ISSN: 3060-7876

SMPS

Stochastic Models in Probability and Statistics

Vol. 1, No. 2, 2024

In the Name of God

Stochastic Models in Probability and Statistics (SMPS)

Volume 1, Issue 2 July 2024

ISSN-Print: —- ISSN-Online: 3060-7876

Publisher: Faculty of Mathematical Sciences, Ferdowsi University of Mashhad
Published by: Ferdowsi University of Mashhad Press
Printing Method: Electronic

Address:
SMPS Journal, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad,
P.O. Box 1159, Mashhad 91775, Iran.

Tel.: +98-51-38806222 **Fax:** +98-51-38807358
E-mail: smps@um.ac.ir
Website: https://smps.um.ac.ir

©2024 All rights reserved. Stochastic Models in Probability and Statistics

Editorial Board

Editor-in-Chief: Jafar Ahmadi
Ferdowsi University of Mashhad, Iran
Email: ahmadi-j@um.ac.ir

Executive Editor: Mahdi Doostparast
Ferdowsi University of Mashhad, Iran
Email: doustparast@um.ac.ir

Associate Editor Members

- **Mohammad Amini**
  Ferdowsi University of Mashhad, Iran
  Email: m-amini@um.ac.ir
- **Majid Asadi**
  University of Isfahan, Iran
  Email: m.asadi@sci.ui.ac.ir
- **Akbar Asgharzadeh**
  University of Mazandaran, Iran
  Email: a.asgharzadeh@umz.ac.ir
- **Narayanaswamy Balakrishnan**
  McMaster University, Canada
  Email: bala@mcmaster.com
- **Erhard Cramer**
  RWTH Aachen University, Germany
  Email: erhard.cramer@rwth-aachen.de
- **Ismihan Bayramoglu**
  Izmir University of Economics, Turkey
  Email: bayramoglu@ieu.edu.tr
- **Anna Dembinska**
  Warsaw Technical University, Poland
  Email: dembinska@mini.pw.edu.pl
- **Antonio Di Crescenzo**
  Università di Salerno, Italy
  Email: adicrescenzo@unisa.it
- **Ali Dolati**
  Yazd University, Iran
  Email: adolati@yazd.ac.ir
- **Serkan Eryilmaz**
  Atilim University, Turkey
  Email: serkan.eryilmaz@atilim.edu.tr
- **Firoozeh Haghighi**
  University of Tehran, Iran
  Email: haghighi@khayam.ut.ac.ir
- **Taizhong Hu**
  University of Science and Technology of China
  Email: thu@ustc.edu.cn
- **Udo Kamps**
  RWTH Aachen University, Germany
  Email: kamps@isw.rwth-aachen.de
- **Bahaedin Khaledi**
  Razi University, Iran & UNC, USA
  Email: bkhaledi@hotmail.com
- **Subhash Kochar**
  Portland State University, USA
  Email: kochar@pdx.edu
- **Xiaohu Li**
  Stevens Institute of Technology, USA
  Email: xli82@stevens.edu
- **Maria Longobardi**
  Università di Napoli Federico II, Italy
  Email: maria.longobardi@unina.it
- **G. R. Mohtashami Borzadaran**
  Ferdowsi University of Mashhad, Iran
  Email: grmohtashami@um.ac.ir
- **Asok Nanda**
  IISER, India
  Email: asok@iiserkol.ac.in
- **Jorge Navarro**
  University of Murcia, Spain
  Email: jorgenav@um.es

- **Hon Keung Tony Ng**
  Bentley University, USA
  Email: tng@bentley.edu
- **Sangun Park**
  Yonsei University, South Korea
  Email: sangun@yonsei.ac.kr
- **Franco Pellerey**
  Politecnico di Torino, Italy
  Email: pellerey@polito.it
- **Mohammad Z. Raqab**
  University of Jordan
  Email: mraqab@ju.edu.jo
- **Alfonso Suarez Llorens**
  Universidad de Cádiz, Spain
  Email: suarez@uca.es
- **Peng Zhao**
  Jiangsu Normal University, China
  Email: zhaop@jsnu.edu.cn

Managing Editor: Mostafa Razmkhah
Ferdowsi University of Mashhad, Iran
Email: razmkhah_m@um.ac.ir

Technical Editor: Hadi Jabbari Nooghabi
Ferdowsi University of Mashhad, Iran
Email: jabbarinh@um.ac.ir

English Text Editor: Zohreh Vasagh
Ferdowsi University of Mashhad, Iran
Email: z_vasagh@yahoo.com

This journal is published under the auspices of Ferdowsi University of Mashhad

---

We would like to acknowledge the help of Miss Narjes Khatoon Zohorian in the preparation of this issue.

Letter from the Editor-in-Chief

It is my great pleasure to present this issue of Stochastic Models in Probability and Statistics (SMPS). As an international, peer-reviewed, open-access journal, SMPS remains committed to advancing both the theoretical foundations and practical applications of stochastic modeling, probability, and statistics—with a particular focus on reliability and related disciplines.
SMPS publishes two issues annually and is proudly supported by the Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad.

The journal's mission is to promote the development and dissemination of innovative methodologies and ideas that bridge stochastic modeling in probability and statistics with real-world applications. In this spirit, SMPS welcomes high-quality contributions across a broad range of topics, including reliability theory, lifetime data analysis, stochastic modeling in industrial and systems engineering, operations research, statistical methods in information theory, and dependence modeling.

On behalf of the editorial board, I would like to express my sincere gratitude to all authors for their valuable scholarly contributions, to the reviewers for their thoughtful and constructive evaluations, and to the editorial team for their unwavering dedication and professionalism. The continued success of SMPS is made possible by the collective efforts of this dynamic community of researchers and practitioners who share a common goal: advancing the field of stochastic modeling and statistical reliability analysis.

Finally, I warmly invite readers and contributors around the world to support SMPS by submitting original research, citing published work, and sharing new ideas that help shape and expand the journal's vision.

Jafar Ahmadi
Editor-in-Chief
Stochastic Models in Probability and Statistics (SMPS)

Contents

Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 119–142. Research Paper

On the use of sampling costs in quality control to select an estimator for the variance and standard deviation

Ayda Amniattalab[†] and J. B. G. Frenk[‡*]

[†]*Industrial Engineering Department, Faculty of Engineering and Natural Sciences, 34956 Istanbul, Sabancı University, Istanbul, Turkey*
[‡]*Industrial Engineering Department, Faculty of Engineering, Gebze Technical University, 41400, Kocaeli, Gebze, Turkey*
Email(s): ayda@sabanciuniv.edu, frenk@gtu.edu.tr

Abstract. In *R* and *S*-charts in quality control applied to a normal population one mostly uses a linear transformation of the range or the sample deviation as unbiased estimators of the unknown process standard deviation. In this paper, we propose related statistics as alternative estimators of the unknown standard deviation and variance having a smaller mean squared error. At the same time, we give a theoretical explanation for the rules of thumb recommended in quality control which unbiased estimators to use. Since obtaining samples from different independent subgroups is costly, we propose a mathematical model for selecting these samples to satisfy the budget constraint. The used estimator for the variance or standard deviation has a minimal mean squared error within a certain class of estimators. It is shown that selecting the whole sample from one particular chosen subgroup is a good strategy for linear sampling costs.

Keywords: Mean squared error; Pooled statistics; Quality control; *R*-estimator; *S*-estimator.

1 Introduction.

In *R* and *S*-charts (see Section 4.7 and the Appendix of Chapter 4 of Ryan (2011) or Montgomery (2009)) applied to a normal population, one uses a linear transformation of the range or the sample standard deviation as unbiased estimators of the unknown process standard deviation. Within quality control, there is a tradition to use only unbiased estimators. However, if one uses as a quality measure of an estimator its mean squared error, there is no need to restrict

*Corresponding author
Received: 22 February 2024 / Accepted: 23 July 2024
DOI: 10.22067/smps.2024.45525

© 2024 Ferdowsi University of Mashhad

to unbiased estimators. In this paper we will consider a class of estimators containing both biased and unbiased estimators from which it is easy to derive a (biased) estimator having a minimal mean squared error among this class. A similar set-up was also followed in Woodall and Montgomery (2000), Mahmoud et al. (2010) and Vardeman (1999). Our analysis extends their approach. Since the theoretical properties of the R and S estimators in R and S-charts are already known for a long time most of the proofs of these results can be found in papers written more than thirty years ago. Also these results coincide with classical results in statistics and as a consequence of this no papers appeared on this topic recently. After the introduction of these estimators in quality control the focus shifted on how to apply these estimators to different industrial applications. For a recent overview on the practical use of these so-called univariate charts to all kind of different settings like healthcare and environmental control the reader is referred to Suman and Prajapati (2018), Zwetsloot et al. (2023) and Arciszewski (2023).

As in statistics it is mostly assumed in quality control models that gathering data does not involve any costs. However, in practice data gathering might be costly and subject to budget restrictions. Due to these restrictions we need to decide beforehand from which available resources we will generate our data. These available resources should be selected in such a way that our constructed estimator will satisfy the quality measure imposed on an estimator. In this paper this can be either an estimator within the class of unbiased estimators having a minimal mean squared error or a biased estimator within a given class of estimators satisfying this property. In this paper we will propose a simple static optimization problem dealing with the selection of the available resources from which the data are generated.

The outline of this paper is now as follows. In the first subsection of Section 2, we present a compact overview of the classical unbiased estimators of the standard deviation used in S and R-charts for single groups. Computing the mean squared error of these estimators we also confirm theoretically some empirical rules of thumb applied in the literature which estimator to use. In the past these rules of thumb were advised without any theoretical explanation.

In the second subsection of Section 2 we propose related biased estimators having a smaller mean squared error than the classical used estimators for both the standard deviation and the variance. At the same time we extend these results for a normal population consisting of either single or multiple independent subgroups to a population having a location-scale family of distributions.

In section 3 of this paper we formulate in the presence of a budget constraint on collecting data a simple static model constructing an estimator for both the variance and standard deviation having the smallest mean squared error. In this set-up we are allowed to sample from different independent subgroups each having different sampling costs. Due to the separable structure of the mean squared error objective function this optimization problem can be solved by a dynamic programming procedure. As a special case this model confirms the intuitively clear result that for linear sampling costs it is a good strategy to generate the whole sample from a single subgroup.

Finally in Section 4 we list our findings in this paper and give in the Appendix the proofs of our main results. Observe that the first subsection of Section 2 can be regarded as an elementary and unifying note on the property of the classical estimators used in R and S-charts discussed in standard textbooks on quality control.

2 On classical and alternative estimators of the standard deviation and variance used in quality control.

In the first subsection we will discuss the main properties of the classical unbiased estimators for the standard deviation used in R and S-charts. The result in Lemma 1 about the normalisation constant used in a S-chart seems to be new. In the second subsection we propose for single and multiple subgroups some alternative biased estimators for both the standard deviation and variance having a smaller mean squared error than the classical unbiased estimators. We will also analyze in this subsection the differences between these estimators and compute the ratios of their mean squared error. In the same subsection encountering multiple independent subgroups of data having different sample sizes the relation of these estimators to so-called pooled estimators (see page 360 of Irwin and Miller (2004) or Arnold (1990)) is also discussed.

2.1 On the main properties of the classical unbiased estimators in R and S-charts.

In this section we first give for completeness an overview of the estimators used in quality control to estimate the unknown standard deviation for normal populations. Similar results can be found in Mahmoud et al. (2010) and Woodall and Montgomery (2000). Introducing in a normal sample $\mathbf{X}^\top = (X_1,...,X_n)$, the largest order statistic $X_{n:n} := \max\{X_1,...,X_n\}$ and the smallest order statistic $X_{1:n} := \min\{X_1,...,X_n\}$, one uses in R-charts within quality control as an unbiased estimator of the unknown standard deviation $\sigma > 0$ in a sample of size n the statistic

$$T_R(\mathbf{X}) := \theta_n R_n(\mathbf{X}), \tag{1}$$

with R_n the so-called sample range given by

$$R_n(\mathbf{X}) := X_{n:n} - X_{1:n},$$

(see page 229 of Montgomery (2009)). Of course, it is assumed this sample is taken from a system in control. To compute the constant θ_n in this unbiased estimator, we observe that

$$R_n(\mathbf{X}) \stackrel{d}{=} \sigma R_{n,s}(\mathbf{Y}), R_{n,s}(\mathbf{Y}) := Y_{n:n} - Y_{1:n}, \tag{2}$$

with $Z_1 \stackrel{d}{=} Z_2$ meaning the random variables Z_1 and Z_2 have the same distribution and $R_{n,s}(\mathbf{Y}) := Y_{n:n} - Y_{1:n}$ the range of the sample $\mathbf{Y}^\top = (Y_1,...,Y_n)$ from a standard normal population. As shown in the remainder of this paper it is easy to generalize the above observations for a normal population to location-scale families of distributions. By the symmetry of a standard normal distribution around zero, we obtain by relations (1) and (2) that

$$\sigma = \mathbb{E}(T_R(\mathbf{X})) = \mathbb{E}(\theta_n \sigma R_{n,s}(\mathbf{Y})) = \theta_n \sigma \mathbb{E}(R_{n,s}(\mathbf{Y})) = 2\sigma\theta_n \mathbb{E}(Y_{n:n}),$$

with $Y_{n:n}$ the largest order statistic of a sample of size n coming from a standard normal distribution. This shows

$$\theta_n = (\mathbb{E}(R_{n,s}(\mathbf{Y})))^{-1} = (2\mathbb{E}(Y_{n:n}))^{-1}. \tag{3}$$

Introducing the mean squared error of an estimator T of some unknown constant b defined by

$$\mathrm{MSE}(T) := \mathbb{E}((T(\mathbf{X}) - b)^2),$$

and the bias of an estimator given by

$$\mathrm{bias}(T) := \mathbb{E}(T(\mathbf{X})) - b,$$

yields by relation (2) and (3) that the mean squared error of the unbiased estimator $T_R(\mathbf{X}) = \theta_n R_n(\mathbf{X})$ is given by

$$\mathrm{MSE}(T_R) = \mathrm{Var}(\theta_n \sigma R_{n,s}(\mathbf{Y})) = \frac{\sigma^2 \mathrm{Var}(R_{n,s}(\mathbf{Y}))}{(\mathbb{E}(R_{n,s}(\mathbf{Y})))^2} = \frac{\sigma^2 \mathrm{Var}(R_{n,s}(\mathbf{Y}))}{4(\mathbb{E}(Y_{n:n}))^2}. \tag{4}$$

Since in general, it seems impossible to derive an elementary expression for the above ratio divided by σ^2 (Arnold et al. (2008)), tables are used within quality control (Montgomery (2009), Harter (1960)). In our computational experiments, we generate sample paths of the stochastic range process $\mathbf{R} = \{R_{n,s} : n \in \mathbb{N}\}$ and use Monte Carlo simulation to compute both $\mathrm{Var}(R_{n,s}(\mathbf{Y}))$ and $\mathbb{E}(R_{n,s}(\mathbf{Y}))$ for $n \leq 50$. The density of the random variable $R_{n,s}(\mathbf{Y})$ is given by David (1970)

$$f_{R_{n,s}(\mathbf{Y})}(t) = n(n-1)\int_{-\infty}^{\infty} \varphi(u)(\Phi(u+t) - \Phi(u))^{n-2}\varphi(u+t)du,\ t \geq 0,$$

with φ the standard normal density and Φ the standard normal cdf. Using relation (2) this implies that the density of the unbiased estimator $T_R(\mathbf{X}) = \theta_n R_n(\mathbf{X})$ is given by

$$f_{T_R(\mathbf{X})}(t) = \frac{1}{\sigma\theta_n} f_{R_{n,s}(\mathbf{Y})}\left(\frac{t}{\sigma\theta_n}\right), t \geq 0,$$

and so it is possible to give a plot of the density of the cdf of this estimator. On page 112 of Montgomery (2009), as a rule of thumb, the unbiased estimator T_R is recommended for samples from a normal population of size n with $n \leq 6$ instead of the unbiased estimator T_S used in a S-chart. In an S-chart, the unbiased estimator $T_S(\mathbf{X}) := \gamma_n S_n(\mathbf{X}), n \geq 2$ is applied with S_n^2 denoting the well-known unbiased sample variance estimator of σ^2 given by

$$S_n^2(\mathbf{X}) := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2, \overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i,$$

with $\overline{X}_n$ representing the sample mean. The constant γ_n is selected in such a way that the estimator $\gamma_n S_n$ is an unbiased estimator of the unknown standard deviation σ. By the properties of a normal distribution it is well known (Casella and Berger (2002)) that

$$S_n(\mathbf{X}) \overset{d}{=} \sigma S_{n,c}(\mathbf{Y}) \text{ and } (n-1)S_{n,c}^2(\mathbf{Y}) \overset{d}{=} Z, \tag{5}$$

with $S_{n,c}$ the sample standard deviation of a sample $\mathbf{Y}^\top = (Y_1, ..., Y_n)$ from a standard normal population and Z a random variable having a chi-square distribution with $n-1$ degrees of freedom. This yields

$$\sigma = \mathbb{E}(\gamma_n S_n(\mathbf{X})) = \mathbb{E}(\gamma_n \sigma S_{n,c}(\mathbf{Y})) = \frac{\gamma_n \sigma}{\sqrt[2]{n-1}}\mathbb{E}(\sqrt[2]{Z}) = \gamma_n \sigma \sqrt[2]{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$

with $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$ the well known gamma function. Hence we obtain for every $n\geq 2$ that

$$\gamma_n=(\mathbb{E}(S_{n,c}(\mathbf{Y})))^{-1}=\sqrt[2]{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \tag{6}$$

Using $S_{n,c}^2$ is an unbiased estimator of the known variance of a standard normal population and so $\mathbb{E}(S_{n,c}^2(\mathbf{Y}))=1$, it follows by relation (5) and (6) that the mean squared error of the estimator $\gamma_n S_n, n\geq 2$ is given by

$$\mathrm{MSE}(\gamma_n S_n)=\mathrm{Var}(\gamma_n\sigma S_{n,c}(\mathbf{Y}))=\sigma^2\frac{\mathrm{Var}(S_{n,c}(\mathbf{Y}))}{(\mathbb{E}(S_{n,c}(\mathbf{Y})))^2}=\sigma^2(\gamma_n^2-1). \tag{7}$$

Since MSE $(\gamma_n S_n)\geq 0$, we obtain for every $n\geq 2$ that $\gamma_n\geq 1$. In the next result, we list some global properties of the sequence $\gamma_n, n\geq 2$. Before discussing these properties, we introduce the following definition.

Definition 1. *The non-negative sequence δ_n, $n\geq 2$ is called log-convex if the first order differences $\Delta\delta_n:=\delta_{n+1}-\delta_n$, $n\geq 2$ are increasing.*

For the sequence $\gamma_n:=(\mathbb{E}(S_{n,c}(\mathbf{Y}))^{-1}$, $n\geq 2$ one can show the following result. Its proof is given in the Appendix.

Lemma 1. *The sequence $\gamma_n, n\geq 2$ is decreasing and log-convex and it satisfies $\gamma_2=\sqrt[2]{\frac{\pi}{2}}$ and for every $n\geq 2$ the first order non-linear recurrence relation*

$$\gamma_{n+1}=\gamma_n^{-1}\sqrt[2]{\frac{n}{n-1}}. \tag{8}$$

Moreover $\lim_{n\uparrow\infty} n(\gamma_n-1)=4^{-1}$ *and for every* $n\geq 3$

$$1\leq\left(\frac{n}{n-1}\right)^{\frac{1}{4}}\leq\gamma_n\leq\left(\frac{n-1}{n-2}\right)^{\frac{1}{4}}. \tag{9}$$

By relation (8), it is easy and numerically stable to compute the constants γ_n, $n\geq 2$. Hence it is not necessary to use tables for the sequence γ_n, $n\in\mathbb{N}$, $n\geq 2$. Since

$$\frac{(n-1)S_n^2(\mathbf{X})}{\sigma^2}\stackrel{d}{=}Z,$$

with Z a random variable having a chi-square distribution with $n-1$ degrees of freedom (Casella and Berger (2002)) having density $f_{\mathbf{Z}}(z)=\frac{e^{-\frac{z}{2}}z^{\frac{n-1}{2}-1}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}$ this yields

$$T_S(\mathbf{X})=\gamma_n S_n(\mathbf{X})\stackrel{d}{=}\frac{\gamma_n\sigma\sqrt[2]{Z}}{\sqrt[2]{n-1}}. \tag{10}$$

Since the density of the non-negative random variable $\sqrt[2]{Z}$ with Z having a chi-square distribution having $n-1$ degrees of freedom is given by

$$f_{\sqrt[2]{Z}}(z)=\frac{z^{\frac{n-1}{2}}e^{-\frac{z^2}{2}}}{2^{\frac{n-3}{2}},\Gamma(\frac{n-1}{2})},\quad z>0.$$

This yields by relation (10) that the density $f_{T_S}(t)$, $t > 0$ of the estimator $T_S(\mathbf{X}) = \gamma_n S_n(\mathbf{X})$ equals

$$f_{T_S(\mathbf{X})}(t) = \frac{\sqrt[2]{n-1}}{\gamma_n \sigma} f_{\sqrt[2]{Z}}\left(\frac{\sqrt[2]{n-1}t}{\gamma_n \sigma}\right) = \frac{\sigma \sqrt[2]{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} f_{\sqrt[2]{Z}}\left(\frac{\sigma \sqrt[2]{2}\Gamma(\frac{n}{2})t}{\Gamma(\frac{n-1}{2})}\right),$$

and so it is possible to give a plot of the density of this estimator. In Figure 1, we give a plot of the functions $n \to \mathrm{MSE}(\gamma_n S_{n,c}) = \gamma_n^2 - 1$ and $n \to \mathrm{MSE}(\theta_n R_{n,c})$ for $2 \leq n \leq 50$. As this figure shows, comparing the mean squared error of both the R and S estimators defined in relations (4) and (7), it is always better to use the estimator $T_S = \gamma_n S_n$ instead of the estimator $T_R = \theta_n R_n$ for $n \geq 7$ while there is a slight preference for the S-estimator when $n \leq 6$. This confirms the rule of thumb given in Montgomery (2009) or Vardeman (1999) and the recommendation of always using the S-estimator in Mahmoud et al. (2010). An additional advantage of the S-estimator is that the constant γ_n can be numerically computed in a stable way applying relation (8) without making use of tables, while θ_n needs to be computed by using tables or (if not available) by simulation or numerical integration. This means computing θ_n can be less accurate.

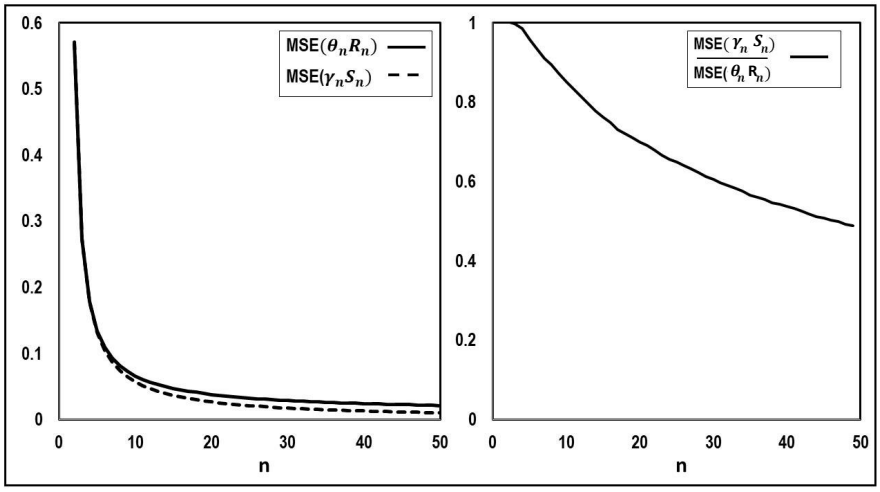


Figure 1: Plot of the mean squared errors of R and S estimators (left panel) for $\sigma = 1$ and their ratio (right panel).

In the next subsection we introduce related unbiased estimators having a smaller mean squared error.

2.2 On some alternative biased estimators for the standard deviation and the variance

It is well known within statistics that biased estimators of a unknown parameter having a smaller mean squared error (Woodall and Montgomery (2000)) can under certain conditions be constructed using a linear transformation of the unbiased estimator of the same parameter. The construction of these alternative estimators for both the range estimator and sample standard

deviation estimator used in R and S-charts is discussed in the following result. Its proof is given in the Appendix.

Lemma 2. *Let $k>0$ and $\mathbf{X}^\top=(X_1,....,X_n), n\geq 2$ be a random vector of size n satisfying*

$$X_i \stackrel{d}{=} a+bY_i, 1\leq i\leq n,$$

with unknown location parameter $a\in\mathbb{R}$ and scale parameter $b>0$ and the random vector $\mathbf{Y}^\top=(Y_1,...,Y_n)$ consisting of independent and identically distributed random variables has a known cdf F having a finite variance. If the statistic $T:\mathbb{R}^n\to\mathbb{R}$ is an estimator of the unknown scale parameter $b>0$ and

$$T(\mathbf{X}) \stackrel{d}{=} bf(\mathbf{Y}).$$

for some function $f:\mathbb{R}^n\to\mathbb{R}$ satisfying $\mathbb{E}(f^k(\mathbf{Y}))>0$ and $\mathbb{E}(f^{2k}(\mathbf{Y}))$ finite, then among the class of statistics $\alpha T^k, \alpha\in\mathbb{R}$, used as an estimator of the unknown parameter b^k the estimator $\alpha^\star T^k$ with

$$\alpha^\star=\frac{\mathbb{E}(f^k(\mathbf{Y}))}{\mathbb{E}(f^{2k}(\mathbf{Y}))},$$

is an estimator of b^k with the smallest mean squared error. Its bias is given by

$$bias(\alpha^\star T^k)=\frac{-b^k\,Var(f^k(\mathbf{Y}))}{\mathbb{E}(f^{2k}(\mathbf{Y}))},$$

and its mean squared error by

$$MSE\left(\alpha^\star T^k\right)=-bias(\alpha^\star T^k)b^k.$$

Since $\mathrm{Var}(f^k(\mathbf{Y}))\geq 0$, we obtain by Lemma 2 that on average for every $k\in\mathbb{N}$ we underestimate the unknown parameter b^k using the estimator $\alpha^\star T^k$ for b^k. The result in Lemma 2 can also be applied to a non-normal population. Since we know the cdf of the random variable Y_1 it is possible by simulation to determine the value of $\alpha^\star$. Observe in a normal population it follows for the standard sample deviation estimator that the constant $\alpha^\star$ is given by an elementary expression. As shown in Corollary 3 this constant is given by γ_n^{-1} with γ_n listed in relation (6). As an example of a non-normal population we mention a nonnegative sample $\mathbf{X}^\top=(X_1,...,X_n)$ satisfying

$$X_i \stackrel{d}{=} aY_i^b, a>0,\ b>0, \tag{11}$$

with Y_i, $i=1,...,n$, nonnegative independent and identically distributed random variables having a known cdf F and the random variables $\ln(Y_i)$ having a finite variance. Well known examples covering this case are Y_1 exponentially distributed with parameter 1. In this case the random variable X_i has a Weibull distribution. Another example is given by $\mathbf{Y}_i$ has a gamma distribution with scale parameter 1 and known shape parameter $\alpha>0$. In this case the random variable X_i has a so-called generalised gamma distribution (Abbaszadehpeivasti and Frenk (2023)). Since by relation (11) it follows

$$\ln(X_i) \stackrel{d}{=} \ln(a)+b\ln(Y_i), \tag{12}$$

we can easily apply Lemma 2 replacing in the used statistic the random variable X_i by the random variable $\ln(X_i)$. In case the nonnegative independent random variables X_i, $i = 1,2,..,n$ are log-normal distributed, it follows by definition that the random variables $\ln(Y_i)$ are standard normal distributed. Another example is given by

$$X_i \stackrel{d}{=} (a+bY_i)^{\beta}, a > 0,\ b > 0,$$

with $\beta > 0$ known and Y_i, $i = 1,...,n$ a sequence of non-negative and identically distributed random variables having a known cdf F and finite variance. In this case it follows that

$$X_i^{\beta^{-1}} \stackrel{d}{=} a + bY_i,$$

and we replace in the used statistic the random variable X_i by the random variable $X_i^{\beta^{-1}}$. The above transformations resemble the Box-Cox transformations used in statistical Process Control transforming non-normal charts into charts which are approximately normal distributed (see section 3 of Figueiredo and Gomes (2006)). If we compare the mean squared error of the unbiased estimator $\alpha_0 T^k$, $\alpha_0 := \mathbb{E}(f^k(\mathbf{Y}))^{-1}$ of the parameter b^k with the minimum mean squared error estimator within the class of estimators αT^k, $\alpha \in \mathbb{R}$ the next corollary follows immediately.

Corollary 1. *If the conditions of Lemma 2 hold and the unbiased estimator $\alpha_0 T^k$ of b^k with $\alpha_0 = \mathbb{E}(f^k(\mathbf{Y}))^{-1}$ is considered, then*

$$\frac{MSE(\alpha^\star T^k)}{MSE(\alpha_0 T^k)} = \frac{\mathbb{E}(f^k(\mathbf{Y}))^2}{\mathbb{E}(f^{2k}(\mathbf{Y}))} \leq 1.$$

Proof. It follows by relation (32) that

$$\mathrm{MSE}(\alpha_0 T^k) = b^{2k} p_k(\alpha_0) = \frac{b^{2k}\mathrm{Var}(f^k(\mathbf{Y}))}{\mathbb{E}(f^k(\mathbf{Y}))^2}.$$

Applying Lemma 2 implies the desired result. □

We now apply Lemma 2 (for $k = 1$) to the classical unbiased estimators used in a R and S-chart for a normal population. The next result is an easy application of Lemma 2.

Corollary 2. *If $\mathbf{X}^\top = (X_1,....,X_n)$, $n \geq 2$ is a random sample of size n from a normal population then the estimator $T_R(\alpha^\star) := \alpha^\star R_n$ of the standard deviation $\sigma > 0$ with*

$$\alpha^\star = \frac{\mathbb{E}(R_{n,s}(\mathbf{Y}))}{\mathbb{E}(R_{n,s}^2(\mathbf{Y}))},$$

and $\mathbf{Y}^\top = (Y_1,...,Y_n)$ a random sample from a standard normal population satisfies

$$\mathbb{E}(\alpha^\star R_n(\mathbf{X})) = \frac{\sigma\mathbb{E}(R_{n,s}(\mathbf{Y}))^2}{\mathbb{E}(R_{n,s}^2(\mathbf{Y}))}.$$

Among the class of estimators $T_R(\alpha) := \alpha R_n, \alpha > 0$ of the standard deviation $\sigma > 0$ in a R-chart applied to a normal population the biased estimator $T_R(\alpha^\star)$ has the smallest mean squared error and this is given by

$$MSE(\alpha^\star R_n) = \sigma^2 \frac{Var(R_{n,s}(\mathbf{Y}))}{E(R_{n,s}(\mathbf{Y})^2)}.$$

Proof. Since in a normal population we know that

$$(X_1,...,X_n) = (\mu + \sigma Y_1,...,\mu + \sigma Y_n),$$

with $\mathbf{Y}^\top = (Y_1,...,Y_n)$ consisting of independent and standard normal distributed random variables we obtain

$$R_n(\mathbf{X}) = \max\{\mu + \sigma Y_1,...,\mu + \sigma Y_n\} - \min\{\mu + \sigma Y_1,...,\mu + \sigma Y_n\} = \sigma R_{n,s}(\mathbf{Y}).$$

Since $f(\mathbf{Y}) = Y_{n:n} - Y_{1:n} = R_{n,s}(\mathbf{Y})$ and $\mathbb{E}(R_{n,s}(\mathbf{Y})) > 0$ for $n \geq 2$ we may apply Lemma 2 substituting $k = 1$. This shows the result. □

As already observed, one uses in a S-chart in quality control the unbiased estimator $T_S = \gamma_n S_n$ with γ_n listed in relation (6). The next result is again an application of Lemma 2 for $k = 1$.

Corollary 3. *If $\mathbf{X}^\top = (X_1,....,X_n)$, $n \geq 2$ is a random sample of size n from a normal population then the estimator $T_S(\alpha^{\star\star}) = \alpha^{\star\star} S_n$ of the standard deviation $\sigma > 0$ with*

$$\alpha^{\star\star} = \gamma_n^{-1} = \sqrt[2]{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},$$

satisfies

$$\mathbb{E}(\alpha^{\star\star} S_n(\mathbf{X})) = \frac{2}{n-1} \frac{\Gamma^2(\frac{n}{2})\sigma}{\Gamma^2(\frac{n-1}{2})} = \gamma_n^{-2}\sigma.$$

Among the class of estimators $T_S(\alpha) = \alpha S_n$, $\alpha > 0$ of the standard deviation $\sigma > 0$ in a normal population the biased estimator $T_S(\alpha^{\star\star})$ has the smallest mean squared error and this is given by

$$MSE(\alpha^{\star\star} S_n) = \sigma^2 \left(1 - \frac{2\Gamma^2(\frac{n}{2})}{(n-1)\Gamma^2(\frac{n-1}{2})}\right) = \sigma^2(1 - \gamma_n^{-2}). \tag{13}$$

Proof. Again using $(X_1,...,X_n) = (\mu + \sigma Y_1,...,\mu + \sigma Y_n)$ with $\mathbf{Y}^\top = (Y_1,...,Y_n)$ consisting of independent and standard normal distributed random variables we obtain $S_n(\mathbf{X}) = \sigma S_{n,c}(\mathbf{Y})$ and so

$$f(\mathbf{Y}) := S_{n,c}(\mathbf{Y}) = \sqrt[2]{\frac{\sum_{i=1}^n (Y_i - \overline{Y}_n)^2}{n-1}}.$$

Since $\mathbb{E}(f(\mathbf{Y})) > 0$ for every $n \geq 2$, the conditions of Lemma 2 are satisfied for $k = 1$ and this shows the result observing $\mathbb{E}\left(f^2(\mathbf{Y})\right) = 1$ and

$$E(f(\mathbf{Y})) = \frac{1}{\sqrt[2]{n-1}} \mathbb{E}\left(\sqrt[2]{Z}\right) = \sqrt[2]{\frac{2}{n-1}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}.$$

with the random variable Z having a chi-square distribution with $n-1$ degrees of freedom. □

Observe the above results also hold for any sequence of independent and identically distributed random variables $\mathbf{X}_i$, $i = 1, ..., n$, for which the normalized random variable $Y_i = b^{-1}(X_i - a)$ has a known cdf. If this holds we cannot compute the optimal α but need to approximate this value by simulation. By relations (7) and (13) we obtain for a normal population

$$\gamma_n^2 \mathrm{MSE}(\alpha^{\star\star} S_n) = \mathrm{MSE}(\gamma_n S_n).$$

Since by relation (9) we know $(\frac{n}{n-1})^{\frac{1}{2}} \leq \gamma_n^2 \leq (\frac{n-1}{n-2})^{\frac{1}{2}}, n \geq 3$, this implies that both the optimal biased and unbiased estimators of the standard variance in a normal population have almost the same mean squared error for already relatively small values of the sample size. Observe the mean squared error also depends linearly on the unknown variance. Hence, if it is suspected that the variance is large and the mean squared error is regarded as a more important quality measure of an estimator than whether an estimator is biased or unbiased, one should apply the biased estimator $\alpha^{\star\star} S_n$. In Figure 2, we now list the mean squared error of the most important three different estimators excluding the classical range estimator for $\sigma = 1$. As expected, the estimator $\alpha^{\star\star} S_n$ has the smallest mean squared error.

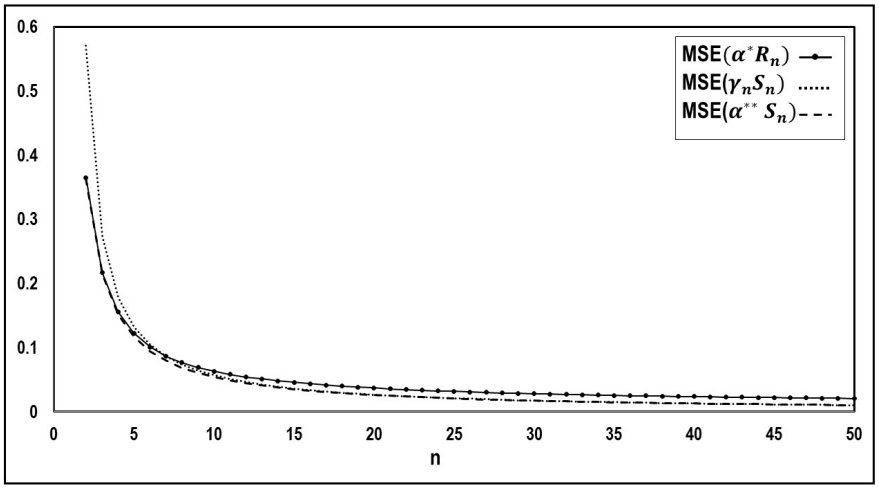


Figure 2: Plot of the mean squared errors of the three estimators of the standard deviation.

In case we like to give an estimation of the variance we identify in the next lemma the best possible estimator with respect to the mean squared error within the class of linear transformations of the sample variance. The next lemma is actually a special case of Lemma 2 for $k = 2$. The proof is listed in the Appendix.

Lemma 3. *If $\mathbf{X}^\top = (X_1, ..., X_n), n \geq 2$, is a random sample of size n from a population satisfying*

$$X_i \stackrel{d}{=} \mu + \sigma Y_i, \tag{14}$$

having unknown expectation μ and standard deviation σ and the random vector $\mathbf{Y}^\top = (Y_1, ..., Y_n)$ consists of independent and identically distributed random variables having a known cdf F and

a finite 4*th moment, then the estimator* $T_S^2(\alpha^{\star\star\star}) = \alpha^{\star\star\star} S_n^2$ *of the variance* $\sigma^2 > 0$ *with*

$$\alpha^{\star\star\star} = \frac{n}{\mathbb{E}(Y_1^4) - \frac{n-3}{n-1} + n},$$

satisfies

$$\mathbb{E}(\alpha^{\star\star\star} S_n^2) = \frac{n\sigma^2}{\mathbb{E}(Y_1^4) - \frac{n-3}{n-1} + n}.$$

Among the class of estimators $T_S^2(\alpha) = \alpha S_n^2$, $\alpha > 0$ *of the variance* σ^2 *the biased estimator* $T_S^2(\alpha^{\star\star\star})$ *has the smallest mean squared error and this is given by*

$$MSE(\alpha^{\star\star\star} S_n^2) = b^4(1 - \alpha^{\star\star\star}) = \frac{b^4((n-1)\mathbb{E}(Y_1^4) - n + 3)}{(n-1)\mathbb{E}(Y_1^4) - n + 3 + n^2}.$$

Suppose we consider M independent samples $\mathbf{X}_m = (X_{1,m}, ..., X_{n_m,m})$ of different or equal sizes $n_m \geq 2$, $m = 1, ..., M$ from location-scale families having possibly different unknown location parameters but the same unknown scale parameter. This set-up often occurs in quality control. Typically, an initial series of independent samples or independent subgroups is used to estimate the mean and standard deviation of a process. During this initial phase, the process should be in control. These estimations are then used to produce control limits in R and S-charts. Since the means might be different in each of the samples we consider for the estimation of the scale parameter the following class of statistics

$$T(\mathbf{X}_1, ..., \mathbf{X}_M) = \sum_{m=1}^{M} \alpha_m T_m^k(\mathbf{X}_m), \alpha_m \in \mathbb{R}, \tag{15}$$

with $T_m : \mathbb{R}^{n_m} \to \mathbb{R}$ the statistic used to estimate the scale parameter in sample subgroup $m = 1, ..., M$. This means we use a weighted combination of estimators for the estimation of the unknown scale parameter. One can now show the following generalization of Lemma 2. This result is also mentioned in Woodall and Montgomery (2000) without proof for the special case $k = 1$ and the normal population applying the sample standard deviation statistic to each subgroup. For completeness its elementary proof is listed in the Appendix.

Lemma 4. *Let* $k > 0$ *and for* $m = 1, ..., M$ *the random vectors*

$$\mathbf{X}_m = (X_{1,m}, ..., X_{n_m,m}),$$

$n_m \geq 2$ *are independent random samples of size* $n_m, m = 1, ..., M$, *satisfying*

$$X_{i,m} \stackrel{d}{=} a_m + bY_{i,m}, \tag{16}$$

for every $1 \leq i \leq n_m$ *with the random vector* $\mathbf{Y}_m = (Y_{1,m}, ..., Y_{n_m,m})$ *consisting of independent and identically distributed random variables having a known cdf* $F_m, m = 1, ..., M$. *If for* $m = 1, ..., M$ *the statistic* $T_m : \mathbb{R}^{n_m} \to \mathbb{R}$ *is an estimator of the unknown scale parameter* b *and*

$$T_m(\mathbf{X}_m) \stackrel{d}{=} bf_m(\mathbf{Y}_m), \tag{17}$$

for some function $f_m : \mathbb{R}^{n_m} \to \mathbb{R}$ *satisfying* $\mathbb{E}(f_m^k(\mathbf{Y}_m)) > 0$ *and* $\mathbb{E}(f_m^{2k}(\mathbf{Y}_m))$ *is finite then among the class of statistics* $\sum_{m=1}^{M} \alpha_m T_m^k, \alpha_m \in \mathbb{R}$ *used as an estimator of the parameter* b^k *the estimator* $\sum_{m=1}^{M} \alpha_m^\star T_m^k$ *with*

$$\alpha_m^\star = \left(1 + \sum_{m=1}^{M} \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m))\right)^2}{Var\left(f_m^k(\mathbf{Y}_m)\right)}\right)^{-1} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{Var(f_m^k(\mathbf{Y}_m))},$$

is an estimator of b^k *having the smallest mean squared error among this class. Its bias equals*

$$bias\left(\sum_{m=1}^{M} \alpha_m^\star T_m^k\right) = -b^k \left(1 + \sum_{m=1}^{M} \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m))\right)^2}{Var(f_m^k(\mathbf{Y}_m))}\right)^{-1},$$

and its mean squared error

$$MSE\left(\sum_{m=1}^{M} \alpha_m^\star T_m^k\right) = -bias\left(\sum_{m=1}^{M} \alpha_m^\star T_m^k\right) b^k.$$

If we restrict ourselves to the class of unbiased estimators within the class of estimators $T = \sum_{k=1}^{K} \alpha_m T_m^k, \alpha_k \in \mathbb{R}$ and select that unbiased estimator with the smallest mean squared error one can easily show the following result. Its proof is listed in the Appendix.

Lemma 5. *Under the same conditions as in Lemma 4 the unbiased estimator of* b^k *with the smallest mean squared error among the class of estimators* $T = \sum_{m=1}^{M} \alpha_m T_m^k, \alpha_m \in \mathbb{R}$ *is given by* $\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k$ *with*

$$\alpha_m^{\star\star} = \left(\sum_{m=1}^{M} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))^2}{Var(f_m^k(\mathbf{Y}_m))}\right)^{-1} \frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{Var(f_m^k(\mathbf{Y}_m))}.$$

Its mean squared error is given by

$$MSE\left(\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k\right) = b^{2k} \left(\sum_{m=1}^{M} \frac{\left(\mathbb{E}\left(f_m^k(\mathbf{Y}_m)\right)\right)^2}{Var\left(f_m^k(\mathbf{Y}_m)\right)}\right)^{-1}.$$

The next result is a generalization of Corollary 1 for independent samples.

Lemma 6. *If the conditions of Lemma 4 hold and* $\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k$ *is the unbiased minimum mean squared error estimator of* b^k *and* $\sum_{m=1}^{M} \alpha_m^\star T_m^k$ *the biased minimum mean squared error estimator of* b^k *then*

$$\frac{MSE\left(\sum_{m=1}^{M} \alpha_m^\star T_m^k\right)}{MSE\left(\sum_{m=1}^{M} \alpha_m^{\star\star} T_m^k\right)} = \frac{\sum_{m=1}^{M} \frac{(\mathbb{E}(f_m^k(\mathbf{Y}_m)))^2}{Var(f_m^k(\mathbf{Y}_m))}}{1 + \sum_{m=1}^{M} \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m))\right)^2}{Var(f_m^k(\mathbf{Y}_m))}}.$$

If the independent random variables $Y_i^{(m)}$, $1 \le m \le M$, $1 \le i \le m$ mentioned in Lemma 4 and 5 have a standard normal distribution then for the estimation of the standard deviation $\sigma > 0$ we apply Lemma 4 with $k = 1$ and $f_m(\mathbf{Y}_m) = S_{n_m,c}(\mathbf{Y}_m)$. By relation (6) it follows that

$$\mathbb{E}(f_m(\mathbf{Y}_m)) = \frac{1}{\gamma_{n_m}}, \mathrm{Var}(f_m(\mathbf{Y}_m) = 1 - \frac{1}{\gamma_{n_m}^2}. \tag{18}$$

By Lemma 4 for $k=1$ it follows using relation (18) that the weights $\alpha_m^\star$, $m=1,...,M$ of the minimum mean squared error estimator of the standard deviation within the class $\sum_{m=1}^M \alpha_m S_{n_m}$, $\alpha_m \in \mathbb{R}$ are given by

$$\alpha_m^\star = \left(1+\sum_{m=1}^M \frac{1}{\gamma_{n_m}^2-1}\right)^{-1} \frac{\gamma_{n_m}}{\gamma_{n_m}^2-1}, 1 \le m \le M. \tag{19}$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\sum_{m=1}^M \alpha_m^\star S_{n_m}\right) = \sigma^2 \left(1+\sum_{m=1}^M \frac{1}{\gamma_{n_m}^2-1}\right)^{-1}. \tag{20}$$

Applying Lemma 5 and relation (18) it follows that the weights $\alpha_m^{\star\star}$, $m=1,...,M$ of the minimum mean squared error unbiased estimator of the standard deviation within the class $\sum_{m=1}^M \alpha_m S_{n_m}$, $\alpha_m \in \mathbb{R}$ equal

$$\alpha_m^{\star\star} = \left(\sum_{m=1}^M \frac{1}{\gamma_{n_m}^2-1}\right)^{-1} \frac{1}{\gamma_{n_m}^2-1}, 1 \le m \le M \tag{21}$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\sum_{m=1}^M \alpha_m^{\star\star} S_{n_m}\right) = \sigma^2 \left(\sum_{m=1}^M \frac{1}{\gamma_{n_m}^2-1}\right)^{-1}. \tag{22}$$

This shows for a normal population that

$$\frac{\mathrm{MSE}\left(\sum_{m=1}^M \alpha_m^\star S_{n_m}\right)}{\mathrm{MSE}\left(\sum_{m=1}^M \alpha_m^{\star\star} S_{n_m}\right)} = \frac{\sum_{m=1}^M \frac{1}{\gamma_{n_m}^2-1}}{1+\sum_{m=1}^M \frac{1}{\gamma_{n_m}^2-1}}. \tag{23}$$

In order to illustrate the relations we consider the following example.

Example 1. Devor et al. (1992) provides on page 165 a data set made up of measurements of an engine block's cylinder bore diameter. The inside of the cylinder bore was measured after the boring operation and the units of measurement are 1/10,000 of an inch. Approximately every 30 minutes, samples of size $n=5$ are collected. The first 35 samples are displayed in Table 1. The precise dimensions are in the range of 3.5205, 3.5202, 3.5204, and so forth. The final three numbers in the measurements are provided by the entries in Table 1. Since we assume that the above non-negative measurements follow the multiplicative model (see relation (11)) satisfying $\ln(X_i)$ is normally distributed we need to compute first as shown in relation (12) the sample $\ln(X_{ij})$, $i=1,...,35$, $j=1,...,5$ before evaluating the value of the estimator of b. For this data set every subgroup $m=1,...,35$ has sample size $n_m=5$ and so

$$\gamma_{n_m} = \gamma_5 = \sqrt[2]{\frac{5-1}{2}} \frac{\Gamma\left(\frac{5-1}{2}\right)}{\Gamma\left(\frac{5}{2}\right)} = 1.0638.$$

Applying relations (20) and (22) we obtain using Excel that the mean squared error of the biased estimator of the parameter b is given by

$$\mathrm{MSE}\left(\sum_{m=1}^M \alpha_m^\star S_{n_m}\right) = 0.0000883 \times b^2,$$

Table 1: Cylinder diameter data.

Sample i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}	Sample i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}
1	205	202	204	207	205	19	207	206	194	197	201
2	202	196	201	198	202	20	200	204	198	199	199
3	201	202	199	197	196	21	203	200	204	199	200
4	205	203	196	201	197	22	196	203	197	201	194
5	199	196	201	200	195	23	197	199	203	200	196
6	203	198	192	217	196	24	201	197	196	199	207
7	202	202	198	203	202	25	204	196	201	199	197
8	197	196	196	200	204	26	206	206	199	200	203
9	199	200	204	196	202	27	204	203	199	199	197
10	202	196	204	195	197	28	199	201	201	194	200
11	205	204	202	208	205	29	201	196	197	204	200
12	200	201	199	200	201	30	203	206	201	196	201
13	205	196	201	197	198	31	203	197	199	197	201
14	202	199	200	198	200	32	197	194	199	200	199
15	200	200	201	205	201	33	200	201	200	197	200
16	201	187	209	202	200	34	199	199	201	201	201
17	202	202	204	198	203	35	200	204	197	197	199
18	201	198	204	201	201						

and the mean squared error of the unbiased estimator of the parameter b by

$$\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star\star}S_{n_m}\right)=0.0000936\times b^2.$$

This shows using relation (23) that the mean squared error of the unbiased estimator is approximately 5% bigger than the mean squared error of the biased estimator.

Since it is shown in Lemma 1 that γ_n satisfies some tight upper and lower bounds we can approximate $\alpha_m^{\star}$ in relation (19) and $\alpha_m^{\star\star}$ in relation (21), $1\leq m\leq M$ by some simpler expressions. By Lemma 1 it follows that

$$\gamma_n^2\simeq\left(\frac{n}{n-1}\right)^{\frac{1}{2}}.$$

This shows using $x^{\frac{1}{2}}-1\simeq\frac{1}{2}(x-1)$ (first order Taylor approximation around $x=1$) that by relation (19)

$$\alpha_m^{\star}\simeq 2\left(1+2\sum_{m=1}^{M}(n_m-1)\right)^{-1}(n_m-1)$$

and by relation (21)

$$\alpha_m^{\star\star}\simeq\left(\sum_{m=1}^{M}(n_m-1)\right)^{-1}(n_m-1)=\left(\sum_{m=1}^{M}n_m-M\right)^{-1}(n_m-1).$$

This shows that the minimal mean squared error unbiased estimator of the standard deviation is approximately the pooled standard deviation estimator. At the same time the mean squared

error of both estimators listed in relation (20) and (22) satisfy

$$\mathrm{MSE}(\sum_{m=1}^{M}\alpha_m^{*}S_{n_m}) \simeq \sigma^2\left(1+2\sum_{m=1}^{M}(n_m-1)\right)^{-1},$$

and

$$\mathrm{MSE}(\sum_{m=1}^{M}\alpha_m^{**}S_{n_m}) \simeq \sigma^2\left(2\sum_{m=1}^{M}(n_m-1)\right)^{-1}.$$

Finally we analyze in this section the estimation of the variance if there exist independent samples from the location-scale families m, $m=1,...,M$ with possibly different means but the same scale parameter. This means that the independent random vectors

$$\mathbf{Y}_m=(Y_{1,m},....,Y_{n_m,m}),$$

consist of independent and identically distributed random variables having mean zero and variance 1. It is assumed that additionally the random variables $Y_{1,m}$, $m=1,...,M$ have a finite fourth moment. It follows introducing $f_m^2(\mathbf{Y}_m)=S_{n_m}^2(\mathbf{Y}_m)$ that by relation (33)

$$\mathbb{E}(f_m^2(\mathbf{Y}_m))=1, \mathrm{Var}(f_m^2(\mathbf{Y}_m))=\frac{1}{n_m}\left(\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-2}\right). \tag{24}$$

By Lemma 4 and using relation (24) we obtain that the weights $\alpha_m^{\star}$ of the minimum mean squared error estimator of the variance within the class $\sum_{m=1}^{M}\alpha_m S_{n_m}^2$, $\alpha_m\in\mathbb{R}$ are given by

$$\alpha_m^{\star}=\left(1+\sum_{m=1}^{M}\frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-1}}\right)^{-1}\frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-1}}, 1\le m\le M,$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star}S_{n_m}^2\right)=\sigma^4\left(1+\sum_{m=1}^{M}\frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-1}}\right)^{-1}. \tag{25}$$

By Lemma 5 and again relation (24) the weights $\alpha_m^{\star\star}$ of the minimum mean squared error unbiased estimator of the variance within the same class equal

$$\alpha_m^{\star\star}=\left(\sum_{m=1}^{M}\frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-1}}\right)^{-1}\frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-1}}, 1\le m\le M, \tag{26}$$

and this estimator has mean squared error

$$\mathrm{MSE}\left(\sum_{m=1}^{M}\alpha_m^{\star\star}S_{n_m}^2\right)=\sigma^4\left(\sum_{m=1}^{M}\frac{n_m}{\mathbb{E}\left(Y_{1,m}^4\right)-\frac{n_m-3}{n_m-1}}\right)^{-1}.$$

Since for every $m=1,...,M$ the cdf of the random variable $Y_{1,m}$ is known we can estimate by simulation the fourth moment or calculate it. Depending on which estimator we prefer to use,

we can thus identify for every m the weights $\alpha_m^{\star\star}$ or $\alpha_m^{\star}$. If the random variable $Y_{1,m}$ for some m has a standard normal cdf it follows that its fourth moment equals 3. In some particular cases it is also possible to approximate the optimal weights by a simpler expression. If in each subgroup the random variable $Y_{1,m}$, $m = 1,...,M$, have the same cdf and the sample size n_m in each subgroup is relatively large implying $\frac{n_m-3}{n_m-1} \simeq 1$ it follows by relation (26) that for every $1 \leq m \leq M$ the optimal weights $\alpha_m^{\star\star}$ approximates

$$\alpha_m^{\star\star} \simeq n_m \left(\sum_{m=1}^{M} n_m\right)^{-1}. \tag{27}$$

Hence by relation (27) we observe that for $Y_{1,m}$, $m = 1,...,M$ having the same cdf that the pooled sample variance estimator has a mean squared error close to the minimum mean squared error unbiased estimator of the variance within the class $\sum_{m=1}^{M} \alpha_m S_{n_m}^2$, $\alpha_m \in \mathbb{R}$. Observe in quality control (see section Section 6.3.2 of Montgomery (2009) for identically distributed samples the biased estimator

$$\left(\frac{\sum_{m=1}^{M}(n_m-1)S_{n_m}^2}{\sum_{m=1}^{M} n_m - M}\right)^{\frac{1}{2}},$$

of the standard deviation is recommended without any explanation. By the previous analysis this estimator of the standard deviation is approximately for n_m large and the number of samples small the square root of the minimum mean squared error unbiased estimator (within the class $\sum_{m=1}^{M} \alpha_m S_{n_m}^2$) of the variance. In the next section we introduce sampling costs.

3 On sampling costs in quality control.

Suppose there exists M different independent subgroups from which we can generate samples needed to estimate the unknown standard deviation or variance . It is assumed that all these subgroups are in control and to collect a sample from each of these subgroups, we need to pay sampling costs. It is assumed that generating a sample of size n from subgroup $m = 1,...,M$ has sampling costs $g_m(n)$ with $g_m : \mathbb{Z}_+ \to \mathbb{N}$ denoting the increasing sampling cost function of generating a sample of size n from this subgroup m. We assume for simplicity that the range of the function g_m is an integer and since we only consider samples with sample sizes bigger or equal to 2 we assume without loss of generality for any of the independent subgroups $m = 1,...M$ that $g_m(0) = g_m(1) = 0$. Since in practice we only deal with cost measured in certain units this integer range assumption is not restrictive and simplifies our proposed dynamic solution procedure for the general sampling cost case. We now like to determine, given our available budget $B \in \mathbb{N}$, from which of the independent subgroup m, $m = 1,...,M$ we should generate a sample and how large this sample should be under the condition that our budget constraint is satisfied and the used estimator for the unknown standard deviation or variance has minimum mean squared error over all possible samples satisfying our budget constraint. Clearly, this is a static allocation optimization problem. To formulate this static allocation problem for both the estimation of the variance and the standard deviation introduce for every $1 \leq m \leq M$ the sequence of random vectors $\mathbf{Y}_m(n) = (\mathbf{Y}_{1,m},...,\mathbf{Y}_{n,m})$, $n \in \mathbb{N}$ and the functions

$$f_m(\mathbf{Y}_m(n)) := S_{n,c}(\mathbf{Y}_m(n)) = \sqrt[2]{\frac{1}{n-1}\sum_{i=1}^{n}(Y_{i,m} - \overline{Y}_{n,m})^2},\ n \geq 2,$$

and let the function $h_{m,k} : \mathbb{N} \to (0,\infty)$, $m = 1,...,m$ be given by

$$h_{m,k}(n) = \begin{cases} 0 & \text{if } n = 0,1 \\ \frac{\left(\mathbb{E}(f_m^k(\mathbf{Y}_m(n)))\right)^2}{\mathrm{Var}\left(f_m^k(\mathbf{Y}_m(n))\right)} & n \geq 3 \end{cases}$$

By Lemma 4 the static allocation problem for the standard deviation estimation problem is given by

$$\min\left\{b^2\left(1+\sum_{m=1}^M h_{m,1}(n_m)\right)^{-1} : \sum_{m=1}^M g_m(n_m) \leq B, n_j \in \mathbb{N}, 1 \leq j \leq m\right\},$$

and for the variance estimation problem by

$$\min\left\{b^4\left(1+\sum_{m=1}^M h_{m,2}(n_m)\right)^{-1} : \sum_{m=1}^M g_m(n_m) \leq B, n_j \in \mathbb{N}, 1 \leq j \leq m\right\}.$$

Since $b > 0$ it is equivalent to consider for both $k = 1$ or $k = 2$ the optimization problems

$$\upsilon(P_k) := \max\left\{\sum_{m=1}^M h_{m,k}(n_m) : \sum_{m=1}^M g_m(n_m) \leq B, n_j \in \mathbb{N}\right\}.$$

For $k = 2$ it follows by relation (25) that the objective functions $h_{m,2}$ for $m = 1,...,M$ are given by

$$h_{m,2}(n) = \frac{n}{\mathbb{E}\left(Y_{1,m}^4\right) - \frac{n-3}{n-1}} = \frac{n}{\mathbb{E}\left(Y_{1,m}^4\right) - 1 + \frac{2}{n-1}}, n \geq 2 \tag{28}$$

It is easy to check that the functions $h_{m,2}$ in relation (28) are increasing. Also for the standard deviation problem we obtain by relation (22) that for normal independent subgroups $m = 1,...,M$ it follows for every $m = 1,...M$ that

$$h_{m,1}(n) = (\gamma_n^2 - 1)^{-1}, \ n \geq 2. \tag{29}$$

By Lemma 1 this function is also increasing. Due to the separability of both the objective function and the restriction it is well known that for any increasing sampling cost functions g_m, $1 \leq m \leq M$ both optimization problems can be solved by the following dynamic programming recursion formulas. Introduce for $j = 1,...,M$ the sequences $w_{j,k} : \mathbb{N} \to \mathbb{R}_+$ given by

$$w_{j,k}(y) = \max\left\{\sum_{m=j}^M h_{m,k}(n_m) : \sum_{m=j}^M g_m(n_m) \leq y, n_j \in \mathbb{N}, j = m,...,M\right\}.$$

Clearly $w_{1,k}(B) = \upsilon(P_k)$ and

$$w_{M,k}(y) = \max\left\{h_{M,k}(n_M) : g_M(n_M) \leq y, n_j \in \mathbb{N}\right\}.$$

Also it follows for every $y \in \{0,...,B\}$ that

$$w_{j,k}(y) = \max\{h_{j,k}(n_j) + w_{j+1,k}(y - g_j(n_j)) : n_j \in \mathbb{N}, g_j(n_j) \in \{0,...,y\}\}.$$

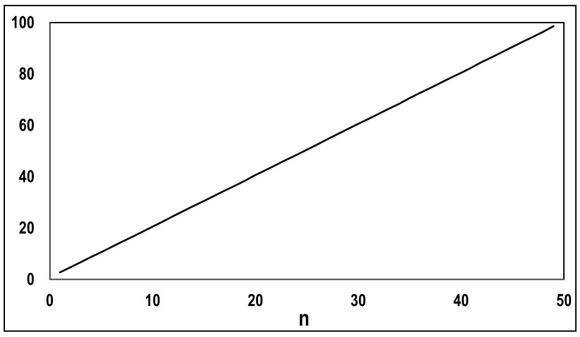


Figure 3: Plot of the function $n \to (\gamma_n^2 - 1)^{-1}$.

Although both allocation problems can be solved by dynamic programming for any set of sampling cost functions by computing iteratively the sequences $w_{j,k}$ from $j = M$ up to 1 it is easy to identify a close to optimal allocation for linear sampling cost functions given by

$$g_m(n) = \begin{cases} 0 & \text{if } n = 0,1 \\ & \\ c_m n & \text{if } n \geq 2 \end{cases}, c_m \in \mathbb{N}.$$

Since for $k = 2$ the functions in relation (28) are approximately linear a good heuristic is to spend the whole budget on generating a sample from subgroup m_* satisfying

$$m_* = \arg\min \left\{ c_m(\mathbb{E}(Y_{1,m}^4) - 1) : 1 \leq m \leq M \right\}.$$

In case we consider the standard deviation estimation allocation problem for independent normal subgroups we observe a similar behavior. As already noticed we conjectured using Lemma 1 that $(\gamma_n^2 - 1)^{-1} \approx 2n$ and so the objective function in (29) is almost linear. In Figure 3 we listed a plot of the function $n \to (\gamma_n^2 - 1)^{-1}$ confirming our approximation.

Hence the functions h_m listed in relation (29) are the same for every m and approximately a linear function. This means that we should generate the samples from the cheapest subgroup to obtain the best possible estimator of the process standard deviation. Since all subgroups generate the same type of probabilistic information and give the same additional increase to the objective function, one should, given the available budget, generate a sample having the largest possible size. Such a sample has on average the smallest mean squared error. This means for linear sampling costs it is a good heuristic to generate samples from the cheapest subgroup. In this particular case one can generate the biggest sample size.

4 Conclusion.

In this paper, we give partly an overview and extension of the classical unbiased estimators and their biased extensions used within R and S-charts for estimating the standard variation for single and multiple independent subgroups. We also consider the case of estimating the variance for single and multiple subgroups and unify both approaches for location-scale parameter families. We also propose a simple mathematical model in case we need to pay sampling costs for

generating samples from different independent subgroups in our effort to construct an estimator within a certain class of estimators having minimal mean squared error. Analyzing this static model (contrary to for example using dynamic Bayesian type control charts (Makis (2008))) confirms the intuition that, given the available budget and linear sampling cost functions, it is a good strategy to generate the whole sample from one particular subgroup.

Acknowledgement

The authors like to thank the constructive comments of the referees which greatly improved the presentation of the paper.

Disclosure statement

The authors report there are no competing interests to declare

Appendix A.

In the first section of this appendix we give the proofs of the main results mentioned in this paper. In the second subsection we list some useful properties of the gamma function needed to show the log convexity of the sequence γ_n, $n \geq 2$.

A.1 Proof of the main results

We start this subsection with a proof of Lemma 1 discussing the global behaviour of the sequence γ_n,$n \geq 2$ occurring in a normal population.

Proof. Since it is well known that $\Gamma(\frac{1}{2}) = \sqrt[2]{\pi}, \Gamma(1) = 1$ and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for every $\alpha > 0$, this shows $\gamma_2 = \sqrt[2]{\frac{\pi}{2}}$. Also by the definition of γ_n listed in relation (6), it follows for every $n \geq 2$

$$\gamma_n \gamma_{n+1} = \sqrt[2]{\frac{n-1}{2}} \sqrt[2]{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n+1}{2})} = \frac{\sqrt[2]{\frac{n-1}{2}}}{\frac{n-1}{2}} \sqrt[2]{\frac{n}{2}} = \sqrt[2]{\frac{n}{n-1}}. \tag{30}$$

Hence for every $n \geq 2$ we obtain

$$\gamma_{n+1} = \gamma_n^{-1} \sqrt[2]{\frac{n}{n-1}},$$

and we have verified relation (8). By Lemma 8, the non-negative sequence γ_n, $n \geq 2$ is decreasing and log-convex. This implies using relation (30) that for every $n \geq 2$

$$\gamma_{n+1}^2 \leq \gamma_n \gamma_{n+1} = \sqrt[2]{\frac{n}{n-1}} \leq \gamma_n^2.$$

and we have verified relation (9). Using the inequalities in relation (9), it is easy to verify that the limit relation holds. □

Next we present a proof of Lemma 2.

Proof. It follows for any $k > 0$

$$\mathrm{MSE}(\alpha T^k) = \mathrm{Var}(\alpha T^k) + \left(\mathbb{E}(\alpha T^k) - b^k\right)^2 = \alpha^2 \mathrm{Var}(T^k) + \left(\alpha \mathbb{E}(T^k) - b^k\right)^2. \tag{31}$$

Since $T(\mathbf{X}) \stackrel{d}{=} bf(\mathbf{Y})$ we obtain for every $k > 0$ that $T^k(\mathbf{X}) \stackrel{d}{=} b^k f^k(\mathbf{Y})$. This implies by relation (31) that

$$MSE(\alpha T^k) = \alpha^2 b^{2k} \mathrm{Var}\left(f^k(\mathbf{Y})\right) + \left(\alpha b^k \mathbb{E}(f^k(\mathbf{Y})) - b^k\right)^2 = b^{2k} p_k(\alpha), \tag{32}$$

with

$$p_k(\alpha) = \alpha^2 \mathbb{E}(f^{2k}(\mathbf{Y})) - 2\alpha \mathbb{E}(f^k(\mathbf{Y})) + 1.$$

Since the function p_k is a convex quadratic function and by Lyapunovs inequality and $\mathbb{E}(f^k(\mathbf{Y})) > 0$ we obtain $\mathbb{E}(f^{2k}(\mathbf{Y})) > 0$ the desired result follows by applying the first order conditions to the function p_k. □

The next proof is a proof of Lemma 3 and is actually a special case of the above result for $k = 2$.

Proof. It follows by relation (14) that $S_n(\mathbf{X}) = \sigma f(\mathbf{Y})$ with

$$f^2(\mathbf{Y}) := \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2.$$

Since S_n^2 is an unbiased estimator of σ^2 and $\mathrm{Var}(Y_i) = 1$ we obtain $\mathbb{E}(f^2(\mathbf{Y})) = 1$. Also it can be shown ((Cho and Cho, 2008),(Wilks, 1962)) that

$$0 \leq \mathrm{Var}\left(\frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2\right) = \frac{1}{n}\left(\mathbb{E}(Y_1^4) - \frac{n-3}{n-1}\right). \tag{33}$$

Since

$$\mathbb{E}(f^4(\mathbf{Y})) = \mathrm{Var}\left(\frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y}_n^2\right) + 1,$$

this yields

$$\mathbb{E}(f^4(\mathbf{Y})) = \frac{1}{n}\left(\mathbb{E}(Y_1^4) - \frac{n-3}{n-1}\right) + 1,$$

and applying Lemma 2 for $k = 2$ we obtain the desired result. □

In the next proof we verify the generalization of the above result to subgroups as expressed in Lemma 4.

Proof. Since the subgroups are independent and hence the random variables T_m, $m = 1, ..., M$ are also independent it follows using relations (15), (16) and (17) and introducing the random vector $\mathbf{Y}_m = (Y_{1,m}, ..., Y_{n_m,m})$ that

$$T(\mathbf{X}_1, ..., \mathbf{X}_M) \stackrel{d}{=} b^k \sum_{m=1}^M \alpha_m f_m^k(\mathbf{Y}_m).$$

This implies

$$\begin{aligned} \mathrm{MSE}(T) &= \mathrm{Var}(T) + \mathrm{bias}(T)^2 \\ &= b^{2k}\left(\sum_{m=1}^{M}\alpha_m^2 \mathrm{Var}\left(f_m^k(\mathbf{Y}_m)\right) + \left(\sum_{m=1}^{M}\alpha_k \mathbb{E}(f_m(\mathbf{Y}_m)) - 1\right)^2\right). \end{aligned} \tag{34}$$

To determine the optimal mean squared error we need to solve the unconstrained strictly convex quadratic minimization problem

$$\min\left\{\sum_{m=1}^{M}\alpha_m^2 \mathrm{Var}\left(f_m^k(\mathbf{Y}_m)\right) + \left(\sum_{m=1}^{M}\alpha_m \mathbb{E}(f_m^k(\mathbf{Y}_m)) - 1\right)^2 : \alpha_m \in \mathbb{R}, 1 \le m \le M\right\}.$$

This optimal solution must be the unique solution of the first order conditions applied to the above objective function and by substituting one can verify that this optimal solution is given by $\alpha^\star = (\alpha_1^\star, ..., \alpha_M^\star)$ with

$$\alpha_m^\star = \left(1 + \sum_{j=1}^{M}\frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{\mathrm{Var}(f_m^k(\mathbf{Y}_m))}\right)^{-1}\frac{\mathbb{E}(f_m^k(\mathbf{Y}_m))}{\mathrm{Var}(f_m^k(\mathbf{Y}_m))}.$$

Substituting this into relation (34) yields the expression for the bias and optimal mean squared error. □

Finally we list the proof of Lemma 5.

Proof. By relation (34) we need to solve the strict convex optimization problem

$$\min\left\{\sum_{m=1}^{M}\alpha_m^2 \mathrm{Var}(f_m^k(\mathbf{Y}^{(m)})) : \sum_{m=1}^{M}\alpha_m \mathbb{E}(f_m^k(\mathbf{Y}^{(m)})) = 1, \alpha_m \in \mathbb{R}, 1 \le m \le M\right\}.$$

Substituting now the suggested solution into the necessary and sufficient KKT conditions, we obtain the desired result. □

A.2 Some useful properties of gamma functions

First, we mention the following well-known definitions.

Definition 2. *The function* $\psi : (0,\infty) \to \mathbb{R}$ *given by*

$$\psi(\alpha) = \frac{\Gamma^{(1)}(\alpha)}{\Gamma(\alpha)},$$

with $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt, \alpha > 0$ *the well-known Gamma function is called the Digamma function.*

Definition 3. *The function* $f : (0,\infty) \to \mathbb{R}$ *is called completely monotone if all its derivatives exist and for every* $x \ge 0$ *and* $m \in \mathbb{Z}_+$ *it holds that*

$$(-1)^m f^{(m)}(x) \ge 0,$$

with $f^{(m)}$ *denoting the* m*th derivative of the function* f*.*

We now list the following result.

Lemma 7. *The function* $p:(0,\infty)\to\mathbb{R}$ *given by*

$$p(\alpha)=\ln\Gamma\left(\frac{\alpha+1}{2}\right)-\ln\Gamma\left(\frac{\alpha}{2}\right), \tag{35}$$

is non-negative and satisfies $(-1)^{m-1}p^{(m)}(\alpha)\geq 0$ *for every* $m\in\mathbb{N}$.

Proof. It follows for every $m\in\mathbb{N}$ and $\alpha\geq 0$ that

$$p^{(m)}(\alpha)=\frac{1}{2^{m-1}}\left(\psi^{(m-1)}\left(\frac{\alpha+1}{2}\right)-\psi^{(m-1)}\left(\frac{\alpha}{2}\right)\right). \tag{36}$$

In (Alzer, 1997) it is shown that the function $h_0:(0,\infty)\to\mathbb{R}$ given by $h_0(\alpha)=\ln(\alpha)-\psi(\alpha)$ is completely monotone on $(0,\infty)$. By the definition of h_0 we obtain $\psi(\alpha)=\ln(\alpha)-h_0(\alpha)$ and so for every $m\in\mathbb{N}$

$$\psi^{(m)}(\alpha)=\frac{(-1)^{m-1}(m-1)!}{\alpha^m}-h_0^{(m)}(\alpha).$$

This shows for every $m\in\mathbb{N}$ that

$$(-1)^{m-1}\psi^{(m)}(\alpha)=\frac{(-1)^{2m-2}(m-1)!}{\alpha^m}+(-1)^m\psi^{(m)}(\alpha)\geq 0. \tag{37}$$

Hence for m even it follows by relation (37) that $\psi^{(m)}(\alpha)\leq 0$, and so the function $\alpha\to\psi^{(m-1)}(\alpha)$ is decreasing. For m odd $\psi^{(m)}(\alpha)\geq 0$, and so the function $\alpha\to\psi^{(m-1)}(\alpha)$ is increasing. This shows by relation (36) the result. □

An easy application of Lemma 7 is given by the following result.

Lemma 8. *The function* $\gamma_0:(1,\infty)\to\mathbb{R}$ *given by* $\gamma_0(x)=\ln(\gamma(x))$ *with*

$$\gamma(x):=\sqrt[2]{\frac{x-1}{2}\frac{\Gamma(\frac{x-1}{2})}{\Gamma(\frac{x}{2})}},$$

is completely monotone.

Proof. Since $\alpha\Gamma(\alpha)=\Gamma(\alpha+1)$ for every $\alpha\geq 0$ we obtain

$$2\gamma_0(x)=\ln(\gamma(x)^2)=\ln\left(\frac{x-1}{2}\frac{\Gamma^2(\frac{x-1}{2})}{\Gamma^2(\frac{x}{2})}\right)=\ln\left(\frac{\Gamma(\frac{x+1}{2})\Gamma(\frac{x-1}{2})}{\Gamma^2(\frac{x}{2})}\right)=p(x)-p(x-1),$$

with the function p listed in relation (35). This implies

$$\gamma_0^{(m(}(x)=\frac{1}{2}\left(p^{(m)}(x)-p^{(m)}(x-1)\right),$$

and since by Lemma 7, we know that $(-1)^m p^{(m+1)}(x)\geq 0$ we obtain the desired result. □

References

Abbaszadehpeivasti, H. and Frenk, J. B. G. (2023). On the method of moment approach applied to a generalized gamma distribution. *Communications in Statistics-Theory and Methods*, **52**, 3685–3708.

Alzer, H. (1997). On some inequalities for the gamma and psi functions. *Mathematics of Computation*, **66**, 373–389.

Arciszewski, T. J. (2023). A review of control charts and exploring their utility for regional environmental monitoring programs. *Environments*, **10**, 78.

Arnold, S. F. (1990). *Mathematical Statistics*. Prentice Hall, Englewood Cliffs.

Arnold, B. C. and Balakrishnan, N. and Nagaraja, H. N. (2008). *A First Course on Order statistics*. SIAM, Philadelphia.

Casella, G. and Berger, R. L. (2002). *Statistical Inference (Second Edition)*. Duxbury, Pacific Grove.

Cho, E., and Cho, M. J. (2008). Variance of sample variance. *Section on Survey Research Methods-JSM*, **2**, 1291–1293.

David, H. A. (1970). *Order statistics*. Wiley, New York.

Devor, R., Sutherland, J. and Chang, T-H. (1992). *Statistical Quality Design and Control: Contemporary Concepts and Methods*. Prentice Hall, Upper Saddle River.

Figueiredo, F. and Gomes, M. I. (2006). Box-Cox transformations and robust control charts in SPC. In *Advanced Mathematical and Computational Tools in Metrology* VII, (pp. 35-46), World Scientific, New Jersey.

Harter, H. L. (1960). Tables of range and studentized range. *The Annals of Mathematical Statistics*, **31**, 1122–1147.

Irwin, M. and Miller, M. (2004). *John E. Freund's Mathematical Statistics: With Applications*. Prentice Hall, New York.

Mahmoud, M. A., Henderson, G. R., Epprecht, E. K., and Woodall, W. H. (2010). Estimating the standard deviation in quality-control applications. *Journal of Quality Technology*, **42**, 348–357.

Makis, V. (2008). Multivariate Bayesian control chart. *Operations Research*, **56**, 487–496.

Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. Wiley, New York.

Ryan, T. P. (2011). *Statistical Methods of Quality Improvement*. Wiley, New York.

Suman, G. and Prajapati, D. (2018). Control chart applications in healthcare: a literature review. *International Journal of Metrology and Quality Engineering*, **9**, 5.

Vardeman, S. B. (1999). A brief tutorial on the estimation of the process standard deviation. *IIE transactions*, **31**, 503–507.

Woodall, W. H., and Montgomery, D. C. (2000). Using ranges to estimate variability. *Quality Engineering*, **13**, 211–217.

Wilks, S. S. (1962). Mathematical Statistics (Second Edition). Wiley, New York.

Zwetsloot, I. M. Jones-Farmer, A. and Woodall, W. H. (2023). Monitoring Univariate processes using control charts: some practical related issues and advice. *Quality Engineering*, **36**, 487–499.

Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 143–151. Research Paper

A probabilistic model for the structure functions of coherent systems

Mohammad Khanjari Sadegh*

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
Email(s): mkhanjari@um.ac.ir

Abstract. The probabilistic idea for the structure-function of a coherent system is suggested in the literature. In this paper, a primary model for this idea is defined. This model leads to an extension of the class of coherent systems with binary (deterministic) structure functions. Also, in the stochastic comparison of the coherent systems with probabilistic structure functions, it is shown that the concept of survival signature can successfully be used in this model. The well-known idea of the system signature is not usable here.

Keywords: Coherent systems; Probabilistic structure function; Survival signature; System signature.

1 Introduction

The mathematical and statistical theory of system reliability is based on the central concept of "structure function", which is a binary function and describes deterministically the state of a system when the states of its components are given. In practical uses and real world applications, it is somewhat restricted as for various reasons the functioning of system components does not always provide absolute certainty that the system will function. In other words, except for the failures of system components, other random factors may cause to system failure. For example in a car system sometimes we have seen that the main and key components of the car are functioning but the car does not work. Some unconsidered factors such as the road, weather or driver circumstances, can lead to uncertainty about whether or not the system meets the actual requirements. In a computer system, there is always some uncertainty about the quality of the compatibility between the computer with the new components and softwares. Generally, in the most of real-world systems, because of lack of our perfect knowledge and our uncertainty about the quality of the system functioning, a generalization of structure function from binary function to a probability may have substantial advantages for realistic system reliability quantification.

*Corresponding author
Received: 27 October 2023 / Accepted: 6 August 2024
DOI: 10.22067/smps.2024.45575

© 2024 Ferdowsi University of Mashhad

This idea was suggested by Coolen and Coolen-Maturi (2016a). A specific model for probabilistic structure function is not given in their work.

In this paper, among various factors that may have effects on system performance, we consider the quality of the links between system components that we think it is an important factor in system functioning. The perfect required performance of a single component when it is putted in a system, is not only depend on its functioning but also to its relationship and link with the other system components. In the following section, we explain our model in detail. Illustrative examples are also given. Finally in Section 3, we show that the survival signature a concept defined in Coolen and Coolen-Maturi (2013), can successfully be used in our model for probabilistic structure function. Particularly its application in stochastic comparison among the coherent systems with the probabilistic structure functions is studied and show that the system signature is deficient here.

2 Probabilistic structure function

In this section, for the sake of completeness we first review the binary structure function and then present our probabilistic model for the structure function. Consider a system consists of n components and assume that all components and the system are in a functioning or failed state. In a fixed point of time, let the state vector $\mathbf{X} = (X_1, \ldots, X_n) \in \{0,1\}^n$, with

$$X_i = \begin{cases} 1, & \text{if } i\text{th component is working,} \\ 0, & \text{otherwise.} \end{cases}$$

The structure function $\phi : \{0,1\}^n \to \{0,1\}$ is defined as

$$\phi(\mathbf{X}) = \phi(X_1, \ldots, X_n) = \begin{cases} 1, & \text{if the system is working,} \\ 0, & \text{otherwise.} \end{cases}$$

The system is said to be coherent if $\phi(\mathbf{X})$ is not decreasing in any X_i and all components be relevant, that is

$$\phi(1_i, \mathbf{X}) - \phi(0_i, \mathbf{X}) = 1,$$

at least for one $\mathbf{X} \in \{0,1\}^{n-1}$.
Obviously $\phi(1, \ldots, 1) = 1$ and $\phi(0, \ldots, 0) = 0$.
Also $\phi(\mathbf{X}) = X_i \phi(1_i, \mathbf{X}) + (1 - X_i)\phi(0_i, \mathbf{X})$, $i = 1, \ldots, n$ (pivotal decomposition).
If $\phi(\mathbf{X}) = 1(0)$ then $\mathbf{X}$ is said to be a path(cut) vector and the corresponding subset $P = \{1 \le i \le n | X_i = 1\} (C = \{1 \le i \le n | X_i = 0\})$ is called a path(cut) set of the system. Note that an arbitrary vector $\mathbf{X} \in \{0,1\}^n$ always is a path vector or a cut vector but not both. Whereas $P \subseteq \{1, \ldots, n\}$ can be both a path and a cut set.

If $P(C) \subseteq \{1, \ldots, n\}$ is a path(cut) set and $Q \subset P(C)$ is not a path(cut) set then $P(C)$ is a minimal path(cut) set of the system. It is known that if $P_1, \ldots, P_r (C_1, \ldots, C_s)$ are all minimal path(cut) sets of the system, then

$$\phi(\mathbf{X}) = \max_{1 \le i \le r} \min_{j \in P_i} X_j = \min_{1 \le i \le s} \max_{j \in C_i} X_j. \tag{1}$$

If $p_i = EX_i = P(X_i = 1)$ is the reliability of component i then

$$h(\mathbf{p}) = h(p_1, \ldots, p_n) = E\phi(\mathbf{X}) = P(\phi(\mathbf{X}) = 1)$$

is the reliability function of the system. For more details on the reliability of coherent systems with binary structure function, see Barlow and Proschan (1975).

Now, we define our model for the structure function as a probability.
Here we still assume that the binary state for the system components. We also take into account the quality of the links between the components as an effective factor in system performance. This factor may cause to failure of the system even if all its components are working. We consider a particular link for each system component and assume that the component is working perfectly, if both the component and its specific link are functioning. Therefore, given the states of components, we consider the state of the system to be a conditional probability as follows

$$\phi_{\mathbf{a}}(\mathbf{x}) = P_{\mathbf{a}}(\text{system is functioning}|\mathbf{X} = \mathbf{x}),$$

where $\mathbf{a} = (a_1, \ldots, a_n)$ with

$$a_i = P(\text{the link of component } i \text{ is functioning}), i = 1, \ldots, n.$$

As usual we assume that the failure of the components leads to the failure of the system definitely. Let

$$S = \begin{cases} 1, & \text{if system is functioning,} \\ 0, & \text{otherwise,} \end{cases}$$

then $\phi_{\mathbf{a}}(\mathbf{x}) = P_{\mathbf{a}}(S = 1|\mathbf{X} = \mathbf{x})$.

Model assumption:
For a given vector $\mathbf{a} = (a_1, \ldots, a_n)$, we assume that
A1. $\phi_{\mathbf{a}}(\mathbf{x})$ is increasing in x_i and a_i, $i = 1, \ldots, n$.
A2. The component i is relevant, that is $\phi_{\mathbf{a}}(1_i, \mathbf{x}) > 0$ and $\phi_{\mathbf{a}}(0_i, \mathbf{x}) = 0$ at least for one $\mathbf{x} \in \{0,1\}^{n-1}$, $i = 1, \ldots, n$.

Under the above conditions, we call the system as a coherent system. In a coherent system we have $\phi_{\mathbf{a}}(0, \ldots, 0) = 0$. Also $\phi_{\mathbf{a}}(1, \ldots, 1) > 0$, which is not necessary equal to 1. It means that even if all components of the system are working, there exists a positive probability of system failure. But if $\mathbf{x}$ is a cut vector we have $\phi_{\mathbf{a}}(\mathbf{x}) = \phi(\mathbf{x}) = 0$ in which $\phi(\mathbf{x})$ is the binary structure function. In fact

$$\phi_{\mathbf{a}}(\mathbf{x}) = h(\mathbf{a}.\mathbf{x}),$$

where

$$\mathbf{a}.\mathbf{x} = (a_1 x_1, \ldots, a_n x_n),$$

and $h(\mathbf{p}) = E\phi(\mathbf{X})$ is the reliability function of the system with binary structure function.
In particular case when $\mathbf{a} = (1, 1, \ldots, 1)$, we have $\phi_{\mathbf{a}}(\mathbf{x}) = h(\mathbf{x}) = \phi(\mathbf{x})$.
Note that $\phi_{\mathbf{a}}(1, \ldots, 1) = h(a_1, \ldots, a_n) \leq 1$.
A3. We assume that the links between components are functioning independently and are independent of the states of components. Also assume that X_i's, $i = 1, \ldots, n$ are independent.

Minimal path(cut) sets
If $\phi_{\mathbf{a}}(\mathbf{x}) > (=)0$ we call $\mathbf{x}$ as a path(cut) vector. The path(cut) sets and the minimal path(cut) sets are defined as before. Although $\phi_{\mathbf{a}}(\mathbf{x})$ is simply satisfied in the pivotal decomposition but the Equation (1) does not hold true for $\phi_{\mathbf{a}}(\mathbf{x})$.

Lemma 1. *For* $\mathbf{p} = (p_1, \ldots, p_n)$ *and* $\mathbf{a} = (a_1, \ldots, a_n)$ *the system reliability is given by*

$$h_{\mathbf{a}}(\mathbf{p}) = E_{\mathbf{a}}S = P_{\mathbf{a}}(S = 1) = \sum_{\mathbf{x}_p} \phi_{\mathbf{a}}(\mathbf{x}_p) P(\mathbf{X} = \mathbf{x}_p).$$

Also

$$1 - h_{\mathbf{a}}(\mathbf{p}) = P_{\mathbf{a}}(S = 0) = \sum_{\mathbf{x}_p} (1 - \phi_{\mathbf{a}}(\mathbf{x}_p)) P(\mathbf{X} = \mathbf{x}_p) + \sum_{\mathbf{x}_c} P(\mathbf{X} = \mathbf{x}_c),$$

where $\mathbf{x}_p(\mathbf{x}_c)$ *is a path(cut) vector of the system.*

Proof. We have

$$h_{\mathbf{a}}(\mathbf{p}) = P_{\mathbf{a}}(S = 1) = \sum_{\mathbf{x}_p} P_{\mathbf{a}}(S = 1 | \mathbf{X} = \mathbf{x}_p) P(\mathbf{X} = \mathbf{x}_p) = \sum_{\mathbf{x}_p} \phi_{\mathbf{a}}(\mathbf{x}_p) P(\mathbf{X} = \mathbf{x}_p).$$

The proof of the second equality given for unreliability of the system is similar. □

Remark 1. *The Lemma 1 shows clearly that the system failure is not only dependent to the component failures but also on the failure of the links between components. The first sum in* $1 - h_{\mathbf{a}}(\mathbf{p})$ *gives in fact the contribution of the link failures between the system components to the failure of the system and the second sum is the same for the component failures.*

Remark 2. *As mentioned before,* $\phi_{\mathbf{a}}(\mathbf{x})$ *reduces to the binary structure function* $\phi(\mathbf{x})$ *when* $\mathbf{a} = (1, \ldots, 1)$*, that is all links between components are functioning. Therefore the class of coherent systems with binary structure functions is a subclass of coherent systems with probabilistic structure functions. Obviously* $\phi_{\mathbf{a}}(\mathbf{x}) \leq \phi(\mathbf{x})$ *and therefore* $h_{\mathbf{a}}(\mathbf{p}) \leq h(\mathbf{p})$.

In our model, in fact

$$h_{\mathbf{a}}(\mathbf{p}) = E(\phi_{\mathbf{a}}(\mathbf{X})) = E(\phi(\mathbf{a}.\mathbf{X})) = h(\mathbf{a}.\mathbf{p}) = h(a_1 p_1, \ldots, a_n p_n).$$

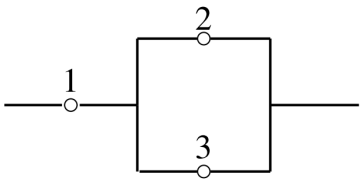
Example 1. We now give some examples.

- Series system: We have $\phi_{\mathbf{a}}(\mathbf{x}) = a_1 a_2 \cdots a_n x_1 x_2 \cdots x_n$ and $h_{\mathbf{a}}(\mathbf{p}) = a_1 a_2 \cdots a_n p_1 p_2 \cdots p_n$.
- Parallel system: In this system we have

  $$\phi_{\mathbf{a}}(\mathbf{x}) = 1 - \prod_1^n (1 - a_i x_i)$$

  and

  $$h_{\mathbf{a}}(\mathbf{p}) = 1 - \prod_1^n (1 - a_i p_i).$$



- 2-out-of-3:G system: In this system we have

$$\phi_{\mathbf{a}}(\mathbf{x}) = a_1a_2x_1x_2 + a_1a_3x_1x_3 + a_2a_3x_2x_3 - 2a_1a_2a_3x_1x_2x_3$$

and

$$h_{\mathbf{a}}(\mathbf{p}) = a_1a_2p_1p_2 + a_1a_3p_1p_3 + a_2a_3p_2p_3 - 2a_1a_2a_3p_1p_2p_3.$$

- Series-parallel system: For this system we have

$$\phi_{\mathbf{a}}(\mathbf{x}) = a_1a_2x_1x_2 + a_1a_3x_1x_3 - a_1a_2a_3x_1x_2x_3$$

and

$$h_{\mathbf{a}}(\mathbf{p}) = a_1a_2p_1p_2 + a_1a_3p_1p_3 - a_1a_2a_3p_1p_2p_3.$$

3 Stochastic comparison

In this section, for the sake of completeness, we first review the concepts of system signature and survival signature and find their relationship. We then define the survival signature of a coherent system with probabilistic structure function and show that it can successfully be used in our model.

The concept of the system signature was introduced by Samaniego (1985). It is a very useful tool and has a wide range of applications in the study of reliability analysis of coherent systems. Let $T = \phi(T_1, \ldots, T_n)$ be the lifetime of a coherent system where T_i is the lifetime of component i. When T_i's are independent and identically distributed (i.i.d.), it is shown that

$$P(T > t) = \sum_{i=1}^{n} s_i P(T_{i:n} > t), \tag{2}$$

where $T_{i:n}$ is the i^{th} ordered component lifetime, $s_i = P(T = T_{i:n})$ and the probability vector $\mathbf{s} = (s_1, \ldots, s_n)$ is the system signature (see, Samaniego (1985)).

Although the system signature is an important tool and has many applications in reliability studies of coherent systems with binary structure functions but it is not the case for the coherent systems with probabilistic structure functions as we have seen in the previous section that the system failure is not determined by the failure of components deterministically. Note that the Equation (2) does not hold for the systems with probabilistic structure functions.

This section shows that how in coherent systems with probabilistic structure functions, the "survival signature" (a concept introduced by Coolen and Coolen-Maturi (2013)), plays the role of the system signature in coherent systems with binary structure functions.

As given in Coolen and Coolen-Maturi (2013), the survival signature for a coherent system of order n with components of r different types is defined in which the system has m_k components of type k, $k = 1,\ldots,r$ and also the components of the same type are exchangeable and the components of different types are independent. For $i_k = 0,\ldots,m_k$ and $k = 1,\ldots,r$ this measure is defined as follow

$$\bar{s}(i_1,\ldots,i_r) = P(\phi(\mathbf{X}) = 1|\text{exactly } i_k \text{ components of type } k \text{ are working})$$
$$= \left[\prod_{k=1}^{r}\binom{m_k}{i_k}\right]^{-1} \sum_{\mathbf{x}\in S_{i_1,\ldots,i_r}} \phi(\mathbf{x}),$$

where $S_{i_1,\ldots,i_r} = \{\mathbf{x}|\sum_{j=1}^{m_k} x_j^k = i_k,\ k = 1,\ldots,r.\}$
Also the system reliability is given by

$$P(T > t) = \sum_{i_1=0}^{m_1}\cdots\sum_{i_r=0}^{m_r} \bar{s}(i_1,\ldots,i_r)\prod_{k=1}^{r} P(C_t^k = i_k),$$

where $C_t^k \in \{0,1,\ldots,m_k\}$ is the number of components of type k that function at time t.
Under the above stated assumptions it is shown that the $\bar{s}$, as like as the system signature is not dependent on the joint distribution of the components (see, Coolen and Coolen-Maturi (2013)). For more details on the survival signatures and their applications, see Coolen and Coolen-Maturi (2016b) and Coolen and Coolen-Maturi (2021).

3.1 Relationship between system signature and survival signature

In this subsection we consider the relationship between the system signature and survival signature in coherent systems with binary structure functions. For simplicity, we assume that all n components of the system are of the same type. That is $r = 1$. Therefore we have

$$\bar{s}(i) = P(\phi(\mathbf{X}) = 1|\text{exactly } i \text{ components are working}) = \frac{\sum_{|\mathbf{x}|=i}\phi(\mathbf{x})}{\binom{n}{i}},$$

where $|\mathbf{x}| = \sum_1^n x_k$. Note that $\bar{s}(0) = 0$ and $\bar{s}(n) = 1$.
The following lemma gives the relationship between s_i and $\bar{s}(i)$.

Lemma 2. *We have*

$$\sum_{k=i+1}^{n} s_k = \bar{s}(n-i) = \frac{\sum_{|\mathbf{x}|=n-i}\phi(\mathbf{x})}{\binom{n}{i}}.$$

Proof. In view of the definition of system signature which is a probability vector, and definition of survival signature, we have

$$\sum_{k=i+1}^{n} s_k = P(\text{system failure occurs after } i\text{th failure of components})$$
$$= P(\text{system is working until } i\text{th failure of components},)$$
$$= P(\phi(\mathbf{X}) = 1|\sum X_k = n-i) = \bar{s}(n-i).$$

□

Note that for $i = 1,\ldots,n$ we have:

$$s_i = \bar{s}(n-i+1) - \bar{s}(n-i).$$

Lemma 3. *Consider two coherent systems with structure functions $\phi_1(\mathbf{X})$ and $\phi_2(\mathbf{X})$. Let $\bar{s}_1$, $\bar{s}_2$, $\mathbf{s}_1$ and $\mathbf{s}_2$ denote their survival signatures and system signatures, respectively. Then*

$$\bar{s}_1(i) \le \bar{s}_2(i), i = 1,2,\ldots,n \quad \textit{if and only if} \quad \mathbf{s}_1 \le_{st} \mathbf{s}_2.$$

Proof. In view of the definition of usual stochastic order, the proof follows from Lemma 2. □

To see the definition of the stochastic order $\mathbf{s}_1 \le_{st} \mathbf{s}_2$, and other stochastic orders, we refer the reader to Shaked and Shanthikumar (2007).
The Lemma 3 shows that in stochastic ordering of coherent systems with binary structure functions, the system signature and survival signature are two equivalent tools. We see in the sequel that, it is not case for the coherent systems with probabilistic structure functions.

Definition 1. *In a coherent system with probabilistic structure function $\phi_{\mathbf{a}}(\mathbf{x})$, we define the survival signature as follow*

$$\bar{s}_{\mathbf{a}}(i) = P_{\mathbf{a}}(\textit{system is functioning}|\textit{the number of working components is } i).$$

The following lemma gives an expression for $\bar{s}_{\mathbf{a}}(i)$ in a coherent system with i.i.d. components. It is useful for determining the reliability function of the system.

Lemma 4. *In a coherent system with i.i.d. components and with probabilistic structure function $\phi_{\mathbf{a}}(\mathbf{x})$, we have*

$$\bar{s}_{\mathbf{a}}(i) = \frac{\sum_{\mathbf{x}:|\mathbf{x}|=i} \phi_{\mathbf{a}}(\mathbf{x})}{\binom{n}{i}}.$$

Proof. We have

$$\begin{aligned}
\bar{s}_{\mathbf{a}}(i) &= P_{\mathbf{a}}(S=1|\sum_{j=1}^{n} X_j = i) \\
&= \frac{\sum_{\mathbf{x}:|\mathbf{x}|=i} P_{\mathbf{a}}(S=1,\mathbf{X}=\mathbf{x})}{P(\sum X_j = i)} \\
&= \frac{\sum_{\mathbf{x}:|\mathbf{x}|=i} P_{\mathbf{a}}(S=1|\mathbf{X}=\mathbf{x})P(\mathbf{X}=\mathbf{x})}{P(\sum X_j = i)} \\
&= \frac{\sum_{\mathbf{x}:|\mathbf{x}|=i} \phi_{\mathbf{a}}(\mathbf{x}) p^i (1-p)^{n-i}}{\binom{n}{i} p^i (1-p)^{n-i}},
\end{aligned}$$

where p is the common reliability of components. □

Lemma 4 shows that the survival signature $\bar{s}_{\mathbf{a}}(i)$ is not dependent on component reliabilities. It is easy to see that the above lemma also holds true when the lifetimes of the system components are exchangeable. Note that when $\mathbf{a} = (1,\ldots,1)$ then $\bar{s}_{\mathbf{a}}(i)$ reduces to $\bar{s}(i)$, the usual survival signature for the coherent systems with binary structure functions.
Now in the next theorem, we obtain the reliability function of the system.

Theorem 1. *Under the assumptions of Lemma 4, we have*

$$h_{\mathbf{a}}(p) = \sum_{i=1}^{n} \bar{s}_{\mathbf{a}}(i) \binom{n}{i} p^i (1-p)^{n-i}. \tag{3}$$

Proof. We have

$$\begin{aligned} h_{\mathbf{a}}(p) &= P_{\mathbf{a}}(S=1) \\ &= \sum_{i=1}^{n} P_{\mathbf{a}}(S=1 | \sum_j X_j = i) \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^{n} \bar{s}_{\mathbf{a}}(i) \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned}$$

□

The following lemma compares the coherent systems with probabilistic structure functions in usual stochastic order.

Lemma 5. *Consider two coherent systems with probabilistic structure functions $\phi^{(1)}_{\mathbf{a}_1}(\mathbf{X})$ and $\phi^{(2)}_{\mathbf{a}_2}(\mathbf{X})$. Let $\bar{s}^{(1)}_{\mathbf{a}_1}$, $\bar{s}^{(2)}_{\mathbf{a}_2}$, $\mathbf{s}_1$, $\mathbf{s}_2$, $h^{(1)}_{\mathbf{a}_1}(p)$ and $h^{(2)}_{\mathbf{a}_2}(p)$ denote their survival signatures, system signatures and reliability functions, respectively. If*

$$\bar{s}^{(1)}_{\mathbf{a}_1}(i) \leq \bar{s}^{(2)}_{\mathbf{a}_2}(i), i = 1, \ldots, n$$

then

$$h^{(1)}_{\mathbf{a}_1}(p) \leq h^{(2)}_{\mathbf{a}_2}(p).$$

Proof. From Theorem 1, the proof is immediate. □

We note that

$$\bar{s}^{(1)}_{\mathbf{a}_1}(i) \leq \bar{s}^{(2)}_{\mathbf{a}_2}(i), i = 1, \ldots, n$$

is not necessary equivalent to $\mathbf{s}_1 \leq_{st} \mathbf{s}_2$, unless $\mathbf{a}_1 = \mathbf{a}_2 = (1, \ldots, 1)$.

The Equation (3) is a similar version of the Equation (2). For example in a series system with i.i.d. components and in view of Example 1, we have $\bar{s}_{\mathbf{a}}(i) = 0$ for $i = 1, \ldots, n-1$ and $\bar{s}_{\mathbf{a}}(n) = (a_1 a_2 \cdots a_n)/\binom{n}{i}$. Hence $h_{\mathbf{a}}(p) = a_1 a_2 \cdots a_n p^n$.

Acknowledgements

The author would like to thanks referees and the editor in chief for giving valuable and constructive comments, which led to a significantly improvement in the revised version of the paper.

Disclosure statement

The authors report there are no competing interests to declare

References

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston.

Coolen, F. P. A. and Coolen-Maturi, T. (2013). Generalizing the signature to systems with multiple types of components. In: *Complex Systems and Dependability. Advances in Intelligent and Soft Computing*, Vol 170, 115–130. Springer, Berlin.

Coolen F. P. A. and Coolen-Maturi T. (2016a). The structure function for system reliability as predictive (imprecise) probability. *Reliability Engineering and System Safety* **154**, 180–187.

Coolen, F.P.A. and Coolen-Maturi T. (2016b). On the structure function and survival signature for system reliability. *Safety and Reliability*, **36**, 77–87.

Coolen, F.P.A. and Coolen-Maturi T. (2021). The survival signature for quantifying system reliability: an introductory overview from practical perspective. In:*Reliability Engineering and Computational Intelligence*, Vol 976, 23–37, Springer, Cham.

Samaniego, F. J. (1985). On closure of the IFR class under formation of coherent systems. *IEEE Transactions on Reliability*, **34**, 69–72.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*, Springer, New York.

Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 153–169. Research Paper

Some copula-based results on optimal number of cold standby components in parallel systems

Motahareh Parsa, Hadi Jabbari* and Jafar Ahmadi

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
Email(s): motahareh.parsa@gmail.com, Jabbarinh@um.ac.ir, ahmadi-j@um.ac.ir

Abstract. This paper deals with a system's profit and efficiency functions consisting of a two-parallel components supported by $(n-2)$ cold standby redundancy. It is supposed that the components are not repairable and the failed component is immediately replaced with one of the cold standby components. Because the lifetimes of active components are dependent, their dependency is modelled by Farlie-Gumbel-Morgenstern (FGM) copula function. Finding the optimal n according to the highest profit function of the system and efficiency values is studied while the effect of the dependence parameter is considered. It is concluded that though increasing n grants higher system reliability it is not always wise as long as the redundancy maintenance costs. Due to the complexity of the formulas, for the large values of n, the results are analyzed numerically and graphically. Finally, some examples are given to illustrate the results.

Keywords: Cold standby redundancy; Copula function; Profit function; Reliability; System efficiency.

1 Introduction

Utilizing the redundant units in the system has always seemed a proper tool to improve the system's structure of more reliable and available systems. The pioneering works like Boland et al. (1988) and Boland et al. (1991) investigated the redundancy allocation in coherent systems. But, to optimize the system's profit, one also needs to consider the costs of spare unites and maintenance. Gupta et al. (1986) analyzed the cost function of server two-unit cold standby. Kong and Frangopol (2004) used a life-cycle cost analysis of deteriorating structures to introduce the optimum maintenance scenario. Ram et al. (2013) employed cost profit analysis for a highly reliable complex system. Ram and Kumar (2015) used cost profit analysis to investigate a standby system incorporating waiting time to repair. Singh and Gulati (2014) studied the

*Corresponding author
Received: 23 November 2023 / Accepted: 25 September 2024
DOI: 10.22067/smps.2024.45746

© 2024 Ferdowsi University of Mashhad

reliability and cost analysis of a two-unit standby redundant complex system under different repair facility. Singh and Rawal (2014) focused on the standby complex system and used profit function to discuss the available maintenance.

The present paper deals with the average profit function of a parallel $(2,n-2)$ system. The parallel $(2,n-2)$ system, as was discussed by Papageorgiou and Kokolakis (2004) and Papageorgiou and Kokolakis (2007) is a system of two initially active units which are supported by $n-2$ cold standbys of the same type. Once one active unit fails, it is replaced by one of the standbys instantly and the process is continued to the final standby. Here, we intend to find the optimum number of units in the system while installing the cold standbys costs. To model the reliability of a system consisting of time-dependent units, we have specifically employed Farlie-Gumbel-Morgenstern (FGM) copula which is known for being well-fitted to data of weak dependence (see, Nelsen (2006)). Domma and Giordano (2013) obtained a closed-form expression for stress-strength reliability using the dependence through FGM copula and Shih and Emura (2019) extended their results using the generalized FGM copula. Yongjin et al. (2018) studied the effect of degrees of dependence among components on system reliability using FGM copula. Zhang and Zhang (2022) proposed a copula-based approach to study the allocation problem of hot standbys in series systems composed of two heterogeneous and dependent components.

The rest of the paper is organized as follows. Section 2 contains some preliminary concepts. Section 3 discusses the main topics related to parallel $(2,n-2)$ system structure and defines the profit function of the system. A criteria is proposed for evaluating the system efficiency. Also, some upper and lower bounds are provided which facilitate to find the optimum strategy for the system structure. In Section 4, some illustrative examples are given in which FGM is applied for modelling the dependence concept of components lifetimes in the system when the component lifetime follows exponential distribution.

2 Preliminaries

Suppose that (X,Y) has distribution function (df) F with absolutely continuous marginal dfs F_1 and F_2. The random vector (X,Y) is said to be positive quadrant dependent (PQD) (negative quadrant dependent (NQD)) if and only if for every $x,y \in R$,

$$F(x,y) \geq (\leq) F_1(x)F_2(y).$$

The notion of 2-copula, denoted by $C(.,.)$, was first introduced by Sklar (1959) which models bivariate distribution of (X,Y) by its marginals as follows

$$F(x,y) = C(F_1(x),F_2(y)). \tag{1}$$

The survival copula is also defined by Nelsen (2006) as

$$\begin{aligned}\bar{F}(x,y) &= P(X>x,Y>y)\\ &= \bar{F}_1(x)+\bar{F}_2(y)-1+C(1-\bar{F}_1(x),1-\bar{F}_2(y))\\ &= \hat{C}(\bar{F}_1(x),\bar{F}_2(y)),\end{aligned}$$

where $\bar{F}_i(x) = 1-F_i(x)$ for $i=1,2$.

The renowned FGM family of bivariate dfs is of the form

$$F(x,y) = F_1(x)F_2(y)\big[1+\alpha\bar{F}_1(x)\bar{F}_2(y)\big],\ -1 \le \alpha \le 1,$$

where α is the dependence parameter, $\alpha \ge 0$ models PQD and $\alpha \le 0$ models NQD. According to (1) for $u,v \in [0,1]$, the FGM 2-copula is given by

$$C(u,v) = uv\big[1+\alpha(1-u)(1-v)\big],\ -1 \le \alpha \le 1.$$

A well-known version of multivariate FGM family had been introduced by Nelsen (2006) as

$$F(x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} F_i(x_i)\left[1+\sum_{k=2}^{n}\sum_{1\le i_1<i_2<\cdots<i_k\le n}\alpha_{i_1 i_2\ldots i_k}\prod_{j=1}^{k}\bar{F}_{i_j}(x_{i_j})\right].$$

Applying n-dimensional version of Sklar's Theorem was given in Schweizer and Sklar (1983), the FGM n-copula family introduced by Johnson and Kotz (1972) is as

$$C(u_1,u_2,\ldots,u_n) = \prod_{i=1}^{n} u_i\left[1+\sum_{k=2}^{n}\sum_{1\le i_1<i_2<\cdots<i_k\le n}\alpha_{i_1 i_2\ldots i_k}\prod_{j=1}^{k}(1-u_{i_j})\right], \tag{2}$$

in which (2^n-n-1) parameters must satisfy the following conditions

$$1+\sum_{k=2}^{m}\sum_{s_1\le r_1<r_2<\cdots<r_k\le s_m}\alpha_{r_1 r_2\ldots r_k}\prod_{j=1}^{n}\xi_{r_j} \ge 0,\ |\xi_{r_j}| \le 1,$$

for $1 \le s_1 < s_m \le n$ and $m = 2,3,\ldots,n$. Here, each k-margin, $2 \le k < n$, of an FGM n-copula is an FGM k-copula (see, Conway (1983)) for more applications of FGM n-copula).

This study chooses the FGM copula to model the dependence structure of component's lifetimes. The main reasons for this choice are modelling weak dependence which is the property of the lifetime data, the capability of modelling both positive and negative dependence, and well-fitting to real reliability data (see, Amini et al. (2011)).

3 Main results

In this section, we suppose a parallel system contains n components that operated initially by two components and the remaining $(n-2)$ components are cold standbys, which is known as the parallel $(2,n-2)$ system. We study the structures and some properties of the systems and define the mean profit and efficiency functions up to time t. Later, a lower bound for the profit function is introduced that simplifies the effect of the number of redundancies.

3.1 System profit and efficiency functions

Let T_i be the lifetime of the ith component of the parallel $(2,n-2)$ system and S_i be standby duration of the cold standby component, for $i = 1,2,\ldots,n$. Consequently, $S_1 = S_2 = 0$. Once the first failure occurs, the third (the first standby) component is replaced and starts its operation

at $S_3 = \min\{T_1, T_2\}$. The forth (the second standby) component is replaced upon failure times of $\max\{T_1, T_2\}$ or $S_3 + T_3$, whichever occurs first i.e.

$$S_4 = \min\{\max\{T_1, T_2\}, T_3 + S_3\}, \tag{3}$$

Also, for $j = 5$ we have

$$S_5 = \min\{\max\{T_1, T_2, T_3 + S_3\}, T_4 + S_4\}. \tag{4}$$

Similarly, for $j \geq 6$

$$S_j = \min\{\max\{T_1, T_2, T_3 + S_3, \ldots, T_{j-2} + S_{j-2}\}, T_{j-1} + S_{j-1}\}, \quad j = 6, \ldots, n. \tag{5}$$

See, Papageorgiou and Kokolakis (2007) for more details. Suppose that T_i's $i = 1, 2, \ldots, n$ are iid with CDF F_T and F_{S_j} is the CDF of S_j. Clearly, T_i's and S_j's are not independent and S_j is independent of T_i for $i \geq j$. Let T_s be the system lifetime. Then as given in Papageorgiou and Kokolakis (2007),

$$T_s = \max\{T_1, T_2, T_3 + S_3, \ldots, T_n + S_n\}.$$

Here, we have supposed that there exists positive or negative dependence between the units lifetimes and the standby periods. Then the reliability function of the proposed parallel $(2, n-2)$ system is given by

$$\begin{aligned} \bar{F}_{sys,n}(x) &= P(T_s > x) \\ &= 1 - C(F_{T_1}(x), F_{T_2}(x), F_{T_3+S_3}(x), \ldots, F_{T_n+S_n}(x)), \end{aligned} \tag{6}$$

where $C(.)$ denotes an appropriate copula function for modelling the dependence between units lifetimes. The cdf of $T_j + S_j$ is given by

$$F_{T_j+S_j}(u) = \int_0^u F_{S_j}(u-s) f_T(s) ds, \quad j = 3, ..., n.$$

For $j = 3$, we have

$$\begin{aligned} F_{T_3+S_3}(x) &= P(T_3 + S_3 \leq x) \\ &= \int_0^x F_{S_3}(x-u) f_T(u) du \\ &= F_T(x) - \int_0^x \bar{F_T}^2(x-u) f_T(u) du. \end{aligned} \tag{7}$$

Let k_1 be the income for reliability of system per unit time and k_2 be the storage and maintenance cost per unit time of standby components. Then, the expected profit function of parallel $(2, n-2)$ system up to time $t > 0$ is given by

$$B_n(t) = In_n(t) - Co_n(t), \tag{8}$$

where

$$In_n(t) = k_1 \int_0^t \bar{F}_{sys,n}(x) dx, \tag{9}$$

denotes the system average income in $[0,t]$ and

$$Co_n(t) = k_2 \sum_{j=3}^{n} \int_0^t \bar{F}_{S_j}(x)dx, \tag{10}$$

is the average cost of the system during the interval $[0,t]$. Ram and Singh (2010) considered a similar criteria for the expected profit of complex systems to do the cost analysis. Ram et al. (2013) also used the same expected profit for cost analyzing of a standby system with including the waiting time to repair in their study.

For a further rational comparisons of systems, one needs to consider the system efficiency as well. We propose the following criteria for evaluating the system efficiency, for $n \geq 3$

$$E_n(t) = \frac{In_n(t)}{In_n(t)+Co_n(t)}, \quad t > 0. \tag{11}$$

By (11), it is obvious that $0 \leq E_n(t) \leq 1$ for $t > 0$. Also, $E_n(t)$ tends to 1 and 0 as $Co_n(t)$ tends to 0 and $+\infty$, respectively. For achieving the simple computations, we consider the lower bound profit function using this strategy for the reliability of system and its standby components in the next subsection.

3.2 Upper and lower bounds

As n is increased, the main formulas (8) and (11) become too complex to follow. In order to consider the best number of components in the system, we can provide a simple lower bound for the profit function. First, we obtain upper bound for reliability function of S_k, $k \geq 3$. For $k = 3$, we have $\bar{F}_{S_3}(u) = [\bar{F}_T(u)]^2$. According to the basic probability and convolution theorem, from (3), we obtain

$$\begin{aligned}
\bar{F}_{S_4}(x) &= P(\min\{\max\{T_1,T_2\},T_3+S_3\} > x) \\
&\leq P(T_3+S_3 > x) \\
&= 1 - \int_{t_3=0}^{x} P(T_3+S_3 \leq x | T_3 = t_3) dF_{T_3}(t_3) \\
&= \bar{F}_T(x) + \int_{t_3=0}^{x} \bar{F}_T^2(x-t_3) f_T(t_3) dt_3.
\end{aligned} \tag{12}$$

Similarly from (4), we obtain

$$\begin{aligned}
\bar{F}_{S_5}(x) &= P(\min\{\max\{T_1,T_2,T_3+S_3\},T_4+S_4\} > x) \\
&\leq P(T_4+S_4 > x) \\
&= 1 - \int_{t_4=0}^{x} P(T_4+S_4 \leq x | T_4 = t_4) dF_{T_4}(t_4) \\
&= \bar{F}_T(x) + \int_{t_4=0}^{x} \bar{F}_{S_4}(x-t_4) f_T(t_4) dt_4 \\
&\leq \bar{F}_T(x) + \int_{t_4=0}^{x} \left[\bar{F}_T(x-t_4) + \int_{t_3=0}^{x-t_4} \bar{F}_T^2(x-t_3-t_4) f_T(t_3) dt_3 \right] f_T(t_4) dt_4 \\
&= \bar{F}_T(x) + \int_{t_4=0}^{x} \bar{F}_T(x-t_4) f_T(t_4) dt_4 + \int_{t_4=0}^{x} \int_{t_3=0}^{x-t_4} \bar{F}_T^2(x-t_3-t_4) f_T(t_3) f_T(t_4) dt_3 dt_4,
\end{aligned} \tag{13}$$

where second inequality in (13) is obtained by (12). Also, using (13), an upper bound for the reliability function of S_6 and S_7 are given by

$$\begin{aligned}\bar{F}_{S_6}(x) &\le P(T_5+S_5>x)\\ &\le \bar{F}_T(x)+\int_{t_5=0}^{x}\Bigg[\bar{F}_T(x-t_5)+\int_{t_4=0}^{x-t_5}[\bar{F}_T(x-t_5-t_4)\\ &\qquad+\int_{t_3=0}^{x-t_4-t_5}\bar{F}_T^2(x-t_3-t_4-t_5)f_T(t_3)dt_3\Bigg]f_T(t_4)dt_4\Bigg]f_T(t_5)dt_5.\end{aligned}$$

and

$$\begin{aligned}\bar{F}_{S_7}(x) &\le P(T_6+S_6>x)\\ &\le \bar{F}_T(x)+\int_{t_6=0}^{x}\Bigg[\bar{F}_T(x-t_6)+\int_{t_5=0}^{x-t_6}\Bigg[\bar{F}_T(x-t_6-t_5)+\int_{t_4=0}^{x-t_5-t_6}[\bar{F}_T(x-t_6-t_5-t_4)\\ &\qquad+\int_{t_3=0}^{x-t_4-t_5-t_6}\bar{F}_T^2(x-t_3-t_4-t_5-t_6)f_T(t_3)dt_3\Bigg]f_T(t_4)dt_4\Bigg]f_T(t_5)dt_5\Bigg]f_T(t_6)dt_6,\end{aligned}$$

respectively. Continuing the same process, from (5), an explicit possible upper bound for $\bar{F}_{S_k}$ can be achieved which is stated in the next lemma. For convenience of notations, throughout the paper, we suppose that $\sum_{r=i}^{j}a_r=0$ for $i>j$.

Lemma 1. *Suppose the assumptions of the proposed model hold. Then, a possible upper bound for the reliability of the mth standby component, $4\le m\le n$, is achieved by*

$$\begin{aligned}\bar{F}_{S_m}(x) &\le \bar{F}_{S_{m-1}+T_{m-1}}(x)\\ &\le \bar{F}_T(x)+\int_{t_{m-1}=0}^{x}\Bigg[\bar{F}_T(x-t_{m-1})+\int_{t_{m-2}=0}^{x-t_{m-1}}[\bar{F}_T(x-t_{m-1}-t_{m-2})+\cdots\\ &+\int_{t_4=0}^{x-\sum_{i=5}^{m-1}t_i}\Bigg[\bar{F}_T\left(x-\sum_{i=4}^{m-1}t_i\right)\\ &+\int_{t_3=0}^{x-\sum_{i=4}^{m-1}t_i}\bar{F}_T^2\left(x-\sum_{i=3}^{m-1}t_i\right)f_T(t_3)dt_3\Bigg]f_T(t_4)dt_4\cdots\Bigg]f_T(t_{m-2})dt_{m-2}\Bigg]f_T(t_{m-1})dt_{m-1}. \qquad (14)\end{aligned}$$

As T_j+S_j is stochastically smaller than $\max\{T_1,T_2,T_3+S_3,\ldots,T_n+S_n\}$, for $1\le j\le n$, we have the following possible lower bound for the system reliability:

$$\begin{aligned}\bar{F}_{sys,n}(x) &= P(\max\{T_1,T_2,T_3+S_3,\ldots,T_n+S_n\}>x)\\ &\ge P(T_j+S_j>x),\quad j=1,2,...,n.\end{aligned}$$

Especially,

$$\bar{F}_{sys,n}(x)\ge\bar{F}_T(x). \qquad (15)$$

Also, using Bonferroni inequality and Lemma 1, we have

$$\begin{aligned}
\bar{F}_{sys,n}(x) &= 1 - P(\max\{T_1, T_2, T_3 + S_3, \ldots, T_n + S_n\} \leq x) \\
&\leq \sum_{m=1}^{n} \bar{F}_{T_m+S_m}(x) \\
&\leq n\bar{F}_T(x) \\
&+ \sum_{m=3}^{n} \left[\int_{t_m=0}^{x} \left[\bar{F}_T(x - t_m) + \int_{t_{m-1}=0}^{x-t_m} [\bar{F}_T(x - t_m - t_{m-1}) + \cdots \right. \right. \\
&+ \int_{t_4=0}^{x-\sum_{i=5}^{m} t_i} \left[\bar{F}_T \left(x - \sum_{i=4}^{m} t_i \right) \right. \\
&\left. \left. \left. + \int_{t_3=0}^{x-\sum_{i=4}^{m} t_i} \bar{F}_T^2 \left(x - \sum_{i=3}^{m} t_i \right) f_T(t_3) dt_3 \right] f_T(t_4) dt_4 \cdots \right] f_T(t_{m-1}) dt_{m-1} \right] f_T(t_m) dt_m \Bigg]. \qquad (16)
\end{aligned}$$

where $S_1 = S_2 = 0$.

Utilizing the upper bound for the kth component survival in (14) and the lower bound for system survival in (15), a possible lower bound for the profit function of system in (8) is achieved which is stated in the following proposition.

Proposition 1. *A possible lower bound for the profit function of parallel* $(2, n-2)$ *system,* $n \geq 4$*, at time* $t > 0$ *is obtained by*

$$\begin{aligned}
B_n(t) &= k_1 \int_{u=0}^{t} \bar{F}_{sys,n}(u) du - k_2 \sum_{m=3}^{n} \int_{u=0}^{t} \bar{F}_{S_m}(u) du \\
&\geq k_1 \int_0^t \bar{F}_T(u) du - k_2 \int_0^t \bar{F}_T^2(u) du - k_2 (n-3) \int_0^t \bar{F}_T(u) du \\
&- k_2 \sum_{m=4}^{n} \int_{u=0}^{t} \left[\int_{t_{m-1}=0}^{u} \left[\bar{F}_T(u - t_{m-1}) + \int_{t_{m-2}=0}^{u-t_{m-1}} [\bar{F}_T(u - t_{m-1} - t_{m-2}) + \cdots \right. \right. \\
&+ \int_{t_4=0}^{u-\sum_{i=5}^{m-1} t_i} \left[\bar{F}_T \left(u - \sum_{i=4}^{m-1} t_i \right) \right. \\
&\left. \left. \left. + \int_{t_3=0}^{u-\sum_{i=4}^{m-1} t_i} \bar{F}_T^2 \left(u - \sum_{i=3}^{m-1} t_i \right) f_T(t_3) dt_3 \right] f_T(t_4) dt_4 \cdots \right] f_T(t_{m-2}) dt_{m-2} \right] f_T(t_{m-1}) dt_{m-1} \Bigg] du.
\end{aligned}$$

It is observed that the effect of dependence parameter is not appeared in the lower bound of profit function given in Proposition 1.

Using the inequalities in (14), (15) and (16), we obtain a lower bound for the propose system efficiency evaluation in (11), which is stated in the next proposition.

Proposition 2. *A lower bound for the propose system efficiency evaluation in (11) of parallel*

$(2,n-2)$ *system,* $n \geq 3$*, at time* $t > 0$ *is given by*

$$\begin{aligned} E_n(t) &= \frac{In_n(t)}{In_n(t)+Co_n(t)} \\ &= \frac{k_1 \int_{x=0}^{t} \bar{F}_{sys,n}(x)dx}{k_1 \int_{x=0}^{t} \bar{F}_{sys,n}(x)dx + k_2 \sum_{m=3}^{n} \int_{x=0}^{t} \bar{F}_{S_m}(x)dx} \\ &\geq \frac{k_1 \int_{x=0}^{t} \bar{F}_T(x)dx}{\int_{x=0}^{t} \left[(nk_1+k_2(n-3))\bar{F}_T(x)+k_2[\bar{F}_T(x)]^2+(k_1+k_2)\sum_{m=3}^{n-1}\phi_m^F(x)+k_1\phi_n^F(x)\right]dx} \\ &= \left[n+(n-3)\frac{k_2}{k_1}+\frac{\int_{x=0}^{t}\left[k_2[\bar{F}_T(x)]^2+(k_1+k_2)\sum_{m=3}^{n-1}\phi_m^F(x)+k_1\phi_n^F(x)\right]dx}{k_1\int_{x=0}^{t}\bar{F}_T(x)dx}\right]^{-1} \end{aligned} \tag{17}$$

where

$$\begin{aligned} \phi_m^F(x) = \int_{t_m=0}^{x} \Bigg[\bar{F}_T(x-t_m) + \int_{t_{m-1}=0}^{x-t_m} \Bigg[\bar{F}_T(x-t_m-t_{m-1}) + \cdots + \int_{t_4=0}^{x-\sum_{i=5}^{m} t_i} \Bigg[\bar{F}_T\left(x-\sum_{i=4}^{m} t_i\right) \\ + \int_{t_3=0}^{x-\sum_{i=4}^{m} t_i} \bar{F}_T^2\left(x-\sum_{i=3}^{m} t_i\right) f_T(t_3)dt_3\Bigg] f_T(t_4)dt_4 \cdots\Bigg] f_T(t_{m-1})dt_{m-1}\Bigg] f_T(t_m)dt_m. \end{aligned} \tag{18}$$

For special case $n = 3$, by Proposition 2, we have an explicit simple expression for the bound of $E_3(t)$, which is stated below

$$\begin{aligned} E_3(t) &= \frac{In_3(t)}{In_3(t)+Co_3(t)} \\ &\geq \frac{k_1 \int_{x=0}^{t} \bar{F}_T(x)dx}{\int_{x=0}^{t}\left[3k_1\bar{F}_T(x)+k_2[\bar{F}_T(x)]^2+k_1\phi_3^F(x)\right]dx} \\ &= \left[3+\frac{1}{k_1\int_{x=0}^{t}\bar{F}_T(x)dx}\int_{x=0}^{t}\left(k_2[\bar{F}_T(x)]^2+k_1\int_{t_3=0}^{x}\bar{F}_T^2(x-t_3)f_T(t_3)dt_3\right)dx\right]^{-1}. \end{aligned} \tag{19}$$

The next section provides some examples for the values of profit and efficiency functions.

4 Illustrative examples

In this section, the results illustrated by two examples, in first example we obtain exact expression and in the second we find a lower bound for $E_n(t)$. The FGM copula model is considered the positive and negative dependencies between components lifetimes and the marginal distributions are assumed to be exponential.

Example 1. (Exact expression for $E_n(t)$)

Let components lifetimes be exponentially distributed, i.e. $F_T(x) = 1-e^{-\lambda x}$, $x > 0$. As $S_3 = \min\{T_1, T_2\}$, then we have

$$\bar{F}_{S_3}(x) = [\bar{F}_T(x)]^2 = e^{-2\lambda x}, \tag{20}$$

also from (7),

$$F_{T_3+S_3}(x) = \int_0^x F_{S_3}(x-u) f_T(u) du$$
$$= 1 + \mathrm{e}^{-2\lambda x} - 2\mathrm{e}^{-\lambda x} = \left(1 - \mathrm{e}^{-\lambda x}\right)^2. \tag{21}$$

Utilizing FGM 3-copula model as the dependence structure of units lifetimes, from (6) we reach the following expressions for the system reliability

$$\bar{F}_{sys,3}(x) = 1 - C(F_T(x), F_T(x), F_{T_3+S_3}(x))$$
$$= 1 - F_T^2(x) F_{T_3+S_3}(x) \left[1 + \alpha_1 \bar{F}_T^2(x) + \alpha_2 \bar{F}_T(x) \bar{F}_{T_3+S_3}(x) + \alpha_3 \bar{F}_T(x) \bar{F}_{T_3+S_3}(x) + \alpha_4 \bar{F}_T^2(x) \bar{F}_{T_3+S_3}(x)\right].$$

Since by assumption T_1 and T_2 are independent, $\alpha_1 = 0$. By assuming $\alpha = \alpha_2 = \alpha_3 = \alpha_4$ for simplicity, then we have

$$\bar{F}_{sys,3}(x) = 1 - F_T^2(x) F_{T_3+S_3}(x) \left[1 + \alpha(2\bar{F}_T(x) \bar{F}_{T_3+S_3}(x) + \bar{F}_T^2(x) \bar{F}_{T_3+S_3}(x))\right]. \tag{22}$$

Substituting (20) and (21) into (22), we get

$$\bar{F}_{sys,3}(x) = \mathrm{e}^{-\lambda x}(\mathrm{e}^{-\lambda x} - 2)\left((\mathrm{e}^{-\lambda x} - 1)^2(\mathrm{e}^{-4\lambda x}\alpha - 4\mathrm{e}^{-2\lambda x}\alpha - 1) - 1\right). \tag{23}$$

From (9) and (23), we obtain

$$In_3(t) = k_1 \int_0^t \bar{F}_{sys,3}(x) dx$$
$$= k_1 \int_0^t \mathrm{e}^{-\lambda x}(\mathrm{e}^{-\lambda x} - 2)\left((\mathrm{e}^{-\lambda x} - 1)^2(\mathrm{e}^{-4\lambda x}\alpha - 4\mathrm{e}^{-2\lambda x}\alpha - 1) - 1\right) dx$$
$$= \frac{k_1}{840\lambda}\left[1750 + 157\alpha - 3360\mathrm{e}^{-\lambda t} + 2520\mathrm{e}^{-2\lambda t} - 1120\mathrm{e}^{-3\lambda t}(2\alpha + 1)\right.$$
$$\left. + 210\mathrm{e}^{-4\lambda t}(20\alpha + 1) - 2352\alpha\mathrm{e}^{-5\lambda t} - 140\alpha\mathrm{e}^{-6\lambda t} + 480\alpha\mathrm{e}^{-7\lambda t} - 105\alpha\mathrm{e}^{-8\lambda t}\right]. \tag{24}$$

Also, from (10) and (20)

$$Co_3(t) = k_2 \int_0^t \bar{F}_{S_3}(x) dx = \frac{k_2}{2\lambda}\left(1 - \mathrm{e}^{-2\lambda t}\right). \tag{25}$$

Take $A_3(t;\alpha,\lambda) = \frac{In_3(t)}{Co_3(t)}$, then by substituting (24) and (25), we have

$$A_3(t;\alpha,\lambda) = \frac{In_3(t)}{Co_3(t)}$$
$$= \frac{\lambda k_1}{420\lambda k_2\left(1 - \mathrm{e}^{-2\lambda t}\right)}\left[1750 + 157\alpha - 3360\mathrm{e}^{-\lambda t} + 2520\mathrm{e}^{-2\lambda t} - 1120\mathrm{e}^{-3\lambda t}(2\alpha + 1)\right.$$
$$\left. + 210\mathrm{e}^{-4\lambda t}(20\alpha + 1) - 2352\alpha\mathrm{e}^{-5\lambda t} - 140\alpha\mathrm{e}^{-6\lambda t} + 480\alpha\mathrm{e}^{-7\lambda t} - 105\alpha\mathrm{e}^{-8\lambda t}\right]. \tag{26}$$

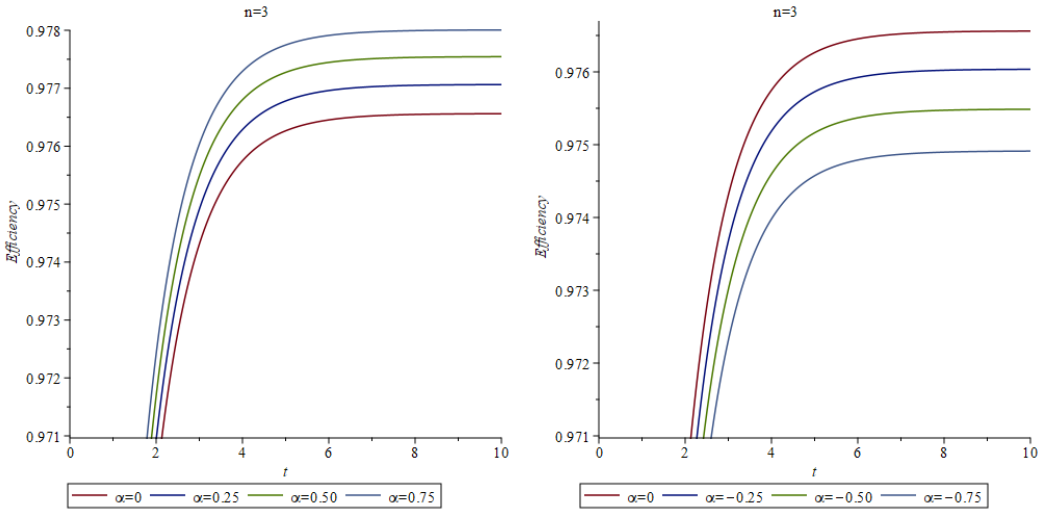


Figure 1: The plot of $E_3(t)$ for $k_1 = 10$, $k_2 = 1$, $\lambda = 1$ and some selected values of α.

Then, from (11) and (26), we can obtain the system efficiency for $n = 3$ as a function of t, α, λ, k_1 and k_2. The plot of $E_3(t)$ has been displayed in Figure 1 for some selected values of α. From Figure 1, it is observed that $E_3(t)$ is increasing in t and the results indicate that it is also increasing in α.

Now we obtain expression for $E_4(t)$. According to the properties of survival copula and using FGM 2-copula model as the dependence structure of units lifetimes, for $n = 4$ we have

$$\begin{aligned}\bar{F}_{S_4}(x) &= \hat{C}(\bar{F}_{\max\{T_1,T_2\}}(x), \bar{F}_{T_3+S_3}(x)) \\ &= \bar{F}_{\max\{T_1,T_2\}}(x) + \bar{F}_{T_3+S_3}(x) - 1 + C(1-\bar{F}_{\max\{T_1,T_2\}}(x), 1-\bar{F}_{T_3+S_3}(x)) \\ &= \bar{F}_{\max\{T_1,T_2\}}(x) + \bar{F}_{T_3+S_3}(x) - 1 + F_{\max\{T_1,T_2\}}(x)F_{T_3+S_3}(x)\left[1+\alpha\bar{F}_{\max\{T_1,T_2\}}(x)\bar{F}_{T_3+S_3}(x)\right]. \quad (27)\end{aligned}$$

From (21) and (27) after simplification, we get

$$\bar{F}_{S_4}(x) = \mathrm{e}^{-2\lambda x}\left[\alpha\left(\mathrm{e}^{-6\lambda x} - 8\mathrm{e}^{-5\lambda x} + 26\mathrm{e}^{-4\lambda x} - 44\mathrm{e}^{-3\lambda x} + 41\mathrm{e}^{-2\lambda x} - 20\mathrm{e}^{-\lambda x} + 4\right) + \left(2-\mathrm{e}^{-\lambda x}\right)^2\right]. \quad (28)$$

From (7) and (28), we obtain

$$\begin{aligned}
F_{T_4+S_4}(x) &= \int_0^x F_{S_4}(x-u) f_T(u) du \\
&= \int_0^u \lambda \Big[(41\alpha+1)\,\mathrm{e}^{-4\lambda(x-u)} - (20\alpha+4)\,\mathrm{e}^{-3\lambda(x-u)} + (4\alpha+4)\,\mathrm{e}^{-2\lambda(x-u)} \\
&\quad + \alpha\left(\mathrm{e}^{-8\lambda(x-u)} - 8\mathrm{e}^{-7\lambda(x-u)} + 26\mathrm{e}^{-6\lambda(x-u)} - 44\mathrm{e}^{-5\lambda(x-u)}\right)\Big] \mathrm{e}^{-\lambda u} du \\
&= \alpha \mathrm{e}^{-\lambda x}\left(-\frac{1}{7}\mathrm{e}^{-7\lambda x} + \frac{28}{21}\mathrm{e}^{-6\lambda x} - \frac{182}{35}\mathrm{e}^{-5\lambda x} + 11\mathrm{e}^{-4\lambda x} + \frac{1435}{105}\mathrm{e}^{-3\lambda x} - 10\mathrm{e}^{-2\lambda x} + 4\mathrm{e}^{-\lambda x} - \frac{71}{105}\right) \\
&\quad + \frac{2}{3}\mathrm{e}^{-4\lambda x} - \mathrm{e}^{-2\lambda x}\left(2-\mathrm{e}^{-\lambda x}\right)^2. \qquad (29)
\end{aligned}$$

From (2) and FGM 4-copula, we acquire the following formula for

$$\begin{aligned}
\bar{F}_{sys,4}(x) &= 1 - C(F_T(x), F_T(x), F_{T_3+S_3}(x), F_{T_4+S_4}(x)) \\
&= 1 - F_T^2(x) F_{T_3+S_3}(x) F_{T_4+S_4}(x)\Big[1 + \alpha_1 \bar{F}_T^2(x) + \alpha_2 \bar{F}_T(x)\bar{F}_{T_3+S_3}(x) + \alpha_3 \bar{F}_T(x)\bar{F}_{T_3+S_3}(x) \\
&+ \alpha_4 \bar{F}_T(x)\bar{F}_{T_4+S_4}(x) + \alpha_5 \bar{F}_T(x)\bar{F}_{T_4+S_4}(x) + \alpha_6 \bar{F}_{T_3+S_3}(x)\bar{F}_{T_4+S_4}(x) + \alpha_7 \bar{F}_T^2(x)\bar{F}_{T_3+S_3}(x) \\
&+ \alpha_8 \bar{F}_T^2(x)\bar{F}_{T_4+S_4}(x) + \alpha_9 \bar{F}_T(x)\bar{F}_{T_3+S_3}(x)\bar{F}_{T_4+S_4}(x) + \alpha_{10}\bar{F}_T(x)\bar{F}_{T_3+S_3}(x)\bar{F}_{T_4+S_4}(x) \\
&+ \alpha_{11}\bar{F}_T^2(x)\bar{F}_{T_3+S_3}(x)\bar{F}_{T_4+S_4}(x)\Big]. \qquad (30)
\end{aligned}$$

By assumption T_1 and T_2 are independent (i.e. $\alpha_1 = 0$) and let us take $\alpha = \alpha_2 = \alpha_3 = \cdots = \alpha_{11}$ for simplicity, then substituting (28) and (29) into (30), we get

$$\begin{aligned}
\bar{F}_{sys,4}(x) &= 1 - \left(\frac{1}{49}\mathrm{e}^{-3\lambda x} - \frac{68}{147}\mathrm{e}^{-2\lambda x} - \frac{10646}{2205}\mathrm{e}^{-\lambda x} - \frac{13522}{441}\right)\alpha^3 \mathrm{e}^{-21\lambda x} \\
&+ \frac{\mathrm{e}^{-\lambda x}}{11025}\Big[-(1451471\alpha+1050)\,\alpha^2\mathrm{e}^{-19\lambda x} + (4412444\alpha+20300)\,\alpha^2\mathrm{e}^{-18\lambda x} \\
&-(9599217\alpha+176120)\,\alpha^2\mathrm{e}^{-17\lambda x} + (14510782\alpha+897400)\,\alpha^2\mathrm{e}^{-16\lambda x} \\
&-\left(13443613\alpha^2+2939510\alpha+1225\right)\alpha\mathrm{e}^{-15\lambda x} + \left(3264300\alpha^2+6296080\alpha+19600\right)\alpha\mathrm{e}^{-14\lambda x} \\
&+\left(8178235\alpha^2-8254750\alpha-134750\right)\alpha\mathrm{e}^{-13\lambda x} - \left(4024202\alpha^2-4416090\alpha-507150\right)\alpha\mathrm{e}^{-12\lambda x} \\
&-\left(24299249\alpha^2-4633930\alpha+1080100\right)\alpha\mathrm{e}^{-11\lambda x} + \left(62839976\alpha^2-9071580\alpha+1081500\right)\alpha\mathrm{e}^{-10\lambda x} \\
&-\Big[\left(85278587\alpha^3+794360\alpha^2-335055\alpha\right)\mathrm{e}^{-\lambda x} - 78298098\alpha^2 - 20302100\alpha + 2194395\Big]\alpha\mathrm{e}^{-8\lambda x} \\
&-\Big[\left(51848470\alpha^3+33413170\alpha^2-2201430\alpha-3675\right)\mathrm{e}^{-\lambda x} - 25063072\alpha^3 - 31040240\alpha^2 + 224420\alpha + 36750\Big]\mathrm{e}^{-6\lambda x} \\
&-\Big[\left(8690913\alpha^3+19047910\alpha^2+1435945\alpha-154350\right)\mathrm{e}^{-\lambda x} - 2058008\alpha^3 - 7960330\alpha^2 - 1564220\alpha + 349125\Big]\mathrm{e}^{-4\lambda x} \\
&-\Big[\left(299052\alpha^3+2213400\alpha^2+892605\alpha-459375\right)\mathrm{e}^{-\lambda x} - 20164\alpha^3 - 375200\alpha^2 - 290920\alpha + 352800\Big]\mathrm{e}^{-2\lambda x} \\
&-\left(29820\alpha^2+28980\alpha-147000\right)\mathrm{e}^{-\lambda x} - (7455\alpha+25725)\Big]. \qquad (31)
\end{aligned}$$

Summing-up, from (20) and (28), one can obtain an expression for $Co_4(t)$, also, from (31), we can find $In_4(t)$. Consequently, substituting in (11), we have the system efficiency for $n = 4$ as a

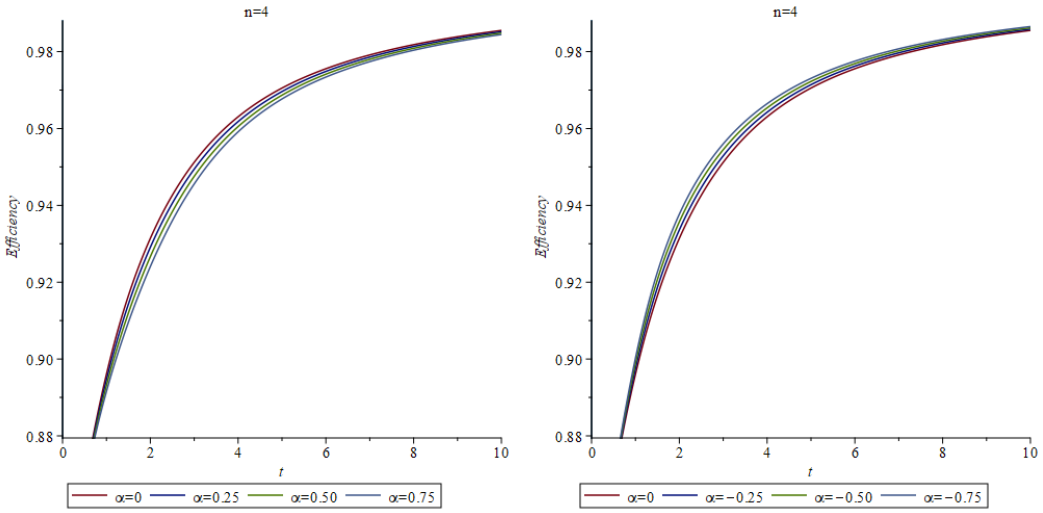


Figure 2: The plot of $E_4(t)$ for $k_1 = 10$, $k_2 = 1$, $\lambda = 1$ and some selected values of α.

function of t, α, λ, k_1 and k_2. The plot of $E_4(t)$ has been displayed in Figure 2 for some selected values of α. From Figure 2, it is observed that $E_4(t)$ is also increasing in t and decreasing in α. Also, Figures 1 and 2 indicate that $E_n(t)$ decreases as n increases.

By continuing the same process step by step, we can get an exact explicit for $E_n(t)$ when $n \geq 5$.

Example 2. (Lower bound for $E_n(t)$)

Assume that components lifetimes are exponentially distributed as $F_T(x) = 1 - \mathrm{e}^{-\lambda x}$, $x > 0$. To obtain the lower bound of $E_n(t)$, denoted by $LE_n(t)$, as appeared in (17), we should first compute $\phi_m^F(x)$ for $3 \leq m \leq n$. Let $n = 3$, then from (18), we have

$$\begin{aligned}\phi_3^F(x) &= \int_{t_3=0}^{x} \bar{F}_T^2(x - t_3) f_T(t_3) dt_3 \\ &= \int_{t_3=0}^{x} \mathrm{e}^{-2\lambda(x-t_3)} \lambda \mathrm{e}^{-\lambda t_3} dt_3 \\ &= \mathrm{e}^{-\lambda x} - \mathrm{e}^{-2\lambda x}. \end{aligned} \tag{32}$$

From (32) and (19), we have a lower bound for $E_3(t)$, as

$$LE_3(t) = \frac{2k_1(1 - e^{-\lambda t})}{7k_1 + k_2 - 8k_1 e^{-\lambda} + (k_1 - k_2)e^{-2\lambda}}. \tag{33}$$

Similarly, from (18), we have

$$\begin{aligned}\phi_4^F(x) &= \int_{t_4=0}^{x}\left[\bar{F}_T(x-t_4)+\int_{t_3=0}^{x-t_4}\bar{F}_T^2(x-t_4-t_3)f_T(t_3)dt_3\right]f_T(t_4)dt_4\\ &= \int_{t_4=0}^{x}\left[\mathrm{e}^{-\lambda(x-t_4)}+\int_{t_3=0}^{x-t_4}\mathrm{e}^{-2\lambda(x-t_4-t_3)}\lambda\mathrm{e}^{-\lambda t_3}dt_3\right]\lambda\mathrm{e}^{-\lambda t_4}dt_4\\ &= \mathrm{e}^{-2\lambda x}+2\lambda x\mathrm{e}^{-\lambda x}-\mathrm{e}^{-\lambda x}. \end{aligned} \tag{34}$$

From (17), for $n=4$, we have

$$E_4(t) \geq \frac{k_1\int_{x=0}^{t}\bar{F}_T(x)dx}{\int_{x=0}^{t}\left[(4k_1+k_2(4-3))\bar{F}_T(x)+k_2[\bar{F}_T(x)]^2+(k_1+k_2)\phi_3^F(x)+k_1\phi_4^F(x)\right]dx} \tag{35}$$

Upon substituting (32) and (34) into (35), we obtain a lower bound for $E_4(t)$, as

$$LE_4(t) = \frac{k_1\left(1-\mathrm{e}^{-\lambda t}\right)}{-2\left(3k_1+k_2\right)+k_1\lambda t)\mathrm{e}^{-\lambda t}+2(3k_1+k_2)}.$$

Similarly, from (18), we have

$$\begin{aligned}\phi_5^F(x) &= \int_{t_5=0}^{x}\left[\bar{F}_T(x-t_5)+\int_{t_4=0}^{x-t_5}\left[\bar{F}_T(x-t_5-t_4)+\int_{t_3=0}^{x-t_4-t_5}\bar{F}_T^2(x-t_5-t_4-t_3)f_T(t_3)dt_3\right]f_T(t_4)dt_4\right]f_T(t_5)dt_5\\ &= \int_{t_4=0}^{x}\left[\mathrm{e}^{-\lambda(x-t_5)}+\int_{t_3=0}^{x-t_5}\left[\mathrm{e}^{-\lambda(x-t_5-t_4)}+\int_{t_3=0}^{x-t_4-t_5}\mathrm{e}^{-2\lambda(x-t_5-t_4-t_3)}\lambda\mathrm{e}^{-\lambda t_3}dt_3\right]\lambda\mathrm{e}^{-\lambda t_4}dt_4\right]\lambda\mathrm{e}^{-\lambda t_5}dt_5\\ &= -\mathrm{e}^{-2\lambda x}+(1+\lambda^2x^2)\mathrm{e}^{-\lambda x}, \end{aligned} \tag{36}$$

and

$$\begin{aligned}\phi_6^F(x) &= \int_{t_6=0}^{x}\left[\bar{F}_T(x-t_6)+\int_{t_5=0}^{x-t_6}\left[\bar{F}_T(x-t_6-t_5)+\int_{t_4=0}^{x-t_5-t_6}\left[\bar{F}_T(x-t_6-t_5-t_4)\right.\right.\right.\\ &\quad\left.\left.\left.+\int_{t_3=0}^{x-t_4-t_5-t_6}\bar{F}_T^2(x-t_6-t_5-t_4-t_3)f_T(t_3)dt_3\right]f_T(t_4)dt_4\right]f_T(t_5)dt_5\right]f_T(t_6)dt_6\\ &= \int_{t_6=0}^{x}\left[\mathrm{e}^{-\lambda(x-t_6)}+\int_{t_5=0}^{x-t_6}\left[\mathrm{e}^{-\lambda(x-t_6-t_5)}+\int_{t_4=0}^{x-t_5-t_6}\left[\mathrm{e}^{-\lambda(x-t_6-t_5-t_4)}\right.\right.\right.\\ &\quad\left.\left.\left.+\int_{t_3=0}^{x-t_4-t_5-t_6}\mathrm{e}^{-2\lambda(x-t_6-t_5-t_4-t_3)}\lambda\mathrm{e}^{-\lambda t_3}dt_3\right]\lambda\mathrm{e}^{-\lambda t_4}dt_4\right]\lambda\mathrm{e}^{-\lambda t_5}dt_5\right]\lambda\mathrm{e}^{-\lambda t_6}dt_6\\ &= \mathrm{e}^{-2\lambda x}+(\frac{1}{3}\lambda^3x^3+2\lambda x-1)\mathrm{e}^{-\lambda x}. \end{aligned} \tag{37}$$

Also

$$\begin{aligned}\phi_7^F(x) &= \int_{t_7=0}^{x}\Bigg[\bar{F}_T(x-t_7)+\int_{t_6=0}^{x-t_7}\Bigg[\bar{F}_T(x-t_7-t_6)+\int_{t_5=0}^{x-t_6-t_7}\Bigg[\bar{F}_T(x-t_7-t_6-t_5)\\ &+\int_{t_4=0}^{x-t_5-t_6-t_7}\Bigg[\bar{F}_T(x-t_7-t_6-t_5-t_4)+\int_{t_3=0}^{x-t_4-t_5-t_6-t_7}\bar{F}_T^2(x-t_7-t_6-t_5-t_4-t_3)f_T(t_3)dt_3\Bigg]\\ &f_T(t_4)dt_4\Bigg]f_T(t_5)dt_5\Bigg]f_T(t_6)dt_6\Bigg]f_T(t_7)dt_7\\ &= \int_{t_7=0}^{x}\Bigg[\mathrm{e}^{-\lambda(x-t_7)}+\int_{t_6=0}^{x-t_7}\Bigg[\mathrm{e}^{-\lambda(x-t_7-t_6)}+\int_{t_5=0}^{x-t_6-t_7}\Bigg[\mathrm{e}^{-\lambda(x-t_7-t_6-t_5)}\\ &+\int_{t_4=0}^{x-t_5-t_6-t_7}\mathrm{e}^{-\lambda(x-t_7-t_6-t_5-t_4)}+\int_{t_3=0}^{x-t_4-t_5-t_6-t_7}\mathrm{e}^{-2\lambda(x-t_7-t_6-t_5-t_4-t_3)}\lambda\mathrm{e}^{-\lambda t_3}dt_3\Bigg]\lambda\mathrm{e}^{-\lambda t_4}dt_4\Bigg]\\ &\lambda\mathrm{e}^{-\lambda t_5}dt_5\Bigg]\lambda\mathrm{e}^{-\lambda t_6}dt_6\Bigg]\lambda\mathrm{e}^{-\lambda t_7}dt_7\\ &= -\mathrm{e}^{-2\lambda x}+(\frac{1}{12}\lambda^4x^4+\lambda^2x^2+1)\mathrm{e}^{-\lambda x}, \end{aligned} \tag{38}$$

and

$$\begin{aligned}\phi_8^F(x) &= \int_{t_8=0}^{x}\Bigg[\bar{F}_T(x-t_8)+\int_{t_7=0}^{x-t_8}\Bigg[\bar{F}_T(x-t_8-t_7)+\int_{t_6=0}^{x-t_7-t_8}\Bigg[\bar{F}_T(x-t_8-t_7-t_6)\\ &+\int_{t_5=0}^{x-t_6-t_7-t_8}\Bigg[\bar{F}_T(x-t_8-t_7-t_6-t_5)+\int_{t_4=0}^{x-t_5-t_6-t_7-t_8}\Bigg[\bar{F}_T(x-t_8-t_7-t_6-t_5-t_4)\\ &+\int_{t_3=0}^{x-t_4-t_5-t_6-t_7-t_8}\bar{F}_T^2(x-t_8-t_7-t_6-t_5-t_4-t_3)f_T(t_3)dt_3\Bigg]f_T(t_4)dt_4\Bigg]f_T(t_5)dt_5\Bigg]f_T(t_6)dt_6\Bigg]\\ &f_T(t_7)dt_7\Bigg]f_T(t_8)dt_8\\ &= \mathrm{e}^{-2\lambda x}+\left(\frac{1}{60}\lambda^5x^5+\frac{1}{3}\lambda^3x^3+2\lambda x-1\right)\mathrm{e}^{-\lambda x}. \end{aligned} \tag{39}$$

Consequently, using (32), (34), (36), (37), (38) and (39) and also (17), we obtain a lower bound for $E_n(t), n=5,6,7,8$ which are given by

$$LE_5(t) = \frac{2k_1\left(\mathrm{e}^{-\lambda t}-1\right)}{\left(\left(2\lambda^2t^2+8\lambda t+20\right)k_1+\left(4\lambda t+8\right)k_2\right)\mathrm{e}^{-\lambda t}+\left(k_2-k_1\right)\mathrm{e}^{-2\lambda t}-19k_1-9k_2},$$

$$LE_6(t) = \frac{3k_1\left(\mathrm{e}^{-\lambda t}-1\right)}{\left(k_1\lambda^3t^3+3\lambda^2t^2\left(2k_1+k_2\right)+12\lambda t\left(2k_1+k_2\right)+42k_1+24k_2\right)\mathrm{e}^{-\lambda t}-42k_1-24k_2},$$

$$LE_7(t) = \frac{12k_1\left(\mathrm{e}^{-\lambda t}-1\right)}{\left(k_1\lambda^4t^4+4\lambda^3t^3\left(2k_1+k_2\right)+24\lambda^2t^2\left(2k_1+k_2\right)+48\lambda t\left(3k_1+2k_2\right)+240k_1+144k_2\right)\mathrm{e}^{-\lambda t}-A},$$

and

$$LE_8(t) = \frac{60k_1\left(\mathrm{e}^{-\lambda t}-1\right)}{\left(t^5\lambda^5k_1+5\lambda^4t^4\left(2k_1+k_2\right)+40\lambda^3t^3\left(2k_1+k_2\right)+120\lambda^2t^2\left(3k_1+2k_2\right)+360\lambda t\left(3k_1+2k_2\right)\right)\mathrm{e}^{-\lambda t}+B},$$

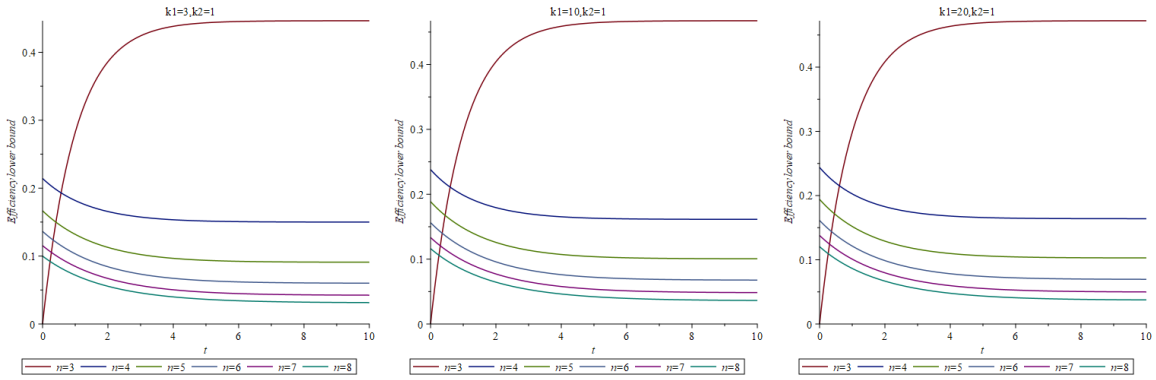


Figure 3: The plot of $LE_n(t)$ for $n = 3,...,8$, $k_2 = 1$ and some selected values of k_1.

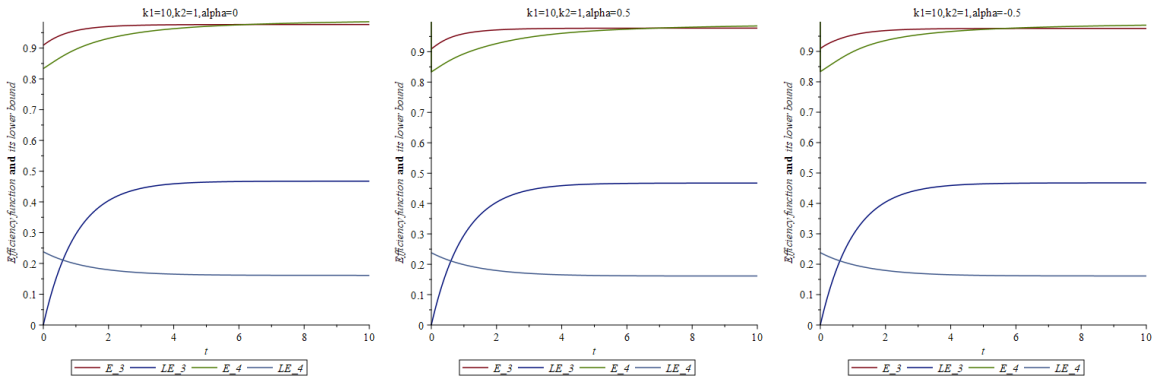


Figure 4: The plots of $E_n(t)$ and $LE_n(t)$ for $n = 3$ and 4, $k_1 = 10$, $k_2 = 1$ and some selected values of α.

respectively, where

$$A = 6(k_1 - k_2)\mathrm{e}^{-2\lambda t} + 234k_1 + 150k_2 \quad \text{and} \quad B = (1560k_1 + 1080k_2)\left(\mathrm{e}^{-\lambda t} - 1\right).$$

The plot of $LE_n(t)$ displayed in Figure 3 for $n = 3,...,8$, $k_2 = 1$ and some selected values of k_1. From this figure, it is concluded that

(i) For $n = 3$, the lower bound $LE_3(t)$ is increasing in t which is obvious analytically from (33).

(ii) For $n \geq 4$, the lower bound $LE_n(t)$ is decreasing in t and n.

(iii) With the change of k_1, the monotony behavior of the graphs is almost similar.

To compare the exact value of $E_n(t)$ and its lower bound $LE_n(t)$, for $n = 3$ and 4, their graphs are plotted in Figure 4. This figure shows that the difference between $E_3(t)$ and $LE_3(t)$ is decreasing in t while it increases for $n = 4$.

5 Conclusion

The allocation problems of cold standby components for a $(2, n-2)$ parallel system have been studied. The components are not repairable ; once a failure occurs, it has been replaced immediately with one of the cold standby components. This causes a dependency between the lifetime of the components after the first failure. So, the copula function has been used to model the mentioned dependency. A profit function and criteria for system efficiency have been defined to find the best allocation policies of one or more standby component(s). To illustrate the proposed policy, numerical computations, and graphical analyses have been presented. Numerical results indicated that one standby component optimizes the system efficiency function.

Acknowledgements

The authors would like to express sincere thanks to two anonymous reviewers, which comments and suggestions helped to improve the manuscript.

References

Amini, M., Jabbari, H., and Mohtashami Borzadaran, G. R. (2011). Aspects of dependence in generalized Farlie-Gumble-Morgenstern distributions. *Communications in Statistics - Simulation and Computation* 40 1192–1205.

Boland, P. J., El-Neweihi, E., and Proschan, F. (1991). Redundancy importance and allocation of spares in coherent systems. *Journal of Statistical Planning and Inference*, **29**, 55–65.

Boland, P. J., El Neweihi, E., and Proschan, F. (1988). Active redundancy allocation in coherent systems. *Probability in the Engineering and Informational Sciences*, **2**, 343–353.

Conway, D. A. (1983). *Farlie-Gumbel-Morgenstern Distributions*, in: Kotz S., Johnson N.L. (Eds.), Ency. Stat. Sci. 3, Wiley, New York 28–31.

Domma, F., and Giordano, S. (2013). A copula-based approach to account for dependence in stress-strength models. *Statistical Papers*, **54**, 807–826.

Johnson, N. L., and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

Gupta, R., Bajaj, C. P., and Sinha, S. M. (1986). Cost-benefit analysis of a multi-component standby system with inspection and slow switch. *Microelectronics Reliability*, **26**, 879–882.

Kong, J. S., and Frangopol, D. M. (2004). Cost-reliability interaction in life-cycle cost optimization of deteriorating structures. *Journal of Structural Engineering*, **130**, 1704–1712.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Second ed., Springer, New York.

Papageorgiou, E., and Kokolakis, G. (2004). A two-unit parallel system supported by (n-2) standbys with general and non-identical lifetimes. *International Journal of Systems Science*, **35**, 1–12.

Papageorgiou, E., and Kokolakis, G. (2007) A two-unit parallel system supported by $(n-2)$ cold standbys-analytic and simulation approach. *European Journal of Operational Research*, **176**, 11016–1032.

Ram, M., and Kumar, A. (2015). Performability analysis of a system under 1-out-of-2: G scheme with perfect reworking. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **37**, 1029–1038.

Ram, M., and Singh, S. B. (2010). Analysis of a complex system with common cause failure and two types of repair facilities with different distributions in failure. *International Journal of Reliability and Safety*, **4**, 381–392.

Ram, M., Singh, S. B., and Singh, V. V. (2013). Stochastic analysis of a standby system with waiting repair strategy. *IEEE Transactions on Systems, man, and cybernetics: Systems*, **43**, 698–707.

Schweizer, B., and Sklar, A. (1983). *Probabilistic Metric Spaces*. North-Holland, New York.

Shih, J. H., and Emura, T. (2019). Bivariate dependence measures and bivariate competing risks models under the generalized FGM copula. *Statistical Papers*, **60**, 1101–1118.

Singh, V. V., and Gulati, J. (2014). Availability and cost analysis of a standby complex system with different types of failure under waiting repair discipline using Gumbel-Hougaard Family copula distribution. *IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques* (ICICT), 180–186.

Singh, V. V., and Rawal, D. K. (2014). Availability, MTTF and Cost analysis of complex system under Preemptive resume repair policy using copula distribution. *Pakistan Journal of Statistics and Operation Research*, 299-312.

Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de I'Institute de Statistique de L'Universite de Paris*, **8**, 229–231.

Yongjin, Z., Youchao, S., Longbiao, L., and Ming, Z. (2018). Copula-based reliability analysis for a parallel system with a cold standby. *Communications in Statistics-Theory and Methods*, **47**, 562–582.

Zhang, J., and Zhang, Y. (2022). A copula-based approach on optimal allocation of hot standbys in series systems. *Naval Research Logistic*, **69**, 902–913.

Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 171–189. Research Paper

SMPS

Varinaccuracy of ranked set sample

Manoj Chacko[†] and Varghese George[‡*]

[†]*Department of Statistics, University of Kerala, Thiruvanthapuram 695 581, India*
[‡]*Department of Statistics, University of Kerala, Thiruvanthapuram 695 581, India*
[‡]*Department of Statistics, St. Stephen's College, Pathanapuram, Kerala, 689 695, India*
Email(s): manojchacko02@gmail.com, varghesessc@gmail.com

Abstract. In this paper, Kerridge's inaccuracy measure and varinaccuracy of the ranked set sample (RSS) are considered. By deriving the expression for Kerridge's inaccuracy measure and varinaccuracy of the rth order statistic, the expression for Kerridge's inaccuracy measure and varinaccuracy of RSS are obtained. Kerridge's inaccuracy measure and varinaccuracy of moving ranked set sample are also obtained.

Keywords: Inaccuracy measure; Order statistics; Ranked set sampling; Varinaccuracy.

1 Introduction

McIntyre (1952) introduced a sampling scheme named ranked set sampling (RSS) as a process of improving the precision of the sample mean as an estimate of the population mean. In some situations, the measurement of the variable of interest is costly and/or time-consuming, but the ranking of variables related to the study variable can be easily done by a judgment method (see Chen et al. (2004)). The procedure of ranked set sampling involves randomly choosing n set of units, each of size n from a population, and then the units in each set are ranked using some inexpensive methods. Then from the first set of n units, select the unit that has the lowest rank. From the second set of n units, select the unit with the second lowest rank. The process continues until the unit with nth rank is selected in the nth set. Then make measurements on the variable of interest of the selected units. Let $X_{(i:n)}$ be the measurement made on the ith selected units, then $X_{(1:n)}, X_{(2:n)}, \ldots, X_{(n:n)}$ constitute the ranked set sample.

Many authors modified the McIntyre (1952) method of ranked set sampling. Al-Odat and Al-Saleh (2001) introduced a modified ranked set sampling scheme named moving extreme ranked set sampling for improving the efficiency of the estimate of the population mean. The procedure

*Corresponding author
Received: 29 August 2024 / Accepted: 26 October 2024
DOI: 10.22067/smps.2024.45946

© 2024 Ferdowsi University of Mashhad

of moving the extreme ranked set sampling method is as follows. Choose n sets of random samples of sizes $1,2,...,n$ respectively, from the population. Rank the units in each set using the judgment method or some inexpensive methods, without making the actual measurement of the variable of interest. Select the unit with rank one from each set and then take actual measurement of selected units, we get a moving lower extreme ranked set sample (MLERSS) and the method is known as moving lower extreme ranked set sampling. If we select the units with maximum rank in each set and then take actual measurements of the selected units, we get a moving upper extreme ranked set sample (MUERSS) and the method is known as moving upper extreme ranked set sampling.

Let X be a non negative random variable with probability density function $f(x)$. Shannon (1948) introduced a measure of uncertainty as an average level of information associated with the random variable X, known as Shannon Entropy and it is defined as

$$H(X) = E_f\big[-\ln[f(X)]\big] = -\int_0^\infty f(x)\ln[f(x)]dx.$$

One of the generalization of Shannon entropy was done by Kerridge (1961). Let X and Y be two absolutely continuous non negative random variables with distribution functions F and G and probability density functions f and g, respectively. If F is the distribution function corresponding to the observations and G is the distribution assigned by the experimenter, then the inaccuracy measure of X and Y, proposed by Kerridge (1961) is given by

$$KI(f,g) = E_f\big[-\ln[g(X)]\big] = -\int_0^\infty f(x)\ln[g(x)]dx. \tag{1}$$

The inaccuracy measure defined in (1) can also be called as cross-entropy of Y on X or the relative distance between X and Y. In reliability analysis, Kerridge (1961) notion of inaccuracy delves into the concept by examining the potential errors in expressing probabilities related to different events in experiments. These errors can stem from two primary sources: one from missing data and another from misspecification of the model. Kerridge's inaccuracy measure is designed to address both types of errors. Kayal and Sunoj (2017) studied a generalized Kerridge's inaccuracy measure for conditionally specified models.Taneja and Tuteja (1986) discussed about the weighted inaccuracy measure. Ghosh and Kundu (2018) discussed the conditional cumulative past version of Kerridge's inaccuracy measure. Taneja et al. (2009) presented the dynamic version of inaccuracy between two residual lifetime distributions. Here the information contents are averaged over a known distribution. Kerridge's inaccuracy measure has a significant role in regression analysis for the Akaike information criteria. Applications of Kerridge's measure in coding theory can be seen in Nath (1968).

Recently, researchers have been interested in studying the scatter of information contents of a random variable. The variability of the information component cannot be explained by inaccuracy, which is the expected information content of suitability. Buono et al. (2021) introduced a dispersion index between two random variables X and Y based on cross-entropy named

varianaccuracy defined as

$$\begin{aligned} VarKI(f,g) &= Var_f\big[-\ln[g(X)]\big] \\ &= E_f[\ln^2 g(X)] - \Big[E_f\big[\ln g(X)\big]\Big]^2 \\ &= \int_0^\infty f(x)\ln^2 g(x)dx - \Big[KI(f,g)\Big]^2. \end{aligned}$$

The variance of information content have many applications in reliability study, estimation etc. Varinaccuracy measures the discrepancy incurred in choosing a reference distribution. Sharma and Kundu (2024) have discussed the residual and past varinaccuracy measures and its application. Balakrishnan et al. (2024) discussed the dispersion indices based on Kerridge inaccuracy measure.

Thapliyal and Taneja (2013, 2015) discussed the inaccuracy and residual inaccuracy of order statistics. Ahmadi (2021) explained some results based on Kerridge's inaccuracy measures of records. Goel et al. (2018) discussed Kerridge's inaccuracy measure for record statistics. Mohammed (2019) discussed the inaccuracy measure in concomitants of ordered random variables under Farlie-Gumbel-Morgenstern family. Mohammadi and Hashempour (2023) studied extropy based inaccuracy measure in order statistics. Chacko and George (2023, 2024) discussed the extropy properties of ranked set samples when sampling is not perfect. George and Chacko (2023) discussed the cumulative residual extropy properties of ranked set samples for Cambanis type bivariate distributions. None of the previous work studied the varinaccuracy based on ranked set samples. In this paper, we discuss Kerridge's inaccuracy measure and varrinacuracy for RSS and its modifications.

The paper is organized as follows. In section 2, Kerridge's inaccuracy measure and varinaccuracy between the distribution of the rth order statistic and parent distribution are explained. Kerridge's measure of inaccuracy of ranked set sample is explained in section 3. In section 4, we explain the varriancccuracy of ranked set sample with examples. Section 5 gives Kerridge's inaccuracy measure of moving extreme ranked set sample. Varinaccuracy of moving extreme ranked set sample is given in section 6. Finally, section 7 gives the concluding remarks.

2 Varinaccuracy between distribution of the rth order statistic and the parent distribution

Let $X_1, X_2, \ldots X_n$ be a random sample from a population with pdf $f(x)$ and cdf $F(x)$. If we arrange X_i's in ascending order of magnitude as $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, then $X_{(r)}$ is called the rth order statistic of X_i's. Then, the pdf of rth order statistic is given by

$$f_{r:n}(x) = \frac{1}{B(r, n-r+1)}(F(x))^{r-1}(1-F(x))^{n-r}f(x), \tag{2}$$

where $B(a,b)$ is the beta function. Therefore Kerridge's inaccuracy measure between $X_{(r)}$ and X can be written as

$$\begin{aligned} KI(f_{r:n}, f) &= -\int_0^\infty f_{r:n}(x)\ln f(x)dx \\ &= E\big[-\ln f(X_{(r)})\big]. \end{aligned}$$

Also, the varinaccuracy between distribution of $X_{(r)}$ and X can be written as

$$VarKI(f_{r:n}, f) = Var\big[-\ln f(X_{(r)})\big].$$

Lemma 1. *Let $M_{r:n}(t)$ be the moment generating function of $\ln f(X_{(r)})$. Then, $M_{r:n}(t)$ is given by*

$$M_{r:n}(t) = E[f(F^{-1}(U))]^t,$$

where U follows a beta distribution with parameters r and $n-r+1$.

Proof. From (2), we have

$$\begin{aligned} M_{r:n}(t) &= E[e^{t\ln f(X_{(r)})}] \\ &= E[f^t(X_{(r)})] \\ &= \int_0^\infty \frac{1}{B(r,n-r+1)}(F(x))^{r-1}(1-F(x))^{n-r} f^{t+1}(x)dx \\ &= \int_0^1 \frac{1}{B(r,n-r+1)} u^{r-1}(1-u)^{n-r} f^t(F^{-1}(u))du \\ &= E[f^t(F^{-1}(U))], \end{aligned}$$

where U follows beta distribution with parameters r and $n-r+1$. □

Theorem 1. *Let $K_{r:n}(t)$ be the cumulant generating function of $\ln f(X_{(r)})$. Then, $KI(f_{r:n}, f) = -K'_{r:n}(0)$ and $VarKI(f_{r:n}, f) = K''_{r:n}(0)$, where $K'_{r:n}(0)$ and $K''_{r:n}(0)$ are the first and second derivatives of $K_{r:n}(t)$ at $t = 0$.*

Proof. We have $K_{r:n}(t) = \ln M_{r:n}(t)$. Therefore

$$K'_{r:n}(0) = E[\ln f(X_{(r)})], \quad \text{and} \quad K''_{r:n}(0) = Var[\ln f(X_{(r)})].$$

Also,

$$KI(f_{r:n}, f) = -K'_{r:n}(0) \quad \text{and} \quad VarKI(f_{r:n}, f) = K''_{r:n}(0).$$

Hence the theorem. □

Example 1. If X follows uniform distribution over (a,b), then $M_{r:n}(t) = \dfrac{1}{(b-a)^t}$. Therefore

$$KI(f_{r:n}, f) = -\ln(b-a) \quad \text{and} \quad VarKI(f_{r:n}, f) = 0.$$

Example 2. If X follows exponential distribution with pdf $f(x) = \theta e^{-\theta x}, x \geq 0, \theta > 0$, then

$$M_{r:n}(t) = \theta^{t-1}\frac{B(r,n-r+t)}{B(r,n-r+1)}.$$

Therefore

$$\begin{aligned} KI(f_{r:n}, f) &= -\Big[\ln\theta + \psi(n-r+1) - \psi(n+1)\Big] \\ &= -\ln\theta + \sum_{i=1}^{r}\frac{1}{n-i+1} \end{aligned}$$

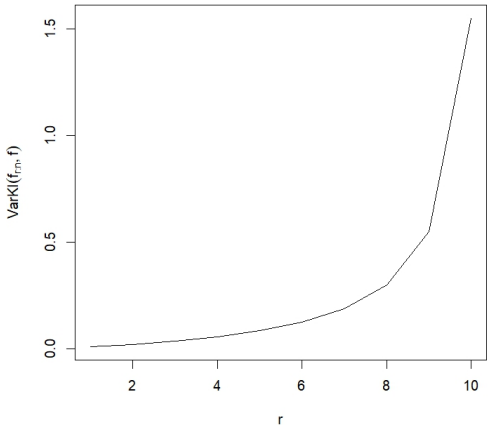


Figure 1: Graph of $VarKI(f_{r:n},f)$ of exponential distribution.

and

$$\begin{aligned} VarKI(f_{r:n},f) &= \psi'(n-r+1)-\psi'(n+1) \\ &= \sum_{i=1}^{r} \frac{1}{(n-i+1)^2}, \end{aligned}$$

where $\psi(.)$ is the digamma function and $\psi'(.)$ is the trigamma function.

Clearly $VarKI(f_{r:n},f)$ of exponential distribution is free from parameter. We have drawn the graph of $VarKI(f_{r:n},f)$ for exponential distribution when $n=10$ and is given in Figure 1.

Example 3. If X follows Pareto distribution with pdf $f(x)=\frac{\lambda\beta^{\lambda}}{x^{\lambda+1}}, x\geq\beta,\lambda>0,\beta>0$, then

$$M_{r:n}(t)=\frac{\lambda^{t-1}}{\beta^{t-1}}\frac{B(r,n-r+(t-1)(1+\frac{1}{\lambda})+1)}{B(r,n-r+1)}.$$

Therefore

$$\begin{aligned} KI(f_{r:n},f) &= -\Big[\ln\lambda-\ln\beta+(1+\frac{1}{\lambda})\psi(n-r+1)-(1+\frac{1}{\lambda})\psi(n+1)\Big] \\ &= \ln\beta-\ln\lambda+(1+\frac{1}{\lambda})\sum_{i=1}^{r}\frac{1}{n-i+1} \end{aligned}$$

and

$$\begin{aligned} VarKI(f_{r:n},f) &= (1+\frac{1}{\lambda})^2\psi'(n-r+1)-(1+\frac{1}{\lambda})^2\psi'(n+1) \\ &= (1+\frac{1}{\lambda})^2\sum_{i=1}^{r}\frac{1}{(n-i+1)^2}. \end{aligned}$$

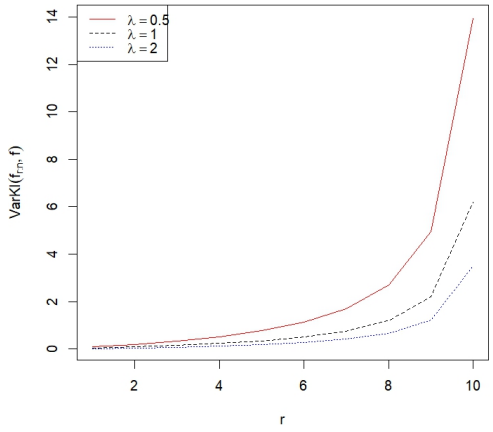


Figure 2: Graph of $VarKI(f_{r:n}, f)$ of Pareto distribution.

We have drawn the graph of $VarKI(f_{r:n}, f)$ for Pareto distribution for different values of λ when $n = 10$ and is given in Figure 2.

Example 4. If X follows standard power distribution with pdf $f(x) = \eta x^{\eta-1}, 0 < x < 1, \eta > 0$, then

$$M_{r:n}(t) = \eta^{t-1} \frac{B(r, n-r+(t-1)(1-\frac{1}{\eta})+1)}{B(r, n-r+1)}.$$

Therefore

$$\begin{aligned} KI(f_{r:n}, f) &= -\Big[\ln\eta + (1-\frac{1}{\eta})\psi(n-r+1) - (1-\frac{1}{\eta})\psi(n+1)\Big] \\ &= -\ln\eta + (1-\frac{1}{\eta})\sum_{i=1}^{r}\frac{1}{n-i+1} \end{aligned}$$

and

$$\begin{aligned} VarKI(f_{r:n}, f) &= (1-\frac{1}{\eta})^2\psi'(n-r+1) - (1-\frac{1}{\eta})^2\psi'(n+1) \\ &= (1-\frac{1}{\eta})^2\sum_{i=1}^{r}\frac{1}{(n-i+1)^2}. \end{aligned}$$

We have drawn the graph of $VarKI(f_{r:n}, f)$ for power distribution for different values of η when $n = 10$ and is given in Figure 3.

Remark 1. *From Table 1, it is observed that if X, Y and Z follow standard power distribution with pdf $f(x) = \eta x^{\eta-1}, 0 < x < 1, \eta > 0$, exponential distribution with pdf $g(y) = \theta e^{-\theta y}, y \geq 0, \theta > 0$ and Pareto distribution with pdf $h(z) = \frac{\lambda\beta^{\lambda}}{z^{\lambda+1}}, \lambda > 0, \beta > 0, z \geq \beta$, respectively and if $1 + \frac{1}{\lambda} \geq |\, 1 - \frac{1}{\eta}\,|$, then $VarKI(f_{r:n}, f) \leq VarKI(g_{r:n}, g) \leq VarKI(h_{r:n}, h)$.*

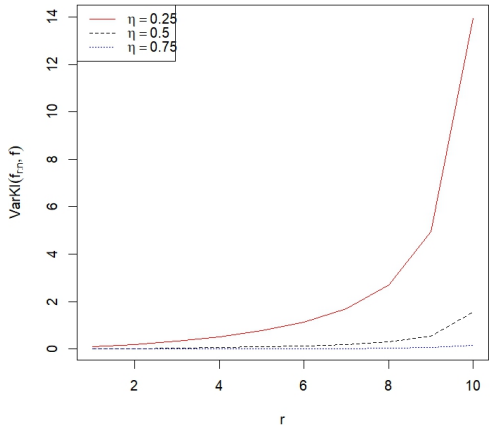


Figure 3: Graph of $VarKI(f_{r:n}, f)$ of standard power distribution.

Table 1: Expressions for $KI(f_{r:n}, f)$ and $VarKI(f_{r:n}, f)$.

Distribution	pdf	$KI(f_{r:n}, f)$	$VarKI(f_{r:n}, f)$
Uniform	$\frac{1}{b-a},\ a<x<b$	$-\ln(b-a)$	0
Exponential	$\theta e^{-\theta x}, x \geq 0, \theta > 0$	$-\ln\theta + \sum_{i=1}^{r} \frac{1}{n-i+1}$	$\sum_{i=1}^{r} \frac{1}{(n-i+1)^2}$
Pareto	$\frac{\lambda\beta^{\lambda}}{x^{\lambda+1}}, x \geq \beta, \lambda > 0, \beta > 0$	$\ln\beta - \ln\lambda + (1+\frac{1}{\lambda})\sum_{i=1}^{r} \frac{1}{n-i+1}$	$(1+\frac{1}{\lambda})^2 \sum_{i=1}^{r} \frac{1}{(n-i+1)^2}$
Standard power	$\eta x^{\eta-1}, 0<x<1, \eta > 0$	$-\ln\eta + (1-\frac{1}{\eta})\sum_{i=1}^{r} \frac{1}{n-i+1}$	$(1-\frac{1}{\eta})^2 \sum_{i=1}^{r} \frac{1}{(n-i+1)^2}$

So if $1+\frac{1}{\lambda} \geq \mid 1-\frac{1}{\eta} \mid$, discrepancy incurred in choosing the rth order statistic of power distribution is less compared to those of exponential distribution and Pareto distribution.

Theorem 2. *Let X and Y be two absolutely continuous non negative random variables with pdf $f_X(x)$ and $g_X(x)$, then*

$$VarKI(f,g) \leq \Big(\int (KI(f,g) + \ln f_X(x))^4 dx \Big)^{\frac{1}{2}} \Big(E(f_X(X)) \Big)^{\frac{1}{2}},$$

where $KI(f,g)$ is the Kerridge's inaccuracy measure of X and Y.
The equality is attained by

$$\left(\frac{KI(f,g) + \ln f_X(x)}{f_X(x)} \right)^2 E(f_X(X)) = \int \big[KI(f,g) + \ln f_X(x) \big]^2 dx.$$

Proof. We have

$$\begin{aligned} VarKI(f,g) &= E_g\left[-\ln f_X(X) - E_g(-\ln f_X(X))\right]^2 \\ &= E\left[-\ln f_X(X) - KI(f,g)\right]^2 \\ &= \int (KI(f,g)+\ln f_X(x))^2 f_X(x)dx. \end{aligned}$$

Then, by Cauchy- Schwartz inequality,

$$\begin{aligned} VarKI(f,g) &\leq \left(\int (KI(f,g)+\ln f_X(x))^4 dx\right)^{\frac{1}{2}} \left(\int (f(x))^2 dx\right)^{\frac{1}{2}} \\ &= \left(\int (KI(f,g)+\ln f_X(x))^4 dx\right)^{\frac{1}{2}} \left(E(f_X(X))\right)^{\frac{1}{2}}. \end{aligned}$$

The equality is attained if and only if there is a constant $\alpha \geq 0$ such that

$$(KI(f,g)+\ln f_X(x))^2 = \alpha (f_X(x))^2.$$

The constant α can be evaluated as

$$\alpha = \frac{\int \left[KI(f,g)+\ln f_X(x)\right]^2 dx}{E(f_X(X))}.$$

Hence the theorem. □

Remark 2. *Balakrishnan et al. (2024) showed that $VarKI(f,g)$ does not affect linear transformation. So clearly $VarKI(f_{r:n},f)$ is independent of scale parameter. This can be evident from Examples 2 and 3 for exponential and Pareto distributions, respectively.*

3 Kerridge's inaccuracy measure of ranked set sample

Let $X_{(r:n)}, r = 1,2,...,n$ be a ranked set sample of size n from a population with pdf $f(x)$ and cdf $F(x)$. If the ranking is perfect, then $X_{(r:n)}$ is nothing but the rth order statistic of a random sample of size n.

Theorem 3. *If $X_{SRS} = \{X_1, X_2, ..., X_n\}$ and $X_{RSS} = \{X_{(1:n)}, X_{(2:n)}, ..., X_{(n:n)}\}$, where $X_1, X_2, ..., X_n$ is a random sample from a population with pdf $f(x)$ and cdf $F(x)$, then the Kerridge's inaccuracy measure associated with ranked set sample and simple random sample can be written as*

$$KI(f_{RSS}, f_{SRS}) = \sum_{r=1}^{n} KI(f_{r:n}, f).$$

Table 2: Expression for $KI(f_{RSS}, f_{SRS})$.

Distribution	pdf	$KI(f_{RSS}, f_{SRS})$
Uniform	$\frac{1}{b-a}, a < x < b$	$-n\ln(b-a)$
Exponential	$\theta e^{-\theta x}, x \geq 0, \theta > 0$	$n(1-\ln\theta)$
Pareto	$\frac{\lambda\beta^{\lambda}}{x^{\lambda+1}}, \lambda > 0, \beta > 0, x \geq \beta$	$n\left(1-\ln\lambda+\ln\beta+\frac{1}{\lambda}\right)$
Standard power	$\eta x^{\eta-1}, 0 < x < 1, \eta > 0$	$n\left(1-\ln\eta-\frac{1}{\eta}\right)$

Proof. Let $f_{SRS}(\underline{x})$ and $f_{RSS}(\underline{x})$ be the pdfs of X_{SRS} and X_{RSS}, respectively, where $\underline{x} = (x_1, x_2, ..., x_n)$. Therefore, $f_{SRS}(\underline{x}) = \prod_{r=1}^{n} f(x_r)$ and $f_{RSS}(\underline{x}) = \prod_{r=1}^{n} f_{r:n}(x_r)$. Then

$$\begin{aligned} KI(f_{RSS}, f_{SRS}) &= -\int_0^{\infty}\int_0^{\infty}\ldots\int_0^{\infty} f_{RSS}(\underline{x})\ln f_{SRS}(\underline{x})dx_1dx_2\ldots dx_n \\ &= E\big[-\ln f_{SRS}(X_{RSS})\big] \\ &= \sum_{r=1}^{n} E\big[-\ln f(X_{(r:n)})\big] \\ &= \sum_{r=1}^{n} KI(f_{r:n}, f). \end{aligned}$$

Hence the theorem. □

Example 5. We obtained $KI(f_{RSS}, f_{SRS})$ for uniform, exponential, Pareto and standard power distributions with the pdfs given as in Table 1, the results are summarized in Table 2.

4 Measure of Varinaccuracy of ranked set sample

Theorem 4. *If $X_{SRS} = \{X_1, X_2, ..., X_n\}$ and $X_{RSS} = \{X_{(1:n)}, X_{(2:n)}, ..., X_{(n:n)}\}$, where $X_1, X_2, ..., X_n$ is a random sample from a population with pdf $f(x)$ and cdf $F(x)$, then the varinaccuracy between ranked set sample and simple random sample can be written as*

$$VarKI(f_{RSS}, f_{SRS}) = \sum_{r=1}^{n} VarKI(f_{r:n}, f).$$

Proof. Let $f_{SRS}(\underline{x})$ and $f_{RSS}(\underline{x})$ be the pdfs of X_{SRS} and X_{RSS}, respectively, where $\underline{x} = (x_1, x_2, ..., x_n)$. Therefore, $f_{SRS}(\underline{x}) = \prod_{r=1}^{n} f(x_r)$ and $f_{RSS}(\underline{x}) = \prod_{r=1}^{n} f_{r:n}(x_r)$. Then

$$\begin{aligned} VarKI(f_{RSS}, f_{SRS}) &= Var\big[-\ln f_{SRS}(X_{RSS})\big] \\ &= Var\Big[-\ln\prod_{r=1}^{n} f(X_{(r:n)})\Big] \\ &= \sum_{r=1}^{n} Var\big[-\ln f(X_{(r:n)})\big] \\ &= \sum_{r=1}^{n} VarKI(f_{r:n}, f). \end{aligned}$$

Table 3: Expression for $VarKI(f_{RSS}, f_{SRS})$.

Distribution	pdf	$VarKI(f_{RSS}, f_{SRS})$
Uniform	$\frac{1}{b-a}, a < x < b$	0
Exponential	$\theta e^{-\theta x}, x \geq 0, \theta > 0$	$\sum_{i=1}^{n} \frac{1}{i}$
Pareto	$\frac{\lambda \beta^{\lambda}}{x^{\lambda+1}}, \lambda > 0, \beta > 0, x \geq \beta$	$(1+\frac{1}{\lambda})^2 \sum_{i=1}^{n} \frac{1}{i}$
Standard power	$\eta x^{\eta-1}, 0 < x < 1, \eta > 0$	$(1-\frac{1}{\eta})^2 \sum_{i=1}^{n} \frac{1}{i}$

□

Example 6. We obtained $VarKI(f_{RSS}, f_{SRS})$ for uniform, exponential, Pareto and standard power distributions with the pdfs given as in Table 1, the results are summarized in Table 3.

We have drawn the graph of $VarKI(f_{RSS}, f_{SRS})$ for exponential, Pareto and standard power distributions and displayed in Figures 4, 5 and 6, respectively.

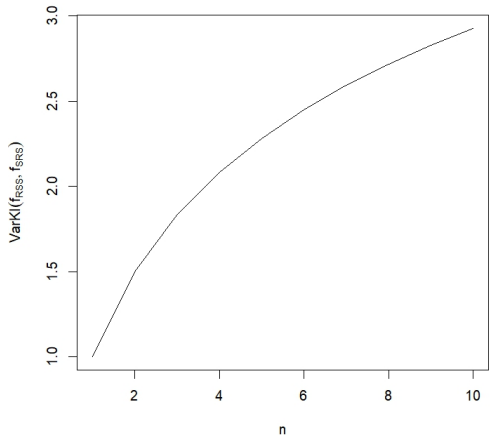


Figure 4: Graph of $VarKI(f_{RSS}, f_{SRS})$ for exponential distribution.

Remark 3. *From Table 3 and the same assumptions of Remark 1, we have*

$$VarKI(f_{RSS}, f_{SRS}) \leq VarKI(g_{RSS}, g_{SRS}) \leq VarKI(h_{RSS}, h_{SRS}).$$

So if $1+\frac{1}{\lambda} \geq |\, 1-\frac{1}{\eta}\,|$, discrepancy incurred in choosing RSS for power distribution is less compared to those of exponential distribution and Pareto distribution.

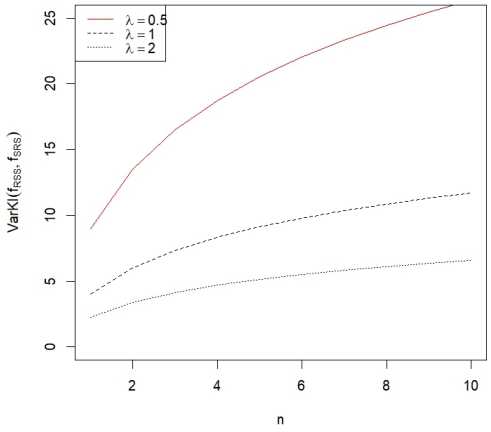


Figure 5: Graph of $VarKI(f_{RSS}, f_{SRS})$ for Pareto distribution.

5 Kerridge's inaccuracy measure of moving extreme ranked set sample

In this section, first we consider the Kerridge's inaccuracy measure of MLERSS. Let $X_{(1:j)}$, $j = 1,2,...n.$ be the measurement of the jth unit of MLERSS.

Theorem 5. *If $X_{SRS} = \{X_1, X_2, ..., X_n\}$ and $X_{MLERSS} = \{X_{(1:j)}, j = 1,2,...,n\}$, where $X_1, X_2, ..., X_n$ is a random sample from a population with pdf $f(x)$ and cdf $F(x)$, then the Kerridge's inaccuracy measure between MLERSS and simple random sample can be written as*

$$KI(f_{MLERSS}, f_{SRS}) = \sum_{i=1}^{n} KI(f_{1:i}, f).$$

Proof. Let $f_{SRS}(\underline{x})$ and $f_{MLERSS}(\underline{x})$ be the pdfs of X_{SRS} and X_{MLERSS}, respectively, where $\underline{x} = (x_1, x_2, ..., x_n)$. Therefore, $f_{SRS}(\underline{x}) = \prod_{i=1}^{n} f(x_i)$ and $f_{MLERSS}(\underline{x}) = \prod_{i=1}^{n} f_{1:i}(x_i)$. Then

$$\begin{aligned} KI(f_{MLERSS}, f_{SRS}) &= -\int_0^\infty \int_0^\infty ... \int_0^\infty f_{MLERSS}(\underline{x}) ln f_{SRS}(\underline{x}) dx_1 dx_2 ... dx_n \\ &= E\left[-\ln f_{SRS}(X_{MLERSS})\right] \\ &= \sum_{i=1}^{n} E\left[-\ln f(X_{(1:i)})\right] \\ &= \sum_{i=1}^{n} KI(f_{1:i}, f). \end{aligned}$$

Hence the theorem. □

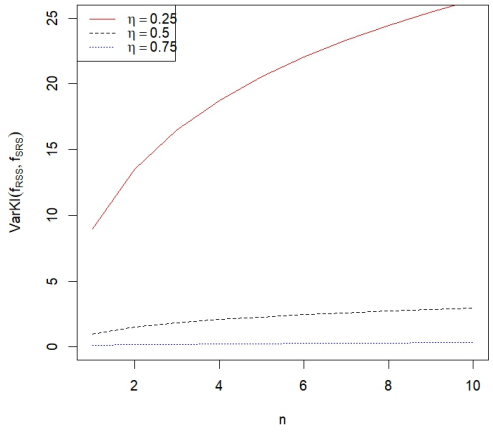


Figure 6: Graph of $VarKI(f_{RSS}, f_{SRS})$ for standard power distribution.

Next, we consider the Kerridge's inaccuracy measure of MUERSS. Let $X_{(j:j)}$, $j = 1,2,...n.$ be the measurement the jth unit of MUERSS.

Theorem 6. *Let $X_{SRS} = \{X_1, X_2, ..., X_n\}$ and $X_{MUERSS} = \{X_{(j:j)}, j = 1,2,...,n\}$ where $X_1, X_2, ..., X_n$ is a random sample from a population with pdf $f(x)$ and cdf $F(x)$, then, the Kerridge's inaccuracy measure between MUERSS and simple random sample can be written as of*

$$KI(f_{MUERSS}, f_{SRS}) = \sum_{i=1}^{n} KI(f_{i:i}, f).$$

Proof. Let $f_{SRS}(\underline{x})$ and $f_{MUERSS}(\underline{x})$ be the pdfs of X_{SRS} and X_{MUERSS} respectively, where $\underline{x} = (x_1, x_2, ..., x_n)$. Therefore, $f_{SRS}(\underline{x}) = \prod_{i=1}^{n} f(x_i)$ and $f_{MUERSS}(\underline{x}) = \prod_{i=1}^{n} f_{i:i}(x_i)$. Then

$$\begin{aligned} KI(f_{MUERSS}, f_{SRS}) &= -\int_0^\infty \int_0^\infty ... \int_0^\infty f_{MUERSS}(\underline{x}) \ln f_{SRS}(\underline{x}) dx_1 dx_2 ... dx_n \\ &= E\left[-\ln f_{SRS}(X_{MUERSS})\right] \\ &= \sum_{i=1}^{n} E\left[-\ln f(X_{(i:i)})\right] \\ &= \sum_{i=1}^{n} KI(f_{i:i}, f). \end{aligned}$$

□

Example 7. We obtained $KI(f_{MLERSS}, f_{SRS})$ and $KI(f_{MUERSS}, f_{SRS})$ for uniform, exponential, Pareto and standard power distributions with the pdfs given as in Table 1, the results are summarized and presented in Table 4.

Table 4: Expression for $KI(f_{MLERSS}, f_{SRS})$ and $KI(f_{MUERSS}, f_{SRS})$.

Distribution	pdf	$KI(f_{MLERSS}, f_{SRS})$	$KI(f_{MUERSS}, f_{SRS})$
Uniform	$\frac{1}{b-a}, a < x < b$	$-n\ln(b-a)$	$-n\ln(b-a)$
Exponential	$\theta e^{-\theta x}, x \geq 0$	$-n\ln\theta + \sum_{i=1}^{n}\frac{1}{i}$	$(n+1)\sum_{i=1}^{n}\frac{1}{i} - n(1+\ln\theta)$
Pareto	$\frac{\lambda\beta^{\lambda}}{x^{\lambda+1}}, x \geq \beta$	$n\ln\beta - n\ln\lambda + (1+\frac{1}{\lambda})\sum_{i=1}^{n}\frac{1}{i}$	$n(\ln\beta - \ln\lambda) + (1+\frac{1}{\lambda})\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{1}{j}$
Standard power	$\eta x^{\eta-1}, 0 < x < 1$	$-n\ln\eta + (1-\frac{1}{\eta})\sum_{i=1}^{n}\frac{1}{i}$	$-n\ln\eta + (1-\frac{1}{\eta})\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{1}{j}$

6 Measure of Varinaccuracy of moving extreme ranked set sample

In this section, first we consider the varentropy of MLERSS. Let $X_{(1:j)},\ j = 1,2,...n.$ be the measurement of the jth unit of MLERSS.

Theorem 7. *If $X_{SRS} = \{X_1, X_2, ..., X_n\}$ and $X_{MLERSS} = \{X_{(1:j)}, j = 1,2,...,n\}$, where $X_1, X_2, ..., X_n$ is a random sample from a population with pdf $f(x)$ and cdf $F(x)$, then the varinaccuracy between MLERSS and simple random sample can be written as of*

$$VarKI(f_{MLERSS}, f_{SRS}) = \sum_{i=1}^{n} VarKI(f_{1:i}, f).$$

Proof. Let $f_{SRS}(\underline{x})$ and $f_{MLERSS}(\underline{x})$ be the pdfs of X_{SRS} and X_{MLERSS} respectively, where $\underline{x} = (x_1, x_2, ..., x_n)$. Therefore, $f_{SRS}(\underline{x}) = \prod_{i=1}^{n} f(x_i)$ and $f_{MLERSS}(\underline{x}) = \prod_{i=1}^{n} f_{i:i}(x_i)$. Then

$$\begin{aligned} VarKI(f_{MLERSS}, f_{SRS}) &= Var\big[-\ln f_{SRS}(X_{MLERSS})\big] \\ &= Var\big[-\ln \prod_{r=1}^{n} f(X_{(1:i)})\big] \\ &= \sum_{i=1}^{n} Var\big[-\ln f(X_{(1:i)})\big] \\ &= \sum_{i=1}^{n} VarKI(f_{1:i}, f). \end{aligned}$$

□

We have drawn the graph of $VarKI(f_{MLERSS}, f_{SRS})$ for exponential, Pareto and standard power distributions, is given in Figures 7, 8 and 9, respectively.

Next, we consider the varentropy of MUERSS. Let $X_{(j:j)},\ j = 1,2,...n.$ be the measurement the jth unit of MUERSS.

Theorem 8. *Let $X_{SRS} = \{X_1, X_2, ..., X_n\}$ and $X_{MUERSS} = \{X_{(j:j)}, j = 1,2,...,n\}$ where $X_1, X_2, ..., X_n$ is a random sample from a population with pdf $f(x)$ and cdf $F(x)$, then, the varinaccuracy between MUERSS and simple random sample can be written as of*

$$VarKI(f_{MUERSS}, f_{SRS}) = \sum_{i=1}^{n} VarKI(f_{i:i}, f).$$

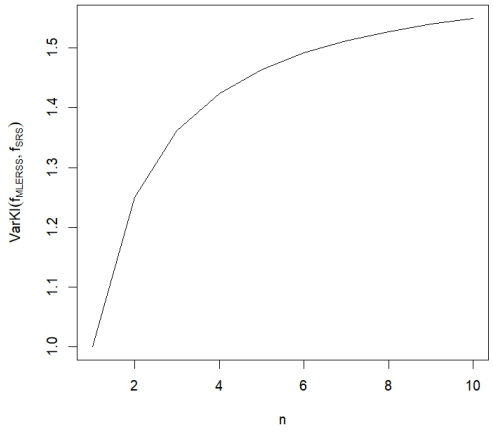


Figure 7: Graph of $VarKI(f_{MLERSS}, f_{SRS})$ for exponential distribution.

Proof. Let $f_{SRS}(\underline{x})$ and $f_{MUERSS}(\underline{x})$ be the pdfs of X_{SRS} and X_{MLERSS} respectively, where $\underline{x} = (x_1, x_2, ..., x_n)$. Therefore, $f_{SRS}(\underline{x}) = \prod_{i=1}^{n} f(x_i)$ and $f_{MUERSS}(\underline{x}) = \prod_{i=1}^{n} f_{i:i}(x_i)$. Then,

$$\begin{aligned} VarKI(f_{MUERSS}, f_{SRS}) &= Var\big[-\ln f_{SRS}(X_{MUERSS})\big] \\ &= Var\big[-\ln \prod_{i=1}^{n} f(X_{(i:i)})\big] \\ &= \sum_{i=1}^{n} Var\big[-\ln f(X_{(i:i)})\big] \\ &= \sum_{i=1}^{n} VarKI(f_{i:i}, f). \end{aligned}$$

□

Example 8. We obtained $VarKI(f_{MLERSS}, f_{SRS})$ and $VarKI(f_{MUERSS}, f_{SRS})$ for uniform, exponential, Pareto and standard power distributions with the pdfs given as in Table 1, the results are summarized and presented in Table 5.

Remark 4. *From Table 5 and the same assumptions of Remark 1, we have*

$$VarKI(f_{MLERSS}, f_{SRS}) \leq VarKI(g_{MLERSS}, g_{SRS}) \leq VarKI(h_{MLERSS}, h_{SRS}),$$

and

$$VarKI(f_{MUERSS}, f_{SRS}) \leq VarKI(g_{MUERSS}, g_{SRS}) \leq VarKI(h_{MUERSS}, h_{SRS}).$$

Figures 10, 11 and 12 show the graph of $VarKI(f_{MUERSS}, f_{SRS})$ for exponential, Pareto and standard power distributions with the pdfs given as in Table 1, respectively.

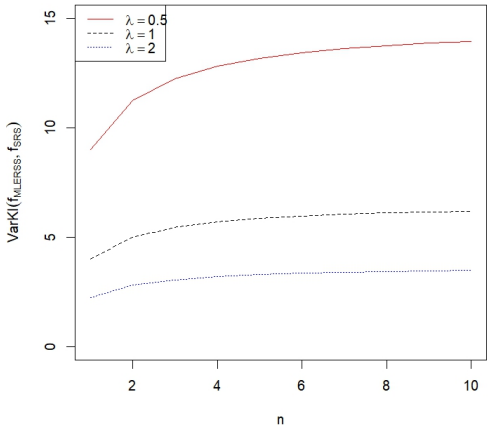


Figure 8: Graph of $VarKI(f_{MLERSS}, f_{SRS})$ for Pareto distribution.

Table 5: Expression for $VarKI(f_{MLERSS}, f_{SRS})$ and $VarKI(f_{MUERSS}, f_{SRS})$

Distribution	pdf	$VarKI(f_{MLERSS}, f_{SRS})$	$VarKI(f_{MUERSS}, f_{SRS})$
Uniform	$\frac{1}{b-a}, a < x < b$	0	0
Exponential	$\theta e^{-\theta x}, x \geq 0, \theta > 0$	$\sum_{i=1}^{n} \frac{1}{i^2}$	$\sum_{i=1}^{n} \frac{i}{(n-i+1)^2}$
Pareto	$\frac{\lambda \beta^{\lambda}}{x^{\lambda+1}}, \lambda > 0, \beta > 0, x \geq \beta$	$(1+\frac{1}{\lambda})^2 \sum_{i=1}^{n} \frac{1}{i^2}$	$(1+\frac{1}{\lambda})^2 \sum_{i=1}^{n} \frac{i}{(n-i+1)^2}$
Standard power	$\eta x^{\eta-1}, 0 < x < 1, \eta > 0$	$(1-\frac{1}{\eta})^2 \sum_{i=1}^{n} \frac{1}{i^2}$	$(1-\frac{1}{\eta})^2 \sum_{i=1}^{n} \frac{i}{(n-i+1)^2}$

7 Conclusion

In this work, we considered Kerridge's inaccuracy measure and varinaccuracy of ranked set sample and its different modifications. If we consider a ranked set sampling when ranking is not perfect, the measurement of ith unit of the ranked set sample is nothing but the ith order statistic, $X_{(i)}$ of the random sample. In this paper, we derived Kerridge's inaccuracy measure between $X_{(i)}$ and X and obtained its varinaccuracy. The expression for varinaccuracy between $X_{(i)}$ and X for the exponential, standard power, and Pareto distributions are derived. We also obtained an upper bound to varinaccuracy.

It is seen that Kerridge's inaccuracy measure between the ranked set sample and simple random sample can be written as the sum of Kerridge's inaccuracy measure between $X_{(i)}$ and X. Similarly, varinaccuracy between a ranked set sample and a simple random sample can be written as the sum of varinaccuracy between $X_{(i)}$ and X. We also obtained the expression for Kerridge's inaccuracy measure and varinacccuracy based on moving ranked set samples. The expression for Kerridge's inaccuracy measure and varinaccuracy of exponential, standard power, and Pareto distributions based on ranked set sample and moving extreme ranked set sample are obtained. It is observed that if $1+\frac{1}{\lambda} \geq \mid 1-\frac{1}{\eta} \mid$, the discrepancy incurred in choosing RSS and

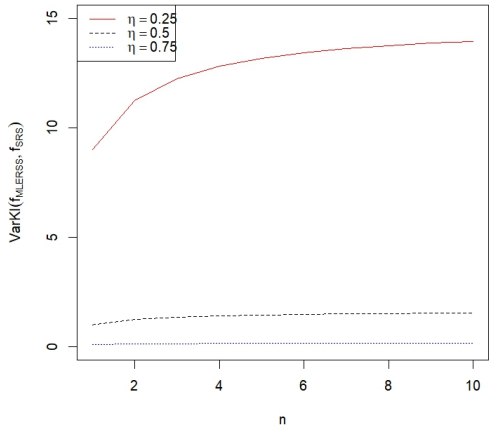


Figure 9: Graph of $VarKI(f_{MLERSS}, f_{SRS})$ for standard power distribution.

moving extreme ranked set sample for power distribution is less than those of exponential and Pareto distribution.

Acknowledgement

The authors are thankful to the anonymous referees for their valuable comments and suggestions which lead to an improved version of the manuscript.

References

Ahmadi, J. (2021). Characterization of continuous symmetric distributions using information measures of records. *Statistical Papers*, **62**, 2603–2626.

Al-Odat, M. T., and Al-Saleh, M. F. (2001). A variation of ranked set sampling. *Journal of Applied Statistical Science*, **10**, 137–146.

Balakrishnan, N., Buono, F., Calì, C., and Longobardi, M. (2024). Dispersion indices based on Kerridge inaccuracy measure and Kullback-Leibler divergence. *Communications in Statistics-Theory and Methods*, **53**, 5574–5592.

Buono, F., Calì, C., and Longobardi, M. (2021). Dispersion indices based on Kerridge inaccuracy and Kullback-Leibler divergence. *arXiv preprint arXiv*:2106. 12292.

Chacko, M., and George, V. (2023). Extropy properties of ranked set sample for Cambanis type bivariate distributions. *Journal of the Indian Society for Probability and Statistics,* **24**, 111–133.

]

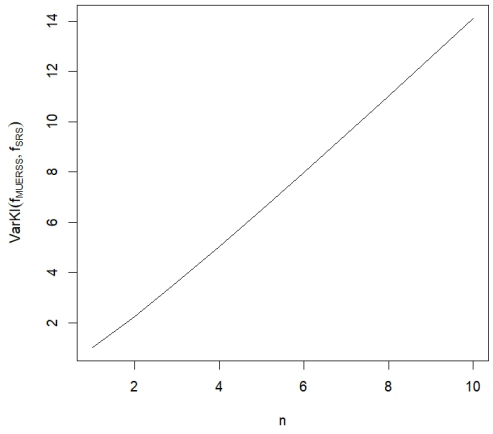


Figure 10: Graph of $VarKI(f_{MUERSS}, f_{SRS})$ for exponential distribution.

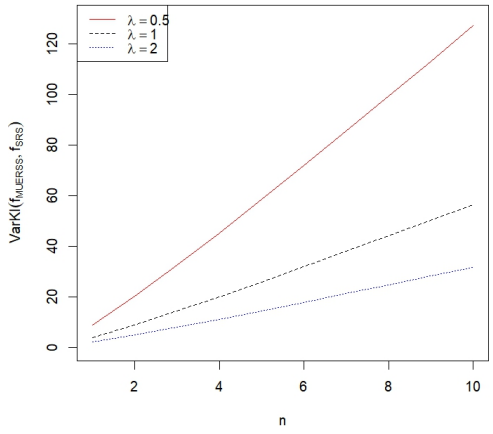


Figure 11: Graph of $VarKI(f_{MUERSS}, f_{SRS})$ for Pareto distribution.

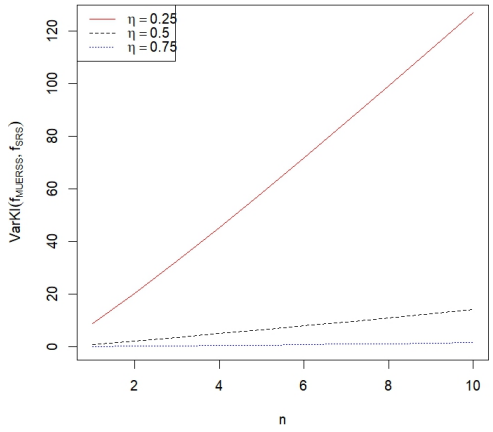


Figure 12: Graph of $VarKI(f_{MUERSS}, f_{SRS})$ for standard power distribution.

Chacko, M., and George, V. (2024). Extropy properties of ranked set sample when ranking is not perfect. *Communications in Statistics-Theory and Methods*, **53**, 3187–3210.

Chen, Z., Bai, Z., and Sinha, B. K. (2004). *Ranked Set Sampling: Theory and Applications* (Vol. 176). Springer, New York.

George, V., and Chacko, M. (2023). Cumulative residual extropy properties of ranked set sample for Cambanis type bivariate distributions. *Journal of the Kerala Statistical Association*, **33**, 50–70.

Ghosh, A., and Kundu, C. (2018). On generalized conditional cumulative past inaccuracy measure. *Applications of Mathematics*, **63**, 167–193.

Goel, R., Taneja, H. C., and Kumar, V. (2018). Kerridge measure of inaccuracy for record statistics. *Journal of Information and Optimization Sciences*, **39**, 1149–1161.

Kayal, S., and Sunoj, S. M. (2017). Generalized Kerridge's inaccuracy measure for conditionally specified models. *Communications in Statistics-Theory and Methods*, **46**, 8257–8268.

Kerridge, D. F. (1961). Inaccuracy and inference. *Journal of the Royal Statistical Society. Series B* (Methodological), 184–194.

McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, **3**, 385–390.

Mohamed, M. S. (2019). A measure of inaccuracy in concomitants of ordered random variables under Farlie-Gumbel-Morgenstern family. *Filomat*, **33**, 4931–4942.

Mohammadi, M., Hashempour, M. (2023). Extropy based inaccuracy measure in order statistics. *Statistics*, **57**, 1490–1510.

Nath, P. (1968). Inaccuracy and coding theory. *Metrika*, **13**, 123–135.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, **27**, 379–423.

Sharma, A., and Kundu, C. (2024). Residual and past varinaccuracy measures. *IMA Journal of Mathematical Control and Information*, **41**, 539–563.

Taneja, H. C., Kumar, V., and Srivastava, R. (2009). A dynamic measure of inaccuracy between two residual lifetime distributions. *International Mathematical Forum*, **4**, 1213–1220.

Taneja, H. C., and Tuteja, R. K. (1986). Characterization of a quantitative-qualitative measure of inaccuracy. *Kybernetika*, **22**, 393–402.

Thapliyal, R., and Taneja, H. C. (2013). A measure of inaccuracy in order statistics. *Journal of Statistical Theory and Applications,* **12**, 200–207.

Thapliyal, R., and Taneja, H. C. (2015). On residual inaccuracy of order statistics. *Statistics and Probability Letters*, **97**, 125–131.

Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 191–209. Research Paper

Bayesian inference of reliability in the case of zero-failure data for the binomial distribution

Yuting Jiang[†] **and Xuefeng Feng**[‡*]

[†]*Southwest Jiaotong University Hope College, Chengdu 610400, China*
[‡]*Department of Statistics, School of Mathematics, Southwest Jiaotong University , Chengdu 611756, China*
Email(s): jiangyuting713@163.com, fxf5693@163.com

Abstract. This article explores Bayesian inference for the reliability of the binomial distribution in cases with zero-failure data. We propose Expected-Bayesian (E-Bayesian) and hierarchical Bayesian estimators of reliability, considering three different joint prior distributions of hyper-parameters in the prior distribution of reliability, which is assumed to follow a beta distribution. We derive closed-form expressions for the E-Bayesian estimators of reliability and propose hierarchical Bayesian estimators, which are subsequently evaluated using Monte Carlo simulations. Furthermore, we study the one-sided modified Bayesian (M-Bayesian) lower credible limits for reliability. The performance of the proposed methods is demonstrated through Monte Carlo simulations. Finally, a real data example is analyzed for illustrative purposes.

Keywords: E-Bayesian; Hierarchical Bayesian; M-Bayesian; Reliability; Zero-failure data.

1 Introduction

With the rapid advancement of manufacturing design techniques, many modern products are designed to operate without failure for several years, decades or more, such as high reliability and longevity products in aerospace, engineering and industry. In this situation, extensive discussions have taken place regarding reliability analysis for zero-failure data. Zero-failure data refers to situations where the test units exhibit no failure within the specified life testing time. Numerous statistical procedures have been proposed in the literature for analyzing reliability in such scenarios. For example, Martz and Waller (1979), Mao and Xia (1992), Han (2009), Chen et al. (1995) have contributed to this body of work. However, limited attention has been given to simultaneously addressing both point and interval estimators of reliability parameters

*Corresponding author
Received: 16 October 2024 / Accepted: 18 November 2024
DOI: 10.22067/smps.2024.45973

© 2024 Ferdowsi University of Mashhad

for zero-failure data. Notable exceptions include the works of Fan and Chang (2009) and Zhang et al. (2019). In the paper, we will introduce and discuss the Expected-Bayesian (E-Bayesian) estimator, hierarchical Bayesian estimator, and one-sided modified Bayesian (M-Bayesian) lower credible limits of reliability for zero-failure data from the binomial distribution. These methods are explored under the assumption that the prior distribution of reliability follows a beta distribution.

For the zero-failure data, Martz and Waller (1979) first proposed a Bayesian zero-failure reliability demonstration testing procedure for an exponential failure-time model and a gamma prior distribution on the failure rate. Following Martz and Waller (1979), special attention has been paid to statistical analysis methods for zero-failure data. Mao and Luo (1989) proposed the matching distribution curve method, and Chen et al. (1993) developed a optimal confidence limit method. In addition, Fan and Chang (2009), Bailey (1997) and Fan et al. (2009) proposed Bayesian methods for estimating the reliability of explosive devices with binary data from destructive tests, where exponential lifetime distributions were used. Han (2007) first proposed an Expected-Bayesian (E-Bayesian) estimation method for estimating the failure probability, in the case of zero-failure data or very few zero-failure data. Jiang et al. (2010) put forward a modified maximum likelihood estimation (MMLE) and shrinkage estimation method to analyze zero-failure data from Weibull distribution. Subsequently, on the base of the thought of hierarchical prior distribution introduced by Lindley and Smith (1972). Han (2011) introduced E-Bayesian and hierarchical Bayesian estimation methods to estimate the reliability of the binomial distribution, however, the proposed methods only considered the case of only one hyper-parameter in the prior distribution of the reliability. Xia (2012) proposed a grey bootstrap method in the information poor theory for the reliability analysis of zero-failure data under the condition of a known or unknown lifetime distribution. Nam et al. (2013) presented a new accelerated zero-failure testing model for rolling bearings. Li et al. (2019) considered the classical Bayesian, E-Bayesian and hierarchical Bayesian estimates of reliability for zero-failure data from Weibull distribution, and confidence interval of reliability is also obtained by the parametric bootstrap method. Zhang et al. (2019) provided a Bayesian reliability evaluation method for very few failure data under the Weibull distribution. Zhang et al. (2021) proposed a Bayesian method for analyzing few or zero failure data using re-parametrization of the Weibull distribution. However, these researches only discussed the case of one of two hyper-parameters is random variable in the prior distribution of failure probability or reliability.

In recent years, the credible (confidence) intervals of reliability parameters based on zero failure data has also made new progress. For example, Han (2008) proposed a two-sided M-Bayesian credible limits method of reliability parameters, in the case of zero-failure data from exponential distribution; furthermore, the properties for two-sided M-Bayesian credible limits of failure rate were discussed. Subsequently, Han (2012) put forward a M-Bayesian credible limits method for estimating the reliability of binomial distribution, in the case of zero-failure data. Tian and Chen (2014) studied interval estimates of failure rate and reliability for the exponential distribution based on zero-failure data, using the method of two-sided M-Bayesian credible limits. Jiang et al. (2015) presented a interval estimation method of failure probability for Weibull distribution under zero-failure data situation. Zhang (2021) proposed a novel method that does not require prior information for estimating Weibull parameters in the absence of failure data, an unbiased estimator and confidence interval for the Weibull scale parameter are

also obtained. However, these researches only considered the case of only one of two hyper-parameters is random in the prior distribution of reliability parameter. In the following, we aim to discuss the Bayesian inference for the reliability of binomial products, in the case of zero-failure data. This discussion encompasses the point estimation method and one-sided M-Bayesian lower credible limits for reliability, based on three different priors, when the prior distribution of reliability is assumed to be a beta distribution with hyper-parameters a and b.

The remainder of this article is organized as follows. In Section 2, we will first introduce the zero-failure data for type-I censored life testing and describe the prior distribution of reliability. Section 3 then presents point estimation of reliability for zero-failure data, including E-Bayesian estimator and hierarchical Bayesian estimator. In Section 4, we discuss one-sided M-Bayesian lower credible limits of reliability in the case of zero-failure data, considering three different joint prior distributions of hyper-parameters. In Section 5, numerical examples are provided, and Section 6 contains some conclusions and several remarks.

2 Zero-failure data for type-I censored life testing

In this section, we first introduce the zero-failure data for highly reliable products in type-I censored life testing. Furthermore, the prior distribution of reliability for zero-failure data is assumed based on engineering experience.

2.1 The zero-failure data

In some cases in reliability engineering, determining the life distribution of products becomes challenging when there is no failure information except for the number of failures. In such situations, reliability estimates can be obtained through the application of nonparametric methods designed for binomial distribution.

In this article, we assume that the distribution of the lifetime T of products is unknown, conduct type-I censored life testing k times, the censoring time of these independent trials are denoted as $t_1, t_2, ..., t_k$ $(0 < t_1 < t_2 < \cdots < t_k)$, respectively. There are n_i testing samples at the censoring time $t_i, i = 1, 2, ..., k$, if these testing units are all without failure at the censoring time t_i, it can be considered that the lifetimes of these testing units is greater than the censoring time $t_i, i = 1, 2, ..., k$. Therefore, the corresponding test data of the k times type-I censored life testing are called the zero-failure data, and denoted as $(t_i, n_i), i = 1, 2, ..., k$. According to the whole testing process of the type-I censored life testing, the following test information can be obtained.

(1) At the beginning time of the type-I censored life testing, the reliability of products can be expressed as $R_0 = Pr(T > 0) = 1$.

(2) At the censoring time t_i, the reliability of products denoted as $R_i = Pr(T > t_i), i = 1, 2, ..., k$, it is clear that $R_0 > R_1 > \cdots > R_k$.

(3) Let $s_i = n_i + n_{i+1} + \cdots + n_k$, it indicates that there are s_i testing units survival at the censoring time t_i, i.e., s_i is the number of products which lifetime is longer than $t_i, i = 1, 2, ..., k$.

Consequently, the above problem can be expressed as evaluating the reliability of products by using the zero-failure data $(t_i, n_i), i = 1, 2, ..., k$, in the type-I censored life testing.

2.2 The prior distribution of reliability

According to some engineering experience and the results of Han (2011), Beta distribution is often used as the conjugate prior distribution of reliability (i.e., the probability of success) in the reliability modeling of zero-failure data for binomial products. We thus take the beta distribution $Beta(a,b)$ as the conjugate prior distribution of the reliability R_i, and the probability density function (pdf) of R_i is

$$\pi(R_i|a,b) = \frac{1}{B(a,b)} R_i^{a-1}(1-R_i)^{b-1}, 0 < R_i < 1, i = 1, 2, ..., k, \tag{1}$$

where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t$ is the beta function, and hyper-parameters $a > 0, b > 0$.

With the continuous improvement of science and technology, the reliability of products is constantly improved, so the possibility of high reliability of products is large, and the possibility of low reliability is small. According to Han (2011), we choose a and b such that $\pi(R_i|a,b)$ is an increasing function of R_i. The derivative of $\pi(R_i|a,b)$ with respect to R_i is

$$\frac{\mathrm{d}\pi(R_i|a,b)}{\mathrm{d}R_i} = \frac{R_i^{a-1}(1-R_i)^{b-1}}{B(a,b)} \left[(a-1)(1-R_i) - (b-1)R_i\right], i = 1, 2, ..., k, \tag{2}$$

Since $a > 0, b > 0$ and $0 < R_i < 1$, from (2), we thus can easily see that $\mathrm{d}\pi(R_i|a,b)/\mathrm{d}R_i > 0$ on $a > 1, 0 < b < 1$, and then $\pi(R_i|a,b)$ is an increasing function of R_i.

According to the results of Berger (1985) and Han (2011) on the Bayesian estimation, from the perspective of the robustness of Bayesian estimation, the thinner the tail of a prior distribution, and the worse the robustness of Bayesian estimation. Therefore, the hyper-parameter a should not be too large, and there should be an upper bound c to be determined, where c is a constant greater than 1. Thereby, the hyper-parameters a and b can be selected in the range of $1 < a < c$ and $0 < b < 1$.

3 Point estimation of reliability

In this section, we investigate point estimation of reliability using the methods of E-Bayesian estimation and hierarchical Bayesian estimation, respectively.

3.1 E-Bayesian estimator of reliability

Referring to the concept of E-Bayesian estimator of failure probability proposed by Han (2007), this method is applicable in cases with either zero or few failure data. In this subsection, we present E-Bayesian estimators of R_i under three different joint prior distributions of hyper-parameters a and b in the prior distribution of R_i.

In the following, we first present the definition of the E-Bayesian estimator of R_i when the prior distribution of R_i is distributed as $Beta(a,b)$.

Definition 1. *Let* $\hat{R}_{iB}(a,b)$ *be continuous, if*

$$\iint_D \left|\hat{R}_{iB}(a,b)\right|\pi(a,b)\mathrm{d}a\mathrm{d}b < \infty,\ \ i = 1,2,...,k,$$

then

$$\hat{R}_{iEB} = \iint_D \hat{R}_{iB}(a,b)\pi(a,b)\mathrm{d}a\mathrm{d}b,\ \ i = 1,2,...,k,$$

is called the Expected Bayesian (E-Bayesian) estimator of R_i*, where* $\hat{R}_{iB}(a,b)$ *is the Bayesian estimator of* R_i *with hyper-parameters* a *and* b*,* $D = \{(a,b)|1 < a < c, 0 < b < 1\}$*,* $\pi(a,b)$ *is the pdf of hyper-parameters* a *and* b *in the region* D*.*

From the Definition 1, it is easy to see that the E-Bayesian estimator of R_i is the mathematical expectation for the Bayesian estimator $\hat{R}_{iB}(a,b)$, that is,

$$\hat{R}_{iEB} = \iint_D \hat{R}_{iB}(a,b)\pi(a,b)\mathrm{d}a\mathrm{d}b = E[\hat{R}_{iB}(a,b)],\ \ i = 1,2,...,k.$$

According to the Bayesian theorem and the Definition 1, we have the following theorem, the proof see Appendix A.

Theorem 1. *For the zero-failure data* (t_i, n_i) *from the* k *times type-I censored life testing, let* $s_i = \sum_{j=i}^{k} n_j$*,* $i = 1,2,...,k$*. If the prior density function* $\pi(R_i|a,b)$ *of* R_i *is presented by the formula* (1)*, then have the following conclusions.*

(1) Using the squared error loss function, the Bayesian estimator of R_i *is* $\hat{R}_{iB}(a,b) = \dfrac{s_i + a}{s_i + a + b}$*,*

(2) For the following three different joint prior distributions of hyper-parameters a *and* b*,*

$$\pi_1(a,b) = \frac{2(c-a)}{(c-1)^2},\ \ 1 < a < c, 0 < b < 1, \tag{3}$$

$$\pi_2(a,b) = \frac{1}{c-1},\ \ 1 < a < c, 0 < b < 1, \tag{4}$$

$$\pi_3(a,b) = \frac{2a}{c^2-1},\ \ 1 < a < c, 0 < b < 1, \tag{5}$$

the corresponding E-Bayesian estimators of R_i *are presented as follows, respectively,*

$$\hat{R}_{iEB1} = \frac{2}{(c-1)^2}\left[G_1(s_i,c) - G_2(s_i,c)\right],\ \ i = 1,2,...,k,$$

$$\hat{R}_{iEB2} = \frac{1}{c-1}\left[G_3(s_i,c) - G_4(s_i,c)\right],\ \ i = 1,2,...,k,$$

$$\hat{R}_{iEB3} = \frac{2}{c^2-1}\left[G_5(s_i,c) - G_6(s_i,c)\right],\ \ i = 1,2,...,k,$$

where $G_1(s_i,c)$ *to* $G_6(s_i,c)$ *and* $q_1(s_i,c)$ *to* $q_5(s_i,c)$ *are provided in Appendix A.*

It is easy to show that the E-Bayesian estimators of R_i in the Theorem 1 have the following properties.

Corollary 1. *In Theorem 1, for the E-Bayesian estimators* $\hat{R}_{iEBj}$ $(i=1,2,...,k, j=1,2,3)$*, the following two properties are hold almost surely,*

(1) For the fixed i, $\hat{R}_{iEB1} < \hat{R}_{iEB2} < \hat{R}_{iEB3}$;

(2) For the fixed i, $\lim\limits_{s_i\to\infty} \hat{R}_{iEB1} = \lim\limits_{s_i\to\infty} \hat{R}_{iEB2} = \lim\limits_{s_i\to\infty} \hat{R}_{iEB3}$.

3.2 Hierarchical Bayesian estimator of reliability

In this subsection, we present the hierarchical Bayesian estimators of R_i $(i=1,2,...,k)$, based on three different joint prior distributions of hyper-parameters a and b. In the following, we first introduce the hierarchical prior distribution of R_i $(i=1,2,...,k)$.

According to the definition of hierarchical prior distribution proposed by Good (1983). If we select $\pi(R_i|a,b)$, which presented by the formula (1), as the prior density function of R_i, and take formulas (3) to (5) as the joint prior distributions of hyper-parameters a and b, respectively, then the hierarchical prior distributions of R_i $(i=1,2,...,k)$ can be easily obtained. For instance,

$$\pi_4(R_i) = \int_0^1\int_1^c \pi(R_i|a,b)\pi_1(a,b)\mathrm{d}a\mathrm{d}b = \frac{2}{(c-1)^2}\int_0^1\int_1^c \frac{(c-a)}{B(a,b)}R_i^{a-1}(1-R_i)^{b-1}\mathrm{d}a\mathrm{d}b, \tag{6}$$

$$\pi_5(R_i) = \int_0^1\int_1^c \pi(R_i|a,b)\pi_2(a,b)\mathrm{d}a\mathrm{d}b = \frac{1}{c-1}\int_0^1\int_1^c \frac{1}{B(a,b)}R_i^{a-1}(1-R_i)^{b-1}\mathrm{d}a\mathrm{d}b, \tag{7}$$

$$\pi_6(R_i) = \int_0^1\int_1^c \pi(R_i|a,b)\pi_3(a,b)\mathrm{d}a\mathrm{d}b = \frac{2}{c^2-1}\int_0^1\int_1^c \frac{a}{B(a,b)}R_i^{a-1}(1-R_i)^{b-1}\mathrm{d}a\mathrm{d}b. \tag{8}$$

We can then use Bayes' theorem to derive the hierarchical Bayesian estimators of R_i, and the results are provided in the next theorem, the proof see Appendix B.

Theorem 2. *For the zero-failure data* (t_i,n_i) *from the* k *times type-I censored life testing, let* $s_i=\sum_{j=i}^k n_j$, $i=1,2,...,k$. *If the hierarchical prior distributions of* R_i *are given by (6) to (8), then under the squared error loss function, the corresponding hierarchical Bayesian estimators of* R_i *are, respectively,*

$$\hat{R}_{iHB1} = \frac{\int_0^1\int_1^c (c-a)\frac{B(a+s_i+1,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c (c-a)\frac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k,$$

$$\hat{R}_{iHB2} = \frac{\int_0^1\int_1^c \frac{B(a+s_i+1,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c \frac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k,$$

$$\hat{R}_{iHB3} = \frac{\int_0^1\int_1^c a\frac{B(a+s_i+1,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c a\frac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k.$$

Taking use of the Monte Carlo simulation method, we can easily get the hierarchical Bayesian estimators of R_i in the Theorem 2.

4 One-sided credible limits of reliability

In this section, we investigate the one-sided modified Bayesian (M-Bayesian) lower credible limits of R_i $(i = 1,2,...,k)$, based on three different joint priors for the hyper-parameters a and b.

4.1 One-sided Bayesian lower credible limits

We first propose the following lemma, which aims to derive the one-sided Bayesian lower credible limits of R_i $(i = 1,2,...,k)$ based on three different priors for the hyper-parameter a and b.

Lemma 1. *For the zero-failure data (t_i, n_i) $(i = 1,2,...,k)$, if the prior distribution of R_i is given by (1), that is R_i is a beta random variable with parameters a and b, we then have $\frac{a+s_i}{b}\frac{1-R_i}{R_i}$ follows a F distribution with the degrees of freedom $2b$ and $2(a+s_i)$, i.e.,*

$$\frac{a+s_i}{b}\frac{1-R_i}{R_i} \sim F(2b, 2(a+s_i)), \quad i = 1,2,..,k.$$

By means of the Lemma 1 and the theory of Bayesian credible limits, we can easily obtain the one-sided Bayesian lower credible limits of R_i. The result is given by the following corollary.

Corollary 2. *For the zero-failure data (t_i, n_i) $(i = 1,2,...,k)$, if the prior distribution of R_i is given by (1), then the $100(1-\alpha)\%$ one-sided Bayesian lower credible limit of R_i is*

$$\hat{R}_{iBL}(a,b) = \frac{a+s_i}{a+s_i+bF_{1-\alpha}(2b, 2(a+s_i))}, \quad i = 1,2,...,k. \tag{9}$$

where $F_{1-\alpha}(2b, 2(a+s_i))$ is the 1-α quantiles of F distribution with the degrees of freedom $2b$ and $2(a+s_i)$.

4.2 One-sided M-Bayesian lower credible limits

In this subsection, we consider the one-sided M-Bayesian lower credible limits of R_i $(i = 1,2,...,k)$. We thus first give the definition of one-sided M-Bayesian lower credible limit of R_i.

Definition 2. *Let $\hat{R}_{iBL}(a,b)$ be continuous, if*

$$\iint_D \left|\hat{R}_{iBL}(a,b)\right| \pi(a,b) \mathrm{d}a\mathrm{d}b < \infty, \quad i = 1,2,...,k,$$

then

$$\hat{R}_{iMBL} = \iint_D \hat{R}_{iBL}(a,b)\pi(a,b)\mathrm{d}a\mathrm{d}b, \quad i = 1,2,...,k,$$

is called the one-sided modified Bayesian (M-Bayesian) lower credible limit of the reliability R_i, where $D = \{(a,b)|1 < a < c, 0 < b < 1\}$ is the domain of a and b, $\hat{R}_{iBL}(a,b)$ is the one-sided Bayesian lower credible limit of R_i with hyper-parameters a and b, $\pi(a,b)$ is the pdf of hyper-parameters a and b in the region D.

We have the following theorem about one-sided M-Bayesian lower credible limits of R_i, for the proof see Appendix C.

Theorem 3. *For the zero-failure data* (t_i, n_i) $i = 1,2,...,k$, *if the prior density function of* R_i *is given by* (1), *then under the credible level of* $1-\alpha$ $(0<\alpha<1)$, *for the three different joint prior distributions* (3) *to* (5) *of the hyper-parameters* a *and* b, *the corresponding one-sided M-Bayesian lower credible limits of* R_i *are, respectively,*

$$\hat{R}_{iMBL1} = \frac{2}{(c-1)^2}\int_0^1\int_1^c \frac{(c-a)(a+s_i)}{a+s_i+bF_{1-\alpha}(2b,2(a+s_i))}\mathrm{d}a\mathrm{d}b,\ \ i=1,2,...,k,$$
$$\hat{R}_{iMBL2} = \frac{1}{c-1}\int_0^1\int_1^c \frac{a+s_i}{a+s_i+bF_{1-\alpha}(2b,2(a+s_i))}\mathrm{d}a\mathrm{d}b,\ \ i=1,2,...,k,$$
$$\hat{R}_{iMBL3} = \frac{2}{c^2-1}\int_0^1\int_1^c \frac{a(a+s_i)}{a+s_i+bF_{1-\alpha}(2b,2(a+s_i))}\mathrm{d}a\mathrm{d}b,\ \ i=1,2,...,k.$$

The one-sided M-Bayesian lower credible limits in Theorem 3 can be evaluated by the Monte Carlo Integration method. Furthermore, the analysis results of the Monte Carlo simulation show that the one-sided M-Bayesian lower credible limits have the following properties.

Corollary 3. *In Theorem 3, for the one-sided M-Bayesian lower credible limits* $\hat{R}_{iMBLj}$ $(i = 1,2,...,k, j = 1,2,3)$, *the following two properties are hold almost surely,*

(1) For the specified i, $\hat{R}_{iMBL1} < \hat{R}_{iMBL2} < \hat{R}_{iMBL3}$,

(2) For the specified i, $\lim\limits_{s_i\to\infty}\hat{R}_{iMBL1} = \lim\limits_{s_i\to\infty}\hat{R}_{iMBL2} = \lim\limits_{s_i\to\infty}\hat{R}_{iMBL3}$.

5 Numerical analysis

5.1 Simulation study

In this subsection, simulation study is conducted to investigate the performance of the point estimators of R_i (including E-Bayesian and hierarchical estimators). According to the Theorems 1 and 2 in Section 3, by simulating the value of s_i $(i = 1,2,...,6)$, we can obtain the E-Bayesian estimators $\hat{R}_{iEBj}$ $(j = 1,2,3)$, and the hierarchical Bayesian estimators $\hat{R}_{iHBj}$ $(j = 1,2,3)$. The results of each $\hat{R}_{iEBj}, \hat{R}_{iHBj}$ are shown in Tables 1 and 2, respectively.

It is clearly that from Tables 1 and 2, for the fixed i $(i = 1,2,...,k)$, the E-Bayesian estimators $\hat{R}_{iEBj}$ $(j = 1,2,3)$ are very close to each other for different values of c (4, 5, 6, 7, 8). In addition, the same conclusion is suitable for the hierarchical Bayesian estimators $\hat{R}_{iHEBj}$ $(j = 1,2,3)$. This indicates that $\hat{R}_{iEBj}$, $\hat{R}_{iHBj}$ $(j = 1,2,3)$ are all robust for different values of c. Therefore, in practical application, we can take the midpoint of interval [4, 8] as the value of c, i.e., $c = 6$. Besides, since $\hat{R}_{iEBj}$, $\hat{R}_{iHBj}$ $(j = 1,2,3)$ are all robust for the specified c (4, 5, 6, 7, 8) under three different joint prior distributions of hyper-parameters a and b, in order to reduce the computational complexity, we thus suggest that the uniform distribution (**??**) can be used as the joint prior distribution of a and b. Finally, for the same values of c and i, the E-Bayesian estimators $\hat{R}_{iEBj}$ are very close to hierarchical Bayesian estimators $\hat{R}_{iHBj}$, thus the E-Bayesian method is preferred to estimate the reliability R_i.

Table 1: The Monte Carlo simulation results of $\hat{R}_{iEBj}$ $(j=1,2,3)$ for different values of c.

		c				
s_i	$\hat{R}_{iEBj}$	4	5	6	7	8
1000	$\hat{R}_{iEB1}$	0.999501	0.999501	0.999502	0.999502	0.999502
	$\hat{R}_{iEB2}$	0.999502	0.999502	0.999502	0.999502	0.999503
	$\hat{R}_{iEB3}$	0.999502	0.999502	0.999502	0.999503	0.999503
500	$\hat{R}_{iEB1}$	0.999005	0.999006	0.999007	0.999007	0.999008
	$\hat{R}_{iEB2}$	0.999006	0.999007	0.999008	0.999009	0.999010
	$\hat{R}_{iEB3}$	0.999007	0.999008	0.999009	0.999011	0.999012
100	$\hat{R}_{iEB1}$	0.995130	0.995145	0.995161	0.995176	0.995191
	$\hat{R}_{iEB2}$	0.995153	0.995176	0.995199	0.995222	0.995244
	$\hat{R}_{iEB3}$	0.995167	0.995197	0.995226	0.995256	0.995285
50	$\hat{R}_{iEB1}$	0.990504	0.990563	0.990620	0.990677	0.990732
	$\hat{R}_{iEB2}$	0.990593	0.990679	0.990762	0.990844	0.990924
	$\hat{R}_{iEB3}$	0.990646	0.990756	0.990864	0.990970	0.991074
20	$\hat{R}_{iEB1}$	0.977918	0.978223	0.978516	0.978798	0.979069
	$\hat{R}_{iEB2}$	0.978385	0.978821	0.979234	0.979627	0.980001
	$\hat{R}_{iEB3}$	0.978665	0.979219	0.979747	0.980249	0.980726
10	$\hat{R}_{iEB1}$	0.960392	0.961329	0.962204	0.963025	0.963797
	$\hat{R}_{iEB2}$	0.961847	0.963139	0.964322	0.965412	0.966419
	$\hat{R}_{iEB3}$	0.962720	0.964345	0.965835	0.967202	0.968459

5.2 An illustrative example

In this subsection, we present a real data set to assess the performances of the proposed methods for estimating reliability based on zero-failure data. We consider the zero-failure data of type-I censored life testing of engines from Han (2007), which is listed in Table 3 (time unit: hour).

According to the results of simulation study, we select $c=6$ as a supper limit of the hyper-parameter a. Therefore, according to the Theorems 1 and 2, we can easily obtain the E-Bayesian estimators $\hat{R}_{iEBj}$ $(j=1,2,3)$, and the hierarchical Bayesian estimators $\hat{R}_{iHBj}$ $(j=1,2,3)$ of R_i at each censoring time t_i $(i=1,2,...,9)$. The results are provided in Table 4 and Figure 1, respectively.

By Theorem 3, with different credible levels $1-\alpha$=0.99, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7, and taking c =6 as a upper limit of hyper-parameter a, we can then obtain the one-sided M-Bayesian lower credible limits $\hat{R}_{iMBLj}$ $(j=1,2,3)$ of R_i at each censoring time t_i $(i=1,2,...,9)$. The corresponding results are presented in Table 5 and Figures 2 and 3.

According to the Table 5 and Figures 2 and 3, we can see that the one-sided M-Bayesian lower credible limits of R_i satisfy the Corollary 3 under the same credible level 1-α. In addition, Table 5 and Figures 2 and 3 show that there ia a very small difference among the three kinds of one-sided M-Bayesian lower credible limits of R_i.

Table 2: The Monte Carlo simulation results of $\hat{R}_{iHBj}$ $(j=1,2,3)$ for different values of c.

		c				
s_i	$\hat{R}_{iHBj}$	4	5	6	7	8
1000	$\hat{R}_{iHB1}$	0.999845	0.999841	0.999838	0.999835	0.999832
	$\hat{R}_{iHB2}$	0.999839	0.999834	0.999830	0.999827	0.999823
	$\hat{R}_{iHB3}$	0.999836	0.999830	0.999825	0.999821	0.999817
500	$\hat{R}_{iHB1}$	0.999655	0.999645	0.999638	0.999631	0.999624
	$\hat{R}_{iHB2}$	0.999641	0.999630	0.999620	0.999612	0.999604
	$\hat{R}_{iHB3}$	0.999633	0.999620	0.999608	0.999598	0.999590
100	$\hat{R}_{iHB1}$	0.997749	0.997690	0.997640	0.997596	0.997556
	$\hat{R}_{iHB2}$	0.997660	0.997590	0.997530	0.997480	0.997438
	$\hat{R}_{iHB3}$	0.997612	0.997528	0.997460	0.997406	0.997357
50	$\hat{R}_{iHB1}$	0.994988	0.994876	0.994785	0.994704	0.994732
	$\hat{R}_{iHB2}$	0.994813	0.994685	0.994589	0.994506	0.994639
	$\hat{R}_{iHB3}$	0.994717	0.994572	0.994464	0.994377	0.994444
20	$\hat{R}_{iHB1}$	0.986136	0.985793	0.985880	0.985824	0.985802
	$\hat{R}_{iHB2}$	0.985863	0.985751	0.985707	0.985708	0.985757
	$\hat{R}_{iHB3}$	0.985725	0.985623	0.985605	0.985642	0.985728
10	$\hat{R}_{iHB1}$	0.971898	0.972036	0.972229	0.972512	0.972796
	$\hat{R}_{iHB2}$	0.971990	0.972361	0.972809	0.973292	0.973778
	$\hat{R}_{iHB3}$	0.972052	0.972548	0.973145	0.973780	0.974458

Table 3: The zero-failure data of some engine.

i	1	2	3	4	5	6	7	8	9
t_i	250	450	650	850	1050	1250	1450	1650	1850
n_i	3	3	3	3	4	4	4	4	4
s_i	32	29	26	23	20	16	12	8	4

Table 4: The estimators of $\hat{R}_{iEBj}$,$\hat{R}_{iHBj}$ $(j=1,2,3)$ of reliability of some engine at t_i $(i=1,2,...,9)$ $(c=6)$.

t_i	250	450	650	850	1050	1250	1450	1650	1850
$\hat{R}_{iEB1}$	0.985833	0.984515	0.982926	0.980974	0.978516	0.974040	0.967197	0.955402	0.930041
$\hat{R}_{iHB1}$	0.991436	0.990469	0.989283	0.987794	0.985878	0.982282	0.976559	0.966194	0.942310
$\hat{R}_{iEB2}$	0.986152	0.984895	0.983386	0.981541	0.979234	0.975075	0.968815	0.958294	0.936684
$\hat{R}_{iHB2}$	0.991199	0.990233	0.989053	0.987583	0.985706	0.982230	0.976807	0.967228	0.946544
$\hat{R}_{iEB3}$	0.986381	0.985167	0.983715	0.981974	0.979747	0.975814	0.969971	0.960359	0.941430
$\hat{R}_{iHB3}$	0.991053	0.990087	0.988911	0.987453	0.985601	0.982198	0.976961	0.967968	0.949221

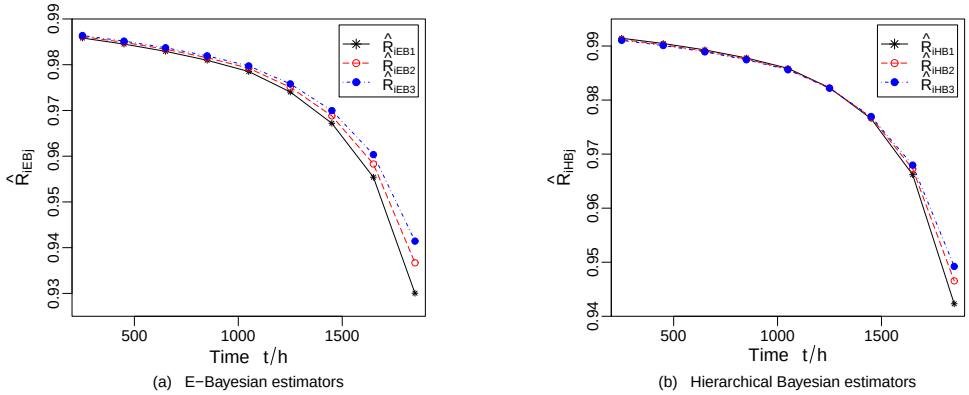


Figure 1: E-Bayesian estimators $\hat{R}_{iEBj}$ and hierarchical Bayesian estimators $\hat{R}_{iHBj}$ of reliability R_i.

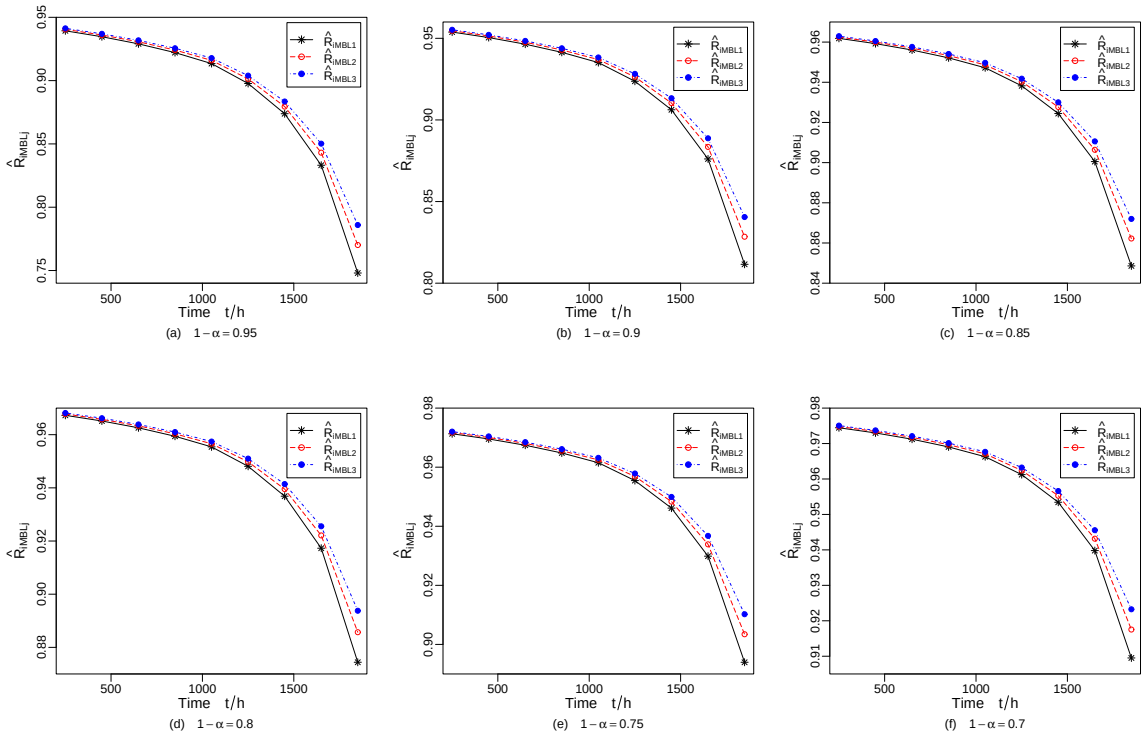


Figure 2: One-sided M-Bayesian lower credible limits of reliability R_i at different credible level 1-α.

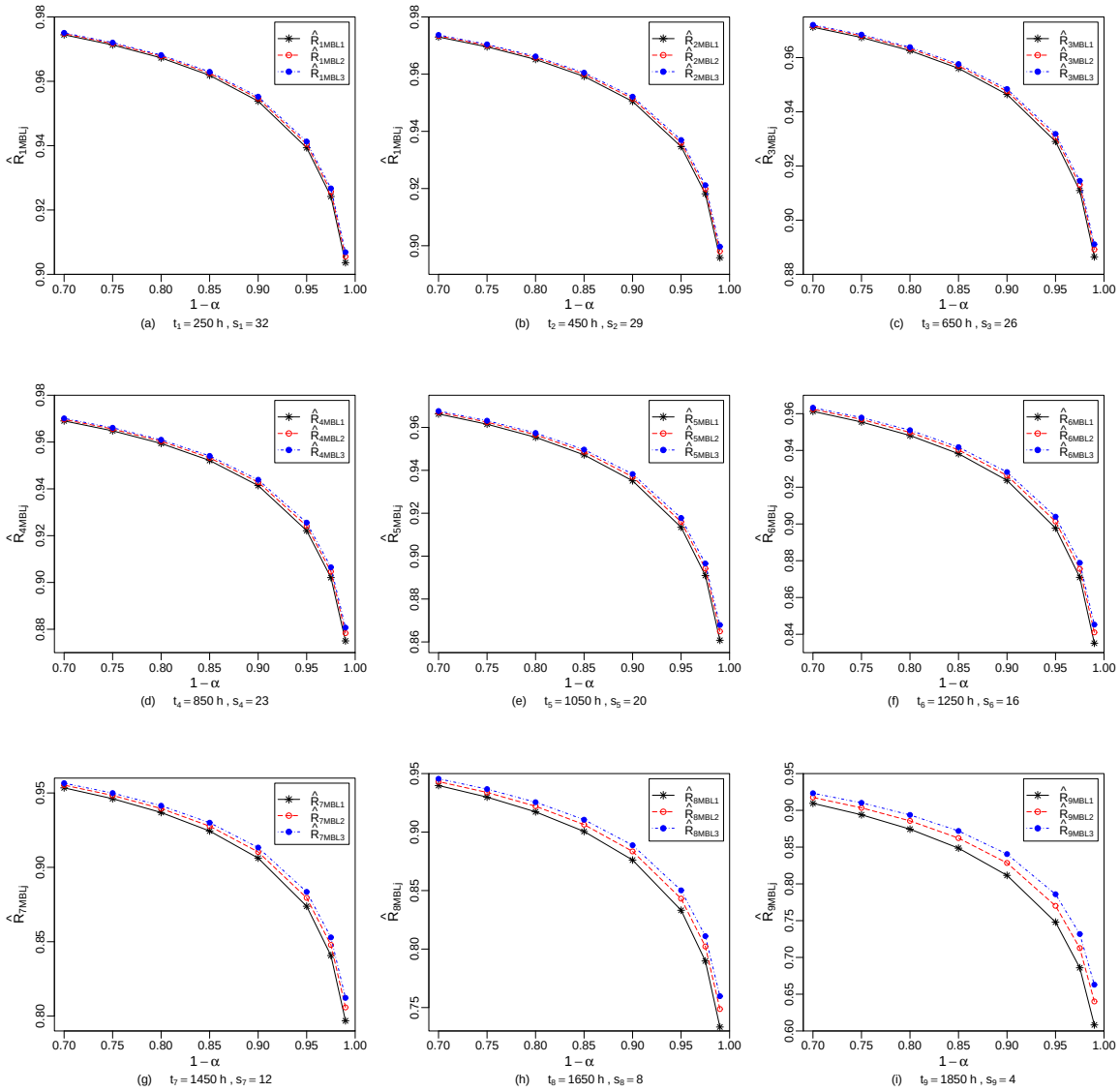


Figure 3: One-sided M-Bayesian lower credible limits of reliability R_i for different credible level at each censoring time.

Table 5: One-sided M-Bayesian credible lower limits of R_i for different credible levels 1-α (c =6).

			$1-\alpha$						
t_i	s_i	$\hat{R}_{iHBj}$	0.99	0.95	0.90	0.85	0.80	0.75	0.7
250	32	$\hat{R}_{iMBL1}$	0.903628	0.939340	0.953827	0.961854	0.967281	0.971307	0.974455
		$\hat{R}_{iMBL2}$	0.905528	0.940475	0.954643	0.962491	0.967797	0.971732	0.974809
		$\hat{R}_{iMBL3}$	0.906885	0.941285	0.955226	0.962947	0.968166	0.972036	0.975062
450	29	$\hat{R}_{iMBL1}$	0.895819	0.934663	0.950460	0.959222	0.965150	0.969549	0.972990
		$\hat{R}_{iMBL2}$	0.898069	0.936012	0.951431	0.959981	0.965765	0.970056	0.973413
		$\hat{R}_{iMBL3}$	0.899677	0.936976	0.952125	0.960523	0.966204	0.970419	0.973715
650	26	$\hat{R}_{iMBL1}$	0.886458	0.929035	0.946401	0.956046	0.962577	0.967424	0.971220
		$\hat{R}_{iMBL2}$	0.889164	0.930665	0.947576	0.956965	0.963322	0.968041	0.971733
		$\hat{R}_{iMBL3}$	0.891098	0.931828	0.948415	0.957622	0.963854	0.968480	0.972099
850	23	$\hat{R}_{iMBL1}$	0.875030	0.922131	0.941412	0.952138	0.959408	0.964807	0.969039
		$\hat{R}_{iMBL2}$	0.878347	0.924138	0.942863	0.953275	0.960330	0.965571	0.969673
		$\hat{R}_{iMBL3}$	0.880716	0.925571	0.943898	0.954086	0.960988	0.966144	0.970127
1050	20	$\hat{R}_{iMBL1}$	0.860764	0.913460	0.935132	0.947212	0.955410	0.961503	0.966282
		$\hat{R}_{iMBL2}$	0.864925	0.915993	0.936967	0.948652	0.956579	0.962472	0.967088
		$\hat{R}_{iMBL3}$	0.867897	0.917803	0.938278	0.949681	0.957414	0.963162	0.967662
1250	16	$\hat{R}_{iMBL1}$	0.835142	0.897735	0.923696	0.938223	0.948103	0.955459	0.961235
		$\hat{R}_{iMBL2}$	0.841038	0.901365	0.926340	0.940303	0.949794	0.956858	0.962404
		$\hat{R}_{iMBL3}$	0.845249	0.903960	0.928228	0.941788	0.951002	0.957861	0.963236
1450	12	$\hat{R}_{iMBL1}$	0.796842	0.873848	0.906212	0.924428	0.936862	0.946142	0.953443
		$\hat{R}_{iMBL2}$	0.805837	0.879485	0.910348	0.927695	0.939525	0.948351	0.955288
		$\hat{R}_{iMBL3}$	0.812262	0.883514	0.913301	0.930027	0.941428	0.949932	0.956608
1650	8	$\hat{R}_{iMBL1}$	0.733382	0.833161	0.876101	0.900522	0.917297	0.929874	0.939803
		$\hat{R}_{iMBL2}$	0.748753	0.843100	0.883482	0.906394	0.922107	0.933878	0.943158
		$\hat{R}_{iMBL3}$	0.759733	0.850202	0.888754	0.910587	0.925544	0.936744	0.945559
1850	4	$\hat{R}_{iMBL1}$	0.608222	0.748032	0.811571	0.848596	0.874404	0.893959	0.909514
		$\hat{R}_{iMBL2}$	0.640133	0.770141	0.828453	0.862237	0.885703	0.903440	0.917528
		$\hat{R}_{iMBL3}$	0.662926	0.785934	0.840513	0.871982	0.893774	0.910220	0.923242

6 Conclusion

This paper investigates point estimation and credible limits for the reliability of the binomial distribution in the case of zero-failure data from a Bayesian perspective.

For the point estimation of reliability based on zero-failure data, the authors consider the E-Bayesian and hierarchical Bayesian estimators, utilizing three different joint prior distributions for the two hyper-parameters in the prior distribution of reliability. Besides, closed-form expressions for the E-Bayesian estimators of reliability are obtained, and the hierarchical Bayesian estimators are evaluated using the Monte Carlo simulation method. The results of numerical examples and a real example indicate that both the E-Bayesian estimators and hierarchical Bayesian estimators are robust for different values of the upper bound c. Therefore, we suggest selecting the midpoint of integer interval [4, 8] as the value of the upper bound c in applications, namely, c=6. Besides, compared to the hierarchical Bayesian estimation method, the E-Bayesian estimation method is much easier to implement in practical engineering applications.

To obtain the credible limits of reliability based on zero-failure data, we study the one-sided M-Bayesian lower credible limits of reliability using three different joint prior distributions for two hyper-parameters. We propose estimation expressions for these one-sided M-Bayesian lower credible limits and discuss the properties of these estimators. The results of an illustrative example show that the proposed one-sided M-Bayesian lower credible limits perform excellently, and the properties of these estimators are stable.

Acknowledgement

The authors would like to thank the editor and the referees for their insightful comments and valuable suggestions, which have led to a substantial improvement in this paper.

Declaration of interest statement

The authors declare that there are no conflicts of interest related to this study.

Appendix A

The proof of Theorem 1. (i) For the zero-failure data (t_i, n_i) $(i = 1,2,...,k)$ from the k times type-I censored life testing. The likelihood function of R_i is

$$L(0|R_i) = R_i^{s_i}, \ \ i = 1,2,...,k.$$

We thus according to the Bayesian theorem, the posterior density function of R_i can be given by,

$$\pi(R_i|s_i) = \frac{L(0|R_i)\pi(R_i|a,b)}{\int_0^1 L(0|R_i)\pi(R_i|a,b)\mathrm{d}R_i} = \frac{R_i^{s_i+a-1}(1-R_i)^{b-1}}{B(s_i+a,b)}, \ \ i = 1,2,...,k,$$

this indicates that the posterior distribution of R_i is the $Beta(s_i+a,b)$ distribution. Therefore, using the squared error loss function, the Bayesian estimator of R_i is

$$\hat{R}_{iB}(a,b)=\int_0^1 R_i\pi(R_i|s_i)\mathrm{d}R_i=\frac{1}{B(s_i+a,b)}\int_0^1 R_i^{s_i+a}(1-R_i)^{b-1}\mathrm{d}R_i=\frac{s_i+a}{s_i+a+b},\ \ i=1,2,...,k.$$

(ii) If the prior distributions of hyper-parameters a and b is presented by (3) to (5), then according to the Definition 1, the E-Bayesian estimators of R_i can be provided as, respectively,

$$\begin{aligned}
\hat{R}_{iEB1}&=\int_1^c\int_0^1\frac{s_i+a}{s_i+a+b}\cdot\frac{2(c-a)}{(c-1)^2}\mathrm{d}b\mathrm{d}a\\
&=\frac{2}{(c-1)^2}\left[\int_1^c(s_i+a)(c-a)\ln(s_i+a+1)\mathrm{d}a-\int_1^c(s_i+a)(c-a)\ln(s_i+a)\mathrm{d}a\right]\\
&=\frac{2}{(c-1)^2}[G_1(s_i,c)-G_2(s_i,c)],\\
\hat{R}_{iEB2}&=\int_1^c\int_0^1\frac{s_i+a}{s_i+a+b}\cdot\frac{1}{c-1}\mathrm{d}b\mathrm{d}a\\
&=\frac{1}{c-1}\left[\int_1^c(s_i+a)\ln(s_i+a+1)\mathrm{d}a-\int_1^c(s_i+a)\ln(s_i+a)\mathrm{d}a\right]\\
&=\frac{1}{c-1}[G_3(s_i,c)-G_4(s_i,c)],
\end{aligned}$$

and

$$\begin{aligned}
\hat{R}_{iEB3}&=\int_1^c\int_0^1\frac{s_i+a}{s_i+a+b}\cdot\frac{2a}{c^2-1}\mathrm{d}b\mathrm{d}a\\
&=\frac{2}{c^2-1}\left[\int_1^c a(s_i+a)\ln(s_i+a+1)\mathrm{d}a-\int_1^c a(s_i+a)\ln(s_i+a)\mathrm{d}a\right]\\
&=\frac{2}{c^2-1}[G_5(s_i,c)-G_6(s_i,c)],
\end{aligned}$$

where

$$\begin{aligned}
G_1(s_i,c)&=\frac{1}{2}(s_i+c+2)q_1(s_i,c)-(s_i+c+1)q_2(s_i,c)-\frac{1}{3}q_3(s_i,c),\\
G_2(s_i,c)&=\frac{1}{2}(s_i+c)q_4(s_i,c)-\frac{1}{3}q_5(s_i,c),\\
G_3(s_i,c)&=\frac{1}{2}q_1(s_i,c)-q_2(s_i,c),\\
G_4(s_i,c)&=\frac{1}{2}q_4(s_i,c),\\
G_5(s_i,c)&=-\frac{1}{2}(s_i+2)q_1(s_i,c)+(s_i+1)q_2(s_i,c)+\frac{1}{3}q_3(s_i,c),\\
G_6(s_i,c)&=\frac{1}{3}q_5(s_i,c)-\frac{s_i}{2}q_4(s_i,c),
\end{aligned}$$

$$
\begin{aligned}
q_1(s_i,c) &= (s_i+c+1)^2\left[\ln(s_i+c+1)-\frac{1}{2}\right]-(s_i+2)^2\left[\ln(s_i+2)-\frac{1}{2}\right],\\
q_2(s_i,c) &= (s_i+c+1)\left[\ln(s_i+c+1)-1\right]-(s_i+2)\left[\ln(s_i+2)-1\right],\\
q_3(s_i,c) &= (s_i+c+1)^3\left[\ln(s_i+c+1)-\frac{1}{3}\right]-(s_i+2)^3\left[\ln(s_i+2)-\frac{1}{3}\right],\\
q_4(s_i,c) &= (s_i+c)^2\left[\ln(s_i+c)-\frac{1}{2}\right]-(s_i+1)^2\left[\ln(s_i+1)-\frac{1}{2}\right],\\
q_5(s_i,c) &= (s_i+c)^3\left[\ln(s_i+c)-\frac{1}{3}\right]-(s_i+1)^3\left[\ln(s_i+1)-\frac{1}{3}\right].
\end{aligned}
$$

Appendix B

The proof of Theorem 2. Since the hierarchical prior densities function of R_i are given by (6) to (8), respectively, we then according to the Bayesian theorem, the posterior densities function of R_i can be expressed as follows, respectively,

$$
\begin{aligned}
h_1(R_i|s_i) &= \frac{L(0|R_i)\pi_4(R_i)}{\int_0^1 L(0|R_i)\pi_4(R_i)\mathrm{d}R_i} = \frac{\int_0^1\int_1^c (c-a)\dfrac{R_i^{s_i+a-1}(1-R_i)^{b-1}}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c (c-a)\dfrac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k,\\
h_2(R_i|s_i) &= \frac{L(0|R_i)\pi_5(R_i)}{\int_0^1 L(0|R_i)\pi_5(R_i)\mathrm{d}R_i} = \frac{\int_0^1\int_1^c \dfrac{R_i^{s_i+a-1}(1-R_i)^{b-1}}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c \dfrac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k,\\
h_3(R_i|s_i) &= \frac{L(0|R_i)\pi_6(R_i)}{\int_0^1 L(0|R_i)\pi_6(R_i)\mathrm{d}R_i} = \frac{\int_0^1\int_1^c a\dfrac{R_i^{s_i+a-1}(1-R_i)^{b-1}}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c a\dfrac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k.
\end{aligned}
$$

We therefore use the square error loss function, the corresponding hierarchical Bayesian estimator of R_i can be presented as, respectively,

$$\hat{R}_{iHB1} = \int_0^1 R_i h_1(R_i|s_i)\mathrm{d}R_i = \frac{\int_0^1\int_1^c (c-a)\dfrac{B(a+s_i+1)}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c (c-a)\dfrac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k,$$

$$\hat{R}_{iHB2} = \int_0^1 R_i h_2(R_i|s_i)\mathrm{d}R_i = \frac{\int_0^1\int_1^c \dfrac{B(a+s_i+1)}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c \dfrac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k,$$

$$\hat{R}_{iHB3} = \int_0^1 R_i h_3(R_i|s_i)\mathrm{d}R_i = \frac{\int_0^1\int_1^c a\dfrac{B(a+s_i+1)}{B(a,b)}\mathrm{d}a\mathrm{d}b}{\int_0^1\int_1^c a\dfrac{B(a+s_i,b)}{B(a,b)}\mathrm{d}a\mathrm{d}b}, \quad i=1,2,...,k.$$

Appendix C

The proof of Theorem 3. For the three different prior densities function $\pi_1(a,b), \pi_2(a,b), \pi_3(a,b)$ of hyper-parameters a and b shown in (3) to (5), respectively, taking use of the Definition 2 and formula (9) of the Corollary 2, the one-sided M-Bayesian lower credible limits of R_i are given by, respectively

$$\hat{R}_{iMBL1} = \iint_D \hat{R}_{iB}(a,b)\pi_1(a,b)\mathrm{d}a\mathrm{d}b = \frac{2}{(c-1)^2}\int_0^1\int_1^c \frac{(c-a)(a+s_i)}{a+s_i+bF_{1-\alpha}(2b,2(a+s_i))}\mathrm{d}a\mathrm{d}b, \quad i=1,2,..,k,$$

$$\hat{R}_{iMBL2} = \iint_D \hat{R}_{iB}(a,b)\pi_2(a,b)\mathrm{d}a\mathrm{d}b = \frac{1}{c-1}\int_0^1\int_1^c \frac{a+s_i}{a+s_i+bF_{1-\alpha}(2b,2(a+s_i))}\mathrm{d}a\mathrm{d}b, \quad i=1,2,..,k,$$

$$\hat{R}_{iMBL3} = \iint_D \hat{R}_{iB}(a,b)\pi_3(a,b)\mathrm{d}a\mathrm{d}b = \frac{2}{c^2-1}\int_0^1\int_1^c \frac{a(a+s_i)}{a+s_i+bF_{1-\alpha}(2b,2(a+s_i))}\mathrm{d}a\mathrm{d}b, \quad i=1,2,..,k.$$

References

Bailey, R.T. (1997). Estimation from zero-failure data. *Risk Analysis*, **17**, 375–380.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd. Springer-Verlagm, New York.

Chen, J. D., Sun, W. L., and Li, B. X. (1993). The confidence limits for reliability parameters in the case of failure data. *In Proceedings of the Asian Conference on Statistics*, **1993**, 17–20.

Chen, J. D., Sun, W. L., and Li, B. X. (1995). On the confidence limits in the case of no failure data. *Acta Mathmaticiae Applicatae Sinica*, **18**, 90–100. (In Chinese)

Fan, T. H., Balakrishnan, N., and Chang, C. C. (2009). The Bayesian approach for highly reliable electro-explosive devices using one-shot device testing. *Journal of Statistical Computation and Simulation*, **79**, 1143–1154.

Fan, T. H., and Chang, C. C. (2009). A Bayesian zero-failure reliability demonstration test of high quality electro-explosive devices. *Quality and Reliability Engineering International*, **25**, 913–920.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. University of Minnesota Press.

Han, M. (2007). E-Bayesian estimation of failure probability and its application. *Mathematical and Computer Modelling*, **45**, 1272–1279.

Han, M. (2008). Two-sided M-Bayesian credible limits method of reliability parameters and its applications. *Communications in Statistics-Theory and Methods*, **37**, 1659–1670.

Han, M. (2009). E-Bayesian estimation and hierarchical Bayesian estimation of failure rate. *Applied Mathematical Modelling*, **33**, 1915–1922.

Han, M. (2011). E-Bayesian estimation of the reliability derived from Binomial distribution. *Applied Mathematical Modelling*, **35**, 2419–2424.

Han, M. (2012). The M-Bayesian credible limits of the reliability derived from binomial distribution. *Communications in Statistics-Theory and Methods*, **41**, 3814–3830.

Jiang, P., Lim, J. H., Zuo, M. J., and Guo, B. (2010). Reliability estimation in a Weibull lifetime distribution with zero-failure field data. *Quality and Reliability Engineering International*, **26**, 691–701.

Jiang, P., Xing, Y. Y., Jia, X., and Guo, B. (2015). Weibull failure probability estimation based on zero-failure data. *Mathematical Problems in Engineering*, **2015**, 681232.

Lindley, D. V., and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**, 1–18.

Li, H. Y., Xie, L. Y., Li, M., Ren, J. G., Zhao, B. F., and Zhang, S. J. (2019). Reliability assessment of high-quality and long-life products based on zero-failure Data. *Quality and Reliability Engineering International*, **35**, 470–482.

Mao, S. S., and Luo, C. B. (1989). Reliability analysis of zero-failure data. *Mathematical Theory and Applied Probability*, **4**, 489–506. (In Chinese)

Mao, S. S., and Xia, J. F. (1992). The hierarchical Bayesian analysis of the zero-failure data. *Applied Mathematics-A Journal of Chinese Universities*, **7**, 411–424.

Martz, H. F., and Waller, R. A. (1979). A Bayesian zero-failure (BAZE) reliability demonstration testing procedure. *Journal of Quality Technology*, **11**, 128–138.

Nam, J. S., Kim, H. E., and Kim, K. U. (2013). A new accelerated zero-failure test model for rolling bearings under elevated temperature conditions. *Journal of Mechanical Science and Technology*, **27**, 1801–1807.

Tian, Q. U., and Chen, Y. P. (2014). Two-sided M-Bayesian credible limits of reliability parameters in the case of zero-failure data for exponential distribution. *Applied Mathematical Modelling*, **38**, 2856–2600.

Xia, X. T. (2012). Reliability analysis of zero-failure data with poor information. *Quality and Reliability Engineering International*, **28**, 981–990.

Zhang, L. L., Jin, G. , and You, Y. (2019). Reliability assessment for very few failure data and Weibull distribution. *Mathematical Problems in Engineering*, **1**, 8947905.

Zhang, C. W. (2021). Weibull parameter estimation and reliability analysis with zero-failure data from high-quality products. *Reliability Engineering and System Safety*, **207**, 107321.

Zhang, C. W., Pan, R., and Goh, T. N. (2021). Reliability assessment of high-Quality new products with data scarcity. *International Journal of Production Research*, **59**, 4175–4187.

Stochastic Models in Probability and Statistics
Vol. 1, No. 2, 2024, pp. 211–228. Research Paper

A two-piece scale mixture normal measurement error models for replicated data

Fereshteh Kahrari*

Department of Statistical Sciences, University of Padova, Italy
Email(s): fereshteh.kahrari@unipd.it

Abstract. We develop a new class of flexible replicated measurement error models (RMEM) based on the normal two-piece scale mixture (TP-SMN) family to model the distribution of the latent variable. In the proposed approach, the replicated observations are jointly modeled by a mixture of two components from a scale mixture skew-normal (SMSN) density. The flexibility of this class can enable the simultaneous accommodation of skewness, outliers, and multimodality. The proposed connection between the unobserved covariates and the response facilitates the construction of an EM-type algorithm to perform maximum likelihood estimation. The effectiveness of the maximum likelihood estimations is studied through the simulation studies. Also, the method is applied to analyze continuing survey data on food intake by individuals on diet habits.

Keywords: ECM algorithm; Equation error; Replicated measurement error model; Two-piece scale mixture normal.

1 Introduction

A broad coverage of the measurement error model (MEM) has been extensively studied in the literature; see, for example, Carroll et al. (2006), Cheng and Van Ness (1999), Fuller (1987), Gustafson (2004) and Buonaccorsi (2010). These studies assumed that the distribution of the random errors and the unobserved covariate is Gaussian, which is sometimes not feasible as the model is sensitive to skewness, outliers, and unobserved heterogeneity in the data. In the context of MEMs, this can lead to some particular problems for these types of models. First, a non-identifiability problem in normal MEMs (Reiersol (1950)) can occur because one can not establish a single relationship between the parameters of the jointly distribution of the observations and the parameters of the model, consequently, no consistent estimator of the key

*Corresponding author
Received: 15 October 2024 / Accepted: 7 December 2024
DOI: 10.22067/smps.2024.46110

© 2024 Ferdowsi University of Mashhad

parameter, i.e., slope parameter, exists. To avoid this difficulty we may use prior knowledge of the error variances or make some assumptions on the error variances in advance, while such prior knowledge/assumptions are usually not easy to justify. Fortunately, the non-identifiability problem can be overcome in the case of replicated measurement error models (RMEMs) as the error variances can be estimated separately or together with other parameters. In this case, we refer the reader to Chan and Mak (1979), Isogawa (1985), and Lin et al. (2004). Second, recently, Lin and Cao (2013) considered a new replicated structural MEM in which the replicated observations jointly follow scale mixtures of normal (SMN) distributions. The scale mixture normal replicated measurement error model (SMN-RMEM) provides an appealing robust alternative to the usual model based on normal distributions, which also takes advantage of SMN distributions to accommodate extreme and outlying observations. Cao et al. (2015) considered models containing both error-prone covariates and predictors measured without errors, under multivariate SMN-RMEM, providing appealing robust and adaptable alternatives to the usual Gaussian assumptions. Third, the normality assumption and even SMN distributions, however, may be violated when a data set contains asymmetric outcomes. Cao et al. (2018) developed a new RMEM in the class of scale mixtures of skew-normal (SMSN) distributions. There can still be problems related to the simultaneous occurrence of skewness, discrepant observations, and multimodality.

Unlike other constructions in RMEMs, our model is formulated by replacing the normal assumption of the classical formulations by a more flexible class of distributions, called two-piece scale mixture normal (TP-SMN) distributions for one of the components, while retaining SMN for the others. Specific distributions in this family, including univariate versions of two-piece normal (TP-N), two-piece t (TP-T) with ν degrees of freedom, two-piece slash (TP-SL) and two-piece contaminated normal (TP-CN), are examined for the unobserved value of the covariates. In this approach, latent covariates and random observational errors are jointly modeled by a two-component mixture of SMSN densities. However, we view our construction as based on the TP-SMN assumption, as highlighted by the title, while the two-component SMSN likelihood is a mathematical implication of the assumption. Moreover, unlike a similar model previously proposed by Cao et al. (2018), the SMSN and TP-SMN families have different properties (for example, different tail behavior). In addition, numerical stability might also be a property that could be used to motivate our proposed model for the explanatory variable. This means that estimating the skewness parameter in the TP-SMN class of distributions is often easier, and more stable than estimating the shape parameter in the SMSN family (which controls the asymmetry, tails, location of the mode, and spread of the density).

The rest of this paper is organized as follows. Section 2 gives a brief description of the TP-SMN and SMSN distributions. Section 3, the replicated structural measurement error model with TP-SMN distributions (TP-SMN-RMEM) is defined, Proposition 3.1 represents the main result for this section. In Section 4, the advantages in terms of efficiency and robustness of MLEs of model (4) based on the ECM technique studied by Meng and Rubin (1993) are considered. A closed form expression is also obtained for the asymptotic covariance matrix of the ML estimators. In Section 5, the performance of the model and the importance of equation error are examined via simulation studies. Section 6 applies the model to analyze the inner relationship between saturated fat and caloric intake in CSFII data. Some conclusions are given in Section 7.

2 Asymmetric heavy-tailed distributions

2.1 Two-piece scale mixture normal (TP-SMN) distributions

Following Andrews and Mallows (1974) the pdf of random variable $X \sim SMN(\mu, \sigma, \nu)$ is denoted as,

$$f_{SMN}(x;\mu,\sigma,\nu) = \int_0^\infty \phi\left(x;\mu,\kappa(u)\sigma^2\right) dH(u;\nu), \quad y \in \mathbb{R},$$

where $\phi(.;\mu,\sigma^2)$ represents the pdf of $N(\mu,\sigma^2)$ distribution, $H(.;\nu)$ is the cdf of the scale mixing random variable U which is indexed by parameter ν. Also, $X \sim SMN(\mu,\sigma,\nu)$ has the stochastic representation given by

$$X \stackrel{d}{=} \mu + \sigma\kappa^{1/2}(U)W, \quad x \in \mathbf{R},$$

where W follows the standard normal distribution and is assumed to be independent of U.

The TP-SMN family is an analogy and alternative to the SMSN family, which contains the light/heavy-tailed and symmetry/asymmetry members including the TP-N, TP-t, TP-SL, and TP-CN distributions.

Definition 1. *Following a general two-piece distribution from Arellano-Valle et al. (2005), the pdf of random variable $Y \sim TP-SMN(\mu,\sigma,\gamma,\nu)$ for $y \in \mathbb{R}$ can be defined as,*

$$f_{TP-SMN}(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2(1-\gamma) f_{SMN}(y;\mu,\sigma(1-\gamma),\nu), & I_{(-\infty,\mu]}(y) \\ 2\gamma f_{SMN}(y;\mu,\sigma\gamma,\nu), & I_{[\mu,\infty)}(y), \end{cases} \tag{1}$$

where $\gamma \in (0,1)$ is the slant parameter, $I_A(x)$ denotes the indicator function of the set A.

Maleki and Mahmoudi (2017) studied the MLE problem for the parameters of the TP-SMN family using an EM-type algorithm. Barkhordar et al. (2022) proposed and examined the performance of a Bayesian approach for a homoscedastic nonlinear regression (NLR) model assuming errors with TP-SMN distributions. Maleki et al. (2019c) and Hoseinzadeh et al. (2021) examined the performance of the TP-SMN family in the context of NLR models (TP-SMN-NLR) using an EM-type algorithm to obtain MLEs for the parameters. Zarei et al. (2022) developed a general class of robust mixture regression model based on two-piece scale mixtures of normal distributions (TP-SMN) . For more details of stochastic representations, statistical inferences, and applications of the TP-SMN family, see Maleki et al. (2019d), Moravveji et al. (2019), Arellano-Valle et al. (2020), and Ghasami et al. (2020).

Corollary 1. *Let $Y \sim TP-SMN(\mu,\sigma,\gamma,\nu)$, then Y has a stochastic representation given by,*

i. $Y \stackrel{d}{=} \mu + \sigma\kappa^{1/2}(U)W_\gamma V$, where V is a continuous random variable with density function $2\phi(v)I_{[0,\infty)}(v)$, the standard half-normal density, and W_γ is an independent discrete random variable with probability function

$$p(w;\gamma) = \gamma^{(1+s)/2}(1-\gamma)^{(1-s)/2} I_{\{-1,1\}}(s), \tag{2}$$

with $s = sign(w)$. Equivalently, if $Y \stackrel{d}{=} \mu + \sigma W_\gamma |X|$, where $X \sim SMN(0,1,\nu)$ and is independent of W_γ, then $Y \sim TP-SMN(\mu,\sigma,\gamma,\nu)$.

2.2 Examples of the TP-SMN distributions

In this section, we present some members of the TP-SMN family which are examined for the unobserved value of the covariates. We consider the case where $\kappa(u) = 1/u$.

Two-piece normal distribution Two-piece normal, TP-N(μ, σ, γ), distribution is obtained from Eq. (1) when $P(U = 1) = 1$.

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2(1-\gamma)\phi(y;\mu,\sigma^2(1-\gamma)^2), & I_{(-\infty,\mu]}(y), \\ 2\gamma\phi(y;\mu,\sigma^2\gamma^2), & I_{[\mu,\infty)}(y). \end{cases}$$

Arellano-Valle et al. (2020) studied the main properties of the TP-N distribution.

Two-piece t distribution The pdf of a two-piece t distribution with ν degree of freedom, TP-t$(\mu, \sigma, \gamma, \nu)$ say, is derived from Eq. (1) by taking U to be distributed as Gamma$(\nu/2, \nu/2)$, $\nu > 0$.

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left(1+\frac{1}{\nu}(\frac{y-\mu}{\sigma(1-\gamma)})^2\right)^{-\frac{\nu+1}{2}}, & I_{(-\infty,\mu]}(y), \\ 2\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left(1+\frac{1}{\nu}(\frac{y-\mu}{\sigma\gamma})^2\right)^{-\frac{\nu+1}{2}}, & I_{[\mu,\infty)}(y). \end{cases}$$

Two-piece slash distribution Two-piece slash distribution, denoted by TP-SL$(\mu, \sigma, \gamma, \nu)$, arises when U has $Beta(\nu, 1)$ distribution. The TP-SL pdf is then given by

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2\nu(1-\gamma)\int_0^1 u^{\nu-1}\phi(y;\mu,u^{-1}\sigma^2(1-\gamma)^2)du, & I_{(-\infty,\mu]}(y), \\ 2\nu\gamma\int_0^1 u^{\nu-1}\phi(y;\mu,u^{-1}\sigma^2\gamma^2)du, & I_{[\mu,\infty)}(y). \end{cases}$$

Two-piece contaminated normal distribution Another member of the TP-SMN family is known as the two-piece contaminated normal distribution, and denoted by $TP-CN(\mu, \sigma, \gamma, \nu_1, \nu_2)$, arises when U is a discrete random variable with probability function

$$h(u;\nu_1,\nu_2) = \nu_1\mathbb{I}_{(u=\nu_2)} + (1-\nu_1)\mathbb{I}_{(u=1)}, \text{ where } 0 < \nu_i < 1 \text{ for } i = 1,2.$$

In this case, the random variable Y has pdf given by

$$f(y;\mu,\sigma,\gamma,\nu) = \begin{cases} 2(1-\gamma)(1-\nu_1)\phi(y;\mu,\sigma^2(1-\gamma)^2), & I_{(-\infty,\mu]}(y), \\ 2\gamma\nu_1\phi(y;\mu,\nu_2^{-1}\sigma^2\gamma^2), & I_{[\mu,\infty)}(y). \end{cases}$$

2.3 Scale mixture skew-normal distributions

Following Branco and Dey (2001) a random vector $\mathbf{Y}$ has a multivariate scale mixture skew-normal distribution (SMSN) with location vector $\boldsymbol{\mu}$, positive definite scale matrix $\boldsymbol{\Omega}$ and skewness/shape vector $\boldsymbol{\lambda}$, denoted by $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \boldsymbol{\nu})$, if its pdf is given by

$$f_{SMSN}(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda},\boldsymbol{\nu}) = \int_0^\infty 2\phi_p(\mathbf{y};\boldsymbol{\mu},\kappa(u)\boldsymbol{\Omega})\,\Phi\left(\kappa^{-1/2}(u)A\right)dH(u;\boldsymbol{\nu}),$$

where $A = \boldsymbol{\lambda}^\top \boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})$. The SMSN random vector $\mathbf{Y}$ can be introduced as the location-scale transformation, following the stochastic representation given by

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{X}, \tag{3}$$

where following Arellano-Valle and Genton (2005), the random vector $\mathbf{X}$ has a multivariate skew-normal (SN) distribution denoted by $\mathbf{X} \sim SN_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda})$ with the pdf given by

$$f_{SN}(\mathbf{x}; \mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{x}; \mathbf{I}_p, \boldsymbol{\lambda})\Phi(\boldsymbol{\lambda}^\top \mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^p.$$

From (3) it follows straightforward that $\mathbf{Y}|U = u \sim SN_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Omega}, \boldsymbol{\lambda})$. Thus, using Eq.(3), it can be shown that the mean vector and variance-covariance matrix of $\mathbf{Y}$ are given, respectively, by

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\mathrm{E}\left(\kappa^{\frac{1}{2}}(U)\right)\boldsymbol{\Delta}, \\ V(\mathbf{Y}) &= \mathrm{E}\left(\kappa(U)\right)\boldsymbol{\Omega} - \frac{2}{\pi}\mathrm{E}\left(\kappa^{\frac{1}{2}}(U)\right)^2 \boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \end{aligned}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$, and $\boldsymbol{\Delta} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\delta}$.

3 The TP-SMN structural RMEM

The proposed model can be described below. Suppose that we observe (X_t, Y_t), as surrogates of true (latent) unobserved variables (x_t, y_t), plus additive measurement errors (δ_t, ξ_t), that is, they satisfy an incomplete linear relationship $y_t = \alpha + \beta x_t + e_t$, in which the equation error e_t means that in some situations, the true variables are not perfectly related if factors other than x_t are responsible for the variation in y_t. Note that the equation error does not need to exist, which means that $e_t = 0$. In this case, the model is known as a no-equation-error model. Suppose that x_t and y_t are observed p and q times (respectively) to make replicated observations $X_t^{(i)}$ and $Y_t^{(j)}$.

$$\begin{cases} X_t^{(i)} = x_t + \delta_t^{(i)}, & i = 1, \ldots, p, \\ Y_t^{(j)} = y_t + \xi_t^{(j)}, & j = 1, \ldots, q, \\ y_t = \alpha + \beta x_t + e_t, & t = 1, \ldots, n. \end{cases} \tag{4}$$

By introducing the vectors $\mathbf{Z}_t = (\mathbf{X}_t^\top, \mathbf{Y}_t^\top)^\top$, where $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(p)})^\top$ and $\mathbf{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(q)})^\top$, and $\boldsymbol{\epsilon}_t = (\delta_t^{(1)}, \ldots, \delta_t^{(p)}, e_t + \xi_t^{(1)}, \ldots, e_t + \xi_t^{(q)})^\top$, $\boldsymbol{a} = (\mathbf{0}_p^\top, \alpha\mathbf{1}_q^\top)^\top$, $\boldsymbol{b} = (\mathbf{1}_p^\top, \beta\mathbf{1}_q^\top)^\top$. Model (4) can be written compactly as

$$\mathbf{Z}_t = \boldsymbol{a} + \boldsymbol{b}x_t + \boldsymbol{\epsilon}_t. \tag{5}$$

Thus, from Eq. (5) the distribution of $\mathbf{Z}_t$ becomes specified once the joint distribution of $\mathbf{r}_t = (x_t, \boldsymbol{\epsilon}_t{}^\top)^\top$ is specified. To obtain robust estimation of the parameters, we propose to replace

the normal assumption by

$$\begin{aligned}
&U_t \overset{iid}{\sim} H(u;\boldsymbol{\nu}), \quad t=1,\ldots,n,\\
&x_t|U_t=u_t \overset{iid}{\sim} TP-N(\mu_x,u_t^{-1/2}\sigma_x,\gamma_x),\ i.e.,\ x_t \overset{iid}{\sim} TP-SMN(\mu_x,\sigma_x,\gamma_x,\boldsymbol{\nu}),\\
&\delta_t^{(i)}|U_t=u_t \overset{iid}{\sim} N(0,u_t^{-1}\phi_\delta),\ i.e.,\ \delta_t^{(i)} \overset{iid}{\sim} SMN(0,\phi_\delta,\boldsymbol{\nu}),\\
&\xi_t^{(j)}|U_t=u_t \overset{iid}{\sim} N(0,u_t^{-1}\phi_\xi),\ i.e.,\ \xi_t^{(j)} \overset{iid}{\sim} SMN(0,\phi_\xi,\boldsymbol{\nu}),\\
&e_t|U_t=u_t \overset{iid}{\sim} N(0,u_t^{-1}\phi_e),\ i.e.,\ e_t \overset{iid}{\sim} SMN(0,\phi_e,\boldsymbol{\nu}),
\end{aligned} \tag{6}$$

which we call the two-piece scale mixture replicated measurement error model (TP-SMN-RMEM). From (5) and (6), since for each t, x_t and $\boldsymbol{\epsilon}_t$ are indexed by the same scale mixing factor U_t, they are not independent. However, x_t and $\boldsymbol{\epsilon}_t$ are conditionally independent, which implies $\mathrm{cov}(x_t,\boldsymbol{\epsilon}_t|U_t)=0$. Model (4) can be specified equivalently in the hierarchical form as

$$\begin{aligned}
&\mathbf{Z}_t|x_t,U_t=u_t \overset{iid}{\sim} N_m(\boldsymbol{a}+\boldsymbol{b}x_t,u_t^{-1}\boldsymbol{\Sigma}),\\
&x_t|U_t=u_t \overset{iid}{\sim} TP-N(\mu_x,u_t^{-1/2}\sigma_x,\gamma_x),\\
&U_t \overset{iid}{\sim} H(u_t;\boldsymbol{\nu}),
\end{aligned} \tag{7}$$

where $m=p+q$, $\boldsymbol{c}=(\mathbf{0}_p^\top,\mathbf{1}_q^\top)^\top$, $\mathbf{D}(.)$ denotes a diagonal matrix and $\boldsymbol{\Sigma}$ is a $m\times m$ block matrix as follows

$$\boldsymbol{\Sigma}=\mathbf{D}(\phi)+\phi_e\mathbf{c}\mathbf{c}^\top=\begin{bmatrix}\boldsymbol{\Sigma}_\delta & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{e,\xi}\end{bmatrix}, \tag{8}$$

with $\boldsymbol{\Sigma}_\delta=\phi_\delta\mathbf{I}_p$ and $\boldsymbol{\Sigma}_{e,\xi}=\phi_\xi\mathbf{I}_q+\phi_e\mathbf{1}_q\mathbf{1}_q{}^\top$.

The following proposition represents the main result for this section. The vector of parameters is denoted by $\boldsymbol{\theta}=(\alpha,\beta,\mu_x,\sigma_x,\gamma_x,\phi_e,\phi_\delta,\phi_\xi)^\top$.

Proposition 1. *Under the hierarchical representation defined by (7), the observed random vectors* $(\mathbf{z}_1,\ldots,\mathbf{z}_n)$ *are drawn independently from the common distribution given by*

$$\begin{aligned}
f(\mathbf{z}_t;\boldsymbol{\theta})=&\gamma_x f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})\\
&+(1-\gamma_x)f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu}), \qquad \mathbf{z}_t\in\mathbb{R}^m,
\end{aligned} \tag{9}$$

where

$$\begin{aligned}
&\boldsymbol{\mu}=\boldsymbol{a}+\boldsymbol{b}\mu_x,\\
&\boldsymbol{\Omega}_{\gamma_x}=\boldsymbol{\Sigma}+{\gamma_x}^2\sigma_x^2\mathbf{b}\mathbf{b}^\top,\\
&\boldsymbol{\lambda}_{\gamma_x}=\frac{\gamma_x\sigma_x}{\sqrt{1-\sigma_x^2\gamma_x^2\mathbf{b}^\top\boldsymbol{\Omega}_{\gamma_x}^{-1}\mathbf{b}}}\boldsymbol{\Omega}_{\gamma_x}^{-1/2}\mathbf{b}.
\end{aligned}$$

Similarly, $\boldsymbol{\Omega}_{1-\gamma_x}$ *and* $\boldsymbol{\lambda}_{1-\gamma_x}$ *could be defined by using* $1-\gamma_x$ *instead of* γ_x.

Using the well-known Sherman-Morrison formula

$$\mathbf{\Omega}_{\gamma_x}^{-1} = \left(\mathbf{\Sigma} + \sigma_x^2\gamma_x^2\mathbf{b}\mathbf{b}^\top\right)^{-1} = \mathbf{\Sigma}^{-1} - \frac{\gamma_x^2\sigma_x^2\mathbf{\Sigma}^{-1}\mathbf{b}\mathbf{b}^\top\mathbf{\Sigma}^{-1}}{1+\sigma_x^2\gamma_x^2\mathbf{b}^\top\mathbf{\Sigma}^{-1}\mathbf{b}}.$$

Since we assume that $x_t \sim TP-SMN(\mu_x,\sigma_x,\gamma_x,\boldsymbol{\nu})$, based on Corollary 1, the stochastic representation of x_t is $x_t \overset{iid}{=} \mu_x + U_t^{-1/2}w_tv_t$ for $w_t \in \{\gamma_x, -(1-\gamma_x)\}$, which is a random variable with probability function $p(w_t;\gamma_x)$ given by (2), and $v_t \sim HN(0,\sigma_x^2;(0,\infty))$, and $U_t \sim H(.;\boldsymbol{\nu})$ then we have the equivalent formulation of our model.

Proposition 2. *The structural TP-SMN-RMEM defined by (7) has a hierarchical representation as follows,*

$$\begin{aligned}
&\mathbf{z}_t|v_t,w_t,u_t \overset{ind.}{\sim} N_m(\mathbf{a}+\mathbf{b}\mu_x+\mathbf{b}v_tw_t,u_t^{-1}\mathbf{\Sigma}),\\
&v_t|u_t \overset{i.d.d.}{\sim} HN(0,u_t^{-1}\sigma_x^2;(0,\infty)),\\
&w_t \overset{i.i.d.}{\sim} p(.;\gamma_x),\\
&U_t \overset{i.i.d.}{\sim} H(.;\boldsymbol{\nu}),
\end{aligned} \tag{10}$$

where $v_t|u_t$ and w_t are independent variables, for $t=1,\ldots,n$.

As a first consequence of Proposition 2, we compute the first two moments of the observation variables $\mathbf{z}_t$. Using Corollary 1, equation (5) and the hierarchical representation given by (7), we obtain

$$\begin{aligned}
\mathrm{E}(\mathbf{Z}_t) &= \mathrm{E}(\mathrm{E}(\mathbf{Z}_t|x_t)) = \mathbf{a}+\mathbf{b}\mu_x+\mathbf{b}\sigma_x(2\gamma_x-1)\sqrt{\frac{2}{\pi}}\mathrm{E}\left(U_t^{-1/2}\right),\\
\mathrm{var}(\mathbf{Z}_t) &= \mathrm{var}(\mathrm{E}(\mathbf{Z}_t|x_t)) + \mathrm{E}(\mathrm{var}(\mathbf{Z}_t|x_t))\\
&= \mathrm{E}\left(U_t^{-1}\right)\left(\mathbf{\Sigma}+[\gamma_x^3+(1-\gamma_x)^3]\sigma_x^2\mathbf{b}\mathbf{b}^\top\right)\\
&\quad -\frac{2}{\pi}\mathrm{E}\left(U_t^{-1/2}\right)^2(2\gamma_x-1)^2\sigma_x^2\mathbf{b}\mathbf{b}^\top.
\end{aligned}$$

Denoting the t-th complete dataset by $\mathbf{z}_{ct}=\{\mathbf{z}_t,v_t,w_t,u_t\}$, according to

$$\begin{aligned}
f(\mathbf{z}_t,v_t,w,u_t;\boldsymbol{\theta}) &= f(\mathbf{z}_t|v_t,w_t,u_t;\boldsymbol{\theta})f(v_t|u_t;\boldsymbol{\theta})p(w_t;\gamma_x)h(u_t;\boldsymbol{\nu})\\
&= \phi_m(\mathbf{z}_t;\mathbf{a}+\mathbf{b}\mu_x+\mathbf{b}v_tw_t,u_t^{-1}\mathbf{\Sigma})\\
&\quad\times 2\phi(v_t;0,u_t^{-1}\sigma_x^2)\times\gamma_x^{(1+s_t)/2}(1-\gamma_x)^{(1-s_t)/2}I_{\{-1,1\}}(s_t)h(u_t;\boldsymbol{\nu}),
\end{aligned} \tag{11}$$

which yields the following proposition.

Proposition 3. *Let us represent the random vector $\mathbf{z}_t \sim TP-SMN-RMEM$ as (9), the structural TP-SMN-RMEM defined by Proposition 2 and the joint distribution of $(\mathbf{z}_t,v_t,w_t,u_t)$ admits (11),*

thus following that,

$$
\begin{aligned}
&(i)\quad p(w_t|\mathbf{z}_t;\boldsymbol{\theta}) = \begin{cases} \pi_{\gamma_x t} & if \quad w_t = \gamma_x \\ \pi_{(1-\gamma_x)t} & if \quad w_t = -(1-\gamma_x), \end{cases} \\
&(ii)\quad f(v_t|w_t,u_t,\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\phi(v_t;\mu_{w_t},\sigma^2_{w_t}(u_t))}{\Phi(\tau_{w_t}(u_t))} I_{\{v_t\}}(0,\infty), \\
&(iii)\quad f(v_t|u_t,\mathbf{z}_t;\boldsymbol{\theta}) = \pi_{\gamma_x t}(u_t)\frac{\phi(v_t;\mu_{\gamma_x t},\sigma^2_{\gamma_x t}(u_t))}{\Phi(\tau_{\gamma_x t}(u_t))} I_{\{v_t\}}(0,\infty) \\
&\qquad\qquad + \pi_{(1-\gamma_x)t}(u_t)\frac{\phi(v_t;\mu_{(1-\gamma_x)t},\sigma^2_{(1-\gamma_x)t}(u_t))}{\Phi(\tau_{-(1-\gamma_x)t}(u_t))} I_{\{v_t\}}(0,\infty), \\
&(iv)\quad f(u_t|w_t,\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{w_t})\Phi(\tau_{w_t}(u_t))h(u_t;\boldsymbol{\nu})}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{w_t},\boldsymbol{\lambda}_{w_t},\boldsymbol{\nu})}, \\
&(v)\quad f(u_t|\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\gamma_x\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\tau_{\gamma_x t}(u_t))h(u_t;\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})} \\
&\qquad\qquad + \frac{(1-\gamma_x)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{1-\gamma_x})\Phi(\tau_{-(1-\gamma_x)t}(u_t))h(u_t;\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})},
\end{aligned}
$$

with $w_t = \gamma_x, -(1-\gamma_x)$, for $t = 1,\ldots,n$, leading to

$$
\mu_{w_t} = \frac{\sigma_x^2 w_t \mathbf{b}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_t-\boldsymbol{\mu})}{1+\sigma_x^2 w_t^2\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}},\ \sigma^2_{w_t}(u_t) = \frac{u_t^{-1}\sigma_x^2}{1+\sigma_x^2 w_t^2\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}}, \tau_{w_t}(u_t) = \frac{\mu_{w_t}}{\sigma_{w_t}(u_t)},
$$

$$
\begin{aligned}
\pi_{\gamma_x t} &= \frac{\gamma_x f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})}, \\
\pi_{(1-\gamma_x)t} &= \frac{(1-\gamma_x) f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})}{f(\mathbf{z}_t;\boldsymbol{\theta})}, \\
\pi_{\gamma_x t}(u_t) &= \frac{\gamma_x\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\tau_{\gamma_x t}(u_t))}{f(\mathbf{z}_t|u_t,\boldsymbol{\theta})}, \\
\pi_{(1-\gamma_x)t}(u_t) &= \frac{(1-\gamma_x)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{1-\gamma_x})\Phi(\tau_{-(1-\gamma_x)t}(u_t))}{f(\mathbf{z}_t|u_t,\boldsymbol{\theta})}, \\
f(\mathbf{z}_t|u_t,\boldsymbol{\theta}) &= \gamma_x\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\tau_{\gamma_x t}(u_t)) \\
&\quad + (1-\gamma_x)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},u_t^{-1}\boldsymbol{\Omega}_{1-\gamma_x})\Phi(\tau_{-(1-\gamma_x)t}(u_t)).
\end{aligned}
$$

4 Maximum likelihood estimation via the EM algorithm

The advantages in terms of efficiency and robustness of MLEs of model (4) based on the ECM technique studied by Meng and Rubin (1993) are considered. We start by considering the hierarchical representation (10) and the fact that

$$
p(w_t;\gamma_x) = \gamma_x{}^{(1+s_t)/2}(1-\gamma_x)^{(1-s_t)/2} I_{\{-1,1\}}(s_t).
$$

Denoting the estimate of $\boldsymbol{\theta}$ at the k-th iteration by $\widehat{\boldsymbol{\theta}}^{(k)}$ and $\mathbf{z}_c = \{\mathbf{z}, \mathbf{v}, \mathbf{w}, \mathbf{u}\}$ be the complete dataset of model (4), where $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$, $\mathbf{v} = \{v_1, \ldots, v_n\}$, $\mathbf{w} = \{w_1, \ldots, w_n\}$, and $\mathbf{u} = \{u_1, \ldots, u_n\}$. With $\kappa(u_t) = 1/u_t$ and using the equation given by (11), the log-likelihood function for $\boldsymbol{\theta}$ based on the t-th complete data, $\mathbf{z}_{ct}$, is in the following form

$$\ell(\boldsymbol{\theta}|\mathbf{z}_c) = \sum_{t=1}^{n} \ell(\boldsymbol{\theta}|\mathbf{z}_{ct}) = \sum_{t=1}^{n} \left(\ell_{\mathbf{z}_t|v_t,w_t,u_t} + \ell_{v_t|u_t} + \ell_{w_t}\right),$$

where

$$\begin{aligned}
\ell_{\mathbf{z}_t|v_t,w_t,u_t} &= -\frac{1}{2} log(|2\pi u_t^{-1}\boldsymbol{\Sigma}|) - \frac{u_t}{2}(\mathbf{z}_t - \mathbf{a})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}_t - \mathbf{a}) \\
&\quad + u_t(\mathbf{z}_t - \mathbf{a})^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}\mu_x + uwv_t(\mathbf{z}_t - \mathbf{a})^\top \boldsymbol{\Sigma}^{-1}\mathbf{b} \\
&\quad - uwv_t\mathbf{b}^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}\mu_x - \frac{u_t}{2}\mathbf{b}^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}\mu_x^2 - \frac{uw^2v^2{}_t}{2}\mathbf{b}^\top \boldsymbol{\Sigma}^{-1}\mathbf{b}, \\
\ell_{v_t|u_t} &= -\frac{1}{2} log(2\pi u_t^{-1}\sigma_x^2) - \frac{u_t}{2}\frac{v_t^2}{\sigma_x^2}, \\
\ell_{w_t} &= \frac{1}{2}(1 + s_t) log\gamma_x + \frac{1}{2}(1 - s_t) log(1 - \gamma_x).
\end{aligned}$$

Obviously, from equation (8), we have that $|\boldsymbol{\Sigma}| = \phi_\delta^p \phi_\xi^{q-1}(\phi_\xi + q\phi_e)$, and

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{22} \end{bmatrix} \tag{12}$$

in which $\boldsymbol{\Sigma}^{11} = \frac{1}{\phi_\delta}\mathbf{I}_p$ and $\boldsymbol{\Sigma}^{22} = \frac{1}{\phi_\xi}\mathbf{I}_q - \frac{\phi_e}{\phi_\xi(\phi_\xi + q\phi_e)}\mathbf{1}_q\mathbf{1}_q{}^\top$. The constants that are independent of $\boldsymbol{\theta}$ in the above expressions can be ignored. The EM algorithm is constructed as follows. Given the current value $\widehat{\boldsymbol{\theta}}^{(k)}$ of $\boldsymbol{\theta}$, the E-step of the EM algorithm calculates the Q-function defined by $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{t=1}^{n} Q_t(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$

$$\begin{aligned}
Q_t(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{p}{2}\log\phi_\delta - \frac{q-1}{2}\log\phi_\xi - \frac{1}{2}\log(\phi_\xi + q\phi_e), \\
&\quad - \frac{1}{2}\widehat{u}_t^{(k)}\left[\frac{1}{\phi_\delta}\sum_{i=1}^{p} X_t^{2(i)} + \frac{1}{\phi_\xi}\sum_{j=1}^{q}\left(Y_t^{(j)} - \alpha\right)^2 - \frac{q^2\phi_e}{\phi_\xi(\phi_\xi + q\phi_e)}(\bar{Y}_t - \alpha)^2\right] \\
&\quad + \widehat{u}_t^{(k)}\mu_x\left[\frac{p}{\phi_\delta}\bar{X}_t + \frac{q\beta}{(\phi_\xi + q\phi_e)}(\bar{Y}_t - \alpha)\right] \\
&\quad + \widehat{uwv}_t^{(k)}\left[\frac{p}{\phi_\delta}\bar{X}_t + \frac{q\beta}{(\phi_\xi + q\phi_e)}(\bar{Y}_t - \alpha)\right] \\
&\quad - \widehat{uwv}_t^{(k)}\mu_x\left[\frac{p}{\phi_\delta} + \frac{q\beta^2}{(\phi_\xi + q\phi_e)}\right] \\
&\quad - \frac{1}{2}\widehat{u}_t^{(k)}\mu_x^2\left[\frac{p}{\phi_\delta} + \frac{q\beta^2}{(\phi_\xi + q\phi_e)}\right] \\
&\quad - \frac{1}{2}\widehat{uw^2v^2}_t^{(k)}\left[\frac{p}{\phi_\delta} + \frac{q\beta^2}{(\phi_\xi + q\phi_e)}\right] \\
&\quad - \log\sigma_x - \frac{1}{2\sigma_x^2}\widehat{uv^2}_t^{(k)} + \frac{1}{2}(1 + \widehat{s}_t^{(k)})\log\gamma_x + \frac{1}{2}(1 - \widehat{s}_t^{(k)})\log(1 - \gamma_x),
\end{aligned} \tag{13}$$

in which $\mathbf{Y}_t = (\mathbf{Y}_t^{(1)}, \ldots, \mathbf{Y}_t^{(q)})^\top$, $\boldsymbol{\mu}_y = (\alpha + \beta\mu_x)\mathbf{1}_q$, $\boldsymbol{\beta} = \beta\mathbf{1}_q$. Furthermore, to obtain the Q-function, we first need to calculate the following conditional expectations, which must be evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$.
E-step:

$$\begin{aligned}
\widehat{s}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{s_t|\mathbf{Z}_t = \mathbf{z}_t\},\\
\widehat{u}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{U_t|\mathbf{Z}_t = \mathbf{z}_t\},\\
\widehat{uv^2}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{U_t v_t^2|\mathbf{Z}_t = \mathbf{z}_t\},\\
\widehat{uwv}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{U_t w_t v_t|\mathbf{Z}_t = \mathbf{z}_t\},\\
\widehat{uw^2v^2}_t^{(k)} &= E_{\widehat{\boldsymbol{\theta}}^{(k)}}\{U_t w_t^2 v_t^2|\mathbf{Z}_t = \mathbf{z}_t\},
\end{aligned}$$

where $E_{\widehat{\boldsymbol{\theta}}^{(k)}}$ means that the expectation is computed at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$. The derivation of these moments will be useful in the implementation of the EM algorithm.

Proposition 4. *The structural TP-SMN-RMEM defined by Proposition 2 and the joint distribution of* $(\mathbf{z}_t, v_t, w_t, u_t)$ *admits* (11)*, thus following that*

$$\begin{aligned}
&(i)\quad E\{s_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \pi_{\gamma_x t} - \pi_{(1-\gamma_x)t},\\
&(ii)\quad E\{v_t^2 U_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \pi_{\gamma_x t}\left\{\frac{\chi_t(\gamma_x)\mu_{\gamma_x t}^2 + \zeta_t(\gamma_x)\mu_{\gamma_x t}\sigma_{\gamma_x t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})} + \sigma_{\gamma_x t}^2\right\}\\
&\qquad + \pi_{(1-\gamma_x)t}\left\{\frac{\chi_t(1-\gamma_x)\mu_{(1-\gamma_x)t}^2 + \zeta_t(1-\gamma_x)\mu_{(1-\gamma_x)t}\sigma_{(1-\gamma_x)t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})} + \sigma_{(1-\gamma_x)t}^2\right\},\\
&(iii)\quad E\{v_t w_t U_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \gamma_x\pi_{\gamma_x t}\left\{\frac{\chi_t(\gamma_x)\mu_{\gamma_x t} + \zeta_t(\gamma_x)\sigma_{\gamma_x t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})}\right\}\\
&\qquad - (1-\gamma_x)\pi_{(1-\gamma_x)t}\left\{\frac{\chi_t(1-\gamma_x)\mu_{(1-\gamma_x)t} + \zeta_t(1-\gamma_x)\sigma_{(1-\gamma_x)t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})}\right\},\\
&(iv)\quad E\{v_t^2 w_t^2 U_t|\mathbf{z}_t;\boldsymbol{\theta}\} = \gamma_x^2\pi_{\gamma_x t}\left\{\frac{\chi_t(\gamma_x)\mu_{\gamma_x t}^2 + \zeta_t(\gamma_x)\mu_{\gamma_x t}\sigma_{\gamma_x t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x},\boldsymbol{\nu})} + \sigma_{\gamma_x t}^2\right\}\\
&\qquad + (1-\gamma_x)^2\pi_{(1-\gamma_x)t}\left\{\frac{\chi_t(1-\gamma_x)\mu_{(1-\gamma_x)t}^2 + \zeta_t(1-\gamma_x)\mu_{(1-\gamma_x)t}\sigma_{(1-\gamma_x)t}}{f_{SMSN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{1-\gamma_x},-\boldsymbol{\lambda}_{1-\gamma_x},\boldsymbol{\nu})} + \sigma_{(1-\gamma_x)t}^2\right\},
\end{aligned}$$

where $t = 1, \ldots, n$, $s_t = sign(w_t) \in \{-1, 1\}$ *and*

$$\chi_t(\gamma_x) = \begin{cases} f_{SN}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\boldsymbol{\lambda}_{\gamma_x}) & \text{if } U_t \overset{d}{=} 1,\\ 2t_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\nu)\frac{\nu+m}{\nu+d_{\gamma_x}}T\left(\sqrt{\frac{\nu+m+2}{\nu+d_{\gamma_x}}}A_{\gamma_x};\nu+m+2\right), & \text{if } U_t \sim \Gamma(\nu/2,\nu/2),\\ 2f_{SL}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x},\nu)\frac{2\nu+m}{d_{\gamma_x}}\frac{P_1(\frac{m+2\nu+2}{2},\frac{d_{\gamma_x}}{2})}{P_1(\frac{m+2\nu}{2},\frac{d_{\gamma_x}}{2})}E\left(\Phi(S_{\gamma_x}^{1/2}A_{\gamma_x})\right), & \text{if } U_t \sim Beta(\nu,1), \end{cases}$$

Also, for

$$h(u_t;\nu_1,\nu_2) = \begin{cases} \nu_2 & \nu_1\\ 1 & 1-\nu_1, \end{cases}$$

we have

$$\chi_t(\gamma_x) = 2\left\{\nu_1\nu_2\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Omega}_{\gamma_x})\Phi(\nu_2^{1/2}A_{\gamma_x}) + (1-\nu_1)\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x})\Phi(A_{\gamma_x})\right\},$$

$$\zeta_t(\gamma_x)=\begin{cases}\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Omega}_{\gamma_x})\phi(A_{\gamma_x}), & \text{if } U_t\overset{d}{=}1,\\ \frac{\sigma_{\gamma_x t}}{\sigma_x}\frac{\Gamma(\frac{m+\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu+2}{2})}\{\nu+d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})\}^{-\frac{1}{2}}t_m(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu), & \text{if } U_t\sim\Gamma(\nu/2,\nu/2),\\ \frac{\sigma_{\gamma_x t}}{\sigma_x}\frac{\Gamma(\frac{m+2\nu+1}{2})}{\sqrt{2\pi}\Gamma(\frac{2\nu+m}{2})}\{\frac{2}{d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})}\}^{\frac{1}{2}} & \\ \times\frac{P_1(\frac{2\nu+m+1}{2},\frac{d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})}{2})}{P_1(\frac{2\nu+m}{2},\frac{d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})}{2})}f_{SL}(\mathbf{z}_t;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) & \text{if } U_t\sim Beta(\nu,1).\end{cases}$$

Also, for

$$h(u_t;\nu_1,\nu_2)=\begin{cases}\nu_2 & \nu_1\\ 1 & 1-\nu_1\end{cases}$$

we have

$$\zeta_t(\gamma_x)=\nu_1\sqrt{\nu_2}\phi_m(\mathbf{z}_t;\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Omega}_{\gamma_x})\phi(\sqrt{\nu_2}A_{\gamma_x}),$$

where

$$\sigma^2_{\gamma_x t}=\frac{\sigma_x^2}{1+\sigma_x^2\gamma_x^2\mathbf{b}^\top\boldsymbol{\Sigma}^{-1}\mathbf{b}},\qquad d_{\mathbf{z}_t}(\boldsymbol{\mu},\boldsymbol{\Sigma})=(\mathbf{z}_t-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{z}_t-\boldsymbol{\mu}).$$

Furthermore, $\pi_{\gamma_x t}$, $\mu_{\gamma_x t}$ are given by Proposition 3, and the distribution of the variable S_s is given above. Also, note that we should use $-\boldsymbol{\lambda}_{1-\gamma_x}$ wherever it is needed.

The *M*-step consists of maximization of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$. For this, we use the faster extension of the original EM, the ECME algorithm, by replacing the *M*-step with a sequence of conditional maximization (CM) steps. CM-steps then conditionally maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as follows:

CM-step Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ with respect to $\boldsymbol{\theta}$. This is accomplished by the following steps.
CM-step.1: Update $\widehat{\gamma}_x^{(k)},\widehat{\sigma}_x^{2(k)},\widehat{\mu}_x^{(k)}$ by

$$\begin{aligned}\widehat{\gamma}_x^{(k)}&=\frac{1+\widehat{\bar{s}}^{(k)}}{2},\\ \widehat{\sigma}_x^{2(k+1)}&=\frac{1}{n}\sum_{t=1}^n\widehat{uv^2}_t^{(k)},\\ \widehat{\mu}_x^{(k+1)}&=\frac{\sum_{t=1}^n\widehat{u}_t\bar{X}_t-\sum_{t=1}^n\widehat{uwv}_t}{\sum_{t=1}^n\widehat{u}_t},\\ \widehat{\mu}_y^{(k+1)}&=\frac{\sum_{t=1}^n\widehat{u}_t\bar{Y}_t-\sum_{t=1}^n\widehat{uwv}_t\widehat{\beta}^{(k)}}{\sum_{t=1}^n\widehat{u}_t},\end{aligned}$$

where $\mu_y=\alpha+\beta\mu_x$, $\widehat{\bar{s}}^{(k)}=\sum_{t=1}^n\widehat{s}_t^{(k)}/n$, $\bar{X}_t=\sum_{i=1}^p X_t^{(i)}/p$, $\bar{Y}_t=\sum_{i=1}^q Y_t^{(j)}/q$,

CM-step.2: With fixed $\widehat{\mu}_x=\widehat{\mu}_x^{(k+1)}$, $\widehat{\mu}_y=\widehat{\mu}_y^{(k+1)}$

1. Update $\widehat{\beta}^{(k)}$ by

$$\widehat{\beta}^{(k+1)}=\frac{\sum_{t=1}^n\widehat{uwv}_t^{(k)}(\bar{Y}_t-\widehat{\mu}_y^{(k+1)})}{\sum_{t=1}^n\widehat{uw^2v^2}_t^{(k)}}.$$

2. Update $\widehat{\phi}_{\delta}^{(k)}$ by

$$\widehat{\phi}_{\delta}^{(k+1)} = \frac{1}{n}\left\{\sum_{t=1}^{n}\frac{1}{p}\widehat{u}_t^{(k)}\sum_{i=1}^{p}\left(X_t^{(i)}-\widehat{\mu}_x^{(k+1)}\right)^2 - 2\sum_{t=1}^{n}\widehat{uwv}_t^{(k)}\left(\bar{X}_t-\widehat{\mu}_x^{(k+1)}\right)\right\} + \frac{1}{n}\sum_{t=1}^{n}\widehat{uw^2v^2}_t^{(k)}$$

3. Update $\widehat{\phi}_{\xi}^{(k)}$ by

$$\widehat{\phi}_{\xi}^{(k+1)} = \frac{1}{n(q-1)}\left\{\sum_{t=1}^{n}\widehat{u}_t^{(k)}\sum_{j=1}^{q}\left(Y_t^{(j)}-\widehat{\mu}_y^{(k+1)}\right)^2 - q\sum_{t=1}^{n}\widehat{u}_t^{(k)}\left(\bar{Y}_t-\widehat{\mu}_y^{(k+1)}\right)^2\right\}.$$

4. Update $\widehat{\phi}_{e}^{(k)}$ by

$$\widehat{\phi}_{e}^{(k+1)} = \frac{1}{n(q-1)}\left\{q\sum_{t=1}^{n}\widehat{u}_t^{(k)}\left(\bar{Y}_t-\widehat{\mu}_y^{(k+1)}\right)^2 - \frac{1}{q}\sum_{t=1}^{n}\widehat{u}_t^{(k)}\sum_{j=1}^{q}\left(Y_t^{(j)}-\widehat{\mu}_y^{(k+1)}\right)^2\right\} + \frac{1}{n}\sum_{t=1}^{n}\widehat{uw^2v^2}_t^{(k)}\widehat{\beta}^{2(k+1)} - \frac{2}{n}\sum_{t=1}^{n}\widehat{uwv}_t^{(k)}\widehat{\beta}^{(k+1)}\left(\bar{Y}_t-\widehat{\mu}_y^{(k+1)}\right).$$

CM-step.3: Update $\widehat{\alpha}^{(k)}$ by

$$\widehat{\alpha}^{(k+1)} = \widehat{\mu}_y^{(k+1)} - \widehat{\beta}^{(k+1)}\widehat{\mu}_x^{(k+1)}.$$

Remark 1. *The iteration of the ECM algorithm is repeated until the difference between two successive log-likelihood values,* $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)};\mathbf{z}_1,\ldots,\mathbf{z}_n)-\ell(\widehat{\boldsymbol{\theta}}^{(k)};\mathbf{z}_1,\ldots,\mathbf{z}_n)|$*, is sufficiently small, say* 10^{-6}*. As initial values for the algorithm, we can use the mean vector* $\bar{\mathbf{z}}$ *for* $\boldsymbol{\mu}$*, the sample variance* $S_{\mathbf{X}}^2 = \frac{1}{n}\sum_{t=1}^{n} S_{\mathbf{X}_t}^2$ *with* $S_{\mathbf{X}_t}^2 = \frac{1}{p}\sum_{i=1}^{p}\left(X_t^{(i)}-\bar{X}_t\right)^2$ *(according to the model given by (4)) for* $\bar{u}(\sigma_x^2+\phi_\delta)$ *with* $\bar{u} = \frac{1}{n}\sum_{t=1}^{n} u_t$*. As for* ϕ_δ*, it can be set equal to some fraction of* σ_x^2*; in our subsequent numerical work, we have started the iterative process with* $\sigma_x^2 = \phi_\delta$*. Similarly, the sample variance* $S_{\mathbf{Y}}^2 = \frac{1}{n}\sum_{t=1}^{n} S_{\mathbf{Y}_t}^2$ *with* $S_{\mathbf{Y}_t}^2 = \frac{1}{q}\sum_{j=1}^{p}\left(Y_t^{(j)}-\bar{Y}_t\right)^2$ *for* $\bar{u}\phi_\xi$*. Regarding* $\boldsymbol{\gamma}$*, an option would be to start from the sample skewness of the observed explanatory variables* $\mathbf{X} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)^\top$ *with* $\mathbf{X}_t = (X_t^{(1)},\ldots,X_t^{(p)})^\top$*, namely* $g_1(\mathbf{X}) = \frac{1}{n}\sum_{t=1}^{n} g_1(\mathbf{X}_t)$*, where* $g_1(\mathbf{X}_t) = \frac{m_{3\mathbf{X}_t}}{m_{2\mathbf{X}_t}}$*, with* $m_{k\mathbf{X}_t} = \frac{1}{p}\sum_{i=1}^{p}\left(X_t^{(i)}-\bar{X}_t\right)^k$*, and invert these values to obtain an initial choice of* γ_x*, In part (ii) Corollary 1 the Pearson measure of skewness was calculated for the TP-SMN family. .*

4.1 Asymptotic covariance matrix of MLEs

According to Arellano-Valle et al. (2020) approach, the empirical information matrix can be approximated as

$$\mathbf{I}_{emp}(\boldsymbol{\theta};\mathbf{z}) = \sum_{t=1}^{n} U(\mathbf{z}_t;\boldsymbol{\theta})U^\top(\mathbf{z}_t;\boldsymbol{\theta}) - n\bar{U}(\mathbf{z};\boldsymbol{\theta})\bar{U}^\top(\mathbf{z};\boldsymbol{\theta}),$$

where $\bar{U}(\mathbf{z};\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} U(\mathbf{z}_t;\boldsymbol{\theta})$ and $U(\mathbf{z}_t;\boldsymbol{\theta})$ is the empirical scoring function for the subject t. The individual score function can be determined as (Louis (1982))

$$U(\mathbf{z}_t;\boldsymbol{\theta}) = \frac{\partial \log f(\mathbf{z}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left\{\frac{\partial \ell(\boldsymbol{\theta};\mathbf{z}_{ct})}{\partial \boldsymbol{\theta}}|\mathbf{z}_t;\theta\right\},$$

where $\mathbf{z}_{ct} = \{\mathbf{z}_t, w_t, v_t, u_t\}$, is the complete data vector from the tth observation and $\ell(\boldsymbol{\theta}; \mathbf{z}_{ct})$ is the corresponding contribution to the log-likelihood function.

$$\mathbf{I}_{emp}(\widehat{\boldsymbol{\theta}}; \mathbf{z}) = \sum_{t=1}^{n} \widehat{U}_t \widehat{U}_t^{\top},$$

where $\widehat{U}_t = U(\mathbf{z}_t; \widehat{\boldsymbol{\theta}}) = (\widehat{U}_{t,\alpha}, \widehat{U}_{t,\beta}, \widehat{U}_{t,\mu_x}, \widehat{U}_{t,\sigma_x}, \widehat{U}_{t,\gamma_x}, \widehat{U}_{t,\phi_\delta}, \widehat{U}_{t,\phi_\xi}, \widehat{U}_{t,\phi_e})^{\top}$, with

$$\begin{aligned}
\widehat{U}_{t,\alpha} &= \frac{q}{\widehat{\phi}_\xi + q\widehat{\phi}_e} \left\{ \widehat{u}_t(\bar{Y}_t - \widehat{\mu}_y) - \widehat{\beta}\widehat{uwv}_t \right\}, \\
\widehat{U}_{t,\beta} &= \frac{q}{\widehat{\phi}_\xi + q\widehat{\phi}_e} \left\{ (\bar{Y}_t - \widehat{\alpha})(\widehat{u}_t\widehat{\mu}_x + \widehat{uwv}_t) - \widehat{\beta}(\widehat{u}_t\widehat{\mu}_x^2 + 2\widehat{uwv}_t\widehat{\mu}_x + \widehat{uw^2v^2}_t) \right\}, \\
\widehat{U}_{t,\mu_x} &= \frac{p}{\widehat{\phi}_\delta} \{\widehat{u}_t(\bar{X}_t - \widehat{\mu}_x) - \widehat{uwv}_t\} + \frac{q\widehat{\beta}}{\widehat{\phi}_\xi + q\widehat{\phi}_e} \left\{ \widehat{u}_t(\bar{Y}_t - \widehat{\mu}_y) - \widehat{\beta}\widehat{uwv}_t \right\}, \\
\widehat{U}_{t,\sigma_x} &= -\frac{1}{\widehat{\sigma}_x} + \frac{1}{\widehat{\sigma}_x^3}\widehat{uv^2}_t, \\
\widehat{U}_{t,\gamma_x} &= \frac{1}{\widehat{\gamma}_x}\left(\frac{1+\widehat{s}_t}{2}\right) - \frac{1}{1-\widehat{\gamma}_x}\left(\frac{1-\widehat{s}_t}{2}\right), \\
\widehat{U}_{t,\phi_\delta} &= \frac{p}{2\widehat{\phi}_\delta^2} \left\{ \frac{\widehat{u}_t}{p} \sum_{i=1}^{p} (X_t^{(i)} - \widehat{\mu}_x)^2 - 2(\bar{X}_t - \widehat{\mu}_x)\widehat{uwv}_t + \widehat{uw^2v^2}_t - \widehat{\phi}_\delta \right\}, \\
\widehat{U}_{t,\phi_\xi} &= \frac{q}{2(\widehat{\phi}_\xi + q\widehat{\phi}_e)^2} \left\{ -q\frac{\widehat{\phi}_e(2\widehat{\phi}_\xi + q\widehat{\phi}_e)}{\widehat{\phi}_\xi^2}(\bar{Y}_t - \widehat{\mu}_y)^2\widehat{u}_t - \widehat{\beta}(\bar{Y}_t - \widehat{\mu}_y)\widehat{uwv}_t + \widehat{\beta}^2\widehat{uw^2v^2}_t \right\} \\
&\quad - \frac{q}{2}\left(\frac{\widehat{\phi}_\xi + (q-1)\widehat{\phi}_e}{\widehat{\phi}_\xi(\widehat{\phi}_\xi + q\widehat{\phi}_e)} \right) + \frac{\widehat{u}_t}{2\widehat{\phi}_\xi^2} \sum_{j=1}^{q} (Y_t^{(j)} - \widehat{\mu}_y)^2, \\
\widehat{U}_{t,\phi_e} &= \frac{q^2}{2(\widehat{\phi}_\xi + q\widehat{\phi}_e)^2} \left\{ (\bar{Y}_t - \widehat{\mu}_y)^2\widehat{u}_t - 2\widehat{\beta}(\bar{Y}_t - \widehat{\mu}_y)\widehat{uwv}_t + \widehat{\beta}^2\widehat{uw^2v^2}_t - \frac{1}{q}(\widehat{\phi}_\xi + q\widehat{\phi}_e) \right\}.
\end{aligned}$$

5 Simulation study

The simulation study is considered in this section. It is well known that misspecification of the model's distribution will lead to biases of parameter estimates. The main goal is to confirm the effectiveness and precision of MLEs under TP-SMN distributions. We consider four skew distributions, including TP-N, TP-t with $\nu = 4$, TP-SL with $\nu = 2$, and TP-CN with $\nu_1 = 0.2, \nu_2 = 0.3$, in RMEM with or without equation error, respectively. Under the circumstances of no equation error, we denote the distributions TP-N, TP-t, TP-SL and TP-CN by TP-N_0, TP-t_0, TP-SL_0 and TP-CN_0 respectively. To maintain the general and consistent form of the estimators among different TP-SMN distributions, the degrees of freedom for the TP-SMN model will not be estimated together with the parameters of interest. Hence, we selected some heavy-tailed distributions with fixed degrees of freedom. In the application, to find the best distribution for the data, the degrees of freedom for different distributions, chosen by the Schwarz (1978) information criterion. As usual, we are interested in the regression parameters α and β in the simulations. We first generate data from the model (4) with the number of replicated observations chosen as $p = 4$ and $q = 3$ under the TP-t distribution with $\nu = 4$. Then, we use the EM algorithm to calculate the MLEs of $\boldsymbol{\theta}$ under the TP-N, TP-t, TP-SL and TP-CN distributions with or without equation error, respectively. We used two indicators to evaluate estimates, including sample bias (BIAS) and standard deviation (SD).

Other parameters of model (4) are set as: $\mu_x = 1.5, \sigma_x = 1, \gamma_x = 0.95, \alpha = 2, \beta = 1, \phi_\delta = 1, \phi_\xi = 0.5$. For comparison, we set $\phi_e = 0.5, 1$, $\phi_e = 1.5$ respectively, which shows that the degree of matching between the true covariate and response tends from strong to weak. Based on 1000 simulations, the results are reported in Table 1. As we expected, the BIAS, and the corresponding SD decrease when the sample size increases from 50 to 100 in all scenarios, and the MLE under TP-t ($\nu = 4$) distribution (the true distribution) is the best estimator using any of the two indicators, which fully reflects the effectiveness and accuracy of the ML estimates. Moreover, performance under the TP-CN and TP-t distributions behaves better than those under the TP-N distribution, which may be attributed to their heavy-tailed features. Thus, we encourage the use of the heavy-tailed model when the data show heavy-tailed features, even if the proposed heavy-tailed distribution is not the true one.

Obviously, the estimates with equation error consistently perform better than the estimates without equation error. Especially, when ϕ_e increases from 0.5 to 1.5, the SD ratios without equation error and with equation error become clearly larger. It states that for the data with skewness or heavy-tailed features, ignoring equation error will bring about a serious deviation for statistical inference. The equation error plays an important role in expressing the uncertain relationship between the true covariates and the response.

6 Application

We give an illustrative example using the CSFII data set. This data set has conducted 24-hour recall measures, as well as three additional 24-hour recall phone interviews of 1827 women who were recorded about their daily diet intake (for example, saturated fat, calories, vitamin, etc.). Carroll et al. (2006) have indicated that saturated fat has a great relationship with the risk of breast cancer and other diseases, but the statistical significance of saturated fat disappeared when adding caloric intake to the logistic regression model. Therefore, it is necessary to reveal the inner relationship between saturated fat and caloric intake. In this illustrative example, we take the calorie intake/5000 as x and the saturated fat intake/100 as y (Carroll et al. (2006); Lin and Cao (2013)). Instead of x and y, the nutrition variables X and Y are calculated using four 24-hour recalls and suppose that they follow the model (4) with $p = q = 4$. For the purpose of verifying the existence of skewness and heavy tails in the latent covariate x, Lin and Cao (2013) fit the data using normal RMEM without equation error. They showed that the latent covariate is positively skewed and heavy-tailed. This indicates that a normal model may not offer a good fit. The degrees of freedom for different distributions, chosen by the Schwarz (1978) information criterion, are also reported in Table 2. Using the profile log-likelihood functions for the three models, getting the highest values of the profile log-likelihood, the degrees of freedom are found as $\nu = 4.3$ for TP-t, $\nu = 1.3$ for TP-SL, and $\nu_1 = 0.29, \nu_2 = 0.22$ for TP-CN. We now consider TP-SMN distributions for x_t and SMN distributions for measurement errors δ_t, ξ_t and equation error e_t in the RMEM, that is, the TP-SMN-RMEM proposed in this paper. We calculate the MLEs of the parameter $\boldsymbol{\theta}$ and their standard errors (SE) and also the Akaike information criterion (AIC) (Akaike (1974)) values based on the RMEM model in the above distributions. These estimates are displayed in Table 2. The AIC value under the TP-t distribution is always the smallest, whether there is an equation error or skewness or not, and it gets the smallest value under the TP-t_0 distribution in all situations. Thus, TP-t-RMEM is more suitable for these data. Note that as another potential competitor, AIC values based on the SMSN-RMEM model have added in Table 2 (in parenthesis). Moreover, the heavy-tailed models show smaller SEs than the normal model. It would be better consider skewness in the models for their smaller AIC values. The practical significance of the parameter β can be described as the positive proportion of saturated fat in calories. The skew parameter γ_x is positive, suggesting that these data are skewed to the right. The values of ϕ_e seem very small and only slightly change between different models, indicating that there is an inapparent random relationship between the intake of calories and saturated fats.

Table 1: Performances of estimators under TPt-RMEM with or without equation error

Parameter	n	Estimator	$\phi_e = 0.5$		$\phi_e = 1$		$\phi_e = 1.5$	
			BIAS	SD	BIAS	SD	BIAS	SD
α	50	$TP-N$	-0.123	0.641	-0.108	0.781	-0.145	0.797
		$TP-t$	-0.052	0.390	-0.033	0.492	-0.054	0.501
		$TP-SL$	-0.072	0.483	-0.051	0.587	-0.074	0.598
		$TP-CN$	-0.069	0.472	-0.047	0.568	-0.073	0.595
		$TP-N_0$	-0.946	1.240	-1.997	2.011	-2.924	2.317
		$TP-t_0$	-0.796	0.542	-1.442	1.042	-1.857	1.545
		$TP-SL_0$	-0.803	0.649	-1.508	1.098	-2.152	1.618
		$TP-CN_0$	-0.801	0.631	-1.492	1.087	-2.138	1.609
	100	$TP-N$	-0.056	0.405	-0.061	0.512	-0.067	0.769
		$TP-t$	-0.020	0.254	-0.021	0.316	-0.028	0.531
		$TP-SL$	-0.031	0.297	-0.032	0.312	-0.043	0.552
		$TP-CN$	-0.029	0.289	-0.031	0.308	-0.042	0.548
		$TP-N_0$	-0.880	0.655	-1.884	1.231	-2.625	1.996
		$TP-t_0$	-0.795	0.382	-1.826	0.692	-1.136	1.379
		$TP-SL_0$	-0.827	0.448	-1.835	0.824	-2.021	1.408
		$TP-CN_0$	-0.820	0.439	-1.831	0.808	-2.018	1.402
β	50	$TP-N$	0.047	0.303	0.041	0.356	0.058	0.397
		$TP-t$	0.026	0.196	0.019	0.221	0.037	0.267
		$TP-SL$	0.034	0.207	0.029	0.268	0.041	0.289
		$TP-CN$	0.032	0.201	0.026	0.259	0.039	0.281
		$TP-N_0$	0.488	0.607	0.996	0.927	1.017	1.128
		$TP-t_0$	0.395	0.211	0.884	0.513	0.998	0.885
		$TP-SL_0$	0.407	0.298	0.941	0.584	1.011	0.916
		$TP-CN_0$	0.405	0.281	0.937	0.571	1.009	0.903
	100	$TP-N$	0.021	0.201	0.026	0.217	0.037	0.228
		$TP-t$	0.010	0.131	0.009	0.152	0.011	0.173
		$TP-SL$	0.014	0.142	0.013	0.167	0.018	0.189
		$TP-CN$	0.013	0.136	0.012	0.161	0.017	0.181
		$TP-N_0$	0.428	0.317	0.976	0.627	1.010	0.638
		$TP-t_0$	0.315	0.182	0.824	0.313	0.971	0.585
		$TP-SL_0$	0.385	0.198	0.901	0.384	0.987	0.416
		$TP-CN_0$	0.379	0.192	0.899	0.380	0.979	0.4160

Table 2: Parameter estimations for *CSFII* data

Models	Degree	AIC	Parameter							
			μ_x	α	β	γ_x	σ_x	ϕ_δ	ϕ_ξ	ϕ_e
$TP-N$	/	-18965(-18842)	0.196	-0.060	0.957	0.793	0.017	0.011	0.014	0.0012
			(0.007)	(0.022)	(0.067)	(0.090)	(0.002)	(0.0002)	(0.0003)	(0.0006)
$TP-t$	$\nu=4$	-21957(-21182)	0.200	-0.052	0.914	0.651	0.011	0.007	0.008	0.0003
			(0.007)	(0.014)	(0.052)	(0.085)	(0.001)	(0.0002)	(0.0002)	(0.0002)
$TP-SL$	$\nu=1.2$	-21837.5(-21066)	0.201	-0.051	0.911	0.701	0.006	0.004	0.005	0.0002
			(0.009)	(0.016)	(0.049)	(0.063)	(0.001)	(0.001)	(0.0001)	(0.0001)
$TP-CN$	$\nu_1=0.4,\nu_2=0.2$	-21591.5(-20954)	0.202	-0.049	0.896	0.672	0.007	0.005	0.006	0.0003
			(0.008)	(0.014)	(0.043)	(0.051)	(0.001)	(0.0001)	(0.0001)	(0.0001)
$TP-N_0$	/	-19010.5(-18996)	0.199	-0.067	0.974	0.768	0.018	0.011	0.014	/
			(0.007)	(0.024)	(0.069)	(0.082)	(0.002)	(0.0003)	(0.0003)	
$TP-t_0$	$\nu=4$	-22041(-21887.5)	0.200	-0.059	0.934	0.642	0.012	0.008	0.009	/
			(0.008)	(0.013)	(0.046)	(0.078)	(0.001)	(0.0002)	(0.0002)	
$TP-SL_0$	$\nu=1.2$	-21921.5(-21153)	0.201	-0.055	0.921	0.691	0.005	0.004	0.004	/
			(0.009)	(0.017)	(0.036)	(0.057)	(0.001)	(0.0001)	(0.0002)	
$TP-CN_0$	$\nu_1=0.4,\nu_2=0.2$	-21880.5(-21497)	0.202	-0.051	0.901	0.613	0.008	0.006	0.006	/
			(0.008)	(0.012)	(0.036)	(0.037)	(0.0007)	(0.0002)	(0.0002)	

7 Conclusion

In this paper, we develop an RMEM in the TP-SMN distribution class, called TP-SMN-RMEM, which is suitable for asymmetric, heavy-tailed, and multimodal data. Also, it includes many special cases, such as nonreplicated MEM under SMN distributions (Lachos et al. (2011)), RMEM under normal distribution (Lin et al. (2004)) or SMN distributions (Lin and Cao (2013)). Moreover, in contrast to similar models, especially, considering SMSN-RMEM (proposed by Cao et al. (2018)), what our proposed model offers exactly is that the SMSN and TP-SMN families have different properties (for instance, different tail behavior). In addition, numerical stability might also be a property that could be used to motivate our proposed model for the explanatory variable. This means that, estimating the skewness parameter in the TP-SMN class of distributions is often easier, and more stable, than estimating the shape parameter in the SMSN family (which controls the asymmetry, tails, location of the mode, and spread of the density). We provide explicit expressions for both the EM-type iterative estimates and the corresponding standard errors. Due to the hierarchical structure of the model, the features of skewness of both covariate and response can be captured under these assumptions. The proposed TP-SMN-RMEM models can reduce the negative impact of distribution misspecification and outliers to some extent. In biological, environmental, chemical, medical and other research areas, measurement error, skewness, multimodality, and outliers commonly exist in real data; thus, the robust TP-SMN-RMEM models would be a good choice to fit such data.

Acknowledgement

The author would like to thank Dr. Marcos Oliveira Prats from the Federal University of Minas Gerais for his help with the R Code.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.

Arellano-Valle, R. B., Azzalini A., Ferreira C.S. and Santoro, K. (2020). A two-piece normal measurement error model. *Computational Statistics & Data Analysis*, **144**, 106863.

Arellano-Valle, R. B., Gomez, H. and Quintana, F.A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, **128**, 427–443.

Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96**, 93–116.

Arellano-Valle, R. B., Ozan, S., Bolfarine, H. and Lachos, V. H. (2005). Skew normal measurement error models. *Journal of Multivariate Analysis*, **96**, 265–281.

Barkhordar, Z., Maleki, M., Khodadadi, Z., Wraith, D. and Negahdari, F. (2020). A Bayesian approach on the two-piece scale mixtures of normal homoscedastic nonlinear regression models. *Journal of Applied Statistics*, **9**, 1305–1322.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113.

Buonaccorsi. J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Chapman and Hall, Boca Raton.

Cao, C. Z., Lin, J. G., Shi, J. Q., Wang, W. and Zhang, X. Y. (2015). Multivariate measurement error models for replicated data under heavy-tailed distributions. *Journal of Chemometrics*, **29**, 457–466.

Cao, C. Z., Wang, W., Shi, J. Q. and Lin, J. G. (2018). Measurement error models for replicated data under asymmetric heavy-tailed distributions. *Computational Economics*, **52**, 531–553.

Carroll, R. J. Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models. A modern perspective (2nd edn)*. Chapman and Hall, Boca Raton.

Chan, L. K. and Mak, T. K. (1979). Maximum likelihood estimation of a linear structural relationship with replication. *Journal of the Royal Statistical Society, Series B*, **41**, 263–268.

Cheng, C. L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. Oxford University Press, London.

Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.

Heteroscedastic nonlinear regression models using asymmetric and heavy tailed two-piece distributions. *AStA Advances in Statistical Analysis*, **105**, 451–467.

Isogava, Y. (1985). Estimating a multivariate linear structural relationship with replication. *Journal of the Royal Statistical Society, Series B*, **47**, 211–215.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology*. Chapman and Hall, Boca Raton.

Ghasami, S., Maleki, M., and Khodadadi, Z. (2020). Leptokurtic and platykurtic class of robust symmetrical and asymmetrical time series models. *Journal of Computational and Applied Mathematics*, **376**, 112806.

Lachos, V. H., Angolini, T., and Abanto-Valle, C. A. (2011). On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers*, **52**, 567–590.

Lin, N., Bailey, B. A. and He, X. M. (2004). Adjustment of measuring devices with linear models. *Technometrics*, **46**, 127–134.

Lin, J. G. and Cao, C. Z. (2013). On estimation of measurement error models with replication under heavy-tailed distributions. *Computational Statistics*, **28**, 802–829.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226–233.

Maleki, M., Barkhordar, Z., Khodadadi, Z., and Wraith, D. (2019c). A robust class of homoscedastic nonlinear regression models. *Journal of Statistical Computation and Simulation*, **89**, 2765–2781

Maleki, M., Contreras-Reyes, J. E., and Mahmoudi, M. R. (2019d). Robust mixture modeling based on two-piece scale mixtures of normal family. *Axioms*, **8**, 38.

Maleki, M., and Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics-Theory and Methods*, **46**, 12356–12369.

Meng, X. L., and Rubin, D. B. (1993). Maximum likelihood estimation via the EM algorithm: a general framework. *Biometrika*, **80**, 267–278.

Moravveji, B., Khodadadi, Z., and Maleki, M. (2019). A Bayesian analysis of two-piece distributions based on the scale mixtures of normal family. *Iranian Journal of Science and Technology, Transactions A: Science*, **43**, 991–1001.

Reiersol, O. (1950). Identifiability of a linear relation between variables which are subject to errors. *Econometrica*, **18**, 375–389.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Zarei, A., Khodadadi, Z., Maleki, M. and Zare, K. (2022). Robust mixture regression modeling based on two-piece scale mixtures of normal distributions. *Advances in Data Analysis and Classification*, **17**, 181–210.

End of SMPS Journal – Volume 1, Issue 2

For manuscript submission and future issues, visit:
www.smps.um.ac.ir