(x_i|\mathsf{x}_{(i)}, \mathsf{y}, \theta) \\ & \le C_i \exp\left\{-\frac{(x_i - x_{i-1} - \mu\lambda_i)^2}{2\sigma^2\lambda_i} - \frac{(x_{i+1} - x_i - \mu\lambda_{i+1})^2}{2\sigma^2\lambda_{i+1}} - \frac{y_i - x_i}{s}\right\} \\ & = C_i * M_1\phi(x_i; u_1, \delta_1^2), \end{aligned} \tag{13}$$

where $\phi(\cdot)$ is the PDF of a normal distribution with mean $u_1 = \frac{x_{i-1}\lambda_{i+1}s + x_{i+1}\lambda_i s + \sigma^2\lambda_i\lambda_{i+1}}{(\lambda_i+\lambda_{i+1})s}$ and variance $\delta_1^2 = \sigma^2\frac{\lambda_i\lambda_{i+1}}{\lambda_i+\lambda_{i+1}}$, and $M_1 = \sqrt{2\pi}\delta_1 \exp\{-\frac{1}{2\delta_1^2}[\frac{A_1}{(\lambda_i+\lambda_{i+1})s} - u_1^2]\}$ is a constant with $A_1 = (x_{i-1} + \mu\lambda_i)^2\lambda_{i+1}s + (x_{i+1} - \mu\lambda_{i+1})^2\lambda_i s + 2\sigma^2\lambda_i\lambda_{i+1}y_i$.

Another upper bound can be also obtained by noting that $1+\exp\{-(y_i-x_i)/s\} > \exp\{-(y_i - x_i)/s\}$, which is

$$\begin{aligned} & p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}(x_i|\mathsf{x}_{(i)}, \mathsf{y}, \theta) \\ & \le C_i \exp\left\{-\frac{(x_i - x_{i-1} - \mu\lambda_i)^2}{2\sigma^2\lambda_i} - \frac{(x_{i+1} - x_i - \mu\lambda_{i+1})^2}{2\sigma^2\lambda_{i+1}} + \frac{y_i - x_i}{s}\right\} \\ & = C_i * M_2\phi(x_i; u_2, \delta_2^2), \end{aligned} \tag{14}$$

where $\phi(\cdot)$ is the PDF of a normal distribution with mean $u_2 = \frac{x_{i-1}\lambda_{i+1}s+x_{i+1}\lambda_i s-\sigma^2\lambda_i\lambda_{i+1}}{(\lambda_i+\lambda_{i+1})s}$ and variance $\delta_2^2 = \sigma^2\frac{\lambda_i\lambda_{i+1}}{\lambda_i+\lambda_{i+1}}$, and $M_2 = \sqrt{2\pi}\delta_2 \exp\{-\frac{1}{2\delta_2^2}[\frac{A_2}{(\lambda_i+\lambda_{i+1})s} - u_2^2]\}$, is a constant with $A_2 = (x_{i-1}+\mu\lambda_i)^2\lambda_{i+1}s + (x_{i+1} - \mu\lambda_{i+1})^2\lambda_i s - 2\sigma^2\lambda_i\lambda_{i+1}y_i$.

Similarly, the marginal conditional PDF for X_n given by (7) also have two upper bounds. One is

$$p_{X_n|\mathsf{X}_{(n)},\mathsf{Y},\theta}(x_n|\mathsf{x}_{(n)},\mathsf{y},\theta) \le C_n * M_3\phi(x_n; u_3, \delta_3^2), \tag{15}$$

where $\phi(\cdot)$ is the PDF of a normal distribution with mean $u_3 = x_{n-1}+\mu\lambda_n+\sigma^2\lambda_n/s$ and variance $\delta_3^2 = \sigma^2\lambda_n$, and $M_3 = \sqrt{2\pi}\delta_3 \exp\{-\frac{1}{2\delta_3^2}[(x_{n-1}+\mu\lambda_n)^2 + \frac{2\sigma^2\lambda_n y_n}{s} - u_3^2]\}$. The other is

$$p_{X_n|\mathsf{X}_{(n)},\mathsf{Y},\theta)}(x_n|\mathsf{x}_{(n)},\mathsf{y},\theta) \le C_n * M_4\phi(x_n; u_4, \delta_4^2), \tag{16}$$

where $\phi(\cdot)$ is the PDF of a normal distribution with mean $u_4 = x_{n-1}+\mu\lambda_n-\sigma^2\lambda_n/s$ and variance $\delta_4^2 = \sigma^2\lambda_n$, and $M_4 = \sqrt{2\pi}\delta_4 \exp\left\{\frac{1}{2\delta_4^2}[(x_{n-1}+\mu\lambda_n)^2 - \frac{2\sigma^2\lambda_n y_n}{s} - u_4^2]\right\}$.

Since we have obtained the bounds for $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$, the rejection method for generating X_i $(i = 1, \ldots, n)$ can be developed. For convenience, denote the explicitly expressed part of $p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$ by $\tilde{p}_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}$, i.e. $\tilde{p}_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta} = p_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}/C_i$, $i = 1, \ldots, n$. Based on (13) and (14), the procedure for generating one value of X_i $(i = 1, \ldots, n-1)$ is as follows:

Step 1 Compute u_1, δ_1^2, M_1 and u_2, δ_2^2, M_2;

Step 2 If $M_1 < M_2$, let $M = M_1$ and $g(x_i) = \phi(x_i|u_1, \delta_1^2)$, otherwise let $M = M_2$ and $g(x_i) = \phi(x_i|u_2, \delta_2^2)$;

Step 3 Generate ξ from the normal distribution $g(x_i)$, and U from the uniform distribution $U(0, 1)$;

Step 4 If $M \cdot U \le \tilde{p}_{X_i|\mathsf{X}_{(i)},\mathsf{Y},\theta}(\xi|\cdot)/g(\xi)$, let $X_i = \xi$, otherwise return to step 3.

To generate X_n, the procedure is similar but with u_1, δ_1^2, M_1 and u_2, δ_2^2, M_2 replaced by u_3, δ_3^2, M_3 and u_4, δ_4^2, M_4 in (15) and (16).

Remark 1 Note that the number of trials before a ξ is accepted has a geometric distribution with mean M, so the algorithm works best if M is small. That is the reason why we choose a smaller value between M_1 and M_2 or between M_3 and M_4.

Plugging-in the above procedure in the Gibbs sample allows us to obtain a sequence of sample $\mathsf{z}^q = (z_1^q, \ldots, z_n^q)'$, denoted by $\{\mathsf{z}^q, q \in \mathsf{N}\}$. To be used in E step, the sequence should be taken into the processes of burn-in and thinning. We discard $Q_0 = 500$ samples for burn-in in the simulation study and the example, where the predetermined Q_0 is the necessary number for the Markov chain $\{\mathsf{Z}^q, q \in$

N} to get a convergence state and can be determined by graphical checks of the MCMC trace plot. We use the autocorrelation function (ACF) plot to diagnose the autocorrelations among the sample sequence. For both the simulation study and the example, a thinned sequence of every 20 samples is taken based on the ACF plots. When a thinned sequence composed of Q_1 samples is obtained, the estimates of various expectations in (9) thus are simply the averages of the corresponding Q_1 values from the thinned sequence; for example,

$$\hat{\mathbb{E}}_{\mathsf{X}}[X_i|\mathsf{y}, \theta^{\text{old}}] = \frac{1}{Q_1}\sum_{q=1}^{Q_1} \tilde{z}_i^q,$$

where $\tilde{z}_i^q$ is the sample value for X_i in the thinned sequence. The other expectation quantities can be estimated in a similar way. Note that Q_1 is also predetermined and we use $Q_1 = 1500$. Once the estimates of all the expectations in (9) are obtained, the estimation procedure based on the EM algorithm, which is given in the previous section, could be carried out.

Remark 2 Instead of the MCEM method, the stochastic EM algorithm (SEM) also can be used for parameter estimation. For example, the method was used by [20] for the Gamma process with normal distributed measurement errors. More discussions on MCEM and SEM can be found in [21, 22].

3.3 Approximate Calculation of Maximum Likelihood Value

Theoretically, the maximum likelihood value could be directly obtained by the plug-in of the obtained estimate $\hat{\theta}$ into the likelihood function. However, the likelihood function for the proposed model is rather complicated, which takes a form of a multiple integral. Even for only one unit with n observations, the likelihood function is an n-fold integral of the joint density (4) over all elements in X. It is equivalently an n-fold integral over the n measurement errors ϵ_i, $i = 1, \ldots, n$, which is after some algebra,

$$L(\theta|\mathsf{y}) = A\int_{\epsilon_1,\cdots,\epsilon_n}\cdots\int\prod_{i=1}^{n}\exp\left\{\frac{2B_i(\epsilon_i-\epsilon_{i-1})-(\epsilon_i-\epsilon_{i-1})^2}{2\sigma^2\lambda_i}\right\}\frac{e^{-\epsilon_i/s}}{s\left(1+e^{-\epsilon_i/s}\right)^2}d\epsilon_1\cdots d\epsilon_n, \tag{17}$$

where $A \equiv A(\mathsf{y}, \theta) = (2\pi\sigma^2)^{-\frac{n}{2}}\prod_{i=1}^{n}\lambda_i^{-\frac{1}{2}}\exp\{-(y_i - y_{i-1} - \mu\lambda_i)^2/(2\sigma^2\lambda_i)\}$, $B_i \equiv B_i(\mathsf{y}, \theta) = y_i - y_{i-1} - \mu\lambda_i$. If there are m independent units, the likelihood function becomes a product of m terms, each of which has the form of (17). Hence a feasible algorithm for the calculation of the likelihood function (17) is desired.

Noting that the measurement errors ϵ_i, $i = 1, \ldots, n$, are i.i.d random variables subject to the logistic PDF $f_\epsilon(\epsilon) = e^{-\epsilon/s}/[s(1 + e^{-\epsilon/s})^2]$, the likelihood function (17) can be regarded as an expectation,

$$L(\theta|\mathsf{y}) = A \cdot \mathbb{E}_{\epsilon_1,\cdots,\epsilon_n}\left[\prod_{i=1}^{n} \exp\left\{\frac{2B_i(\epsilon_i - \epsilon_{i-1}) - (\epsilon_i - \epsilon_{i-1})^2}{2\sigma^2\lambda_i}\right\}\right] \tag{18}$$

$$= A \cdot \mathbb{E}_{\epsilon_1}\left[\varphi(\epsilon_1) \cdot \mathbb{E}_{\epsilon_2|\epsilon_1} \cdots \mathbb{E}_{\epsilon_{n-1}|\epsilon_{n-2}}\left[\varphi(\epsilon_{n-2}, \epsilon_{n-1}) \cdot \mathbb{E}_{\epsilon_n|\epsilon_{n-1}}[\varphi(\epsilon_{n-1}, \epsilon_n)]\right]\right]$$

where

$$\varphi(\epsilon_{i-1}, \epsilon_i) = \exp\left\{\left(\frac{B_i}{\sigma^2\lambda_i} - \frac{B_{i+1}}{\sigma^2\lambda_{i+1}}\right)\epsilon_i - \left(\frac{1}{2\sigma^2\lambda_i} + \frac{1}{2\sigma^2\lambda_{i+1}}\right)\epsilon_i^2 + \frac{\epsilon_{i-1}\epsilon_i}{\sigma^2\lambda_i}\right\},$$

$$\varphi(\epsilon_{n-1}, \epsilon_n) = \exp\left\{\frac{B_n}{\sigma^2\lambda_n}\epsilon_n - \frac{1}{2\sigma^2\lambda_n}\epsilon_n^2 + \frac{\epsilon_{n-1}\epsilon_n}{\sigma^2\lambda_n}\right\},$$

$$\varphi(\epsilon_1) = \exp\left\{\left(\frac{B_1}{\sigma^2\lambda_1} - \frac{B_2}{\sigma^2\lambda_2}\right)\epsilon_1 - \left(\frac{1}{2\sigma^2\lambda_1} + \frac{1}{2\sigma^2\lambda_2}\right)\epsilon_1^2\right\}.$$

The above formula having successive conditional expectations implies that, in order to obtain the likelihood value, we need to integrate the functions $\varphi(\epsilon_{i-1}, \epsilon_i)$, $i = 1, \ldots, n$, inside out based on the given observations y and the MLE $\hat{\theta}$. The integration operations are almost impossible, hence an estimation or approximation method is expected.

Bearing in mind the principle that sample mean is the best estimator for the expectation, we develop an iterative approximation method for the calculation of the likelihood function at MLE based on the conditional expectation form (18). The approximation method takes the following steps:

Step 1 For each ϵ_i ($i = 1, \ldots, n$), generate a sample of size K according to the PDF $f_\epsilon(\epsilon|\hat{\theta})$, denoted by $(e_{i,1}, \ldots, e_{i,K})$;

Step 2 For $\mathbb{E}_{\epsilon_n|\epsilon_{n-1}}[\varphi(\epsilon_{n-1}, \epsilon_n)]$, use $(e_{n,1}, \ldots, e_{n,K})$ to calculate the sample mean of $\varphi(\epsilon_{n-1}, \epsilon_n)$ for each $e_{n-1,k}$, $k = 1, \ldots, K$. That is, for a fixed $e_{n-1,k}$, calculate $\hat{\mathbb{E}}_{\epsilon_n|\epsilon_{n-1}=e_{n-1,k}}[\varphi(\epsilon_{n-1}, \epsilon_n)] = \frac{1}{K}\sum_{j=1}^{K}\varphi(e_{n-1,k}, e_{n,j})$;

Step 3 For $\mathbb{E}_{\epsilon_i|\epsilon_{i-1}}[\cdot]$, $i = n-1, \ldots, 2$, iteratively calculate the sample mean of $\varphi(\epsilon_{i-1}, \epsilon_i)$ for each $e_{i-1,k}$, $k = 1, \ldots, K$, but using $(e_{i,1}, \ldots, e_{i,K})$ and $\hat{\mathbb{E}}_{\epsilon_{i+1}|\epsilon_i}[\cdot]$ obtained in the previous step. That is, for a fixed $e_{i-1,k}$, calculate $\hat{\mathbb{E}}_{\epsilon_i|\epsilon_{i-1}=e_{i-1,k}}[\cdot] = \frac{1}{K}\sum_{j=1}^{K}\varphi(e_{i-1,k}, e_{i,j}) \cdot \hat{\mathbb{E}}_{\epsilon_{i+1}|\epsilon_i=e_{i,j}}[\cdot]$;

Step 4 For the last expectation, $\hat{\mathbb{E}}_{\epsilon_1}[\cdot] = \frac{1}{K}\sum_{j=1}^{K}\varphi(e_{1,j}) \cdot \hat{\mathbb{E}}_{\epsilon_2|\epsilon_1=e_{1,j}}[\cdot]$;

Step 5 The likelihood value at the MLE $\hat{\theta}$ is calculated by $L(\hat{\theta}|\mathsf{y}) \approx A(y, \hat{\theta})\hat{\mathbb{E}}_{\epsilon_1}[\cdot]$, and the corresponding log-likelihood is $\ell(\hat{\theta}|\mathsf{y}) \approx \log(A(y, \hat{\theta})) + \log(\hat{\mathbb{E}}_{\epsilon_1}[\cdot])$.

Since the observations y and the MLE $\hat{\theta}$ are already known, the above algorithm is able to yield the maximum likelihood value rather fast. Moreover, the algorithm can be easily modified and applied to other Wiener-based degradation models with measurement errors, as long as the step 1 is designed to generate measurement errors

according to the assumed distribution. Based on a simulation experiment for the model with Gaussian noises,[1] it is found that the sample size K can be chosen to be 1000 or more.

3.4 Statistical Inference

For statistical inference, we need not only the estimates for the unknown parameters, but also the variance-covariance matrix. As indicated in Bedair et al. [23], the Louis's formula in Louis [24] can be used for finding estimates of the variance-covariance matrix. In this work, for the parameter vector $\theta' = (\mu, \sigma, s)$ or $\theta' = (\mu, \sigma, b, s)$,

$$I_{\hat{\theta}} = \mathbb{E}_{\mathrm{X}}\left[-\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}|\mathrm{Y}, \hat{\theta}\right] - \mathbb{E}_{\mathrm{X}}\left[\frac{\partial \ell(\theta)}{\partial\theta}\frac{\partial \ell(\theta)}{\partial\theta'}|\mathrm{Y}, \hat{\theta}\right], \tag{19}$$

where $\mathrm{X} = (\mathsf{X}_1, \ldots, \mathsf{X}_m)$ and $\mathrm{Y} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_m)$. The formulae for the first and the second derivatives are given in the Appendix. To obtain $I_{\hat{\theta}}$, the expectation terms in (19) can be estimated using samples of the last iteration in the MCEM procedure.

Based on the above matrix, the variance-covariance matrix of θ is obtained as $\Sigma_{\hat{\theta}} = I_{\hat{\theta}}^{-1}$, and the diagonal elements are the estimated variances for the parameter estimates $\hat{\mu}$, $\hat{\sigma}$, $\hat{b}$ and $\hat{s}$. Due to the asymptotic normality of the MLE $\hat{\theta}$, we have $\hat{\theta} \dot{\sim} N(\theta, \Sigma_{\hat{\theta}})$, which can be used for further statistical inferences for θ, such as statistical hypothesis testing and construction of confidence interval (CI).

Besides the above asymptotic method, the bootstrap method provides another approach for statistical inferences. In our work, parametric bootstrapping can be adopted, which generates samples for Y by treating the MLE $\hat{\theta}$ as the true model parameter. For each of the samples, a separate estimate of θ can be obtained. Then all the estimates together will yield the empirical distribution for the MLE $\hat{\theta}$, and the upper and lower percentiles of the distribution can be used to test hypothesis and construct confidence interval.

4 Estimation of Remaining Useful Life

In a degradation process, a so-called soft failure of a product is often as when the degradation process hits a fixed threshold D. Thus the failure time of the product can be presented by the first passage time T_D of the process, which is,

$$T_D = \inf\{t : X(t) \geq D | X(0) < D\}. \tag{20}$$

[1] This famous model, based on the Wiener process assuming normal distributed measurement errors, has an explicitly expressed likelihood function [7].

For the Wiener process, the first passage time is a random variable subject to a transformed Inverse Gaussian (IG) distribution, that is, $\Lambda(T_D) \sim \text{IG}(D/\mu, D^2/\sigma^2)$. The CDF and PDF of T_D are, respectively,

$$F_{T_D}(t) = \Phi\left(\sqrt{\frac{D^2}{\sigma^2\Lambda(t)}}\left(\frac{\mu\Lambda(t)}{D} - 1\right)\right) + \exp\left(\frac{2D\mu}{\sigma^2}\right)\Phi\left(-\sqrt{\frac{D^2}{\sigma^2\Lambda(t)}}\left(\frac{\mu\Lambda(t)}{D} + 1\right)\right) \tag{21}$$

and

$$f_{T_D}(t) = \sqrt{\left(\frac{D^2}{2\pi\sigma^2\Lambda(t)^3}\right)}\exp\left[-\frac{(\mu\Lambda(t) - D)^2}{2\sigma^2\Lambda(t)}\right]\Lambda'(t). \tag{22}$$

As an important concept, remaining useful life is defined as the length of time from the current time point until a failure and indicates how long a product could survive based on the obtained information [25]. The RUL is actually an conditional random variable and has different expressions from different points of view. If a product is considered to generally represent a population and its degradations follow a Wiener process, the RUL can be written as

$$L_D^1(t) = T_D - t | T_D > t, \tag{23}$$

and the corresponding PDF can be obtained based on (21) and (22), which is $f_{L_D^1(t)}(l) = f_{T_D}(l + t)/(1 - F_{T_D}(t))$. If the focus is on a particular product with observable true degradation data $\mathsf{x} = (x_1, \ldots, x_n)'$ at $\mathsf{t} = (t_1, \ldots, t_n)'$, the RUL at t_n can be better defined as

$$L_D^2(t_n) = \inf\{t : X(t_n + t) \geq D | x_1 < D, \ldots, x_n < D\}. \tag{24}$$

Since the Wiener process owns the Markov property, $L_D^2(t_n)$ is equivalent to the first passage time to the threshold $D - x_n$, i.e. $L_D^2(t_n) \overset{d}{=} T_{D-x_n}$. If the available observations are only $\mathsf{y} = (y_1, \ldots, y_n)'$, which are contaminated by measurement errors, the RUL at t_n should be modified as

$$L_D^3(t_n) = \inf\{t : X(t_n + t) \geq D | y_1, \ldots, y_n\}.$$

In particular, when the Wiener process is the true model, the RUL at t_n can be calculated by

$$L_D^3(t_n) = \mathbb{E}_{X_n}[T_{D-X_n} | y_1, \ldots, y_n], \tag{25}$$

with the PDF

$$f_{L_D^3(t_n)}(l) = \int_{-\infty}^{D} f_{T_{D-x_n}}(l)\ p_{X_n|\mathbf{Y},\theta}(x_n|\mathbf{y},\theta)dx_n, \tag{26}$$

where $p_{X_n|\mathbf{Y},\theta}(x_n|\mathbf{y},\theta)$ is the conditional density of X_n, the true degradation at t_n.

With the parameter estimate $\hat{\theta}$, the PDFs of the first two RULs (23) and (24) can be easily obtained by simply replacing the unknown parameters (μ, σ, b) by $(\hat{\mu}, \hat{\sigma}, \hat{b})$ in the Wiener process. However, for the third RUL (25), the PDF shown in (26) includes an integral over the product of an IG PDF and a complicated density, for which a direct computation may be time-consuming. Alternatively we can utilize the Z_n's already generated from the Gibbs sampling to make an approximation of $f_{L_D^3(t_n)}(l)$, which is

$$\hat{f}_{L_D^3(t_n)}(l) = \frac{1}{Q_1}\sum_{q=1}^{Q_1} f_{T_{D-\tilde{z}_n^q}}(l), \tag{27}$$

where Q_1 is a predetermined number as the same as the previous.

5 Monte Carlo Simulations

To investigate the performance of the proposed degradation model, Monte Carlo simulation experiments are conducted based on a Wiener process with the drift rate $\mu = 5$, the diffusion coefficient $\sigma = 1$, the transformed time scale $\Lambda(t) = t^b$ with $b = 1$. For illustration purpose, the measurement error is assumed to follow a mixture of a normal distribution and a logistic distribution,

$$\epsilon_t \sim (1-w)\epsilon_{t,1} + w\epsilon_{t,2},$$
$$\epsilon_{t,1} \sim N(0,4),\ \epsilon_{t,2} \sim Logistic(0,2),\ w \sim B(p),$$

where w is a Bernoulli variable with $P(W = 1) = p$. Set $p = 0.5$.

In the entire simulation study, we assume that m units are subject to a degradation test under the same testing conditions and measured at n time points, $\mathbf{t} = (1, \ldots, n)'$. To show the effect of number of units and number of measurements for each unit on the estimation efficiency, different combinations of (m, n) are considered and the degradation data are generated accordingly. For comparison purpose, two models are used to fit data and estimate the model parameters. One is the proposed model, $Y(t) = X(t) + \epsilon_t$, $\epsilon_t \sim Logistic(0, s)$, and the other is the traditional one based on the Weiner process with normal distributed measurement errors or Gaussian noises, $Y(t) = X(t) + \epsilon_t$, $\epsilon_t \sim N(0, \gamma^2)$. For all the experiments, 5000 replications are run in each simulation. The convergence criterion adopted to terminate the EM algorithm is $|(\theta^{\text{new}} - \theta^{\text{old}})/\theta^{\text{old}}| \leq 0.005$.

Table 1 reports the estimation results for the three parameters of the Wiener process, which is the true degradation, as well as the Kullback-Leibler divergence (KL divergence) between the true and the estimated PDFs of the failure time or first passage time T_D given in (22). That is, if we denote the true parameters by θ_0,

$$\mathrm{KL}(f_{T_D}(t;\theta_0) \parallel f_{T_D}(t;\hat{\theta})) = -\int_0^\infty f_{T_D}(t;\theta_0) \ln\left\{\frac{f_{T_D}(t;\hat{\theta})}{f_{T_D}(t;\theta_0)}\right\} dt.$$

A smaller value of the KL divergence implies a better estimation of the PDF.

As described above, the data are generated from the model with mixed measurement errors, while the models used for estimations are assumed to have logistic distributed and normal distributed measurement errors respectively, both of which are wrong models. Thus the results in Table 1 tend to indicate the robustness of different model assumptions.

Table 1 summarizes the empirical mean and root-mean-square-error (rmse) of the estimated parameters as well as the empirical mean and standard error (se) of the KL divergence under different combinations of the model and the values of m and n. From the results in the table, we see that the proposed degradation model, i.e. the model with logistic distributed measurement errors, is more robust than the model with normal errors in almost all cases we considered. In terms of the empirical mean, the estimates of the parameters (μ, σ, b) are comparative under the two models,

Table 1 The estimation results for the two models: mean(rmse) for $(\hat{\mu}, \hat{\sigma}, \hat{b})$ and mean(se) for KL divergence

	Logistic error				Normal error			
(m, n)	$\hat{\mu}$	$\hat{\sigma}$	$\hat{b}$	KL	$\hat{\mu}$	$\hat{\sigma}$	$\hat{b}$	KL
(1, 25)	5.0888	0.6457	0.9967	29.202	5.0936	0.5262	0.9966	97.902
	(0.8030)	(0.5345)	(0.0488)	(66.593)	(0.8154)	(0.7023)	(0.0496)	(148.52)
(1, 50)	5.0189	0.7543	1.0010	3.2220	5.0167	0.6458	1.0012	14.329
	(0.6333)	(0.3767)	(0.0319)	(7.3366)	(0.6392)	(0.5393)	(0.0323)	(29.075)
(2, 10)	5.1367	0.7303	0.9929	79.217	5.3410	0.6006	0.9736	189.61
	(0.8541)	(0.4307)	(0.0756)	(214.11)	(0.9690)	(0.5412)	(0.0833)	(475.44)
(2, 25)	5.0351	0.8271	0.9994	3.9737	5.0338	0.7963	0.9996	7.5851
	(0.5485)	(0.3067)	(0.0340)	(8.9491)	(0.5561)	(0.3865)	(0.0346)	(21.616)
(2, 50)	5.0146	0.8988	1.0002	0.6098	5.0134	0.8806	1.0002	0.9798
	(0.4387)	(0.2010)	(0.0219)	(1.2783)	(0.4409)	(0.2476)	(0.0220)	(3.4132)
(5, 10)	5.0664	0.8891	0.9965	12.287	5.0814	0.8894	0.9954	16.4223
	(0.5304)	(0.2591)	(0.0468)	(22.781)	(0.5566)	(0.3052)	(0.0493)	(35.112)
(5, 25)	5.0206	0.9292	0.9995	0.7790	5.0231	0.9554	0.9994	0.8438
	(0.3481)	(0.1585)	(0.0213)	(1.1583)	(0.3544)	(0.1889)	(0.0217)	(1.4650)
(5, 50)	5.0135	0.9653	0.9997	0.1554	5.0146	0.9748	0.9996	0.1549
	(0.2795)	(0.0881)	(0.0141)	(0.2099)	(0.2802)	(0.0944)	(0.0141)	(0.2115)

[a]The true parameters are $\mu_0 = 5$, $\sigma_0 = 1$, and $b_0 = 1$. And $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$

although the proposed model outperforms the model with normal errors more often. However, if the rmse is concerned, the assumption of logistic error is clearly superior to the normality assumption as the former has smaller rmse values. Moreover, except the case of (5, 50), the KL divergence for the proposed model is generally less than that with normal errors with respect to both the mean and the standard error. This means that the failure time distribution obtained from the proposed model could better represent the true one. So it may be concluded that the proposed model is robust in dealing with non-normal measurement errors.

To further demonstrate the robustness of the proposed model, the estimation of the RUL is also considered for a comparison between the two models. Since the RUL is a random variable, it expectation named 'mean residual useful life' (MRUL) is used for estimation. The MRUL at t_n is the expectation of the RUL $L_D^3(t_n)$, and can be obtained by integrating the PDF (26), which is

$$\text{MRUL}(t_n) = \int_0^\infty l \cdot f_{L_D^3(t_n)}(l) dl.$$

The estimated MRUL at t_n for the degradation process can be calculated by substituting the parameter θ by $\hat{\theta}$, i.e. $\widehat{\text{MRUL}}(t_n) = \text{MRUL}(t_n)|_{\theta=\hat{\theta}}$. If the PDF $f_{L_D^3(t_n)}(l)$ is complicated, the estimated MRUL also can be given by integrating over (27), $\widehat{\text{MRUL}}(t_n) = \int_0^\infty l \cdot \hat{f}_{L_D^3(t_n)}(l) dl = \frac{1}{Q_1} \sum_{q=1}^{Q_1} \int_0^\infty l \cdot f_{T_{D-\tilde{z}_n^q}}(l) dl$.

With the MLEs of the parameters of the two models, the corresponding MRULs could be easily estimated. The performances of the two MRUL estimates are measured by the average relative error to a benchmark RUL,

$$\text{average relative error} = \frac{1}{N} \sum_{k=1}^{N} \left| \frac{\widehat{\text{MRUL}}_k(t_n)}{\text{RUL}_{bm,k}} - 1 \right|$$

where the subscript k denotes the kth replication of total N replications and $N = 5000$ in our experiments. For a better comparison, two benchmarks are chosen, the real RUL, $D - x_n$, and the real MRUL obtained from (22), $\int_0^\infty t \cdot f_{T_{D-x_n}}(t) dt$, where x_n is the observation at t_n and also the last observation for a unit. The average relative errors of the estimated MRUL's to the two benchmarks are reported in Tables 2 and 3. (We set the threshold $D = 300$ for all cases.)

From both tables, we observe that the proposed model has smaller average relative errors in all cases than the model with normal errors. This implies that, in term of the RUL prediction, the assumption of logistic distribution for the measurement error is better than the assumption of normal distribution, and thus the proposed degradation model can be considered more robust. In particular, the robustness of the proposed model is rather significant when the number of observations for a unit is small, such as the cases with $n = 10$. So we may conclude again that the model with logistic distributed measurement errors is preferred when modelling degradation data with non-normal measurement errors.

Table 2 The average relative errors of the estimated MRUL's to the real RUL

(m^b, n)	Model[a]	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
(1, 25)	1	0.0543				
	2	0.0567				
(1, 50)	1	0.0720				
	2	0.0722				
(2, 10)	1	0.1304	0.1312			
	2	0.1633	0.1663			
(2, 25)	1	0.0414	0.0412			
	2	0.0432	0.0432			
(2, 50)	1	0.0654	0.0656			
	2	0.0656	0.0657			
(5, 10)	1	0.0787	0.0791	0.0785	0.0786	0.0781
	2	0.0827	0.0833	0.0825	0.0829	0.0823
(5, 25)	1	0.0330	0.0332	0.0333	0.0337	0.0336
	2	0.0332	0.0335	0.0336	0.0340	0.0340
(5, 50)	1	0.0601	0.0602	0.0596	0.0608	0.0596
	2	0.0603	0.0605	0.0598	0.0610	0.0597

[a]1—The model with logistic errors, 2—the model with normal errors
[b]The value of m represents the number of units

Table 3 The average relative errors of the estimated MRUL's to the real MRUL

(m^b, n)	Model[a]	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
(1, 25)	1	0.0490				
	2	0.0519				
(1, 50)	1	0.0512				
	2	0.0514				
(2, 10)	1	0.1292	0.1293			
	2	0.1621	0.1650			
(2, 25)	1	0.0308	0.0308			
	2	0.0330	0.0329			
(2, 50)	1	0.0406	0.0402			
	2	0.0408	0.0404			
(5, 10)	1	0.0753	0.0754	0.0754	0.0751	0.0753
	2	0.0796	0.0797	0.0797	0.0795	0.0796
(5, 25)	1	0.0197	0.0195	0.0195	0.0196	0.0193
	2	0.0201	0.0200	0.0199	0.0201	0.0197
(5, 50)	1	0.0319	0.0316	0.0314	0.0314	0.0317
	2	0.0323	0.0319	0.0318	0.0318	0.0321

[a]1—The model with logistic errors, 2—the model with normal errors
[b]The value of m represents the number of units

6 An Example

6.1 Model Fitting

In this section, we return to the example on the capacity loss of a lithium-ion battery mentioned in the introduction. The data collected at 168 time points are fitted by three models, including the Wiener process without measurement error and the two models discussed in the previous section, which are the Weiner process with normal distributed measurement errors and the proposed model. The transformed time scale is also set as $\Lambda(t) = t^b$. Since all three models have the Weiner process describing the true degradations, their expected degradation processes are of a same form, which is $\mathrm{E}[Y(t)] = \mathrm{E}[X(t)] = \mu t^b$.

Table 4 reports the parameter MLEs, the corresponding maximum log-likelihood values and the Akaike information criterion (AIC) values for the three models. It is found that the AIC value for the proposed model is the smallest, and the other two values are slightly larger. This observation implies that the proposed model outperforms the other two models in fitting the data of battery capacity. Another interesting thing is that, the Wiener process without measurement error seems to have a little bit better performance than the model with normal errors, although their AIC values are rather close to each other.

To intuitively compare the performances of the three models, their estimated expected degradations are plotted against number of cycles in Fig. 2, together with the observation data. The expected degradations for the three models are $0.0339t^{0.6296}$, $0.0320t^{0.6413}$, and $0.0272t^{0.6738}$ respectively. The figure shows that the expected degradation process obtained from the proposed model (red line) is relatively closer to the path consisting of the observation data and less affected by the data fluctuation.

For further comparison, three estimated density functions of the failure time or first passage time T_D are plotted in Fig. 3 and some characteristic values are listed in Table 5. As indicated in [26], a lithium-ion battery is considered to reach the end-of-life criteria once there is a 20% fade in capacity. Since the original capacity of Battery #6 is 2.0353 Ahr, the threshold for degradation models is set as $D = 2.0353 \times 20\% = 0.4071$. So the associated failure time T_D as a random variable can be represented by its density function (22) with $D = 0.4071$ and the parameters given by the estimates in Table 4. As a direct result, we can obtain the estimated failure time densities for the three models, as well as their mean, mode, media, and 5%-, 95%-percentiles.

Table 4 The fitting results for the three models

Model	MLE	Log-likelihood	AIC
Wiener	$\hat{\mu} = 0.0339, \hat{\sigma} = 0.0609, \hat{b} = 0.6296$	396.2370	−789.474
Wiener + normal	$\hat{\mu} = 0.0320, \hat{\sigma} = 0.0537, \hat{b} = 0.6413, \hat{\gamma} = 0.0065$	396.6125	−789.225
Wiener + logistic	$\hat{\mu} = 0.0272, \hat{\sigma} = 0.0334, \hat{b} = 0.6738, \hat{s} = 0.0061$	403.2261	−802.452

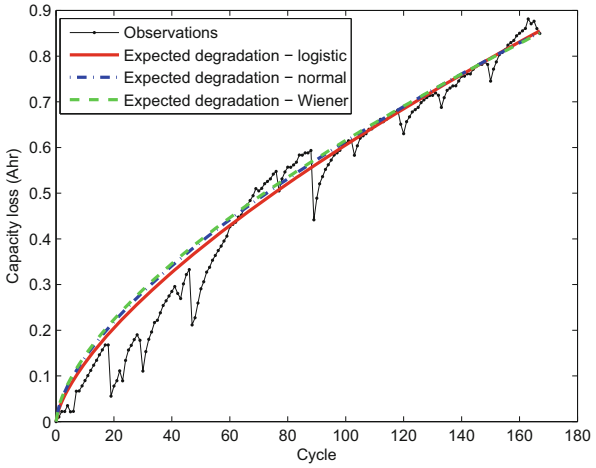


Fig. 2 The estimation results for the expected degradation of the capacity loss

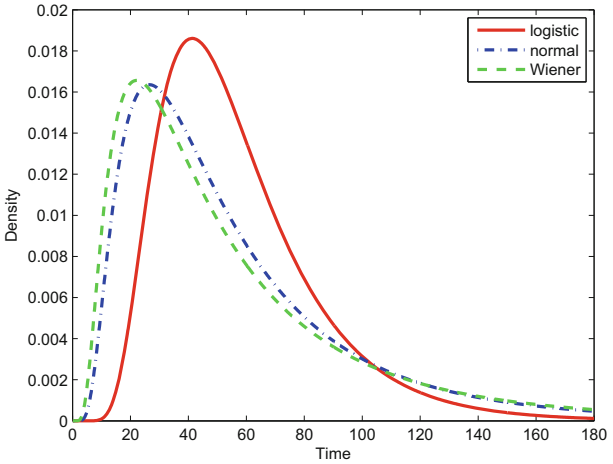


Fig. 3 The estimated density functions of T_D: $D = 0.4071$

Table 5 Some characteristics of the density functions of T_D: $D = 0.4071$

	Mean	Mode	Media	5%-percentile	95%-percentile
Wiener	57.9	22.1	42.6	12.1	155.5
Wiener + normal	57.6	26.5	44.9	14.4	142.9
Wiener + logistic	57.4	41.3	51.6	24.3	110.3

As shown in Fig. 3, the density function for the proposed model (red line) is thinner and taller than the other two density functions. A similar phenomenon is also observed when comparing the 5%- and 95%-percentiles. Clearly, the 5%-percentile for the proposed model is larger than those for the other two models, while the 95%-percentile is smaller. This means that the 5%–95% range of failure time provided by the proposed model is narrower than the other two ranges. The reason for the difference in the shape of the three densities is that, the proposed model assumes the errors following the logistic distribution, which has relatively heavy tails and is able to absorb the large fluctuations in the data, and thus the true degradations described the Wiener process has a smaller diffusion parameter, leading to a first passage time with a smaller variance.

Other interesting findings are that, the density function for the proposed model moves right forward compared to the other two densities, and the corresponding mode and media take much larger values. This also can be explained by the assumption of the logistic measurement error in the proposed model, which reduces the impact caused by the large drops in capacity loss and then lifts up the lower tail of the failure time density.

Although the expected degradation paths and the means are rather close for the three models, the different assumptions on measurement error would undoubtedly result in different judgments in decision-making. For example, the 95% percentiles for the Wiener-based models without or with normal errors may tell that, about 5% of the #6-type batteries would be of low quality with the lifetime less than 15 cycles, but this judgment may be too pessimistic if the proposed model is used for analysis, which implies that even the 5% worst batteries could last 24 cycles. In this case decisions for burn-in, maintenance, etc., should be carefully made.

To make the analysis more complete, the residuals are also computed for the two models assuming the existence of measurement errors, and tested for their independence and distribution. As estimates for the measurement errors, the residuals at the $n = 168$ time points are calculated by $\hat{\epsilon}_i = y_i - E(X_i|y_1, \ldots, y_n; \hat{\theta})$, $i = 1, \ldots, n$, where $E(X_i|y_1, \ldots, y_n; \hat{\theta})$ is the expected true degradation at the ith time point given the observed degradations and the MLEs for parameters. We use the ACF plot and Ljung-Box test to diagnose the autocorrelations in the residual series. Surprisingly, the error terms for the proposed model, i.e. the model with logistic distributed measurement errors, exhibit low autocorrelations with absolute values less than 0.2 for greater than 0 lags. There is not enough evidence to reject the null hypothesis that the residuals are not autocorrelated since the p-value of the test is 0.5510. However, for the model assuming normal distributed errors, the test is rejected as the autocorrelations among the residuals are relatively high especially when the lag is one. Thus it is more reasonable to treat the error terms for the proposed model i.i.d.. We further use the Kolmogorov-Smirnov test to test whether the residuals could be fitted by the logistic or normal distribution. Unfortunately, both tests are rejected, which indicate that both distributions could not really fit the error terms well.

Although the logistic distribution and the normal distribution are not ideal in capturing the feature of the error terms, we may still suggest to adopt the proposed

model in modeling degradation data with large measurement errors and further decision-making. This is because the proposed model has a better performance in model-fitting, which is then likely to yield a more accurate density function of failure time.

6.2 The Confidence Intervals

Once the proposed model is chosen, further statistical inferences could be made with the variance-covariance matrix of the unknown parameters. We focus on the construction of confidence interval here. As discussed in Sect. 3.4, the variance-covariance matrix can be estimated by either the Louis's method or the bootstrap method, which are respectively

$$\hat{\Sigma}_{\hat{\theta}} = \begin{pmatrix} 4.2604 & 2.7454 & -27.9223 & -0.0331 \\ 2.7454 & 2.0880 & -19.6299 & -0.0399 \\ -27.9223 & -19.6299 & 199.5917 & 0.2385 \\ -0.0331 & -0.0399 & 0.2385 & 0.0064 \end{pmatrix} \times 10^{-4},$$

$$\hat{\Sigma}_{\hat{\theta},boot} = \begin{pmatrix} 2.0745 & 1.1462 & -13.8931 & -0.0022 \\ 1.1462 & 0.9115 & -9.8911 & -0.0105 \\ -13.8931 & -9.8911 & 133.9039 & 0.0025 \\ -0.0022 & -0.0105 & 0.0025 & 0.0052 \end{pmatrix} \times 10^{-4},$$

where $\hat{\Sigma}_{\hat{\theta}}$ is computed based on $Q_1 = 1500$ samples of the last iteration in the MCEM procedure, and $\hat{\Sigma}_{\hat{\theta},boot}$ is the sample covariance matrix for 500 separate estimates of θ, which are obtained from 500 bootstrapped samples. For both matrices, the root of the diagonal elements are the estimated standard errors for the MLEs, $\hat{\mu}$, $\hat{\theta}$, $\hat{b}$ and $\hat{s}$.

Table 6 summarizes the parameter estimates, the estimated se's as well as the 5% and 95% sample percentiles from the bootstrapped $\hat{\theta}$'s. We find that, with the assumption of asymptotic normality for MLE, the Louis's se leads to a 90% CI with a negative left-end-point for $\hat{\mu}$, i.e. $(-0.0066, 0.0610)$. This CI is odd since either Figs. 1 or 2 evidently indicates a positive drift rate parameter μ. A possible reason for the negative left-end-point may be that, there is only one battery used for analysis, which provides only limited information for the drift rate, and thus results in a large se. Furthermore the limited information would make the normality assumption not really valid. From the 500 separate estimates of θ, we calculate out the sample skewness and sample excess kurtosis, which are 0.8367 and 0.5417 respectively. This implies that the distribution of $\hat{\mu}$ is skewed to the right and has heavier tails than the normal distribution. Therefore it may be improper to use the common, normal-distribution-based approach to construct CI for $\hat{\mu}$ in this example. An alternative way for the CI construction is to directly utilize the sample percentiles

Table 6 Estimation results for the proposed model

	Estimate	Se (Louis)	Se (Bootstrap)	(5%-, 95%-) Sample percentiles (Bootstrap)
μ	0.0272	0.0206	0.0144	(0.0065, 0.0524)
σ	0.0334	0.0144	0.0095	(0.0164, 0.0479)
b	0.6738	0.1413	0.1157	(0.5465, 0.9304)
s	0.0061	0.0008	0.0007	(0.0050, 0.0073)

of the bootstrapped estimates of μ. Since there are 500 estimates, we take the 25th and 475th smallest estimates as the 5% and 95% sample percentiles and construct a 90% CI with the two sample percentiles as end points; that is (0.0065, 0.0524) for $\hat{\mu}$ as shown in Table 6. For the other three parameters σ, b and s, the CIs could be obtained either by the normal approximation or from their bootstrapped estimates in a similar way as μ. The latter are given in the last column of Table 6.

7 Conclusions

In this chapter, a degradation model based on the Wiener process was developed by assuming measurement errors subject to the logistic distribution, which has longer tails than the normal distribution. Considering the complication of the associated likelihood function, the MCEM method together with the Gibbs sampling were used for parameter estimation. And an iterative algorithm was developed for a quick calculation of the likelihood value at the MLE. Besides, the estimation of remaining useful life was also discussed. The estimation results for the model parameter and the RUL showed that the proposed degradation model has better performances than the Wiener process with Gaussian noises, and seems more robust in dealing with non-normal measurement errors. The robustness of the proposed model was also demonstrated by an example.

Although only the Wiener process is considered is this work, other traditional degradation processes, such as the Gamma process and the inverse Gaussian process, also can be extended by assuming logistic distributed measurement errors. But it is clear that the parameter estimation would be still a difficult job when using this type of extended models. As shown in the chapter, a feasible solution to the parameter estimation problem goes to the combination of the EM algorithm and a proper sampling method. Once the parameter estimates are attained, the corresponding likelihood value can be calculated by the algorithm proposed in this chapter. Furthermore, in simulation experiments, more degradation data with various measurement error assumptions could be considered for testing, so that the performances of different models could be better evaluated.

The work in this chapter was motivated by the performance degradation data of lithium-ion battery, but used only the degradation data itself for modelling. In fact,

batteries are usually used in time-varying environments, where exist dynamic life-affecting environmental variables, such as voltage, temperature and humidity. Hence it must be rather interesting to include these dynamic variables into degradation models like Hong et al. [2] did so that batteries' reliability can be predicted more accurately.

Acknowledgements This work was published on Quality and Reliability Engineering International. This work was supported by the National Natural Science Foundation of China [Grant Number 71401146]; the MOE (Ministry of Education) Key Laboratory of Econometrics; and Fujian Key Laboratory of Statistical Sciences, China.

Appendix 1

This appendix presents the technical details for the inference in Sect. 3.4. Recall the parameter $\theta = (\mu, \sigma, s)'$. The first and second partial derivatives of the loglikelihood function in (8) with respect to θ are as follows. Let $\kappa_{l,i} = x_{l,i} - x_{l,i-1} - \mu\lambda_{l,i}$.

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\kappa_{l,i}$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{1}{\sigma}\sum_{l=1}^{m} n_l + \frac{1}{\sigma^3}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{\kappa_{l,i}^2}{\lambda_{l,i}}$$

$$\frac{\partial \ell}{\partial s} = \frac{1}{s^2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}(y_{l,i} - x_{l,i}) - \frac{1}{s}\sum_{i=1}^{m} n_l - \frac{2}{s^2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{y_{l,i} - x_{l,i}}{1 + e^{(y_{l,i}-x_{l,i})/s}}$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{1}{\sigma^2}\sum_{l=1}^{m}\Lambda_{l,n_l}$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{1}{\sigma^2}\sum_{l=1}^{m} n_l - \frac{3}{\sigma^4}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{\kappa_{l,i}^2}{\lambda_{l,i}}$$

$$\frac{\partial^2 \ell}{\partial s^2} = -\frac{2}{s^3}\sum_{l=1}^{m}\sum_{i=1}^{n_l}(y_{l,i} - x_{l,i}) + \frac{1}{s^2}\sum_{i=1}^{m} n_l + \frac{4}{s^3}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{y_{l,i} - x_{l,i}}{1 + e^{(y_{l,i}-x_{l,i})/s}}$$
$$-\frac{2}{s^4}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{(y_{l,i} - x_{l,i})^2 e^{(y_{l,i}-x_{l,i})/s}}{[1 + e^{(y_{l,i}-x_{l,i})/s}]^2}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\kappa_{l,i}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial s} = \frac{\partial^2 \ell}{\partial \sigma \partial s} = 0.$$

If the time transformation function $\Lambda(t)$ contains an extra parameters b, i.e. $\Lambda(t) = \Lambda(t; b)$, we also need the first and second partial derivatives of the loglikelihood function (8) with respect to b.

$$\frac{\partial \ell}{\partial b} = -\frac{1}{2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{1}{\lambda_{l,i}}\frac{\partial \lambda_{l,i}}{\partial b} + \frac{1}{\sigma^2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\left(\frac{\kappa_{l,i}\mu}{\lambda_{l,i}} + \frac{\kappa_{l,i}^2}{2\lambda_{l,i}^2}\right)\frac{\partial \lambda_{l,i}}{\partial b}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial b^2} &= \frac{1}{2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{1}{\lambda_{l,i}^2}\left(\frac{\partial \lambda_{l,i}}{\partial b}\right)^2 - \frac{1}{2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{1}{\lambda_{l,i}}\frac{\partial^2 \lambda_{l,i}}{\partial b^2} \\ &\quad + \frac{1}{\sigma^2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\left(\frac{\kappa_{l,i}\mu}{\lambda_{l,i}} + \frac{\kappa_{l,i}^2}{2\lambda_{l,i}^2}\right)\frac{\partial^2 \lambda_{l,i}}{\partial b^2} \\ &\quad - \frac{1}{\sigma^2}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\left(\frac{\mu^2}{\lambda_{l,i}} + \frac{2\kappa_{l,i}\mu}{\lambda_{l,i}^2} + \frac{\kappa_{l,i}^2}{\lambda_{l,i}^3}\right)\left(\frac{\partial \lambda_{l,i}}{\partial b}\right)^2\end{aligned}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial b} = -\sum_{l=1}^{m}\sum_{i=1}^{n_l}\frac{\mu}{\sigma^2}\frac{\partial \lambda_{l,i}}{\partial b}$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial b} = -\frac{2}{\sigma^3}\sum_{l=1}^{m}\sum_{i=1}^{n_l}\left(\frac{\kappa_{l,i}\mu}{\lambda_{l,i}} + \frac{\kappa_{l,i}^2}{2\lambda_{l,i}^2}\right)\frac{\partial \lambda_{l,i}}{\partial b}$$

$$\frac{\partial^2 \ell}{\partial s \partial b} = 0.$$

Appendix 2

This appendix presents part of codes for Monte Carlo simulations. All codes were compiled based Matlab software, and the related files are available upon request.

```
%%
%%%%%%%%%%%%%%%% Main file %%%%%%%%%%%%%%%%
%%


%% Define parameters and original process

Nrep = 5000;                    %% number of replications
M = 2;                            %% number of units
N = 25;                           %% sample size
NSEM = 1000;
thin = 20;
Nuse = 1500;
Nskip = 500;
Nkeep = 1 + (Nuse - 1)*thin;
NGib = Nkeep + Nskip;      %% numbers to control MCMC and Gibbs
```

```
D = 300;                          %% predetermined threshold

mu = 5;                           %% drift parameter
sigma = 1;                        %% diffusion parameter
b = 1;
s = 2;
u = 0;
gamma = 2;
p=0.5;
time = (1:N)';
log_time = log(time);

for nr = 1:Nrep DATA = cell(M,1);

  DATAA = cell(M,1);
  ZMAT = cell(M,1);
  FPT = zeros(M,1);

  for mm = 1:M
    index_neg = 1;
    while index_neg>0
      [dataA,fpt] = BrownM(N,mu,sigma,b,D);
      terror1 = -s*log(1./rand(N,1)-1) + u;   %% logistic error
      terror2 = randn(N,1)*gamma;                 %% normal error
      pp = binornd(1,p,[N,1]);
      terror = pp.*terror1+(1-pp).*terror2;   %% mixture
      data = dataA + terror;
      index_neg = sum(data(1:10,:)<0);
    end
    DATA{mm,1}=data;
    DATAA{mm,1} = dataA;
    FPT(mm,1) = fpt;
  end


%% MLE assuming Gaussian error

  theta1 =[mu,sigma,b,s];
  theta_1 = mleWienError(N,M,DATA,time,theta1);


%% MLE assuming logistic error

  mu0 =data(end)/time(end);
  sigma0 = sigma;
  b0 = b;
  s0 = s;
  theta =[mu0,sigma0,b0,s0];
  Z = mu0*time.^b0;
  k=1;
  while k<=NSEM
    for mm=1:M
      data = DATA{mm,1};
      Zmat0 = zeros(Nkeep,N);
      LAMBDA = time.^theta(3);
      lambda = LAMBDA - [0;LAMBDA(1:end-1)];
      for i=1:NGib
        for j=1:N
          if j==1
            Z(1) = Gibbs1(0,Z(2),lambda(1),lambda(2),data(1),theta);
```

```
        elseif j<N
          Z(j) = Gibbs1(Z(j-1),Z(j+1),lambda(j),lambda(j+1),data(j),theta);
        else
          Z(N) = Gibbs2(Z(N-1),lambda(N),data(N),theta);
        end
      end
      if i>Nskip
        Zmat0(i-Nskip,:)=Z';
      end
    end
    Zmat = zeros(Nuse,N);
    for i = 1:Nuse
      Zmat(i,:) = Zmat0(1+(i-1)*thin,:);
    end
    ZMAT{mm,1}=Zmat;
  end

  theta_new =mleMCEM(N,M,DATA,time,Nuse,ZMAT,theta);
  diff = max(abs((theta_new-theta)./theta));
  if diff<=0.005
    theta = theta_new;
    break
  end
  theta = theta_new;
  k = k+1;

end


%% MLE of each parameter for the two models

muh = theta(1);
sigmah = theta(2);
bh = theta(3);
sh = theta(4);

muh1 = theta_1(1);
sigmah1 = theta_1(2);
bh1 = theta_1(3);
gamh1 = theta_1(4);


%% KL divergence

tempfun = (@(t) -sqrt(D^2/2/pi/sigma^2./t.^(3*b)).*exp(-(mu*t.^b-D).^2...
/2/sigma^2./t.^b)*b.*t.^b.*(log(sigma/sigmah) + 1/2*log(t.^(3*b-3*bh))...
-(muh*t.^bh-D).^2/2/sigmah^2./t.^bh + (mu*t.^b-D).^2/2/sigma^2./t.^b...
+ log(bh/b) + (bh-b).*log(t)));

tempfun1 = (@(t) -sqrt(D^2/2/pi/sigma^2./t.^(3*b)).*exp(-(mu*t.^b-D).^2...
/2/sigma^2./t.^b)*b.*t.^b.*(log(sigma/sigmah1) + 1/2*log(t.^(3*b-3*bh1))...
-(muh1*t.^bh1-D).^2/2/sigmah1^2./t.^bh1 + (mu*t.^b-D).^2/2/sigma^2./t.^b...
 + log(bh1/b) + (bh1-b).*log(t)));

KL = quadgk(tempfun,0,Inf);

KL1 = quadgk(tempfun1,0,Inf);


results = [theta theta_1 KL KL1];
fid = fopen('estimate.txt','a+');
fprintf(fid,'%4.4f %4.4f %4.4f %4.4f %4.4f %4.4f %4.4f
%4.4f %4.4f %4.4f \r\n',results');
```

```
  fclose(fid);
end                          %% end of repetition


%%
%%%%%%%%%%%%%%%% Function BrownM %%%%%%%%%%%%%%%%
%%

function [data,FPT] = BrownM(Npoints,Mean,Std,b,D)
time = (1:Npoints)';
time_cont = ((Npoints+0.0001):0.0001:2*Npoints)';
time_temp = [time;time_cont];
N_temp = length(time_temp);
lambda = time_temp.^b;
dt = lambda - [0;lambda(1:N_temp-1)];
temp = normrnd(Mean*dt,Std*sqrt(dt));
data_temp = cumsum(temp);
data = data_temp(1:Npoints);
pos_temp = find(data_temp>=D);
FPT = time_temp(pos_temp(1));
end


%%
%%%%%%%%%%%%%%%% Function mleWienError %%%%%%%%%%%%%%%%
%%

function f = mleWienError(N,M,DATA,time,theta0)
one1 = ones(N-1,1);
DDATA = cell(M,1);
for nm=1:M
  data = DATA{nm,1};
  DDATA{nm,1} = data - [0;data(1:end-1)];
end

  function f1 = mle2(theta1)
    b = theta1(1);
    phi = theta1(2);
    LAM = time.^b;
    lam = LAM - [0;LAM(1:end-1)];
    D = diag(lam);
    P = diag([1;2*one1]) + diag(-one1,-1) + diag(-one1,1);
    PHI = phi*D+P;
    invPHI = PHI^(-1);
    temp1 = 0;
    temp2 = 0;
    temp3 = 0;
    for nm = 1:M
      Ddata = DDATA{nm,1};
      temp1 = temp1 + lam'*invPHI*Ddata;
      temp2 = temp2 + lam'*invPHI*lam;
    end
    mu = temp1/temp2;
    for nm = 1:M
      Ddata = DDATA{nm,1};
      temp = Ddata - mu*lam;
      temp3 = temp3 + temp'*invPHI*temp;
    end
```

```
    gam2 = temp3/N/M;
    f1 = -1/2*M*log(det(PHI)) - N*M/2*log(gam2);
    f1 = -f1;
  end

th10 = [theta0(3),theta0(2)/theta0(4)/1.5];
th1hat = fmincon(@mle2,th10,[0,-1],0);
bh = th1hat(1);
phih = th1hat(2);
LAM = time.^bh;
lam = LAM - [0;LAM(1:end-1)];
D = diag(lam);
P = diag([1;2*one1]) + diag(-one1,-1) + diag(-one1,1);
PHI = phih*D+P;
invPHI = inv(PHI);
temp4 = 0;
temp5 = 0;
temp6 = 0;
for nm=1:M
  Ddata = DDATA{nm,1};
  temp4 = temp4 + lam'*invPHI*Ddata;
  temp5 = temp5 + lam'*invPHI*lam;
end
muh = temp4/temp5;
for nm=1:M
  Ddata = DDATA{nm,1};
  temp = Ddata - muh*lam;
  temp6 = temp6 + temp'*invPHI*temp;
end
gam2h = temp6/N/M;
sigma2h = gam2h*phih;
gamh = sqrt(gam2h);
sigmah = sqrt(sigma2h);
f = [muh sigmah bh gamh];
end


%%
%%%%%%%%%%%%%%% Function Gibbs1 %%%%%%%%%%%%%%%
%%

function f = Gibbs1(a1,a2,t1,t2,y1,theta)
mu = theta(1);
sigma = theta(2);
b = theta(3);
s = theta(4);
sigma2 = sigma^2;
u2 = 1;
temp2 = 0;
temp4 = sigma*sqrt(t1*t2/(t1+t2));
temp31 = (a1*t2*s + a2*t1*s - sigma2*t1*t2)/(t1+t2)/s;
A1 = (a1+mu*t1)^2*t2*s + (a2-mu*t2)^2*t1*s - 2*sigma2*t1*t2*y1;
tempM1 = -(A1/(t1+t2)/s - temp31^2);
temp32 = (a1*t2*s + a2*t1*s + sigma2*t1*t2)/(t1+t2)/s;
A2 = (a1+mu*t1)^2*t2*s + (a2-mu*t2)^2*t1*s + 2*sigma2*t1*t2*y1;
tempM2 = -(A2/(t1+t2)/s - temp32^2);
if tempM1 < tempM2
  temp3 = temp31;
```

```
  M = exp(tempM1/temp4^2/2)*temp4*sqrt(2*pi);
else
  temp3 = temp32;
  M = exp(tempM2/temp4^2/2)*temp4*sqrt(2*pi);
end
while (u2>temp2)
  u1 = normrnd(temp3,temp4);
  temp1 = exp(-(y1-u1)/s);
  temp2 = (exp(-(u1-a1-mu*t1)^2/(2*sigma2*t1)...
  -(a2-u1-mu*t2)^2/(2*sigma2*t2))*temp1/(1+temp1)^2)/normpdf(u1,temp3,temp4);
  u2 = M*rand;
end
f = u1;
end


%%
%%%%%%%%%%%%%%% Function Gibbs2 %%%%%%%%%%%%%%%
%%

function f = Gibbs2(a1,t1,y1,theta)
mu = theta(1);
sigma = theta(2);
b = theta(3);
s = theta(4);
sigma2 = sigma^2;
u2 = 1;
temp2 = 0;
temp4 = sigma*sqrt(t1);
temp31 = a1 + mu*t1 - sigma2*t1/s;
A1 = (a1+mu*t1)^2 - 2*sigma2*t1*y1/s;
tempM1 = -(A1-temp31^2);
temp32 = a1 + mu*t1 + sigma2*t1/s;
A2 = (a1+mu*t1)^2 + 2*sigma2*t1*y1/s;
tempM2 = -(A2-temp32^2);
if tempM1 < tempM2
  temp3 = temp31;
  M = exp(tempM1/temp4^2/2)*temp4*sqrt(2*pi);
else
  temp3 = temp32;
  M = exp(tempM2/temp4^2/2)*temp4*sqrt(2*pi);
end
while u2 > temp2
  u1 = normrnd(temp3,temp4);
  temp1 = exp(-(y1-u1)/s);
  temp2 = exp(-(u1-a1-mu*t1)^2/(2*sigma2*t1))...
  *temp1/(1+temp1)^2/normpdf(u1,temp3,temp4);
  u2 = M*rand;
end
f = u1;
end


%%
%%%%%%%%%%%%%%% Function mleMCEM %%%%%%%%%%%%%%%
%%

function f=mleMCEM(N,M,DATA,time,Nuse,ZMAT,theta0)
```

```
one = ones(N,1);
ZMEAN = cell(M,1);
ZEND = zeros(M,1);
DZ2MEAN = cell(M,1);
for nm=1:M
  Zmat = ZMAT{nm,1};
  Zmean = mean(Zmat)';
  ZMEAN{nm,1} = Zmean;
  ZEND(nm,1) = Zmean(N);
  DZ = Zmat - [zeros(Nuse,1), Zmat(:,1:end-1)];
  DZ2MEAN{nm,1} = mean(DZ.^2)';
end

  function f1 = mles(s)
    f1=0;
    for nm=1:M
      data = DATA{nm,1};
      Zmat = ZMAT{nm,1};
      Zmean = ZMEAN{nm,1};
      temp1 = sum(data-Zmean);
      temp2 = repmat(data',Nuse,1) - Zmat;
      temp3 = mean(temp2./(1+exp(temp2/s)))';
      f1 = f1+ temp1 -N*s - 2*one'*temp3;
    end
  end

sh = fsolve(@mles,theta0(4));

  function f2 = mleb(b)
    LAM = time.^b;
    lam = LAM - [0;LAM(1:end-1)];
    mu = sum(ZEND)/LAM(end)/M;
    f2 = 0;
    temp4 = 0;
    for nm = 1:M
      DZ2mean = DZ2MEANnm,1;
      temp4 = temp4 + one'*(DZ2mean./lam)-mu^2*LAM(end);
    end
    temp5 = temp4/N/M;
    f2 = f2 -M*N*log(temp5) - M*one'*log(lam);
    f2 = -f2;
  end

bh = fminbnd(@mleb,0,100);

LAM = time.^bh;
lam = LAM - [0;LAM(1:end-1)];
muh = sum(ZEND)/LAM(end)/M;
temp6 = 0;
for nm = 1:M
  DZ2mean = DZ2MEANnm,1;
  temp6 = temp6 + one'*(DZ2mean./lam)-muh^2*LAM(end);
end
sigmah2 = temp6/N/M;
sigmah = sqrt(sigmah2);
f = [muh sigmah bh sh];
end
```

References

1. Wang, X. (2010). Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis, 101*,(2), 340–351.
2. Hong, Y. L., Duan, Y. Y., Meeker, W. Q., Stanley, D. L., & Gu, X. H. (2015). Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data. *Technometrics, 57*(2), 180–193.
3. Xu, Z. B., Hong, Y. L., & Jin, R. (2016). Nonlinear general path models for degradation data with dynamic covariates. *Applied Stochastic Models in Business and Industry, 32*(2), 153–167.
4. Sari, J. K., Newby, M. J., Brombacher, A. C., & Tang, L. C. (2009). Bivariate constant stress degradation model: Led lighting system reliability estimation with two-stage modelling. *Quality and Reliability Engineering International, 25*(8), 1067–1084.
5. Meeker, W. Q., & Escobar, L. (1998). *Statistical methods for reliability data*. New York: Wiley.
6. Ye, Z. S., & Xie, M. (2015). Stochastic modelling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry, 31*(1), 16–32.
7. Whitmore, G. A., & Schenkelberg, F. (1997). Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Analysis, 3*(1), 27.
8. Hu, C. H., Lee, M. Y., & Tang, J. (2015). Optimum step-stress accelerated degradation test for Wiener degradation process under constraints. *European Journal of Operational Research, 241*(2), 412–421.
9. Zhai, Q. Q., Ye, Z. S. (2017). RUL prediction of deteriorating products using an adaptive Wiener process model. *IEEE Transactions on Industrial Informatics, 13*(6), 2911–2921.
10. Wang, X. L., Jiang, P., Guo, B., & Cheng, Z. J. (2014). Real-time reliability evaluation with a general Wiener process-based degradation model. *Quality and Reliability Engineering International, 30*(2), 205–220.
11. Whitmore, G. A. (1995). Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Analysis, 1*(3), 307–319.
12. Ye, Z. S., Wang, Y., Tsui, K. L., & Pecht, M. (2013). Degradation data analysis using Wiener processes with measurement errors. *IEEE Transactions on Reliability, 62*(4), 772–780.
13. Tang, S. J., Yu, C. Q., Wang, X., Guo, X. S., & Si, X. S. (2014). Remaining useful life prediction of lithium-ion batteries based on the Wiener process with measurement error. *Energies, 7*(2), 520–547.
14. Li, J. X., Wang, Z. H., Zhang, Y. B., Fu, H. M., Liu, C. R., & Krishnaswamy, S. (2017). Degradation data analysis based on a generalized Wiener process subject to measurement error. *Mechanical Systems and Signal Processing, 94*, 57–72.
15. Zhai, Q. Q., & Ye, Z. S. (2017). Robust degradation analysis with non-Gaussian measurement errors. *IEEE Transactions on Instrumentation and Measurement, 66*(11), 2803–2812.
16. Saha, B., & Goebel, K. (2007). *Battery data set*. NASA Ames Prognostics Data Repository.
17. Friedman, J. H. (2016). Robust location and scale estimation with censored data. Available via http://poloclub.gatech.edu/idea2016/slides/keynote-friedman-kdd-idea16.pdf.
18. Wei, G. C. G., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association, 85*(411), 699–704.
19. Ripley, B. D. (1987). *Stochastic simulation*. New York: Wiley.
20. Son, K. L., Fouladirad, M., & Barros, A. (2016). Remaining useful lifetime estimation and noisy gamma deterioration process. *Reliability Engineering and System Safety, 149*, 76–87.
21. Celeux, G., & Diebolt, J. (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly, 2*, 73–82.
22. Celeux, G., Chauveau, D., & Diebolt, J. (1996). Stochastic versions of the EM algorithm: An experimental study in the mixture case. *Journal of Statistical Computation and Simulation, 55*(4), 287–314.

23. Bedair, K., Hong, Y. L., Li, J., & Al-Khalidi, H. R. (2016). Multivariate frailty models for multi-type recurrent event data and its application to cancer prevention trial. *Computational Statistics and Data Analysis, 101*, 161–173.
24. Louis, A. T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B: Methodological, 44*(2), 226–233.
25. Zhang, J. X., Hu, C. H., Si, X. S., Zhou, Z. J., & Du, D. B. (2014). Remaining useful life estimation for systems with time-varying mean and variance of degradation processes. *Quality and Reliability Engineering International, 30*(6), 829–841.
26. Dubarry, M., & Liaw, B. Y. (2009). Identify capacity fading mechanism in a commercial LiFePO4 cell. *Journal of Power Sources, 194*(1), 541–549.

An Introduction of Generalized Linear Model Approach to Accelerated Life Test Planning with Type-I Censoring

Rong Pan and Kangwon Seo

Abstract Accelerated life tests are often expensive and difficult to conduct. Failure time censoring is anticipated because some test units do not fail over the testing period even under the accelerated stress condition. Therefore, a test plan must be carefully designed to maximize its statistical efficiency. This chapter presents an approach to optimal test planning based on the proportional hazard model of accelerated life test data. It is shown that this approach can accommodate multiple stress factors and is applicable to any failure time distribution.

1 Problem Description

Life testing is commonly used for obtaining the lifetime information of a product. However, the test stress applied on the product needs to be elevated to a higher-than-normal-use stress level in order to observe test unit failures within a limited testing period. This is the so-called accelerated life test (ALT), which is, oftentimes, the only way to obtain a product's lifetime data beside of its field failure time observations. ALTs have been used for both production reliability quantification and product failure analysis. They are typically very expensive, demanding special testing chambers and a large amount of energy. If a test plan were not well thought-out before conducting the test, it could be possible that very few failures, even no failures, were observed during the testing period, thus no useful conclusions could be drawn on the product. Therefore, conducting an ALT requires a carefully designed test plan.

The literature on optimal ALT planning is vast. For a comprehensive review of ALT literature up to early 2000's, readers may refer to [21, 22, 24]. To briefly

R. Pan (✉)
School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA
e-mail: rong.pan@asu.edu

K. Seo
University of Missouri, Columbia, MO, USA
e-mail: seoka@missouri.edu

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_14

summarize it, Nelson and his collaborators [13, 25, 26] first introduced the optimal experimental design theory to planning ALTs in the late 1970's. Different failure time distributions were discussed and multiple failure time censoring schemes were considered. Islam and Ahmad [10] and Yum and Choi [40] developed the optimal ALT planning with periodic inspection and type-I censoring. Seo and Yum [36] extended it to intermittent inspection scheme. All of these test plans were for a single stress factor. Escobar and Meeker [7], Park and Yum [31] and Sitter and Torsney [37] presented the optimal test plans with two stress factors. In [27], the way of finding optimal test plans for different optimal criteria by deriving the expected Fisher information matrix was discussed in detail. This approach was further discussed in [2, 5, 6, 28]. Other extensions of ALT plans can be found in, e.g., [32, 33].

The optimal test plans aforementioned are developed based on failure time regression models, which are also referred as failure time acceleration models in reliability engineering, where a stress factor at some higher-than-normal-use level is supposed to directly shorten a product's lifetime. When a Weibull failure time distribution is in use, the failure time acceleration model becomes a Weibull regression model. The proportional hazard (PH) regression model for analyzing censored ALT data, however, assumes that the effect of a stress variable is on altering the failure rate or hazard function of a failure time distribution. According to the Cox's proportional hazard function, this effect is in an exponential function form, multiplicative to the baseline hazard function [4]. Aitkin and Clayton [1] showed that the PH regression model with censored survival data can be re-written in a generalized linear model (GLM) form and this model is applicable to multiple distributions, including exponential, Weibull, and extreme value distributions. In addition, [3] analyzed ALT results using a piecewise exponential distribution with the GLM approach. Finkelstein [8] and Finkelstein and Wolfe [9] considered the GLM approach to analyzing interval censored survival data. More recently, researchers started to use this approach for designing optimal ALT test plans. Monroe et al. [16, 17] provided the formulation and the optimal design process for type-I censored exponential and Weibull distributions. Yang and Pan [39] developed the GLM approach to ALT planning with interval censored data. They further utilized this approach to planning tests with more than one optimal criterion and to planning tests for discriminating different acceleration models, respectively [29, 30]. Seo and Pan [35] considered special constraints to ALT planning such as test chamber assignments and test unit clusters and modeled them as random effects in a generalized linear mixed model (GLMM). They also developed an optimal ALT design tool, an R package named *ALTopt*, that follows the GLM approach and is able to handle multiple stress factors and different censoring schemes [34].

In this chapter, we will give a succinct presentation of the GLM approach to ALT planning. Generally speaking, the design parameters to be considered in an ALT may include:

- number of test units,
- manufacturing information of test units,
- total test duration,

- censoring scheme,
- stress factors,
- stress levels,
- testing conditions as stress factor-level combinations,
- allocation of test units to each testing condition,
- number of test chambers and their capacities,
- sequence of ALT experiments.

This list could go even longer to include some managerial and implementational aspects of the test, such as the logistics in planning a test and the capability of test chamber; however, we would ignore these considerations at the moment because they are not directly related to the statistical efficiency of a test plan, which is the focal point of this chapter. Even with this list, we have to narrow down to a few key decision variables, while assuming other items are either pre-specified in advance or can be ignored with sufficient resource available. Here, we will focus on the selection of testing conditions and the allocation of test units to each testing condition. This is a typical experimental design problem; i.e., to find the appropriate experimental factor-level combinations and to decide the sample size for each factor-level combination.

The problem to be tackled in this chapter has one response variable, which is the failure time or lifetime of the product being tested, and one or more stress variables, which are applied to shorten the product's lifetime by certain chemical or mechanical failure mechanisms. The type-I censoring scheme is assumed in this chapter, which implies that the test ends at a predetermined time. Product's lifetime is continuously monitored. Exact failure times are collected from the ALT experiment, although they can be censored because some test units survive the entire testing period. Denote the response variable as Y and the stress variables as x_1, x_2, etc. A regression function can be established between Y and x's. But, unlike the standard regression model based on the normality assumption, Y, as product lifetime, is often modeled by a non-normal distribution, such as exponential, Weibull, gamma or lognormal distribution. Consequentially, the parameter estimation is done by maximum likelihood estimation (MLE), instead of least squares estimation (LSE).

2 Accelerated Failure Time Model

Although the PH regression model will be used for ALT planning in this chapter, for the completeness of ALT modeling, we will first introduce the accelerated failure time (AFT) model, which is still the most commonly used model in the reliability engineering field.

The AFT model assumes that the lifetime of a product is changed by a stress variable by an acceleration factor, AF, when it is subject to an accelerated life test.

That is,

$$Y_s = \frac{Y_u}{AF}, \tag{1}$$

where Y_s is the lifetime under the accelerated stress condition and Y_u is under the normal use condition. Therefore, the cumulative distribution function (CDF) of lifetime is given by

$$F_s(y) = Pr(Y_s \leq y) = Pr(Y_u \leq (AF \times y)) = F_u(AF \times y). \tag{2}$$

Accordingly, the probability density function (PDF) and hazard function of lifetime under the accelerated stress condition can be derived as, respectively,

$$f_s(y) = AF \times f_u(AF \times y) \tag{3}$$

and

$$h_s(y) = AF \times h_u(AF \times y). \tag{4}$$

Assume that a product's lifetime follows exponential distribution with its CDF as

$$F_u(y) = 1 - e^{-(\frac{y}{\mu_u})}, \tag{5}$$

where μ_u is the mean failure time, or mean-time-to-failure (MTTF), of the product under its normal use condition. Then, it is easy to see that, under an accelerated stress condition, this MTTF becomes

$$\mu_s = \frac{\mu_u}{AF}. \tag{6}$$

Equivalently, the failure rate λ of the product, which is the reciprocal of MTTF, is given by

$$\lambda_s = AF \times \lambda_u. \tag{7}$$

Note that acceleration factor is indeed a function of stress variables, thus it can be written as $AF(\boldsymbol{x})$, where $\boldsymbol{x}$ is the vector of stress variables. Equation (6) is used when exponential lifetime distribution is a proper lifetime model. When Weibull distribution or gamma distribution is in use, their scale parameters (i.e., the characteristic life of Weibull distribution or the first parameter of gamma distribution) will replace μ in Eq. (6). It can be shown that, based on the accelerated failure time model, only the scale parameter of lifetime distribution is affected by the acceleration factor (see [38], Ch. 8).

How can an elevated stress level shorten a product's lifetime? It is in fact related to the nature of failure process, or physics of failure, of the product that is affected by the particular stress factor. These physical models are known as Arrhenius model, Eying model, Pack model, etc. ([23]). For example, considering a thermal stress-induced failure process, the following Arrhenius law is typically employed to explain the effect of thermal stress on a product's lifetime; that is,

$$Y \propto \exp\left(\frac{E_a}{kT}\right), \tag{8}$$

where E_a is the material's activation energy, k the Boltzmann constant, and T the temperature in degrees Kelvin. Combining it with the exponential failure time distribution, we have

$$\mu_s = \mu_u \exp\left(\frac{E_a}{k}\left(\frac{1}{T_s} - \frac{1}{T_u}\right)\right). \tag{9}$$

Taking the normal-use condition as a reference and applying natural logarithm on both sides of Eq. (9), a log-linear regression function appears as

$$\log \mu(x) = \beta_0 + \beta_1 x, \tag{10}$$

where β_0 and β_1 are regression coefficients and they are related to the mean-time-to-failure under the normal-use condition and the material's activation energy, respectively, while x is defined as the natural stress variable. Here, for the thermal stress, the natural stress variable x is defined as $1/kT$.

The log-linear function for failure distribution parameter, as discussed above, is valid for other stress variables too, as long as some proper transformations of stress variables are employed. For example, by the Eyring's law an additional stress variable (pressure or voltage) can be introduced into Eq. (8) and the log-linear function of Eq. (10) becomes

$$\log \mu(x_1, x_2) = \beta_1 + \beta_1 x_1 + \beta_2 x_2, \tag{11}$$

where x_1 is the natural thermal stress variable as defined above and x_2 is the second natural stress variable, which is defined as $x_2 = \log s$ and s is pressure or voltage.

In general, the log-linear function is used to model the relationship between a failure time distribution parameter and natural stress variables, such as

$$\log \theta = \boldsymbol{x}^T \boldsymbol{\beta}, \tag{12}$$

where θ is the scale parameter of failure time distribution, $\boldsymbol{x}$ the vector of natural stress variables with the first element being 1, and $\boldsymbol{\beta}$ the vector of regression coefficients. In particular, when Weibull failure time distribution is in use, θ is the characteristic life parameter of Weibull distribution and the AFT model is given by

$$Y \sim Weibull(\alpha, \theta(\boldsymbol{x})), \ or, \ F(y) = 1 - e^{-(\frac{y}{\theta(\boldsymbol{x})})^{\alpha}} \tag{13}$$

and

$$\log \theta(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\beta}. \tag{14}$$

This AFT is called the Weibull regression model, which is one of the most popular failure time regression models. Notice that the exponential regression model is a special case of Eq. (13) when the shape parameter α equals to 1.

3 Proportional Hazard Model

In contrast to AFT, the PH model assumes the effect of a stress variable will be applied on the failure rate, or hazard function, of a product, instead of its failure time. With the PH assumption, the hazard function under a stress condition is given by

$$h_s(y) = h_0(y) e^{\boldsymbol{x}^T \boldsymbol{\beta}}, \tag{15}$$

where $h_0(y)$ is the baseline hazard function and it can be any function of time y. Specifically, if the baseline hazard function is a constant, then failure time distribution is an exponential distribution, or, if the function is a power function of time, then the failure time distribution becomes a Weibull distribution.

The PH model is in fact a semi-parametric model, as the baseline hazard function needs not to be defined for the estimation of effects of stress variables. The partial likelihood estimation method is typical inference method for this task. However, when the failure time distribution is fully defined as Weibull distribution or exponential distribution, it is easy to see that the PH model is the same as the Weibull regression model as specified in Eqs. (13) and (14), except that the sign of coefficient β could be negated. This property is not applicable to other failure distributions.

By the PH model, the hazard function of Weibull distribution is given by

$$h(y, \boldsymbol{x}) = \lambda(\boldsymbol{x}) y^{\alpha - 1} = \lambda_0 e^{\boldsymbol{x}^T \boldsymbol{\beta}} y^{\alpha - 1}, \tag{16}$$

where λ_0 is the baseline failure rate and α is the shape parameter, same as in Eq. (13). The intrinsic failure rate $\lambda(\boldsymbol{x})$ can be converted to the characteristic life parameter in Eq. (13) by $\lambda = \alpha/\theta^{\alpha}$ or $\log \lambda = \log \alpha - \alpha \log \theta$. Therefore, the log-linear equation, Eq. (14), can be re-written as

$$\log \lambda = \boldsymbol{x}^T \boldsymbol{\beta}'. \tag{17}$$

Here, the regression coefficient vector $\boldsymbol{\beta}'$ is a reparameterization of $\boldsymbol{\beta}$ in Eq. (14). To simplify the notation, later in the chapter when a PH model is in use, the regression coefficient vector will be simply denoted as $\boldsymbol{\beta}$.

4 GLM Formulation

Assume that a type-I censored ALT experiment is conducted, so there are some test units that have survived at the end of the reliability test; i.e., their failure times are censored at the test terminal time. The likelihood function of ALT data can be constructed. The contributions of these censored units to the total likelihood function is their reliability functions, while for those observed failure times, the contributions are their PDFs. The total likelihood function is given by

$$L = \prod_{i=1}^{n} (R_i(y_i))^{1-c_i} (f(y_i))^{c_i} = \prod_{i=1}^{n} (h(y_i))^{c_i} R(t_i), \tag{18}$$

where subscript i represents the index of test unit and c_i is an indicator variable with $c_i = 1$ if the observed time y_i is a failure time, or 0 if it is a survival time.

Based on the PH model Eq. (15), the log-likelihood function can be derived as

$$\log L = \sum_{i=1}^{n} [c_i (\log h_0(y_i) + \boldsymbol{x}_i^T \boldsymbol{\beta}) + e^{\boldsymbol{x}_i^T \boldsymbol{\beta}} \log R_0(y_i)]. \tag{19}$$

Let $\mu_i = e^{\boldsymbol{x}_i^T \boldsymbol{\beta}} (-\log R_0(y_i))$, then

$$\log L = \sum_{i=1}^{n} [c_i \log \mu_i - \mu_i + c_i (\log h_0(y_i) - \log(-\log R_0(y_i)))]. \tag{20}$$

Note that the first two terms of the right-hand-side of Eq. (20) consist of a Poisson distribution likelihood function, while the last term is a constant, or invariant to any stress variable x. Therefore, by considering the first two terms only, the indicator variable c_i can be treated as following a Poisson distribution with mean μ_i. By maximizing this Poisson likelihood function, the estimation of stress variable effects, β's, can be obtained.

Recasting the whole formula into a GLM fashion ([20]), it becomes

- The response variable, c, are independently Poisson distributed as $Poisson(\mu_i)$;
- The linear predictor is given by $\eta_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$;
- The link function is a canonical link function given by $\ln \mu_i = \eta_i + a_i$, where $a_i = \log(-\log R_0(y_i))$ is an offset term.

To substantiate the formulation above, let failure times follow a Weibull distribution with the hazard function of Eq. (16), then the offset term becomes

$$a_i = \log \lambda_0 + \alpha \log y_i. \tag{21}$$

Note that this offset term will not affect the estimation of stress variable effects if the partial likelihood function maximization method is used. However, it does affect the full likelihood function and is needed for estimating the Weibull shape parameter. In addition, to design ALT experiments, if the concern is about how to estimate the stress effects efficiently or how to predict the characteristic life or MTTF at the normal-use condition more precisely, this offset term becomes a nuisance term and can be ignored. In the next section, the design problems of ALT will be discussed based on this GLM formulation.

5 ALT Design Problems

Planning an ALT is a special type of design of experiments (DOE) with the experimental factors being stress variables and the response being failure time. The consideration of failure time censoring must be included into ALT planning. The most common censoring scheme is the type-I censoring, where the test will be terminated at a pre-specified time. For the test units that do not fail, their failure times are right censored at the test terminal time. To plan an ALT test, one needs to keep in mind that this type of test are very expensive and both test units and testing equipment could be limited; therefore, the test plan should be designed to extract as much useful information from the test as possible. In general, one should reduce the chance of generating no failure observations from the test and one should place one test stress condition as close to the normal-use condition as possible. Mathematically, an objective function is proposed and the ALT planning problem is formulated as to find the test plan that can optimize this objective function, thus it is referred as the optimal test plan or optimal design.

The standard factorial design and classical designs like central composite design and Box-Behnken design are widely used for product quality and reliability improvement [18, 19]. However, in some circumstances it is hard or impossible to apply them in ALTs. First, the design region of an ALT experiment may not be a regular space. This is particularly true for an ALT that is constrained by the acceleration factor requirement and the stress chamber's capability limitation. Monroe and Pan [15] gave a good example that illustrated these constraints. Second, an ALT acceleration model can be formulated with engineering domain knowledge, thus it may not confirm with the regular linear or polynomial models used in regular DOE. Third, due to the limitation of time and budget, ALT experiments can only have small number of runs, to which the classical designs may not be applicable. In order to tackle the design problems with considerations mentioned above, we follow the idea of optimal design, initially developed by [11, 12].

An optimal design is computer-generated with respect to a mathematically optimal criterion. The most commonly used optimal design criterion is the D-optimal criterion [19], which is defined by

$$\xi^* := \arg\max_{\xi} det(\boldsymbol{X}(\xi)^T \boldsymbol{W} \boldsymbol{X}(\xi)), \tag{22}$$

where ξ denotes a test plan and ξ^* is the D-optimal test plan, $\boldsymbol{X}(\xi)$ is the expanded design matrix of a test plan, $\boldsymbol{W}$ is the weight matrix. Both $\boldsymbol{X}$ and $\boldsymbol{W}$ can be derived from the GLM formulation above. In particular, suppose there are n independent test units, then matrix $\boldsymbol{X}$ is a $n \times p$ matrix that includes all regression terms and all test units, while matrix $\boldsymbol{W}$ is a $n \times n$ diagonal matrix with weight elements to be μ_i. For a Weibull distribution with known shape parameter, this weight element can be found to be

$$w_{ii} = \mu_i = 1 - e^{-\lambda_0 \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) t^\alpha}, \tag{23}$$

where t is the test termination time.

Matrix $\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}$ is called the information matrix and the variance-covariance matrix of GLM parameter estimation is proportional to the inverse of this matrix. Maximizing the determinant of this matrix is equivalent to minimizing the general variance of model parameter estimators. Therefore, the D-optimal test plan found by Eq. (22) is a statistically efficient plan that is most precise at estimating GLM parameters, or the effects of stress variables, on the product's failure time.

To maximize the determinant, an optimization algorithm needs to be employed. The R package, *ALTopt*, utilizes the function "stats:optim" with the "L-BFGS-B" method to perform optimization. This function allows box constraints on design variables. For example, if there are 3 stress variables, a cuboidal design region with each variable's range from 0 to 1 can be assigned as the design region, then *ALTopt* can find the design points (testing conditions) in this region, as well as the number of test units for each design point, such that the test plan becomes D-optimal.

The *ALTopt* package can also derive U-optimal and I-optimal test plans. These test plans are optimal for minimizing the prediction variance of a reliability parameter at the product's normal-use condition and over a range of normal-use conditions, respectively. For more information of *ALTopt*, readers may refer to [34].

6 Examples

Let us consider the adhesive bond ALT example from Meeker and Escobar [14] to demonstrate the generation process of a D-optimal test plan with single stress factor by the GLM approach. The following information is given for the test.

- Stress factor: Temperature
- Failure time distribution: Weibull

- Use condition: 50 °C
- Highest stress limit: 120 °C
- Number of test units: 300 test units
- Testing time (Censoring time): 180 days
- Assumed shape parameter value: $\alpha = 1.667$
- Assumed regression coefficients: $\beta_0 = 27.89$, $\beta_1 = -1.21$ (These values were obtained by re-parameterization from those in the original example via $\beta_0 = -1.667 \times -16.733 = 27.89$ and $\beta_1 = -1.667 \times 0.7265 = -1.21$.)

Besides the above information, the lowest stress limit is specified as 80 °C in order to guarantee enough failure time data from the test. Applying the Arrhenius acceleration model, the natural stress variable is defined as $S = 1/kT = 11{,}605/(273.15 + T)$. Accordingly, the use condition and the lowest and highest stress limits are transformed to be as $S_U = 11{,}605/(273.15 + 50) = 35.91$, $S_L = 11{,}605/(273.15 + 80) = 32.86$, and $S_H = 11{,}605/(273.15 + 120) = 29.52$, respectively. In DOE, it is a common practice to use the coded variable instead of the original one. As such, the natural stress variable is further transformed to the coded variable via $x = (S - S_H)/(S_L - S_H)$ so that the design region is constrained to [0, 1]. Figure 1 illustrates the variable transformation schemes.

The following linear predictor model is obtained in terms of the coded variable,

$$\ln \mu = \beta_0 + \beta_1 S = \beta_0 + \beta_1\{(S_L - S_H)x + S_H\} = -7.855 - 4.049x. \quad (24)$$

Now, run the R package *ALTopt* to generate the D-optimal design. The software package basically runs the optimization routine. First, its algorithm assigns each of 300 design points randomly into the design region, and then the algorithm has these design points moved in the direction of an improved objective function value at each iteration. The iterative process stops when the objective function value has converged. Since the result depends on the initial random assignment, it is

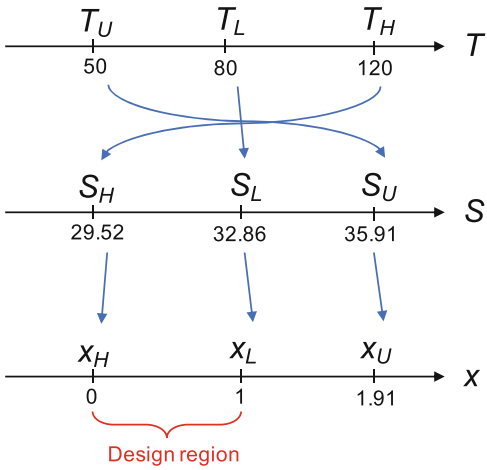


Fig. 1 Variable transformation

Table 1 The D-optimal design for the adhesive bond ALT

x	Number of test units
0.000	173
0.557	127

recommended to run the routine several times and pick the best one. Table 1 shows the D-optimal design generated by the following lines of code, which has the highest objective function value among 30 different runs of the optimization routine.

```
R> library("ALTopt")
R> set.seed(200)
R> altopt.rc("D", N = 300, tc = 180, nf = 1,
+ alpha = 1.667, formula = ~x1,  coef =
+ c(-7.855, -4.049), nOpt = 30)
```

It shows the two distinguished design points and the number of test units allocated to each point. We can transform the test locations back to the original variable, which results in 173 test units allocated to 120 °C and 127 test units allocated to 96.7 °C. We can show that 154 or 155 test units are expected to be failed at 120 °C, and so are 26 or 27 test units at 96.7 °C before the end of the testing period if the assumed parameter values were correct. There numbers are reasonably large enough for model parameter estimation.

Some readers may have a doubt about the assumed parameter values. Where do those values come from and why we need those? In fact the weight matrix $\boldsymbol{W}$ depends on the regression coefficients and the shape parameter, as one can see from Eq. (23). Therefore, we need some a priori knowledge about these parameters, which is often obtained from engineering experience, domain experts' opinions, previous similar tests and literature. If the assumed model parameter values are in fact incorrect, we might end up generating a completely different and inefficient test plan. For example, suppose the linear predictor model in Eq. (24) is true but we have assumed the following model, with β_1 being -2.428 instead of -4.049, incorrectly,

$$\ln \mu = -7.855 - 2.428x. \tag{25}$$

The D-optimal design based on the incorrect model of Eq. (25) is shown in Table 2. First, a substantial change of the lowest testing stress condition from 0.557 (96.7 °C) to 0.928 (82.6 °C) has been suggested. Second, there is a significant difference for the test unit allocations. In specific, 46 more test units were allocated at the higher stress level in Table 1, while 22 more test units were allocated at the lower stress level in Table 2. As a result only 8 test units are expected to be failed at the lower

Table 2 The D-optimal design based on the incorrect linear predictor model

x	Number of test units
0.000	139
0.928	161

stress level (i.e., 82.6 °C), which might not be enough number of failures for an accurate parameter estimation.

Expanding the aforementioned example, we now consider an ALT with two stress factors. We add another (coded) stress factor x_2 while the testing time is shortened to 120 days. Suppose the following linear predictor model.

$$\ln \mu = -7.855 - 4.049x_1 - 1.217x_2. \tag{26}$$

The design region becomes the square of $[0, 1] \times [0, 1]$, where the highest and lowest stress limits correspond to $(0, 0)$ and $(1, 1)$, respectively. Table 3 shows the output of ALTopt software package. Similar to the previous example, the optimization routine of the software package has been run 30 times. It determines the three distinguished design points which are necessary to estimate the three unknown regression coefficients. Figure 2 illustrates the generated design in the 2-dimensional design space, where the areas of the circles are proportional to the number of test units allocations. The following codes generated the results.

```
R> set.seed(300)
R> d_opt_design <- altopt.rc("D", N = 300, tc = 120,
+ nf = 2, alpha = 1.667, formula = ~x1 + x2,
+ coef = c(-7.855, -4.049, -1.217), nOpt = 30)
R> design.plot(d_opt_design$opt.design.rounded,
+ xAxis = x1, yAxis = x2)
```

Table 3 The D-optimal design with two stress factors

x_1	x_2	Number of test units
0.000	0.000	124
0.528	0.000	60
0.000	1.000	116

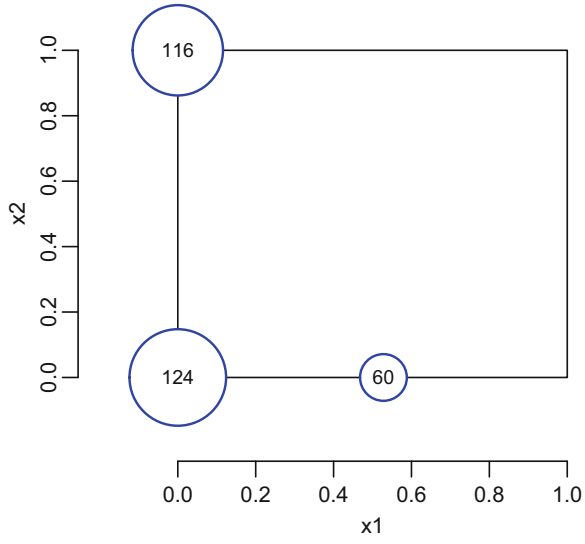


Fig. 2 Design plot of the D-optimal design with two stress factors

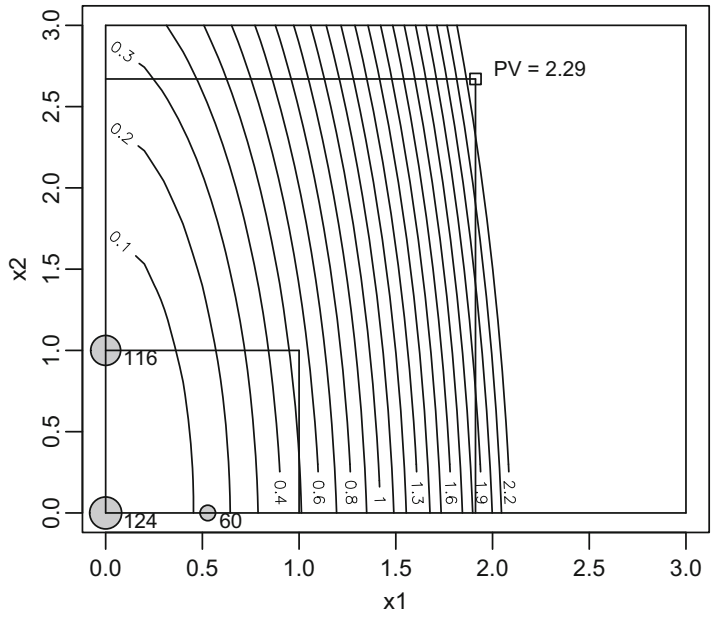


Fig. 3 Prediction variance contour plot of the D-optimal design

One of possible ways to evaluate a given ALT design or compare two or more ALT plans is to calculate the prediction variance at the use condition. For the aforementioned example, let us assume the use condition is given as, for a coded unit, $x_1 = 1.912$ and $x_2 = 2.670$. We now can evaluate the D-optimal design of Fig. 2 in terms of the prediction variance at the use condition. The following code creates the prediction variance contour plot, which is shown by Fig. 3.

```
R> pv.contour.rc(d_opt_design$opt.design.rounded,
+ xAxis = x1, yAxis = x2, tc = 120, nf = 2,
+ alpha = 1.667, formula = ~ x1 + x2, coef =
+ c(-7.855, -4.049, -1.217), useCond = c(1.912, 2.670))
```

If one is interested in the prediction variance, the D-optimal design may not be the best test plan. Instead, we may want to create a design with the minimum prediction variance at the use condition, which we call the U-optimal design. The following lines of code generate the U-optimal design and its prediction variance contour plot, which is shown by Fig. 4. Note that the prediction variance of the U-optimal design is 1.19 at the use condition, which is smaller than one, 2.29, of the D-optimal design.

```
R> u_opt_design <- altopt.rc("U", N = 300, tc = 120,
+ nf = 2, alpha = 1.667, formula = ~x1 + x2,
+ coef = c(-7.855, -4.049, -1.217), useCond =
+ c(1.912, 2.670), nOpt = 30)
```

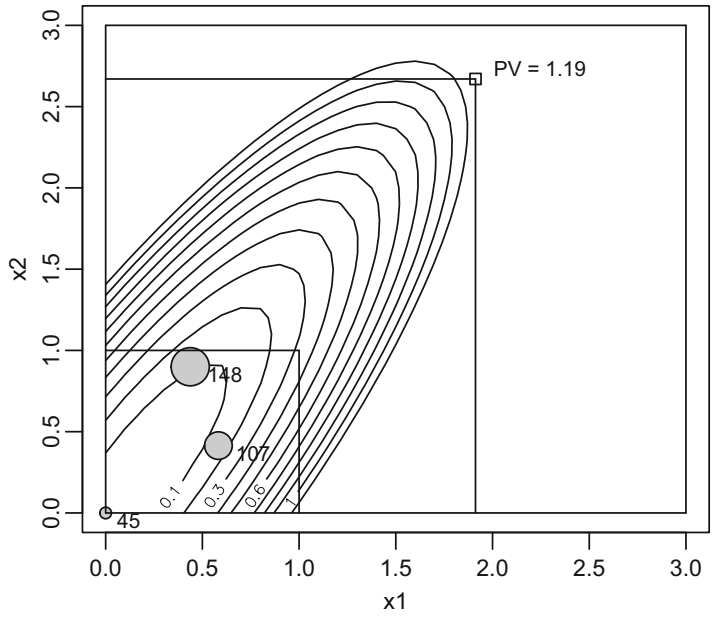


Fig. 4 Prediction variance contour plot of the U-optimal design

```
R> pv.contour.rc(u_opt_design$opt.design.kmeans,
+ xAxis = x1, yAxis = x2, tc = 120, nf = 2,
+ alpha = 1.667, formula = ~ x1 + x2, coef =
+ c(-7.855, -4.049, -1.217), useCond = c(1.912, 2.670))
```

7 Conclusions

In this chapter we introduce a GLM approach to ALT planning. This approach is different from the failure time regression approach in the sense that it is based on the PH model or accelerated failure rate model, instead of the AFT model. However, for Weibull or exponential failure time distribution, both models arrive at the same formulation. The GLM approach takes care of failure time censoring, which is unique and unavoidable phenomenon in reliability tests, and provides a uniform way for constructing the information matrix of a test. The D-optimal ALT design is discussed in this chapter. For readers who are interested in other optimal criteria and/or other constraints on ALT planning, we suggest the following publications: [16, 30, 35, 39].

References

1. Aitkin, M., & Clayton, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics, 29*(2), 156–163.
2. Balakrishnan, N. (2009). A synthesis of exact inferential results for exponential step-stress models and associated optimal accelerated life-tests. *Metrika*, 69–351.
3. Barbosa, E., Colosimo, E., & Louzada-Neto, F. (1996). Accelerated life tests analyzed by a piecewise exponential distribution via generalized linear models. *IEEE Transactions on Reliability, 45*(4), 619–623.
4. Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological) 34*(2), 187–220.
5. Dahmen, K., Burkschat, M., & Cramer, E. (2012). A-and D-optimal progressive Type-II censoring designs based on Fisher information. *Journal of Statistical Computation and Simulation, 82*(6), 879–905.
6. Ding, C., & Tse, S. K. (2013). Design of accelerated life test plans under progressive Type II interval censoring with random removals. *Journal of Statistical Computation and Simulation, 83*(7), 1330–1343.
7. Escobar, L. A., & Meeker, W. Q. (1995). Planning accelerated life tests with two or more experimental factors. *Technometrics, 37*(4), 411–427.
8. Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics, 42*(4), 845–854.
9. Finkelstein, D. M., & Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics, 41*(4), 933–945.
10. Islam, A., & Ahmad, N. (1994). Optimal design of accelerated life tests for the Weibull distribution under periodic inspection and type I censoring. *Microelectronics Reliability, 34*(9), 1459–1468.
11. Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society B, 21*, 272–304.
12. Kiefer, J. (1961). Optimum designs in regression problems. *Annals of Mathematical Statistics, 32*, 298–325.
13. Meeker, W. Q., & Nelson, W. (1975). Optimum accelerated life-tests for the Weibull and extreme value distributions. *IEEE Transactions on Reliability, R-24*(5), 321–332.
14. Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. New York: John Wiley and Sons Inc.
15. Monroe, E. M., & Pan, R. (2008). Experimental design considerations for accelerated life tests with nonlinear constraints and censoring. *Journal of Quality Technology, 40*(4), 355–367.
16. Monroe, E. M., Pan, R., Anderson-Cook, C., Montgomery, D. C., & Borror, C. M. (2010). Sensitivity analysis of optimal designs for accelerated life testing. *Journal of Quality Technology, 42*(2), 121–135.
17. Monroe, E. M., Pan, R., Anderson-Cook, C., Montgomery, D. C., & Borror, C. M. (2011). A generalized linear model approach to designing accelerated life test experiments. *Quality and Reliability Engineering International, 27*(4), 595–607.
18. Montgomery, D. C. (2012). *Design and analysis of experiments* (8th ed.). Hoboken: Wiley.
19. Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology: Process and product optimization using designed experiments* (3rd ed.). Hoboken: Wiley.
20. Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General), 135*(3), 370–384.
21. Nelson, W. (2005a). A bibliography of accelerated test plans. *IEEE Transactions on Reliability, 54*(2), 194–197.
22. Nelson, W. (2005b). A bibliography of accelerated test plans part II – references. *IEEE Transactions on Reliability, 54*(3), 370–373.

23. Nelson, W. (2009). *Accelerated testing: statistical models, test plans, and data analysis* (Vol. 344). Hoboken: John Wiley & Sons.
24. Nelson, W. (2015). An updated bibliography of accelerated test plans. In *2015 Annual Reliability and Maintainability Symposium (RAMS)*, 1–6.
25. Nelson, W., & Kielpinski, T. J. (1976). Theory for optimum censored accelerated life tests for normal and lognormal life distributions. *Technometrics, 18*(1), 105–114.
26. Nelson, W., & Meeker, W. Q. (1978). Theory for optimum accelerated censored life tests for Weibull and extreme value distributions. *Technometrics, 20*(2), 171–177.
27. Ng, H. K. T., Chan, P. S., & Balakrishnan, N. (2004). Optimal progressive censoring plans for the Weibull distribution. *Technometrics, 46*(4), 470–481.
28. Ng, H. K. T., Kundu, D., & Chan P. S. (2009). Statistical analysis of exponential lifetimes under an adaptive Type? II progressive censoring scheme. *Naval Research Logistics, 56*(8), 687–698.
29. Pan, R. & Yang, T. (2014). Design and evaluation of accelerated life testing plans with dual objectives. *Journal of Quality Technology 46*(2), 114.
30. Pan, R., Yang, T., & Seo, K. (2015). Planning constant-stress accelerated life tests for acceleration model selection. *IEEE Transactions on Reliability 64*(4), 1356–1366.
31. Park, J.-W., & Yum, B.-J. (1996). Optimal design of accelerated life tests with two stresses. *Naval Research Logistics (NRL), 43*(6), 863–884.
32. Pascual, F. (2007). Accelerated life test planning with independent Weibull competing risks with known shape parameter. *IEEE Transactions on Reliability, 56*(1), 85–93.
33. Pascual, F. (2008). Accelerated life test planning with independent Weibull competing risks. *IEEE Transactions on Reliability, 57*(3), 435–444.
34. Seo, K., & Pan, R. (2015). ALTopt: An R package for optimal experimental design of accelerated life testing. *R Journal, 7*, 2.
35. Seo, K., & Pan, R. (2018). Planning accelerated life tests with random effects of test chambers. *Applied Stochastic Models in Business and Industry, 34*, 224–243.
36. Seo, S.-K., & Yum, B.-J. (1991). Accelerated life test plans under intermittent inspection and type-I censoring: The case of Weibull failure distribution. *Naval Research Logistics (NRL), 38*(1), 1–22.
37. Sitter, R. R., & Torsney, B. (1995). Optimal designs for binary response experiments with two design variables. *Statistica Sinica, 5*, 495–419.
38. Tobias, P. A., & Trindade, D. (2011). *Applied reliability*. Boca Raton: CRC Press.
39. Yang, T. & Pan, R. (2013). A novel approach to optimal accelerated life test planning with interval censoring. *IEEE Transactions on Reliability 62*(2), 527–536.
40. Yum, B.-J., & Choi, S.-C. (1989). Optimal design of accelerated life tests under periodic inspection. *Naval Research Logistics (NRL), 36*(6), 779–795.

Robust Design in the Case of Data Contamination and Model Departure

Linhan Ouyang, Chanseok Park, Jai-Hyun Byun, and Mark Leeds

Abstract In robust design, it is usually assumed that the experimental data are normally distributed and uncontaminated. However, in many practical applications, these assumptions can be easily violated. It is well known that normal model departure or data contamination can result in biased estimation of the optimal operating conditions of the control factors in the robust design framework. In this chapter, we investigate this possibility by examining these estimation effects on the optimal operating condition estimates in robust design. Proposed estimation methodologies for remedying the difficulties associated with data contamination and model departure are provided. Through the use of simulation, we show that the proposed methods are quite efficient when the standard assumptions hold and outperform the existing methods when the standard assumptions are violated.

Keywords Robust design · Outlier-resistance · Contamination · Model departure · Relative efficiency

L. Ouyang
Department of Management Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China
e-mail: ouyang@nuaa.edu.cn

C. Park (✉)
Applied Statistics Laboratory, Department of Industrial Engineering, Pusan National University, Busan, South Korea

J.-H. Byun
Department of Industrial and Systems Engineering, Gyeongsang National University, Jinju, Gyeongnam, South Korea
e-mail: jbyun@gnu.ac.kr

M. Leeds
Statematics Consulting, New York, NY, USA

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_15

1 Introduction

Robust design is widely used for quality improvement purposes. Many engineering problems use robust design methods in order to improve product quality, process performance, etc. in various fields including the automotive, manufacturing, plastic, and information technology industries. Interested readers are referred to the articles [1, 2] for more examples. One of the important goals in robust design is to obtain the optimal operating conditions using a specific and well designed objective function. In order to achieve such a goal, one can incorporate Taguchi's methods into the robust design [3–5] framework. However, this approach has been shown to be ineffective in certain instances due to various statistical issues. Although there is no serious disagreement among engineers and practitioners about Taguchi's basic philosophy, drawbacks embodied in his approach are addressed by several authors [6–15]. There has been a great deal of research involved in improving the approach and eliminating the various drawbacks. Response surface methods for robust design were eventually developed as a more effective alternative to Taguchi's methods. The response surface approach facilitates understanding of the system by modelling the response functions for process mean and variance separately, and treating them as dual responses. For more details on the response surface approach, one is referred to articles [9, 16, 17]. It is noteworthy that the robust design framework based on the response surface approach has been studied thoroughly by several authors [18–32] among others. In this chapter, unless otherwise specified, 'robust design' refers to the robust design method based on the response surface approach.

In order to obtain the response functions for process mean and variance in the robust design framework, the mean and variance need to be estimated at each design point. The response functions for the process mean and variance are then obtained using the method of classical linear regression with the mean and variance estimates at each design point. The two fundamental assumptions in robust design are that the experimental data are normally distributed and that all of the observations are uncontaminated. However, these fundamental assumptions may not hold in practice because the observations can often be collected under volatile operating conditions, etc. The normality and no contamination assumptions in robust design are related to the normal model departure and outlier-resistance concepts in the statistics literature. If either one of the two assumptions is violated in a serious manner, the estimated optimal operating conditions of the control factors may be located far from the true optimal operating conditions. Thus, a statistical approach which is less sensitive to model departure and outliers is clearly warranted.

For the case in which some of the observations are contaminated, Park and Cho [33] proposed a robust design methodology using outlier-resistant estimators of location and scale. The authors investigated the effectiveness of the median, the median absolute deviation (MAD) and inter-quartile range (IQR). These estimators were incorporated into the robust design and their behavior was shown to be quite effective in the sense that, when the data were contaminated, the bias and mean

squared error (MSE) of the estimated optimal operating conditions obtained was significantly lower than the resulting bias and MSE based on the sample mean and sample variance.

Although contamination is a frequent problem, the normality assumption can also be violated in practice. If the normality assumption does not hold in the sense that the histogram of the observations exhibits skewness or kurtosis or both, it is quite possible that the estimated optimal operating conditions of the control factors will be severely biased. Thus, in the same investigation, [33] also showed that the same estimators were also less sensitive to model departure. Recently, the research by Park and Leeds [31] and Park et al. [32] extended the work of [33] by incorporating the mean, median and Hodges–Lehmann (HL) estimators of the mean response and the standard deviation, MAD, IQR and Shamos estimators of the response standard deviation into the robust design framework. Using these estimators, they proposed seven methods which use various pairs of the aforementioned location and scale estimators. The respective methods were labelled Method A (the sample mean and the sample variance), Method B (sample median and the squared MAD), Method C (sample median and the squared IQR), Method D (HL and the squared Shamos estimator), Method E (sample median and the squared Shamos estimator), Method F (HL and the squared MAD), and Method G (HL and the squared IQR). The results in [32] illustrated that, using bias and MSE as the performance criteria, Method D clearly outperformed the other estimation methods in the case of data contamination or model departure. Specifically, it was shown that, among all of the methods considered, the estimated optimal operating conditions obtained using Method D resulted in the smallest bias and MSE.

One important statistical issue regarding outlier resistance is the concept of breakdown point [34]. The breakdown point discussion will be taken up in more detail but, simply speaking, the breakdown point of an estimator represents the percentage of contaminated observations that can exist before the estimation result becomes non-sensical and incorrect. It can be shown that Methods A, B, C, D, E, F, G, which are proposed by Park et al. [32], have 0%, 50%, 25%, 29%, 29%, 29%, 25% as breakdown points respectively. Therefore, even though Method D resulted in the best performance compared to all the other estimation methods based on the bias and MSE criterion, Method-D also has a lower breakdown point than Method B. This motivates the need for the development of another estimation method which has a higher breakdown point than that of Method D. This method can then be tested to see if its performance is at least similar to that of Method D. Therefore, in the estimation method comparison that will follow, we consider a method that uses two estimators with a 50% breakdown point, namely Huber's location M-estimator [35, 36] for the mean response and Rousseeuw-Croux (RC) estimator [37] for the response standard deviation.

Note that Method A is known to be optimal under the ideal assumption that there is no data contamination or normal model departure. Similarly, Method A is also expected to be inferior to the other methods when the assumption is violated. The goal of the comparative study that follows is to measure the performance of the proposed methods relative to the performance of the optimal method under

ideal assumptions. Thus, we include Method A by Park et al. [32] in the current investigation. Also, for clarity, in the comparison that follows we re-label Methods A, C and D by Park et al. [32] to Methods 1, 2 and 3 and label the new method that uses the M-estimator and the RC estimator as Method 4. Thus, we consider the following four methods in this chapter.

- Method 1: sample mean and sample variance (baseline. 0% breakdown).
- Method 2: sample median and MAD (50% breakdown, low efficiency).
- Method 3: HL and Shamos (29% breakdown, low efficiency).
- Method 4: Huber location M-estimator and RC estimator (50% breakdown with high efficiency).

As noted, Method 1 is optimal under ideal conditions and is included in the comparison for that reason. However, it is also well known that any estimation method that includes the sample mean or the sample variance will not perform well in the case of data contamination [31] and normal model departure [32]. Thus, except from the baseline case (Method 1) where the sample mean and sample variance estimators are used, it was decided that none of the estimation methods used in the investigation would involve the sample mean or the sample variance.

2 Robust Design Based on Dual Response Surfaces

In this section, we briefly review the standard robust design framework. Note that the assumptions, notation and framework used in this study are the same as those used in [28, 31, 32]. For convenience, the assumptions are summarized below.

(a) The response Y depends on the levels of the k control factors given by $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. Y is therefore a function of the control factors $\mathbf{x}$ and this is expressed as $Y = F(x_1, x_2, \ldots, x_k)$. The function $F(\cdot)$ is generally unknown and therefore needs to be estimated.
(b) Each control factor x_i for $i = 1, 2, \ldots, k$ has levels which are continuous and quantitative.
(c) The experimenter has the ability to set the levels of each control factor x_i where $i = 1, 2, \ldots, k$.
(d) The robust design framework uses the dual response surface approach to estimate the process mean response and variance response. The process mean response and variance response functions can be written as

$$M(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j}^{k} \beta_{ij} x_i x_j + \epsilon_M \tag{1}$$

and

$$V(\mathbf{x}) = \eta_0 + \sum_{i=1}^{k} \eta_i x_i + \sum_{i=1}^{k} \eta_{ii} x_i^2 + \sum_{i<j}^{k} \eta_{ij} x_i x_j + \epsilon_V, \tag{2}$$

respectively. The usual assumption is that the error terms ϵ_M and ϵ_V in (1) and (2) are normally distributed [9].

In order to estimate the dual response surfaces, we obtain the process mean response $M(\mathbf{x})$ and the process variance response $V(\mathbf{x})$ using the classical regression model where the control factors, $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ are the predictors. Clearly, before we can estimate the regression models for $M(\mathbf{x})$ and $V(\mathbf{x})$, we must first estimate the mean and variance responses at each design point $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. Now, let us assume that there are m replicated response observations at the ith design point and that Y_{ij} represents the jth response observation at the ith design point where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. The usual estimates of the mean and variance at each design point are the sample mean and sample variance. These estimates are calculated using the well known relations

$$\bar{Y}_i = \frac{1}{m} \sum_{j=1}^{m} Y_{ij} \tag{3}$$

and

$$S_i^2 = \frac{1}{m-1} \sum_{j=1}^{m} (Y_{ij} - \bar{Y}_i)^2. \tag{4}$$

at the ith design point, respectively.

Suppose that we denote $\hat{M}(\mathbf{x})$ and $\hat{V}(\mathbf{x})$ as the estimated dual response functions of $M(\mathbf{x})$ and $V(\mathbf{x})$ respectively. If we assume a second-order polynomial for the response functions, we can then construct the dual response surfaces as follows:

$$\hat{M}(\mathbf{x}) = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_{ii} x_i^2 + \sum_{i<j}^{k} \hat{\beta}_{ij} x_i x_j \tag{5}$$

and

$$\hat{V}(\mathbf{x}) = \hat{\eta}_0 + \sum_{i=1}^{k} \hat{\eta}_i x_i + \sum_{i=1}^{k} \hat{\eta}_{ii} x_i^2 + \sum_{i<j}^{k} \hat{\eta}_{ij} x_i x_j. \tag{6}$$

In order to estimate $M(\mathbf{x})$ and $V(\mathbf{x})$ in (5) and (6), we need to obtain $\hat{M}(\mathbf{x})$ and $\hat{V}(\mathbf{x})$. The standard approach for obtaining them is to use the sample mean in (3) and sample variance in (4) at each design point $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, respectively.

Although it is straightforward to estimate (5) and (6), the ultimate goal in robust design is to estimate the optimal operating conditions of the control factors, $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. These estimated optimal operating conditions can be obtained by optimizing the squared-loss model [17] below. That is, we want to

$$\text{minimize } \left\{\hat{M}(\mathbf{x}) - T_0\right\}^2 + \hat{V}(\mathbf{x}) \tag{7}$$

subject to any constraints on the control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. In (7), T_0 is the target value for the quality characteristic of interest which is usually defined by the analyst. Any imposed constraints are used to specify the feasible joint region of the control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. Note that in the specific case of factorial designs where the levels of the k control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ are used, the feasibility constraints become $x_j \in [L_j, U_j]$ for $j = 1, 2, \ldots, k$. The control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ that result from the optimization problem under the constraint are the optimal control factor estimates which are often referred to as the optimal operating conditions. Note that it can be easily shown that the objective function in (7) is equivalent to the MSE. Therefore, the optimal control factor estimates obtained also minimize the MSE. It should also be pointed out that there is also an alternative formulation of the dual-response optimization model [9] which can be used in the optimization step. This formulation is

$$\text{minimize } \hat{V}(\mathbf{x}) \tag{8}$$

subject to $\hat{M}(\mathbf{x}) = T_0$ and any constraints on the control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. However, a drawback of the model formulation in (8) is that it imposes a zero-bias condition while the squared-loss model in (7) does not. Therefore, since (7) is more flexible in that it allows for bias, it often results in less variability and was the chosen formulation used in this chapter. For detailed information regarding the squared-loss model, the readers are referred to [17].

Finally, we should point out one technical detail associated with the estimation of the response function in (6). Rather than using the sample variance on the left hand side of the regression model in (6), we instead used the logarithm-transformed values of the sample variances (i.e., $\log S_i^2$). By using the log transform of the variance during estimation and then subsequently transforming the variance estimates back by using exponentiation, we can be certain that the estimated variances will always be positive. Note that if such a transformation is not used, then it is possible for any of the estimated variance responses to be negative which is clearly nonsensical. For more details regarding the logarithmic transformation, interested readers are referred to [38] and Example 11.7 of [13]. Finally, after estimating the logarithm of the process variance, $V_{\log}(\mathbf{x})$, we obtain the optimal operating conditions by minimizing

$$\{\hat{M}(\mathbf{x}) - T_0\}^2 + \exp\left(\hat{V}_{\log}(\mathbf{x})\right) \tag{9}$$

subject to any constraints on the control factors $\mathbf{x} = (x_1, x_2, \ldots, x_k)$.

3 The Proposed Methods and Their Properties

3.1 Outlier-Resistant Estimation

In what follows, we briefly review the outlier-resistant estimation procedures used in [31, 32]. As described in Sect. 2 of this chapter, we assume that there are m replicates at each design point. Let Y_{ij} denote the jth response at the ith design point where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. In the standard robust design approach, the sample mean and sample variance are calculated at the ith design point using (3) and (4), respectively, and these are subsequently used to construct the response surfaces, namely, $M(\mathbf{x})$ and $V(\mathbf{x})$.

Now it is well known that the sample mean and the sample variance estimators are extremely sensitive to outliers. Therefore, using them as estimators of $M(\mathbf{x})$ and $V(\mathbf{x})$ can be problematic when there is the possibility that outliers do exist. Note that the terms *contamination* and *outlier* are often used interchangeably in statistical robustness studies so their meanings in this study need to be explained in detail. We define a contaminated observation as a sample observation whose value was either mis-recorded or whose value was measured under conditions that were not the usual or normal conditions. Similarly, we define an observation from a sample as an outlier when its value is unusually large or unusually small relative to the majority of observations in the sample. One possible reason for the existence of an outlier is that the particular observation is contaminated. Yet, another possibility is that the underlying distribution of the sample has heavier tails or a lack of symmetry due to the fact that the underlying distribution is not that which is assumed. Thus, an outlier can be due to contamination but there is also the possibility that it is due to an erroneous distributional assumption. Since it is often assumed that a sample has a normal distribution, the case where this assumption is assumed yet invalid is referred to as normal model departure. Section 8.2 of [39] also considers the delicate concept of extreme observations and outliers. For more details on outliers, their causes and their impact on statistical inference, the reader is referred to the articles in [34, 40–43].

In order to illustrate how outliers can effect estimation and inference, suppose that we have a sample, $Y_{i1}, Y_{i2}, \ldots, Y_{im}$. Clearly, the sample mean $\bar{Y}_i = (1/m)\sum_{j=1}^{m} Y_{ij}$ will go to infinity if even just one of the Y_{ij} goes to infinity. This is the reason why it is said that the mean and variance estimators have zero as their breakdown point. Fortunately, there exist well known outlier-resistant location

estimators which can be used in place of the sample mean and sample variance. As an alternative to the sample mean, outlier-resistant location estimators include the median and the HL and Huber estimators. Similarly, alternative outlier-resistant estimators of the variance, often referred to as scale estimators, include the MAD, IQR, Shamos [44], and the RC estimator [37]. Next, we briefly review some of the properties of these outlier-resistant estimators.

3.1.1 Outlier-Resistant Location Estimators

We provide a brief review of two well known and outlier-resistant location estimators, namely the HL and Huber's M-estimator. First we introduce the HL estimator. This location estimator is defined as

$$\underset{i<j}{\text{median}}\left(\frac{Y_i + Y_j}{2}\right).$$

and is a particularly useful location estimator because it is outlier-resistant and, at the same time, still nearly as efficient as the sample mean. It should also be noted that the HL estimator has greater relative efficiency than the median when the normal distribution assumption holds. In fact, when the normal distribution assumption holds, the HL estimator is nearly as efficient as the sample mean. For more details on the concept of *efficiency*, the readers are referred to Section 6.2 of [45]. The HL estimator also has a reasonable breakdown point of 29.3% [46]. Finally, it should be noted that the HL estimator can be viewed as a compromise between the mean and the median since it is a function of the sample mean of the paired observations and the median of all the pairs. For more details on the properties of the HL estimator, see the articles in [47, 48].

Huber's M-estimator [35] which has a breakdown point of 50%, is also nearly efficient as the sample mean. A short derivation of the estimator provides useful insight. Let $Y_1, Y_2, \ldots, Y_m$ be independent and identically-distributed from a location family with the probability density function $f(y - \mu)$. We assume that the function $f(\cdot)$ is symmetric at zero so that μ is the location parameter of the distribution of Y_i. Now suppose that one wants to estimate the location parameter μ. Let y_i be a realization of the random variable Y_i. Then it is straightforward to show that the maximum likelihood estimate (MLE) can be obtained by solving the following equation for μ:

$$\min_{\mu} \sum_{i=1}^{m} \{-\log f(y_i - \mu)\} = \min_{\mu} \sum_{i=1}^{m} \rho(y_i - \mu),$$

where $\rho(\cdot) = -\log f(\cdot)$. The derivative of the negative of the log of the density function, $\psi(\cdot) = \rho'(\cdot)$, is known as Huber's ψ function. It easily follows that the MLE of the log density is the solution of the following estimating equation

$$\sum_{i=1}^{m} \psi(y_i - \mu) = 0,$$

with $\psi(t) = t$. In order to obtain an outlier-resistant estimator, Huber [49] suggested the following bounded monotone ψ function:

$$\psi(t) = \begin{cases} -c & t < -c \\ x & |t| \leq c \\ c & t > c \end{cases},$$

which is also known as metric Winsorizing [50]. Notice that when $c \to 0$, the solution of the estimating equation results in the well known median estimate and when $c \to \infty$, the estimating equation results in the usual mean estimate.

3.1.2 Outlier-Resistant Scale Estimators

In this section we briefly review the MAD, Shamos, and RC estimators of scale. First, we introduce the median absolute deviation (MAD) estimator. The MAD is a well known outlier-resistant estimator for the standard deviation and is given by

$$\text{MAD}(Y_1, \ldots, Y_m) = \underset{1 \leq i \leq m}{\text{median}} \left\{ |Y_i - \tilde{\mu}| \right\}$$

where $\tilde{\mu} = \text{median}(Y_1, \ldots, Y_m)$. Note that the MAD as defined above is not consistent under the normal distribution. The normal-consistent MAD is easily derived as

$$\frac{\underset{1 \leq i \leq m}{\text{median}} \left\{ |Y_i - \tilde{\mu}| \right\}}{\Phi^{-1}(3/4)} \approx \frac{\underset{1 \leq i \leq m}{\text{median}} \left\{ |Y_i - \tilde{\mu}| \right\}}{0.6744898} \tag{10}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the normal cumulative distribution function. For more details on the normal-consistent MAD, see [33]. Going forward, when the term MAD is used in the current discussion, it will refer to the normal-consistent MAD in (10).

As an alternative to the MAD estimator, Shamos [44] proposed the estimator

$$\underset{i<j}{\text{median}} \left(|Y_i - Y_j| \right)$$

for the scale. Notice that one can view the Shamos estimator as the scale analogue of the HL location estimator. As was the case with the MAD estimator, the Shamos estimator is also not consistent under the normal distribution. A normal-consistent

Shamos estimator can be obtained by dividing the estimator above by $\sqrt{2}\cdot\Phi^{-1}(3/4)$ which results in:

$$\frac{\underset{i<j}{\text{median}}\left(|Y_i - Y_j|\right)}{\sqrt{2}\,\Phi^{-1}(3/4)} \approx \frac{\underset{i<j}{\text{median}}\left(|Y_i - Y_j|\right)}{0.9538726}. \tag{11}$$

For more details on the normal consistency constant, see Lèvy-Leduc et al. [51]. Just as was the case with the MAD estimator, going forward, when the term Shamos estimator is used in the current discussion, it will refer to the normal-consistent version in (11). It should be noted that the Shamos estimator has the same breakdown point of 29.3% as the HL estimator [37].

Rousseeuw and Croux [37] proposed yet another alternative scale estimator which is defined as

$$\underset{i<j}{\text{Quartile}_1}\left(|Y_i - Y_j|\right),$$

where Quartile_1 denotes the first quartile. It is straightforward to show that the normal-consistent version of the RC estimator is defined as

$$\frac{\text{Quartile}_{1\,i<j}\left(|Y_i - Y_j|\right)}{\sqrt{2}\Phi^{-1}(5/8)}. \tag{12}$$

For more details on the normal consistency constant, see Equation (3.7) of Rousseeuw and Croux [37].

3.2 Outlier-Resistance and Breakdown Point

One of the important measures of the robustness of any estimator is its breakdown point. The finite-sample breakdown point is the maximum proportion of incorrect observations (i.e., arbitrarily large observations) that an estimator can handle without leading to an egregiously incorrect estimate. For example, the finite-sample breakdown point of the sample mean is zero because the sample mean will become arbitrarily large when any one of the sample observations becomes arbitrarily large. In what follows, we derive the breakdown points for the various location and scale estimators discussed previously. In cases where the derivation is either extremely straightforward or extremely complicated, the breakdown point is only stated. For more details regarding the finite-sample breakdown point, one is referred to Definition 2 in Section 2.2 of Hampel et al. [34] and Chapter 11 of Huber [36].

3.2.1 Location Estimator Breakdown Points

In order to provide more intuition to the concept of breakdown point, we provide a glimpse into the breakdown point derivations for the median, HL and Huber M-location estimators.

In the case of the median, if we have m observations, then we can make up to $\lfloor (m-1)/2 \rfloor$ of the observations arbitrarily large and still obtain a reasonable estimate of the median. The median estimate will not become arbitrarily large. Note that $\lfloor \cdot \rfloor$ denotes the floor function such that $\lfloor x \rfloor$ is the largest integer not exceeding x. Thus, the finite-sample breakdown point of the median is

$$\epsilon_m = \frac{\lfloor (m-1)/2 \rfloor}{m}. \tag{13}$$

In order to avoid confusion, it should be noted that the breakdown point used in Definition 1 in Section 2.2 of Hampel et al. [34] is obtained by taking the limit of the finite-sample breakdown point as $m \to \infty$. For example, suppose we have a total of m observations and assume that we make $\lfloor (m-1)/2 \rfloor$ of the m observations arbitrarily large. Then, since $\lfloor x \rfloor$ can be rewritten as $\lfloor x \rfloor = x - \delta$ where $0 \leq \delta < 1$, we have

$$\epsilon_m = \frac{\lfloor (m-1)/2 \rfloor}{m} = \frac{1}{2} - \frac{1}{2m} - \frac{\delta}{m}.$$

Thus, we obtain the breakdown point, $\epsilon = \lim_{m \to \infty} \epsilon_m = 1/2 = 50\%$

In the case of the HL estimator, the derivation of the finite-sample breakdown point is somewhat more complex than that derived for the median. In order to simply this derivation, we first consider the following variant of the HL estimator where all pairs are considered:

$$\underset{\forall (i,j)}{\text{median}} \left(\frac{Y_i + Y_j}{2} \right). \tag{14}$$

The idea is to derive the breakdown point for this estimator and then use a similar approach to derive the breakdown point for the more standard version of the HL estimator, $\text{median}_{i<j}(Y_i + Y_j)/2$. To begin, suppose we have a total of m observations and assume that the we make m^* of the m observations arbitrarily large. The critical step in the breakdown point derivation is to notice that this implies that there are $m \times m$ paired average terms:

$$\frac{Y_i + Y_j}{2}$$

where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$. Because the HL estimator is the median of the $m \times m$ values, the finite-sample breakdown point cannot be greater than $\lfloor (m^2 - 1)/2 \rfloor / m^2$. This results in the following relation:

$$\frac{m^2-(m-m^*)^2}{m^2} \leq \frac{1}{m^2}\left\lfloor \frac{m^2-1}{2} \right\rfloor.$$

Now, for convenience, let

$$m^{**} = \max\left\{m^* : 1-\left(1-\frac{m^*}{m}\right)^2 \leq \frac{1}{m^2}\left\lfloor \frac{m^2-1}{2} \right\rfloor\right\}.$$

Then the finite-sample breakdown point of the HL estimator is $\epsilon_m = m^{**}/m$, which implies that

$$1-(1-\epsilon_m)^2 \leq \frac{1}{m^2}\left(\frac{m^2-1}{2}-\delta\right) = \frac{1}{2}-\frac{1}{2m^2}-\frac{\delta}{m^2},$$

where $0 \leq \delta < 1$ again. Next, let $\epsilon = \lim_{m\to\infty}\epsilon_m$. Them, taking the limit as $m \to \infty$, we obtain

$$1-(1-\epsilon)^2 \leq \frac{1}{2}.$$

Therefore, in order to obtain the breakdown point of this variant of the HL estimator, we need to find the largest value of ϵ that satisfies this inequality. It is straightforward to show that largest value of ϵ is $1-1/\sqrt{2} \approx 29.3\%$. Given this breakdown point derivation, we can derive a similar result for the standard version of the HL estimator, namely,

$$\underset{i<j}{\text{median}}\left(\frac{Y_i+Y_j}{2}\right),$$

Using a similar but more involved argument to that used when all pairs were considered, it can be shown that the following relation holds:

$$\frac{m(m-1)/2-(m-m^*)(m-m^*-1)/2}{m(m-1)/2} \leq \frac{1}{m(m-1)/2}\left\lfloor \frac{m(m-1)/2-1}{2} \right\rfloor.$$

Now, let m^{**} be the maximum value of m that satisfying this inequality. Clearly, given m^{**}, the finite-sample breakdown point is given by $\epsilon_m = m^{**}/m$. Then, omitting the details of how one obtains m^{**}, it can again be shown that $\epsilon = 1-1/\sqrt{2} \approx 29.3\%$ in the case of the standard version of the HL estimator.

Finally, Huber's M-estimator turns out to have a breakdown point of 50% but its derivation is complicated so the details have been omitted. For more information on the derivation, see [36, 52].

3.2.2 Scale Estimator Breakdown Points

Using a derivation technique similar to that used for the HL location estimator, it can be shown that the Shamos estimator also has a breakdown point $\epsilon = 1 - 1/\sqrt{2} \approx 29.3\%$. In the case of the RC estimator, the finite-sample breakdown point can be obtained by using a similar technique to that used in Theorem 5 of [37]. Notice that the relation derived earlier for the median, $\epsilon_m = \lfloor (m-1)/2 \rfloor / m$, also holds for the MAD which implies that it too has a breakdown point of 50%. The sample standard derivation estimator is clearly analogous to the mean estimator which implies that it also has a breakdown point of zero. Finally, using a similar argument to that used for the median, it is straightforward to show that the breakdown point for the IQR estimator is 25%.

3.3 Outlier-Resistance and Relative Efficiency

Another important property of an outlier resistant estimator is its efficiency [48]. Relative efficiency is a metric for comparing the effectiveness of the two estimators. In practice, the usual method for comparing the dispersion of two estimators (say, $\hat{\theta}_2$ and $\hat{\theta}_1$), is to calculate the ratio of the variances of the respective estimators. This ratio is known as the relative efficiency of $\hat{\theta}_2$ and $\hat{\theta}_1$ (see Section 2.2 of Lehmann [47]), and is defined as

$$\text{RE}(\hat{\theta}_2, \hat{\theta}_1) = \frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)} \times 100\% \tag{15}$$

where $\hat{\theta}_1$ is often a reference or baseline estimator. The asymptotic relative efficiency (ARE) is then defined as the limit of the RE as the sample size $n \to \infty$.

For brevity reasons, we omit the ARE derivations and summarize the AREs of the outlier-resistant estimators under consideration along with their corresponding breakdown points in Table 1. For more details, see Section 2.2 of Lehmann [47], and the studies in [37, 48, 53].

Table 1 Breakdown Points and AREs of the estimators under consideration

	Location				Scale				
	Mean	Median	HL	Huber	SD	IQR	MAD	Shamos	RC
Breakdown	0%	50%	29%	50%	0%	25%	50%	29%	50%
ARE	100%	64%	96%	95%	100%	38%	37%	86%	82%

4 Estimator Performance Assessment: Two Simulation Studies

The goal of the study that follows is to assess whether any of the estimation methods, namely methods 1–4, stands out as far as its effectiveness when dealing with the issue of outliers. As explained earlier, there are two different causes for the existence of outliers in the sample observations. One reason for their existence is because some of the observations have been contaminated. Another possibility is that the underlying distribution of the sample is actually not that distribution which is being assumed. Given that the common distributional assumption is that the underlying distribution is normal, this outlier cause is usually referred to as normal model departure.

A well known approach for assessing the effect of proposed estimators is Monte Carlo simulation. Given that there are two possible causes for the existence of outliers, our investigation into the effectiveness of the outlier-resistant methods consists of two different large scale simulations.

The first simulation, referred to as simulation A, will assess the performance of the proposed estimators under two different scenarios. Scenario one is referred to as the baseline case where the simulated data is normally distributed and not contaminated. Scenario two is referred to as the contaminated case where the simulated data is normally distributed but contaminated. Both scenarios need to be considered because an outlier-resistant estimation method should work well under both scenarios.

The second simulation, referred to as simulation B, the goal is identical to that of simulation A except that rather than assuming that the data is contaminated, the underlying assumption that there is normal model departure. This simulation will also consider the estimator performance under two different scenarios. Scenario one is again referred to as the baseline case where there is no model departure in that the observations are normally distributed. Scenario two is referred to as the normal departure case where the simulated data exhibits normal model departure in various ways which will be explained in detail.

Notice that both simulation A and simulation B have the same goals with the only difference being the manner in which the outliers arise. In what follows, we describe the various components of the simulations. Unless specified otherwise, the various components of the description apply to simulation A and simulation B along with each of the two scenarios in each simulation.

4.1 The Estimation Methods

Although, there were a total of 12 possible combinations of location and scale estimators, as aforementioned in Sect. 1, each simulation included the following four methods:

- Method 1: The sample mean and the sample variance. This is the baseline case with each estimator having a zero breakdown point. The relative efficiency of each estimator is obviously 100%.
- Method 2: The sample median and the MAD. Each estimator has a 50% breakdown point. The median has an ARE of 64% and the MAD has an ARE of 37%.
- Method 3: The Hodges–Lehmann and Shamos estimators. Each of these estimators has a 29% breakdown point. The Hodges–Lehmann estimator has an ARE of 96% and the ARE of the Shamos estimator is 86%.
- Method 4: Huber's M-estimator and the RC estimator. Each estimator has a 50% breakdown point. Huber's M-estimator has an ARE of 96% and the ARE of the Shamos estimator is 82%.

4.2 The Response Functions

When running simulations in which parameters are estimated, it is generally necessary to view the true underlying parameters as known. We can achieve this by using a similar approach to that in [31]. Specifically, we make the assumption that the true process mean $M(\mathbf{x})$ and variance $V(\mathbf{x})$ are known to be the following:

$$M(\mathbf{x}) = T_0 + 5(x_1^2 + x_2^2) \tag{16}$$

and

$$V(\mathbf{x}) = 1 + (x_1 - 1)^2 + (x_2 - 1)^2, \tag{17}$$

where the target $T_0 = 50$. It is then immediate from (16) and (17) that we have

$$\begin{aligned} \phi(\mathbf{x}) &= \{M(\mathbf{x}) - T_0\}^2 + V(\mathbf{x}) \\ &= 25(x_1^2 + x_2^2)^2 + 1 + (x_1 - 1)^2 + (x_2 - 1)^2. \end{aligned} \tag{18}$$

We can then obtain the true optimal operating conditions, by finding the solution to $\partial\phi/\partial x_1 = 0$ and $\partial\phi/\partial x_2 = 0$. Solving these equations for x_1 and x_2, we obtain $x_1 = x_2 = 1/5$. For more details, see Section 5 in [31]. Thus, in our simulation framework $x_1 = x_2 = 1/5$ are the true theoretical optimal operating conditions when the process mean and variance are known and are denoted as $x_1^{\text{oc}} = 1/5$ and $x_2^{\text{oc}} = 1/5$. The goal of the two simulations is to evaluate the performance of the proposed Methods 1–4 comparing the estimated optimal operating conditions with the true theoretical optimal operating conditions.

4.3 Simulation Design Points

In both simulations, the design points used are exactly the same. We considered a standard 3×3 factorial design with three levels. That is, the design points are located at (x_{1i}, x_{2i}) with $x_{1,i} = -1, 0, 1$ and $x_{2,i} = -1, 0, 1$. Therefore, nine design points are used in each simulation. Also, it is assumed that there are $m = 50$ replicates at each design point.

4.4 The Optimization Formulation

Note that the estimated operating conditions are obtained using the formulation described earlier, namely, the minimization of the squared-loss optimization model: $\left\{\hat{M}(\mathbf{x}) - T_0\right\}^2 + \exp\left(\hat{V}_{\log}(\mathbf{x})\right)$ where T_0 is the customer-identified target value for the quality characteristic of interest. This formulation results in the optimal operating conditions which minimize the MSE in (9) for each of Methods 1–4.

We should point out that the "`optim`" function in R Core Team [54] using the option for box-constrained optimization along with the Nelder–Mead algorithm is used to minimize the objective function. In each simulation, the number of repetitions is set to $I = 10{,}000$ so that 10,000 estimated optimal operating conditions are obtained for each of the proposed Methods 1–4.

4.5 Generating the Sample Observations in Each Simulation

As explained previously, in simulation A, the observations are generated using a contamination scheme and, in simulation B, the observations are generated based on a normal model departure scheme. We provide the details of the two approaches in what follows.

4.5.1 Simulation A: Contamination

Generating the sample observations for the baseline scenario where there is no contamination is quite straightforward. Recall that, at each design point, the response functions, $M(\mathbf{x}_i)$ and variance $V(\mathbf{x}_i)$, are known exactly. Therefore, for the case of no contamination, a random sample of size $m = 50$ is generated from the normal distribution with mean $M(\mathbf{x}_i)$ and variance $V(\mathbf{x}_i)$ at each design point $\mathbf{x}_i = (x_{1i}, x_{2i})$ where $i = 1, 2, \ldots, 9$.

Clearly, the goal of the simulation is to investigate the robustness properties of the proposed estimation methods. One way to achieve this goal is to use an approach analogous to that often used for studying the robustness properties of the

empirical influence function [55] and the ϵ-influence function [42, 56]. Essentially, our approach assesses how the relative efficiency of the respective estimator is influenced by the artificial contamination level of the sample.

For the scenario where the data set is assumed to be contaminated, instead of using the pure normal data set, we contaminate 10% of the responses. Again, since the response functions are known exactly, we implement the contamination scheme by using the following mixture model for the observations:

$$Y \sim (1-\pi)\cdot N\big(M(\mathbf{x}), V(\mathbf{x})\big) + \pi \cdot N\big(M_c^*(\mathbf{x}), V(\mathbf{x})\big), \tag{19}$$

where $\pi = 0.1$ (contamination proportion) and $M_c^*(\mathbf{x}) = T_c^* + 5(x_1^2 + x_2^2)$ (contaminated mean response) with $T_c^* = 60, 70, 80$. In this manner, the mean response is contaminated by increasing the target value characteristic of $T_0 = 50$ first by 10, then 20 and finally 30.

The contamination scenario was then replicated $I = 10{,}000$ times resulting in 10,000 optimal operating conditions for Methods 1–4 at each of the three levels of contamination: $T_c^* = 60, 70, 80$. Notice that the contamination scheme used here is slightly different from that used in [31] where a delta noise scheme is used rather than a mixture model.

4.5.2 Simulation B: Normal Model Departure

Since the response functions are identical to those in simulation A, generating the sample observations for the baseline scenario where there is no model departure is done exactly in the same way as it was done for the baseline case of simulation A. A random sample of size $m = 50$ is generated from the normal distribution with mean $M(\mathbf{x}_i)$ and variance $V(\mathbf{x}_i)$ at each design point $\mathbf{x}_i = (x_{1i}, x_{2i})$ where $i = 1, 2, \ldots, 9$.

For the scenario where normal departure is assumed, we consider six alternative distributions. These are

1. The Student t-distribution with three degrees of freedom,
2. The Student t-distribution with four degrees of freedom,
3. The Student t-distribution with five degrees of freedom,
4. The Laplace distribution,
5. The logistic distribution, and
6. The uniform distribution.

Again, at each design point, the response functions, $M(\mathbf{x}_i)$ and variance $V(\mathbf{x}_i)$, are known exactly and are the same as those in simulation A. Therefore, generating the sample observations such that they come from the respective distributions is reasonably straightforward aside from one difficulty. Recall that the goal is to generate a random sample of size $m = 50$ from the respective distribution with mean $M(\mathbf{x}_i)$ and variance $V(\mathbf{x}_i)$ at each design point $\mathbf{x}_i = (x_{1i}, x_{2i})$ where

$i = 1, 2, \ldots, 9$. Therefore, when generating samples in the normal model departure simulation, we need to generate each of the samples from the respective non-normal distribution in such a manner that the mean at the design point x_i equals $M(\mathbf{x}_i)$ and the variance at the same design point equals $V(\mathbf{x}_i)$. Specifically, we need to determine the parameter values associated with the respective distributions such that we satisfy

$$E\big(Y_{ij}\big) = E\big[M(\mathbf{x}_i) + U_{ij}\big] = M(\mathbf{x}_i) \tag{20}$$

and

$$\mathrm{Var}\big(Y_{ij}\big) = \mathrm{Var}\big[M(\mathbf{x}_i) + U_{ij}\big] = \mathrm{Var}\big(U_{ij}\big) = V(\mathbf{x}_i) \tag{21}$$

at each design point $\mathbf{x}_i$.

An example should clarify the issue. Consider the normal departure scenario where one wants to generate samples from the Laplace distribution. By definition, a standard Laplace random variable, U_{ij}, has a mean equal to zero and variance equal to two. Thus, in the case of the Laplace distribution, if we randomly generate our samples, Y_{ij}, using $Y_{ij} = M(\mathbf{x}_i) + 1/\sqrt{2} \cdot V(\mathbf{x}_i)^{1/2} \cdot U_{ij}$, we obtain $E\big(Y_{ij}\big) = M(\mathbf{x}_i)$ and $\mathrm{Var}\big(Y_{ij}\big) = V(\mathbf{x}_i)$. Notice that this is exactly what we want the mean and variance to be at the design point $\mathbf{x}_i$ as described in (20) and (21).

A similar approach can be used for the other distributions. In the case of the logistic distribution, a standard logistic random variable, U_{ij}, has a mean equal zero and variance equal to $\pi^2/3$. Therefore, if we generate our logistic samples using $Y_{ij} = M(\mathbf{x}_i) + \sqrt{3}/\pi \cdot V(\mathbf{x}_i)^{1/2} \cdot U_{ij}$, then clearly the mean of Y_{ij} will be equal to $M(\mathbf{x}_i)$ and the variance of Y_{ij} will be equal to $V(\mathbf{x}_i)$. In the case of the three t-distributions, a similar argument applies. A t-distributed random variable with ν degrees of freedom, U_{ij}, has mean equal to zero and variance $\nu/(\nu - 2)$. Therefore, we obtain the desired mean and variance by generating samples using $Y_{ij} = M(\mathbf{x}_i) + \{(\nu - 2)/\nu\}^{1/2} \cdot V(\mathbf{x}_i)^{1/2} \cdot U_{ij}$. Finally, consider the case of the uniform distribution. If U_{ij} is a uniform random variable in $(-1, 1)$, then it is straightforward to show that its variance is equal to $1/3$. Therefore, by using $Y_{ij} = M(\mathbf{x}_i) + \sqrt{3} \cdot V(\mathbf{x}_i)^{1/2} \cdot U_{ij}$, we will obtain samples with mean equal to $M(\mathbf{x}_i)$ and variance equal to $V(\mathbf{x}_i)$.

4.6 Simulation Performance Results

The two simulations, A and B, were run and the 10,000 optimal operating conditions were obtained for both the baseline case and non-baseline case. In the next two sections, we report various results associated with simulation A and simulation B respectively.

4.6.1 Simulation A: Contamination

In Fig. 1, we plotted the optimal operating conditions for both the baseline case and non-baseline case using each of the proposed Methods 1–4. It turned out that the results in simulation A were quite similar for $T_c^* = 60$, 70 and 80 so we only plotted the results where $T_c^* = 60$. Also, only the first 100 out of the 10,000 resulting data points were included in Fig. 1 because it becomes quite difficult to see the pattern if one uses the complete set of 10,000 data points.

Also, in Fig. 1, the circles denote the optimal operating conditions without contamination, the baseline case, while the crosses denote the optimal operating conditions with contamination, the non-baseline case. Recall that the true optimal operating conditions are located at $(x_1^{oc}, x_2^{oc}) = (1/5, 1/5)$. Therefore, it is clear that, under Method 1, the optimal operating conditions obtained under contamination are shifted away from the true optimal operating condition locations.

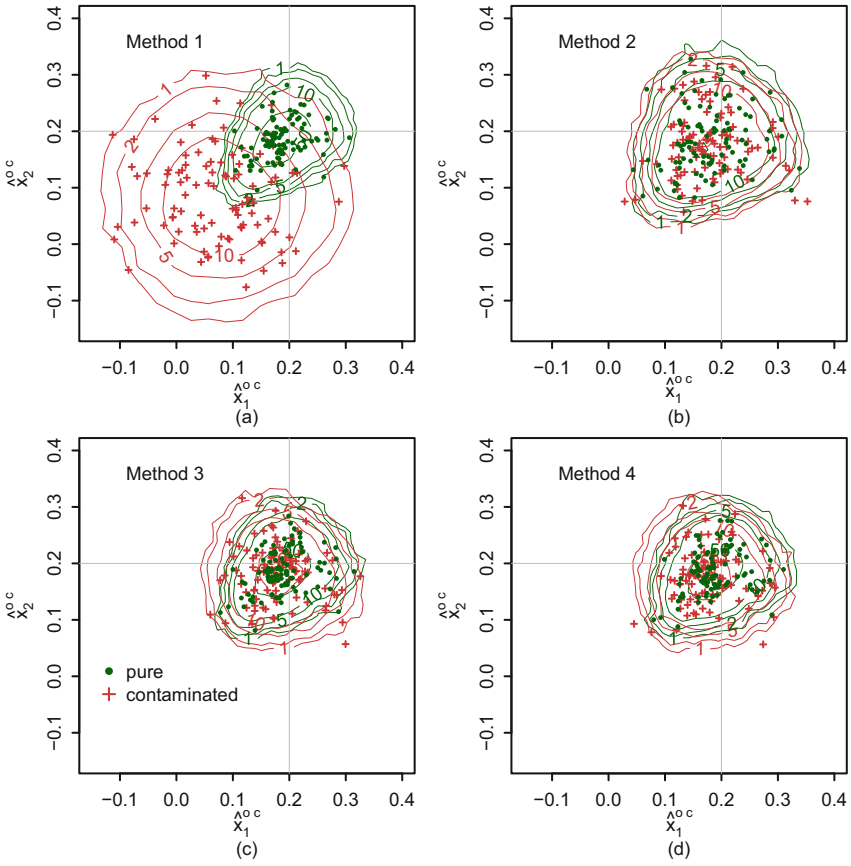


Fig. 1 The optimal operating conditions obtained using Methods 1–4

Conversely, those obtained using Methods 2–4 do not result in the same shift. This result is consistent with the fact that the sample mean and the sample variance are extremely sensitive to contamination while the other methods are less sensitive.

In order to obtain additional insight, in Fig. 1, we also plotted the contours from the two-dimensional kernel densities [57] of the optimal operating conditions. Unlike in the case of the scatter plots, we were able to use the complete set of 10,000 optimal operating conditions when drawing the contours. In the case of Method 1, the contours are clustered more closely than the other methods which is consistent with the notion that the sample mean and the sample variance are efficient estimators in the case of no contamination.

Although the plots in Fig. 1 illustrate the utility of Methods 2–4, they do not offer the ability to discern which methods stand out in terms of their performance. A more useful approach would be one that expresses the relative performance of Methods 2–4 to the performance of Method 1 when Method 1 is used in the no contamination case. This approach would be particularly effective because it is already known that the performance of Method 1 in the no contamination case is optimal.

Recall that, in Sect. 2, we introduced the concept of relative efficiency as defined in (15). Unfortunately, the usual relative efficiency measure can be used only when the estimates being compared are univariate. However, the estimated optimal operating conditions obtained in our simulations are bivariate. In order to deal with this multivariate issue, we use the generalized variance of each estimator which is the determinant of the covariance matrix of the bivariate estimates. For more details on the generalized variance, one is referred to [58, 59]. The first step to calculating the generalized variance is to calculate the variance-covariance matrix associated with each Method ℓ using the true optimal operating conditions, $(x_1^{\text{oc}}, x_2^{\text{oc}}) = (1/5, 1/5)$, as the population location-scale parameters. The variance-covariance matrix resulting from the estimates using Method ℓ is easily obtained as

$$\mathbf{S}_\ell = \begin{pmatrix} \frac{1}{I}\sum_{i=1}^{I}\left(\hat{x}_{\ell,1,i} - \frac{1}{5}\right)^2 & \frac{1}{I}\sum_{i=1}^{I}\left(\hat{x}_{\ell,1,i} - \frac{1}{5}\right)\left(\hat{x}_{\ell,2,i} - \frac{1}{5}\right) \\ \frac{1}{I}\sum_{i=1}^{I}\left(\hat{x}_{\ell,1,i} - \frac{1}{5}\right)\left(\hat{x}_{\ell,2,i} - \frac{1}{5}\right) & \frac{1}{I}\sum_{i=1}^{I}\left(\hat{x}_{\ell,2,i} - \frac{1}{5}\right)^2 \end{pmatrix}, \quad (22)$$

where I is the replication number in the Monte Carlo simulation and $(\hat{x}_{\ell,1,i}, \hat{x}_{\ell,2,i})$ denotes the estimated optimal operations conditions using Method ℓ in the ith iteration. Then, using (22), the relative efficiency of Method ℓ to Method 1 based on the generalized variance [31, 32] is defined as

$$\text{RE}_{\text{GV}}(\text{Method } \ell \mid \text{Method } 1) = \frac{\|\mathbf{S}_1\|}{\|\mathbf{S}_\ell\|}, \quad (23)$$

where $\|\cdot\|$ denotes the determinant of a matrix and $\ell = 1, 2, 3, 4$. It should be noted that the variance-covariance matrix, $\mathbf{S}_1$, in the numerator of (23) is obtained by using Method 1 with no data contamination.

Table 2 Generalized variances ($\times 10^6$)

Method	Method 1	Method 2	Method 3	Method 4
No contamination	2.43013	10.11812	2.952882	3.401389
Contamination ($T_c^* = 60$)	280.16099	12.38782	7.226525	7.204433
Contamination ($T_c^* = 70$)	4643.16815	12.36313	7.073645	7.074748
Contamination ($T_c^* = 80$)	26015.40587	12.36313	7.369051	7.074732

Table 3 Relative efficiencies (%) of each method to Method 1 (without contamination) based on the generalized variance with $T_c^* = 60, 70, 80$

Method	Method 1	Method 2	Method 3	Method 4
No contamination	100.00%	24.02%	82.30%	71.45%
Contamination ($T_c^* = 60$)	0.87%	19.62%	33.63%	33.73%
Contamination ($T_c^* = 70$)	0.05%	19.66%	34.35%	34.35%
Contamination ($T_c^* = 80$)	0.01%	19.66%	32.98%	34.35%

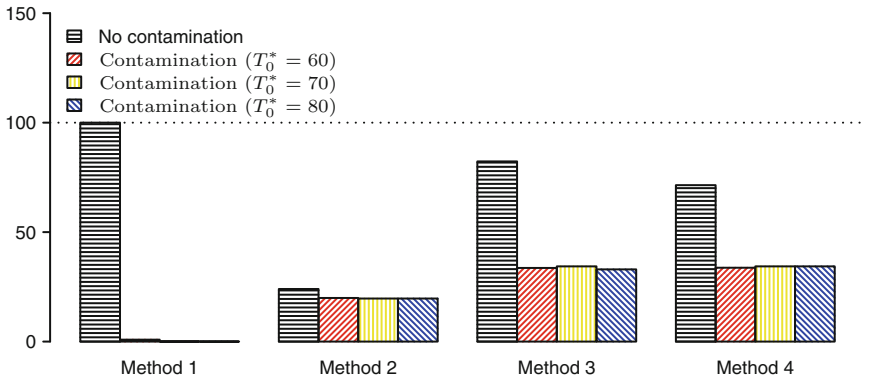


Fig. 2 Relative efficiencies (%) of each method under consideration to Method 1 (without contamination)

Using (23), we calculate the generalized variances and provide them in Table 2. More importantly, the relative efficiencies of all the methods to Method 1 were calculated and are provided in Table 3. In order to obtain a visual illustration of the relative efficiencies, we summarize them again using the bar-chart in Fig. 2. Two narratives emerge from Tables 2, 3 and particularly Fig. 2. The first is that Methods 2–4 clearly outperform Method 1 in terms of their relative efficiency when the data is contaminated. In fact, as one would expect, the performance of Method 1 is extremely poor because of the zero breakdown point of the sample mean and the sample variance. As expected, when the data is not contaminated, Method 1 expectedly stands out as the ideal estimation method. Yet, even in the case of no contamination, Method 3 and Method 4 are quite competitive with relative efficiencies of 81.30% and 71.45% respectively.

4.6.2 Simulation B: Normal Model Departure

The scatter plots of the estimated optimal operating conditions using the generated observations were plotted in Park et al. [32]. Thus, for the scatter plots generated by Methods 1–3, the reader is referred to Figures 2–8 in [32]. Also, since the shape of the scatter plot of the estimated optimal operating conditions using Method 4 is quite similar to those in [32], we decided to omit the Method 4 scatter plot also. Just as was the case in Simulation A, a single-number representing a performance metric is generally more objective and useful than the use of scatter plots. Therefore, we decided to calculate the relative efficiencies of Methods 1–4 resulting from simulation B.

The generalized variances resulting from the simulations are shown in Table 4. Notice that the relative efficiency for simulation B is calculated using a slightly different baseline approach than that used in simulation A. Since simulation B is a model departure simulation, obviously the concept of contamination versus no-contamination does not exist. Therefore, the relative efficiency is calculated so that the same distribution is used in the numerator and the denominator. Using this approach, only the estimation methods are being compared. An example should clarify the calculation. Consider the relative efficiency calculation for Method 2 when the underlying distribution is uniform. In this case, the numerator is the determinant of the generalized variance resulting when Method 1 is used and the underlying distribution is uniform. Similarly, the denominator is the determinant of the generalized variance resulting when Method 2 is used and the underlying distribution is also uniform. Therefore, in general, the simulation B relative efficiency is calculated so that the same underlying distribution assumption is used in both the numerator and the denominator. Therefore, the baseline can be viewed as Method 1 rather than not contaminated (i.e., the normal distribution) as was the case in simulation A.

The simulation B relative efficiencies are provided in Table 5 and also summarized using the bar-chart in Fig. 3. By definition, the relative efficiencies in the Method 1 column of Table 5 are always 100% and therefore provide no additional information. The key result is that, in the case of model departure, Methods 3 and 4 exhibit superior relative performance when compared with Methods 1 and 2. Notice that, even in the case of no model departure where the underlying distribution is

Table 4 Generalized variances ($\times 10^6$) of each method under consideration

Distribution	Method 1	Method 2	Method 3	Method 4
Normal	2.3305	9.6682	2.8258	3.2882
$t(\text{df} = 3)$	24.0699	14.6284	7.2100	8.0266
$t(\text{df} = 4)$	10.1848	12.2450	5.0687	5.7177
$t(\text{df} = 5)$	6.5478	11.3380	4.3306	4.9551
Laplace	6.1039	16.3641	6.1995	8.0322
Logistic	3.6768	10.5310	3.6259	4.2265
Uniform	1.0187	13.8284	1.2664	1.1388

Table 5 Relative efficiencies (%) of each method using the respective Method 1 distribution in the numerator

Distribution	Method 1	Method 2	Method 3	Method 4	$\kappa - 3$	τ_4
Normal	100.0%	24.1%	82.5%	70.9%	0	0.1226
$t\,(\mathrm{df} = 3)$	100.0%	164.5%	333.8%	299.9%	∞	0.2612
$t\,(\mathrm{df} = 4)$	100.0%	83.2%	200.9%	178.1%	∞	0.2168
$t\,(\mathrm{df} = 5)$	100.0%	57.8%	151.2%	132.1%	6	0.1936
Laplace	100.0%	37.3%	98.5%	76.0%	3	0.2357
Logistic	100.0%	34.9%	101.4%	87.0%	1.2	0.1667
Uniform	100.0%	7.4%	80.4%	89.5%	−1.2	0

The kurtosis of each distribution is shown in the last column

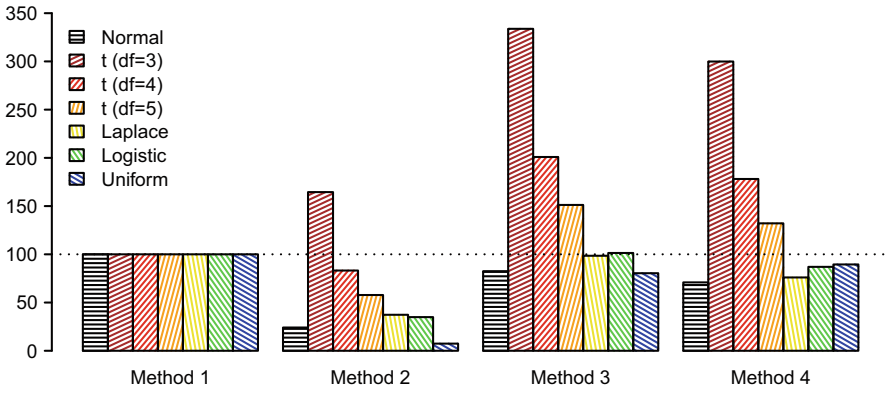


Fig. 3 Relative efficiencies (%) of each method under consideration to Method 1

normal, the performance of Method 3 and Method 4 is still quite reasonable with resulting relative efficiencies of 82.50% and 70.90% respectively.

The other interesting narrative that Table 5 hints at is the possible systematic relation between the method estimation performance and the kurtosis of the respective underlying distribution. Park et al. [32] suggested the use of kurtosis as means of measuring the degree to which a non-normal distribution signifies normal model departure. An informal argument is that a positive kurtosis value implies heavier tails than those of the normal distribution and a negative value implies lighter tails. Therefore, from a statistical inference point of view, heavier tails tend to be much more problematic than lighter tails which supports the idea that positive values of kurtosis can be associated with greater levels of normal departure. For more detailed discussions regarding the properties of kurtosis, one is referred to [32, 60, 61].

Notice that we include the kurtosis of each distribution in the sixth column of Table 5. The pattern that emerges from Table 5 is that, as the kurtosis increases, the performance of Method 3 and Method 4 improves dramatically. What this means is that Methods 3 and 4 are extremely efficient in the case where the normal model

departure is more severe in the sense of heavier tails. This particularly evident in the case of the Student t-distributions where the kurtosis is extremely large and Method 3 and Method 4 exhibit excellent relative performance.

The one problem with the notion that the kurtosis is positively related to the performance of Methods 3 and 4 is that the standard kurtosis measure tends to approach infinity quite quickly. Any distribution with heavier tails than the Student t-distribution with four degrees of freedom will have a value of kurtosis equal to ∞. Therefore, using the conventional measure of kurtosis is not an effective way of measuring the degree of normal model departure, particularly for heavier tailed distributions. In order to circumvent this problem, we suggest the use of the L-kurtosis which was originally developed by Hosking [62]. Similar to the conventional kurtosis measure, the L-kurtosis measures the heaviness of the tails of a distribution. The advantage of the L-kurtosis is that, although the definition is somewhat more complex, it does not possess the ∞ behavior of the conventional kurtosis measure. The L-kurtosis is defined as

$$\tau_4 = \frac{\lambda_4}{\lambda_2}.$$

Here, λ_2 and λ_4 are defined as

$$\lambda_2 = \frac{1}{2} E\Big(X_{(2:2)} - X_{(1:2)}\Big),$$

and

$$\lambda_4 = \frac{1}{4} E\Big(X_{(4:4)} - 3X_{(3:4)} + 3X_{(2:4)} - X_{(1:4)}\Big),$$

where $X_{(k:n)}$ is an order statistic representing the kth smallest value in a sample of size n. For more details, see Hosking [62].

The last column of Table 5, contains the L-kurtosis measure of each respective distribution. Note that the positive relationship between the L-kurtosis and the performance of Methods 3 and 4 is still quite evident.

5 Concluding Remarks

In this chapter, we have provided the background behind a dual response surface methodology that incorporates robust design. We also introduced and described estimators and procedures that can be used for outlier-resistant location and scale estimation of the dual response surfaces.

Through the use of two large scale simulations, we provided empirical evidence that, Method 3, which uses Hodges–Lehmann estimator for scale and the Shamos

estimator for scale, results in solid estimation performance when contamination or normal model departure exists in the underlying sample. The same solid performance also resulted from Method 4 which uses the Huber estimator for location and the Rousseeuw-Croux estimator for scale.

In the case of normal departure, our simulation indicated a systematic connection between the level of kurtosis in the true distribution and the performance of Methods 3 and 4. Given that the performance of Methods 3 and 4 were quite similar and that the breakdown points of the Method 3 estimators and Method 4 estimators are 29% and 50% respectively, we strongly suggest the use of Method 4 if one is interested in incorporating a robust design methodology into the dual response surface framework.

Acknowledgements The work of Professor Ouyang was supported by the National Natural Science Foundation of China under grants NSFC-71811540414 & 71702072 and the Natural Science Foundation for Jiangsu Institutions under grant BK20170810. The work of Professor Park was supported under the framework of the international cooperation program managed by the National Research Foundation of Korea (2018K2A9A2A06019662). The work of Professor Byun was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (2016R1D1A1B03935397).

References

1. Bendell, A., Disney, J., & Pridmore, W. A. (1987). *Taguchi methods: Applications in world industry*. London, UK: IFS Publications.
2. Dehnad, K. (1989). *Quality control, robust design and the Taguchi method*. Pacific Grove, CA: Wadsworth and Brooks/Cole.
3. Taguchi, G., & Wu, Y. (1980). *Introduction to off-line quality control*. Nagoya: Central Japan Quality Control Association.
4. Taguchi, G. (1986). *Introduction to quality engineering*. White Plains, NY: Asian Productivity Organization.
5. Ross, P. J. (1988). *Taguchi techniques for quality engineering*. New York, NY: McGraw-Hill Book Company.
6. Kenett, R., & Zacks, S. (1998). *Modern industrial statistics: Design and control of quality and reliability*. Pacific Grove, CA: Duxbury.
7. Box, G. E. P. (1985). Discussion of off-line quality control, parameter design and the Taguchi methods. *Journal of Quality Technology, 17*, 198–206.
8. Box, G. E. P. (1988). Signal-to-noise ratios, performance criteria, and transformations. *Technometrics, 30*, 1–17.
9. Vining, G. G., & Myers, R. H. (1990). Combining Taguchi and response surface philosophies: A dual response approach. *Journal of Quality Technology, 22*, 38–45.
10. Pignatiello, J., & Ramberg, J. S. (1991). Top ten triumphs and tragedies of Genichi Taguchi. *Quality Engineering, 4*, 221–225.
11. Myers, R. H., Khuri, A. I., & Vining, G. G. (1992). Response surface alternatives to the Taguchi robust design problem. *The American Statistician, 46*, 131–139.
12. Tiao, G. C., Bisgaard, S., Hill, W. J., Peña, D., & Stigler, S. M. (Eds.). (2000). *Box on quality and discovery with design, control and robustness*. New York, NY: Wiley.
13. Myers, R. H., & Montgomery, D. C. (2002). *Response surface methodology*, 2nd Edn. New York, NY: Wiley.

14. Gauri, S. K., & Pal, S. (2014). The principal component analysis (PCA)-based approaches for multi-response optimization: Some areas of concerns. *The International Journal of Advanced Manufacturing Technology, 70*(9), 1875–1887.
15. Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). *Response surface methodology: Process and product optimization using designed experiments*, 4th Edn. New York, NY: Wiley.
16. Del Castillo, E., & Montgomery, D. C. (1993) . A nonlinear programming solution to the dual response problem. *Journal of Quality Technology, 25*, 199–204.
17. Lin, D. K. J., & Tu, W. (1995). Dual response surface optimization. *Journal of Quality Technology, 27*, 34–39.
18. Copeland, K. A. F., & Nelson, P. R. (1996). Dual response optimization via direct function minimization. *Journal of Quality Technology, 28*, 331–336.
19. Kim, K., & Lin, D. K. J. (1998). Dual response surface optimization: A fuzzy modeling approach. *Journal of Quality Technology, 30*, 1–10.
20. Borror, C. M., & Montgomery, D. C. (2000). Mixed resolution designs as alternative to Taguchi inner/outer array designs for robust design problems. *Quality and Reliability Engineering Inetrnational, 16*, 1–11.
21. Scibilia, B., Kobi, A., Barreau, A., & Chassagnon, R. (2003). Robust designs for quality improvement. *IIE Transactions, 35*, 487–492.
22. Oyeyemi, G. M. (2004). Treatment of non-normal responses from designed experiments. *Journal of the Nigerian Statistical Association, 17*, 8–19.
23. Lee, M. K., Kwon, H. M., Kim, Y. J., & J. Bae. (2005). Determination of optimum target values for a production process based on two surrogate variables. In O. Gervasi, M. L. Gavrilova, V. Kumar, A. Laganá, H. P. Lee, Y. Mun, D. Taniar, & C. J. K. Tan (Eds.), *Computational Science and Its Applications – ICCSA 2005*. Lecture Notes in Computer Science (Vol. 3483, pp. 232–240). Berlin, Springer.
24. Lee, S. B., & Park, C. (2006). Development of robust design optimization using incomplete data. *Computers & Industrial Engineering, 50*, 345–356.
25. Lee, M. K., Kwon, H. M., Hong, S. H., & Kim, Y. J. (2007). Determination of the optimum target value for a production process with multiple products. *International Journal of Production Economics, 107*(1), 173–178. Special Section on Building Core-Competence through Operational Excellence.
26. Lee, S. B., Park, C., & Cho, B.-R. (2007). Development of a highly efficient and resistant robust design. *International Journal of Production Research, 45*, 157–167.
27. Ardakani, M. K., & Wulff, S. S. (2013). An overview of optimization formulations for multiresponse surface problems. *Quality and Reliability Engineering International, 29*(1), 3–16.
28. Park, C. (2013). Determination of the joint confidence region of optimal operating conditions in robust design by bootstrap technique. *International Journal of Production Research, 51*, 4695–4703.
29. Ouyang, L., Ma, Y., Byun, J.-H., Wang, J., & Tu, Y. (2016). An interval approach to robust design with parameter uncertainty. *International Journal of Production Research, 54*(11), 3201–3215.
30. Ardakani, M. K. (2016). The impacts of errors in factor levels on robust parameter design optimization. *Quality and Reliability Engineering International, 32*(5), 1929–1944.
31. Park, C., & Leeds, M. (2016). A highly efficient robust design under data contamination. *Computers & Industrial Engineering, 93*, 131–142.
32. Park, C., Ouyang, L., Byun, J.-H., & Leeds, M. (2017). Robust design under normal model departure. *Computers & Industrial Engineering, 113*, 206–220.
33. Park, C., & Cho, B.-R. (2003). Development of robust design under contaminated and non-normal data. *Quality Engineering, 15*, 463–469.
34. Hampel, F. R., Ronchetti, E., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York, NY: Wiley.

35. Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 35*, 73–101.
36. Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics*, 2nd Edn. New York, NY: Wiley.
37. Rousseeuw, P., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association, 88*, 1273–1283.
38. Bartlett, M. S., & Kendall, D. G. (1946). The statistical analysis of variance-heterogeneity and the logarithmic transformation. *Journal of the Royal Statistical Society, 8*, 128–138.
39. Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics, 22*, 1081–1114.
40. Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448–485). Stanford: Stanford University Press.
41. Park, C., Basu, A., & Basu, S. (1995). Robust minimum distance inference based on combined distances. *Communications in Statistics: Simulation and Computation, 24*, 653–673.
42. Park, C., & Basu, A. (2003). The generalized Kullback-Leibler divergence and robust inference. *Journal of Statistical Computation and Simulation, 73*, 311–332.
43. Basu, A., Shioya, H., & Park, C. (2011). *Statistical inference: The minimum distance approach*. Monographs on Statistics and Applied Probability. London: Chapman & Hall.
44. Shamos, M. I. (1976). Geometry and statistics: Problems at the interface. In J. F. Traub (Ed.), *Algorithms and complexity: New directions and recent results* (pp. 251–280). New York, NY: Academic Press.
45. Hogg, R. V., McKean, J. W., & Craig, A. T. (2013). *Introduction to mathematical statistics*, 7th Edn. London: Pearson.
46. Hettmansperger, T. P., & McKean, J. W. (2010). *Robust nonparametric statistical methods*, 2nd Edn. Boca Raton, FL: Chapman & Hall/CRC.
47. Lehmann, E. L. (1999). *Elements of large-sample theory*. New York, NY: Springer.
48. Serfling, R. J. (2011). Asymptotic relative efficiency in estimation. In M. Lovric (Ed.), *Encyclopedia of statistical science, Part I* (pp. 68–82). Berlin: Springer.
49. Huber, P. J. (1981). *Robust statistics*. New York, NY: Wiley.
50. Dixon, W. J. (1960). Simplified estimation for censored normal samples. *Annals of Mathematical Statistics, 31*, 385–391.
51. Lèvy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M. S., & Reisen, V. A. (2011). Large sample behaviour of some well-known robust estimators under long-range dependence. *Statistics, 45*, 59–71.
52. Huber, P. J. (1984). Finite sample breakdown of *M*- and *P*-estimators. *Annals of Statistics, 12*, 119–126.
53. Staudte, R. G., & Sheather, S. J. (1900). *Robust estimation and testing*. New York, NY: Wiley.
54. R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
55. Hampel, F. R., Marazzi, A., Ronchetti, E., Rousseeuw, P. J., Stahel, W. A., & Welsch, R. E. (1982). Handouts for the instructional meeting on robust statistical methods. In *The 15th European Meeting of Statisticians*, Palermo, Italy.
56. Park, C., & Basu, A. (2011). Minimum disparity inference based on tangent disparities. *International Journal of Information and Management Sciences, 22*, 1–25.
57. Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.
58. Anderson, T. W. (1993). *An introduction to multivariate statistical analysis*. London: Wiley.
59. Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*, 6th Ed. Englewood Cliffs, NJ: Prentice-Hall.
60. DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*, 292–307.
61. Westfall, P. H. (2014). Kurtosis and peakedness, 1095–2014. R.I.P. *The American Statistician, 68*, 191–195.
62. Hosking, J. R. M. (1990). *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B, 52*, 105–124.

Defects Driven Yield and Reliability Modeling for Semiconductor Manufacturing

Tao Yuan, Suk Joo Bae, and Yue Kuo

Abstract Manufacturing processes of modern ultra-large-scale integrated circuits are highly complex and costly. Defects generated in the manufacturing processes are unavoidable and affect not only manufacturing yield but also device reliability. In this reason, accurate modeling of the spatial defects distribution is imperatively important for yield and reliability estimation as well as process improvement. Defects on semiconductor wafers tend to cluster, which introduces excessive zeros, causing over-dispersion in defect count data. This chapter discusses some latest development in modeling the non-homogeneously distributed spatial defect counts, focusing on Bayesian spatial regression approaches based on Poisson models, negative binomial models, and zero-inflated models. Real wafer map data are used to evaluate the performance of these models. In addition, the yield models are extended to build extrinsic reliability models based on a defect-growth concept.

1 Introduction

The semiconductor industry has been flourished along with the development of smart devices, e.g., smart phones and tablet PCs. Because the price of semiconductor products decline rapidly throughout the life of a new technology, the ability to ramp yield quickly after the introduction of new technology is fundamental to earning high revenues. Yield is usually defined as the ratio of the number of usable items after completion of production to the number of potentially usable items at the

T. Yuan
Department of Industrial and Systems Engineering, Ohio University, Athens, OH, USA
e-mail: yuan@ohio.edu

S. J. Bae (✉)
Department of Industrial Engineering, Hanyang University, Seoul, South Korea
e-mail: sjbae@hanyang.ac.kr

Y. Kuo
Department of Chemical Engineering, Texas A&M University, College Station, TX, USA
e-mail: yuekuo@tamu.edu

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0_16

beginning of production [15]. It is an important performance measure in semiconductor manufacturing and has been widely used as a key index of profitability and productivity for business [1]. To secure high revenues in a challenging environment, the need for accurate yield prediction in order to evaluate productivity and estimate production costs is essential.

Integrated circuits (ICs) are highly vulnerable to defects generated in the complex and costly semiconductor fabrication processes. When a defect, defined as a variation in quality, is located in a defect-sensitive area (called "critical area"), it results in yield loss and is called a fault (or called "yield defect"). On the other hand, defects that do not cause immediate yield loss may sometimes cause potential reliability issues and are called latent defects or reliability defects. Accurate modeling of the spatial defects distribution is imperatively important for yield and reliability estimation as well as process improvement.

In general, defective IC chips occur close to one another on a semiconductor wafer. Similarly, defect-free chips are found adjacent to one another. These defect patterns cause over-dispersion in defect count data with excess zeros. Models ignoring such clustered defects patterns may risk grossly underestimating the true yield. This chapter discusses spatial modeling approaches for IC yield estimation, mainly via Bayesian regression models. Moreover, the yield models are extended to build extrinsic reliability models via a defect-growth concept.

1.1 Yield Modeling

To give a brief review on yield models based on statistical modeling of defect counts, we first define the term of fault probability, which is the probability of a random defect being a fault. This probability depends on the defect's size and location. Let the random variable S denote the random defect size with probability density function $f(s)$, and let $A_c(s)$ be the critical area of defect size s. The average critical area is then given by $A_c = \int_0^\infty A_c(s) f(s) ds$. The fault probability is defined as $\Phi = A_c/B$, where B is the total area on which a defect can fall. The conventional Poisson yield model assumes that defects are randomly distributed and the occurrence of a defect at any location is independent of the occurrence of any other defects. The probability that a device contains x defects is calculated by the Poisson distribution

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots, \tag{1}$$

where the random variable X denotes the nonnegative defect count on a randomly selected device and $\lambda > 0$ is the average number of defects on a device. The Poisson yield is the probability of zero faults on a device given by

$$Y = e^{-\Phi\lambda}, \tag{2}$$

where $\Phi\lambda$ is the average number of faults on a device. However, defects on semiconductor wafers tend to cluster [8] and as a result the observed defect counts usually exhibit over-dispersion with an excessive fraction of zeros that exceeds the Poisson probability of zero. For this reason, the Poisson yield model usually predicts a yield that is lower than the true yield [15]. To account for the over-dispersed defect counts, compound Poisson yield models have been developed. A compound Poisson yield model assumes that the defect density is a random variable instead of a constant as assumed by the Poisson yield model. Different defect density distributions lead to various compound Poisson yield models, e.g., the negative binomial (NB) model, Murphy's model, and Seed's model [15]. The widely used NB yield model is obtained by assuming that the defect density is a gamma random variable, and the random number of defects on a device follows the NB distribution

$$P(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\Gamma(x + 1)} \left(\frac{\lambda}{\lambda + \alpha}\right)^x \left(\frac{\alpha}{\lambda + \alpha}\right)^\alpha, \quad x = 0, 1, 2, \ldots,$$

where $\Gamma(\cdot)$ is the Gamma function and $\alpha > 0$ is the clustering coefficient. The limiting case of $\alpha \to \infty$ becomes the Poisson distribution. The NB yield is given by

$$Y = \left(\frac{\alpha}{\Phi\lambda + \alpha}\right)^\alpha = \left(1 + \frac{\Phi\lambda}{\alpha}\right)^{-\alpha}.$$

Besides the Poisson and NB distributions, other count distributions, e.g., the generalized Poisson (GP) distribution and generalized double-Poisson distribution, have been applied to yield modeling [20, 22]. The GP and NB distributions, in many situations, provide very close results [11]. The generalized double-Poisson distribution assumes that centers of defect clusters are randomly distributed and the defects within a cluster are also homogeneously distributions.

To consider the excess zeros in the clustered defect counts, yield models based on zero-inflated and hurdle-at-zero distributions have been proposed [22]. It has been found that zero-inflated and hurdle-at-zero distributions always produce maximum likelihood yield estimates that are equal to the observed yields. Recognizing that defect distribution exhibits spatial variation, spatial regression yield models have been developed. Bae et al. [1] proposed Poisson regression, NB regression, and zero-inflated Poisson (ZIP) regression yield models, including spatial locations of chips as covariates. Maximum likelihood method was employed to estimate model parameters and yield. Later, Yuan et al. [23] studied those three regression models and a new zero-inflated NB (ZINB) regression yield model in a hierarchical Bayesian modeling framework.

There have been other yield prediction and improvement approaches that utilize artificial intelligence or machine learning techniques to identify critical process variables influencing yield [3, 12, 17, 19, 21]. Additionally, Hwang et al. [9] recently proposed a yield model using the spatial nonhomogeneous Poisson process (NHPP) to model the spatial distribution of defects. Those yield modeling approaches are excluded from the discussion in this chapter.

1.2 Yield-Reliability Relation Modeling

Because defects are responsible for both yield losses and potential early reliability failures, there have been attempts to directly estimate the device reliability from the yield information. Note that the focus of the yield-reliability relation modeling is on the prediction of extrinsic reliability, which is determined by the manufacturing defects. Huston and Clarke [6] presented a yield-reliability relation by

$$R = Y^{A_R/A_Y},$$

where Y is the Poisson probability of zero yield defects, R is the Poisson probability of zero reliability defects, and A_R and A_Y are the reliability and yield critical areas, respectively. Kuper et al. [16] developed a different yield-reliability relation given by

$$R = (Y/\delta)^{D_R/D_Y},$$

where δ is a parameter for clustering effects and edge exclusions, and D_R and D_Y are, respectively, the reliability and yield defect densities. The above two definitions of reliability, however, are not time-dependent. Kim and Kuo [13] later proposed a time-dependent Poisson yield-reliability relation model given by

$$R(t) = Y^{c(t)}$$

where the time-dependent factor $c(t)$ is determined by the defect growth during $[0,\ t]$. That is, the defect size (or severity) increases with time under operation. When a defect's size exceeds a failure threshold, a reliability failure occurs. A similar gamma yield-reliability relation model based on the NB yield model was also developed [13, 14]. Hwang [7] developed a statistical defect-growth model and proposed an extrinsic reliability model based on the spatial NHPP yield model presented in [9]. Recently, Yuan et al. [24] described the defect-growth process by a random-coefficient degradation model and proposed new Poisson and NB yield-reliability relation models.

1.3 Organization of This Chapter

The remainder of this chapter is organized as follows. Section 2 presents Bayesian spatial regression yield models including the Poisson, NB, ZIP, and ZINB regression models and compares them using two examples. Section 3 presents new spatial regression yield-reliability relation models and illustrate those models using one example. Finally, Sect. 4 concludes this chapter.

2 Yield Modeling via Bayesian Spatial Regression of Defect Counts

This section presents the Bayesian spatial regression yield models based on the Poisson and NB distributions, including the Poisson regression, NB regression, ZIP regression, and ZINB regression models. Section 2.1 introduces those regression models for individual wafers, and Sect. 2.2 discusses hierarchical Bayesian regression models for multiple wafers.

2.1 Bayesian Spatial Regression Models for Individual Wafers

This section presents four Bayesian spatial regression models for defect counts on a single wafer. Suppose that a wafer is partitioned into n mutually exclusive chips. Let $\boldsymbol{z}_i$ and X_i, for $i = 1, 2, \ldots, n$, denote the spatial location of the ith chip's center on the wafer and random defect count in the ith chip, respectively. Denote $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ the observed defect count dataset on a wafer.

The Poisson regression model assumes that the random defect count X_i follows the Poisson distribution with a mean λ_i that is dependent on the location $\boldsymbol{z}_i$, that is,

$$P(X_i = x_i) = \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}, \quad x_i = 0, 1, 2, \ldots,$$

where λ_i is related to $\boldsymbol{z}_i$ via a generalized linear function $g(\lambda_i) = \boldsymbol{f}(\boldsymbol{z}_i)^T \boldsymbol{\beta}$ [1]. Herein $g(\cdot)$ denotes a link function and the natural logarithm link is selected for the positive Poisson mean parameter, i.e., $\log(\lambda_i) = \boldsymbol{f}(\boldsymbol{z}_i)^T \boldsymbol{\beta}$ [1]. $\boldsymbol{f}(\boldsymbol{z}_i)$ represents the vector of spatial covariates and $\boldsymbol{\beta}$ is the vector of regression coefficients. We choose $\boldsymbol{z}_i = (r_i, \phi_i)$, the polar coordinates of the ith chip center relative to the wafer center, and $\boldsymbol{f}(\boldsymbol{z}_i) = (1, \ r_i, \ \cos\phi_i, \ \sin\phi_i, \ r_i \cos\phi_i, \ r_i \sin\phi_i)^T$ [1]. To complete the Bayesian model specification, prior distributions are needed for the model parameters, which will be discussed at the end of this section.

The NB regression model computes the probability $P(X_i = x_i)$ according to the NB distribution

$$P(X_i = x_i) = \frac{\Gamma(\alpha + x_i)}{\Gamma(\alpha)\Gamma(x_i + 1)} \left(\frac{\lambda_i}{\lambda_i + \alpha}\right)^{x_i} \left(\frac{\alpha}{\lambda_i + \alpha}\right)^{\alpha}, \quad x_i = 0, 1, 2, \ldots,$$

where the mean parameter λ_i is again assumed to be dependent on the spatial covariates via $\log(\lambda_i) = \boldsymbol{f}(\boldsymbol{z}_i)^T \boldsymbol{\beta}$. The unknown parameters of the NB regression model include $\boldsymbol{\beta}$ and α.

The ZIP regression model is specified as

$$P(X_i = 0) = \pi_i + (1 - \pi_i)e^{-\lambda_i},$$
$$P(X_i = x_i) = (1 - \pi_i)\frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}, \quad x_i = 1, 2, \ldots,$$

where the mean parameter λ_i and probability of extra zeros π_i are related to the spatial covariates via $\log(\lambda_i) = \boldsymbol{f}(z_i)^T\boldsymbol{\beta}$ and $\text{logit}(\pi_i) = \boldsymbol{f}(z_i)^T\boldsymbol{\gamma}$, respectively.

The ZINB regression model is given by

$$P(X_i = 0) = \pi_i + (1 - \pi_i)\left(\frac{\alpha}{\lambda_i + \alpha}\right)^{\alpha},$$
$$P(X_i = x_i) = (1 - \pi_i)\frac{\Gamma(\alpha + x_i)}{\Gamma(\alpha)\Gamma(x_i + 1)}\left(\frac{\lambda_i}{\lambda_i + \alpha}\right)^{x_i}\left(\frac{\alpha}{\lambda_i + \alpha}\right)^{\alpha}, \quad x_i = 1, 2, \ldots,$$

where, again, $\log(\lambda_i)$ and $\text{logit}(\pi_i)$ are linear combinations of the spatial covariates $\boldsymbol{f}(z_i)$, i.e., $\log(\lambda_i) = \boldsymbol{f}(z_i)^T\boldsymbol{\beta}$ and $\text{logit}(\pi_i) = \boldsymbol{f}(z_i)^T\boldsymbol{\gamma}$.

Prior distributions are crucial components of Bayesian models. The prior distributions should accurately reflect one's prior knowledge on the model parameters. Non-informative prior distributions would be desired when there is a lack of prior knowledge. For the regression coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, we assume independent multivariate normal (MVN) prior distributions, i.e., $\boldsymbol{\beta} \sim MVN_\nu(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ and $\boldsymbol{\gamma} \sim MVN_\nu(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$, where $\nu = 6$ is the length of the regression coefficient vectors, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are, respectively, the MVN mean vector and variance-covariance matrix. The $MVN_\nu(\mathbf{0}_\nu, 10^6\boldsymbol{I}_\nu)$ distribution is a proper non-informative prior for regression coefficient vectors, where $\mathbf{0}_\nu$ and $\boldsymbol{I}_\nu$ are a zero vector of length ν and a $\nu \times \nu$ identity matrix, respectively.

For the positive NB dispersion parameter α, we choose a Gamma prior distribution $G(a, b)$, where the a and b are shape and scale parameters, respectively. If a non-informative prior is needed for the dispersion parameter, one can choose $G(0.001, 0.001)$ or $G(1, 0.001)$ [4].

Bayesian inference on the model parameters is based on the posterior distribution derived according to the Bayes' theorem

$$f(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}),$$

where $f(\boldsymbol{\theta}|\boldsymbol{x})$, $f(\boldsymbol{x}|\boldsymbol{\theta})$ and $f(\boldsymbol{\theta})$ denote the joint posterior distribution, likelihood function and joint prior distribution, respectively. Herein the vector $\boldsymbol{\theta}$ denotes a model's parameters. For example, $\boldsymbol{\theta} \equiv \boldsymbol{\beta}$ and $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \boldsymbol{\gamma}, \alpha)$ for the Poisson and ZINB regression models, respectively. The likelihood function $f(\boldsymbol{x}|\boldsymbol{\theta})$ of the observed data $\boldsymbol{x}$ for a given model is computed by

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} P(X_i = x_i|\boldsymbol{\theta}),$$

where $P(X_i = x_i|\boldsymbol{\theta})$ for different models have been presented above in this section.

The yield estimated by a model conditioning on its model parameters $\boldsymbol{\theta}$ is given by [1]

$$Y = \frac{1}{n} \sum_{i=1}^{n} P(X_i = 0|\boldsymbol{\theta}). \tag{3}$$

In principal, the posterior distribution of yield, denoted by $f(Y|\boldsymbol{x})$, can be derived from $f(\boldsymbol{\theta}|\boldsymbol{x})$ via transformations of random variables according to the relationship given by Eq. (3).

In theory, marginal posterior distributions for model parameters $f(\boldsymbol{\beta}|\boldsymbol{x})$, $f(\boldsymbol{\gamma}|\boldsymbol{x})$, and $f(\alpha|\boldsymbol{x})$ as well as for yield $f(Y|\boldsymbol{x})$ can be derived from $f(\boldsymbol{\theta}|\boldsymbol{x})$. However, the derivation of those marginal posterior distributions involves high-dimensional integrals. Instead, one may implement Markov chain Monte Carlo (MCMC) simulation based sampling algorithms, e.g., Gibbs sampling, to obtain random samples from those marginal posterior distributions [5]. Each iteration of Gibbs sampling draws a value of one parameter from its conditional posterior distribution conditioning on the latest values of all other parameters. When the number of iterations is large enough, the values drawn from one parameter's conditional posterior distribution can be regarded as a random sample generated from its marginal posterior distribution. In each iteration, a yield value can be calculated using the parameter values drawn in that iteration according to Eq. (3). After the completion of Gibbs sampling, all the calculated yield values form a random sample from $f(Y|\boldsymbol{x})$. Sample statistics can then be used to construct point and interval estimates.

2.2 *Hierarchical Bayesian Spatial Regression Models for Multiple Wafers*

This section extends the regression models presented in Sect. 2.1 to build hierarchical Bayesian regression models. We assume that defect count data on m wafers having identical pattern of n chips on each wafer are available and denote the defect count data by $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$, where the defect count data on the jth wafer $\boldsymbol{x}_j = \{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ with x_{ij} being the number of defects in the ith chip at location $\boldsymbol{z}_{ij}$ on the jth wafer, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. In addition, let the random variable X_{ij} denote the random defect count in the ith chip on the jth wafer.

We first illustrate how to construct a three-stage hierarchical Bayesian Poisson regression model. The first-stage model specifies a Poisson regression for each wafer as

$$P(X_{ij} = x_{ij}|\boldsymbol{\vartheta}_j) = \frac{e^{-\lambda_{ij}}\lambda_{ij}^{x_{ij}}}{x_{ij}!}, \quad \text{with } \log(\lambda_{ij}) = \boldsymbol{f}(z_{ij})\boldsymbol{\beta}_j,$$

for $x_{ij} = 0, 1, \ldots, i = 0, 1, \ldots, n$, and $j = 1, 2, \ldots, m$. Each individual wafer is assumed to have its own Poisson regression parameters $\boldsymbol{\vartheta}_j \equiv \boldsymbol{\beta}_j$.

In the second stage, the Poisson regression parameters of individual wafers are assumed to form a random sample from a multivariate distribution $f(\boldsymbol{\vartheta}|\boldsymbol{\varphi})$ with hyperparameters $\boldsymbol{\varphi}$, that is, $\boldsymbol{\vartheta}_j \sim f(\boldsymbol{\vartheta}|\boldsymbol{\varphi})$, for $j = 1, 2, \ldots, m$. The MVN distribution is usually employed in random-effects and hierarchical Bayesian models to describe unit-to-unit variations, that is,

$$\boldsymbol{\vartheta}_j \sim MVN_\nu(\boldsymbol{\mu}_{\boldsymbol{\vartheta}}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}), \quad j = 1, 2, \ldots, m.$$

Herein $\boldsymbol{\varphi} = (\boldsymbol{\mu}_{\boldsymbol{\vartheta}}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}})$.

The third stage specifies a joint prior distribution for the hyperparameters $\boldsymbol{\mu}_{\boldsymbol{\vartheta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$. A MVN prior distribution is assumed for the mean vector $\boldsymbol{\mu}_{\boldsymbol{\vartheta}}$, i.e., $\boldsymbol{\mu}_{\boldsymbol{\vartheta}} \sim MVN_\nu(\boldsymbol{u}, \boldsymbol{C})$, with $\boldsymbol{u}$ and $\boldsymbol{C}$ denoting the mean vector and variance-covariance matrix, respectively. A scaled-inverse-Wishart prior [2] is employed for the variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$. The $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$ is first decomposed as

$$\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}} = \boldsymbol{\Delta}\boldsymbol{Q}\boldsymbol{\Delta},$$

where $\boldsymbol{\Delta}$ is a diagonal matrix with positive diagonal terms denoted by $\delta_1, \delta_2, \ldots, \delta_\nu$, and $\boldsymbol{Q}$ is a symmetric positive-definite matrix. Next, independent inverse-Wishart (IW) and Gamma prior distributions are specified for the $\boldsymbol{Q}$ matrix and δ_k terms:

$$\boldsymbol{Q} \sim IW(\rho, \boldsymbol{W}) \text{ and } \delta_k \sim G(a_k, b_k), \ k = 1, 2, \ldots, \nu,$$

where ρ and $\boldsymbol{W}$ are the degrees-of-freedom and scale matrix of IW distribution, and a_k and b_k are shape and scale parameters of Gamma distribution. Especially, one can choose the $IW(\nu + 1, \boldsymbol{I}_\nu)$ and $G(0.001, 0.001)$ priors for $\boldsymbol{Q}$ and δ_k's, respectively, to form a noninformative prior for $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$ [2].

Three-stage hierarchical Bayesian NB, ZIP, and ZINB regression models can be similarly constructed. The first stage specifies a regression model for each individual wafer:

$$P(X_{ij} = x_{ij}|\boldsymbol{\vartheta}_j), \quad x_{ij} = 0, 1, 2, \ldots,$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

For the NB regression model

$$P(X_{ij} = x_{ij}|\boldsymbol{\vartheta}_j) = \frac{\Gamma(\alpha_j + x_{ij})}{\Gamma(\alpha_j)\Gamma(x_{ij}+1)} \left(\frac{\lambda_{ij}}{\lambda_{ij}+\alpha_j}\right)^{x_{ij}} \left(\frac{\alpha_j}{\lambda_{ij}+\alpha_j}\right)^{\alpha_j}, x_{ij} = 0, 1, 2, \ldots,$$

where $\log(\lambda_{ij}) = \boldsymbol{f}(\boldsymbol{z}_{ij})^T \boldsymbol{\beta}_j$ and $\boldsymbol{\vartheta}_j \equiv (\boldsymbol{\beta}_j,\ \log(\alpha_j))$ with $\nu = 7$.

For the ZIP regression model

$$P(X_{ij} = x_{ij}|\boldsymbol{\vartheta}_j) = \begin{cases} \pi_{ij} + (1-\pi_{ij})e^{-\lambda_{ij}}, & x_{ij} = 0, \\ (1-\pi_{ij})\dfrac{e^{-\lambda_{ij}}\lambda_{ij}^{x_{ij}}}{x_{ij}!}, & x_{ij} = 1, 2, \ldots, \end{cases}$$

where $\text{logit}(\pi_{ij}) = \boldsymbol{f}(\boldsymbol{z}_{ij})^T \boldsymbol{\gamma}_j$, $\log(\lambda_{ij}) = \boldsymbol{f}(\boldsymbol{z}_{ij})^T \boldsymbol{\beta}_j$, and $\boldsymbol{\vartheta}_j \equiv (\boldsymbol{\alpha}_j,\ \boldsymbol{\beta}_j)$ with $\nu = 12$.

For the ZINB regression model

$$P(X_{ij} = x_{ij}|\boldsymbol{\vartheta}_j) = \begin{cases} \pi_{ij} + (1-\pi_{ij})\left(\frac{\alpha_j}{\lambda_{ij}+\alpha_j}\right)^{\alpha_j}, & x_{ij} = 0, \\ (1-\pi_{ij})\frac{\Gamma(\alpha_j+x_{ij})}{\Gamma(\alpha_j)\Gamma(x_{ij}+1)}\left(\frac{\lambda_{ij}}{\lambda_{ij}+\alpha_j}\right)^{x_{ij}}\left(\frac{\alpha_j}{\lambda_{ij}+\alpha_j}\right)^{\alpha_j}, & x_{ij} = 1, 2, \ldots, \end{cases}$$

where $\text{logit}(\pi_{ij}) = \boldsymbol{f}(\boldsymbol{z}_{ij})^T \boldsymbol{\gamma}_j$, $\log(\lambda_{ij}) = \boldsymbol{f}(\boldsymbol{z}_{ij})^T \boldsymbol{\beta}_j$, $\boldsymbol{\vartheta}_j \equiv (\boldsymbol{\alpha}_j,\ \boldsymbol{\beta}_j,\ \log(\alpha_j))$, and $\nu = 13$.

The second stage describes the wafer-to-wafer variation by assuming $\boldsymbol{\vartheta}_j \sim MVN_\nu(\boldsymbol{\mu}_{\boldsymbol{\vartheta}},\ \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}})$, for $j = 1, 2, \ldots, m$. When constructing the NB, ZIP, and ZINB regression models, the previous study [23] assumed that $\boldsymbol{\beta}_j$, $\boldsymbol{\gamma}_j$, and $\log(\alpha_j)$ follow independent distributions (i.e., MVNs for $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_j$, and univariate normal for $\log(\alpha_j)$) in the second-stage model; while this study uses one joint MVN for all stage-one parameters $\boldsymbol{\vartheta}$.

Finally, the third stage specifies the MVN and scaled-inverse-Wishard priors for $\boldsymbol{\mu}_{\boldsymbol{\vartheta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}$, respectively. The joint posterior distribution of all model parameters can be derived as

$$\begin{aligned} &f(\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_m, \boldsymbol{\mu}_{\boldsymbol{\vartheta}}, \boldsymbol{Q}, \delta_1, \ldots, \delta_\nu|\boldsymbol{x}) \\ &\quad \propto \left[\prod_{j=1}^{m}\prod_{i=1}^{n} P(X_{ij} = x_{ij}|\boldsymbol{\vartheta}_j)\right] \times \exp\left[-\frac{\sum_{j=1}^{m}(\boldsymbol{\vartheta}_j - \boldsymbol{\mu}_{\boldsymbol{\vartheta}})^T \boldsymbol{\Delta}^{-1}\boldsymbol{Q}^{-1}\boldsymbol{\Delta}^{-1}(\boldsymbol{\vartheta}_j - \boldsymbol{\mu}_{\boldsymbol{\vartheta}})}{2}\right] \\ &\quad \times \exp\left[-\frac{(\boldsymbol{\mu}_{\boldsymbol{\vartheta}} - \boldsymbol{u})^T \boldsymbol{V}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\vartheta}} - \boldsymbol{u})}{2}\right] \times |\boldsymbol{Q}|^{-\frac{\nu+\rho+1}{2}} \exp\left[-\frac{tr(\boldsymbol{S}\boldsymbol{Q}^{-1})}{2}\right] \\ &\quad \times \prod_{k=1}^{\nu} \delta_k^{-a_k-1} e^{-b_k\delta_k}. \end{aligned}$$

The Gibbs sampling algorithm is employed to obtain random samples from the marginal posterior distributions $f(\boldsymbol{\mu}_{\boldsymbol{\vartheta}}|\boldsymbol{x})$, $f(\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}|\boldsymbol{x})$, $f(\boldsymbol{\vartheta}_j|\boldsymbol{x})$ and $f(Y_j|\boldsymbol{x})$, for $j = 1, 2, \ldots, m$, where Y_j is the yield of the jth wafer.

2.3 *Illustrative Examples*

This section uses two examples to compare the yield models discussed in Sect. 2. Example 1 involves defect count data on one wafer, and the individual Bayesian models are implemented and compared with the maximum likelihood method. Example 2 considers defect counts on three wafers. The maximum likelihood, individual Bayesian, and hierarchical Bayesian are applied to Example 2. For convenience, we assume that every defect is a fault in the examples, i.e., $\Phi = 1$.

2.3.1 Example 1

Figure 1 shows the defect-count data on a wafer map in Example 1. Tyagi and Bayoumi [20] applied the generalized double-Poisson yield model to analyze this data set. Bae et al. [1] later adopted this wafer map to demonstrate the Poisson regression, NB regression, and ZIP regression yield models using the maximum likelihood method. There are $n = 400$ chips with $n_0 = 317$ defect-free chips. The observed yield is $n_0/n = 0.7925$. The sample mean and variance of the defect counts are 0.625 and 2.731, respectively. The conventional Poisson yield is $e^{-0.625} = 0.5353$, which is significantly lower than 0.7925.

Table 1 lists the yield estimates by the Bayesian Poisson, NB, ZIP, and ZINB regression models presented in Sect. 2.1. Non-informative prior distributions are used, and the Gibbs sampling algorithm runs for 10^6 iterations. Sample means of the posterior samples produced by the Gibbs sampling algorithm are used as point estimates, and 95% Bayesian intervals are constructed using the 2.5th and 97.5th sample percentiles. Because defect counts on one wafer are available, only the individual Bayesian models in Sect. 2.1 are applicable here. Maximum likelihood estimates (MLEs) and 95% normal-approximation confidence intervals of the yield are also presented for the purpose of comparison. For a description of the maximum likelihood method, please refer to Yuan et al. [22] and Bae et al. [1].

Overall, the Bayesian and maximum likelihood methods produce similar yield estimates for the Poisson and ZIP regression models. However, the Bayesian method produces noticeably better yield estimation than the maximum likelihood method for the NB and ZINB regression models. The Poisson regression model significantly underestimates the yield, while the other three models have much higher the estimation accuracy.

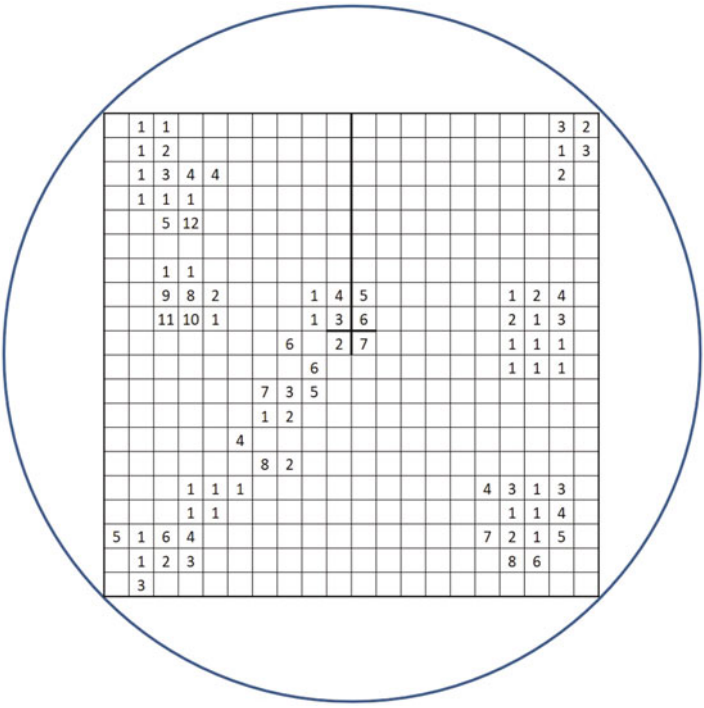


Fig. 1 Example 1 defect count data [20]

Table 1 Yield estimates for Example 1 (observed yield 0.7925)

Inference Method	Estimates	Regression models			
		Poisson	NB	ZIP	ZINB
Individual Bayesian	Posterior mean	0.5597	0.7925	0.7921	0.7921
	95% Bayesian interval	(0.5203, 0.5988)	(0.7524, 0.8300)	(0.7519, 0.8297)	(0.7532, 0.8287)
Maximum likelihood	MLE	0.5573	0.7906	0.7926	0.7894
	95% normal-appx. confidence interval	(0.5175, 0.5963)	(0.7483, 0.8274)	(0.7507, 0.8290	(0.7487, 0.8250)

2.3.2 Example 2

Figure 2 depicts the defect-count data on three wafer maps provided a DRAM manufacturer. Yuan et al. [23] analyzed these data sets using hierarchical Bayesian Poisson, NB, ZIP, and ZINB regression models.

Table 2 compares the yield estimates by the maximum likelihood method, individual Bayesian inference (Sect. 2.1), and hierarchical Bayesian approach (Sect. 2.2). 95% normal-approximate confidence intervals for the maximum likelihood method and 95% Bayesian intervals for the Bayesian methods are shown

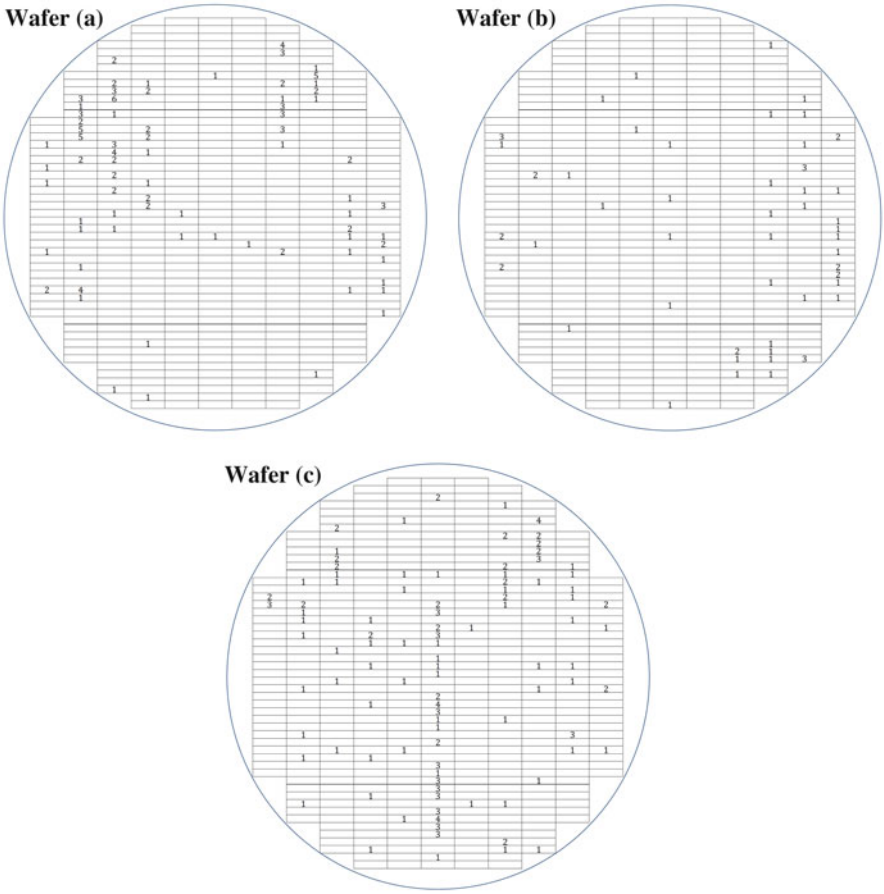


Fig. 2 Example 2 defect count data on (**a**–**c**) wafer maps [23]

in parentheses. Table 3 compares the relative bias of yield estimation, which is defined as [23]

$$\frac{\text{estimated yield} - \text{observed yield}}{\text{observed yield}} \times 100\%.$$

Table 4 compares the average of absolute relative bias of yield estimation by the three methods.

For the Poisson, NB, and ZINB regression models, the individual Bayesian method provides more accurate yield estimates than the maximum likelihood method; while the maximum likelihood method has better yield estimates than the individual Bayesian method for the ZIP regression models. Between the hierarchical Bayesian and individual Bayesian methods, the hierarchical Bayesian method tend

Table 2 Yield estimates for Example 2

Wafer	Observed yield	Method	Regression models			
			Poisson	NB	ZIP	ZINB
(a)	0.8436	Maximum likelihood	0.7629 (0.7285, 0.7943)	0.8419 (0.8070, 0.8715)	0.8336 (0.7993, 0.8629)	0.8389 (0.8048, 0.8681)
		Individual Bayesian	0.7643 (0.7311, 0.7962)	0.8433 (0.8104, 0.8736)	0.8300 (0.7930, 0.8588)	0.8425 (0.8101, 0.8727)
		Hierarchical Bayesian	0.7623 (0.7286, 0.7945)	0.8311 (0.7985, 0.8619)	0.8347 (0.8031, 0.8642)	0.8366 (0.8061, 0.8649)
(b)	0.8985	Maximum likelihood	0.8831 (0.8535, 0.9073)	0.8975 (0.8672, 0.9214)	0.8894 (0.8608, 0.9127)	0.9007 (0.8712, 0.9241)
		Individual Bayesian	0.8839 (0.8561, 0.9090)	0.8974 (0.8689, 0.9228)	0.8846 (0.8570, 0.9094)	0.8975 (0.8694, 0.9227)
		Hierarchical Bayesian	0.8809 (0.8523, 0.9069)	0.8991 (0.8730, 0.9230)	0.8950 (0.8683, 0.9190)	0.8993 (0.8731, 0.9224)
(c)	0.7928	Maximum likelihood	0.7251 (0.6871, 0.7601)	0.7905 (0.7516, 0.8248)	0.7848 (0.7462, 0.8189)	0.7898 (0.7519, 0.8232)
		Individual Bayesian	0.7266 (0.6897, 0.7621)	0.7924 (0.7550, 0.8274)	0.7815 (0.7417, 0.8153)	0.7920 (0.7548, 0.8282)
		Hierarchical Bayesian	0.7270 (0.6898, 0.7626)	0.7988 (0.7648, 0.8313)	0.7928 (0.7561, 0.8268)	0.7948 (0.7603, 0.8270)

Table 3 Relative bias of yield estimation for Example 2 (%)

Wafer	Method	Regression models			
		Poisson	NB	ZIP	ZINB
(a)	Maximum likelihood	−9.57	−0.20	−1.19	−0.56
	Individual Bayesian	−9.40	−0.04	−1.61	−0.13
	Hierarchical Bayesian	−9.64	−1.48	−1.06	−0.83
(b)	Maximum likelihood	−1.71	−0.11	−1.01	0.24
	Individual Bayesian	−1.62	−0.12	−1.55	−0.11
	Hierarchical Bayesian	−1.96	0.07	−0.39	0.09
(c)	Maximum likelihood	−8.54	−0.29	−1.01	−0.38
	Individual Bayesian	−8.35	−0.05	−1.43	−0.10
	Hierarchical Bayesian	−8.30	0.76	0.00	0.25

Table 4 Average of absolute relative bias of yield estimation for Example 2 (%)

Method	Regression models			
	Poisson	NB	ZIP	ZINB
Maximum likelihood	6.61	0.20	1.07	0.39
Individual Bayesian	6.46	0.07	1.53	0.11
Hierarchical Bayesian	6.63	0.77	0.48	0.39

to produce less accurate yield estimates than the individual Bayesian method, with the ZIP regression model as an exception. This is probably due to a so-called "shrink effect" known for the hierarchical Bayesian approach [18], which tends to bring the yield estimates of the three wafers close to each other. Among the four hierarchical Bayesian models, the ZINB regression model tends to produce the most accurate yield estimates.

3 A Yield-Reliability Relation Modeling Approach

This section discusses a spatial regression yield-reliability relation modeling approach based on the defect-growth concept. An example based on Example 1 in Sect. 2.3.1 is used to illustrate the proposed approach. We focus on modeling the extrinsic reliability, which is entirely determined by the manufacturing defects [14].

3.1 Spatial Regression Yield-Reliability Relation Models

Consider a defect whose size at time 0 (i.e., at the end of manufacturing process) is denoted by s^0. Under a set of operating and environmental conditions, the defect size (or severity) grows and may potentially cause a reliability failure [10]. The expected defect size at time t, denoted by s^t, is described by a general degradation-path model

$$s^t = h(t; s^0, \boldsymbol{\eta}),$$

where $h(\cdot)$ denotes the expected degradation-path function with parameter vector $\boldsymbol{\eta}$ to describe the defect-growth process. The defect is not a yield defect if it is not located in the critical area $A_c(s^0)$ at time 0. Similarly, it will not cause a reliability failure by time t if it is not located in the critical area $A_c(s^t)$ determined by its size at time t. One example of analytical defect-growth models is

$$s^t = s^0 + \frac{V}{G} \log \frac{t}{\tau} \tag{4}$$

for time-dependent dielectric breakdown [7, 14], where V is the voltage stress across the dielectric film, and G and τ are materials related constants.

Based on the defect-growth concept, we next define a time-dependent fault probability Φ^t. The fault probability at time zero is defined as

$$\Phi^0 = \int_0^\infty \frac{A_c(s^0)}{B} f(s^0) ds^0.$$

Similarly, the fault probability at time t, defined as the probability that an arbitrary defect will cause device failure by time t, is calculated according to

$$\Phi^t = \int_0^\infty \frac{A_c[h(t; s^0, \boldsymbol{\eta})]}{B} f(s^0) ds^0.$$

If a non-regression yield model with the defect-count distribution $P(X = x)$ is assumed, the yield and extrinsic reliability are given by

$$Y = \sum_{x=0}^{\infty} (1 - \Phi^0)^x P(X = x)$$

and

$$R_E(t) = \frac{1}{1 - P(X = 0)} \sum_{x=1}^{\infty} \left(\frac{1 - \Phi^t}{1 - \Phi^0} \right)^x P(X = x),$$

respectively [24]. The yield is simply the probability that no defects cause device failure at time 0, and the reliability at time t is the probability that no defects cause device failure at time t for those devices with defects that have not failed at time 0. It can be shown that $R_E(0) = 1$ and $R_E(\infty) = 0$.

Assuming a regression yield model with a defect-count distribution $P(X_i = x_i)$, $i = 1, 2, \ldots, n$ and $x_i = 0, 1, \ldots$, the extrinsic reliability is then computed by

$$R_E(t) = \frac{1}{n}\sum_{i=1}^{n} R_{E,i}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{1-P(X_i=0)}\sum_{x_i=1}^{\infty}\left(\frac{1-\Phi^t}{1-\Phi^0}\right)_i^{x} P(X_i=x_i)\right].$$

For example, when $P(X=x)$ is given by the Poisson distribution (Eq. (1)) it can be shown that

$$R_E(t) = \frac{e^{-\lambda}}{1-e^{-\lambda}}\cdot\frac{1-e^{-\lambda\kappa(t)}}{e^{-\lambda\kappa(t)}},$$

where $\kappa(t) = (1-\Phi^t)/(1-\Phi^0)$. When $P(X=x)$ is given by the NB distribution (Eq. (2)), the extrinsic reliability is then given by

$$R_E(t) = \frac{\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha}}{1-\left(\frac{\alpha}{\alpha+\lambda}\right)^{\alpha}}\cdot\frac{1-\left(1-\frac{\lambda\kappa(t)}{\alpha+\lambda}\right)^{\alpha}}{\left(1-\frac{\lambda\kappa(t)}{\alpha+\lambda}\right)^{\alpha}}.$$

Similarly, it can be shown that the Poisson regression and ZIP regression reliability models have the same form of

$$R_E(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{e^{-\lambda_i}}{1-e^{-\lambda_i}}\cdot\frac{1-e^{-\lambda_i\kappa(t)}}{e^{-\lambda_i\kappa(t)}}\right],$$

where $\log(\lambda_i) = \boldsymbol{f}(\boldsymbol{z}_i)^T\boldsymbol{\beta}$. Although the two regression models have the same form of $R_E(t)$ their different $\boldsymbol{\beta}$ estimates result in different $R_E(t)$ estimates. $R_E(t)$ functions for the NB regression and ZINB regression are given by

$$R_E(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left(\frac{\alpha}{\alpha+\lambda_i}\right)^{\alpha}}{1-\left(\frac{\alpha}{\alpha+\lambda_i}\right)^{\alpha}}\cdot\frac{1-\left(1-\frac{\lambda_i\kappa(t)}{\alpha+\lambda_i}\right)^{\alpha}}{\left(1-\frac{\lambda_i\kappa(t)}{\alpha+\lambda_i}\right)^{\alpha}}\right].$$

3.2 An Illustrative Example

The Example 1 in Sect. 2.3.1 is employed here to illustrate the proposed spatial regression reliability models. We assume that device failures are caused by the time-dependent dielectric breakdown and adopt the defect-growth model given by Eq. (4). The dielectric thickness is denoted by w. In addition, we assume the following

Table 5 Parameters in the illustrative example [14]

Defect size distribution			Defect-growth model		Thickness	Voltage
q	p	s_0	τ	G	w	V
3	1	10 Å	1×10^{-11} s	3.5 V/Å	32 Å	1.5 V

Table 6 Reliability prediction by various yield models

	"Expected"	Poisson regression	NB regression	ZIP regression	ZINB regression
R (1 year)	0.536	0.716	0.571	0.508	0.567
R (10 years)	0.478	0.664	0.516	0.443	0.512

widely cited defect size distribution [14, 15]

$$f(s) = \begin{cases} cs_0^{-q-1}s^q, & 0 \le s \le s_0 \\ cs_0^{p-1}s^{-p}, & s_0 \le s \le \infty \end{cases}$$

where $p \neq 1$, $q > 0$, $c = (q+1)(p-1)/(p+q)$, and s_0 is a critical size of the defect with the highest probability of occurrence. Table 5 lists the parameter values adopted from Kim et al. [14].

When a defect's size exceeds the thickness of dielectric film, a breakdown failure occurs. Thus

$$\Phi^0 = P(s^0 > w) = \int_w^\infty f(s)ds$$

and

$$\Phi^t = P(s^t > w) = P[s^0 > w - V/G \cdot \log(t/\tau)].$$

Table 6 compares the reliability predictions by various yield models. As a comparison, we define the "expected" reliability based on the observed defect counts as

$$R_E^*(t) = \frac{1}{n_p}\sum_{i=1}^{n} \phi(t)^{x_i} I_{\{x_i>0\}},$$

where $n_p = \sum_{i=1}^{n} I_{\{x_i>0\}}$, i.e., the number of chips with defects. The NB, ZIP, and ZINB regression models provide more accurate reliability predictions than the Poisson regression model. This is consistent with the result obtained for yield prediction in Sect. 2.3.1. Note that this comparison only confirms that those three models produce more accurate prediction of the defect count distribution than the Poisson regression model. In order to validate the defect-growth model given by Eq. (4), the predicted reliability needs to be compared with reliability obtained from conventional reliability life tests, which will deserve more study and experimentation in the future.

4 Conclusion

This chapter discussed some latest development in IC yield models via spatial regression modeling of clustered defect counts. It focused on the Bayesian modeling and analysis approach. The spatial regression yield models were extended to create new yield-reliability relation models for predicting extrinsic device reliability. The reliability defined by those models is consistent with conventional definition of reliability. However, more experiments and studies are needed to establish and validate defect-growth processes for different failure mechanisms of ICs.

Acknowledgements This work is partially supported by the National Science Foundation projects 1633500 and 1633580. The work of S. J. Bae was supported by the Technological Innovation R&D Program (S2480669) funded by the Small and Medium Business Administration(SMBA), Korea.

References

1. Bae, S. J., Hwang, J. Y., & Kuo, W. (2007). Yield prediction via spatial modeling of clustered defect counts across a wafer map. *IIE Transactions, 39*, 1073–1083.
2. Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviation and correlations, with application to shrinkage. *Stat Sinica, 10*, 1281–1311.
3. Chen, A., & Hong, A. (2010). Sample-efficient regression true (SERT) for semiconductor yield loss analysis. *IEEE Transactions on Semiconductor Manufacturing, 23*, 358–369.
4. Congdon, P. (1992). *Bayesian statistical modeling*. New York: Wiley.
5. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis*, 2nd Edn. Boca Raton, FL: Chapman & Hall/CRC.
6. Huston, H. H., & Clarke, C. P. (1992). Reliability defect detection and screening during processing—theory and implementation. In *Proceedings of the International Reliability Physics Symposium* (pp. 268–275). Piscataway: IEEE.
7. Hwang, J. Y. (2004). Spatial stochastic processes for yield and reliability management with applications to nano electronics. PhD dissertation. Texas A&M University, College Station.
8. Hwang, J. Y., & Kuo, W. (2007). Model-based clustering for integrated circuit yield enhancement. *European Journal of Operational Research, 178*, 143–153.
9. Hwang, J. Y., Kuo, W., & Ha, C. (2011). Modeling of integrated circuit yield using a spatial nonhomogeneous Poisson process. *IEEE Transactions on Semiconductor Manufacturing, 24*, 377–384.
10. Jensen, F. (1991). Yield, quality and reliability—a natural correlation? In: R. H. Matthews (Ed.), *Reliability '91*. Boca Raton, FL: CRC Press.
11. Joe, H., & Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparision with negative binomial. *The Biochemical Journal, 47*, 219–229.
12. Kang, B. S., Lee, J. H., Shin, C. K., Yu, S. J., & Park, S. C. (1998). Hybrid machine learning system for integrated yield management in semiconductor manufacturing. *Expert Systems with Applications, 15*, 123–132.
13. Kim, T., & Kuo, W. (1999). Modeling manufacturing yield and reliability. *IEEE Transactions on Semiconductor Manufacturing, 12*, 485–492.
14. Kim, K. O., Kuo, W., & Luo, W. (2004). A relation model of gate oxide yield and reliability. *Microelectronics and Reliability, 44*, 425–434.
15. Kuo, W., Chien, W. T. K., & Kim, T. (1998). *Reliability, yield, and stress burn-in*. Boston, MA: Kluwer Academic Publishers.

16. Kuper, F., Van der Pol, J., Ooms, E., Johnson, T., Wijburg, R., Koster, W., & Johnston, D. (1996). Relation between yield and reliability of integrated circuits: Experimental results and application to continuous early failure rate deduction program. In *Proceedings of the International Reliability Physics Symposium* (pp. 17–21). Piscataway: IEEE.
17. Lee, J. H., & Ha, S. H. (2009). Recognizing yield patterns through hybrid applications of machine learning techniques. *Information Sciences, 179*, 844–850.
18. León, R. V., Ramachandran, R., Ashby, A. J., & Thyagarajan, J. (2007). Bayesian modeling of accelerated life tests with random effects. *Journal of Quality Technology, 39*, 3–16.
19. Li, T. S., Huang, C. L., & Wu, Z. Y. (2006). Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system. *Journal of Intelligent Manufacturing, 17*, 335–361.
20. Tyagi, A., & Bayoumi, M. A. (1992). Defect clustering viewed through generalized Poisson distribution. *IEEE Transactions on Semiconductor Manufacturing, 5*, 196–206.
21. Weiss, S. M., Baseman, R. J., Tipu, F., Collins, C. N., Davies, W. A., Singh, R., & Hopkins, J. W. (2009). Rule-based data mining for yield improvement in semiconductor manufacturing. *Applied Intelligence, 33*, 318–329.
22. Yuan, T., Bae. S. J., & Yue, K. (2018). Statistical models of over-dispersed spatial defects for predicting integrated circuits yield. *IEEE Transactions on Reliability*. Submitted.
23. Yuan, T., Ramadan, S. Z., & Bae, S. J. (2011). Yield prediction for integrated circuits manufacturing through hierarchical Bayesian modeling of spatial defects. *IEEE Transactions on Reliability, 60*, 729–741.
24. Yuan, T., Zhu, X., & Kuo, Y. (2017). A yield-reliability relation modeling approach based on random effects degradation models. In *Proceedings of the Asia Pacific Conference of the Prognostics and Health Management Society* (pp. 791–794).

Index

Y. Lio et al. (eds.), *Statistical Quality Technologies*, ICSA Book Series in Statistics,
https://doi.org/10.1007/978-3-030-20709-0