ISSN: 3060-7876

SMPS

Stochastic Models in Probability and Statistics

Vol. 2, No. 1, 2025

In the Name of God

Stochastic Models in Probability and Statistics (SMPS)

Volume 2, Issue 1 July 2025

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

Editorial Board

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. 2, No. 1, 2025, pp. 1–29. Research Article

SMPS

On some bivariate integer-valued distributions on Z^2

Nada Nadim Alkhatib and Emad-Eldin Aly Ahmed Aly*

Department of Statistics and Operations Research, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait

Email(s): engalkhatib.92@hotmail.com and eealy50@gmail.com

Abstract. We proposed and studied a new bivariate random sign transformation of nonnegative bivariate integer-valued distributions. This transformation develops new bivariate integer-valued distributions on Z^2. We applied the new transformation to the bivariate Poisson and the bivariate geometric distributions. As an illustration, we fitted a real-life data set developed based on the results of the 2019 UEFA Europa League using the new distributions.

Keywords: Bivariate RST; ML estimators; MM estimators; Monte Carlo simulations.

1 Introduction

The development of nonnegative integer-valued bivariate distributions has received considerable attention in the literature. For important results and reviews on this topic, we refer the reader to Kocherlakota and Kocherlakota (1992), Johnson et al. (1997), Lai (2006) and Sarabia Alegría and Gómez Déniz (2008). Some recent important results in this area include Odhah (2013), Genest and Mesfioui (2014), Bulla et al. (2015), Omair et al. (2016) and Karlis and Mamode Khan (2023).

Chesneau et al. (2018) noted that changes in intra-daily stock prices take both positive and negative integer values and that the price change is therefore characterized by discrete positive and negative jumps. This motivated Chesneau et al. (2018) to propose and study some bivariate integer-valued distributions on Z^2. Omair et al. (2022) proposed some bivariate integer-valued distributions on Z^2 and applied their models to fit the two real life data sets; the difference in the number of casualties to the number of employees on duty on railroads and the difference in the number of goals scored in the English Premier League in different years.

In this paper, we proposed and studied a new bivariate random sign transformation (BRST) of nonnegative bivariate integer valued distributions. The BRST is an extension of the random

*Corresponding author

Received: 09 December 2024 / Accepted: 10 February 2025
DOI: 10.22067/smps.2025.91162.1039

sign transformation (RST) of Aly (2018). The BRST is also a generalization of the transformation used in Chesneau et al. (2018). We used the BRST to introduce and study new families of bivariate integer-valued distributions on Z^2. The first family is developed based on the bivariate Poisson distribution (BPD). The second family is developed based on the bivariate geometric distribution (BGD).

In Section 2, we review some important bivariate integer-valued distributions. In Sections 3, we introduce and study three versions of the BRST. In Section 4, we apply the transformations of Sections 3 to the BPD. In Section 5, we apply the transformations of Sections 3 to the BGD. In Section 6, we report the results of Monte Carlo simulation studies conducted to evaluate the estimators of the parameters of the models of Sections 4 and 5. In Section 7, we apply the models of Sections 4 and 5 to a real life data set developed based on the results of the 2019 UEFA Europa League.

A random vector (or variable) will be denoted by RV. The probability mass function of discrete RV will be abbreviated by *pmf* and the joint probability mass function of a discrete RV will be abbreviated by *jpmf*. The univariate Bernoulli distribution with parameter θ will be denoted by $Ber(\theta)$. The geometric distribution with $pmf, g(x) = (1-\theta)^x\theta, x = 0,1,\ldots, 0 < \theta < 1$, will be denoted by $Geo(\theta)$. The Poisson distribution with parameter $\lambda > 0$ will be denoted by $Poi(\lambda)$.

2 Some bivariate integer-valued distributions

2.1 Some bivariate Bernoulli distributions

Definition 1. *Assume that* $\underline{\beta} = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$, *where* $0 \le \beta_{ij} \le 1$ *and* $\sum\sum_{i,j=0,1} \beta_{ij} = 1$. *The RV* (U_1, U_2) *with jpmf*,

$$P(U_1 = i, U_2 = j) = \beta_{ij}, \qquad 0 \le i, j \le 1 \tag{1}$$

is said to have the Bivariate Bernoulli (BVBer) distribution denoted by $BVBer(\underline{\beta})$.

Lemma 1. *Assume that* $U_1 \sim Ber(\beta_{11} + \beta_{10}), V_2 \sim Ber(\frac{\beta_{11}}{\beta_{11}+\beta_{10}})$ *and* $V_3 \sim Ber(\frac{\beta_{01}}{1-\beta_{11}-\beta_{10}})$ *are independent. Let*

$$U_2 = U_1V_2 + (1-U_1)V_3, \tag{2}$$

then, (U_1, U_2) *has the* $BVBer(\underline{\beta})$ *distribution of (1).*

Definition 2. *The RV* (U_1, U_2) *with jpmf*,

$$g(u_1,u_2) = \pi^{u_1}\overline{\pi}^{1-u_1}\alpha^{(1-u_1)(1-u_2)+u_1u_2}\overline{\alpha}^{u_1(1-u_2)+u_2(1-u_1)}, \qquad u_1, u_2 = 0,1 \tag{3}$$

is said to have the two parameters BVBer distribution denoted by $BVBer(\pi, \alpha)$.

Note that the $BVBer(\pi,\alpha)$ of (3) is the special case of (1) when $\beta_{11} = \alpha\pi, \beta_{01} = \overline{\alpha}\overline{\pi}, \beta_{10} = \overline{\alpha}\pi, \beta_{00} = \alpha\overline{\pi}$. To generate (U_1, U_2) from the $BVBer(\pi,\alpha)$ of (3), we independently generate $U_1 \sim Ber(\pi), V_2 \sim Ber(\alpha)$ and $V_3 \sim Ber(\overline{\alpha})$ and use (2).

Note that if $(U_1, U_2) \sim BVBer(\pi,\alpha)$, then $U_1 \sim Ber(\pi), U_2 \sim Ber(\pi\alpha + \overline{\pi}\overline{\alpha}), Cov(U_1, U_2) = \pi\overline{\pi}(2\alpha - 1)$ and U_1, U_2 are independent if and only if $\alpha = 0.5$.

Definition 3. *The RV (U_1,U_2) with jpmf,*

$$g(u_1,u_2)=\frac{1}{2}\beta^{1-u_1-u_2+2u_1u_2}\overline{\beta}^{u_1+u_2-2u_1u_2},\qquad 0\le\beta\le 1,u_1,u_2=0,1, \tag{4}$$

is said to have the one-parameter BVBer distribution denoted by $BVBer(\beta)$.

Note that $BVBer(\beta)$ is the special case of $BVBer(\underline{\beta})$ when $\beta_{00}=\beta_{11}=\frac{1}{2}\beta$ and $\beta_{01}=\beta_{10}=\frac{1}{2}\overline{\beta}$. Note also that if $(U_1,U_2)\sim BVBer(\beta)$, then $U_i\sim Ber(\frac{1}{2}),i=1,2,Cov(U_1,U_2)=\frac{2\beta-1}{4}$and U_1,U_2 are independent if and only if $\beta=\frac{1}{2}$.

To generate one realization (U_1,U_2) from $BVBer(\beta)$ of (4), we independently generate $U_1\sim Ber(\frac{1}{2}),V_2\sim Ber(\beta)$ and $V_3\sim Ber(\overline{\beta})$ and use (2).

2.2 The BPD

Assume that $W_j\sim Poi(\lambda_j),j=1,2,3$ are independent RV. It is well known that $(X_1=W_1+W_3,X_2=W_2+W_3)$ has the bivariate Poisson distribution $(BPD(\underline{\lambda}))$ with *jpmf*

$$p(s,t;\underline{\lambda})=e^{-(\lambda_1+\lambda_2+\lambda_3)}\frac{\lambda_1^s}{s!}\frac{\lambda_2^t}{t!}\sum_{i=0}^{\min(s,t)}\binom{s}{i}\binom{t}{i}i!\left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^i,\qquad s,t=0,1,2,\ldots. \tag{5}$$

Note that

$$E(X_i)=Var(X_i)=\lambda_i+\lambda_3\text{ and }Cov(X_1,X_2)=\lambda_3. \tag{6}$$

For a comprehensive treatment of the *BPD*, we refer to Kocherlakota and Kocherlakota (1992) and Johnson et al. (1997). The *jpmf* of (5) can be computed by using the R function "pbivpois" of Karlis and Ntzoufras (2005). Let $(X_{1,i},X_{2,i}),i=1,2,\ldots,n$ be a random sample from (5). The *MLE* of λ_1,λ_2 and λ_3 can be obtained by using the R function "simple.bp" of Karlis and Ntzoufras (2005). The method of moments estimators (*MME*) of λ_1,λ_2 and λ_3 are given by

$$\widetilde{\lambda}_j=\overline{X}_j-\widetilde{\lambda}_3,\qquad j=1,2, \tag{7}$$

and

$$\widetilde{\lambda}_3=\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i}-\overline{X}_1\right)\left(X_{2,i}-\overline{X}_2\right). \tag{8}$$

2.3 The BGD of Phatak and Sreehari

The RV (X_1,X_2) with *jpmf*,

$$q(s,t;\underline{\theta})=\binom{s+t}{s}\delta_1^s\delta_2^t(1-\delta_1-\delta_2),\qquad s,t=0,1,2,\ldots \tag{9}$$

where $0<\delta_1,\delta_2<1$ and $0<1-\delta_1-\delta_2<1$, is said to follow the BGD of Phatak and Sreehari (1981), denoted by $BGD(\underline{\delta})$.

Note that for the $BGD(\underline{\delta})$, the following results hold (see, Krishna and Pundir (2009) and Hogg et al. (2005)):

1. $X_1 \sim Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_2}\right)$ and $X_2 \sim Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_1}\right)$.
2. Let $(X_{1,i},X_{2,i}), i = 1,2,\ldots,n$ be a random sample from (9).
   (a) The *MLE* of δ_1 and δ_2 are given by
   $$\widehat{\delta}_j = \frac{\overline{X}_j}{1+\overline{X}_1+\overline{X}_2}, \qquad j = 1,2, \tag{10}$$
   where $\overline{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{j,i}, j = 1,2$.
   (b) Using the result that $q(0,0;\underline{\delta}) = 1-\delta_1-\delta_2$, the *MME* of δ_1 and δ_2 are obtained as follows:
   $$\widetilde{\delta}_j = \frac{\overline{X}_j \times \sum_{i=1}^{n} I(X_{1,i}=0, X_{2,i}=0)}{n}, \qquad j = 1,2. \tag{11}$$
3. We may generate a realization (X_1,X_2) from the $BGD(\underline{\delta})$ as follows:
   (a) Generate X_2 from $Geo(1-\frac{\delta_2}{1-\delta_1})$.
   (b) Given that $X_2 = y$, generate $V_1,V_2,\ldots,V_{y+1}$ independently from $Geo(1-\delta_1)$ and set $X_1 = \sum_{i=1}^{y+1} V_i$.

3 The BRST

Definition 4. *Assume that (U_1,U_2) has a BVBer distribution, that (X_1,X_2) is a nonnegative integer-valued RV independent of (U_1,U_2) with jpmf, $f(s,t)$. Let $f_i(\cdot)$ be the marginal pmf of $X_i, i = 1,2$. Then, the BRST of (X_1,X_2) is defined as*

$$Z_i = (2U_i - 1)X_i, \qquad i = 1,2. \tag{12}$$

3.1 BRST based on the $BVBer(\underline{\beta})$

Assume that $(U_1,U_2) \sim BVBer(\underline{\beta})$ of (1). In this case, the *jpmf* of Z_1 and Z_2 is given as follows:

$$h(0,0) = f(0,0), \tag{13}$$

$$h(s,0) = \begin{cases} (\beta_{11}+\beta_{10})f(s,0), & s = 1,2,\ldots \\ (\beta_{00}+\beta_{01})f(-s,0), & s = -1,-2,\ldots \end{cases} \tag{14}$$

$$h(0,t) = \begin{cases} (\beta_{11}+\beta_{01})f(0,t), & t = 1,2,\ldots \\ (\beta_{00}+\beta_{10})f(0,-t), & t = -1,-2,\ldots \end{cases} \tag{15}$$

and

$$h(s,t) = f(|s|,|t|) \times \begin{cases} \beta_{00}, & s,t = -1,-2,\ldots \\ \beta_{10}, & s = 1,2,\ldots, t = -1,-2,\ldots \\ \beta_{01}, & s = -1,-2,\ldots, t = 1,2,\ldots \\ \beta_{11}, & s,t = 1,2,\ldots. \end{cases} \tag{16}$$

For $i=1,2$, the marginal *pmf* of Z_i is given by

$$h_i(s)=\begin{cases}(\beta_{11}+\beta_{10})\,f_i(s), & s=1,2,\ldots\\ f_i(0), & s=0,\\ (\beta_{00}+I(i=1)\beta_{01}+I(i=2)\beta_{10})\,f_i(-s), & s=-1,-2,\ldots.\end{cases} \tag{17}$$

Lemma 2. *It holds that*

$$E(Z_1^nZ_2^m)=E(X_1^nX_2^m)\times\begin{cases}1, & \text{if } m \text{ and } n \text{ are even},\\ (2\beta_{11}+2\beta_{10}-1), & \text{if } m \text{ is even and } n \text{ is odd},\\ (2\beta_{11}+2\beta_{01}-1), & \text{if } m \text{ is odd and } n \text{ is even},\\ (1-2\beta_{10}-2\beta_{01}), & \text{if } m \text{ and } n \text{ are odd}.\end{cases} \tag{18}$$

Proof. Note that (18) follows from the result that for $m,n=0,1,2,\ldots$, we have

$$Z_1^nZ_2^m=X_1^nX_2^m\times\begin{cases}1, & \text{if } m \text{ and } n \text{ are even},\\ (2U_1-1), & \text{if } m \text{ is even and } n \text{ is odd},\\ (2U_2-1), & \text{if } m \text{ is odd and } n \text{ is even},\\ (2U_1-1)(2U_2-1), & \text{if } m \text{ and } n \text{ are odd}.\end{cases}$$

□

Corollary 1. *By (18), we have*

$$E(Z_1^n)=E(X_1^n)\times\begin{cases}1, & \text{if } n \text{ is even},\\ (2\beta_{11}+2\beta_{10}-1), & \text{if } n \text{ is odd},\end{cases} \tag{19}$$

$$E(Z_2^m)=E(X_2^m)\times\begin{cases}1, & \text{if } m \text{ is even},\\ (2\beta_{11}+2\beta_{01}-1), & \text{if } m \text{ is odd},\end{cases} \tag{20}$$

and

$$E(Z_1Z_2)=(1-2\beta_{10}-2\beta_{01})\,E(X_1X_2). \tag{21}$$

Hence, by (19)-(21), for $i=1,2$,

$$E(Z_i^2)=E(X_i^2),$$

$$E(Z_i)=(2\beta_{11}+2I(i=1)\beta_{10}+2I(i=2)\beta_{01}-1)\,E(X_i), \tag{22}$$

$$\begin{aligned}Var(Z_i)=&Var(X_i)+4\,(\beta_{11}+I(i=1)\beta_{10}+I(i=2)\beta_{01})\\ &\times(1-\beta_{11}-I(i=1)\beta_{10}-I(i=2)\beta_{01})\,(E(X_i))^2\end{aligned} \tag{23}$$

and

$$\begin{aligned}Cov(Z_1,Z_2)=&(1-2\beta_{10}-2\beta_{01})\,Cov(X_1,X_2)\\ &+4E(X_1)E(X_2)\,(\beta_{11}-(\beta_{11}+\beta_{10})\,(\beta_{11}+\beta_{01}))\,.\end{aligned} \tag{24}$$

Corollary 2. *In the special case when U_1 and U_2 are independent (i.e., when $\beta_{11}=\alpha_1\alpha_2,\beta_{10}=\alpha_1\overline{\alpha}_2,\beta_{01}=\overline{\alpha}_1\alpha_2$ and $\beta_{00}=\overline{\alpha}_1\overline{\alpha}_2$ with $0<\alpha_1,\alpha_2<1$) (18)-(24) reduce to the corresponding results of Chesneau et al. (2018).*

Let $EN(W_1,W_2)$ be Shannon's entropy (see Shannon (1951)) of the RV (W_1,W_2) and let $EN(V)$ be Shannon's entropy of the RV V. Then, we have the following lemma.

Lemma 3. *It holds that*

$$\begin{aligned} EN(Z_1,Z_2) &= EN(X_1,X_2)+EN(U_1,U_2)\{1-f_1(0)-f_2(0)+f(0,0)\} \\ &\quad +(f_1(0)-f(0,0))EN(U_2)+(f_2(0)-f(0,0))EN(U_1). \end{aligned} \tag{25}$$

Proof.

$$EN(Z_1,Z_2) = -\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} h(i,j)\ln h(i,j) = \sum_{r=1}^{9} S_r, \tag{26}$$

where

$$S_1 = -f(0,0)\ln f(0,0), \tag{27}$$

$$S_2 = -\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} h(i,j)\ln h(i,j), \quad S_3 = -\sum_{i=-1}^{-\infty}\sum_{j=1}^{\infty} h(-i,j)\ln h(-i,j),$$

$$S_4 = -\sum_{i=1}^{\infty}\sum_{j=-1}^{-\infty} h(i,-j)\ln h(i,-j), \quad S_5 = -\sum_{i=-1}^{-\infty}\sum_{j=-1}^{-\infty} h(-i,-j)\ln h(-i,-j),$$

$$S_6 = -\sum_{j=1}^{\infty} h(0,j)\ln h(0,j), \quad S_7 = -\sum_{j=-1}^{-\infty} h(0,-j)\ln h(0,-j),$$

$$S_8 = -\sum_{i=1}^{\infty} h(i,0)\ln h(i,0), \quad \text{and} \quad S_9 = -\sum_{j=-1}^{-\infty} h(-i,0)\ln h(-i,0).$$

For S_2, we have

$$S_2 = -\beta_{11}\ln\beta_{11}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j) - \beta_{11}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)\ln f(i,j) = -\beta_{11}T_1\ln\beta_{11}+\beta_{11}T_2,$$

where

$$T_1 = 1-f_1(0)-f_2(0)+f(0,0) \quad \text{and} \quad T_2 = -\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)\ln f(i,j).$$

For S_3, we have

$$\begin{aligned} S_3 &= -\sum_{i=-1}^{-\infty}\sum_{j=1}^{\infty}\beta_{01}f(-i,j)\{\ln\beta_{01}+\ln f(-i,j)\} \\ &= -\beta_{01}\ln\beta_{01}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j) - \beta_{01}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(i,j)\ln f(i,j) \\ &= -\beta_{01}\ln\beta_{01}T_1+\beta_{01}T_2. \end{aligned}$$

Similarly,

$$S_4 = -\beta_{10}\ln\beta_{10}T_1+\beta_{10}T_2 \quad \text{and} \quad S_5 = -\beta_{00}\ln\beta_{00}T_1+\beta_{00}T_2.$$

Hence,

$$\sum_{r=2}^{5} S_r = EN(U_1,U_2)T_1 + T_2. \tag{28}$$

We can show that

$$T_2 = EN(X_1,X_2) - f(0,0)\ln f(0,0) + \sum_{i=0}^{\infty}[f(0,i)\ln f(0,i) + f(i,0)\ln f(i,0)]. \tag{29}$$

For S_6, we have

$$\begin{aligned}
S_6 &= -\sum_{j=1}^{\infty}(\beta_{11}+\beta_{01})\,f(0,j)\,\{\ln(\beta_{11}+\beta_{01}) + \ln f(0,j))\} \\
&= -(\beta_{11}+\beta_{01})\ln(\beta_{11}+\beta_{01})\sum_{j=1}^{\infty} f(0,j) - (\beta_{11}+\beta_{01})\sum_{j=1}^{\infty} f(0,j)\ln f(0,j) \\
&= -(\beta_{11}+\beta_{01})\ln(\beta_{11}+\beta_{01})\,(f_1(0) - f(0,0)) + (\beta_{11}+\beta_{01})\,f(0,0)\ln f(0,0) \\
&\quad -(\beta_{11}+\beta_{01})\sum_{i=0}^{\infty} f(0,i)\ln f(0,i).
\end{aligned}$$

For S_7, we have

$$\begin{aligned}
S_7 &= -(\beta_{00}+\beta_{10})\ln(\beta_{00}+\beta_{10})\,(f_1(0) - f(0,0)) + (\beta_{00}+\beta_{10})\,f(0,0)\ln f(0,0) \\
&\quad -(\beta_{00}+\beta_{10})\sum_{i=0}^{\infty} f(0,i)\ln f(0,i).
\end{aligned}$$

Hence

$$S_6 + S_7 = -\sum_{i=0}^{\infty} f(0,i)\ln f(0,i) + (f_1(0) - f(0,0))\,EN(U_2) + f(0,0)\ln f(0,0). \tag{30}$$

Similarly,

$$S_8 + S_9 = -\sum_{i=0}^{\infty} f(i,0)\ln f(i,0) + (f_2(0) - f(0,0))\,EN(U_1) + f(0,0)\ln f(0,0). \tag{31}$$

By (26)-(31), we obtain (25).

□

3.1.1 Maximum likelihood estimators (MLE)

Assume that $\underline{\theta}$ is the unknown parameter vector in the *jpmf* of (X_1,X_2) and hence in the *jpmf* of (Z_1,Z_2). In what follows, the presence of $\underline{\theta}$ will be made explicit in both f and h. Let $(Z_{1,i},Z_{2,i}), i = 1,2,\ldots,n$ be a random sample from $h(\cdot,\cdot;\underline{\theta})$ of (13)-(16). Define

$$n_0 = \sum_{i=1}^{n} I(Z_{1,i}=0, Z_{2,i}=0), \quad n_{+,0} = \sum_{i=1}^{n} I(Z_{1,i}>0, Z_{2,i}=0), \tag{32}$$

$$n_{-,0}=\sum_{i=1}^{n} I\left(Z_{1,i}<0, Z_{2,i}=0\right), \quad n_{0,+}=\sum_{i=1}^{n} I\left(Z_{1,i}=0, Z_{2,i}>0\right), \tag{33}$$

$$n_{0,-}=\sum_{i=1}^{n} I\left(Z_{1,i}=0, Z_{2,i}<0\right), \quad n_{+,+}=\sum_{i=1}^{n} I\left(Z_{1,i}>0, Z_{2,i}>0\right), \tag{34}$$

$$n_{-,+}=\sum_{i=1}^{n} I\left(Z_{1,i}<0, Z_{2,i}>0\right), \quad n_{+,-}=\sum_{i=1}^{n} I\left(Z_{1,i}>0, Z_{2,i}<0\right), \tag{35}$$

and

$$n_{-,-}=\sum_{i=1}^{n} I\left(Z_{1,i}<0, Z_{2,i}<0\right). \tag{36}$$

We can show that

$$E(n_0)=nP(Z_1=0,Z_2=0)=nf(0,0;\underline{\theta}), \tag{37}$$

$$E\left(n_{+,0}\right)=n\left(\beta_{11}+\beta_{10}\right)\left(f_2(0;\underline{\theta})-f(0,0;\underline{\theta})\right), \tag{38}$$

$$E\left(n_{-,0}\right)=n\left(1-\beta_{11}-\beta_{10}\right)\left(f_2(0;\underline{\theta})-f(0,0;\underline{\theta})\right), \tag{39}$$

$$E\left(n_{0,+}\right)=n\left(\beta_{11}+\beta_{01}\right)\left(f_1(0;\underline{\theta})-f(0,0;\underline{\theta})\right), \tag{40}$$

$$E\left(n_{0,-}\right)=n\left(1-\beta_{11}-\beta_{01}\right)\left(f_1(0;\underline{\theta})-f(0,0;\underline{\theta})\right), \tag{41}$$

$$E\left(n_{+,+}\right)=n\beta_{11}\left(1-f_1(0;\underline{\theta})-f_2(0;\underline{\theta})+f(0,0;\underline{\theta})\right), \tag{42}$$

$$E\left(n_{-,+}\right)=n\beta_{01}\left(1-f_1(0;\underline{\theta})-f_2(0;\underline{\theta})+f(0,0;\underline{\theta})\right), \tag{43}$$

$$E\left(n_{+,-}\right)=n\beta_{10}\left(1-f_1(0;\underline{\theta})-f_2(0;\underline{\theta})+f(0,0;\underline{\theta})\right), \tag{44}$$

and

$$E\left(n_{-,-}\right)=n\left(1-\beta_{11}-\beta_{10}-\beta_{01}\right)\left(1-f_1(0;\underline{\theta})-f_2(0;\underline{\theta})+f(0,0;\underline{\theta})\right). \tag{45}$$

Lemma 4. *Assume that $\underline{T}(\underline{X}_1,\underline{X}_2)$ is the MLE of $\underline{\theta}$ based on a random sample from $f(\cdot,\cdot;\underline{\theta})$, and let $I_{X_1,X_2}(\underline{\theta})$ be the corresponding Fisher Information Matrix. Let $(Z_{1,i},Z_{2,i}), i=1,2,\ldots,n$ be a random sample from $h(\cdot,\cdot;\underline{\theta})$ and let $n_{+,0},\ldots,n_{-,-}$ be as in (32)-(36). Then, for the MLE of $\underline{\theta},\beta_{11},\beta_{10}$ and β_{01}, we have*

1. $$\widehat{\underline{\theta}}=\underline{T}(|\underline{Z_1}|,|\underline{Z_2}|).$$

2. *$\widehat{\beta}_{11},\widehat{\beta}_{10}$ and $\widehat{\beta}_{01}$ are obtained by maximizing*

$$\begin{aligned} l_1 =& n_{+,0}\ln\left(\beta_{11}+\beta_{10}\right)+n_{-,0}\ln\left(1-\beta_{11}-\beta_{10}\right)+n_{0,+}\ln\left(\beta_{11}+\beta_{01}\right) \\ &+n_{0,-}\ln\left(1-\beta_{11}-\beta_{01}\right)+n_{+,+}\ln\beta_{11}+n_{-,+}\ln\beta_{01} \\ &+n_{+,-}\ln\beta_{10}+n_{-,-}\ln\left(1-\beta_{11}-\beta_{10}-\beta_{01}\right) \end{aligned} \tag{46}$$

subject to the constraints,

$$0\le\beta_{11}\le 1, \quad 0\le\beta_{10}\le 1, \quad 0\le\beta_{01}\le 1, \text{ and } 0\le\beta_{11}+\beta_{10}+\beta_{01}\le 1.$$

3.

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_{11}-\beta_{11} \\ \widehat{\beta}_{10}-\beta_{10} \\ \widehat{\beta}_{01}-\beta_{01} \\ \widehat{\underline{\theta}}-\underline{\theta} \end{pmatrix} \xrightarrow{D} MVN(\underline{0}, \begin{bmatrix} \Sigma_1^{-1} & \underline{0} \\ \underline{0} & I^{-1}_{X_1,X_2}(\underline{\theta}) \end{bmatrix}, \tag{47}$$

where

$$\sum_1 = \begin{bmatrix} \sigma_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3 \end{bmatrix},$$

$$\sigma_{23} = \frac{1-f_1(0;\underline{\theta})-f_2(0;\underline{\theta})+f(0,0;\underline{\theta})}{1-\beta_{11}-\beta_{10}-\beta_{01}},$$

$$\sigma_1 = \frac{f_2(0;\underline{\theta})-f(0,0;\underline{\theta})}{(\beta_{11}+\beta_{10})(1-\beta_{11}-\beta_{10})} + \frac{f_1(0;\underline{\theta})-f(0,0;\underline{\theta})}{(\beta_{11}+\beta_{01})(1-\beta_{11}-\beta_{01})} + \frac{(1-\beta_{10}-\beta_{01})\sigma_{23}}{\beta_{11}},$$

$$\sigma_2 = \frac{f_2(0;\underline{\theta})-f(0,0;\underline{\theta})}{(\beta_{11}+\beta_{10})(1-\beta_{11}-\beta_{10})} + \frac{(1-\beta_{11}-\beta_{01})\sigma_{23}}{\beta_{10}},$$

$$\sigma_3 = \frac{f_1(0;\underline{\theta})-f(0,0;\underline{\theta})}{(\beta_{11}+\beta_{01})(1-\beta_{11}-\beta_{01})} + \frac{(1-\beta_{11}-\beta_{10})\sigma_{23}}{\beta_{01}},$$

$$\sigma_{12} = \frac{f_2(0;\underline{\theta})-f(0,0;\underline{\theta})}{(\beta_{11}+\beta_{10})(1-\beta_{11}-\beta_{10})} + \sigma_{23},$$

and

$$\sigma_{13} = \frac{f_1(0;\underline{\theta})-f(0,0;\underline{\theta})}{(\beta_{11}+\beta_{01})(1-\beta_{11}-\beta_{01})} + \sigma_{23}.$$

Proof. The log-likelihood function (Log-LF) of the sample is given by

$$l = l_1 + l_2, \tag{48}$$

where l_1 is as in (46) and

$$l_2 = \sum_{i=1}^{n} \ln f(|z_{1,i}|, |z_{2,i}|; \underline{\theta}). \tag{49}$$

It is clear from (48),(46), and (49) that the MLE of $\underline{\theta}$ is obtained by maximizing l_2 and the MLE of $\beta_{11}\beta_{10}$ and β_{01} are obtained by maximizing l_1 subject to the constraints,

$$0 \le \beta_{11} \le 1,\ 0 \le \beta_{10} \le 1,\ 0 \le \beta_{01} \le 1,\ \text{and } 0 \le \beta_{11}+\beta_{10}+\beta_{01} \le 1.$$

We will use the R function "constrOptim" to obtain the MLE of $\beta_{11}\beta_{10}$ and β_{01}. To obtain $\sum_1$ of (47) we use (37)-(45) and the following results:

$$\frac{\partial^2 l}{\partial \beta_{11}^2} = \frac{\partial^2 l_1}{\partial \beta_{11}^2} = -\frac{n_{+,0}}{(\beta_{11}+\beta_{10})^2} - \frac{n_{-,0}}{(1-\beta_{11}-\beta_{10})^2} - \frac{n_{0,+}}{(\beta_{11}+\beta_{01})^2} - \frac{n_{0,-}}{(1-\beta_{11}-\beta_{01})^2} - \frac{n_{+,+}}{\beta_{11}^2} - \frac{n_{-,-}}{(1-\beta_{11}-\beta_{10}-\beta_{01})^2},$$

$$\frac{\partial^2 l}{\partial \beta_{10}^2} = \frac{\partial^2 l_1}{\partial \beta_{10}^2} = -\frac{n_{+,0}}{(\beta_{11}+\beta_{10})^2} - \frac{n_{-,0}}{(1-\beta_{11}-\beta_{10})^2} - \frac{n_{+,-}}{\beta_{10}^2} - \frac{n_{-,-}}{(1-\beta_{11}-\beta_{10}-\beta_{01})^2},$$

$$\frac{\partial^2 l}{\partial \beta_{01}^2} = \frac{\partial^2 l_1}{\partial \beta_{01}^2} = -\frac{n_{0,+}}{(\beta_{11}+\beta_{01})^2} - \frac{n_{0,-}}{(1-\beta_{11}-\beta_{01})^2} - \frac{n_{-,+}}{\beta_{01}^2} - \frac{n_{-,-}}{(1-\beta_{11}-\beta_{10}-\beta_{01})^2},$$

$$\frac{\partial^2 l}{\partial \beta_{10}\partial \beta_{11}} = \frac{\partial^2 l_1}{\partial \beta_{10}\partial \beta_{11}} = -\frac{n_{+,0}}{(\beta_{11}+\beta_{10})^2} - \frac{n_{-,0}}{(1-\beta_{11}-\beta_{10})^2} - \frac{n_{-,-}}{(1-\beta_{11}-\beta_{10}-\beta_{01})^2},$$

$$\frac{\partial^2 l}{\partial \beta_{01}\partial \beta_{11}} = \frac{\partial^2 l_1}{\partial \beta_{01}\partial \beta_{11}} = -\frac{n_{0,+}}{(\beta_{11}+\beta_{01})^2} - \frac{n_{0,-}}{(1-\beta_{11}-\beta_{01})^2} - \frac{n_{-,-}}{(1-\beta_{11}-\beta_{10}-\beta_{01})^2},$$

and

$$\frac{\partial l}{\partial \beta_{10}\partial \beta_{01}} = \frac{\partial l_1}{\partial \beta_{10}\partial \beta_{01}} = -\frac{n_{-,-}}{(1-\beta_{11}-\beta_{10}-\beta_{01})^2}.$$

□

3.1.2 Method of moments estimators (MME)

Lemma 5. *Assume that $\widetilde{\underline{T}}(\cdot,\cdot)$ is the MME of $\underline{\theta}$ based on a random sample from $f(\cdot,\cdot;\underline{\theta})$. Then, for the MME of $\underline{\theta}, \beta_{11}, \beta_{10}$ and β_{01} we have*

$$\widetilde{\underline{\theta}} = \widetilde{\underline{T}}(|\underline{Z_1}|, |\underline{Z_2}|), \tag{50}$$

$$\widetilde{\beta}_{10} = \frac{1}{4}\{1 + A_1 - A_2 - C\}, \tag{51}$$

$$\widetilde{\beta}_{01} = \frac{1}{4}\{1 - A_1 + A_2 - C\}, \tag{52}$$

and

$$\widetilde{\beta}_{11} = \frac{1}{4}\{1 + A_1 + A_2 + C\}, \tag{53}$$

where

$$A_j = \frac{\sum_{i=1}^n Z_{j,i}}{\sum_{i=1}^n |Z_{j,i}|}, \qquad j = 1,2, \tag{54}$$

and

$$C = \frac{\sum_{i=1}^n Z_{1,i} Z_{2,i}}{\sum_{i=1}^n |Z_{1,i} Z_{2,i}|}. \tag{55}$$

Proof. Note that $\widetilde{\beta}_{11}, \widetilde{\beta}_{10}$ and $\widetilde{\beta}_{01}$ are obtained as follows. We start by solving

$$\frac{1}{n}\sum_{i=1}^n Z_{1,i} = (2\beta_{11} + 2\beta_{10} - 1)\left\{E(X_1)\Big|_{\underline{\theta}=\widetilde{\underline{\theta}}}\right\}, \tag{56}$$

$$\frac{1}{n}\sum_{i=1}^n Z_{2,i} = (2\beta_{11} + 2\beta_{01} - 1)\left\{E(X_2)\Big|_{\underline{\theta}=\widetilde{\underline{\theta}}}\right\}, \tag{57}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Z_{1,i}Z_{2,i} = (1-2\beta_{01}-2\beta_{10})\left\{E(X_1X_2)\Big|_{\underline{\theta}=\widetilde{\underline{\theta}}}\right\}. \tag{58}$$

Note that

$$X_1X_2 = |Z_1Z_2| \text{ and } E(X_1X_2) = E(|Z_1Z_2|).$$

Hence, (58) can be replaced by

$$\frac{1}{n}\sum_{i=1}^{n} Z_{1,i}Z_{2,i} = (1-2\beta_{01}-2\beta_{10})\frac{1}{n}\sum_{i=1}^{n}|Z_{1,i}Z_{2,i}|. \tag{59}$$

We can show that (56), (57), and (59) are, respectively, equivalent to

$$A_1 = 2\widetilde{\beta}_{11}+2\widetilde{\beta}_{10}-1, \tag{60}$$

$$A_2 = 2\widetilde{\beta}_{11}+2\widetilde{\beta}_{01}-1, \tag{61}$$

and

$$C = 1-2\widetilde{\beta}_{01}-2\widetilde{\beta}_{10}. \tag{62}$$

By solving (60)-(62), we obtain (51)-(53). □

Consider the special case when X_1 and X_2 are independent. In this case, the *MME* estimators of β_{10},β_{01} and β_{11} are as given in (51)-(53) after replacing C of (55) with

$$C_1 = \frac{n\sum_{i=1}^{n} Z_{1,i}Z_{2,i}}{\left(\sum_{i=1}^{n}|Z_{1,i}|\right)\left(\sum_{i=1}^{n}|Z_{2,i}|\right)}.$$

3.2 BRST based on the $BVBer(\pi,\alpha)$ distribution

Assume that (U_1,U_2) has the $BVBer(\pi,\alpha)$ distribution of (3). Define (Z_1,Z_2) as in (12). Then, the *jpmf* of Z_1 and Z_2 is given as follows:

$$h(0,0) = f(0,0), \tag{63}$$

$$h(s,0) = \begin{cases} \pi f(s,0), & s=1,2,\ldots, \\ \overline{\pi} f(-s,0), & s=-1,-2,\ldots \end{cases} \tag{64}$$

$$h(0,t) = \begin{cases} (\alpha\pi+\overline{\alpha\pi})\, f(0,t), & t=1,2,\ldots, \\ (\overline{\alpha}\pi+\alpha\overline{\pi})\, f(0,-t), & t=-1,-2,\ldots \end{cases} \tag{65}$$

and

$$h(s,t) = f(|s|,|t|)\times\begin{cases} \overline{\pi}\alpha, & s,t=-1,-2,\ldots \\ \overline{\alpha\pi}, & s=1,2,\ldots,t=-1,-2,\ldots \\ \overline{\alpha\pi}, & s=-1,-2,\ldots,t=1,2,\ldots \\ \alpha\pi, & s,t=1,2,\ldots. \end{cases} \tag{66}$$

The marginal $pmf's$ of Z_1 and Z_2 are given by

$$h_1(s) = \begin{cases} \pi f_1(s), & s = 1,2,\ldots \\ f_1(0), & s = 0, \\ \overline{\pi} f_1(-s), & s = -1,-2,\ldots \end{cases} \tag{67}$$

and

$$h_2(t) = \begin{cases} (\alpha\pi + \overline{\alpha\pi})\, f_2(t), & t = 1,2,\ldots, \\ f_2(0), & t = 0, \\ (\overline{\alpha}\pi + \alpha\overline{\pi})\, f_2(-t), & t = -1,-2,\ldots. \end{cases} \tag{68}$$

Lemma 6. *It holds that*

$$E(Z_1^n Z_2^m) = E(X_1^n X_2^m) \times \begin{cases} 1, & \text{if } m \text{ and } n \text{ are even}, \\ (2\pi - 1), & \text{if } m \text{ is even and } n \text{ is odd}, \\ (2\alpha - 1)(2\pi - 1), & \text{if } m \text{ is odd and } n \text{ is even}, \\ (2\alpha - 1), & \text{if } m \text{ and } n \text{ are odd}, \end{cases}$$

$$E(Z_1^n) = E(X_1^n) \times \begin{cases} 1, & \text{if } n \text{ is even}, \\ (2\pi - 1), & \text{if } n \text{ is odd}, \end{cases}$$

$$E(Z_2^m) = E(X_2^m) \times \begin{cases} 1, & \text{if } m \text{ is even}, \\ (2\alpha - 1)(2\pi - 1), & \text{if } m \text{ is odd}, \end{cases}$$

$$E(Z_1) = (2\pi - 1) E(X_1),$$

$$E(Z_2) = (2\alpha - 1)(2\pi - 1) E(X_2),$$

$$Var(Z_1) = Var(X_1) + 4\pi\overline{\pi}\,(E(X_1))^2,$$

$$Var(Z_2) = Var(X_2) + 2\,(\alpha\overline{\pi} + \overline{\alpha}\pi)\,(E(X_2))^2,$$

$$E(Z_1 Z_2) = (2\alpha - 1) E(X_1 X_2),$$

and

$$Cov(Z_1, Z_2) = (2\alpha - 1)\{Cov(X_1, X_2) + 4\pi\overline{\pi} E(X_1) E(X_2)\}.$$

3.2.1 MLE estimators

Assume that $(Z_{1,i}, Z_{2,i}), i = 1,2,\ldots,n$ is a random sample from $h(\cdot,\cdot;\underline{\theta})$ of (63)-(66). Let l_2 be as in (49) and let $n_{\pm,0}, n_{\pm,-}, n_{\pm,+}$ be as in (32)-(36) and

$$n_{\pm,\cdot} = n_{\pm,0} + n_{\pm,-} + n_{\pm,+}.$$

The Log-LF of the sample is given by

$$l_3 = l_2 + l_4,$$

where

$$\begin{aligned} l_4 = & n_{+,\cdot} \ln \pi + n_{-,\cdot} \ln \overline{\pi} + (n_{+,+} + n_{-,-}) \ln(\alpha) + (n_{-,+} + n_{+,-}) \ln \overline{\alpha} \\ & + n_{0,+} \ln(\alpha\pi + \overline{\alpha\pi}) + n_{0,-} \ln(\alpha\overline{\pi} + \overline{\alpha}\pi). \end{aligned}$$

Lemma 7. *Assume that $\underline{T}(\cdot,\cdot)$ is the MLE of $\underline{\theta}$ based on a random sample from $f(\cdot,\cdot;\underline{\theta})$, and let $I_{X_1,X_2}(\underline{\theta})$ be the corresponding Fisher information Matrix. Let $\widehat{\underline{\theta}},\widehat{\pi}$ and $\widehat{\alpha}$ be the MLE of $\underline{\theta},\pi$ and α. Then,*
1.

$$\widehat{\underline{\theta}} = \underline{T}(|\underline{Z_1}|,|\underline{Z_2}|).$$

2. $\widehat{\pi}$ and $\widehat{\alpha}$ are obtained by maximizing l_4 subject to the constraints,

$$0 \le \alpha \le 1 \text{ and } 0 \le \pi \le 1.$$

3. As $n \longrightarrow \infty$,

$$\sqrt{n}\begin{pmatrix} \widehat{\pi}-\pi \\ \widehat{\alpha}-\alpha \\ \widehat{\underline{\theta}}-\underline{\theta} \end{pmatrix} \xrightarrow{D} MVN(\underline{0},\begin{bmatrix} \Sigma_2^{-1} & \underline{0} \\ \underline{0} & I^{-1}_{X_1,X_2}(\underline{\theta}) \end{bmatrix},$$

where

$$\sum_2 = \begin{bmatrix} \sigma_1^* & \sigma_{12}^* \\ \sigma_{12}^* & \sigma_2^* \end{bmatrix},$$

$$\sigma_1^* = \frac{(f_1(0;\underline{\theta}_1) - f(0,0;\underline{\theta}_1,\underline{\theta}_2))(2\alpha-1)^2}{(\alpha\pi+\overline{\alpha}\overline{\pi})(\alpha\overline{\pi}+\overline{\alpha}\pi)} + \frac{1-f_1(0;\underline{\theta}_1)}{\pi\overline{\pi}},$$

$$\sigma_2^* = \frac{(f_1(0;\underline{\theta}_1) - f(0,0;\underline{\theta}_1,\underline{\theta}_2))(2\pi-1)^2}{(\alpha\pi+\overline{\alpha}\overline{\pi})(\alpha\overline{\pi}+\overline{\alpha}\pi)} + \frac{1-f_1(0;\underline{\theta}_1)-f_2(0;\underline{\theta}_2)+f(0,0;\underline{\theta}_1,\underline{\theta}_2))}{\alpha\overline{\alpha}},$$

and

$$\sigma_{12}^* = \frac{(f_1(0;\underline{\theta}_1) - f(0,0;\underline{\theta}_1,\underline{\theta}_2))(2\alpha-1)(2\pi-1)}{(\alpha\pi+\overline{\alpha}\overline{\pi})(\alpha\overline{\pi}+\overline{\alpha}\pi)}.$$

Proof. To obtain $\sum_2$, we use (38)-(45), (69)-(71),

$$\frac{\partial^2 l_4}{\partial\alpha\partial\pi} = 2\left\{\frac{n_{0,+}}{\alpha\pi+\overline{\alpha}\overline{\pi}} - \frac{n_{0,-}}{\alpha\overline{\pi}+\overline{\alpha}\pi}\right\} - (2\alpha-1)(2\pi-1)\left\{\frac{n_{0,+}}{(\alpha\pi+\overline{\alpha}\overline{\pi})^2} + \frac{n_{0,-}}{(\alpha\overline{\pi}+\overline{\alpha}\pi)^2}\right\}, \tag{69}$$

$$\frac{\partial^2 l_4}{\partial\pi^2} = -\frac{n_{+,\cdot}}{\pi^2} - \frac{n_{-,\cdot}}{\overline{\pi}^2} - (2\alpha-1)^2\left\{\frac{n_{0,+}}{(\alpha\pi+\overline{\alpha}\overline{\pi})^2} + \frac{n_{0,-}}{(\alpha\overline{\pi}+\overline{\alpha}\pi)^2}\right\}, \tag{70}$$

and

$$\frac{\partial^2 l_4}{\partial\alpha^2} = -(2\pi-1)^2\left\{\frac{n_{0,+}}{(\alpha\pi+\overline{\alpha}\overline{\pi})^2} + \frac{n_{0,-}}{(\alpha\overline{\pi}+\overline{\alpha}\pi)^2}\right\} - \frac{n_{+,+}+n_{-,-}}{\alpha^2} - \frac{n_{-,+}+n_{+,-}}{\overline{\alpha}^2}. \tag{71}$$

□

3.2.2 MME estimators

Lemma 8. *Assume that $\widetilde{\underline{T}}(\cdot,\cdot)$ is the MME of $\underline{\theta}$ based on a random sample from $f(\cdot,\cdot;\underline{\theta})$. For the MME of $\underline{\theta},\pi$, and α, we have*

$$\widetilde{\underline{\theta}} = \widetilde{\underline{T}}(|\underline{Z_1}|,|\underline{Z_2}|),$$

$$\widetilde{\pi} = \frac{\sum_{i=1}^{n} Z_{1,i} + \sum_{i=1}^{n} |Z_{1,i}|}{2\sum_{i=1}^{n} |Z_{1,i}|}, \tag{72}$$

and

$$\widetilde{\alpha} = \frac{1}{2}\left\{ \frac{(\sum_{i=1}^{n} |Z_{1,i}|) \times (\sum_{i=1}^{n} Z_{2,i})}{(\sum_{i=1}^{n} Z_{1,i}) \times (\sum_{i=1}^{n} |Z_{2,i}|)} + 1 \right\}. \tag{73}$$

Note that $\widetilde{\pi}$ and $\widetilde{\alpha}$ are obtained by solving

$$\sum_{i=1}^{n} Z_{1,i} = (2\pi - 1)\sum_{i=1}^{n} |Z_{1,i}|$$

and

$$\sum_{i=1}^{n} Z_{2,i} = (2\alpha - 1)(2\pi - 1)\sum_{i=1}^{n} |Z_{2,i}|.$$

3.3 BRST based on $BVBer(\beta)$

Assume that (U_1,U_2) has the $BVBer(\beta)$ distribution of (4). Define (Z_1,Z_2) as in (12). Then, the *jpmf* of Z_1 and Z_2 is given as follows:

$$h(0,0) = f(0,0), \tag{74}$$

$$h(s,0) = \frac{1}{2} \times \begin{cases} f(s,0), & s = 1,2,\ldots \\ f(-s,0), & s = -1,-2,\ldots \end{cases} \tag{75}$$

$$h(0,t) = \frac{1}{2} \times \begin{cases} f(0,t), & t = 1,2,\ldots \\ f(0,-t), & t = -1,-2,\ldots \end{cases} \tag{76}$$

and

$$h(s,t) = \frac{1}{2} f(|s|,|t|) \times \begin{cases} \beta, & s,t = -1,-2,\ldots \\ \overline{\beta}, & s = 1,2,\ldots,t = -1,-2,\ldots \\ \overline{\beta}, & s = -1,-2,\ldots,t = 1,2,\ldots \\ \beta, & s,t = 1,2,\ldots. \end{cases} \tag{77}$$

The marginal *pmf's* of Z_1 and Z_2 are given, for $i = 1,2$, by

$$h_i(s) = \begin{cases} \frac{1}{2} f_i(s), & s = 1,2,\ldots \\ f_i(0), & s = 0, \\ \frac{1}{2} f_i(-s), & s = -1,-2,\ldots. \end{cases} \tag{78}$$

Lemma 9. *It holds that*

$$E(Z_1^n Z_2^m) = E(X_1^n X_2^m) \times \begin{cases} 1, & \text{if } m \text{ and } n \text{ are even,} \\ 0, & \text{if } m \text{ is even and } n \text{ is odd,} \\ 0, & \text{if } m \text{ is odd and } n \text{ is even,} \\ 2\beta - 1, & \text{if } m \text{ and } n \text{ are odd,} \end{cases} \tag{79}$$

$$E(Z_i^n) = E(X_i^n) \times \begin{cases} 1, & \text{if } n \text{ is even,} \\ 0, & \text{if } n \text{ is odd,} \end{cases} \qquad i = 1,2, \tag{80}$$

$$Var(Z_i) = Var(X_i) + (E(X_i))^2, \qquad i = 1,2, \tag{81}$$

and

$$Cov(Z_1, Z_2) = (2\beta - 1)\{Cov(X_1, X_2) + E(X_1)E(X_2)\}. \tag{82}$$

Proof. Note that for $i = 1,2$ and $r = 0,1,2,\ldots,$

$$Z_i^r = \begin{cases} X_i^r, & \text{if } r \text{ is even,} \\ (2U_i - 1)X_i^r, & \text{if } r \text{ is odd.} \end{cases}$$

Consequently, for $m,n = 0,1,2,\ldots,$

$$Z_1^n Z_2^m = X_1^n X_2^m \times \begin{cases} 1, & \text{if } m \text{ and } n \text{ are even,} \\ (2U_1 - 1), & \text{if } m \text{ is even and } n \text{ is odd,} \\ (2U_2 - 1), & \text{if } m \text{ is odd and } n \text{ is even,} \\ (2U_1 - 1)(2U_2 - 1), & \text{if } m \text{ and } n \text{ are odd.} \end{cases}$$

Hence, we obtain (79) and (80). Using (79) and (80) ,we obtain

$$E(Z_1) = E(Z_2) = 0,$$

$$E(Z_i^2) = E(X_i^2), \qquad i = 1,2,$$

and

$$E(Z_1 Z_2) = (2\beta - 1)E(X_1 X_2).$$

Hence, we obtain (81) and (82). □

Remark 1. *1. For the MLE, we use the notation of Lemma 4. We can show that the log-LF is given by*

$$l = C + (n_{+,+} + n_{-,-}) \ln \beta + (n_{+,-} + n_{-,+}) \ln \overline{\beta} + l_2,$$

where l_2 is as in (49). Hence, the MLE of $\underline{\theta}$ is as in Lemma 4 and the MLE of β is given by

$$\widehat{\beta} = \frac{n_{+,+} + n_{-,-}}{n_{+,+} + n_{-,-} + n_{+,-} + n_{-,+}}. \tag{83}$$

In addition, as $n \longrightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \widehat{\beta} - \beta \\ \widehat{\underline{\theta}} - \underline{\theta} \end{pmatrix} \xrightarrow{D} MVN\left(\underline{0}, daig\left\{\beta\overline{\beta}, I^{-1}_{X_1,X_2}(\underline{\theta})\right\}\right).$$

2. The MME of $\underline{\theta}$ is as in (50) and the MME of β is given by

$$\widehat{\beta}_m = \frac{1}{2}\left\{\frac{\sum Z_{1,i}Z_{2,i}}{\sum |Z_{1,i}Z_{2,i}|} + 1\right\}. \tag{84}$$

4 BRST of the BPD

Assume that (U_1, U_2) has a *BVBer* distribution, that (X_1, X_2) has the *BPD* of (5) and that (X_1, X_2) is independent of (U_1, U_2). In the three models of this section, we may estimate λ_1, λ_2 and λ_3 using the following alternatives:

1. The *MLE* estimators obtained as in Lemma 3 using the R function "simple.bp" of Karlis and Ntzoufras (2005) on $(|Z_{1,i}|, |Z_{2,i}|), i = 1, 2, \ldots, n$.
2. The *MME* of (8) and (7) expressed in terms of $(|Z_{1,i}|, |Z_{2,i}|), i = 1, 2, \ldots, n$.

4.1 Models based on the $BVBer(\underline{\beta})$ distribution

Assume here that (U_1, U_2) has the $BVBer(\underline{\beta})$ distribution of (1). In this case, the *jpmf* of Z_1 and Z_2 is given as in (13)-(16) after replacing $f(\cdot,\cdot)$ by $p(\cdot,\cdot;\underline{\lambda})$ of (5). The marginal *pmf's* of Z_1 and Z_2 are given as in (13) and (16) after replacing $f_i(\cdot)$ by the *pdf* of $Poi(\lambda_i + \lambda_3), i = 1, 2$. Using (22), (23), (24), and (6) we obtain

$$\begin{aligned} Cov(Z_1, Z_2) = &(1 - 2\beta_{10} - 2\beta_{01})\,\lambda_3 + 4\,(\lambda_1 + \lambda_3)\,(\lambda_2 + \lambda_3) \\ &\times (\beta_{11} - (\beta_{11} + \beta_{10})\,(\beta_{11} + \beta_{01}))\,, \end{aligned}$$

$$E(Z_i) = (2\beta_{11} + 2I(i=1)\beta_{10} + 2I(i=2)\beta_{01} - 1)\,(\lambda_1 I(i=1) + \lambda_2 I(i=2) + \lambda_3)\,,$$

and

$$\begin{aligned} Var(Z_i) = &(\lambda_1 I(i=1) + \lambda_2 I(i=2) + \lambda_3) + 4\,(\beta_{11} + I(i=1)\beta_{10} + I(i=2)\beta_{01}) \\ &\times (1 - \beta_{11} - I(i=1)\beta_{10} - I(i=2)\beta_{01})\,(\lambda_1 I(i=1) + \lambda_2 I(i=2) + \lambda_3)^2\,. \end{aligned}$$

For the estimation of β_{11}, β_{10}, and β_{01}, we may use the following alternatives:

1. The *MLE* estimators obtained as in Lemma 3 using the R function "constrOptim".
2. The *MME* of (51)-(53).

In each of Figures 1 and 2, we give the scatter plot of a random sample of size $n = 100$ from the *BRST* of the *BPD* for selected values of $\underline{\beta}$ and $\underline{\lambda}$.

4.2 Models based on the $BVBer(\pi, \alpha)$ distribution

Assume that (U_1, U_2) has the $BVBer(\pi, \alpha)$ distribution of (1). In this case, the *jpmf* of Z_1 and Z_2 is given as in (63)-(66) after replacing $f(\cdot,\cdot)$ by $p(\cdot,\cdot;\underline{\lambda})$ of (5). The marginal *pmf's* of Z_1 and Z_2 are given as in (67) and (68) after replacing $f_i(\cdot)$ by the *pdf* of $Poi(\lambda_i + \lambda_3), i = 1, 2$. In addition,

$$\begin{aligned} E(Z_1) &= (2\pi - 1)\,(\lambda_1 + \lambda_3)\,, \\ E(Z_2) &= (2\alpha - 1)(2\pi - 1)\,(\lambda_2 + \lambda_3)\,, \\ Var(Z_1) &= (\lambda_1 + \lambda_3) + 4\pi\overline{\pi}\,(\lambda_1 + \lambda_3)^2\,, \end{aligned}$$

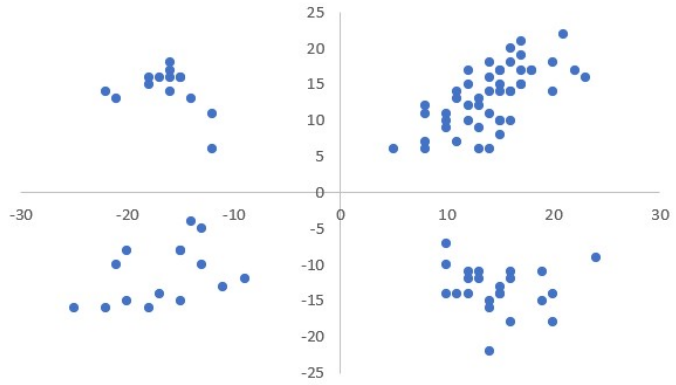


Figure 1: *BRST* of *BPD* with $\underline{\beta} = (0.4, 0.22, 0.17), \underline{\lambda} = (9, 7, 2), \rho = 0.137$, and $n = 100$.

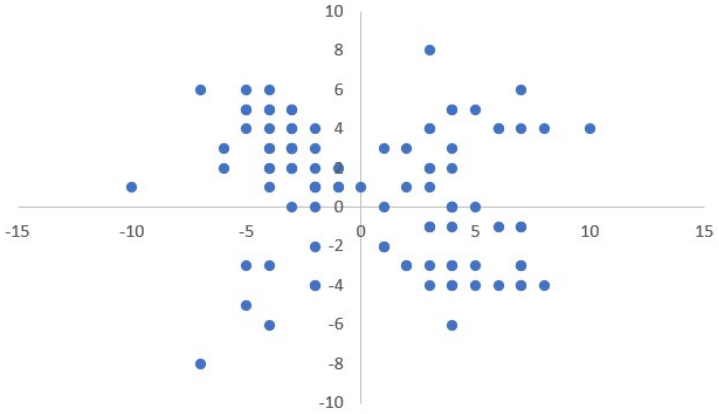


Figure 2: *BRST* of *BPD* with $\underline{\beta} = (0.2, 0.3, 0.35)$, $\underline{\lambda} = (3, 2, 1), \rho = -0.22$, and $n = 100$.

$$Var(Z_2) = (\lambda_2+\lambda_3)+2\,(\overline{\alpha}\pi+\alpha\overline{\pi})\,(\lambda_2+\lambda_3)^2,$$

and

$$Cov(Z_1,Z_2) = \lambda_3+4\pi\overline{\pi}(2\alpha-1)\,(\lambda_1+\lambda_3)\,(\lambda_2+\lambda_3).$$

For the estimation of π and α, we may use the following alternatives:

1. The *MLE* estimators obtained as in Lemma 6 using the R function "constrOptim".
2. The *MME* of (72)-(73).

In each of Figures 3 and 4, we give the scatter plot of a random sample of size $n = 100$ from the *BRST* of the *BPD* for selected values of π, α and $\underline{\lambda}$.

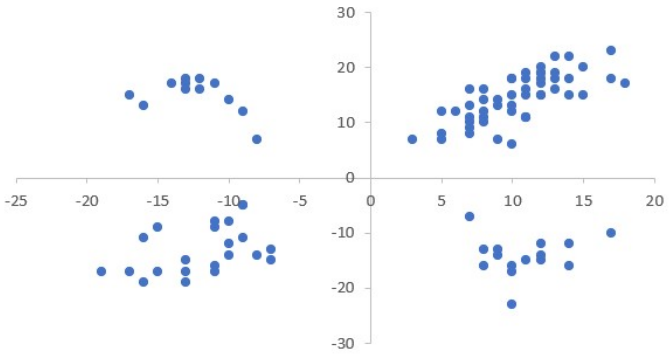


Figure 3: *BRST* of *BPD* with $\pi = 0.6, \alpha = 0.7, \underline{\lambda} = (5,8,6), \rho = 0.379$, and $n = 100$.

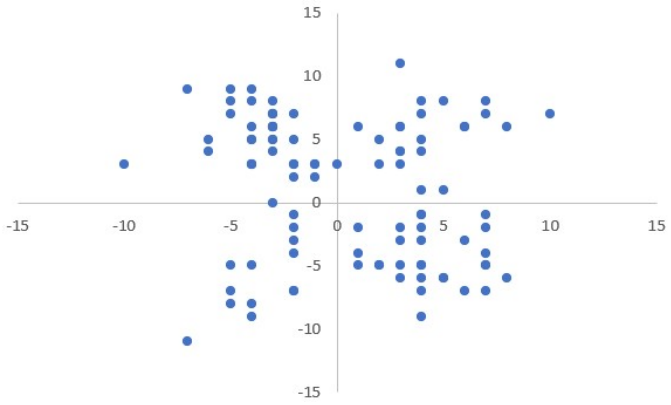


Figure 4: *BRST* of *BPD* with $\pi = 0.5, \alpha = 0.4, \underline{\lambda} = (3,4,1), \rho = -0.178$, and $n = 100$.

4.3 Models based on the $BVBer(\beta)$ distribution

Assume that (U_1,U_2) has the $BVBer(\beta)$ distribution of (4). In this case, the $jpmf$ of Z_1 and Z_2 is as given in (74)-(77) after replacing $f(\cdot,\cdot)$ by $p(\cdot,\cdot;\underline{\lambda})$ of (5). The marginal $pmf's$ of Z_1 and Z_2 are given as in (78) after replacing $f_i(\cdot)$ by the pdf of $Poi(\lambda_i+\lambda_3), i=1,2$. In addition,

$$E(Z_1)=E(Z_2)=0,$$

$$Var(Z_i)=(\lambda_i+\lambda_3)+(\lambda_i+\lambda_3)^2, \qquad i=1,2,$$

and

$$Cov(Z_1,Z_2)=(2\beta-1)\left(\lambda_3+(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)\right).$$

For the estimation of β, we may use the following alternatives:

1. The *MLE* estimator of (83).
2. The *MME* of (84).

In each of Figures 5 and 6, we give the scatter plot of a random sample of size $n=100$ from the *BRST* of the *BPD* for selected values of β and $\underline{\lambda}$.

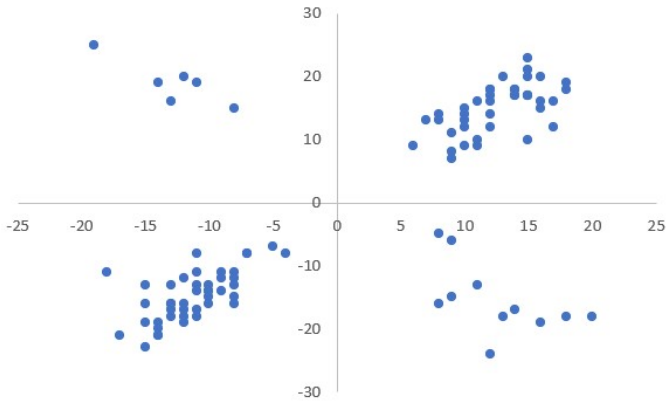


Figure 5: *BRST* of *BPD* with $\beta=0.75, \underline{\lambda}=(4,7,8), \rho=0.594$, and $n=100$.

5 BRST of the BGD

Assume that (U_1,U_2) has a *BVB* distribution, that (X_1,X_2) has the *BGD* with the $jpmf$ of (9) and that (X_1,X_2) is independent of (U_1,U_2). In the three models of this section, we may estimate δ_1 and δ_2 using the following alternatives:

1. The *MLE* estimators obtained as in Lemma 3 using (10).
2. The *MME* of (11) expressed in terms of $(|Z_{1i}|,|Z_{2i}|), i=1,2,\ldots,n$.

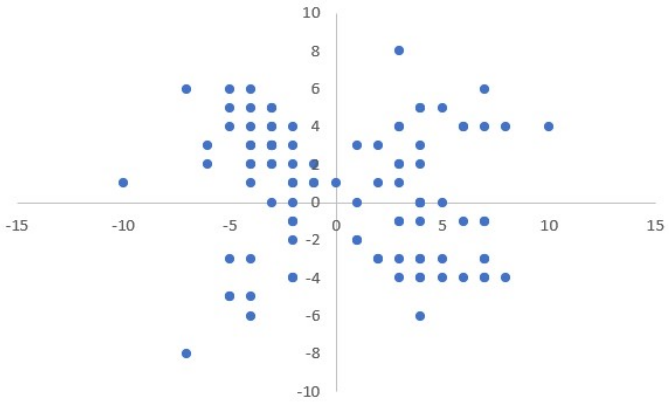


Figure 6: *BRST* of *BPD* with $\beta = 0.4, \underline{\lambda} = (3,2,1), \rho = -0.144$, and $n = 100$.

5.1 Models based on the $BVBer(\underline{\beta})$ distribution

Assume that (U_1, U_2) has the $BVBer(\underline{\beta})$ distribution of (1). In this case, the *jpmf* of Z_1 and Z_2 is given as in (13)-(16) after replacing $f(\cdot,\cdot)$ by the *jpmf* of (9). The marginal *pmf's* of Z_1 and Z_2 are given as in (13) and (16) after replacing $f_1(\cdot)$ by the *pdf* of $Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_2}\right)$ and $f_2(\cdot)$ by the *pdf* of $Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_1}\right)$. We can show that

$$E(Z_i) = (2\beta_{11} + 2I(i=1)\beta_{10} + 2I(i=2)\beta_{01} - 1)\frac{\delta_i}{\delta_3},$$

$$\begin{aligned} Var(Z_i) =& \left(\frac{\delta_i}{1-\delta_1-\delta_2}\right)\left(1+\frac{\delta_i}{1-\delta_1-\delta_2}\right) + 4\left(\beta_{11} + I(i=1)\beta_{10} + I(i=2)\beta_{01}\right) \\ & \times \left(1 - (\beta_{11} + I(i=1)\beta_{10} + I(i=2)\beta_{01})\right)\left(\frac{\delta_i}{1-\delta_1-\delta_2}\right)^2, \end{aligned}$$

and

$$Cov(Z_1, Z_2) = \frac{\delta_1\delta_2}{(1-\delta_1-\delta_2)^2}\left\{(1-2\beta_{10}-2\beta_{01}) + 4Cov(U_1,U_2)\right\},$$

where

$$Cov(U_1,U_2) = \beta_{11} - (\beta_{11}+\beta_{10})(\beta_{11}+\beta_{01}).$$

For the estimation of β_{11}, β_{10}, and β_{01}, we may use the following alternatives:

1. The *MLE* estimators obtained as in Lemma 3 using the R function "constrOptim".
2. The *MME* of (51)-(53).

In each of Figures 7 and 8, we give the scatter plot of a random sample of size $n = 100$ from the *BRST* of the *BGD* for selected values of $\underline{\beta}$ and $\underline{\theta}$.

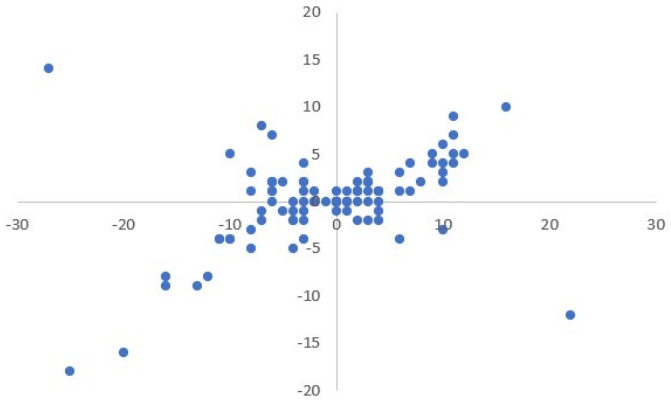


Figure 7: *BRST* of *BGD* with $\underline{\theta} = (0.6, 0.3, 0.1), \underline{\beta} = (0.4, 0.22, 0.17), \rho = 0.3993$, and $n = 100$.

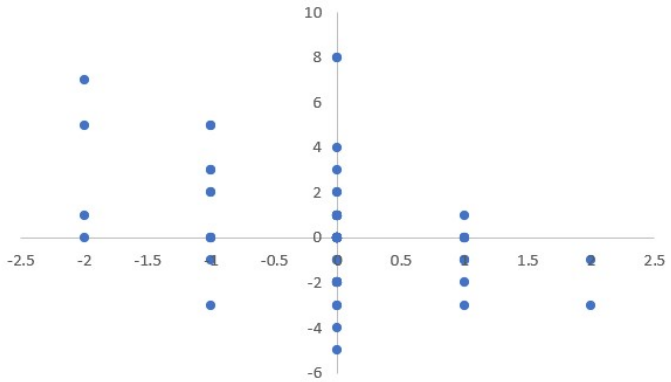


Figure 8: *BRST* of *BGD* with $\underline{\theta} = (0.2, 0.48, 0.32), \underline{\beta} = (0.2, 0.3, 0.35), \rho = -0.387$, and $n = 100$.

5.2 Models based on the $BVBer(\pi,\alpha)$ distribution

Assume that (U_1,U_2) has the $BVBer(\pi,\alpha)$ distribution of (1). In this case, the $jpmf$ of Z_1 and Z_2 is given as in (63)-(66) after replacing $f(\cdot,\cdot)$ by the $jpmf$ of (9). The marginal $pmf's$ of Z_1 and Z_2 are given as in (67) and (68) after replacing $f_1(\cdot)$ by the pdf of $Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_2}\right)$ and $f_2(\cdot)$ by the pdf of $Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_1}\right)$. We can show that

$$E(Z_1)=(2\pi-1)\frac{\delta_1}{1-\delta_1-\delta_2},$$

$$E(Z_2)=(2\alpha-1)(2\pi-1)\frac{\delta_2}{1-\delta_1-\delta_2},$$

$$Var(Z_1)=\left(\frac{\delta_1}{1-\delta_1-\delta_2}\right)\left(1+\frac{\delta_1}{1-\delta_1-\delta_2}\right)+4\pi\overline{\pi}\left(\frac{\delta_1}{1-\delta_1-\delta_2}\right)^2,$$

$$Var(Z_2)=\left(\frac{\delta_2}{1-\delta_1-\delta_2}\right)\left(1+\frac{\delta_2}{1-\delta_1-\delta_2}\right)+2\left(\overline{\alpha}\pi+\alpha\overline{\pi}\right)\left(\frac{\delta_2}{1-\delta_1-\delta_2}\right)^2,$$

and

$$Cov(Z_1,Z_2)=\left\{\frac{\delta_1\delta_2}{(1-\delta_1-\delta_2)^2}+4\pi\overline{\pi}(2\alpha-1)\left(\frac{\delta_1\delta_2}{(1-\delta_1-\delta_2)^2}\right)\right\}.$$

For the estimation of π and α, we may use the following alternatives:

1. The MLE estimators obtained as in Lemma 6 using the R function "constrOptim".
2. The MME of (72)-(73).

In each of Figures 9 and 10, we give the scatter plot of a random sample of size $n=100$ from the $BRST$ of the BGD for selected values of α,π and $\underline{\lambda}$.

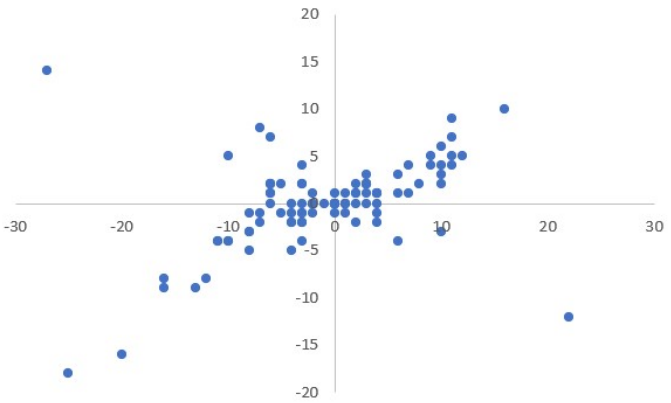


Figure 9: $BRST$ of BGD with $\underline{\theta}=(0.6,0.3,0.1),\pi=0.6,\alpha=0.7,\rho=0.4917$, and $n=100$.

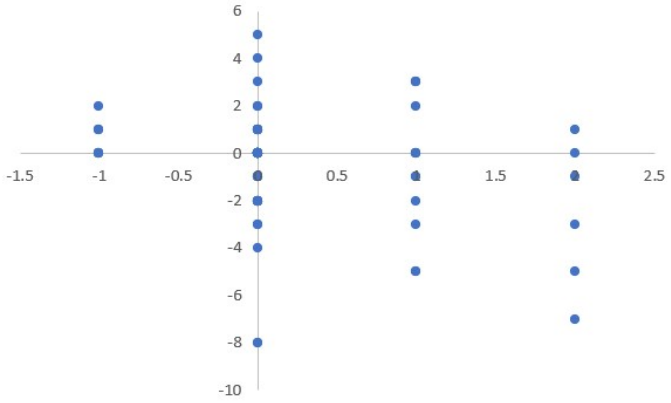


Figure 10: *BRST* of *BGD* with $\underline{\theta} = (0.2, 0.48, 0.32), \pi = 0.7, \alpha = 0.5, \rho = -0.24661$, and $n = 100$.

5.3 Models based on the $BVBer(\beta)$ distribution

Assume that (U_1, U_2) has the $BVBer(\beta)$ distribution of (4). In this case, the $jpmf$ of Z_1 and Z_2 is as given in (74)-(77) after replacing $f(\cdot,\cdot)$ by the $jpmf$ of (9). The marginal $pmf's$ of Z_1 and Z_2 are given as in (78) after replacing $f_1(\cdot)$ by the pdf of $Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_2}\right)$ and $f_2(\cdot)$ by the pdf of $Geo\left(\frac{1-\delta_1-\delta_2}{1-\delta_1}\right)$. We can show that

$$E(Z_i) = 0, \qquad i = 1, 2,$$

$$Var(Z_i) = \frac{\delta_i}{1-\delta_1-\delta_2}\left(1 + \frac{2\delta_i}{1-\delta_1-\delta_2}\right), \qquad i = 1, 2,$$

and

$$Cov(Z_1, Z_2) = \frac{2(2\beta-1)\delta_1\delta_2}{(1-\delta_1-\delta_2)^2}.$$

For the estimation of β, we may use the following alternatives:

1. The *MLE* estimator of (83).
2. The *MME* of (84).

In each of Figures 11 and 12, we give the scatter plot of a random sample of size $n = 100$ from the *BRST* of the *BGD* for selected values of β and $\underline{\theta}$.

6 Simulations

We have conducted 12 simulation studies to asses the performance of the *MLE* and the *MME* estimators of the model parameters. In each simulation, we used 10,000 realizations of samples of

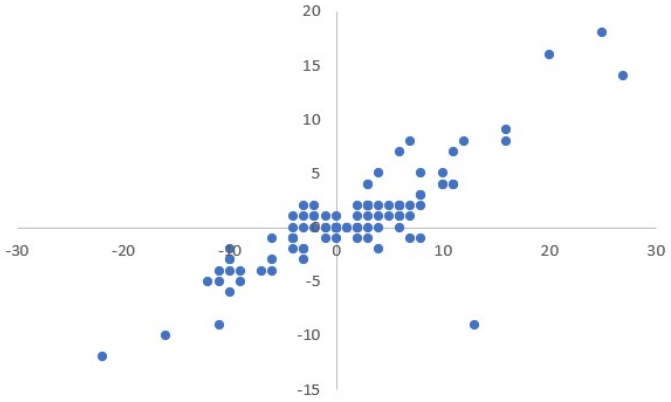


Figure 11: *BRST* of *BGD* with $\underline{\theta} = (0.6, 0.3, 0.1), \beta = 0.8, \rho = 0.837$, and $n = 100$.

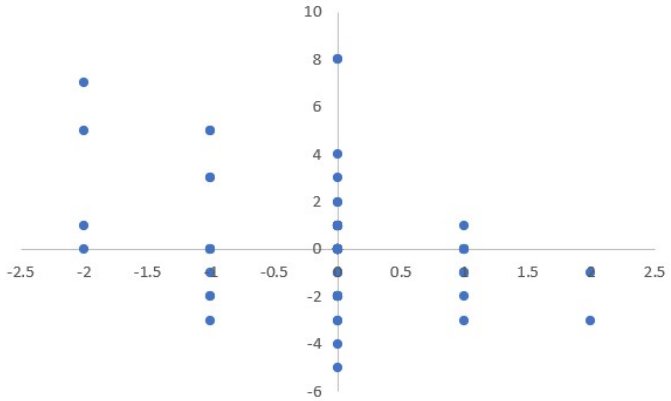


Figure 12: *BRST* of *BGD* with $\underline{\theta} = (0.2, 0.48, 0.32), \beta = 0.4, \rho = -0.3335$, and $n = 100$.

Table 1: BRST of BPD using $BVBer(\underline{\beta})$.

	$\lambda_1 = 3$	$\lambda_2 = 2$	$\lambda_3 = 1$
MLE	2.998(0.199)	1.996(0.192)	1.002(0.182)
MME	3.005(0.228)	2.003(0.222)	0.995(0.214)
	$\beta_{11} = 0.2$	$\beta_{10} = 0.3$	$\beta_{01} = 0.35$
MLE	0.205(0.016)	0.283(0.021)	0.357(0.017)
MME	0.2(0.028)	0.306(0.033)	0.356(0.033)

Table 2: BRST of BPD using $BVBer(\pi, \alpha)$.

	$\lambda_1 = 3$	$\lambda_2 = 2$	$\lambda_3 = 1$	$\pi = 0.7$	$\alpha = 0.5$
MLE	2.998(0.199)	1.996(0.192)	1.002(0.182)	0.685(0.023)	0.5(0.021)
MME	3.005(0.228)	2.003(0.222)	0.995(0.214)	0.7(0.029)	0.501(0.086)

Table 3: BRST of BPD using $BVBer(\beta)$.

	$\lambda_1=3$	$\lambda_2=2$	$\lambda_3=1$	$\beta=0.4$
MLE	2.998(0.199)	1.996(0.192)	1.002(0.182)	0.4(0.031)
MME	3.005(0.228)	2.003(0.222)	0.995(0.214)	0.4(0.041)

Table 4: BRST of BGD using $BVBer(\beta)$.

	$p_0=0.2$	$p_1=0.48$	$p_2=0.32$
MLE	0.2(0.013)	0.48(0.016)	0.321(0.015)
MME	0.2(0.013)	0.48(0.016)	0.321(0.015)
	$\beta_{11}=0.2$	$\beta_{10}=0.3$	$\beta_{01}=0.35$
MLE	0.2(0.0111)	0.277(0.024)	0.35(0.011)
MME	0.201(0.056)	0.299(0.056)	0.349(0.059)

Table 5: BRST of BGD using $BVBer(\pi,\alpha)$.

	$p_0=0.2$	$p_1=0.48$	$p_2=0.32$	$\pi=0.7$	$\alpha=0.5$
MLE	0.2(0.013)	0.48(0.016)	0.321(0.015)	0.675(0.037)	0.502(0.026)
MME	0.2(0.013)	0.48(0.016)	0.321(0.015)	0.699(0.05)	0.5(0.018)

Table 6: BRST of BGD using $BVBer(\beta)$.

	$p_0=0.2$	$p_1=0.48$	$p_2=0.32$	$\beta=0.4$
MLE	0.2(0.013)	0.48(0.016)	0.321(0.015)	0.4(0.051)
MME	0.2(0.013)	0.48(0.016)	0.321(0.015)	0.4(0.086)

Table 7: BRST of independent Poisson RV using $BVBer(\beta)$.

	$\lambda_1=3$	$\lambda_2=2$	$\beta_{11}=0.2$	$\beta_{10}=0.3$	$\beta_{01}=0.35$
MLE	2.999(0.099)	2(0.082)	0.201(0.008)	0.286(0.016)	0.351(0.008)
MME	2.999(0.099)	2.(0.082)	0.201(0.03)	0.3(0.033)	0.349(0.034)

Table 8: BRST of independent Poisson RV using $BVBer(\pi,\alpha)$.

	$\lambda_1=3$	$\lambda_2=2$	$\pi=0.7$	$\alpha=0.5$
MLE	2.999(0.099)	2(0.082)	0.684(0.025)	0.502(0.017)
MME	2.999(0.099)	2.(0.082)	0.7(0.031)	0.502(0.092)

Table 9: BRST of independent Poisson RV using $BVBer(\beta)$.

	$\lambda_1=3$	$\lambda_2=2$	$\beta=0.4$
MLE	2.999(0.099)	2(0.082)	0.4(0.031)
MME	2.999(0.099)	2.(0.082)	0.4(0.040)

Table 10: BRST of independent geometric RV using $BVBer(\underline{\beta})$.

	$\theta_1 = 0.5$	$\theta_2 = 0.6$	$\beta_{11} = 0.2$	$\beta_{10} = 0.3$	$\beta_{01} = 0.35$
MLE	0.501(0.02)	0.601(0.022)	0.199(0.003)	0.284(0.022)	0.35(0.005)
MME	0.501(0.02)	0.601(0.022)	0.236(0.064)	0.263(0.066)	0.32(0.067)

Table 11: BRST of independent geometric RV using $BVBer(\pi, \alpha)$.

	$\theta_1 = 0.5$	$\theta_2 = 0.6$	$\pi = 0.7$	$\alpha = 0.5$
MLE	0.501(0.02)	0.601(0.022)	0.683(0.033)	0.504(0.059)
MME	0.501(0.02)	0.601(0.022)	0.701(0.046)	0.5(0.01)

size 300 from the considered model. The mean (the standard deviation) of the 10,000 estimators of each parameter are reported next in Tables 1 to 12.

The results of the Tables 1-12 suggest that both the *MLE* and the *MME* estimators perform well in all the considered models. However, for most of the models, the *MLE* estimators have smaller standard deviations than the corresponding *MME* estimators for all parameters.

7 Data analysis

The data of this example are based on the results of the 2019 UEFA Europa League. The 48 teams of this competition are divided into 12 groups of four teams each. Each team plays one home match and one away match against the other three teams of its group. For each team, we obtained one observation computed by taking the difference between a) the sum of scores of its three home matches and b) the sum of scores of its three away matches. For example, the observation of team Apoel of Group A (Apoel, Dudelange, Qarabağ and Sevilla) is obtained as follows. The sum of Apoel's three home scores $((3,4),(2,1)$, and $(1,0))$ is $(6,5)$ and the sum of Apoel's three away scores $((2,0),(2,2)$, and $(0,1))$ is $(4,3)$. Hence, the difference $((6,5)-(4,3))$ is $(2,2)$. The resulting bivariate data of 48 observations is as follows:

(0,-2)	(0,-2)	(0,4)	(0,-4)	(0,5)	(0,7)	(-1,0)	(1,1)	(1,-1)	(-1,-1)	(1,3)	(-1,-3)
(-1,-3)	(1,-4)	(1,-4)	(1,-5)	(-1,-6)	(2,0)	(2,-1)	(-2,1)	(2,2)	(2,-3)	(-2,-3)	(-2,-3)
(2,-4)	(-2,4)	(-2,-4)	(-3,-1)	(3,2)	(3,3)	(3,-3)	(4,-1)	(4,-2)	(4,-2)	(4,3)	(4,-4)
(-4,-5)	(5,-2)	(-5,2)	(6,-2)	(6,-2)	(6,4)	(2,-3)	(3,-5)	(-1,3)	(2,-2)	(3,-5)	(7,0).

To fit the above data we will explore the following bivariate BRST models:

1. P1 (BPD based on $BBer(\underline{\beta})$), P2 (BPD based on $BBer(\pi, \alpha)$) and P3 (BPD based on $BBer(\beta)$).

Table 12: BRST of independent geometric RV using $BVBer(\beta)$.

	$\theta_1 = 0.5$	$\theta_2 = 0.6$	$\beta = 0.4$
MLE	0.501(0.02)	0.601(0.022)	0.4(0.063)
MME	0.501(0.02)	0.601(0.022)	0.4(0.092)

2. P4 (independent Poisson RV based on $BBer(\underline{\beta})$), P5 (independent Poisson RV based on $BBer(\pi,\alpha)$) and P6 (independent Poisson RV based on $BBer(\beta)$).

3. P7 (BGD based on $BBer(\underline{\beta})$), P8 (BGD based on $BBer(\pi,\alpha)$) and P9 (BGD based on $BBer(\beta)$).

4. P10 (independent Geometric RV based on $BBer(\underline{\beta})$), P11 (independent Geometric RV based on $BBer(\pi,\alpha)$) and P12 (independent Geometric RV based on $BBer(\beta)$).

In all of the above models, we obtained the MLE. For models P1, P4, and P10, $\widehat{\beta}_{11} = 0.19800, \widehat{\beta}_{10} = 0.46751$, and $\widehat{\beta}_{01} = 0.113099$. For models P2, P5, and P11, $\widehat{\pi} = 0.659675$ and $\widehat{\alpha} = 0.418447$. For models P3, P6, and P12, $\widehat{\beta} = 0.421$. For models P1-P3, $\widehat{\lambda}_1 = 1.634857, \widehat{\lambda}_2 = 2.572357$, and $\widehat{\lambda}_3 = 0.7193102$. For models P4-P6, $\widehat{\lambda}_1 = 2.35$ and $\widehat{\lambda}_2 = 2.83$. For models P7-P9, $\widehat{\theta}_1 = 0.38, \widehat{\theta}_2 = 0.457$, and $\widehat{\theta}_3 = 0.162$. For models P10-P12, $\widehat{\theta}_1 = 0.298$ and $\widehat{\theta}_2 = 0.261$.

We divided Z^2 into the nine mutually exclusive areas corresponding to $n_0, n_{0,\mp}, n_{\mp,0}, n_{\pm,-}$, and $n_{\mp,+}$ of (32)-(36). The expected counts of each of these areas are computed using (37)-(45). For example, the computations for model P_1 are given next. Note that in this case,

$$f(0,0;\underline{\widehat{\lambda}}) = 0.0072517, \quad f_1(0;\underline{\widehat{\lambda}}) = 0.094973, \text{ and } f_2(0;\underline{\widehat{\lambda}}) = 0.037192. \tag{85}$$

Hence, by (37)-(45) and (85), we obtain the following Table 13 for Model P1.

Table 13: Observed and expected counts for Model P1.

P1	n_0	$n_{+,0}$	$n_{-,0}$	$n_{0,+}$	$n_{0,-}$	$n_{+,+}$	$n_{-,+}$	$n_{+,-}$	$n_{-,-}$
Observed	0	2	1	3	4	7	4	18	9
Expected	0.348	0.946	0.492	1.31	2.901	8.326	4.742	19.31	9.6266

For each of the 12 models, we computed the Chi-square test statistic and the corresponding P-value. The P-values for models P2, P3, and P6-P12 are all less than 0.05. The Chi-square test statistics and the corresponding P-values of the remaining models are as follows: P1 $(5.104, 0.078)$, P4 $(3.0386, 0.386)$, and P5 $(7.8953, 0.096)$. It is clear from these results that model P4 provides the best fit for the considered data.

8 Conclusions

We extended the RST of Aly (2018) to produce bivariate integer-valued random vectors on Z^2. The proposed Bivariate RST (BRST) is also an extension of the family of bivariate discrete distributions on Z^2 of Chesneau et al. (2018). We studied in details our proposed family and considered, in particular, a number of new bivariate integer-valued distributions on Z^2. The proposed BRST can be applied to other bivariate nonnegative integer-valued random vectors to produce new families of bivariate integer-valued random vectors on Z^2.

As an illustration, we applied the proposed families to a real data set developed based on the results of the 2019 UEFA Europa League. One of our proposed models provided an excellent fit of this data.

Acknowledgments

The authors wish to thank the two referees for their valuable comments and suggestions which resulted in improving the presentation of this paper. This paper is based on the first author MSc thesis at Kuwait University.

References

Aly, E.-E. (2018). A unified approach for developing Laplace-type distributions. *Journal of the Indian Society for Probability and Statistics*, **19**, 245–269.

Bulla, J., Chesneau, C. and Kachour, M. (2015). On the bivariate Skellam distribution. *Communications in Statistics-Theory and Methods*, **44**, 4552–4567.

Chesneau, C., Kachour, M. and Bakouch, H. S. (2018). A family of bivariate discrete distributions on Z^2 based on the Rademacher distribution. *In ProbStat Forum*, **11**, 53–66.

Genest, C. and Mesfioui, M. (2014). Bivariate extensions of Skellam's distribution. *Probability in the Engineering and Informational Sciences*, **28**, 401–417.

Hogg, R. V., McKean, J. W. and Craig, A. T. (2005). *Introduction to Mathematical Statistics, 6th ed.* Pearson Education, New Delhi.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete multivariate distributions.* Wiley Series in Probability and Statistics, **165**, Wiley, New York.

Karlis, D. and Mamode Khan, N. (2023). Models for integer data. *Annual Review of Statistics and Its Application*, **10**, 297–323.

Karlis, D. and Ntzoufras, I. (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software*, **14** (i10).

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions.* Statistics: textbooks and monographs, **132**, Marcel Dekker, New York.

Krishna, H. and Pundir, P. S. (2009). A bivariate geometric distribution with applications to reliability. *Communications in Statistics–Theory and Methods*, **38**, 1079–1093.

Lai, C.D. (2006). *Constructions of Discrete Bivariate Distributions.* In Advances in distribution theory, order statistics, and inference, 29–58, Birkhäuser.

Odhah, O.H. (2013). Probability Distributions on the Class of Integer Numbers. Doctoral dissertation, King Saud University.

Omair M.A., Alomani, G. A. and Alzaid A. (2022). Bivariate distributions on Z^2. *Bulletin of the Malaysian Mathematical Sciences Society*, **45**, 425–444.

Omair, M, A, Alzaid A., Odhah, O. H. (2016). A trinomial difference distribution. *Revista Colombiana de Estadística,* **39**, 1–15.

Phatak, A. G. and Sreehari, M. (1981). Some characterizations of a bivariate geometric distribution. *Journal of Indian Statistical Association*, **19**, 141–146.

Sarabia Alegría, J. M. and Gómez Déniz, E. (2008). Construction of multivariate distributions: a review of some recent results. SORT, **32**, 3–36.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, **30**, 50–64.

Stochastic Models in Probability and Statistics
Vol. 2, No. 1, 2025, pp. 31–39. Research Article

SMPS

On shrinkage estimation under divergence loss

Mohammad Arashi[†*], Hidekazu Tanaka[‡], Morteza Amini[§]

[†] *Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran*
[‡] *Graduate School of Science, Osaka Prefecture University, Sakai, Japan*
[§] *Department of Statistics, School of Mathematics, Statistics, and Computer Science, College of Science, University of Tehran, Tehran, Iran*

Email(s): arashi@um.ac.ir, tanaka@ms.osakafu-u.ac.jp, morteza.amini@ut.ac.ir

Abstract. In this paper, the superiority conditions for a general class of shrinkage estimators in the estimation problem of the normal mean are established under divergence loss. This approach is an extension of the work of Ghosh and Mergel (2009).

Keywords: Divergence loss; James-Stein estimate; Shrinkage estimate; Superharmonic function.

1 Introduction

In the line of seminal works of Stein (1956) and James and Stein (1961), there is growing interest in modifying and generalizing the latter work to bring a new estimator of shrinkage type in order to outperform the sample mean. Interesting studies may include in the couple of works done by Baranchik (1970), Efron and Morris (1973), Strawderman (1971), Faith (1978), Stein (1981), Brandwein and Strawderman (1980), Casella (1990), George (1991), Shao et al. (1994), Maruyama (2004), Srivastava and Kubokawa (2005), Ghosh et al. (2008), Wells and Zhou (2008), Ghosh and Mergel (2009) and Arashi and Tabatabey (2010) under different settings.

This work is arisen from the recent study due to Ghosh and Mergel (2009). They considerably investigated on the superiority conditions of Baranchik-type estimators over the sample mean in multivariate normal model, with divergence loss. In this paper, a minor extension is then carried out for another class of shrinkage estimators.

For the precise setup, first of all suppose that $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 I_p)$, where $\boldsymbol{\theta}$ and σ^2 are both unknown. Further, $S \sim (\sigma^2/(m+2))\chi^2_m$ is independent of $\boldsymbol{X}$. The aim of this work is to establish

*Corresponding author
Received: 01 November 2024 / Accepted: 15 March 2025
DOI: 10.22067/smps.2025.90474.1034

conditions in which the general class of shrinkage estimators given by

$$\boldsymbol{\delta}(\boldsymbol{X}) = \boldsymbol{X} + S\boldsymbol{g}(\boldsymbol{X}) \tag{1}$$

outperforms $\boldsymbol{X}$, where $\boldsymbol{g}: \mathbb{R}^p \to \mathbb{R}^p$ satisfies some regularity conditions which will be given later.

The outline of this paper is as follows: In Section 2, some preliminary results as well as some notations are given, while the main results are exhibited in Section 3. Some important remarks are also given in Section 4.

2 Preliminaries

Before revealing the main results, we express some useful notations. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$, let

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{j=1}^{p} x_j y_j, \quad \|\boldsymbol{x}\|^2 = \sum_{i=1}^{p} x_i^2.$$

Definition 1. *A function $h: \mathbb{R}^p \to \mathbb{R}$ is said to be almost differentiable if there exists a function $\blacktriangledown h: \mathbb{R}^p \to \mathbb{R}^p$ such that, for all $\boldsymbol{z} \in \mathbb{R}^p$,*

$$h(\boldsymbol{x}+\boldsymbol{z}) - h(\boldsymbol{x}) = \int_0^1 \boldsymbol{z} \cdot \blacktriangledown h(\boldsymbol{x}+t\boldsymbol{z}) dt$$

for (Lebesgue measure) almost all $\boldsymbol{x} \in \mathbb{R}^p$. A function $\boldsymbol{g}: \mathbb{R}^p \to \mathbb{R}^p$ is almost differentiable if all its coordinate functions are almost differentiable. Essentially, $\blacktriangledown$ is the vector differential operator of first partial derivatives with i^{th} coordinate $\triangledown_i = \partial/\partial x_i$.

Definition 2. *A function $\boldsymbol{g}: \mathbb{R}^p \to \mathbb{R}^p$ is said to be homogeneous of degree -1, if it satisfies*

$$\boldsymbol{g}(\lambda \boldsymbol{x}) = \frac{1}{\lambda} \boldsymbol{g}(\boldsymbol{x})$$

for all real $\lambda \neq 0$ and for all $\boldsymbol{x} \in \mathbb{R}^p$.

As an example satisfying the regularity condition in Definition 2, consider the reciprocal function $\boldsymbol{g}(\boldsymbol{x}) = (g(x_1), \ldots, g(x_p))'$ where $g(x) = x^{-1}$. Obviously, $g(\lambda x) = 1/(\lambda x) = \lambda^{-1} g(x)$.

In this paper, we employ the expectation of divergence loss, which includes Kullback-Leibler (KL for short) loss and Bhattacharyya-Hellinger (BH) loss, as a measurement. The divergence loss has been considered by many authors in other contexts. Among others, we refer to Amari (1982) and Cressie and Read (1984).

Definition 3. *Suppose that $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\theta}, \sigma^2 I_p)$ denotes the probability density function of $\mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 I_p)$. For an estimator $\boldsymbol{a}$ of $\boldsymbol{\theta}$, the divergence loss is defined by*

$$\begin{aligned} L_\beta(\boldsymbol{\theta}; \boldsymbol{a}) &= \frac{1}{\beta(1-\beta)} \left(1 - \int_{\mathbb{R}^p} \mathcal{N}^{1-\beta}(\boldsymbol{x}|\boldsymbol{\theta}, \sigma^2 I_p) \mathcal{N}^{\beta}(\boldsymbol{x}|\boldsymbol{a}, \sigma^2 I_p) d\boldsymbol{x}\right) \\ &= \frac{1}{b}\left(1 - \exp\left[-\frac{b}{2\sigma^2}\|\boldsymbol{a}-\boldsymbol{\theta}\|^2\right]\right), \end{aligned} \tag{2}$$

where $b = \beta(1-\beta)$ and $\beta \in (0,1)$.

The second equality in (2) is a consequence of Lemma 2.2 of Ghosh et al. (2008). The KL loss and the BH loss occur as special cases of the divergence loss when $\beta \to 0$ or $\beta \to 1$, and $\beta = 1/2$, respectively. Some graphical displays are presented in Figure 1, illustrating the relative behavior of the loss function for $\sigma = 0.5$. In this figure, β varies within the range $(0,1)$, while $\|\boldsymbol{a}-\boldsymbol{\theta}\|^2$ changes from small to large values across different panes to demonstrate the effect of its magnitude. As shown, for moderate values of $\|\boldsymbol{a}-\boldsymbol{\theta}\|^2$, the shape of the loss function is clearly bath-shaped.

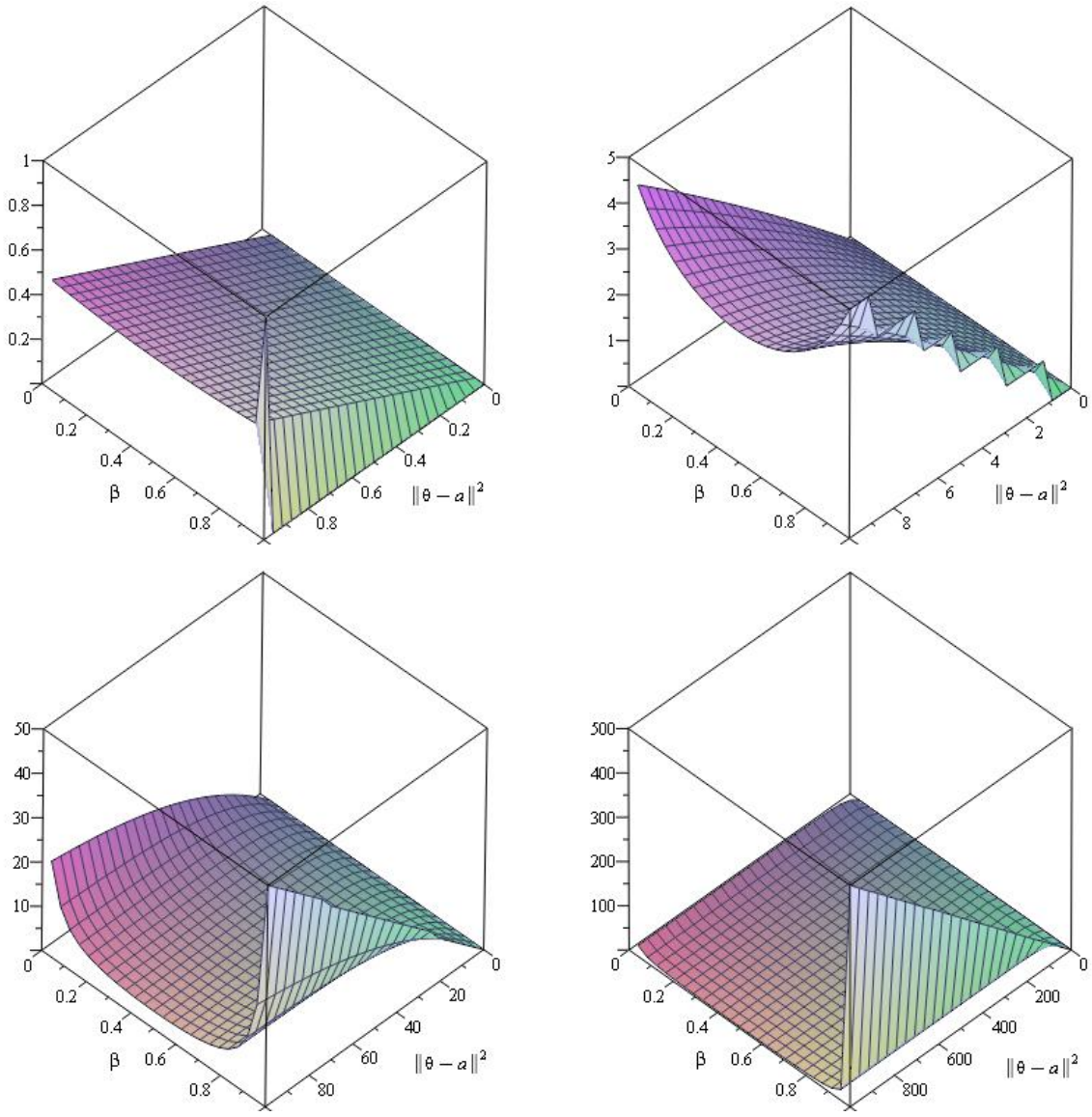


Figure 1: Behavior of the divergence loss relative to β and $\|\boldsymbol{a}-\boldsymbol{\theta}\|^2$.

To close this section, note that since $\|\boldsymbol{X}-\boldsymbol{\theta}\|^2 \sim \sigma^2\chi^2_p$, under the loss, it can be directly concluded that the risk of $\boldsymbol{X}$ is given by

$$R_\beta(\boldsymbol{\theta};\boldsymbol{X}) = \frac{1-(1+b)^{-p/2}}{b}. \tag{3}$$

3 Main results

In this section, we first give sufficient conditions in which the estimator $\boldsymbol{\delta}(\boldsymbol{X})$ given by (1) dominates $\boldsymbol{X}$ for the case of $\Sigma = \sigma^2 I_p$ (Case I). Also, for the case of the general framework Σ (Case II), we give sufficient conditions so that an estimator dominates $\boldsymbol{X}$.

Case I: ($\Sigma = \sigma^2 I_p$, σ^2 is unknown)

Theorem 1. *Assume $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 I_p)$ and $S \sim (\sigma^2/(m+2))\chi^2_m$ are independent, where $\boldsymbol{\theta}$ and σ^2 are both unknown. Further, assume that for a given estimator $\boldsymbol{\delta}(\boldsymbol{X})$ in (1), $\boldsymbol{g}$ is almost differentiable and homogenous function of degree -1 satisfying $E|(\partial/\partial W_i)g_i(\boldsymbol{W})| < \infty$ for all $\boldsymbol{\eta} \in \mathbb{R}^p$ and for $i = 1, \cdots, p$, where $\boldsymbol{W} \sim \mathcal{N}_p(\boldsymbol{\eta}, I_p/(b+1))$ is independent of S. Then, the estimator $\boldsymbol{\delta}(\boldsymbol{X})$ has smaller risk than $\boldsymbol{X}$, under the divergence loss (2), provided that*

$$\|\boldsymbol{g}(\boldsymbol{w})\|^2 + \frac{2}{b+1}\blacktriangledown \cdot \boldsymbol{g}(\boldsymbol{w}) \le 0$$

for all $\boldsymbol{w} \in \mathbb{R}^p$.

Proof. Let $\boldsymbol{Y} = \boldsymbol{X}/\sigma$, $\boldsymbol{\eta} = \boldsymbol{\theta}/\sigma$, $S^* = S/\sigma^2$. Then, by the homogeneity of $\boldsymbol{g}$ and the inequality $e^x - e^y \ge e^y(x-y)$ $(x, y \in \mathbb{R})$, the risk difference between $\boldsymbol{X}$ and $\boldsymbol{\delta}(\boldsymbol{X})$ is given by

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{X}) - R_\beta(\boldsymbol{\theta};\boldsymbol{\delta}(\boldsymbol{X})) &= \frac{1}{b}E\left[\exp\left(-\frac{b}{2\sigma^2}\|\boldsymbol{X}+S\boldsymbol{g}(\boldsymbol{X})-\boldsymbol{\theta}\|^2\right) - \exp\left(-\frac{b}{2\sigma^2}\|\boldsymbol{X}-\boldsymbol{\theta}\|^2\right)\right] \\ &\ge -\frac{1}{2\sigma^2}E\left[\exp\left(-\frac{b}{2\sigma^2}\|\boldsymbol{X}-\boldsymbol{\theta}\|^2\right)\left(S^2\|\boldsymbol{g}(\boldsymbol{X})\|^2 + 2S(\boldsymbol{X}-\boldsymbol{\theta})\cdot\boldsymbol{g}(\boldsymbol{X})\right)\right] \\ &= -\frac{1}{2}E\left[\exp\left(-\frac{b}{2}\|\boldsymbol{Y}-\boldsymbol{\eta}\|^2\right)\left(S^{*2}\|\boldsymbol{g}(\boldsymbol{Y})\|^2 + 2S^*(\boldsymbol{Y}-\boldsymbol{\eta})\cdot\boldsymbol{g}(\boldsymbol{Y})\right)\right]. \end{aligned} \tag{4}$$

Since $\boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\eta}, I_p)$ and $S^* \sim (m+2)^{-1}\chi^2_m$ are independent, and $E(S^*) = E(S^{*2}) = m/(m+2)$, the right hand side in (4) is rewritten by

$$\begin{aligned} &-\frac{m}{2(m+2)}E\left[\exp\left(-\frac{b}{2}\|\boldsymbol{Y}-\boldsymbol{\eta}\|^2\right)\left(\|\boldsymbol{g}(\boldsymbol{Y})\|^2 + 2(\boldsymbol{Y}-\boldsymbol{\eta})\cdot\boldsymbol{g}(\boldsymbol{Y})\right)\right] \\ &\quad = -\frac{m}{2(m+2)}(b+1)^{-p/2}E\left[\|\boldsymbol{g}(\boldsymbol{W})\|^2 + 2\boldsymbol{g}(\boldsymbol{W})\cdot(\boldsymbol{W}-\boldsymbol{\eta})\right]. \end{aligned} \tag{5}$$

By the Stein identity, we have

$$E\left[\boldsymbol{g}(\boldsymbol{W})\cdot(\boldsymbol{W}-\boldsymbol{\eta})\right] = \frac{1}{b+1}E\left[\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{W})\right]. \tag{6}$$

Substituting (5) and (6) in (4) concludes that

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{X}) - R_\beta(\boldsymbol{\theta};\boldsymbol{\delta}(\boldsymbol{X})) &\geq -\frac{m}{2(m+2)}(b+1)^{-p/2}E\left[\|\boldsymbol{g}(\boldsymbol{W})\|^2 + \frac{2}{b+1}\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{W})\right] \\ &\geq 0 \end{aligned} \tag{7}$$

for all $\boldsymbol{\theta}\in\mathbb{R}^p$ and for all $\sigma^2>0$. Here, we have to show that the inequality in (7) holds strictly for some $\boldsymbol{\theta}$ and σ^2. Since $P\{S\boldsymbol{g}(\boldsymbol{X})+2(\boldsymbol{X}-\boldsymbol{\theta})=0\}=0$, the equality in (4) holds only if $\boldsymbol{g}(\boldsymbol{X})=0$ (a.s.), that is, $\boldsymbol{\delta}(\boldsymbol{X})=\boldsymbol{X}$ (a.s.). This completes the proof. □

As a supplement, the earlier result can be proposed in a more general situation. In this regard, we have the following essential consequence.

Theorem 2. *Suppose that $\boldsymbol{X}\sim\mathcal{N}_p(\boldsymbol{\theta},\sigma^2 I_p)$ and $S\sim(\sigma^2/(m+2))\chi^2_m$ are independent, where $\boldsymbol{\theta}$ and σ^2 are both unknown. Consider the class of shrinkage estimators*

$$\boldsymbol{\delta}^*(\boldsymbol{X}) = \boldsymbol{X} + cS\boldsymbol{g}(\boldsymbol{X}),$$

where $\boldsymbol{g}$ is as in Theorem 1. Then, $\boldsymbol{\delta}^*(\boldsymbol{X})$ outperforms $\boldsymbol{X}$ under divergence loss provided that

(i) there exists a function $h(.)$ such that, for all $\boldsymbol{x}\in\mathbb{R}^p$

$$\frac{1}{2}\|\boldsymbol{g}(\boldsymbol{x})\|^2 \leq h(\boldsymbol{x}) \leq -\frac{1}{1+b}\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{x}),$$

(ii) $0<c\leq 1$.

Proof. By the same way as the proof of Theorem 1, the risk difference between $\boldsymbol{X}$ and $\boldsymbol{\delta}^*(\boldsymbol{X})$ is given by

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{X}) - R_\beta(\boldsymbol{\theta};\boldsymbol{\delta}^*(\boldsymbol{X})) &\geq -\frac{m}{2(m+2)}(b+1)^{-p/2}E\left[c^2\|\boldsymbol{g}(\boldsymbol{W})\|^2 + \frac{2c}{b+1}\blacktriangledown\cdot\boldsymbol{g}(\boldsymbol{W})\right] \\ &\geq -\frac{m}{m+2}(b+1)^{-p/2}c(c-1)E\left[h(\boldsymbol{W})\right] \\ &\geq 0 \end{aligned} \tag{8}$$

for all $c\in(0,1]$. By the same reason as Theorem 1, the equality in (8) holds for all $\boldsymbol{\theta}$ and for all σ^2 only if $\boldsymbol{\delta}(\boldsymbol{X})=\boldsymbol{X}$ (a.s.). □

Now, let $\mathscr{S}(p)$ be the set of all positive definite matrices of order p.

Case II: (Unknown $\Sigma\in\mathscr{S}(p)$)

In this case, first of all, let $n-1$ mutually independent random vectors $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{n-1}\sim\mathcal{N}_p(\boldsymbol{0},\Sigma)$ be independent of $\boldsymbol{Z}_n\sim\mathcal{N}_p(\boldsymbol{\theta},\Sigma)$, in which $p\geq 3$. We deal with a more general type of improved shrinkage estimators, based on the sufficient statistic $(\boldsymbol{Z}_n,S)$, of the form

$$\boldsymbol{\gamma}(\boldsymbol{Z}_n,S) = \boldsymbol{Z}_n + \boldsymbol{F}(\boldsymbol{Z}_n,S), \tag{9}$$

where $S=\sum_{j=1}^{n-1}\boldsymbol{Z}_j\boldsymbol{Z}_j'$.

Definition 4. *Suppose that* $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\theta},\Sigma)$ *denotes the probability density function of* $\mathcal{N}_p(\boldsymbol{\theta},\Sigma)$. *For an estimator* $\boldsymbol{a}$ *of* $\boldsymbol{\theta}$, *the generalized divergence loss (GDL) function is defined by*

$$\begin{aligned} L_\beta^\Sigma(\boldsymbol{\theta};\boldsymbol{a}) &= \frac{1}{\beta(1-\beta)}\left(1-\int_{\mathbb{R}^p}\mathcal{N}^{1-\beta}(\boldsymbol{x}|\boldsymbol{\theta},n^{-1}\Sigma)\mathcal{N}^{\beta}(\boldsymbol{x}|\boldsymbol{a},n^{-1}\Sigma)d\boldsymbol{x}\right) \\ &= \frac{1}{b}\left(1-\exp\left[-\frac{nb}{2}\|\boldsymbol{a}-\boldsymbol{\theta}\|_\Sigma^2\right]\right), \end{aligned} \tag{10}$$

where $\|\boldsymbol{a}-\boldsymbol{\theta}\|_\Sigma^2=(\boldsymbol{a}-\boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{a}-\boldsymbol{\theta})$.

Theorem 3. *Suppose that*

$$E\left[\boldsymbol{F}'(\boldsymbol{U},S)\boldsymbol{F}(\boldsymbol{U},S)\right]<\infty$$

for all $\boldsymbol{\theta}\in\mathbb{R}^p$ *and for all* $\Sigma\in\mathscr{S}(p)$, *where* $\boldsymbol{U}\sim\mathcal{N}_p(\boldsymbol{\theta},\Sigma/(nb+1))$ *is independent of* S. *Then, the estimator* $\boldsymbol{\gamma}(\boldsymbol{Z}_n,S)$ *given by* (9) *has smaller risk than* $\boldsymbol{Z}_n$, *under GDL function given by* (10) *provided that*

$$2\blacktriangledown_{\boldsymbol{u}}\cdot\boldsymbol{F}(\boldsymbol{u},S)+(n-p+2)(nb+1)\boldsymbol{F}'(\boldsymbol{u},S)S^{-1}\boldsymbol{F}(\boldsymbol{u},S)\le 0,$$

where $\blacktriangledown_{\boldsymbol{u}}$ *states getting derivative with respect to* $\boldsymbol{u}$.

Proof. By the same way as the proof of Theorem 1, the risk difference between $\boldsymbol{Z}_n$ and $\boldsymbol{\gamma}(\boldsymbol{Z}_n,S)$ is given by

$$R_\beta(\boldsymbol{\theta};\boldsymbol{Z}_n)-R_\beta(\boldsymbol{\theta};\boldsymbol{\gamma}(\boldsymbol{Z}_n,S))\ge-\frac{n}{2}E\left[\left(\|\boldsymbol{F}(\boldsymbol{Z}_n,S)\|_\Sigma^2+2\boldsymbol{F}'(\boldsymbol{Z}_n,S)\Sigma^{-1}(\boldsymbol{Z}_n-\boldsymbol{\theta})\right)e^{\left(-\frac{nb}{2}\|\boldsymbol{Z}_n-\boldsymbol{\theta}\|_\Sigma^2\right)}\right]. \tag{11}$$

Let $\boldsymbol{V}_i=\boldsymbol{Z}_i/\sqrt{nb+1}$ for $i=1,\cdots,n-1$, $T=\sum_{j=1}^{n-1}\boldsymbol{V}_j\boldsymbol{V}_j'$, $\boldsymbol{G}(\boldsymbol{u},T)=\boldsymbol{F}(\boldsymbol{u},(nb+1)T)$ and $\Lambda=\Sigma/(nb+1)$. Then, the right hand side in (11) is rewritten by

$$-\frac{n}{2}(nb+1)^{-1-p/2}E\left[\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}\boldsymbol{G}(\boldsymbol{U},T)+2\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}(\boldsymbol{U}-\boldsymbol{\theta})\right].$$

By applying Lemma 1 of Fourdrinier et al. (2003), we see that

$$\begin{aligned} &E\left[\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}\boldsymbol{G}(\boldsymbol{U},T)+2\boldsymbol{G}'(\boldsymbol{U},T)\Lambda^{-1}(\boldsymbol{U}-\boldsymbol{\theta})\right] \\ &\quad\le E\left[2\blacktriangledown_{\boldsymbol{U}}\cdot\boldsymbol{G}(\boldsymbol{U},T)+(n-p+2)\boldsymbol{G}(\boldsymbol{U},T)'T^{-1}\boldsymbol{G}(\boldsymbol{U},T)\right]. \end{aligned} \tag{12}$$

Since $T=S/(nb+1)$, the right hand side in (12) is expressed as

$$E\left[2\blacktriangledown_{\boldsymbol{U}}\cdot\boldsymbol{F}(\boldsymbol{U},S)+(n-p+2)(nb+1)\boldsymbol{F}'(\boldsymbol{U},S)S^{-1}\boldsymbol{F}(\boldsymbol{U},S)\right]. \tag{13}$$

From (11) to (13), we have

$$\begin{aligned} R_\beta(\boldsymbol{\theta};\boldsymbol{Z}_n)-R_\beta(\boldsymbol{\theta};\boldsymbol{\gamma}(\boldsymbol{Z}_n,S)) &\ge -\frac{n}{2}(nb+1)^{-1-p/2} \\ &\quad\times E\left[2\blacktriangledown_{\boldsymbol{U}}\cdot\boldsymbol{F}(\boldsymbol{U},S)+(n-p+2)(nb+1)\boldsymbol{F}'(\boldsymbol{U},S)S^{-1}\boldsymbol{F}(\boldsymbol{U},S)\right] \\ &\ge 0 \end{aligned} \tag{14}$$

for all $\boldsymbol{\theta}\in\mathbb{R}^p$ and $\Sigma\in\mathscr{S}(p)$. By the same reason as Theorem 1, the equality in (14) holds only if $\boldsymbol{F}(\boldsymbol{Z}_n,S)=0$ (a.s.). □

4 Conclusions

In this paper, we establish the conditions under which a general class of shrinkage estimators outperforms the consistent estimator of a multivariate normal population mean when using a divergence loss function. The results obtained apply to both known and unknown covariance scenarios. Furthermore, our findings extend the earlier work of Ghosh and Mergel (2009) to a broader class of dominant estimators. The proofs of the proposed theorems are presented in a more straightforward manner than in the previous reference. This method can also be applied to exponential-type loss functions, such as LINEX, and reflected normal losses. Importantly, the conditions for superiority are robust concerning the squared error loss function, which further validates our derivations, as the divergence loss encompasses the square error loss as a special case.

4.1 Easy understanding

In conclusion, we provide a simple approach for making decisions regarding the superiority conditions based on divergence loss. As previously mentioned, the square error loss is a specific case of divergence loss. It is important to note that the graphs of both losses are parabolic. Additionally, the graph of the risk of $\mathbf{X}$ in Equation (3) also follows a parabolic form (see Figure 2). From a graphical perspective, for any estimator of $\boldsymbol{\theta}$ to be superior to $\mathbf{X}$, it must have a risk that is lower than that of the parabolic shape. Therefore, we can conclude that determining the superiority conditions for the proposed shrinkage estimators can be achieved by examining them under the square error loss. Interestingly, the conditions derived in this study align exactly with those found in the work of Ghosh and Mergel (2009), as they are the same under square error loss when $b \to 0$.

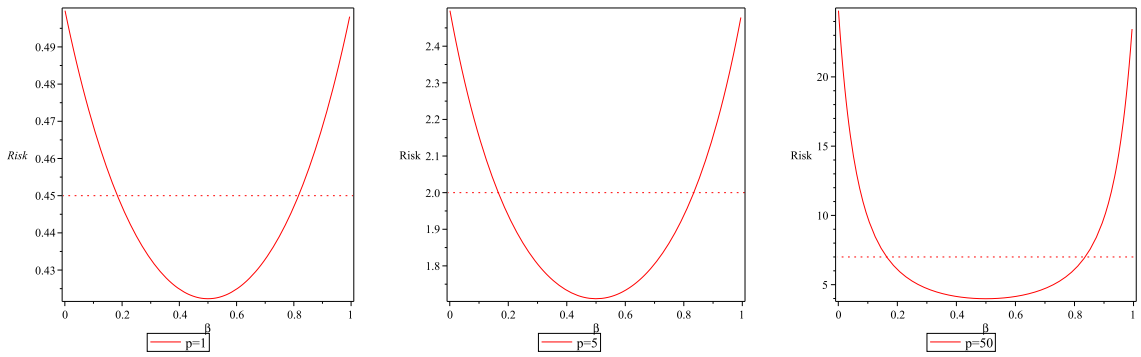


Figure 2: Risk of $\boldsymbol{X}$ under the divergence loss.

Acknowledgments

We would like to express our gratitude to the two anonymous reviewers for their constructive comments, which prompted us to include many additional details in the paper and significantly

improve its presentation. Mohammad Arashi's work is based on the research supported in part by the Iran National Science Foundation (INSF) grant No. 4015320.

References

Amari, S. (1982). Differential geometry of curved exponential families-curvatures and information loss. *The Annals of Statistics*, **10**, 357–387.

Arashi, M. and Tabatabaey, S. M. M. (2010). A note on classical Stein-type estimators in elliptically contoured models. *Journal of Statistical Planning and Inference*, **140**, 1206–1213.

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *The Annals of Mathematical Statistics*, **41**, 642–645.

Brandwein, A. C. and Strawderman, W. E. (1980). Minimax estimators of location parameters for spherically symmetric distributions with concave loss. *The Annals of Statistics*, **8**, 279–284.

Casella, G. (1990). Estimators with nondecreasing risk: Application of a chi-square identity. *Statistics and Probability Letters*, **10**, 107–109.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **46**, 440–464.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors-An empirical Bayes approach. *Journal of the American Statistical Association*, **68**, 117–130.

Faith, R. E. (1978). Minimax Bayes estimators of a multivariate normal mean. *Journal of Multivariate Analysis*, **8**, 372–379.

Fourdrinier, D., Strawderman, W. E. and Wells, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *Journal of Multivariate Analysis*, **85**, 24–39.

George, E. I. (1991). Shrinkage domination in a multivariate common mean problem. *The Annals of Statistics*, **19**, 952–960.

Ghosh, M., Mergel, V. and Datta, G. S. (2008). Estimation, prediction and the Stein phenomenon under divergence loss, *Journal of Multivariate Analysis*, **99**, 1941–1961.

Ghosh, M. and Mergel, V. (2009). On the Stein phenomenon under divergence loss and an unknown variance-covariance matrix. *Journal of Multivariate Analysis*, **100**, 2331–2336.

James, W. and Stein, C. (1961). Estimation of quadratic loss. *In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 361–379.

Maruyama, Y. (2004). Stein's idea and minimax admissible estimation of a multivariate normal mean, *Journal of Multivariate Analysis*, **88**, 320–334.

Shao, P. Yi-Shi and Strawderman, W. E. (1994). Improving on the James-Stein positive-part estimator. *The Annals of Statistics*, **22**, 1517–1538.

Srivastava, M. S. and Kubokawa, T. (2005). Minimax multivariate empirical Bayes estimators under multicollinearity. *Journal of Multivariate Analysis*, **93**, 394–416.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9**, 1135–1151.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 197–206.

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*, **42**, 385–388.

Wells, M. T. and Zhou, G. (2008). Generalized Bayes minimax estimators of the mean of multivariate normal distribution with unknown variance. *Journal of Multivariate Analysis*, **99**, 2208–2220.

Stochastic Models in Probability and Statistics
Vol. 2, No. 1, 2025, pp. 41–56. Research Article

Evaluation of evidences for dynamic systems based on Bayes factors with an application

Majid Hashempour[†*], **and Mahdi Doostparast**[‡]

[†] *Department of Statistics, School of Basic Sciences, University of Hormozgan, Bandar Abbas, Iran*

[‡] *Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran*

Email(s): ma.hashempour@hormozgan.ac.ir, doustparast@um.ac.ir

Abstract. This paper deals with the computation of Bayes factors (BFs) based on sequential order statistics arising from homogeneous exponential populations. Explicit expressions for the BFs are derived from the chi-square and the Poisson distribution functions. Some approximations for the derived BFs are also proposed. A simulated data set is analyzed using the obtained results. Open problems are also mentioned. The findings of this paper may be used for assessing evidence in the available data in various fields such as reliability analyses of engineering systems and life testing experiments.

Keywords: Bayes Factor; Exponential model; Hypotheses testing; Likelihood ratio; Sequential order statistics

1 Introduction

Let $X_1,\ldots,X_n$ be independent and identically distributed (IID) random variables with a common distribution function (DF), say F, and abbreviated by $X_1,\ldots,X_n \overset{i.i.d.}{\sim} F$. Denote in magnitude order of $X_1,\ldots,X_n$ by $X_{1:n} \leq \cdots \leq X_{n:n}$, known as *order statistics* (OSs). The theory of OSs has been widely used in literature. For example, in system reliability analyses, lifetimes of r-out-of-n systems coincide to $X_{r:n}$, where $X_1,\ldots,X_n$ stand for component lifetimes. For more information, see Barlow and Proschan (1981) and David and Nagaraja (2003) and references therein. There are some generalizations of OSs such as *fractional order statistics* and *generalized order statistics*, which are useful for providing a framework to unify similar results in the related literature; see David and Nagaraja (2003) for more information. This paper deals with another unified concept,

*Corresponding author

Received: 14 October 2024 / Accepted: 08 April 2025
DOI: 10.22067/smps.2025.90268.1029

called the *sequential order statistics* (SOS), which has also a motivation in reliability analyses of engineering systems. Specifically, when the component lifetimes are IID, the OSs are suitable for describing r-out-of-n system lifetime. Here failing a component does not change the DFs of lifetimes of surviving components. Motivated by Cramer and Kamps (1996), the failure of a component may result in a higher load on the surviving components and hence causes the lifetime distributions change. More precisely, suppose that F_j, for $j=1,\ldots,n$, denotes the common DF of the component lifetimes when $n-j+1$ components are working. The components begin to work independently at time $t=0$ with the common DF F_1. When at time x_1, the first component failure occurs, the remaining $n-1$ components are working with the left truncated common DF F_2 at x_1. This process continues up to rth component failure and hence the system fails. The mentioned system is called *sequential r-out-of-n system* and its lifetime is then rth component failure time, denoted by $X^{\star}_{(r)}$. In the literature, $(X^{\star}_{(1)},\ldots,X^{\star}_{(n)})$ is called SOSs. Statistical inferences on the basis of SOSs have been studied in literature. For example, Bedbur (2010) obtained the uniformly most powerful unbiased test under a conditional proportional hazard rates (CPHR) model via a decision-theoretic approach. To describe the CPHR model, let $\bar{F}_j(t)=\bar{F}_0^{\alpha_j}(t)$, for $j=1,\ldots,r$, where $\bar{F}_0(t)=1-F_0(t)$ is a given baseline DF. In this case, the hazard rate function of the DF F_j, defined by $h_j(t)=f_j(t)/\bar{F}_j(t)$ for $t>0$ and $j=1,\ldots,n$, is proportional to the hazard rate function of the baseline DF F_0, that is, $h_j(t)=\alpha_j h_0(t)$. See also, Cramer and Kamps (2001a,b), Beutner and Kamps (2009), Schenk et al. (2011), Burkschat and Navarro (2011), Esmailian and Doostparast (2014), Hashempour and Doostparast (2017) and references therein. In this paper, we consider that the DF $F_0(t)$ is the exponential distribution, denoted by $Exp(\sigma)$, that is,

$$F_0(t;\sigma)=1-\exp\Big\{-\Big(\frac{t}{\sigma}\Big)\Big\},\qquad t>0,\quad \sigma>0. \tag{1}$$

The problem of hypotheses testing for exponential populations on the basis of $s(\geq 2)$ multiple and independent SOS samples under the CPHR model via a Bayesian approach is here studied. The available data are denoted by

$$\mathbf{x}=\begin{bmatrix} x_{11} & \ldots & x_{1r} \\ \vdots & \ddots & \vdots \\ x_{s1} & \ldots & x_{sr} \end{bmatrix}, \tag{2}$$

where the ith row of the matrix $\mathbf{x}$ in (2) stands for the SOS sample coming from the ith population $1\leq i\leq s$. In general, the likelihood function (LF) of the available data (2) reads

$$\begin{aligned} L(F_j^{[i]};1\leq i\leq s,1\leq j\leq r) &= \Big(\frac{\Gamma(n+1)}{\Gamma(n-r+1)}\Big)^s\prod_{i=1}^{s}\Bigg(\prod_{j=1}^{r-1}f_j^{[i]}(x_{ij})\Big(\frac{\bar{F}_j^{[i]}(x_{ij})}{\bar{F}_{j+1}^{[i]}(x_{ij})}\Big)^{n-j}\Bigg) \\ &\quad\times f_r^{[i]}(x_{ir})\bar{F}_r^{[i]}(x_{ir})^{n-r}, \end{aligned} \tag{3}$$

where $\bar{F}_j^{[i]}(x)=1-F_j^{[i]}(x)$, and $F_j^{[i]}$ calls for the common DF of the component lifetimes in the ith sequential r-out-of-n sample. For more details, refer to Hashempour and Doostparast (2017). Upon substituting (1) into (3), the LF (3) under the CPHR model reduces to

$$L(\sigma,\alpha;\mathbf{x})=\Big(\frac{\Gamma(n+1)}{\Gamma(n-r+1)}\Big)^s\Big(\prod_{i=1}^{s}\frac{1}{\sigma_i}\Big)^r\exp\Big\{-\sum_{i=1}^{s}\sum_{j=1}^{r}\Big(\frac{x_{ij}m_j}{\sigma_i}\Big)\Big\}, \tag{4}$$

where $\sigma=(\sigma_1,\ldots,\sigma_r)^T$, $\alpha=(\alpha_1,\ldots,\alpha_r)^T$ and $m_j=(n-j+1)\alpha_j-(n-j)\alpha_{j+1}$ for $j=1,\ldots,r$, with convention $\alpha_{r+1}\equiv 0$. For the special case $\sigma_1=\cdots=\sigma_s=\sigma$, the LF (4) simplifies to

$$L(\sigma,\alpha;\mathbf{x})=\left(\frac{\Gamma(n+1)}{\Gamma(n-r+1)}\right)^s\left(\prod_{j=1}^{r}\alpha_j\right)^s\left(\frac{1}{\sigma}\right)^{sr}\exp\left\{-\left(\frac{\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j}{\sigma}\right)\right\}, \tag{5}$$

where σ is the common unknown mean of the baseline exponential DF in (1). In what follows, the following lemma is utilized.

Lemma 1. *Let $X^{\star}_{(1)},\ldots,X^{\star}_{(n)}$ be SOSs under the CPHR model with the baseline $Exp(\sigma)$-distribution. Then, for $r=1,\ldots,n$,*

$$\sum_{j=1}^{r}(n-j+1)\alpha_j D_{ij}=\sum_{j=1}^{r}X_{ij}m_j\sim gamma(r,\sigma), \tag{6}$$

where $D_{ij}=X_{ij}-X_{i,j-1}$, for $j=1,\ldots,r$, $gamma(a,b)$ calls for the gamma distribution with density $f(x;a,b)=(\Gamma(a)b^a)^{-1}x^{a-1}\exp\{-(x/b)\}$, for $x>0$,
and $\Gamma(a)=\int_0^{\infty}x^{a-1}e^{-x}dx$ is the complete gamma function.

For more details, refer to Hashempour and Doostparast (2017). To the best of the anthers knowledge, Bayes factors (BFs) on the basis of SOSs has not been studied in the literature. This paper deals with this problem by emphasizing on SOSs coming for exponential baseline distribution under the CPHR model. So, the rest of this paper is organized as follow: In Section 2 , a review on BF and Bayes is given. A general form for BF is also derived for various hypotheses. In Section 3, BFs for SOS coming from exponential populations under the CPHR are provided. In Section 4, some approximations for the derived BF are proposed. These approximations are useful for numerical evaluations of the BFs specially in big data analyses. In Section 5, simulation studies based on SOS are provided. In Section 6, a real data set on failure times of aircraft components for a life test is analyzed. Section 7 concludes.

2 A review on BF

The BF is a Bayesian approach alternative to the frequentest one for comparing multiple candidate models based on the available data, say $\mathbf{x}$.

2.1 BF for simple hypothesis

Presence of nuisance parameters case the definition of the BF vague and complicated. Thus, in what follows, we consider two cases. As mentioned by Cowles (2013), in the Bayesian analysis when there are only two possible states of the world, M_1 and M_2 (or equivalently, two simple hypotheses H_1 and H_2), one may interest to compare the models with the prior probability $\pi(M_1)=1-\pi(M_2)$. Thus, the prior odd in favour of M_1(or H_1) is $\pi(M_1)/\pi(M_2)$ (*or* $\pi(H_1)/\pi(H_2)$). The posterior odd in favour of a model (or a hypothesis) is derived as the analogous ratio of posterior probabilities: $\pi(M_1|\mathbf{x})/\pi(M_2|\mathbf{x})$ (*or* $\pi(H_1|\mathbf{x})/\pi(H_2|\mathbf{x})$).

The BF in favour of a model or hypothesis is the ratio of the posterior odds to the prior odds. Thus, the BF in favour of M_1 versus M_2 is

$$BF = \frac{\pi(M_1|\mathbf{x})/\pi(M_2|\mathbf{x})}{\pi(M_1)/\pi(M_2)}. \tag{7}$$

The BF (7) is a summary of the evidence provided by the data $\mathbf{x}$ in favour of one scientific theory, represented by a statistical model, as opposed to another one. The BF usually is reported on the $\log_{10}$ scale. A review paper by Kass and Raftery (2012) recommends the interpretations of intervals of values of the BF as in Table 1.

Table 1: Interpretation of the strength of evidence

BF	Evidence against H_1
$0 < \text{BF} \leq \frac{1}{10}$	Strong against
$\frac{1}{10} < \text{BF} \leq \frac{1}{3}$	Substantial against
$\frac{1}{3} < \text{BF} \leq 1$	Barely worth mentioning against
$1 < \text{BF} \leq 3$	Barely worth mentioning
$3 < \text{BF} \leq 10$	Substantial
$10 < \text{BF} < \infty$	Strong

Let $f(\mathbf{y}|M_i)$, $(i = 1,2)$ stand for the probability density function (PDF) of $\mathbf{y}$ given the ith model. Then, the BF (7) for comparing to models M_1 and M_2, or equivalently for testing H_1. The model M_1 is correct against the alternative H_2. The model M_2 is correct, is simplified as

$$BF = \frac{\frac{f(\mathbf{y}|M_1)\Pi(M_1)}{f(\mathbf{y}|M_2)\Pi(M_2)}}{\frac{\Pi(M_1)}{\Pi(M_2)}} = \frac{f(\mathbf{y}|M_1)}{f(\mathbf{y}|M_2)}. \tag{8}$$

Equation (8) means that, the BF is the ratio of the likelihoods under the two simple hypotheses. In other words, it is the evidence contained in the data alone (uninfluenced by the prior) in favour of one model over the other.

Example 1. On the basis of the observed data $\underline{x}$ in (2) and under the CPHR model, described in the preceding section with the baseline $Exp(\sigma)$-distribution, consider the problem of hypotheses testing

$$H_1 : \sigma = \sigma_1 \quad v.s \quad H_2 : \sigma = \sigma_2, \tag{9}$$

where σ_1 and σ_2 are known positive constants and $0 < \sigma_1 < \sigma_2$. Equations (5) and (8) get

$$\begin{aligned} BF = \frac{L(\sigma_1;\underline{x})}{L(\sigma_2;\underline{x})} &= \left(\frac{\sigma_2}{\sigma_1}\right)^{sr} \exp\left\{-\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right)\sum_{i=1}^{s}\sum_{j=1}^{r} x_{ij}m_j\right\} \\ &= \left(\frac{\sigma_2}{\sigma_1}\right)^{sr} \exp\left\{-\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right)\sum_{j=1}^{r}(n-j+1)\alpha_j D_{ij}\right\}. \end{aligned} \tag{10}$$

Note that, BF (10) in the simple versus simple case is the weight of evidence contained in the data alone in favour of M_1 versus M_2. Thus, it ignores any information provided by the priors. For more details, see Hashempour and Doostparast (2016, 2017).

2.2 BF for composite hypotheses

In presence of unknown parameters, say $\boldsymbol{\theta}$, the BF given by (7) is not useful. For these cases, the marginal likelihoods may be used. The numeric value of a marginal likelihood is determined by the data and the entire Bayesian model (the form of the likelihood and all levels of priors). To do this, suppose we wish to compare two families of models, denoted by $\mathscr{M}_1$ and $\mathscr{M}_2$, based on the observed data $\mathbf{y}$. The two families may have different likelihoods, different numbers of unknown parameters, and so on. The BF in the general case is the ratio of the marginal likelihoods under the two candidate families of models. Let $\boldsymbol{\theta_i}$ $(i=1,2)$ denote parameters for the family $\mathscr{M}_i$. The marginal likelihoods under the family $\mathscr{M}_i$ is defined by

$$P(\mathbf{y}|\mathscr{M}_i) = \int P(\mathbf{y}|\theta_i)P(\theta_i|\mathscr{M}_i)d\theta_i. \tag{11}$$

Therefore, (8) motivates us to define the BF as

$$BF = \frac{P(\mathbf{y}|\mathscr{M}_1)}{P(\mathbf{y}|\mathscr{M}_2)}. \tag{12}$$

The suggested BF (12) cannot be interpreted as the evidence in the data alone, since clearly the priors affect each marginal likelihood and therefore the BF itself. For more details, refer to Lewis and Raftery (1997) and Klauer et al. (2024).

3 SOS-based BF for exponential populations

In general, we are interested in comparing composite hypotheses $H_1 : \sigma \in \Omega_1$ against the alternative $H_2 : \sigma \in \Omega_2$ where Ω is the parameter space, $\Omega = \Omega_1 \cup \Omega_2$ and $\Omega_1 \cap \Omega_2 = \emptyset$. Here "$\emptyset$" stands for the empty set. Suppose that $\pi(\sigma)$ is the prior density on the parameter space Ω. To derive the BF on the basis of data x in (2), assume that the parameter vector α in (4) is known, and it is suggested to consider the conjugate prior distribution for the scale parameters σ as $\sigma \sim IG(a,b)$, which is the inverse gamma distribution with shape and scale parameters a and b, respectively. The PDF σ is defined as follows:

$$\pi(\sigma) = \frac{b^a}{\Gamma(a)}\sigma^{-(a+1)}\exp\left\{-\left(\frac{b}{\sigma}\right)\right\}, \qquad \sigma > 0, \quad a > 0, \; b > 0. \tag{13}$$

Equations (4) and (13) imply the posterior distribution of σ given $\underline{x}$ as

$$\sigma \mid \underline{x} \sim IG\left(a_i + r, \sum_{j=1}^{r} x_{ij}m_j + b_i\right), \qquad i = 1,\ldots,s. \tag{14}$$

Remark 1. *Under the squared error loss (SEL) function, that is, $L(\theta,\delta) = (\delta-\theta)^2$, where θ is the parameter of interest and δ is an estimate of θ, the Bayes estimate of the parameter θ is the posterior mean. Thus, the Bayes estimate of σ is*

$$\hat{\sigma}_B = \frac{r\hat{\sigma} + b}{a + r - 1}, \tag{15}$$

where $\hat{\sigma}$ *is the ML estimate of* σ *given by* $\hat{\sigma}=\sum_{j=1}^{r}x_{ij}m_j/r=\sum_{j=1}^{r}(n-j+1)\alpha_j D_{ij}/r$. *Note that the Bayes estimate* (15) *is a weighted mean of the mean of the prior* (13) *and the ML estimate above; that is,* $\hat{\sigma}_B=E(\sigma)w+(1-w)\hat{\sigma}$, *where* $w=(a-1)/(a+r-1)$. *For* $r=n$ *and* $\alpha_1=\cdots=\alpha_n=1$, *we have* $\hat{\sigma}_n=\sum_{j=1}^{n}x_{ij}/n$ *and* $\hat{\sigma}_B=(\sum_{j=1}^{n}x_{ij}+b)/(a+n-1)$, *which are, respectively, the well-known ML and the Bayes estimates of the exponential parameters on the basis of a random sample of size* n; *see, for example, Lawless (2003) and Hashempour and Doostparast (2016).*

Proposition 1. *Let* $\pi_1(\sigma)$ *and* $\pi_2(\sigma)$ *be two proper densities over* Ω_1 *and* Ω_2, *respectively. Then the BF for* $H_1:\sigma\in\Omega_1$ *against* $H_2:\sigma\in\Omega_2$ *is*

$$BF_{1,2}=\frac{\Gamma(a_2)\Gamma(a_1^*)b_1^{a_1}(b_2^*)^{a_2^*}}{\Gamma(a_1)\Gamma(a_2^*)b_2^{a_2}(b_1^*)^{a_1^*}}\left(\frac{P(\chi_{2a_1^*}\in 2b_1^*.\Omega_1^{[-1]})}{P(\chi_{2a_2^*}\in 2b_2^*.\Omega_2^{[-1]})}\right)\left(\frac{P(\chi_{2a_2}\in 2b_2.\Omega_2^{[-1]})}{P(\chi_{2a_1}\in 2b_1.\Omega_1^{[-1]})}\right), \tag{16}$$

where $2b_i^*.\Omega_i^{[-1]}=\{2b_i^*\theta^{-1}:\theta\in\Omega\}$, $2b_i.\Omega_i^{[-1]}=\{2b_i\theta^{-1}:\theta\in\Omega\}$, $b_i^*=b_i+\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j$ *and* $a_i^*=a_i+sr$, *for* $i=1,2$ *and* χ_ν *stands for the chi-square distribution with* ν *degrees of freedom.*

Proof. From (12), (13), and (14), the BF of H_1 against H_2 is

$$\begin{aligned}
BF_{1,2}&=\left(\frac{\int_{\Omega_1}L(\sigma|\underline{\mathbf{x}})\pi_1(\sigma)d\sigma}{\int_{\Omega_2}L(\sigma|\underline{\mathbf{x}})\pi_2(\sigma)d\sigma}\right)\left(\frac{\int_{\Omega_2}\pi_2(\sigma)d\sigma}{\int_{\Omega_1}\pi_1(\sigma)d\sigma}\right)\\
&=\frac{\int_{\Omega_1}A^s\left(\prod_{j=1}^{r}\alpha_j\right)^s\left(\prod_{i=1}^{s}\frac{1}{\sigma}\right)^r\exp\left\{-\sum_{i=1}^{s}\sum_{j=1}^{r}\left(\frac{x_{ij}m_j}{\sigma_i}\right)\right\}\frac{b_1^{a_1}}{\Gamma(a_1)}\sigma^{-(a_1+1)}}{\int_{\Omega_2}A^s\left(\prod_{j=1}^{r}\alpha_j\right)^s\left(\prod_{i=1}^{s}\frac{1}{\sigma}\right)^r\exp\left\{-\sum_{i=1}^{s}\sum_{j=1}^{r}\left(\frac{x_{ij}m_j}{\sigma_i}\right)\right\}\frac{b_2^{a_2}}{\Gamma(a_2)}\sigma^{-(a_2+1)}}\\
&\quad\times\frac{\exp\left\{-\left(\frac{b_2}{\sigma}\right)\right\}d\sigma}{\exp\left\{-\left(\frac{b_1}{\sigma}\right)\right\}d\sigma}\frac{\int_{\Omega_2}\frac{b_2^{a_2}}{\Gamma(a_2)}\sigma^{-(a_2+1)}\exp\left\{-\left(\frac{b_2}{\sigma}\right)\right\}d\sigma}{\int_{\Omega_1}\frac{b_1^{a_1}}{\Gamma(a_1)}\sigma^{-(a_1+1)}\exp\left\{-\left(\frac{b_1}{\sigma}\right)\right\}d\sigma}\\
&=\frac{\Gamma(a_2)b_1^{a_1}}{\Gamma(a_1)b_2^{a_2}}\frac{\int_{\Omega_1}\sigma^{-(a_1+sr+1)}\exp\{-\frac{1}{\sigma}\left(\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j+b_1\right)\}d\sigma}{\int_{\Omega_2}\sigma^{-(a_2+sr+1)}\exp\{-\frac{1}{\sigma}\left(\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j+b_2\right)\}d\sigma}\\
&\quad\times\frac{\int_{\Omega_2}\frac{b_2^{a_2}}{\Gamma(a_2)}\sigma^{-(a_2+1)}\exp\left\{-\left(\frac{b_2}{\sigma}\right)\right\}d\sigma}{\int_{\Omega_1}\frac{b_1^{a_1}}{\Gamma(a_1)}\sigma^{-(a_1+1)}\exp\left\{-\left(\frac{b_1}{\sigma}\right)\right\}d\sigma}\\
&=\frac{\Gamma(a_2)\Gamma(a_1+sr)b_1^{a_1}\left(b_2+\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j\right)^{a_2+sr}}{\Gamma(a_1)\Gamma(a_2+sr)b_2^{a_2}\left(b_1+\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j\right)^{a_1+sr}}\frac{P(\chi_{2(a_1+sr)}\in\Omega_1)}{P(\chi_{2(a_2+sr)}\in\Omega_2)}\times\frac{P(\chi_{2a_1}\in\Omega_2)}{P(\chi_{2a_2}\in\Omega_1)}\\
&=\frac{\Gamma(a_2)\Gamma(a_1^*)b_1^{a_1}(b_2^*)^{a_2^*}}{\Gamma(a_1)\Gamma(a_2^*)b_2^{a_2}(b_1^*)^{a_1^*}}\left(\frac{P(\chi_{2a_1^*}\in 2b_1^*.\Omega_1^{[-1]})}{P(\chi_{2a_2^*}\in 2b_2^*.\Omega_2^{[-1]})}\right)\left(\frac{P(\chi_{2a_2}\in 2b_2.\Omega_2^{[-1]})}{P(\chi_{2a_1}\in 2b_1.\Omega_1^{[-1]})}\right).\\
&=A\left(\frac{P(\chi_{2a_1^*}\in 2b_1^*.\Omega_1^{[-1]})}{P(\chi_{2a_2^*}\in 2b_2^*.\Omega_2^{[-1]})}\right)\left(\frac{P(\chi_{2a_2}\in 2b_2.\Omega_2^{[-1]})}{P(\chi_{2a_1}\in 2b_1.\Omega_1^{[-1]})}\right),
\end{aligned}$$

where $A=\frac{\Gamma(a_2)\Gamma(a_1^*)b_1^{a_1}(b_2^*)^{a_2^*}}{\Gamma(a_1)\Gamma(a_2^*)b_2^{a_2}(b_1^*)^{a_1^*}}$. □

In what follows, for the proper prior $\pi_i(\sigma)(i=1,2)$ in Proposition 1, the truncated inverse

gamma distributions to the parameter of spaces is assumed, that is,

$$\pi_i(\sigma) = \frac{b_i^{a_i}}{\Gamma(a_i)}\sigma^{-(a_i+1)}\exp\left\{-\left(\frac{b_i}{\sigma}\right)\right\} \times \frac{1}{\int_{\Omega_i} IG(a_i,b_i)d\sigma}, \quad i=1,2,\ \sigma\in\Omega_i.$$

A general form for the SOS-based BF in (16) is derived in terms of the chi-square DF. For some common hypotheses, the proposed BF in (16) may be simplified. Similar to Lehmann and Romano (2005, Ch 4.), the following hypotheses are considered and the corresponding simplified BFs are displayed in Table 2:

$$H_3 : \sigma \le \sigma_0 \text{ v.s } H_4 : \sigma > \sigma_0 \tag{17}$$

$$H_5 : \sigma \ge \sigma_0 \text{ v.s } H_6 : \sigma < \sigma_0 \tag{18}$$

$$H_7 : \sigma = \sigma_0 \text{ v.s } H_8 : \sigma > \sigma_0 \tag{19}$$

$$H_9 : \sigma = \sigma_0 \text{ v.s } H_{10} : \sigma < \sigma_0 \tag{20}$$

$$H_{11} : \sigma_1 \le \sigma \le \sigma_2 \text{ v.s } H_{12} : \sigma > \sigma_2 \text{ or } \sigma < \sigma_1 \tag{21}$$

$$H_{13} : \sigma \ge \sigma_2 \text{ or } \sigma \le \sigma_1 \text{ v.s } H_{14} : \sigma_1 < \sigma < \sigma_2 \tag{22}$$

$$H_{15} : \sigma = \sigma_0 \text{ v.s } H_{16} : \sigma \ne \sigma_0. \tag{23}$$

Here, σ_0, σ_1, and σ_2 are known positive constants and $\sigma_1 < \sigma_2$. In Table 2, we have

$$A = \left(\Gamma(a_2)\Gamma(a_1^*)b_1^{a_1}(b_2^*)^{a_2^*}\right) / \left(\Gamma(a_1)\Gamma(a_2^*)b_2^{a_2}(b_1^*)^{a_1^*}\right)$$

and

$$B = \left(\Gamma(a_2)(b_2^*)^{a_2^*}\right) / \left(\Gamma(a_2^*)b_2^{a_2}\right),$$

and F_{χ_ν} stands for the DF of the chi-square distribution with ν degrees of freedom.

Lemma 2 (Johnson et al. (1994)). *For $t > 0$,*

$$F_{\chi_\nu}(t) = 1 - \exp\left\{-\frac{t}{2}\right\}\sum_{i=0}^{\nu-1}\frac{(\frac{t}{2})^i}{i!}. \tag{24}$$

Lemma 2 gives an alternative method for the expression of the BFs associated with the hypotheses (17)-(23); see Table 3.

4 Approximate BF

The BFs in (16) and Table 2 involve the DF of the chi-square distribution. In this section, some approximations for the BF defined by (16) are proposed, which may be useful for numerical evaluations indeed in big data analyses. To do this, some lemmas are given. The first lemma is based on the cumulative distribution function (CDF) of the standard normal CDF, that is,

$$\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\{-u^2/2\}du. \tag{25}$$

Table 2: BFs

$BF_{3,4} = A\left(\frac{1-F_{\chi_{2a_1^*}}(\frac{2b_1^*}{\sigma_0})}{1-F_{\chi_{2a_1}}(\frac{2b_1}{\sigma_0})}\right)\left(\frac{F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_0})}{F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_0})}\right)$
$BF_{5,6} = A\left(\frac{1-F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_0})}{1-F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_0})}\right)\left(\frac{F_{\chi_{2a_1^*}}(\frac{2b_1^*}{\sigma_0})}{F_{\chi_{2a_1}}(\frac{2b_1}{\sigma_0})}\right)$
$BF_{7,8} = \frac{B}{\sigma_0^{sr}}\exp\left\{-\frac{1}{\sigma_0}\left(\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j\right)\right\}\left(\frac{F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_0})}{F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_0})}\right)$
$BF_{9,10} = \frac{B}{\sigma_0^{sr}}\exp\left\{-\frac{1}{\sigma_0}\left(\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j\right)\right\}\left(\frac{1-F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_0})}{1-F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_0})}\right)$
$BF_{11,12} = A\left(\frac{F_{\chi_{2a_1^*}}(\frac{2b_1^*}{\sigma_1})-F_{\chi_{2a_1^*}}(\frac{2b_1^*}{\sigma_2})}{F_{\chi_{2a_1}}(\frac{2b_1}{\sigma_1})-F_{\chi_{2a_1}}(\frac{2b_1}{\sigma_2})}\right)\left(\frac{1-F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_1})+F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_2})}{1-F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_1})+F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_2})}\right)$
$BF_{13,14} = A\left(\frac{1-F_{\chi_{2a_1^*}}(\frac{2b_1^*}{\sigma_1})+F_{\chi_{2a_1^*}}(\frac{2b_1^*}{\sigma_2})}{1-F_{\chi_{2a_1}}(\frac{2b_1}{\sigma_1})+F_{\chi_{2a_1}}(\frac{2b_1}{\sigma_2})}\right)\left(\frac{F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_1})-F_{\chi_{2a_2}}(\frac{2b_2}{\sigma_2})}{F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_1})-F_{\chi_{2a_2^*}}(\frac{2b_2^*}{\sigma_2})}\right)$
$BF_{15,16} = \frac{B}{\sigma_0^{sr}}\exp\left\{-\frac{1}{\sigma_0}\left(\sum_{i=1}^{s}\sum_{j=1}^{r}x_{ij}m_j\right)\right\}$

Lemma 3 (Johnson et al. (1994), page 426)**.** *As* $\nu \to +\infty$*, we have for all* $t > 0$

(I) $F_{\chi_\nu}(t) \approx \Phi\left(\frac{t-\nu}{\sqrt{2\nu}}\right)$,

(II) $F_{\chi_\nu}(t) \approx \Phi\left(\sqrt{2t}-\sqrt{2\nu-1}\right)$.

In Lemma 3, the second approximation is better than the first one; see, for example, Johnson et al. (1994). Meanwhile, we provide some approximations for BF with both of them. So, Lemma 3 gives

$$BF_{11,12}^{[I]} = A\left(\frac{\Phi\left(\frac{\frac{2b_1^*}{\sigma_1}-2a_1^*}{\sqrt{4a_1^*}}\right)-\Phi\left(\frac{\frac{2b_1^*}{\sigma_2}-2a_1^*}{\sqrt{4a_1^*}}\right)}{\Phi\left(\frac{\frac{2b_1}{\sigma_1}-2a_1}{\sqrt{4a_1}}\right)-\Phi\left(\frac{\frac{2b_1}{\sigma_2}-2a_1}{\sqrt{4a_1}}\right)}\right)\left(\frac{1-\Phi\left(\frac{\frac{2b_2}{\sigma_1}-2a_2}{\sqrt{4a_2}}\right)+\Phi\left(\frac{\frac{2b_2}{\sigma_2}-2a_2}{\sqrt{4a_2}}\right)}{1-\Phi\left(\frac{\frac{2b_2^*}{\sigma_1}-2a_2^*}{\sqrt{4a_2^*}}\right)+\Phi\left(\frac{\frac{2b_2^*}{\sigma_2}-2a_2^*}{\sqrt{4a_2^*}}\right)}\right), \tag{26}$$

$$\begin{aligned} BF_{11,12}^{[II]} &= A\left(\frac{\Phi\left(\sqrt{\frac{4b_1^*}{\sigma_1}}-\sqrt{4a_1^*-1}\right)-\Phi\left(\sqrt{\frac{4b_1^*}{\sigma_2}}-\sqrt{4a_1^*-1}\right)}{\Phi\left(\sqrt{\frac{4b_1}{\sigma_1}}-\sqrt{4a_1-1}\right)-\Phi\left(\sqrt{\frac{4b_1}{\sigma_2}}-\sqrt{4a_1-1}\right)}\right) \\ &\quad\times\left(\frac{1-\Phi\left(\sqrt{\frac{4b_2}{\sigma_1}}-\sqrt{4a_2-1}\right)+\Phi\left(\sqrt{\frac{4b_2}{\sigma_2}}-\sqrt{4a_2-1}\right)}{1-\Phi\left(\sqrt{\frac{4b_2^*}{\sigma_1}}-\sqrt{4a_2^*-1}\right)+\Phi\left(\sqrt{\frac{4b_2^*}{\sigma_2}}-\sqrt{4a_2^*-1}\right)}\right), \end{aligned} \tag{27}$$

Table 3: BFs based on Lemma 2

$BF_{3,4} = A \left(\frac{\exp\left\{-\frac{b_1^*}{\sigma_0}\right\} \sum_{i=0}^{a_1^*-1} \frac{(\frac{b_1^*}{\sigma_0})^i}{i!}}{\exp\left\{-\frac{b_1}{\sigma_0}\right\} \sum_{i=0}^{a_1-1} \frac{(\frac{b_1}{\sigma_0})^i}{i!}} \right) \left(\frac{1-\exp\left\{-\frac{b_2}{\sigma_0}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_0})^i}{i!}}{1-\exp\left\{-\frac{b_2^*}{\sigma_0}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_0})^i}{i!}} \right)$
$BF_{5,6} = A \left(\frac{\exp\left\{-\frac{b_2}{\sigma_0}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_0})^i}{i!}}{\exp\left\{-\frac{b_2^*}{\sigma_0}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_0})^i}{i!}} \right) \left(\frac{1-\exp\left\{-\frac{b_1^*}{\sigma}\right\} \sum_{i=0}^{a_1^*-1} \frac{(\frac{b_1^*}{\sigma_0})^i}{i!}}{1-\exp\left\{-\frac{b_1}{\sigma_0}\right\} \sum_{i=0}^{a_1-1} \frac{(\frac{b_1}{\sigma_0})^i}{i!}} \right)$
$BF_{7,8} = \frac{B}{\sigma_0^{sr}} \exp\left\{-\frac{1}{\sigma_0}\left(\sum_{i=1}^{s}\sum_{j=1}^{r} x_{ij}m_j\right)\right\} \left(\frac{1-\exp\left\{-\frac{b_2}{\sigma_0}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_0})^i}{i!}}{1-\exp\left\{-\frac{b_2^*}{\sigma_0}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_0})^i}{i!}} \right)$
$BF_{9,10} = \frac{B}{\sigma_0^{sr}} \exp\left\{-\frac{1}{\sigma_0}\left(\sum_{i=1}^{s}\sum_{j=1}^{r} x_{ij}m_j\right)\right\} \left(\frac{\exp\left\{-\frac{b_2}{\sigma_0}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_0})^i}{i!}}{\exp\left\{-\frac{b_2^*}{\sigma_0}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_0})^i}{i!}} \right)$
$BF_{11,12} = A \left(\frac{\exp\left\{-\frac{b_1^*}{\sigma_1}\right\} \sum_{i=0}^{a_1^*-1} \frac{(\frac{b_1^*}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_1^*}{\sigma_2}\right\} \sum_{i=0}^{a_1^*-1} \frac{(\frac{b_1^*}{\sigma_2})^i}{i!}}{\exp\left\{-\frac{b_1}{\sigma_1}\right\} \sum_{i=0}^{a_1-1} \frac{(\frac{b_1}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_1}{\sigma_2}\right\} \sum_{i=0}^{a_1-1} \frac{(\frac{b_1}{\sigma_2})^i}{i!}} \right) \times \left(\frac{1+\exp\left\{-\frac{b_2}{\sigma_1}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_2}{\sigma_2}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_2})^i}{i!}}{1+\exp\left\{-\frac{b_2^*}{\sigma_1}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_2^*}{\sigma_2}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_2})^i}{i!}} \right)$
$BF_{13,14} = A \left(\frac{1+\exp\left\{-\frac{b_1^*}{\sigma_1}\right\} \sum_{i=0}^{a_1^*-1} \frac{(\frac{b_1^*}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_1^*}{\sigma_2}\right\} \sum_{i=0}^{a_1^*-1} \frac{(\frac{b_1^*}{\sigma_1})^i}{i!}}{1+\exp\left\{-\frac{b_1}{\sigma_1}\right\} \sum_{i=0}^{a_1-1} \frac{(\frac{b_1}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_1}{\sigma_2}\right\} \sum_{i=0}^{a_1-1} \frac{(\frac{b_1}{\sigma_2})^i}{i!}} \right) \times \left(\frac{\exp\left\{-\frac{b_2}{\sigma_1}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_2}{\sigma_2}\right\} \sum_{i=0}^{a_2-1} \frac{(\frac{b_2}{\sigma_2})^i}{i!}}{\exp\left\{-\frac{b_2^*}{\sigma_1}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_1})^i}{i!} - \exp\left\{-\frac{b_2^*}{\sigma_2}\right\} \sum_{i=0}^{a_2^*-1} \frac{(\frac{b_2^*}{\sigma_2})^i}{i!}} \right)$
$BF_{15,16} = \frac{B}{\sigma_0^{sr}} \exp\left\{-\frac{1}{\sigma_0}\left(\sum_{i=1}^{s}\sum_{j=1}^{r} x_{ij}m_j\right)\right\}$

$$BF_{3,4}^{[I]} = A \left(\frac{1 - \Phi\left(\frac{\frac{2b_1^*}{\sigma_0} - 2a_1^*}{\sqrt{4a_1^*}} \right)}{1 - \Phi\left(\frac{\frac{2b_1}{\sigma_0} - 2a_1}{\sqrt{4a_1}} \right)} \right) \left(\frac{\Phi\left(\frac{\frac{2b_2}{\sigma_0} - 2a_2}{\sqrt{4a_2}} \right)}{\Phi\left(\frac{\frac{2b_2^*}{\sigma_0} - 2a_2^*}{\sqrt{4a_2^*}} \right)} \right), \tag{28}$$

$$BF_{3,4}^{[II]} = A \left(\frac{1 - \Phi\left(\sqrt{\frac{4b_1^*}{\sigma_0}} - \sqrt{4a_1^* - 1} \right)}{1 - \Phi\left(\sqrt{\frac{4b_1}{\sigma_0}} - \sqrt{4a_1 - 1} \right)} \right) \left(\frac{\Phi\left(\sqrt{\frac{4b_2}{\sigma_0}} - \sqrt{4a_2 - 1} \right)}{\Phi\left(\sqrt{\frac{4b_2^*}{\sigma_0}} - \sqrt{4a_2^* - 1} \right)} \right). \tag{29}$$

The next lemma presents another approximation the CDF of the χ_ν distribution base on an infinite series. So, the CDF can be approximated by computing the summation for some finite elements.

Lemma 4 (Johnson et al. (1994)). *For $x > 0$,*

$$F_{\chi_\nu}(t) = \frac{2(2t)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}\sum_{i=0}^{\infty}(-1)^i\frac{t^i}{(\nu+2i)2^i i!}. \tag{30}$$

The proposed BFs can be approximated by Lemma 4. For example,

$$\begin{aligned} BF_{11,12} = A&\frac{\frac{2^{a_1^*}}{\Gamma(a_1^*)}\sum_{i=0}^{N}\left\{\frac{(-2)^i(2b_1^*)^{i+a_1^*}}{i!(a_1^*+i)}\left(\left(\frac{1}{\sigma_1}\right)^{i+a_1^*}-\left(\frac{1}{\sigma_2}\right)^{i+a_1^*}\right)\right\}}{\frac{2^{a_1}}{\Gamma(a_1)}\sum_{i=0}^{N}\frac{(-2)^i(2b_1)^{i+a_1}}{i!(a_1+i)}\left(\left(\frac{1}{\sigma_1}\right)^{i+a_1}-\left(\frac{1}{\sigma_2}\right)^{i+a_1}\right)} \\ &\times\frac{1+\frac{2^{a_2}}{\Gamma(a_2)}\sum_{i=0}^{N}\frac{(-2)^i(2b_2)^{i+a_2}}{i!(a_2+i)}\left(\left(\frac{1}{\sigma_2}\right)^{i+a_2}-\left(\frac{1}{\sigma_1}\right)^{i+a_2}\right)}{1+\frac{2^{a_2^*}}{\Gamma(a_2^*)}\sum_{i=0}^{N}\frac{(-2)^i(2b_2^*)^{i+a_2^*}}{i!(a_2^*+i)}\left(\left(\frac{1}{\sigma_2}\right)^{i+a_2^*}-\left(\frac{1}{\sigma_1}\right)^{i+a_2^*}\right)}, \end{aligned} \tag{31}$$

where N is a large number. Other approximations methods such as the Laplace method can be used to approximate the F-distribution function; see, for example, Johnson et al. (1994). A similar approach then is used to obtain an approximate BF based on SOSs.

5 Simulation studies

To examine the accuracy of the proposed BF, we performed a Monte Carlo simulation study in the well-known statistical software R. For generating an SOS-sample from the exponential population with $\sigma = 1$ under the CPHR model, an algorithm proposed by Cramer and Kamps (1996) was performed. Here, we considered the hypothesis $H_{11} : \sigma_1 \leq \sigma \leq \sigma_2$ v.s $H_{12} : \sigma > \sigma_2$ or $\sigma < \sigma_1$.

In Table 4 and Figure 1, the mean of the BFs based on 10^4 iterations for some selected reduces of n and r are displayed. Appr 1 and Appr 2 stand for the approximations based on Lemma (3) and (4), respectively.

Table 5 and Figure 2 represent the mean absolute among approximate and exact BFs.

Empirical results are

- increasing r more effective than increasing the copy s;
- approximations tend to the actual value as r/n increasing;
- Appr 1 dominates Appr 2;
- As $n \longrightarrow \infty$ and r/n goes to unify, the BF determines successfully the correct hypothesis.

Table 4: Exact values and the corresponding approximates for the BF on the basis of a SOS-sample from the exponential population under the CPHR model for some selected values of n and r.

	Exact	Appr1	Appr2
3	1.9843	2.0852	2.6624
4	1.5754	1.6474	2.0768
5	1.2117	1.2621	1.5760
6	0.9860	1.0200	1.2533
7	0.6417	0.6731	0.8562
8	0.5612	0.5880	0.7453
9	0.3924	0.4116	0.5234
10	0.3719	0.4053	0.4997
11	0.2948	0.3078	0.3863
12	0.2341	0.2442	0.3054

(a) $n = 20,\ r = 10$

	Exact	Appr1	Appr2
3	1.3629	1.4212	1.7790
4	0.8295	0.8679	1.0981
5	0.6116	0.6389	0.8052
6	0.4283	0.4472	0.5619
7	0.2965	0.3103	0.3921
8	0.2345	0.2451	0.3080
9	0.1707	0.1780	0.2224
10	0.1221	0.1271	0.1579
11	0.1067	0.1109	0.1371
12	0.0791	0.0821	0.1010

(b) $n = 20,\ r = 15$

	Exact	Appr1	Appr2
3	2.5126	2.6767	3.5636
4	2.4532	2.5998	3.4039
5	2.3153	2.4416	3.1484
6	1.9026	2.0020	2.5661
7	1.7864	1.8722	2.3728
8	1.5225	1.5929	2.0113
9	1.3309	1.3891	1.7427
10	1.1886	1.2388	1.5503
11	0.9422	0.9870	1.2512
12	0.8756	0.9137	1.1486

(c) $n = 20,\ r = 5$

	Exact	Appr1	Appr2
3	2.4105	2.5696	3.4268
4	2.4171	2.5623	3.3578
5	2.1453	2.2650	2.9317
6	1.8881	1.9863	2.5442
7	1.7795	1.8660	2.3694
8	1.5757	1.6483	2.0789
9	1.3484	1.4092	1.7730
10	1.1382	1.1905	1.5024
11	1.0839	1.1308	1.4183
12	0.9575	0.9928	1.2282

(d) $n = 10,\ r = 5$

6 Aircraft data set

To demonstrate the results obtained in the preceding sections, we present an illustrative example. Smith (2002) gave failure times of aircraft components for a life-test, originally due to Mann and Fertig (1973). In the test, $n = 13$ components were placed in a Type-II censored life test in which the failure times of first 10 components to fail were observed (in hours) as $0.22, 0.50, 0.88, 1.00, 1.32, 1.33, 1.54, 1.76, 2.50, 3.00$. Following Hashempour et al. (2019), it is assumed that the lifetimes of the components are IID with an exponential distribution. We considered two simple hypothesis tests based on the ML estimate of the σ in (15). Also, we ran the SOS example for $r = 3,4$ and $s = 3,4,5$. The BF is approximated using (31) for the failure time of aircraft components. The results on Table 6 show that as r or s increases, BF determines the correct hypothesis more successfully.

Table 5: Absolute errors of the approximations in Table 4 for the BF on the basis of a SOS-sample from the exponential population under the CPHR model for some selected values of n and r.

	Appr1	Appr2
3	0.1025	0.6823
4	0.0743	0.5052
5	0.0586	0.3724
6	0.0496	0.2989
7	0.0327	0.2175
8	0.0279	0.1867
9	0.0203	0.1334
10	0.0147	0.1297
11	0.0141	0.0936
12	0.0109	0.0732

(a) $n = 20,\ r = 10$

	Appr1	Appr2
3	0.0631	0.4197
4	0.0420	0.2717
5	0.0319	0.2030
6	0.0206	0.1361
7	0.0148	0.0978
8	0.0113	0.0754
9	0.0080	0.0534
10	0.0056	0.0373
11	0.0048	0.0318
12	0.0035	0.0231

(b) $n = 20,\ r = 15$

	Appr1	Appr2
3	0.1661	1.0566
4	0.1484	0.9558
5	0.1280	0.8377
6	0.1010	0.6677
7	0.0873	0.5903
8	0.0727	0.4928
9	0.0632	0.4154
10	0.0565	0.3698
11	0.0493	0.3151
12	0.0436	0.2808

(c) $n = 20,\ r = 5$

	Appr1	Appr2
3	0.1610	1.0216
4	0.1470	0.9456
5	0.1215	0.7911
6	0.0998	0.6605
7	0.0881	0.5940
8	0.0749	0.5070
9	0.0653	0.4283
10	0.0581	0.3716
11	0.0536	0.3444
12	0.0469	0.2913

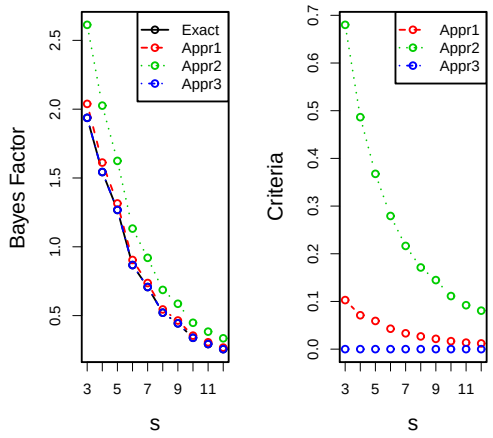
(d) $n = 10,\ r = 5$

7 Conclusion

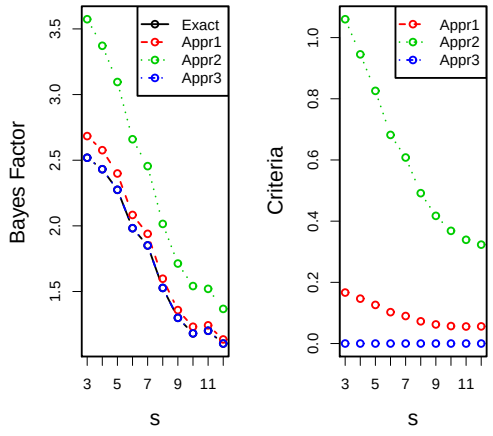
This paper focused on calculating BFs using sequential order statistics arising from homogeneous exponential DFs. Also various approximations for these BFs were proposed. A simulation study was conducted, and real data set was illustrated. The discoveries presented in this paper have practical applications in evaluating evidence in various domains, including reliability analysis of engineering systems and life testing experiments.

Acknowledgements

The authors thank the editor-in-chief, the associate editor, and the anonymous reviewers for their useful comments on the earlier version of this paper.



(a) $n = 20$, $r = 10$



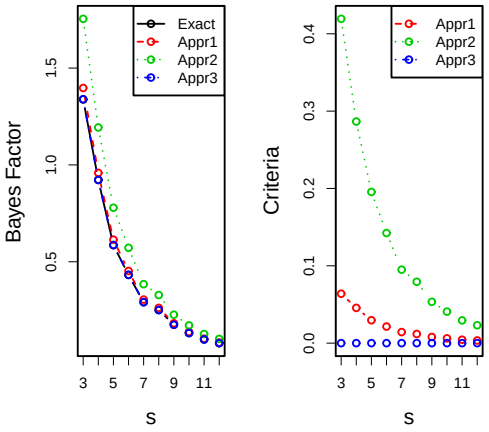
(b) $n = 20$, $r = 5$

Figure 1: BF and criteria.

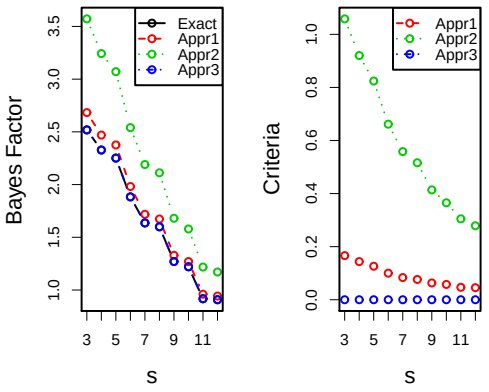
References

Balakrishnan, N., and Kateri, M. (2008). On the maximum likelihood estimation of parameters of Weibull distribution based on complete and censored data. *Statistics and Probability Letters*, **78**, 2971–2975.

Barlow, R.E., and Proschan, F. (1981). *Statistical Theory of Reliability and life Testing: Probability Models.* Springer, Second Edition, New York.



(a) $n = 20,\ r = 15$



(b) $n = 10,\ r = 5$

Figure 2: BF and criteria.

Bedbur, S. (2010). UMPU tests based on sequential order statistics. *Journal of Statistical Planning and Inference*, **140**, 2520–2530.

Beutner, E., and Kamps, U. (2009). Order restricted statistical inference for scale parameters based on sequential order statistics. *Journal of Statistical Planning and Inference*, **139**, 2963–2969.

Burkschat, M., and Navarro, J. (2011). Ageing properties of sequential order statistics. *Probability in the Engineering and Informational Science*, **25**, 449–467.

Cramer, E., and Kamps, U. (1996) Sequential order statistics and k-out-of-n systems with se-

Table 6: Approximations of the BF for failure times of aircraft components.

H_1 vs H_2	r	s	BF
$H_1 : \sigma = 5$	3	3	2.112
$H_2 : \sigma = 6$		4	2.649
		5	6.235
	4	3	4.886
		4	13.537
		5	16.942
$H_1 : \sigma = 5$	3	3	10.784
$H_2 : \sigma = 7$		4	22.507
		5	81.136
	4	3	49.640
		4	301.139
		5	463.789

quentially adjusted failure rates. it Annals of the Institute of Statistical Mathematics, **48**, 535–549.

Cramer. E., and Kamps, U. (2001a). Estimation with sequential order statistics from exponential distributions. *Annals of the Institute of Statistical Mathematics*, **53**, 307–324.

Cramer, E. and Kamps, U. (2001b). Sequential *k*-out-of-*n* systems. In N. Balakrishnan and E. Rao, editors, *Handbook of Statistics, Advances in Reliability*, volume 20, Chapter 12, 301–372.

Cowles, M. K. (2013). *Applied Bayesian Statistics With R and OpenBUGS Examples*. John Wiley & Sons, New York.

David, H. A., and H. N. Nagaraja (2003). it Order Statistics. John Wiley & Sons, New York.

Esmailian, M. and Doostparast, M. (2014). Estimation based on sequential order statistics with random removals. *Probability and Mathematical Statistics*, **34**, 81–95.

Hashempour, M., and Doostparast, M. (2016). Statistical evidences in sequential order statistics arising from a general family of lifetime distributions. *Istatistik: Journal of the Turkish Statistical Association*, **9**, 29–41.

Hashempour, M., and Doostparast, M. (2017). Bayesian inference on multiply sequential order statistics from heterogeneous exponential populations with GLR test for homogeneity. *Communications in Statistics-Theory and Methods*, **46**, 8086–8100.

Hashempour, M., Doostparast, M. and Velayati Moghaddam, E. (2019). Weibull analysis with sequential order statistics under a power trend model for hazard rates with application in arcraft data analysis. *Journal of Statistical Research of Iran*, **16**, 535–557.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. John Wiley & Sons, Inc. Volume 1. Second Edition, New York.

Kass, R. E., and A. E. Raftery. (2012). Bayes factors. *Journal of the American Statistical Association*, **30**, 773–795.

Klauer, K. C., Meyer-Grant, C. G. and Kellen, D. (2024). On Bayes factors for hypothesis tests. *Psychonomic Bulletin and Review*, **32**, 1070–1094.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. 2-th Edition, John Wiley & Sons, Hoboken, New Jersey.

Lehmann, E.L., and Romano, J. (2005). *Testing Statistical Hypothesis*. 3-th Edition, Springer, New York.

Lewis, S. M. and Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, **92**, 648–665.

Mann, N. R. and Fertig, K. W. (1973). Tables for obtaining Weibull confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme value distribution. *Technometrics*, **15**, 87–101.

Schenk, N., Burkschat, M., Cramer, E. and Kamps, U. (2011). Bayesian estimation and prediction with multiply iype-II censored samples of sequential order statistics from one- and two-Parameter exponential distributions. *Journal of Statistical Planning and Inference*, **141**, 1575–1587.

Smith, P. J. (2002). *Analysis of Failure and Survival Data.* Chapman & Hall/CRC, New York.

Stochastic Models in Probability and Statistics
Vol. 2, No. 1, 2025, pp. 57–74. Research Article

Log transformed transmuted exponential distribution: an increasing hazard rate model to deal with cancer patients data

Md. Tahir*, Sanjay K. Singh and Abhimanyu Singh Yadav

Department of Statistics, Banaras Hindu University, Varanasi, India

Email(s): tahirmansori@gmail.com, singhsk64@gmail.com, asybhu10@gmail.com

Abstract. In this paper, we introduce a novel distribution called the log transformed transmuted exponential (LTTE), which is derived by applying a log transformation to the transmuted exponential distribution as the baseline model. We derive several key mathematical and statistical properties of the LTTE distribution, including its moments, quantile function, skewness, kurtosis, reliability function, and hazard rate, along with their respective shapes. The maximum likelihood estimation method is used to estimate the parameters of the distribution. The practical applicability of the LTTE distribution is demonstrated by fitting it to three real-life datasets related to cancer patients. The results indicate that the LTTE distribution offers a superior fit, as evidenced by better values of AIC, BIC, and the Kolmogorov–Smirnov (KS) statistic, when compared to other existing lifetime models.

Keywords: Maximum likelihood estimation, Moments, Transmuted exponential, Log transformation.

1 Introduction

It is an indisputable fact that the world is facing an epidemic of noncommunicable diseases, with cancer cases continuing to grow at an alarming rate. Cancer is currently ranked as the second leading cause of death, following cardiovascular diseases; see Jemal et al. (2008). The GLOBOCAN 2018 report recorded 18.1 million new cancer cases and 9.6 million cancer-related deaths globally. Emerging challenges such as rapid urbanization, population aging, unhealthy lifestyles, and indoor and outdoor air pollution are contributing to the growing cancer burden worldwide, particularly in middle and low-income countries, for example, India. According to the WHO 2020 ranking on cancer burden, India ranks third in terms of new yearly cancer cases,

*Corresponding author

Received: 03 January 2025 / Accepted: 24 April 2025
DOI: 10.22067/smps.2025.91480.1041

following China and United States. The Indian Council of Medical Research's National Cancer Registry Program (ICMR-NCRP), an initiative by the government of India to estimate cancer incidence in the country, reported 1.39 million new cancer cases in 2020 and 1.46 million in 2022. The GLOBOCAN projection estimates that cancer cases in India will rise to 2.08 million by 2040, representing a nearly 50% increase from 2020, Sathish Kumar et al. (2022). Among all the cancer cases, Breast cancer is the leading cause of cancer incidence and mortality in India, accounting for 13.5% of new cases and 10.6% of cancer-related deaths in 2020, Mehrotra and Yadav (2022). Urban factors such as sedentary lifestyle, high obesity rates, delayed marriage, and childbirth, and minimal breastfeeding contribute to its higher burden in urban areas. Another type of hazardous cancer is bladder cancer which ranks as the ninth most common cancer, representing 3.9% of all cancer cases in India, Prakash et al. (2019). It is primarily linked to tobacco use and exposure to industrial chemicals. Also, leukemia, a cancer of the blood-forming tissues, compromises the body's ability to fight infections. It accounts for 27% to 52% of childhood cancers in males and 19% to 52% in females across various population-based registries, Bhutani et al. (2004).

With such a large number of cancer cases being reported, there is a vast amount of data available to perform statistical analyses to identify root causes and develop better treatments for cancer patients. This large-scale data can be effectively analyzed with the help of statistical models, which play a critical role in data interpretation. Statistical models help in quickly and accurately gaining information about the population, often at a lower cost. Once a model is identified, inferences can be drawn from the sample data to understand the broader population trends. Therefore, the development and construction of suitable models are essential for solving complex real-world problems, such as cancer data analysis. Many statisticians have developed various models to address the data from different types of cancer, such as breast cancer, bladder cancer, and leukemia, namely, Al-Kadim and Mahdi (2018) developed the exponentiated transmuted exponential model for analyzing breast cancer survival times, outperforming models like log normal, log logistic, and exponential. Khan et al. (2013) proposed the transmuted inverse Weibull model for bladder cancer data, showing strong performance compared to other models. Kumar et al. (2015) introduced the DUS exponential model, which surpassed the transmuted inverse Weibull and other models. Elbatal et al. (2013) developed the transmuted generalized linear exponential model for leukemia, enhancing the understanding of survival patterns in leukemia patients.

In this discussion, we will explore some of the statistical models developed for analyzing various types of cancer data and their applications. However, it is important to note that not every model is always perfectly suited to real-life phenomena. This is because real-life situations are dynamic and subject to change over time and several inherent factors may influence the outcomes. These factors may include shifts in environmental conditions, advances in medical treatments or changes in patient demographics and lifestyles. As a result, existing models may not always capture the complexities or evolving trends inherent in these phenomena. Given these challenges, there is an ongoing need to update or refine existing models to ensure they accurately reflect current realities. In some cases, this might involve modifying an existing model to account for new factors or changes in the underlying data. In other cases, the development of entirely new models may be necessary to address the limitations of current models and improve their predictive accuracy and applicability. With this motivation in mind, the present study aims to

develop a new statistical model that can effectively analyze data from multiple types of cancer, namely, breast cancer, bladder cancer, and leukemia. The goal is to create a model that not only fits the unique characteristics of these cancer datasets but also outperforms many existing models in terms of different model comparison criteria (e.g., AIC, BIC). By doing so, this study seeks to contribute to more effective analyses of cancer data, leading to better insights into the survival patterns, risk factors, and treatment outcomes for cancer patients across different types of cancer.

In the field of statistical modeling, numerous methodologies have been proposed to create new distributions by modifying or extending an existing baseline distribution. These approaches often involve introducing additional parameters or applying transformations to the baseline distribution, allowing for greater flexibility and adaptability to a wide range of real-world data scenarios. The motivation behind these methods is to capture the complexities and nuances of different types of data that cannot be adequately modeled by standard distributions alone. Some common techniques include applying a parameterized transformation to the cumulative distribution function (CDF) or probability density function (PDF) of the baseline distribution, compounding it with another distribution or incorporating additional shape or scale parameters. Few of them are discussed here; namely, Gupta et al. (1998) proposed a method for generalizing the existing distribution by taking power of the CDF of any baseline probability distribution. Verma et al. (2024) has also proposed a new distribution using the generalization technique.

In recent years, various transformation techniques have been introduced to develop new probability models. The quadratic rank transmutation map (QRTM) technique is widely used but often increases computational complexity by adding parameters; see Shaw and Buckley (2009). In contrast, the DUS transformation technique, introduced by Kumar et al. (2015), enhances baseline distribution flexibility while remaining parsimonious in parameters, reducing estimation complexity. Similarly, the log transformation technique proposed by Maurya et al. (2016) combines parameter parsimony with increased distributional flexibility. These advancements simplify parameter estimation while maintaining robust modeling capabilities.

The primary objective of this article is to introduce a novel probability distribution, termed the log transformed transmuted exponential (LTTE) distribution. This new model is derived using the log transformation technique, which has been recognized for its ability to enhance the flexibility of baseline distributions while maintaining parameter parsimony. Specifically, the LTTE distribution is developed by applying the log transformation to the transmuted exponential distribution, which serves as the baseline. The transmuted exponential distribution (see Owoloko et al. (2015)) is a generalized version of the standard exponential distribution, introduced to provide greater flexibility in modeling data. This distribution is obtained by applying the QRTM to the exponential distribution, thereby adding a single parameter that enhances its ability to capture diverse data behaviors. The CDF and PDF of this distribution are given by

$$G(x;\lambda,\alpha) = (1-e^{-\lambda x})\left(1+\alpha e^{-\lambda x}\right),$$

and

$$g(x;\lambda,\alpha) = \lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right), \quad x \geq 0, \lambda > 0, |\alpha| \leq 1,$$

respectively.

Now by considering the above defined base line distributions, the logarithmic transformed transmuted exponential distribution is given with the following CDF and PDF

$$F(x) = 1 - \frac{1}{\log 2} \log\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right],$$

and

$$f(x) = \frac{\lambda e^{-\lambda x}(1 - \alpha + 2\alpha e^{-\lambda x})}{[2 - (1 - e^{-\lambda x})(1 + \alpha e^{-\lambda x})]\log 2}, \quad x \geq 0, \lambda > 0, |\alpha| \leq 1,$$

respectively. The above proposed distribution is denoted by LTTE(x; α, λ), where α and λ are

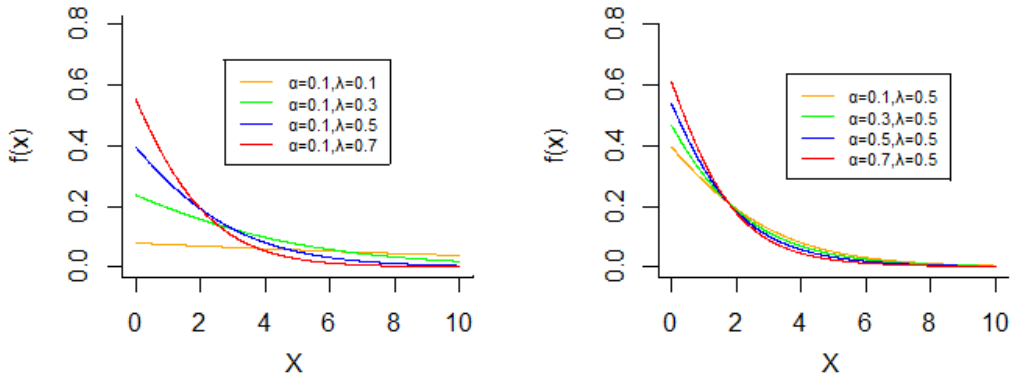


Figure 1: PDF of LTTE for some selected values of λ and α.

the shape and scale parameters, respectively. The shape of the PDF of the proposed distribution is presented in Figure 1. By leveraging the properties of log transformation, the proposed distribution aims to address limitations in existing models, offering improved adaptability to various data sets and practical applications. This approach not only expands the family of transmuted distributions but also contributes to the growing repertoire of tools for statistical modeling and analysis.

The structure of the paper is as follows: Section 1 provides an introduction to the study. Section 2, along with its subsections, explores the distributional properties of the proposed model in detail. Section 3 discusses the parameter estimation using the maximum likelihood estimation (MLE) technique. Section 4 presents simulation studies to evaluate the performance of the estimators. In Section 5, the applicability of the proposed model is demonstrated using three real datasets related to cancer patients. Finally, Section 6 concludes the paper with a summary of the findings and key conclusions.

2 Distributional properties

A new probability distribution is characterized by considering its associated properties. Each of the properties of PDF provides valuable insights and behavior of the random variable it represents. Thus in this section, different distributional properties have been derived for the proposed probability distribution.

2.1 Survival characteristics

In this section, we will discuss the survival function and hazard rate of the proposed model.

- The survival function $S(x)$ is the probability that an equipment/item survived at least time x and it is defined as

$$S(x) = P(X > x) = \frac{1}{\log 2} \log\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right].$$

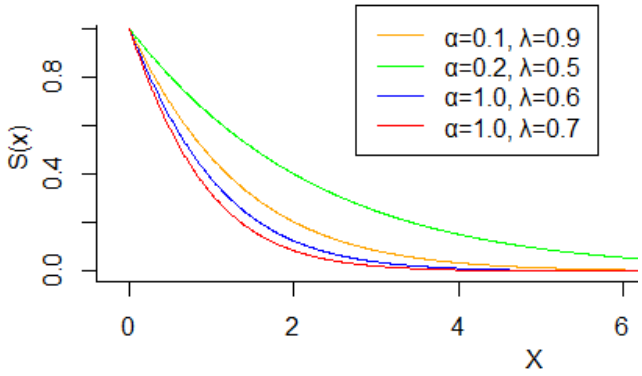


Figure 2: Survival function of LTTE for varying values of α and λ.

- The hazard rate function $h(x)$ is the instantaneous failure rate and it is defined by

$$h(x) = \frac{\lambda e^{-\lambda x}\left(1 - \alpha + 2\alpha e^{-\lambda x}\right)}{\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right] \log\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right]}.$$

- The reverse hazard rate $\tilde{h}(x)$ is obtained as

$$\tilde{h}(x) = \frac{\frac{\lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]\log 2}}{1 - \frac{\log\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]}{\log 2}}.$$

- The cumulative hazard function for LTTE is given by

$$H(x) = -\log S(x) = -\log\left[\frac{\log\left\{2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right\}}{\log 2}\right].$$

The graphical representation of the $S(x)$ and $h(x)$ of the proposed model for varying values of model parameters α and λ are presented in the Figures 2 and 3, respectively. We have examined the nature of hazard for different combinations of model parameters α and λ, and from the graph, it is evident that the model is of increasing hazard nature (except the parameters combination where values of α and λ are very high and it showing nonmonotone hazard rate).

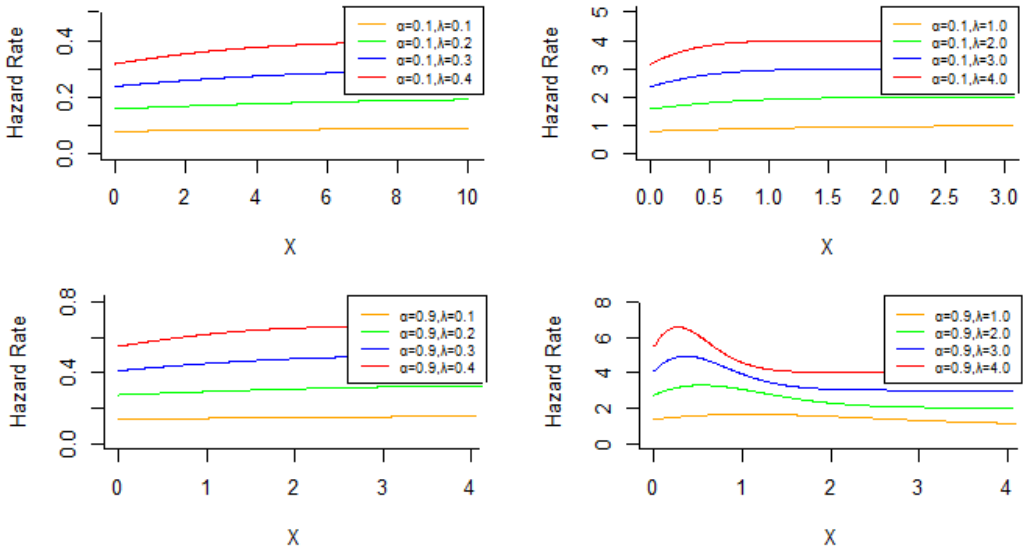


Figure 3: Hazard rate of LTTE for varying values of α and λ.

2.2 Moments

Moments are fundamental properties of any distribution and are widely used to analyze its features and characteristics. The rth moment about the origin for the proposed distribution is expressed as

$$\mu_r' = E(X^r) = \int_0^\infty x^r \frac{\lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]\log 2} dx$$
$$= \frac{\lambda}{\log 2}\sum_{i=0}^{\infty}(-1)^i \frac{\lambda^i}{i!}\int_0^\infty x^{i+r} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]} dx.$$

The respective moments are obtained by putting the values of r.

2.3 Quantile function

The pth quantile function denoted by $Q(p)$ of LTTE $(x;\ \alpha,\ \lambda)$ is obtained by solving

$$F\left[Q(p)\right] = p;$$

and after simplification, the expression for quantile function is given by

$$Q(p) = -\frac{1}{\lambda}\log\left[\sqrt{\left\{\left(\frac{1-\alpha}{2\alpha}\right)^2 - \left(\frac{1-2^{1-p}}{\alpha}\right)\right\}} - \left(\frac{1-\alpha}{2\alpha}\right)\right]. \tag{1}$$

The respective values of quantile can be obtained by putting different values of $p \in (0,1)$ in the above expression for known values of model parameters.

2.4 Skewness and kurtosis

The skewness and kurtosis are commonly employed to analyze the asymmetry and sharpness of a probability distribution. However, their computation often relies on moments, which may not exist for certain distributions. To address these limitations, alternative measures based on the quantile function have been proposed. Notably, Bowley (1920) and Moors (1988) introduced coefficients of skewness and kurtosis that depend on quantile. Bowley's coefficient of skewness, in particular, is defined as follows:

$$B = \frac{Q(3/4) + Q(1/4) - 2Q(1/2)}{Q(3/4) - Q(1/4)}.$$

Similarly, the Moors' coefficient of kurtosis is given by

$$M = \frac{Q(3/8) - Q(1/8) + Q(7/8) - Q(5/8)}{Q(6/8) - Q(2/8)}.$$

Using (1), the coefficients of skewness and kurtosis can be calculated.

2.5 Order statistics

The order statistics are very crucial in statistical analysis as in this case, the analysis of data is performed in ascending or descending order. They are very helpful specially when dealing with extreme observations (e.g., minimum, maximum). Here we will find the expressions of PDFs for 1st, rth, and nth order statistics when the sample follows the proposed distribution.

Suppose that $X_1, X_2, \ldots, X_n$ are a random sample of size n from the proposed distribution and their corresponding order statistics are $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$. The PDF of the rth order statistic $X_{r:n}$, say $f_{r:n}(x)$, for the proposed distribution is given by

$$f_{r:n}(x) = \frac{n!}{(r-1)!\,(n-r)!\,[\log 2]^n} \left[\log \left\{ \frac{2}{\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right]} \right\} \right]^{r-1} \times \lambda e^{-\lambda x} \frac{\left[\log\left\{2 - (1 - e^{-\lambda x})(1 + \alpha e^{-\lambda x})\right\}\right]^{n-r} (1 - \alpha + 2\alpha e^{-\lambda x})}{\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right]}.$$

For $r = 1$ and $r = n$, it simplifies as, respectively,

$$f_{1:n}(x) = \frac{n\lambda e^{-\lambda x}}{[\log 2]^n} \left[\log \left\{ 2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right) \right\} \right]^{n-1} \frac{\left(1 - \alpha + 2\alpha e^{-\lambda x}\right)}{\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right]}$$

and

$$f_{n:n}(x) = \frac{n\lambda e^{-\lambda x}}{[\log 2]^n} \left[\log \left\{ \frac{2}{\left\{2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right\}} \right\} \right]^{n-1} \frac{\left(1 - \alpha + 2\alpha e^{-\lambda x}\right)}{\left[2 - \left(1 - e^{-\lambda x}\right)\left(1 + \alpha e^{-\lambda x}\right)\right]}.$$

2.6 Entropy

Entropy is a measure of the uncertainty or randomness associated with a random variable X characterized by its PDF $f(x)$. It is a significant concept with applications in various fields, including communication, physics and reliability. Among the various measures of entropy, the Rényi entropy, introduced by Rényi (1961), is one of the most widely used. For the proposed distribution, the Rényi entropy is defined as

$$R_\rho = \frac{1}{(1-\rho)} \log \int_{-\infty}^{+\infty} [f(x)]^\rho \, dx \qquad \rho \neq 1.$$

After simplification, the expression for R_ρ for proposed distribution is given by

$$R_\rho = (-1)^i \frac{1}{1-\rho} \log \left\{ \sum_{i=0}^{\infty} \int_0^{\infty} \frac{(x\rho)^i \lambda^{i+\rho}}{i!} \left[\frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\log 2 \left\{ 2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right) \right\}} \right]^\rho dx \right\}, \qquad \rho \neq 1,$$

where ρ is the order of entropy.

2.7 Bonferroni and Lorenz curve

The Lorenz curve and the Bonferroni curve are graphical tools used to analyze inequality in the distribution of resources, such as income or wealth. The Lorenz curve plots the cumulative share of a quantity held by the bottom $x\%$ of the population, highlighting overall inequality, with greater deviations from the diagonal line indicating higher inequality, for more detail see Lorenz (1905). In contrast, the Bonferroni curve focuses on the proportional share of resources held by the lower part of the population, making it more sensitive to changes at the lower end of the distribution; see for more detail Bonferroni (1941). Both curves complement each other in understanding inequality, with the Lorenz curve offering a broader perspective and the Bonferroni curve providing detailed insights into the distribution among the less advantaged. The expression for Lorenz and Bonferroni curves for the proposed distribution are given by

$$L(p) = \frac{1}{\mu} \int_0^p x f(x) \, dx = \frac{\lambda}{\mu \log 2} \int_0^p x e^{-\lambda x} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]} dx$$

and

$$B(p) = \frac{1}{\mu p} \int_0^p x f(x) \, dx = \frac{\lambda}{\mu p \log 2} \int_0^p x e^{-\lambda x} \frac{\left(1-\alpha+2\alpha e^{-\lambda x}\right)}{\left[2-\left(1-e^{-\lambda x}\right)\left(1+\alpha e^{-\lambda x}\right)\right]} dx,$$

respectively.

3 Parameter estimation

Parameter estimation is the process of determining the unknown parameters of a statistical model based on observed data, with the goal of identifying the parameters that best describe the underlying distribution or process. Common methods include MLE, which maximizes the likelihood function; the method of moments, which matches sample moments with theoretical

moments; least squares estimation, which minimizes the sum of squared differences between observed and predicted values. The choice of method depends on the data, model and assumptions about the distribution. Here, the MLE method has been considered for the estimation of the model parameters.

3.1 MLE

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $LTTE(x;\alpha,\lambda)$. Then the likelihood function is given by

$$L(\alpha,\lambda|\mathbf{x}) = \frac{\lambda^n}{(\log 2)^n} e^{-\lambda\sum_{i=1}^n x_i} \frac{\prod_{i=1}^n \left(1-\alpha+2\alpha e^{-\lambda x_i}\right)}{\prod_{i=1}^n \left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right]}.$$

Therefore, the log-likelihood function after ignoring the constant term is given by

$$\log L = n\log\lambda - \lambda\sum_{i=1}^n x_i + \sum_{i=1}^n \log\left(1-\alpha+2\alpha e^{-\lambda x_i}\right) - \sum_{i=1}^n \log\left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right].$$

The maximum likelihood estimates $(\hat{\alpha},\hat{\lambda})$ of the parameter α, and λ can be obtained by differentiating the above equation with respect to α and λ, respectively, and equating them to zero. The following normal equations are obtained:

$$\sum_{i=1}^n \frac{2e^{-\lambda x_i}-1}{\left(1-\alpha+2\alpha e^{-\lambda x_i}\right)} + \sum_{i=1}^n e^{-\lambda x_i} \frac{1-e^{-\lambda x_i}}{\left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right]} = 0 \tag{2}$$

and

$$\frac{n}{\lambda} - \sum_{i=1}^n x_i - 2\alpha\sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{\left(1-\alpha+2\alpha e^{-\lambda x_i}\right)} + \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}\left(1-\alpha+2\alpha e^{-\lambda x_i}\right)}{\left[2-\left(1-e^{-\lambda x_i}\right)\left(1+\alpha e^{-\lambda x_i}\right)\right]} = 0, \tag{3}$$

respectively. The nonlinear equations (2) and (3) are challenging to solve analytically, as they cannot be expressed in closed form. Various methods have been proposed to address such equations, with the Newton–Raphson method being one of the most widely used techniques for iterative and numerical solutions. Using the Newton–Raphson method, the estimates of α and λ (denoted as $\hat{\alpha}$ and $\hat{\lambda}$, respectively) can be obtained by solving these nonlinear equations iteratively.

3.2 Asymptotic confidence interval

Since the explicit distributions of the ML estimators are not available in closed form, thus the asymptotic confidence intervals are constructed in this subsection; see Singh et al. (2014). To achieve this, the Fisher information matrix is derived to facilitate the computation of the asymptotic confidence intervals for the parameters α and λ. The resulting expressions for the Fisher information matrix are provided as follows:

$$I(\alpha,\lambda) = E\begin{bmatrix} -\frac{\delta^2 LogL}{\delta\alpha^2} & -\frac{\delta^2 LogL}{\delta\alpha\delta\lambda} \\ -\frac{\delta^2 LogL}{\delta\lambda\delta\alpha} & -\frac{\delta^2 LogL}{\delta\lambda^2} \end{bmatrix}.$$

All the above derivatives are evaluated at $(\hat{\alpha}, \hat{\lambda})$. The asymptotic variance-covariance matrix of the maximum likelihood estimators is obtained by inverting the Fisher information matrix. The diagonal elements of $I^{-1}(\alpha, \lambda)$ provide the asymptotic variances of α and λ. Using large sample theory, a two-sided $100(1-\beta)\%$ asymptotic confidence intervals for the parameters α and λ are constructed as

$$\hat{\alpha} \mp Z_{\beta/2}\sqrt{\widehat{var}(\alpha)},$$

and

$$\hat{\lambda} \mp Z_{\beta/2}\sqrt{\widehat{var}(\lambda)},$$

respectively, where $Z_{\beta/2}$ is the tabulated value of standard normal distribution at $\beta/2\%$ level of significance. The width of a confidence interval reflects its precision, with narrower intervals indicating more precise estimates. The average width is the mean of all interval widths computed across simulations. Coverage probability measures how often the intervals contain the true parameter value, indicating their reliability. Ideally, intervals should have a small width for precision and a coverage probability close to the nominal level $(1-\beta)$ for accuracy.

3.3 MLE of survival function and hazard function

If $\hat{\alpha}$ and $\hat{\lambda}$ are the maximum likelihood estimates of the parameters α and λ, respectively, then by the invariance property of likelihood estimators, the estimates of the survival function and the hazard function for any mission time $t > 0$ can also be obtained. According to this property, the survival function $S(t)$ and the hazard function $h(t)$, which are functions of α and λ, can be estimated by substituting their maximum likelihood estimates into the respective expressions. Thus, the estimated survival function and the estimated hazard function are given by

$$\hat{S}(x) = \frac{1}{\log 2}\log\left[2 - \left(1 - e^{-\hat{\lambda}x}\right)\left(1 + \hat{\alpha}e^{-\hat{\lambda}x}\right)\right],$$

and

$$\hat{h}(x) = \frac{\hat{\lambda}e^{-\hat{\lambda}x}\left(1 - \hat{\alpha} + 2\hat{\alpha}e^{-\hat{\lambda}x}\right)}{\left[2 - \left(1 - e^{-\hat{\lambda}x}\right)\left(1 + \hat{\alpha}e^{-\hat{\lambda}x}\right)\right]\log\left[2 - \left(1 - e^{-\hat{\lambda}x}\right)\left(1 + \hat{\alpha}e^{-\hat{\lambda}x}\right)\right]},$$

respectively. These estimates provide practical insights into the reliability and risk associated with different mission times.

4 Simulation study

In this section, a Monte Carlo simulation is conducted to evaluate the performance of the proposed point estimators and interval estimates of the parameters. The simulation examines the behavior of the estimators under varying sample sizes and varying model parameter values. Specifically, the sample sizes considered are $n = 30, 60, 90, 120, 150$, and the parameter combinations include $(\alpha = 0.1, \lambda = 0.9)$, $(\alpha = 0.2, \lambda = 0.8)$, $(\alpha = 0.06, \lambda = 0.7)$ and $(\alpha = 0.1, \lambda = 1.0)$. These parameter settings are chosen to cover a wide range of scenarios for assessing the performance of the estimators. The performance of the point estimators is assessed based on their

mean square errors (MSEs), while interval estimators are evaluated using average width (AW) and coverage probability (CP). 95% asymptotic confidence intervals for the parameters are constructed for the same variation, and their corresponding AWs and CPs are also reported. All computations are carried out using the open-source statistical R Software (2024) ensuring transparency and reproducibility of the study. Each combination of sample size and parameter values is examined through 10,000 replications and the results-comprising average estimates (AE), biases and MSEs which are summarized in tabular form (Table 1). The simulation results indicate that the MSEs of the estimators decreases to zero as the sample size n increases, accompanied by a reduction in the widths of the confidence intervals. This behavior demonstrates that the estimators become more precise and their estimated values converge to the true parameter values as the sample size grows. Furthermore, the negligible bias observed across all scenarios supports the conclusion that the proposed estimators are asymptotically unbiased.

Table 2 presents the computed estimates of the survival and hazard functions for varying values of model parameters, demonstrating how the reliability and risk change across different parameters and mission time scenarios.

Table 1: AE, bias, and MSEs along with AW and CPs for the parameters for varying values of α and λ with different sample sizes n.

n	$\alpha = 0.1$					$\lambda = 0.9$				
	AE	Bias	MSE	AW	CP	AE	Bias	MSE	AW	CP
30	-0.0395	-0.1395	0.2692	1.9696	0.8745	0.9852	0.0852	0.0859	1.0690	0.9151
60	0.0450	-0.0550	0.1650	1.5552	0.8680	0.9324	0.0324	0.0449	0.8076	0.9021
90	0.0785	-0.0215	0.1271	1.3574	0.8737	0.9141	0.0141	0.0337	0.6932	0.8939
120	0.0940	-0.0060	0.1077	1.2173	0.8740	0.9061	0.0061	0.0275	0.6169	0.8953
150	0.1037	0.0037	0.0951	1.1213	0.8788	0.9012	0.0012	0.0237	0.5656	0.8959
n	$\alpha = 0.2$					$\lambda = 0.8$				
30	0.0223	-0.1777	0.2758	1.9693	0.8742	0.8980	0.0980	0.0787	0.9997	0.9105
60	0.1140	-0.0860	0.1653	1.5746	0.8686	0.8449	0.0449	0.0404	0.7644	0.8993
90	0.1565	-0.0435	0.1289	1.3731	0.8681	0.8235	0.0235	0.0305	0.6562	0.8868
120	0.1745	-0.0255	0.1077	1.2537	0.8693	0.8147	0.0147	0.0246	0.5941	0.8878
150	0.1888	-0.0112	0.0960	1.1569	0.8696	0.8080	0.0080	0.0213	0.5455	0.8832
n	$\alpha = 0.06$					$\lambda = 0.7$				
30	-0.0663	-0.1263	0.2673	1.9719	0.8718	0.7597	0.0597	0.0490	0.8160	0.9177
60	0.0143	-0.0457	0.1652	1.5440	0.8680	0.7211	0.0211	0.0259	0.6121	0.9037
90	0.0464	-0.0136	0.1276	1.3383	0.8760	0.7078	0.0078	0.0195	0.5221	0.8965
120	0.0593	-0.0007	0.1070	1.2025	0.8770	0.7027	0.0027	0.0158	0.4653	0.8987
150	0.0667	0.0067	0.0933	1.1075	0.8849	0.6997	-0.0003	0.0134	0.4263	0.9020
n	$\alpha = 0.01$					$\lambda = 1.0$				
30	-0.0995	-0.1095	0.2661	1.9585	0.8708	1.0740	0.0740	0.0935	1.1327	0.9200
60	-0.0250	-0.0350	0.1658	1.5259	0.8687	1.0238	0.0238	0.0499	0.8453	0.9073
90	0.0030	-0.0070	0.1263	1.3209	0.8781	1.0074	0.0074	0.0373	0.7199	0.9027
120	0.0150	0.0050	0.1066	1.1766	0.8832	1.0008	0.0008	0.0305	0.6363	0.9033
150	0.0195	0.0095	0.0909	1.0781	0.8929	0.9982	-0.0018	0.0254	0.5802	0.9106

Table 2: Estimate of the reliability and hazard functions for varying values of t for different sample sizes n and model parameters α and λ.

n	t	$\alpha=0.1,\lambda=0.9$		$\alpha=0.2,\lambda=0.8$		$\alpha=0.06,\lambda=0.7$		$\alpha=0.1,\lambda=1.0$	
		$\hat{R}(t)$	$\hat{h}(t)$	$\hat{R}(t)$	$\hat{h}(t)$	$\hat{R}(t)$	$\hat{h}(t)$	$\hat{R}(t)$	$\hat{h}(t)$
30		0.4705	0.8723	0.4925	0.8068	0.5754	0.6357	0.4280	0.9858
60		0.4708	0.8384	0.4918	0.7775	0.5734	0.6155	0.4292	0.9451
90	1.00	0.4698	0.8278	0.4904	0.7689	0.5715	0.6099	0.4286	0.9319
120		0.4690	0.8230	0.4893	0.7650	0.5703	0.6074	0.4280	0.9257
150		0.4684	0.8200	0.4884	0.7627	0.5694	0.6060	0.4276	0.9218
30		0.3796	0.9037	0.4039	0.8331	0.4913	0.6605	0.3361	1.0200
60		0.3817	0.8641	0.4051	0.7980	0.4913	0.6356	0.3390	0.9734
90	1.25	0.3816	0.8509	0.4045	0.7867	0.4901	0.6279	0.3393	0.9574
120		0.3813	0.8446	0.4038	0.7812	0.4893	0.6243	0.3392	0.9498
150		0.3810	0.8406	0.4033	0.7779	0.4887	0.6221	0.3391	0.9448
30		0.3047	0.9280	0.3298	0.8537	0.4175	0.6812	0.2628	1.0455
60		0.3080	0.8845	0.3323	0.8142	0.4191	0.6527	0.2664	0.9950
90	1.50	0.3086	0.8693	0.3324	0.8008	0.4187	0.6433	0.2673	0.9771
120		0.3086	0.8621	0.3322	0.7941	0.4183	0.6389	0.2675	0.9685
150		0.3086	0.8574	0.3320	0.7899	0.4180	0.6361	0.2676	0.9629
30		0.2439	0.9465	0.2686	0.8695	0.3536	0.6983	0.2049	1.0640
60		0.2476	0.9004	0.2718	0.8269	0.3563	0.6672	0.2086	1.0112
90	1.75	0.2486	0.8838	0.2725	0.8119	0.3565	0.6565	0.2098	0.9920
120		0.2489	0.8759	0.2726	0.8042	0.3564	0.6514	0.2102	0.9828
150		0.2491	0.8707	0.2726	0.7994	0.3563	0.6482	0.2105	0.9766

5 Real data applications

In this section, three real datasets related to different types of cancer patients have been utilized to illustrate the practical applicability of the proposed log transformed transmuted exponential distribution (LTTED). The first step involved evaluating whether the selected cancer datasets were appropriate for modeling with the proposed distribution. This suitability assessment was carried out by comparing the performance of the LTTE model with some other widely used lifetime distributions. The comparisons were made using well-established model selection criteria, including the Akaike information criterion (AIC), the Bayesian information icriterion (BIC), the KS statistic and the corresponding p-value obtained from the KS test. The performance of the proposed LTTE model was compared against four competing lifetime distributions: the transmuted exponential distribution (TED), the exponentiated exponential distribution (EED), the Weibull distribution (WD), and the log-exponential distribution (LED). These distributions were chosen due to their popularity and applicability in modeling lifetime data. The model comparison was based on specific criteria: lower values of AIC, BIC, and KS statistic indicated a better fit to the data, while higher p-values from the KS test suggested greater conformity of the data sets to the theoretical distribution. In general, a model demonstrating smaller AIC,

BIC, and KS values alongside a larger p-value was considered the most suitable. To provide a clear understanding, the PDFs of the competing models are detailed below:

- Exponential distribution (see Owoloko et al. (2015)) with PDF

$$f(x,\lambda,\alpha) = \lambda e^{-\lambda x}\left(1-\alpha+2\alpha e^{-\lambda x}\right), \qquad x \geq 0, \alpha, \lambda > 0.$$

- Exponentiated exponential distribution with PDF

$$f(x,\lambda,\alpha) = \alpha\lambda e^{-\lambda x}\left(1-e^{-\lambda x}\right)^{\alpha-1}, \qquad x \geq 0, \alpha, \lambda > 0.$$

- Weibull distribution with PDF

$$f(x,\alpha,\lambda) = \alpha\lambda\,(\lambda x)^{\alpha-1} e^{-(\lambda x)^{\alpha}}, \qquad x \geq 0, \alpha, \lambda > 0.$$

- Log exponential distribution (Maurya et al. (2016)) with PDF

$$f(x,\lambda) = \frac{\lambda e^{-\lambda x}}{\left(1+e^{-\lambda x}\right)\log 2}, \qquad x \geq 0, \lambda > 0.$$

5.1 Data set-I: Breast cancer data

The first dataset represents the survival times of 121 patients diagnosed with breast cancer, collected from a large hospital over the period 1929 to 1938 and is mentioned in Lee (1992). This dataset has been previously analyzed and discussed in studies by Al-Kadim and Mahdi (2018). The dataset is as follows:
0.3, 0.3, 4.0, 5.0, 5.6, 6.2, 6.3, 6.6, 6.8, 7.4, 7.5, 8.4, 8.4, 10.3, 11.0, 11.8, 12.2, 12.3, 13.5, 14.4, 14.4, 14.8, 15.5, 15.7, 16.2, 16.3, 16.5, 16.8, 17.2, 17.3, 17.5, 17.9, 19.8, 20.4, 20.9, 21.0, 21.0, 21.1, 23.0, 23.4, 23.6, 24.0, 24.0, 27.9, 28.2, 29.1, 30.0, 31.0, 31.0, 32.0, 35.0, 35.0, 37.0, 37.0, 37.0, 38.0, 38.0, 38.0, 39.0, 39.0, 40.0, 40.0, 40.0, 41.0, 41.0, 41.0, 42.0, 43.0, 43.0, 43.0, 44.0, 45.0, 45.0, 46.0, 46.0, 47.0, 48.0, 49.0, 51.0, 51.0, 51.0, 52.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 60.0, 60.0, 61.0, 62.0, 65.0, 65.0, 67.0, 67.0, 68.0, 69.0, 78.0, 80.0, 83.0, 88.0, 89.0, 90.0, 93.0, 96.0, 103.0, 105.0, 109.0, 109.0, 111.0, 115.0, 117.0, 125.0, 126.0, 127.0, 129.0, 129.0, 139.0, 154.0

The total time on test (TTT) plot for the considered real dataset is presented in Figure 4. The plot indicates that the data exhibits an increasing hazard rate, which aligns with the hazard rate of the proposed model, making it potentially suitable for modeling such data. Al-Kadim and Mahdi (2018) introduced the exponentiated transmuted exponential (ETE) distribution to analyze this dataset and demonstrated that ETE outperformed three other lifetime models namely lognormal (LN), log-logistic (LL) and exponential distribution (ED) based on its lower AIC (1169.63) and BIC (1178.02). In this study, we further compare the proposed LTTE model with above mentioned four other competing models and observed that the proposed LTTE model exhibits the lowest AIC, BIC, and KS statistic values among the fitted models, including the ETE model, see Table 3. This strongly suggests that the LTTE model provides the best fit for this dataset and can be considered the most appropriate model for analyzing the survival times of breast cancer patients in this context.

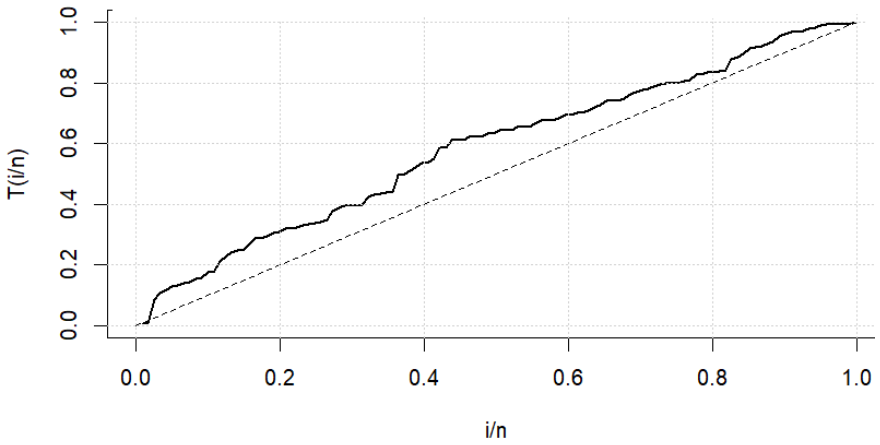


Figure 4: TTT plot of breast cancer dataset.

5.2 Data set-II: Bladder cancer data

The second dataset represents the remission times of 128 bladder cancer patients, extracted from Lee and Wang (2003) and is given by
0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.8,2 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34,,14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69

The TTT plot of the considered dataset (Figure 5) shows that the data is of nonmonotone hazard nature even though from Table 3, it can be seen that our proposed model fits the data very well. Khan et al. (2013) analyzed this dataset using the transmuted inverse Weibull (TIW) distribution, comparing it with the transmuted inverse Rayleigh (TIR), transmuted inverted exponential (TIE), and inverse Weibull (IW) distributions. Based on AIC, BIC, and KS values, they concluded that the TIW distribution provided the best fit. Kumar et al. (2015) proposed the DUS exponential distribution and demonstrated that it outperformed the TIW model, achieving lower AIC (834.044) and BIC (836.896) values. In this study, we extend the comparison to include the proposed LTTE distribution. Table 3 shows that the LTTE model achieves the lowest AIC and BIC values among all the models considered by Khan et al. (2013) and Kumar et al. (2015), establishing it as the best fit for this dataset.

Table 3: The values of the negative of log-likelihood $-\log L$, AIC, BIC, and KS value along with p-values for all the considered data sets.

Models	$-\log L$	AIC	BIC	KS	p-value
Data set-I: Breast Cancer Data					
LTTED	578.943	**1161.885**	1167.477	0.054	0.866
TED	578.978	1161.955	1167.547	0.068	0.638
EED	580.094	1164.187	1169.779	0.080	0.414
WD	579.024	1162.047	1167.639	0.060	0.779
LED	582.319	1166.639	1169.435	0.103	0.153
Data set-II: Bladder Cancer Data					
LTTED	310.064	**624.128**	627.506	0.080	0.382
TED	311.441	626.882	630.260	0.095	0.195
EED	310.156	624.311	627.689	0.073	0.511
WD	322.056	648.111	651.489	0.070	0.557
LED	318.921	639.843	641.532	0.078	0.411
Data set-III: Leukemia Dataset					
LTTED	412.784	**829.568**	835.272	0.185	0.113
TED	413.497	830.995	836.699	0.188	0.105
EED	413.078	830.155	835.859	0.165	0.201
WD	414.087	832.174	837.878	0.307	0.001
LED	414.962	831.923	834.775	0.293	0.002

5.3 Data set-III: Leukemia dataset

The dataset-III is taken from Abouammoh et al. (1994), which represents the ordered lifetimes of 40 patients suffering from leukemia, collected from one of the Ministry of Health hospitals in Saudi Arabia. The data set is
115 181 255 418 441 461 516 739 743 789 807 865 924 983 1024 1062 1063 1165 1191 1222 1222 1251 1277 1290 1357 1369 1408 1455 1478 1549 1578 1578 1599 1603 1605 1696 1735 1799 1815 1852

The TTT plot for this dataset is presented in Figure 6. The plot indicates that the data exhibits an increasing hazard rate, which suggests that models with an increasing hazard rate, such as the proposed LTTE model, may be particularly suitable for accurately representing the underlying data behavior. The results of this comparison, as summarized in the third part of Table 3, reveal that the LTTE model achieves the lowest AIC and BIC values among all the competing models. These lower values indicate that the LTTE model provides the best balance between model complexity and goodness-of-fit for the leukemia data set. Consequently, the LTTE model is determined to be the most appropriate and effective distribution for analyzing this dataset, reinforcing its potential as a robust tool for modeling lifetime data with an increasing hazard rate.

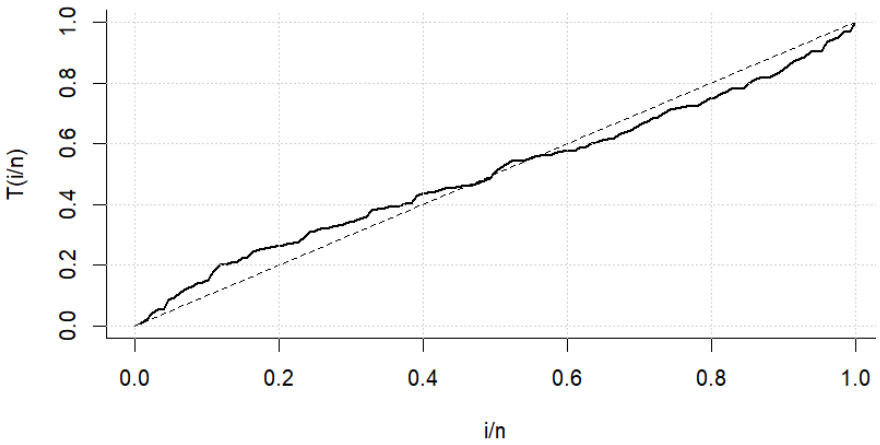


Figure 5: TTT plot of bladder cancer dataset.

6 Summary and conclusions

This research paper formulates a new lifetime probability model, named log transformed transmuted exponential by the extension of transmuted exponential via log transformation. The proposed distribution has an increasing hazard rate. Numerous significant properties of the new distribution are discussed including moments, skewness, kurtosis, order statistics, entropy, quantile function, reliability function and hazard rate. The maximum likelihood estimation procedure is employed to estimate the model parameters. Lastly, we considered three real-life datasets of cancer patients and four other distributions namely transmuted exponential, exponentiated exponential, log exponential and Weibull distributions. It is observed that the proposed LTTE distribution fits the considered datasets very well. The AIC, BIC, and KS-test values illustrate that proposed distribution is better than the above mentioned existing distributions and can be used as an alternate model for the cancer patients datasets.

Acknowledgments

We are grateful to the editors and reviewers for their valuable feedback.

References

Abouammoh, A. M., Abdulghani, S. A., and Qamber, I. S. (1994). On partial orderings and testing of new better than renewal used classes. *Reliability Engineering & System Safety*, **43**, 37–41.

Al-Kadim, K. A., and Mahdi, A. A. (2018). Exponentiated transmuted exponential distribution. *Journal of University of Babylon for Pure and Applied Sciences*, **26**, 78–90.

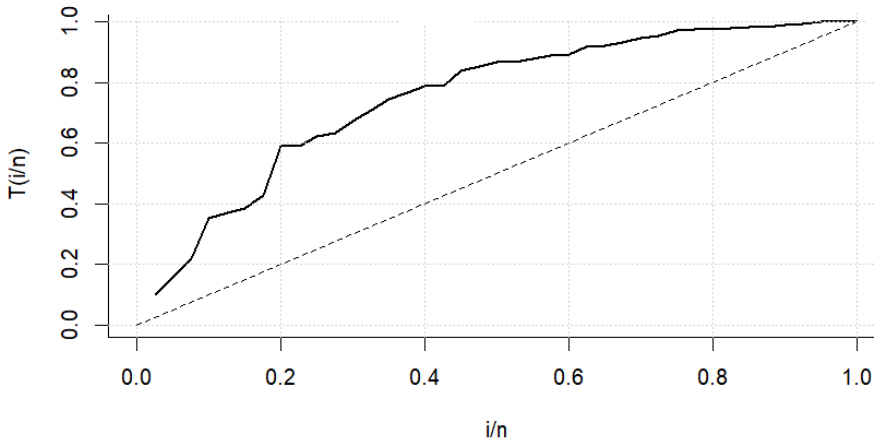


Figure 6: TTT plot of Leukemia dataset

Bhutani, M. A. N. I. S. H. A., Kochupillai, V. I. N. O. D., and Bakshi, S. (2004). Childhood acute lymphoblastic leukemia: Indian experience. *Indian J Med Paediatr Oncol*, **20**, 3–8.

Bonferroni, C. E. (1941). *Elementi di Statistica Generale*. Gili, Torino.

Bowley, A. L. (1926). *Elements of Statistics*. P. S. King & Son Ltd, London.

Elbatal, I., Diab, L. S., and Alim, N. A. (2013). Transmuted generalized linear exponential distribution. *International Journal of Computer Applications*, **83**, 29–37.

Gupta, R. C., Gupta, P. L., and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and Methods*, **27**, 887–904.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., and Thun, M. J. (2007). Cancer statistics, 2007. *CA: A Cancer Journal for Clinicians*, **57**, 43–66.

Khan, M. S., King, R., and Hudson, I. (2013). Characterizations of the transmuted inverse Weibull distribution. *Anziam Journal*, **55**, C197–C217.

Kumar, D., Singh, U., and Singh, S. K. (2015). A method of proposing new distribution and its application to Bladder cancer patients data. *Journal of Statistics Applications and Probability Letters*, **2**, 235–245.

Lee, E. T. (1992) *Statistical Methods for Survival Data Analysis*, 2nd Edition, Wiley Interscience, New York.

Lee, E. T and Wang, J. (2003). *Statistical Methods for Survival Data Analysis*. John Wiley & Sons, New York.

Lorenz, M. O (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**, 209–219.

Maurya, S. K., Kaushik, A., Singh, R. K., Singh, S. K., and Singh, U. (2016). A new method of proposing distribution and its application to real data. *Imperial Journal of Interdisciplinary Research*, **2**, 1331–1338.

Mehrotra, R., and Yadav, K. (2022). Breast cancer in India: Present scenario and the challenges ahead. *World Journal of Clinical Oncology*, **13**, 209.

Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **37**, 25–32.

Owoloko, E. A., Oguntunde, P. E., and Adejumo, A. O. (2015). Performance rating of the transmuted exponential distribution: an analytical approach. *SpringerPlus*, **4**, 818.

Prakash, G., Pal, M., Odaiyappan, K., Shinde, R., Mishra, J., Jalde, D., ... & Bakshi, G. (2019). Bladder cancer demographics and outcome data from 2013 at a tertiary cancer hospital in India. *Indian Journal of Cancer*, **56**, 54–58.

Rényi, A. (1961). On measures of entropy and information. *In Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, Contributions to the Theory of Statistics*, **4**, 547–562.

R Core Team, (2024). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Sathishkumar, K., Chaturvedi, M., Das, P., Stephen, S., and Mathur, P. (2022). Cancer incidence estimates for 2022 & projection for 2025: result from National Cancer Registry Programme, India. *Indian Journal of Medical Research*, Medknow, **156(4& 5)**, 598–607.

Shaw, W. T. and Buckley, I. R. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *arXiv preprint arXiv:0901.0434*.

Singh, S. K., Singh, U., and Yadav, A. S. (2014). Parameter estimation in Marshall-Olkin exponential distribution under type-I hybrid censoring scheme. *Journal of Statistics Applications & Probability*, **3**, 117–127.

Verma, E., Singh, S. K., and Yadav, S. (2025). A new generalized class of Kavya–Manoharan distributions: inferences and applications. *Life Cycle Reliability and Safety Engineering*, **14**, 79–91.

Stochastic Models in Probability and Statistics
Vol. 2, No. 1, 2025, pp. 75–99. Research Article

Bayesian change point inference in time series analysis of COVID-19 pandemic dynamics

Masoud Majidizadeh*

Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

Email(s): tm_majidizadeh@sbu.ac.ir

Abstract. The present study aims to estimate multiple change points in the time series data of confirmed COVID-19 cases and deaths, as well as to assess trends within the identified multiple change points in various countries. The data were analyzed using Poisson time series models that incorporate exogenous variables and autoregressive components, and the estimation of change points was conducted using the reversible jump Markov chain Monte Carlo method. Using the proposed method, we analyze the trajectory of cumulative COVID-19 cases and deaths in these countries, uncovering significant patterns that may have important implications for the effectiveness of pandemic responses across different nations. Furthermore, utilizing a change point detection algorithm in conjunction with a flexible time series model, we apply a forecasting method for COVID-19 and demonstrate its effectiveness in predicting the number of deaths in Japan.

Keywords: Multiple change points, Poisson time series data, Posterior inferences, Reversible jump Markov chain Monte Carlo.

1 Introduction

In this paper, we propose a modeling approach for the time series of confirmed COVID-19 cases and deaths in some countries using a Poisson autoregressive regression model (the formal definition is provided later). Specifically, we aim to model the mean of the infections and deaths as a log-linear model that accommodates an unknown number of potential changes in both the intercept and the slope. This approach is warranted, as it is reasonable to anticipate that the spread of COVID-19 may progress through several distinct phases. Initially, the growth rate is

*Corresponding author
Received: 30 December 2024 / Accepted: 03 May 2025
DOI: 10.22067/smps.2025.91421.1040

typically rapid due to the absence of immunity and insufficient preparedness. Subsequently, the dynamics may transition into phases characterized by slower growth, influenced by government interventions and public health responses aimed at flattening the curve. The estimation of this model can be framed as a change point detection problem.

In recent years, change point analysis has emerged as a vibrant area of research within statistics and econometrics, owing to its diverse applications across various fields. Notable examples include bioinformatics (Fan and Mackey, 2017), climate science (Gromenko et al., 2017), economics (Bai, 1994, 1997; Cho and Fryzlewicz, 2015), finance (Fryzlewicz, 2014), medical science (Chen and Gupta, 2011), and signal processing (Chen and Gu, 2018). Recent reviews on this topic can be found in the works of Perron (2006), Aue and Horváth (2013), and Truong et al. (2020). However, much of the existing literature on change point analysis operates under the assumption of piecewise stationarity. This assumption posits that while the time series in question may be (potentially) nonstationary, it can be segmented into distinct intervals where each segment is stationary and characterized by a common parameter of interest, such as the mean or variance. Although the piecewise stationarity assumption has proven to be effective for many applications, methods developed within this framework are often inadequate for addressing time series with inherent nonstationarity, such as the cumulative infection or death curves of COVID-19.

Following the emergence of the COVID-19 epidemic, several authors have employed change point methods to analyze related data; see, for example, Jiang et al. (2022), ST et al. (2022), Jiang et al. (2023), Dehning et al. (2020), Majidizadeh and Taheriyoun (2024) and Majidizadeh (2024). These methods provide valuable insights into the dynamics of the epidemic, allowing for the identification of significant shifts in trends and patterns over time. By detecting change points, researchers can better understand the impact of various factors, such as government interventions and public health measures, on the progression of the virus.

1.1 The Data

The scope of our analysis encompasses four countries: Iran, Spain, the United States, and Japan. The temporal domains for data segmentation are as follows: first, from April 10, 2020, to October 30, 2021, for the analysis of data in Iran and Spain; second, from April 1, 2021, to November 30, 2021, to examine the effects of vaccination and nonvaccination public health measures in the United States; and third, for short-term forecasting in Japan, the period is May 2020.

During the period from April 10, 2020, to October 30, 2021, both Iran and Spain experienced significant challenges due to the COVID-19 pandemic. This timeframe saw multiple waves of infections, the introduction of various public health measures, and the rollout of vaccination campaigns.

Iran experienced several waves of COVID-19 infections during this period. The first wave began in February 2020, with a significant increase in cases noted in March and April 2020. The second wave peaked in late June and early July 2020, followed by a third wave starting in November 2020, which continued into early 2021. The fourth wave began in April 2021, largely driven by the Delta variant, which became prevalent in mid-2021 (World Health Organization, 2021). The Iranian government implemented various restrictions throughout these waves, in-

cluding lockdowns, travel bans, and the closure of schools and nonessential businesses. For instance, in response to the surge in cases during the third wave, authorities imposed stricter measures in November 2020 and continued to adapt these measures based on the epidemiological situation (Iran Ministry of Health, 2021). Iran began its vaccination campaign in February 2021, initially using the Russian Sputnik V vaccine and later incorporating other vaccines such as Sinopharm and AstraZeneca. By October 2021, the vaccination rate was gradually increasing, but challenges related to vaccine supply and public hesitancy remained (Iranian Red Crescent Society, 2021).

Spain faced several waves of COVID-19 infections, with the first wave peaking in April 2020. The second wave began in late summer 2020, peaking in January 2021. A third wave occurred in March 2021, driven by the emergence of new variants, including the Alpha variant. The Delta variant began to spread in mid-2021, contributing to a fourth wave that peaked in July 2021 (Spanish Ministry of Health, 2021). The Spanish government implemented strict lockdown measures during the initial wave in March 2020, which were gradually eased in the summer. However, restrictions were reintroduced in response to subsequent waves, including curfews, limits on gatherings, and the closure of nightlife venues. In late 2020 and early 2021, measures were adapted based on regional epidemiological data (Government of Spain, 2021). Spain's vaccination campaign began in December 2020, with a rapid rollout of vaccines, primarily the Pfizer-BioNTech and Moderna vaccines. By October 2021, Spain had one of the highest vaccination rates in Europe, with over 80% of the adult population fully vaccinated (European Centre for Disease Prevention and Control, 2021). Figure 1 presents the graphs depicting the total confirmed COVID-19 cases and total deaths in Iran and Spain from April 10, 2020, to October 30, 2021.

From April 1, 2021, to November 30, 2021, the United States faced significant challenges due to the COVID-19 pandemic, marked by multiple waves of infections, the emergence of variants, the implementation of government restrictions, and a nationwide vaccination campaign. The United States experienced several distinct waves of COVID-19 infections during this period. The initial wave peaked in April 2020, with a rapid increase in cases and deaths, particularly in New York and other urban areas. A resurgence of cases occurred in the late summer and fall of 2020, peaking in January 2021. This wave was characterized by increased hospitalizations and deaths, driven by social gatherings and holiday travel (Centers for Disease Control and Prevention, 2021). Following a decline in early 2021, a third wave began in March 2021, driven by the emergence of new variants, particularly the Alpha variant. This wave peaked in late April and early May 2021 (Centers for Disease Control and Prevention, 2021). The Delta variant became the dominant strain in mid-2021, leading to a significant increase in cases during the summer months, particularly among unvaccinated populations. This surge peaked in late July and early August 2021 (World Health Organization, 2021). The vaccination campaign in the United States began in December 2020, with the rollout of the Pfizer-BioNTech and Moderna vaccines. By April 2021, vaccination efforts were expanded to include all adults, and by mid-2021, the vaccination rate significantly increased. By October 2021, approximately 70% of adults had received at least one dose of a COVID-19 vaccine, with over 60% fully vaccinated (Centers for Disease Control and Prevention, 2021). Throughout the pandemic, various government restrictions were implemented to curb the spread of the virus: In March 2020, many states implemented stay-at-home orders and closed nonessential businesses. As cases declined in mid-

2020, many states began to ease restrictions, allowing businesses to reopen with capacity limits and social distancing measures. In response to surges in cases during the fall and winter of 2020, many states reinstated restrictions, including mask mandates and limits on gatherings (National Conference of State Legislatures, 2021). By late 2021, several states and employers began implementing vaccine mandates to encourage vaccination among employees and the public (The White House, 2021).

2 Count time series model

Suppose that $\{Y_t\}$ is a time series of counts and that $\mathscr{F}_{Y,\lambda_t}$ represents the σ-field generated by $\{Y_0,\ldots,Y_t,\lambda_0\}$. Specifically, we define the σ-field as follows: $\mathscr{F}_{Y,\lambda_t} = \sigma(Y_s, s \le t, \lambda_0)$, where σ denotes the smallest σ-algebra generated by the random variables Y_s for $s \le t$ and the constant λ_0. This σ-field captures all the information available up to time t.

Furthermore, we note that the collection of σ-fields $\{\mathscr{F}_{Y,\lambda_t}\}_{t\ge 0}$ forms a filtration. A filtration is an increasing family of σ-fields, which represents the accumulation of information over time. In this context, it reflects how the information about the counts Y_t and the intensity process λ_t evolves as t increases, thus allowing for the modeling of dependencies over time. In the following, we develop a regression model that incorporates exogenous variables and past experiences, expressed by a nonlinear model. Consider the model given by

$$\begin{aligned} Y_t \mid \mathscr{F}_{Y,\lambda_{t-1}} &\sim \text{Poisson}(\lambda_t), \\ \log(\lambda_t) &= \tau(a_1 \log(\lambda_{t-1}) + b_1 \log(Y_{t-1}+1)) + (1-\tau) \\ &\quad \times \left(d + \sum_{i=2}^{p} a_i \log(\lambda_{t-i}) + \sum_{i=2}^{q} b_i \log(Y_{t-i}+1) + \boldsymbol{c}^\top \mathbf{x}_t \right), \end{aligned} \tag{1}$$

for $t \ge 1$, where the parameters d, a_i, and $b_i \in \mathbb{R}$, $\mathbf{x}_t$ is the vector of time-varying covariates and $\boldsymbol{c} \in \mathbb{R}^r$ is the regression coefficient parameters, and let $\boldsymbol{a} = (a_1,\ldots,a_p)^\top$ and $\boldsymbol{b} = (b_1,\ldots,b_q)^\top$. Also, τ is a smoothing parameter $(0 < \tau < 1)$ that limits the changes in λ_t from one time step to the next. In addition, we assume that λ_0 and Y_0 are fixed. We consider a pre-identified smoothing parameter in our study, which ensures that we highly account for the effect of the rate of the parameter in our model. This is particularly important in epidemiological contexts, where rates can have a significant impact on the dynamics of the epidemic.

For the Poisson distribution, the conditional mean is equal to the conditional variance, that is,

$$E[Y_t \mid \mathscr{F}_{Y,\lambda_{t-1}}] = \text{Var}[Y_t \mid \mathscr{F}_{Y,\lambda_{t-1}}] = \lambda_t.$$

Thus, the proposed modeling is based on the evolution of the mean of the Poisson distribution, rather than its variance. For a comprehensive review of count time series models, we refer readers to Weiß (2018) and Fokianos (2012).

Remark 1. *It would be advantageous to treat τ, p, and q as parameters, allowing the observations and methods to estimate them. In this scenario, estimating p and q would require a separate the reversible jump Markov chain Monte Carlo (RJMCMC) method sampling algorithm and incorporate additional complexity into the model. However, for the purpose of our study and to manage computational costs, we assume that these parameters have been identified.*

2.1 A priori assumptions for the model without change point

We assume that $p(\boldsymbol{a}^\top,\boldsymbol{b}^\top,\mathbf{c}^\top,d) = p(\boldsymbol{a}^\top)\times p(\boldsymbol{b}^\top)\times p(\mathbf{c}^\top)\times p(d)$. This independence plays an important role in the computational cost of posterior computation. Let us define $\boldsymbol{\xi}^{(p,q)} = (\boldsymbol{a}^\top,\boldsymbol{b}^\top,\mathbf{c}^\top,d)^\top$. We propose $N(0,\alpha^2 I_k)$ prior for $\boldsymbol{\xi}^{(p,q)}$, where I_k is a $k\times k$ identity matrix and α^2 is variance parameter (hyper-parameter). We consider a uniform prior $\mathscr{U}(0,c_\alpha)$ for α^2, where c_α is a known large value.

Our Bayesian computation is a fusion of Gibbs sampler and RJMCMC, which requires the full conditional distributions. The conditional distribution of $\boldsymbol{\xi}^{(p,q)}$ is

$$p(\boldsymbol{\xi}^{(p,q)},\mid \alpha^2,x,\boldsymbol{y}) \propto f(Y|\boldsymbol{\xi}^{(p,q)})\exp\left(-\frac{1}{2\alpha^2}{\boldsymbol{\xi}^{(p,q)}}^\top\boldsymbol{\xi}^{(p,q)}\right), \tag{2}$$

where $f(Y|\boldsymbol{\xi}^{(p,q)})$ is the likelihood function in (1) and note that

$$p(\alpha^2\mid\boldsymbol{\xi}^{(p,q)}) \propto (\alpha^2)^{-k/2}\exp\Big(-\frac{1}{2\alpha^2}{\boldsymbol{\xi}^{(p,q)}}^\top\boldsymbol{\xi}^{(p,q)}\Big), \quad \alpha^2\in(0,c], \tag{3}$$

which is the truncated inverse gamma distribution, $IG(k/2-1,-\frac{1}{2}{\boldsymbol{\xi}^{(p,q)}}^T\boldsymbol{\xi}^{(p,q)})$ (Majidizadeh, 2024), Majidizadeh and Taheriyoun (2024).

3 Segmentation of count responses

3.1 Model and notations

Consider a Markov chain whose finite-dimensional distributions change at $K-1$ unknown time points, where K is also unknown. Given a partition into K segments, denote the unknown change points by $\boldsymbol{\varepsilon} = (\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_K)^\top$, with $\varepsilon_0 = 0$ and $\varepsilon_K = n$. Here, $K=1$ indicates that there are no change points. Let $\mathscr{Y}_s = \{y_t;\varepsilon_{s-1}+1\le t\le\varepsilon_s\}$ represent the set of all observed values of the response variable in the sth segment for $s=1,\ldots,K$. Our main goal is to estimate the unknown number of change points K, the change points $\boldsymbol{\varepsilon}$, and the corresponding parameters within each segment.

Let $n_s = \#\{t:\varepsilon_{s-1}+1\le t\le\varepsilon_s\}$ for $s=1,\ldots,K$ denote the number of observations in the sth segment. In this model, the observed count responses are partitioned into K segments, where the parameters may differ from those in neighboring segments. The parameters for the sth segment are denoted by the subscript s:

$$\begin{aligned}\log(\lambda_t) &= \tau\times(a_{1,s}\log(\lambda_{t-1,s})+b_{1,s}\log(Y_{t-1,s}+1))+(1-\tau)\\ &\quad\times\left(d_s+\sum_{i=2}^{p}a_{i,s}\log(\lambda_{t-i,s})+\sum_{i=2}^{q}b_{i,s}\log(Y_{t-i,s}+1)+\boldsymbol{c}_s^\top\mathbf{x}_t\right),\end{aligned} \tag{4}$$

and ${\boldsymbol{\xi}^{(p,q)}}_s = (\boldsymbol{a}_s^\top,\boldsymbol{b}_s^\top,\boldsymbol{c}_s^\top,d_s)^\top$, where $s=1,\ldots,K$.

3.2 Model priors

We can also assume that the number of segments K is a priori distributed as a truncated negative binomial distribution given by

$$Pr(K=k) = \frac{1}{c_{r,p,k_{\max}}} \binom{k+r-1}{k} p^k (1-p)^r, \quad k = 1, \ldots, k_{\max},$$

for appropriate choices of parameters r (the number of successes until the experiment is stopped) and p (the success probability), where $c_{r,p,k_{\max}}$ is a normalizing constant. This distribution allows for overdispersion relative to the Poisson distribution and can be useful in modeling scenarios where the variance exceeds the mean. A conservative guideline for $k_{\max}$ is "large enough" but a large value of $k_{\max}$ obviously causes a high computational cost. Employing an expert's idea or preprocessing with other frequentist methods is useful in determining the value of $k_{\max}$. We suggest the following values for the parameters of the truncated negative binomial distribution:

1. $r = 5$, this value provides moderate overdispersion, allowing for variability in the counts.
2. $p = 0.3$, a 30% chance of success in each trial allows for a wider spread in the distribution, appropriate for count data with larger counts being less frequent.
3. $c_{r,p,k_{\max}}$, the normalizing constant ensures that the probabilities sum to 1 over the truncated range of K.

$$c_{r,p,k_{\max}} = \sum_{k=1}^{k_{\max}} \binom{k+r-1}{k} p^k (1-p)^r.$$

This value is computed based on the chosen r, p, and $k_{\max}$. For example, if $k_{\max} = 10$, we would compute $c_{5,0.3,10}$ using the formula above.

Concerning the prior on the locations of change points, we assume that $n_s \geq n_{\min}$ for $s = 1, \ldots, K$, where $n_{\min}$ is the minimum segment length taken to be large enough to avoid sparsity. We further assume that the location of the first change point, ε_1, is a priori distributed according to a uniform distribution over $\{n_{\min}, \ldots, n-(K-1)n_{\min}\}$. This distribution is characterized by a constant probability density function, indicating that all locations within the specified range are equally likely. The prior on the jth change point, ε_j, given ε_{j-1}, is also modeled using a uniform distribution on $\{\varepsilon_{j-1}+n_{\min}, \ldots, n-\varepsilon_{j-1}-(K-j)n_{\min}\}$ for $j = 2, \ldots, K-1$.

3.3 Sampling scheme

Define $\mathbf{E} = (\boldsymbol{\varepsilon}^\top, \boldsymbol{\alpha}^{\mathbf{2}\top}, \boldsymbol{\xi}^{(p,q)^\top})^\top$ as the collection of all parameters, where $\boldsymbol{\alpha}^{\mathbf{2}} = (\alpha^2{}_1, \ldots, \alpha^2{}_K)^\top$ and $\boldsymbol{\xi}^{(p,q)} = (\boldsymbol{\xi}^{(p,q)^\top}_1, \ldots, \boldsymbol{\xi}^{(p,q)^\top}_K)^\top$. Thus, $\mathbf{E}$ has a varying dimension during the algorithm's runs. Each MCMC iteration alternates between two updating steps: the within-model (WM) movements and the switching-model (SM) movements, which are outlined below. The complete algorithm for the comprehensive RJMCMC scheme, which facilitates the detection of change points and the estimation of model parameters, is presented in Algorithm 1. For simplicity, we consider $\boldsymbol{\lambda}_s$ as the first element of $\boldsymbol{\xi}^{(p,q)}{}_s$ for $s = 1, \ldots, K$.

In this algorithm and overall, the prime symbol is used for the *proposed* values and $\mathscr{T}$ is the number of iterations. In what follows and particularly in calculation of the acceptance probabilities, we use functions $p(\cdot)$ for priors and marginal distributions. The arguments of this function discriminate that the density function is calculated for which random variable or vector. Similarly, $p(\cdot \mid \cdot)$ is employed to represent the conditionals and likelihoods and $q(\cdot \mid \cdot)$ for proposal density functions.

3.3.1 Switching-model (SM) movement

We aim to propose new values for the parameters, $(\mathbf{E}', k')$, using the proposal density $q(\mathbf{E}', k' \mid \mathbf{E}, k)$ based on the current parameters $(\mathbf{E}, k)$. This update involves proposing transitions between competing models. The proposed number of segments may either increase by one (birth) or decrease by one (death). We denote k' as the proposed number of segments, which is randomly chosen from $k' = k+1$ or $k' = k-1$ with the following proposal density:

$$q(k' = s \mid k) = \begin{cases} 1/2 & \text{if } s = k+1, \text{ and } k \neq k_{\max},\ n_j \geq 2n_{\min} \text{ for at least one } j, \\ 1/2 & \text{if } s = k-1, \text{ and } k \neq 1, \\ 1 & \text{if } s = k-1 \text{ and } k = k_{\max}, \\ 1 & \text{if } s = k+1 \text{ and } k = 1. \end{cases}$$

Birth $(k' = k+1)$**:**
This step involves creating a new segment by adding an additional change point to the existing set of change points. To this end, a segment, denoted as s^*, is randomly selected from all segments that can be partitioned into two. The new change point in this segment is determined by generating a random number from a uniform distribution over the set $\{\varepsilon_{s^*-1} + n_{\min}, \ldots, n - \varepsilon_{s^*-1} - (K-j)n_{\min}\}$. Let ε'_{s^*} be the generated change point, which must be included in the proposed vector $\boldsymbol{\varepsilon}'$.

The point $\varepsilon_{s^*}{}'$ is chosen to satisfy the conditions $\varepsilon_{s^*}{}' - \varepsilon_{s^*-1} \geq n_{\min}$ and $\varepsilon_{s^*} - \varepsilon'_{s^*} \geq n_{\min}$. Given $\varepsilon_{s^*}{}'$, we need two new hyper-parameters for the variance of the conditional distribution of the coefficients $\boldsymbol{\xi}^{(p,q)'}_{s^*}$ and $\boldsymbol{\xi}^{(p,q)'}_{s^*+1}$ for each new segment. To ensure positivity and simplify acceptance probability calculations, we propose new hyper-parameters $\alpha^{2\prime}_{s^*}$ and $\alpha^{2}_{s^*+1}{}'$, generated using an auxiliary variable $u \sim \mathscr{U}(0,1)$ and deterministic functions of u and $\alpha^2_{s^*}$ as follows:

$$\alpha^{2\prime}{}_{s^*} = \alpha^2{}_{s^*}\left(\frac{u}{1-u}\right)^{\frac{\varepsilon_{s^*+1}-\varepsilon_{s^*}}{\varepsilon_{s^*+1}-\varepsilon_{s^*-1}}}, \quad \alpha^{2\prime}{}_{s^*+1} = \alpha^2{}_{s^*}\left(\frac{1-u}{u}\right)^{\frac{\varepsilon_{s^*}-\varepsilon_{s^*-1}}{\varepsilon_{s^*+1}-\varepsilon_{s^*-1}}}. \tag{5}$$

Two new coefficients are proposed based on these new hyper-parameters. The acceptance probability of the algorithm for this step is Green (1995)

$$\min\Big\{1, (\text{ likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times (\text{Jacobian})\Big\}.$$

Thus, the birth movement is accepted with probability $\min\{1, A_1\}$, where

$$A_1 \;=\; \frac{f(\mathbf{E}' \mid \mathscr{Y}', k')}{f(\mathbf{E} \mid \mathscr{Y}', k)} \frac{p(\mathbf{E}' \mid k')p(k')}{p(\mathbf{E} \mid k)\, p(k)} \frac{q(k \mid k')q(\varepsilon \mid k', k)}{q(k' \mid k)q(\varepsilon' \mid k, k')p(u)} |J|,$$

where $p(u) = 1$, $u \in [0,1]$, and it is straightforward to show that

$$\begin{aligned} \frac{p(k')}{p(k)} &= \frac{p(k+5)}{k+1}, \\ \frac{p(\varepsilon' \mid k')}{p(\varepsilon \mid k)} &= 1, \\ \frac{q(\varepsilon \mid k,k')}{q(\varepsilon' \mid k,k')} &= 1, \end{aligned}$$

where $k' = k+1$. Note that, although we are calculating the current or proposed values in the prior ratio in (6), we have

$$p(k,\mathbf{E}) = p(k,\varepsilon,\alpha^2,\boldsymbol{\xi}^{(p,q)}) = p(k)p(\varepsilon \mid k)p(\alpha^2 \mid \varepsilon,k)p(\boldsymbol{\xi}^{(p,q)} \mid \alpha^2,\varepsilon,k).$$

The denominator in the proposal ratio refers to the conditional density of the proposed number of segments and parameters given the current state, as follows:

$$\begin{aligned} q(k',\mathbf{E}' \mid k,\mathbf{E}) &= q(k' \mid k)q(\mathbf{E}' \mid k,k',\mathbf{E}) \\ &= q(k' \mid k)q(\varepsilon',\boldsymbol{\alpha}^{\mathbf{2}\prime},{\boldsymbol{\xi}^{(p,q)}}' \mid k,k',\mathbf{E}) \\ &= q(k' \mid k)q(\varepsilon' \mid k,k',\mathbf{E})q(\boldsymbol{\alpha}^{\mathbf{2}\prime},{\boldsymbol{\xi}^{(p,q)}}' \mid \varepsilon',k,k',\mathbf{E})q({\boldsymbol{\xi}^{(p,q)}}' \mid \boldsymbol{\alpha}^{\mathbf{2}\prime},\varepsilon',k,k',\mathbf{E}). \end{aligned}$$

After acceptance of this move we update ${\boldsymbol{\xi}^{(p,q)}}'_{s^*}$ and ${\boldsymbol{\xi}^{(p,q)}}'_{s^*+1}$ using the Adaptive Rejection Sampling (ARS) (see Gilks and Wild, 1992).

The determinant of the Jacobian of the transformation between the parameters of the two models, $|J|$, is

$$|J| = \left|\frac{d(\alpha'_{s^*},\alpha'_{s^*+1})}{d(\alpha_{s^*},u)}\right| = \frac{\left(\alpha'_{s^*}+\alpha'_{s^*+1}\right)^2}{\alpha_{s^*}}.$$

Death ($k' = k-1$):

One of the change points, ε_{s^*}, is randomly chosen from the set $\{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_{k-1}\}$ to be removed. After selecting the hyper-parameters ${\alpha^2}_{s^*}$ and ${\alpha^2}_{s^*+1}$, and using the reversing labeling described in (5), the proposed hyper-parameter, denoted as ${\alpha^2}'_{s^*}$, is constructed. This is achieved by reversing the process explained in the Birth step. Once again, due to the lack of a computationally suitable form for the conditional posterior, we update the coefficient vector ${\boldsymbol{\xi}^{(p,q)}}'_{s^*}$ using the random walk Metropolis algorithm. Note that the acceptance probability, $\min\{1,A_2\}$, for the death movement is simply obtained using the inverse of that for the birth and is summarized as

$$A_2 = \frac{f(\mathbf{E}' \mid \mathscr{Y}',k')}{f(\mathbf{E} \mid \mathscr{Y}',k)} \frac{p(\mathbf{E}' \mid k')p(k')}{p(\mathbf{E} \mid k)\,p(k)} \frac{q(k \mid k')q(\varepsilon \mid k',k)}{q(k' \mid k)q(\varepsilon' \mid k,k')p(u)}|J|,$$

where $p(u) = 1$, $u \in [0,1]$, $|J| = (\alpha_{s^*}+\alpha_{s^*+1})^{-2}\alpha'_{s^*}$ and

$$\begin{aligned} \frac{p(k')}{p(k)} &= \frac{k+1}{p(k+5)}, \\ \frac{p(\varepsilon' \mid k')}{p(\varepsilon \mid k)} &= 1, \\ \frac{q(\varepsilon \mid k,k')}{q(\varepsilon' \mid k,k)} &= 1, \end{aligned}$$

where $k' = k-1$ and

$$\alpha'_{s^*} = \left(\alpha_{s^*}\right)^{\frac{\varepsilon_{s^*}-\varepsilon_{s^*-1}}{\varepsilon_{s^*+1}-\varepsilon_{s^*-1}}} \times \left(\alpha_{s^*+1}\right)^{\frac{\varepsilon_{s^*+1}-\varepsilon_{s^*}}{\varepsilon_{s^*+1}-\varepsilon_{s^*-1}}}. \tag{6}$$

After accepting the move we update ${\boldsymbol{\xi}^{(p,q)}}'_{s^*}$ using ARS.

3.3.2 Within-model (WM) movement

This moving scheme involves sampling the parameters of the current model using Metropolis-Hastings updates. In this step, the number of segments remains unchanged, so $k' = k$. A change point is randomly selected for relocation. Specifically, we first select a change point, ε_{s^*}, and then propose a new location from the interval $[\varepsilon_{s^*-1}, \varepsilon_{s^*+1}]$. The corresponding acceptance probability for this move is given by $\min\{1, A_3\}$, where

$$A_3 = \frac{f({\boldsymbol{\xi}^{(p,q)}}'_{s^*}, {\boldsymbol{\xi}^{(p,q)}}'_{s^*+1} \mid \mathscr{Y}'_{s^*}, \mathscr{Y}'_{s^*+1}, k')}{f({\boldsymbol{\xi}^{(p,q)}}_{s^*}, {\boldsymbol{\xi}^{(p,q)}}_{s^*+1} \mid \mathscr{Y}_{s^*}, \mathscr{Y}_{s^*+1}, k, {\alpha^2}_{s^*}, {\alpha^2}_{s^+1*})} \frac{p({\boldsymbol{\xi}^{(p,q)}}'_{s^*} \mid k')p({\boldsymbol{\xi}^{(p,q)}}'_{s^*+1} \mid k')}{p({\boldsymbol{\xi}^{(p,q)}}_{s^*} \mid k')p({\boldsymbol{\xi}^{(p,q)}}_{s^*+1} \mid k')}.$$

Note also that the hyper-parameters ${\alpha^2}'_{s^*}$ and ${\alpha^2}'_{s^*+1}$ are updated via Gibbs sampler, and also ${\boldsymbol{\xi}^{(p,q)}}'_{s^*}$ and ${\boldsymbol{\xi}^{(p,q)}}'_{s^*}$ are updated via ARS.

In our RJMCMC algorithm, we define the stopping time based on the generation of a total of 50,000 pseudo-random samples. The procedure involves the following steps:

- **Total samples:** The algorithm is designed to generate a total of $N_{\text{total}} = 50,000$ samples.
- **Burn-in period:** The first $N_{\text{burn}} = 10,000$ samples are excluded from analysis to allow the Markov chain to reach convergence. These samples are considered the burn-in period and will not be used for parameter estimates.

  The algorithm will stop once the total number of generated samples reaches N_{total}. The valid samples used for inference will thus be the last $N_{\text{total}} - N_{\text{burn}} = 40,000$ samples.

Example 1. To demonstrate the functionality of the presented method in both scenarios—with and without a change point—we generate one realization from the model. In the following simulation studies, we consider $p = q = r = 1$, and also we set $\tau = 0.6$. We construct $x_t = 1 - 0.5 \cdot \mathscr{N}(0,1)$ as the covariate and generate two datasets, each consisting of 300 observations with $\varepsilon_1 = 150$. For these datasets, we set

$$\boldsymbol{\xi}^{(p,q)} = (\mathscr{U}(-0.9, 0.2), \mathscr{U}(-0.5, 0.5), \mathscr{U}(0,1), \mathscr{U}(-0.7, 0.7))^\top + c,$$

where $c = 0$ for dataset 1 and $c = 0.9$ for dataset 2. Therefore, dataset 1 can be effectively considered to have no change point, while dataset 2 exhibits a pronounced change point in the parameters. Figure 2 presents a realization for dataset 2. Figures 3a and 3b display the posterior distributions of the number of change points for the models without and with the change point, respectively. According to Figure 3a, the method demonstrates no false positive performance in models without a change point. Figure 3c shows the estimated posterior distribution of the change point. This represents a single application of the Bayesian method; the average behavior across multiple replications will be explored in the following section.

4 Simulation studies

To assess the performance of our methodology, we apply it to two different simulated data settings.

In the first simulation setting, we examine the estimator for the single change point problem. Accordingly, we consider the following parameters setting for the model defined in Section 3.1:

$$\begin{cases} \boldsymbol{\xi}^{(p,q)}{}_1 = (\mathscr{U}(-1,1), \mathscr{U}(-0.5,0.5), \mathscr{U}(0,0.8), \mathscr{U}(-2,2))^\top, & \text{for} \quad 1 \le x \le 150, \\ \boldsymbol{\xi}^{(p,q)}{}_2 = (\mathscr{U}(1.2,1.5), \mathscr{U}(-1,0), \mathscr{U}(1.2,1.4), \mathscr{U}(-3,-2))^\top, & \text{for} \quad 150 < x \le 300. \end{cases} \tag{7}$$

Using 300 observaosterior samples. The algorithm generates 50,000 pseudo-random numbers, excluding the first 10,000 values as a burn-in period for all parameter estimates. According to Figure 4a, the maximum a posteriori (MAP) estimator of the number of change points is 2 across all replications. Figure 4c illustrates the empirical posterior density of the single change point location, confirming that the method provides an accurate estimate of the change point location at 150. This simulation study focuses on estimation loss, so we do not fix the parameter values. The mean of the empirical squared errors (MESE) of the parameters is presented in Table 1.

We use the Bayesian information criteria (BIC) for model comparison. The resulting average BIC based on the obtained estimates is 1040.175. By eliminating the possibility of a change point in this study, the average BIC increases to 1173.404, confirming the accuracy of the model with change point even though its number of parameters is more than twice that of the model without change point.

A further simulation study is conducted with three change points, using the following parameter settings:

$$\begin{cases} \boldsymbol{\xi}^{(p,q)}{}_1 = (\mathscr{U}(-1,0.5), \mathscr{U}(-0.5,0.8), \mathscr{U}(0.2,0.8), \mathscr{U}(-0.5,0.5))^\top, & \text{for} \quad 1 \le x \le a_1, \\ \boldsymbol{\xi}^{(p,q)}{}_2 = (\mathscr{U}(0.7,1.3), \mathscr{U}(0.8,0.9), \mathscr{U}(-0.9,0), \mathscr{U}(-1.3,-0.7))^\top, & \text{for} \quad a_1 < x \le a_2, \\ \boldsymbol{\xi}^{(p,q)}{}_3 = (\mathscr{U}(-1,0), \mathscr{U}(0.2,0.6), \mathscr{U}(-1.5,-0.5), \mathscr{U}(0,0.7))^\top, & \text{for} \quad a_2 < x \le a_3, \\ \boldsymbol{\xi}^{(p,q)}{}_4 = (\mathscr{U}(0.2,0.6), \mathscr{U}(-1,0), \mathscr{U}(0.1,0.8), \mathscr{U}(1,1.5))^\top & \text{for} \quad a_3 < x \le 1000. \end{cases} \tag{8}$$

The a_i values are randomly chosen from the intervals $[225,275]$, $[475,525]$, and $[725,775]$ for $i = 1,2,3$, thereby randomizing the change point locations. The simulation setting is replicated 200 times, each with 1,000 observations, and we set $k_{\max} = 10$ and $n_{\min} = 50$. A posteriori bar plot of the number of change points is illustrated in Figure 4b, suggesting that the MAP estimate is three in all replicates. The empirical posterior densities are shown in Figure 4d. The estimates of change point locations are generally reliable; in 97.30%, 97.40%, and 98.30% of simulations, the estimated values of ε_1, ε_2, and ε_3 fall within the original intervals, respectively. The mean of the evaluated empirical loss for this setting is provided in Table 2. An independent numeric study is conducted with the same settings but with fixed change points at 250, 500, and 750, with a single replication. The MAP estimates of the change points vector $\boldsymbol{\varepsilon}$ are $(238,529,774)$, while the posterior sample mean is $(246,508,764)$ in this recent simulation study. The bar plot of the number of change point and the empirical density values of the generated change points are shown in Figures 4e and 4f. Based on the BIC, our analysis confirms that the three change point model demonstrates a superior fit compared to both the two change point and four change point models. This finding suggests that the three change point model effectively

balances model complexity and goodness of fit, providing a more parsimonious representation of the underlying data structure. Furthermore, the BIC serves as a robust criterion for model selection, penalizing excessive complexity while rewarding models that adequately capture the data's essential features. The preference for the three change point model indicates that it captures the significant shifts in the data while avoiding overfitting, which can occur in more complex models.

Remark 2. *We also employed the methodologies presented by Ko et al. (2015) and Fearnhead (2006), both of which demonstrated a high level of accuracy in detecting change points and estimating model parameters. However, our findings indicate that our proposed method exhibits slightly superior accuracy in the estimation of these parameters.*

5 COVID-19 data analysis using the proposed method

In this section, we analyze the COVID-19 pandemic using the proposed RJMCMC method and the Poisson time series model. Section 5.1 presents a segmentation of the coronavirus infection and mortality curves across four countries. Building on the proposed algorithm, Section 5.2 evaluates the impact of COVID-19 vaccination and public health measures on the number of confirmed cases and deaths in the USA. Finally, Section 5.3 provides forecasts for deaths in Japan based on the proposed Poisson time series model.

5.1 Segmentation of COVID-19 confirmed cases and deaths

The primary objective of this study is to develop a robust technique for accurately identifying change points in the dynamics of COVID-19 outbreaks. To achieve this, we employed an RJMCMC-based method applied to Model (1) for a comprehensive analysis of data from Iran, and Spain, covering the period from April 10, 2020, to October 30, 2021. This analysis focuses on segmenting daily confirmed cases and COVID-19-related deaths, thereby enhancing our understanding of the evolving patterns of the pandemic across different regions. In this section, we analyze the relationship between the total number of deaths attributed to COVID-19 and various covariates. Specifically, we consider the logarithm of total tests administered (x_1) and the logarithm of total confirmed cases (or alternatively, the total deaths, (x_2) as covariates in our model.

Figure 5 presents the timing of the change points identified by the proposed model in the case and death curves for both Iran and Spain. The following paragraph offers a series of detailed explanations regarding the detected change points and their relationship to various factors, including government restrictions, initiatives by the Ministry of Health, vaccination efforts, and the emergence of different strains of the coronavirus.

The detected change points in COVID-19 deaths in Iran from April 10, 2020, to October 30, 2021, reflect significant shifts in the pandemic's trajectory influenced by public health measures and vaccination efforts. Key dates include May 2, 2020, marking the decline after the initial peak due to lockdowns (World Health Organization, 2020); June 3, 2020, and July 5, 2020, indicating resurgence as restrictions were relaxed (Ministry of Health and Medical Education, Iran, 2020); and October 4, 2020, signaling the onset of a second wave (WHO Iran, 2020). The introduction

of vaccines in December 2020 and the subsequent vaccination campaign beginning January 2021 contributed to a notable decrease in deaths (CNN, 2020); (Al Jazeera, 2021). However, new variants and public fatigue with health measures led to further spikes, particularly in August and September 2021 (The Guardian, 2021). These change points underscore the dynamic nature of the pandemic in Iran, necessitating ongoing public health interventions (Health Ministry of Iran, 2021).

The identified change points in COVID-19 deaths in Spain from April 1, 2021, to November 30, 2021, reflect critical shifts in the pandemic's dynamics influenced by various public health measures and vaccination efforts. The change point on April 24, 2020, marks the beginning of a decline in deaths following strict lockdown measures implemented in March (World Health Organization, 2020). On May 17, 2020, a gradual easing of restrictions led to a resurgence in cases and subsequent deaths, indicating the challenges of reopening (Ministry of Health, Spain, 2020). The change point on August 1, 2020, corresponds to a notable increase in deaths as the summer wave emerged (El País, 2020). Subsequent change points in September and November 2020 reflect the impact of the second wave, exacerbated by increased social interactions and the onset of colder weather (The Lancet, 2020). The December 3, 2020, change point coincides with the approval of vaccines, shifting the focus towards vaccination campaigns (CNN, 2020). The early 2021 change points, particularly January 8 and February 3, highlight the peak of the third wave, leading to significant mortality (The Guardian, 2021). As vaccination efforts ramped up in March and April 2021, deaths began to decline, with further stabilization observed by June (Al Jazeera, 2021). However, new variants and public compliance issues led to fluctuations in deaths through late summer and fall, as indicated by the change points in August and October 2021 (Reuters, 2021). These change points illustrate the complex interplay between public health interventions, seasonal effects, and vaccination strategies in managing the COVID-19 pandemic in Spain.

In Table 3, the parameters of the Bayesian change point model for COVID-19 deaths in Iran exhibit significant variability across the 17 segments, reflecting the dynamic nature of the epidemic. In the initial segments, the parameter a shows both negative and positive values, indicating fluctuations in the influence of previous predicted counts on current counts. The parameter b generally remains positive, suggesting a reinforcing effect of past observed deaths, particularly strong in segments 2, 6, and 15. However, as the segments progress, both parameters demonstrate stabilization, with a approaching zero in segments 10 and 11, while b shows a gradual decline in its influence. The variability in these parameters highlights the changing relationships between past predictions, observed counts, and covariates, emphasizing the model's ability to capture the evolving dynamics of the epidemic in response to public health measures and changing circumstances.

5.2 Impact of public health measures and vaccination on reducing COVID-19 deaths in the USA

In this subsection, we investigate the impact of public health measures and vaccination efforts on the reduction of COVID-19 deaths in the USA . Utilizing the Bayesian change point for Model (1), we aim to identify significant shifts in the trends of confirmed cases and mortality rates in response to various interventions. By analyzing data from this country, we seek to quantify

the effectiveness of these measures over time, providing insights into their role in controlling the pandemic and informing future public health strategies. In this analysis, we examine the total number of deaths attributed to COVID-19 as the response variable. To explore the factors influencing this outcome, we include several covariates: the logarithm of the total number of vaccinations administrated (x_1), the duration of school closures (x_2), and the extent of stay-at-home orders implemented during the pandemic (x_3).

Figure 6 illustrates the detected change points for COVID-19 related deaths in the United States during the specified period. Below, we provide a detailed explanation of the timing and context surrounding these identified change points.

The change points identified from April 1, 2021, to November 30, 2021, illustrate critical transitions in COVID-19 mortality trends in the United States, largely driven by vaccination efforts and public health interventions. The first change point on `2021-04-20` aligns with the expansion of vaccine eligibility, particularly for adults, which significantly increased vaccination rates (Centers for Disease Control and Prevention, 2021). By `2021-05-25`, the CDC updated its guidance to reflect the growing number of vaccinated individuals, allowing for relaxed mask mandates, which likely influenced public behavior and contributed to a decline in deaths (Centers for Disease Control and Prevention, 2021). The change point on `2021-06-19` coincides with the onset of summer, traditionally associated with lower transmission rates, although by `2021-07-15`, the emergence of the Delta variant began to reverse these trends, leading to increased cases and deaths (World Health Organization, 2021). The subsequent change point on `2021-08-19` marked a resurgence in deaths as schools reopened and community transmission increased, while by `2021-09-13`, public health officials noted a concerning rise in hospitalizations and deaths among unvaccinated populations (Centers for Disease Control and Prevention, 2021). On `2021-10-18`, the impact of ongoing vaccination campaigns was evident, yet rising cases persisted, culminating in the change point on `2021-11-09` with the authorization of the Pfizer vaccine for children aged 5-11, marking a significant shift in vaccination strategy aimed at reducing mortality in younger populations (U.S. Food and Drug Administration, 2021).

The parameter estimates from the Poisson time series model for COVID-19 deaths in the U.S. reveal significant variability across the nine segments in Table 4. The parameter a fluctuates from a slight positive value in Segment 1 (0.0012) to a notable negative value in Segment 2 (-0.30231), indicating a shift in the influence of previous predictions on current counts. The parameter b shows a strong positive influence in Segment 1 (1.8623) and an even higher value in Segment 2 (10.1215), reflecting a surge in deaths, but drops significantly in later segments, indicating stabilization. The covariate parameters c_1, c_2, and c_3 exhibit varying effects, with c_2 showing a strong negative impact in Segment 4 (-1.5203), suggesting effective public health measures, while d fluctuates, indicating changes in the baseline death rate. Overall, these changes highlight the complex interactions between vaccination, public health interventions, and COVID-19 mortality dynamics.

5.3 Forecasting

The coronavirus pandemic exhibits a series of distinct epidemic phases, as demonstrated by the nonuniform and fluctuating growth rates of confirmed cases. This variability indicates that any prediction or forecasting model that assumes stationarity and stability within the time

series is likely to be inaccurate. Traditional models that do not account for these fluctuations may fail to capture the underlying dynamics of the epidemic, leading to misleading forecasts and ineffective public health responses. From the perspective of change point detection, a more natural and straightforward approach is to first segment the time series into periods that display relative stability in their behavior. This segmentation allows for the identification of distinct phases of the epidemic, each characterized by its own growth patterns and trends. By isolating these segments, analysts can better understand the factors influencing each phase, such as public health interventions, changes in population behavior, and the emergence of new variants. Following this segmentation, forecasts can be generated based on observations from the most recent segment. This methodology not only enhances the accuracy of predictions but also provides valuable insights into the evolving nature of the pandemic. Moreover, it enables policymakers and public health officials to tailor their strategies to the specific conditions of each phase, ultimately improving the effectiveness of their responses. This approach is supported by the works of Pesaran and Timmermann (2002), Bauwens et al. (2015) and Jiang et al. (2022), which emphasize the importance of recognizing structural changes in the data for more accurate modeling and forecasting in the context of complex and dynamic phenomena such as the coronavirus pandemic.

The method was backtested to generate short-term forecasts of COVID-19 deaths in Japan. In this analysis, daily COVID-19 cases in the country were treated as covariates. Specifically, forecasts were generated for a one-month period, commencing on May 1, 2020. Panel b of Figure 7 displays the actual values alongside the forecasted values for Total COVID-19 deaths in Japan from May 1 to May 30, 2020, as generated by Model 1. Overall, the model provides reasonable short-term forecasts with acceptable accuracy. However, it is important to note that the method is less effective for long-term forecasting, particularly when the data exhibits significant changes. As observed, the forecasting values are highly valid and reliable during the initial days of the month; however, by the end of the month, the forecasts demonstrate a degree of inaccuracy. This discrepancy can be attributed to changes in the behavior of the pandemic. Therefore, it is crucial to consider segmented methods for forecasting and predicting the trajectory of contagious diseases, as they can better account for the dynamic nature of such outbreaks. Panel a of Figure 7 presents the Total deaths in Japan form April 1, 2020 to July 15, 2020, highlighting three detected change points during this period.

6 Conclusion

In this paper, we presented an autoregressive regression time series model tailored for Poisson responses, aimed at analyzing COVID-19 infection and mortality curves across various covariates. We introduced an adapted RJMCMC segmentation algorithm to estimate multiple change points. Our simulation studies and real-world applications illustrated the model's accuracy and its potential to enhance public health decision-making during the COVID-19 pandemic.

Applying our methodology to COVID-19 data from multiple countries, we successfully identified both significant and subtle trends, demonstrating high sensitivity to changes. Notably, the detected change points correlated with public health interventions and vaccination campaigns, indicating that the effects of these factors on disease spread evolved over time. This framework

Table 1: The squared root of the mean of evaluated squared errors of each parameter in (7).

Segment 1		Segment 2	
Parameters	$\sqrt{\text{MESE}}$	Parameters	$\sqrt{\text{MESE}}$
a_1	0.058	a_2	0.9044
b_1	0.022	b_2	0.079
c_1	0.049	c_2	0.178
d_1	0.5908	d_2	0.3035

Table 2: The squared root of the mean of the evaluated squared errors of each parameter in (8).

Segment 1		Segment 2		Segment 3		Segment 4	
parameters	$\sqrt{\text{MESE}}$	parameters	$\sqrt{\text{MESE}}$	parameters	$\sqrt{\text{MESE}}$	parameters	$\sqrt{\text{MESE}}$
a_1	0.025	a_2	0.018	a_3	0.042	a_4	0.010
b_1	0.030	b_2	0.007	b_3	0.072	b_4	0.033
c_1	0.034	c_2	0.016	c_3	0.147	c_4	0.082
d_1	0.133	d_2	0.002	d_3	0.049	d_4	0.110

not only elucidates the drivers of COVID-19 transmission but also aids in formulating effective mitigation strategies.

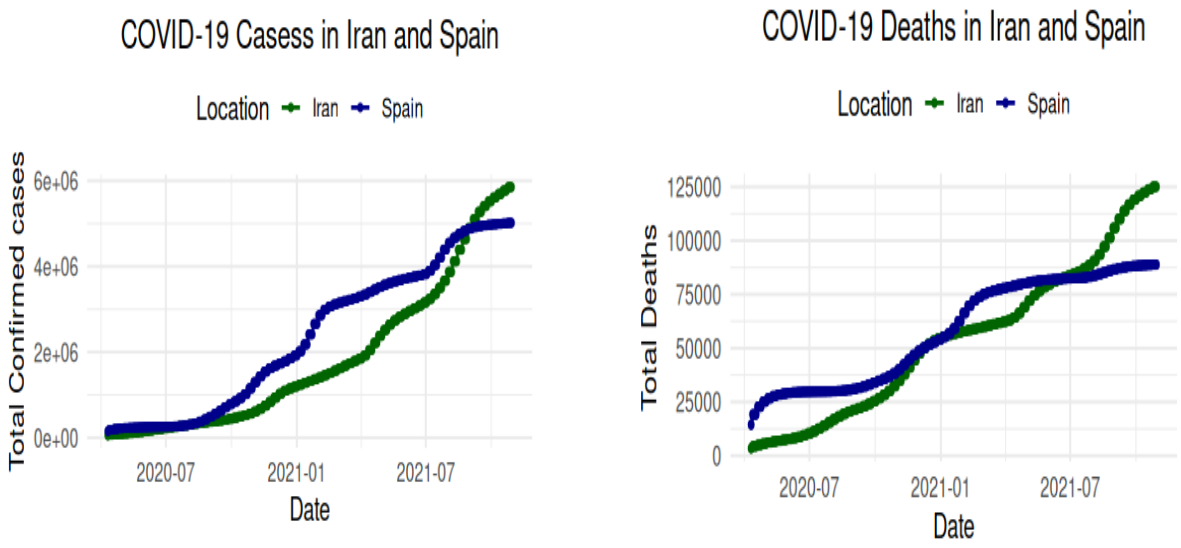
While our approach is based on certain assumptions and limitations related to prior and proposal functions, it relies solely on publicly accessible data. We believe our model enriches the existing literature on COVID-19 by complementing mechanistic models with robust in-sample and out-of-sample predictions. Furthermore, the autoregressive regression framework and RJMCMC method may be applicable to other infectious disease outbreaks characterized by dynamic parameter shifts.

Data availability statement

The COVID-19 data used in this analysis are available in the R-package `COVID19` accessible from `CRAN`. The data is originally sourced from COVID-19 Data Hub; see Guidotti (2022).

Acknowledgements

I would like to express my sincere gratitude to Dr. Alireza Taheriyoun from The George Washington University for his invaluable guidance to this paper.



(a) Total confirmed cases

(b) Total number of deaths

Figure 1: Panel a: Total confirmed cases in Iran and Spain from April 10, 2020, to October 30, 2021. Panel b: Total deaths in Iran and Spain from April 10, 2020, to October 30, 2021.

References

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, **34**, 1–16.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, **15**, 453–472.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, **79**, 551–563.

Bauwens, L., Koop, G., Korobilis, D. and Rombouts, J. V. (2015). The contribution of structural break models to forecasting macroeconomic series. *Journal of Applied Econometrics*, **30**, 596–620.

Chen, J. and Gupta, A. K. (2011). *Parametric Statistical Change Point Analysis: with Applications to Genetics, Medicine, and Finance*. Springer Science & Business Media.

Chen, Y. J. Y. and Gu, Y. (2018). Subspace change-point detection: A new model and solution. *IEEE Journal of Selected Topics in Signal Processing*, **6**, 1224–1239.

Algorithm 1 The Proposed RJMCMC Algorithm

Require: $(\boldsymbol{\xi}^{(p,q)}, \alpha^2, k, \boldsymbol{\varepsilon}, k_{max})$ either randomly or deterministically.

0: **for** $i = 1, \ldots, \mathscr{T}$ **do**
0: **SM movement**- Update the number and the location of change points with proposed k' and $\boldsymbol{\varepsilon}'$ from the proposal densities.
0: $(\boldsymbol{\xi}^{(p,q)}, \boldsymbol{\alpha^2}, k, \boldsymbol{\varepsilon}) \leftarrow (\boldsymbol{\xi}^{(p,q)^{i-1}}, \boldsymbol{\alpha^2}^{i-1}, k^{i-1}, \boldsymbol{\varepsilon}^{i-1})$
0: **while** $i \leq N$ **do**
0: Generate k'
0: **if** $k' = k+1$ **then** {Birth step}
0: Randomly select the segment number s^* to split.
0: Generate $\boldsymbol{\varepsilon}_{s^*}$ at random in the s^*th segment and update $\boldsymbol{\varepsilon}$.
0: Update $\boldsymbol{\alpha^{2'}}$ using the appropriate formula.
0: Compute acceptance probability A_1 and generate $\nu \sim \mathscr{U}(0,1)$.
0: **if** $A_1 \geq \nu$ **then**
0: Generate $\boldsymbol{\xi}^{(p,q)'}_s$, $s = s^*, s^*+1$, by ARS algorithm.
0: Save $\mathbf{E}' := (\boldsymbol{\xi}^{(p,q)'}, \boldsymbol{\alpha^{2'}}, k', \boldsymbol{\varepsilon}')$ and set $i \leftarrow i+1$.
0: **else**
0: Go back to select s^*.
0: **end if**
0: **else**$[k' = k-1]$ {Death step}
0: Randomly choose $\boldsymbol{\varepsilon}_{s^*}$ and remove it from $\boldsymbol{\varepsilon}$ to obtain $\boldsymbol{\varepsilon}'$.
0: $\alpha^{2'}_{s^*} \leftarrow \sqrt{\alpha^2_{s^*}\alpha^2_{s^*+1}}$ and generate $\boldsymbol{\xi}^{(p,q)'}_{s^*}$.
0: Compute acceptance probability A_2 and generate $\nu \sim \mathscr{U}(0,1)$.
0: **if** $A_2 \geq \nu$ **then**
0: Generate $\boldsymbol{\xi}^{(p,q)'}_{s^*}$ by ARS algorithm.
0: Save $\mathbf{E}' := (\boldsymbol{\xi}^{(p,q)'}, \boldsymbol{\alpha^{2'}}, k', \boldsymbol{\varepsilon}')$ and set $i \leftarrow i+1$.
0: **else**
0: Go back to remove $\boldsymbol{\varepsilon}_{s^*}$.
0: **end if**
0: **end if**
0: **end while**
0: **WM movement**- use the data within each segment to generate the updates $\boldsymbol{\xi}^{(p,q)'}$ and $\alpha^{2'}$.
0: Select one of the change points, $\boldsymbol{\varepsilon}_{s^*}$, and propose a new location for it.
0: Update $\boldsymbol{\varepsilon}$ with the new value for $\boldsymbol{\varepsilon}_{s^*}$.
0: Calculate acceptance probability A_3 and generate $\nu \sim \mathscr{U}(0,1)$.
0: **if** $A_3 \geq \nu$ **then**
0: Update $\boldsymbol{\alpha^2}$ via the Gibbs sampler.
0: Update the coefficients $\boldsymbol{\xi}^{(p,q)}$ using ARS.
0: Save $\mathbf{E} := (\boldsymbol{\xi}^{(p,q)'}, \boldsymbol{\alpha^{2'}}, k', \boldsymbol{\varepsilon}')$.
0: **else**
0: Go back to select a new change point.
0: **end if**
0: **end for**=0

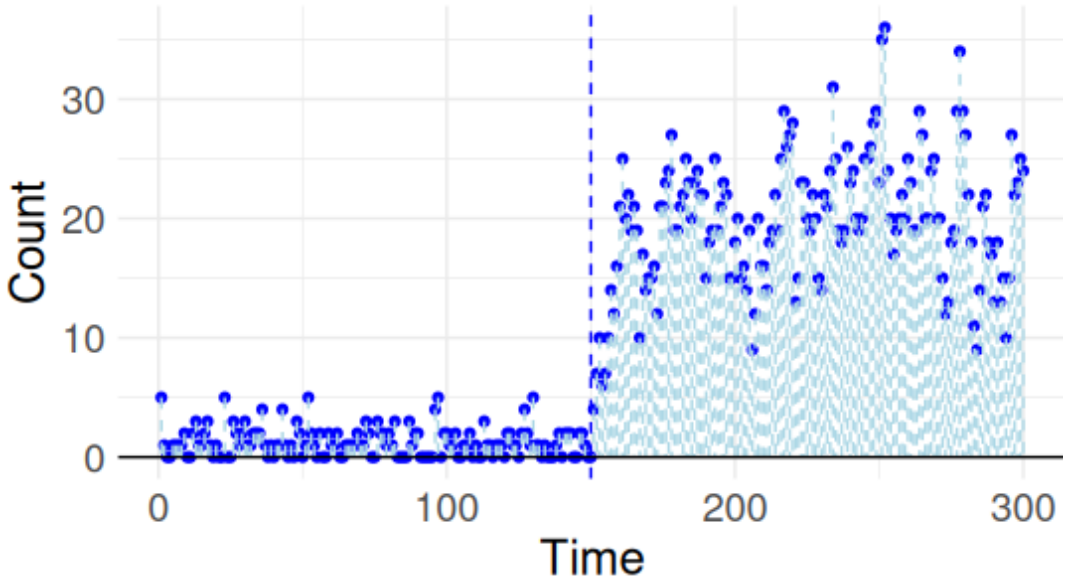


Figure 2: A realization for the dataset 2 in 1.

Cho, H. and Fryzlewicz, P. (2015). Multiple change-point detection for high-dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 475–507.

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J., P., Wilczek, M. and Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, **369**, eabb9789.

Jiang, F., Zhao, Z., and Shao, X. (2023). Time series analysis of COVID-19 infection curve: A change-point perspective. *Journal of Econometrics*, **232**, 1–17.

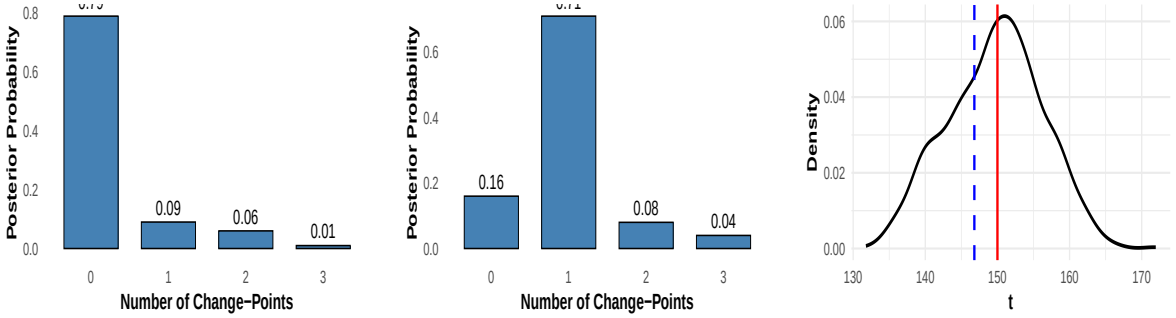
Fan, Z. and Mackey, L. (2017). An empirical bayesian analysis of simultaneous changepoints in multiple data sequences. *The Annals of Applied Statistics*, **11**, 2200–2221.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, **16**, 203–213.

Fokianos, K. (2012) Count time series models. *Handbook of Statistics*, **30**, 315–347.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, **42**, 2243–2281.

Jiang, F., Zhao, Z., and Shao, X. (2022). Modelling the COVID-19 infection trajectory: A piecewise linear quantile trend model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **84**, 1589–1607.



(a) EPRF of K (b) EPRF of K (c) Detected change points

Figure 3: Panel (a) and Panel (b) display the empirical posterior relative frequency (EPRF) of the number of change points, considering $k_{\max} = 4$ and $n_{\min} = 20$, for datasets without and with change points, respectively. Panel (c) illustrates the posterior densities for the location of the change point, where the actual change point and the detected change point are represented by solid and dashed lines, respectively.

Gilks, W. R and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **41**, 337–348.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Gromenko, O. and Kokoszka, P. and Reimherr, M. (2017). Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**, 29–50.

Guidotti, E. (2022). A worldwide epidemiological database for COVID-19 at fine-grained spatial resolution. *Scientific Data*, **9**, 112.

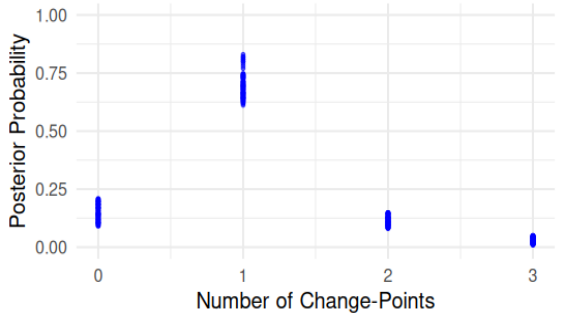
Ko, S. I. M. and Chong, T. T. L. and Ghosh, P. (2015). Dirichlet process hidden Markov multiple change-point model, *Bayesian Analysis*, **10**, 275–296.

Majidizadeh, M. and Taheriyoun, A. R. (2024). Bayesian multiple change-points detection in autocorrelated binary process with application to COVID-19 infection pattern. *Journal of Statistical Computation and Simulation*, **94**, 3723–3749.

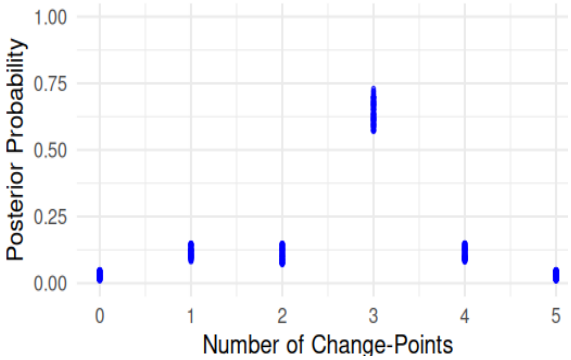
Majidizadeh, M. (2024). Bayesian analysis of the COVID-19 pandemic using a Poisson process with change-points. *Monte Carlo Methods and Applications*, **30**, 449–465.

Perron, P. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics*, **1**, 278–352.

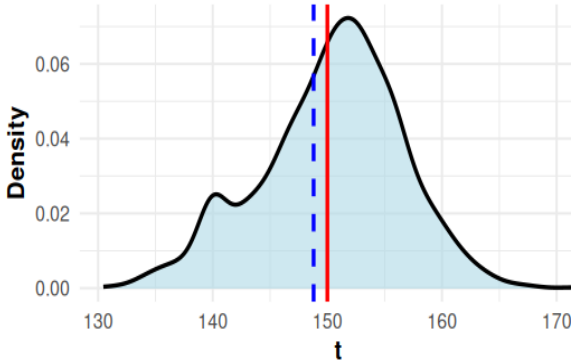
Pesaran, M. H. and Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, **9**, 495–510.



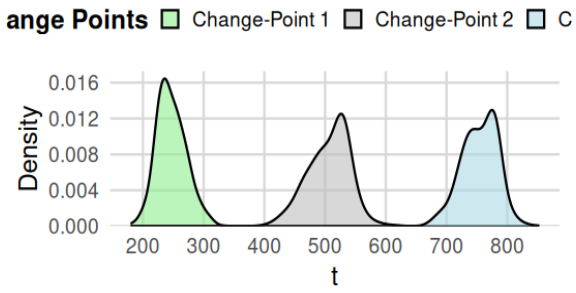
(a) EPRF of K in (7)



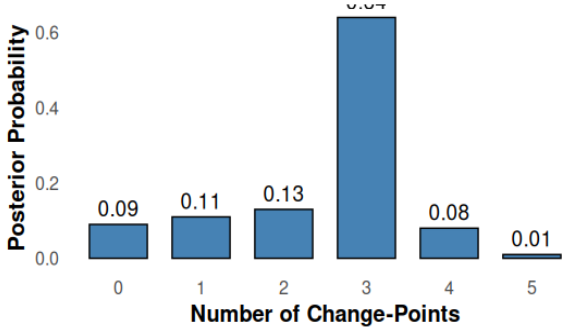
(b) EPRF of K in (8)



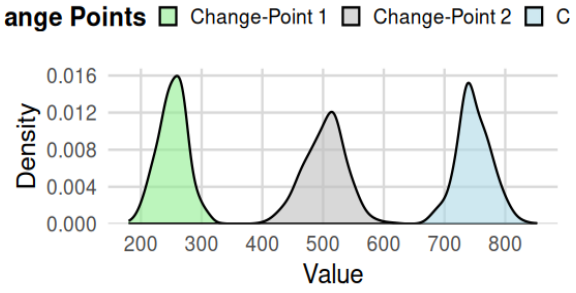
(c) Detected change points



(d) Detected change points for (8)

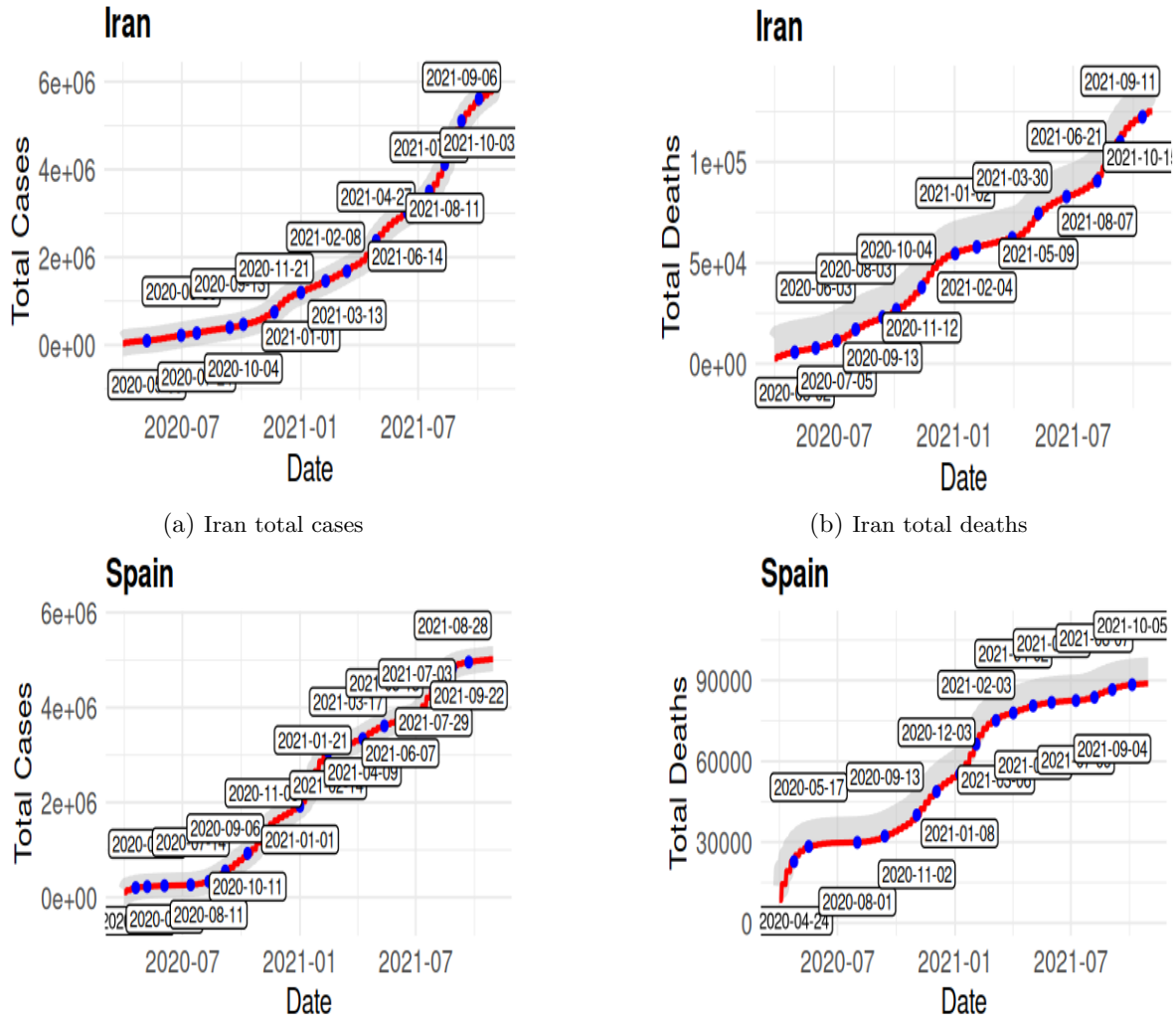


(e) EPRF of K for (8) with constant change points



(f) Posterior change points of (8) with constant change points

Figure 4: Panels (a), (b), and (e): The estimated posterior frequencies of the number of segments for models (7) and (8), and the fixed change point, respectively. Panels (c), (d), and (f): The posterior densities for the location of the change points regarding to model (7), model (8) with random change points and constant change points, respectively.



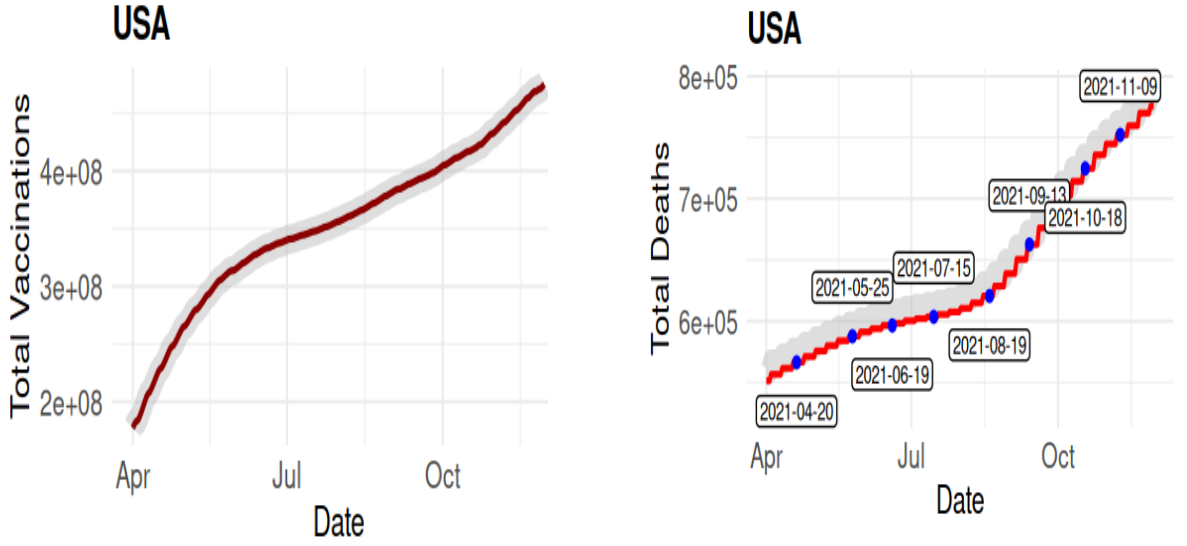
(a) Iran total cases

(b) Iran total deaths

(c) Spain total cases

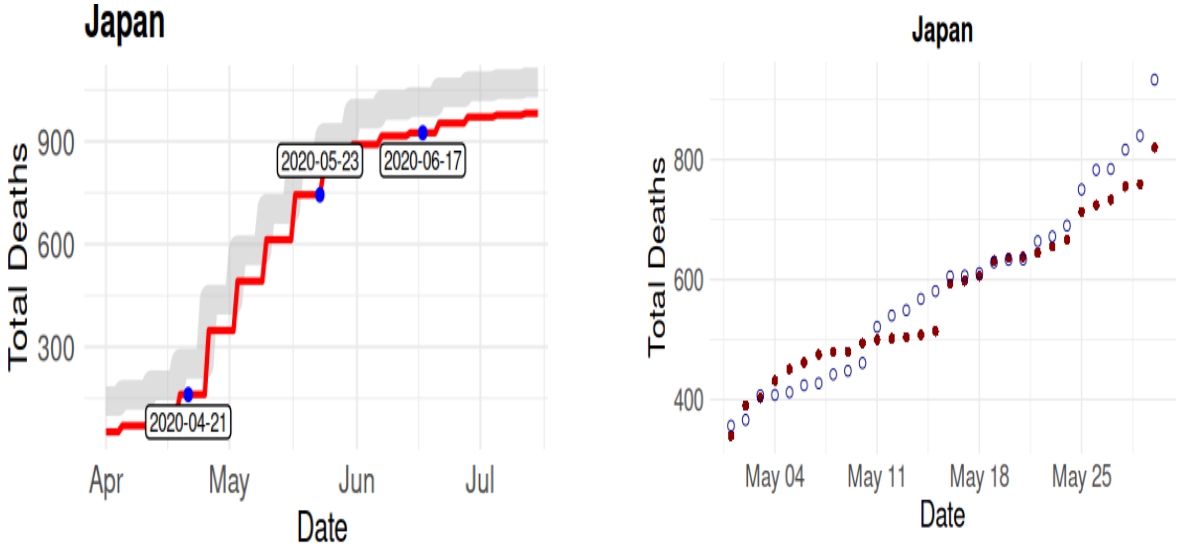
(d) Spain total deaths

Figure 5: The estimated change point locations, based on the Poisson time series model, indicate changes in the confirmed cases and deaths for the total cases in Iran (see Figure a), total deaths in Iran (see Figure b), confirmed cases in Spain (see Figure c), and total deaths in Spain (see Figure d) from April 10, 2020, to October 30, 2021. The points and dates identified in each plot were determined using parameters $k_{\max} = 20$ and $n_{\min} = 20$.



(a) Total vaccination doses administered (b) Total number of deaths

Figure 6: Panel a: Total number of vaccine doses administered in the United States from April 1, 2021, to November 30, 2021. Panel b: Cumulative number of mortality cases during this period, with labels indicating the detected change points based on the parameters $k_{\max} = 5$ and $n_{\min} = 20$.



(a) Total vaccination doses administered (b) Total number of deaths

Figure 7: Panel a: Total deaths in Japan from April 1, 2020, to July 15, 2020, and detected change points. Panel b: Actual (solid points) and forecasted (hollow points) COVID-19 deaths data in Japan from May 1, 2020, to May 30, 2020.

ST, P. K. and Lahiri, B. and Alvarado, R. (2022). Multiple change point estimation of trends in Covid-19 infections and deaths in India as compared with WHO regions, *Spatial Statistics*, **49**, 100538.

Truong, C. and Oudre, L. and Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, **167**, 107299.

Zhang, N. R. and Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, **63**, 22–32.

Weiß, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons, New York.

Table 3: Estimations, including standard errors, of the Poisson time series model applied to the COVID-19 deaths dataset in Iran. This analysis utilizes the logarithm of the total administered tests and the total confirmed cases as covariates. The dataset spans from April 10, 2020, to October 30, 2021, and is divided into 17 segments, with five parameters estimated for each segment.

Segment	a	b	c_1	c_2	d
1	-1.0807 (0.5437)	0.7473 (0.2543)	-0.0011 (0.6207)	0.0015 (0.8405)	0.7703 (0.7890)
2	-0.3087 (0.0746)	1.9667 (0.3354)	0.0034 (0.0470)	0.0014 (0.1890)	0.3283 (1.0526)
3	1.1003 (0.3456)	0.6947 (0.7543)	0.0051 (0.9012)	-0.0070 (0.4001)	-1.3630 (0.2342)
4	-1.4946 (1.0276)	2.1091 (0.6002)	-0.0095 (0.3451)	-0.0162(0.0050)	0.4031 (0.0703)
5	0.4769 (0.9205)	1.1886 (1.020)	0.0011 (0.02)	0.0009 (0.4409)	0.3078 (0.5802)
6	-1.4659 (0.2708)	2.3436 (0.0214)	-0.0046 (0.3882)	0.0010 (0.6634)	-0.5215 (1.03)
7	-1.602 (0.3945)	1.0993 (0.8002)	-0.0079 (0.2500)	0.0012 (0.5432)	1.1756 (0.7365)
8	1.3020 (0.3500)	0.3862 (0.2901)	0.0051 (0.6345)	0.0009 (0.0550)	0.3757 (0.9102)
9	0.2840 (0.3756)	1.4862 (1.0452)	0.0019 (0.1987)	0.00103 (0.7432)	-1.3106 (1.0210)
10	0.00109 (0.0324)	0.6483 (0.5278)	0.0099 (0.7000)	0.0009 (0.0264)	-0.0809 (0.4932)
11	0.0777 (0.8008)	1.0347 (0.6543)	0.0032 (0.8765)	0.013 (0.0902)	-1.8303 (0.1456)
12	0.5987 (0.1267)	1.32054 (0.4534)	0.0025 (0.0278)	0.01390 (0.5789)	-1.8617 (0.0334)
13	-0.4265 (0.0524)	0.5533 (0.7500)	-0.8977 (0.3845)	0.0175 (0.8342)	-0.1452 (0.6500)
14	-0.5959 (0.2109)	0.8422 (0.8300)	-0.0076 (0.7654)	0.0120 (0.3098)	1.3942 (0.6457)
15	-1.1527 (0.0950)	2.6188 (0.9987)	-0.0019 (0.4251)	0.0168 (0.1005)	0.4474 (0.2056)
16	1.7230 (0.4789)	0.8007 (0.2505)	0.0014 (0.5987)	0.0084 (0.3054)	-1.4603 (0.6843)
17	-0.4094 (0.02)	2.1373 (0.02)	-0.0021 (0.02)	0.0079 (0.02)	-0.8996 (0.02)

Table 4: Estimations (including standard errors) of the Poisson time series model for the COVID-19 deaths dataset in the USA, utilizing the logarithm of the total number of vaccinated individuals, along with school closures and stay-at-home measures as covariates. The dataset from April 1, 2021, to November 30, 2021, is divided into nine segments, with six parameters estimated for each segment.

Segment	a	b	c_1	c_2	c_3	d
1	0.0012 (0.01)	1.8623 (0.015)	-0.0017 (0.02)	0.0069 (0.025)	-1.1399 (0.03)	-1.2166 (0.035)
2	-0.30231 (0.02)	10.1215 (0.025)	0.00085 (0.03)	-0.0122 (0.035)	-1.1402 (0.04)	-0.0211 (0.045)
3	-0.0071 (0.015)	0.0904 (0.02)	0.00084 (0.025)	0.0017 (0.03)	-1.0756 (0.035)	-0.0199 (0.04)
4	-0.0012 (0.025)	0.0859 (0.03)	0.00083 (0.035)	-1.5203 (0.04)	0.4031 (0.045)	0.0116 (0.05)
5	-0.0246 (0.02)	0.0969 (0.025)	0.00082 (0.03)	-0.0024 (0.035)	-1.5203 (0.04)	-0.0194 (0.045)
6	-0.7692 (0.04)	1.5803 (0.045)	0.00012 (0.05)	-2.2715 (0.055)	-1.5238 (0.06)	0.1145 (0.065)
7	0.0081 (0.02)	1.5522 (0.025)	-0.0079 (0.03)	0.0183 (0.035)	-0.0474 (0.04)	0.9090 (0.045)
8	-0.0010 (0.01)	1.7845 (0.015)	-0.00008 (0.02)	0.1578 (0.025)	0.1582 (0.03)	-0.9739 (0.035)
9	-0.0334 (0.03)	0.0751 (0.035)	0.00016 (0.04)	0.1563 (0.045)	0.1536 (0.05)	-0.01962 (0.055)

Stochastic Models in Probability and Statistics
Vol. 2, No. 1, 2025, pp. 101–117. Research Article

Varextropy measure with application

Santosh Kumar Chaudhary*

Department of Statistics, Central University of Jharkhand, Cheri-Manatu, Ranchi, Jharkhand, 835222, India

Email(s): skchaudhary1994@gmail.com; santosh.chaudhary@cuj.ac.in

Abstract. In statistical analysis, understanding and quantifying uncertainty is fundamental. Measures such as entropy, extropy, varentropy, and varextropy provide valuable insights into the characteristics of probability distributions. This paper focuses on the concept of varextropy and presents a novel characterization of the uniform distribution, showing that the varextropy of a random variable is zero if and only if the variable is uniformly distributed on the unit interval. Building on this property, we propose a new goodness-of-fit test for uniformity based on a nonparametric estimator of varextropy, denoted by $\widehat{\Delta}$, as introduced by Noughabi and Noughabi (2024). The test statistic is shown to be consistent, and its distribution under the null hypothesis is explored via Monte Carlo simulations. Critical values are tabulated for various sample sizes and tuning parameters, and the test's power is empirically evaluated against alternatives such as the Beta(1,2) distribution, demonstrating superior performance in detecting departures from uniformity. The proposed method is further applied to a real-world environmental dataset of vinyl chloride concentrations, where the transformed data, via the probability integral transform, are shown to conform to a uniform distribution. Overall, this study not only extends the theoretical understanding of varextropy but also introduces a practical and effective tool for uniformity testing in both simulated and real data contexts.

Keywords: Entropy, Extropy, Goodness-of-fit, Monte Carlo simulation, Nonparametric estimator, Order statistics, Uniformity test, Varextropy.

1 Introduction

Quantifying uncertainty in random variables is a central theme in probability theory, information theory, and statistical inference. Various measures have been developed to capture different

*Corresponding author
Received: 11 March 2025 / Accepted: 07 June 2025
DOI: 10.22067/smps.2025.92576.1044

aspects of uncertainty, variability, and information content associated with probability distributions. Among the most widely known and utilized is Shannon's entropy, introduced by Shannon (1948), which provides a foundational measure of the average uncertainty or information content in a random variable. For an absolutely continuous random variable, entropy reflects the expected value of the negative logarithm of its probability density function (pdf) and plays a critical role in areas such as data compression, communication theory, and statistical modeling. Complementing entropy is the concept of *extropy*, proposed by Lad et al. (2015) as a dual measure of uncertainty. While entropy captures average surprise or unpredictability, extropy is designed to assess the regularity and concentration in the distribution of a continuous random variable. Defined in terms of the squared density function, extropy offers a different perspective on information, with applications in decision theory and statistical diagnostics. To understand not just the mean behavior but also the variability of information content, the notion of *varentropy* is employed. Varentropy, introduced by Arıkan (2016), is the variance of the information content (i.e., the log-density). It quantifies the dispersion around the average uncertainty and is particularly useful in finite blocklength information theory, where variability in data coding and transmission must be accounted for. Varentropy has also gained attention in statistical contexts as a more sensitive alternative to classical measures such as kurtosis, especially when analyzing continuous distributions. Building on these concepts, the measure of *varextropy* has recently been introduced to extend the idea of extropy by incorporating variability. Analogous to varentropy, varextropy captures the variance of the density function itself, providing insights into the fluctuation of distribution concentration. This new measure broadens the information-theoretic toolkit for studying distributional properties and can offer useful characterizations of specific distributions, such as the uniform distribution. The interplay between these four measures—entropy, extropy, varentropy, and varextropy—opens up new avenues for theoretical exploration and practical applications, particularly in statistical testing, distribution characterization, and information processing. This paper focuses on the properties and applications of varextropy, particularly in the context of testing for uniformity.

Testing for uniformity is a fundamental problem in statistical analysis with wide-ranging applications across various fields, including quality control, cryptography, simulation, and goodness-of-fit testing. The uniform distribution often serves as a benchmark or null model in many statistical procedures. For example, in simulation studies, ensuring that random number generators produce values that are uniformly distributed is essential for the validity of results. Similarly, in goodness-of-fit testing, the uniform distribution is commonly used to assess whether observed data deviate significantly from a theoretical model. Moreover, many statistical transformations and procedures assume an underlying uniformity, especially in the context of probability integral transforms. As such, reliable tests for uniformity are crucial for validating assumptions, detecting structure in data, and supporting the development of robust statistical methodologies. This motivates the exploration of new approaches, such as those based on information-theoretic measures like varextropy, to enhance the sensitivity and applicability of uniformity tests.

Although varextropy is a relatively recent addition to the family of information-theoretic measures, it has begun to draw interest for its potential applications in characterizing probability distributions. Previous studies have explored the mathematical properties of varextropy and demonstrated its sensitivity to distributional shape and concentration. However, its use in formal hypothesis testing, particularly for assessing uniformity, remains limited in the literature.

Existing uniformity tests are primarily based on classical approaches such as the Kolmogorov–Smirnov test, Cramér-von Mises criterion, and entropy-based methods. In contrast, this work introduces a novel test procedure that leverages the variance of the squared density, *varextropy*, as a means to detect deviations from the uniform distribution. By establishing a new characterization of the uniform distribution through varextropy, we extend its utility beyond descriptive analysis and into inferential statistics. Our method differs from earlier work in that it provides a nonparametric, information-theoretic framework for uniformity testing, offering a potentially more sensitive alternative to traditional approaches. Furthermore, we evaluate the effectiveness of the proposed test through both theoretical derivations and empirical analyses using real-world data, thereby demonstrating its practical relevance. The entropy of a discrete probability distribution $P = \{p_1, \ldots, p_n\}$ is defined as (Shannon, 1948) $H(P) = -\sum_{i=1}^{n} p_i \ln p_i$. The varentropy of a discrete probability distribution $P = \{p_1, \ldots, p_n\}$ is defined as (see Arıkan (2016); De Crescenzo et al. (2025); Maadani et al. (2022))

$$VH(P) = \sum_{i=1}^{n} p_i (\ln p_i)^2 - \left(\sum_{i=1}^{n} p_i \ln p_i \right)^2.$$

Varentropy serves as a measure of the variability in the information content.
Lad et al. (2015) introduced the concept of *extropy*, which is the complement of Shannon entropy. The extropy of a discrete probability distribution $P = \{p_1, \ldots, p_n\}$ is defined as

$$J(P) = -\sum_{i=1}^{n} (1 - p_i) \ln(1 - p_i).$$

Let X be an absolutely continuous random variable with common cumulative distribution function (cdf) F_X and probability density function (pdf) f_X. Let $l_X = \inf\{x \in \mathbb{R} : F_X(x) > 0\}, u_X = \sup\{x \in \mathbb{R} : F_X(x) < 1\}$ and $S_X = (l_X, u_X)$. Then, Shannon (1948) defined differential entropy as a measure of uncertainty

$$H(X) = -\int_{S_X} f_X(x) \log f_X(x)\, dx.$$

Varentropy of X is defined as (Arıkan, 2016; Maadani et al., 2022)

$$\begin{aligned} VH(X) &= \mathrm{Var}[-\log f_X(X)] \\ &= \int_{S_X} f_X(x) (\log f_X(x))^2\, dx - \left(\int_{S_X} f_X(x) \log f_X(x)\, dx \right)^2. \end{aligned}$$

This varentropy measure is widely used in data compression, finite blocklength information theory, and statistics, as it aids in determining ideal code lengths, source dispersion, and other relevant quantities. In statistics, it has proven to be a superior alternative to the kurtosis measure for continuous density functions; see (Arıkan, 2016; Dudewicz and van der Meulen, 1981; Hazeb et al., 2021; Maadani et al., 2022) studied entropy- and extropy-based goodness-of-fit tests for uniformity. An alternative measure of uncertainty, *extropy*, for a nonnegative absolutely continuous random variable X, defined by Lad et al. (2015), is given by

$$J(X) = \mathbb{E}\left(-\frac{1}{2} f_X(X) \right) = -\frac{1}{2} \int_{S_X} f_X^2(x)\, dx.$$

The primary objective of this study is to investigate the properties of varextropy and demonstrate its potential in testing for the uniformity of continuous probability distributions. Specifically, we aim to derive and explore the theoretical properties of varextropy, provide a characterization of the uniform distribution based on this measure, and develop a nonparametric estimator for varextropy from observed data. Additionally, we propose a novel test for uniformity that leverages varextropy, extending its applicability beyond descriptive analysis into inferential statistics. Finally, we evaluate the performance of the proposed test through both theoretical analysis and empirical validation using real-world data to assess its effectiveness in detecting deviations from the uniform distribution.

The main purpose of this paper is to obtain a test of uniformity using the derived characterization of the uniform distribution based on the varextropy of a continuous random variable. This paper is organized as follows. Section 2 contains some properties of varextropy. A characterization of the uniform distribution using varextropy is given in Section 3. A nonparametric estimator is given in Section 4. A test of uniformity is presented in Section 5, and Section 6 contains an application to real data.

2 Varextropy

The varextropy of a discrete probability distribution $P = \{p_1, \ldots, p_n\}$ is defined as (see, Goodarzi (2024); Vaselabadi et al. (2021))

$$VJ(P) = \sum_{i=1}^{n} (1-p_i)\left(\ln((1-p_i))\right)^2 - \left(\sum_{i=1}^{n} (1-p_i)\ln((1-p_i))\right)^2.$$

Varextropy also serves as a measure of the variability in the information content. Varextropy of absolutely continuous random variables X is defined as

$$\begin{aligned} VJ(X) = \mathrm{Var}\left(-\frac{1}{2}f_X(X)\right) &= E\left(-\frac{1}{2}f_X(X) - J(X)\right)^2 \\ &= \frac{1}{4}E(f_X^2(X)) - \frac{1}{4}[E(f_X(X))]^2 \\ &= \frac{1}{4}\int_{S_X} f_X^3(x)\,dx - \frac{1}{4}\left(\int_{S_X} f_X^2(x)\,dx\right)^2. \end{aligned}$$

Note that $VJ(X) \geq 0$, for any random variable X. Vaselabadi et al. (2021) obtained several varextropy properties, as well as conditional varextropy properties based on order statistics, record values, and proportional hazard rate models. The article contains some comparative results regarding varextropy and varentropy. Goodarzi (2024) provided lower bounds for varextropy, obtained the varextropy of a parallel system, and used the varextropy of order statistics to construct a symmetry test. Zaid et al. (2022) computed the entropy, varentropy, and varextropy measures in closed form for generalized and q-generalized extreme value distributions. Varentropy is sometimes independent of the model parameters, whereas the varextropy measure is more adaptable, for example, when X has a normal distribution with mean μ and variance σ^2 (see Vaselabadi et al. (2021)).

Chacko and Grace (2024) investigated the varextropy measure for the nth upper and lower k-record values, deriving expressions for both the measure and its residual and past forms. They applied this to estimate the varextropy of a two-parameter Weibull distribution using maximum likelihood estimations (MLEs) and Bayes estimates based on upper k-record values, with MCMC used for the Bayes estimates. Their simulation results showed that mean squared errors (MSEs) decreased as n increased, and Bayes estimates outperformed MLEs. Among the Bayes estimators, those using the SEL function performed better, and the lowest MSE was achieved using Prior 1. Goodarzi (2022) derived the conditional covariance and variance for a parallel system with n identical, independent components, assuming all components are still functioning at time x. A lower bound for the conditional variance was also provided. Additionally, lower bounds for varextropy were established, and the varextropy of a parallel system was calculated. The results were applied to create a symmetry test, with a real dataset used to illustrate the test statistics. Vaselabadi et al. (2021) explored several properties of the varextropy measure VJ, highlighting its use in quantifying information volatility in residual and past lifetimes. They examined its behavior in relation to order statistics, record values, and proportional hazard rate models. An approximate expression for $VJ(X)$ was also derived using a Taylor series expansion. Additionally, they introduced the concept of conditional varextropy and proposed a new stochastic order called varextropy ordering. Noughabi and Noughabi (2024) investigated the varextropy of a random variable and introduced consistent estimators for it, highlighting their location-invariant variance and mean squared error. Through Monte Carlo simulations, they evaluated the estimators' bias and RMSE under different distributions, showing that the proposed methods performed reliably across various scenarios.

In some situations, two random variables can have the same extropy, which prompts the age-old question, "Which of the extropies is a more appropriate criterion for measuring the uncertainty?" For example, consider random variables U and V (see, Balakrishnan et al. (2020)) with pdfs

$$f_U(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise}, \end{cases} \qquad \text{and} \qquad f_V(x) = \begin{cases} 2e^{-2x}, & x > 0, \\ 0, & \text{otherwise}. \end{cases}$$

We get $J(U) = J(V) = -1/2$, $VJ(U) = 0$, and $VJ(V) = 1/12$. This is the motivation behind considering the variance of $-\frac{1}{2}f(x)$, which is known as the varextropy of a random variable X. So, varextropy can also play a role in measuring uncertainty. The varextropy for some standard distributions are given in Table 1; for more example, see Vaselabadi et al. (2021).

Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed observations. An observation X_j will be called an upper record value if its value exceeds that of all previous observations. Thus, X_j is an upper record if $X_j > X_i$ for every $j > i$. See, Arnold et al. (1998) for more details about record values. A random variable X is said to be smaller than Y in the dispersive ordering ($X \leq_{disp} Y$) if $F_Y^{-1}(F_X(x)) - x$ is increasing in $x \geq 0$. Belzunce et al. (2001) showed that if $X \leq_{disp} Y$, then $U_n^X \leq_{disp} U_n^Y$, where U_n^X and U_n^Y are the nth upper records of X and Y, respectively. Qiu (2017) showed that if $X \leq_{disp} Y$, then $J(X) \leq J(Y)$ and $J(U_n^X) \leq J(U_n^Y)$. Vaselabadi et al. (2021) showed that if $X \leq_{disp} Y$, then $VJ(X) \geq VJ(Y)$. In view of these results, it is conclude that $X \leq_{disp} Y$, then $VJ(U_n^X) \geq VJ(U_n^Y)$, for $n \geq 1$. It is obvious that if X and Y are identically distributed, that is, $X \stackrel{d}{=} Y$, then $VJ(X) = VJ(X)$, $VJ(X_{i:n}) = VJ(Y_{i:n})$ and $VJ(U_n^X) =$

Table 1: Expression for $VJ(X)$.

Distribution	pdf	$VJ(X)$
Uniform	$\frac{1}{b-a}$, $a<x<b$	0
Exponential	$\lambda e^{-\lambda x}$, $x\geq 0, \lambda>0$	$\lambda^2/48$
Weibull distribution	$2xe^{-x^2}$, $x>0$	$\frac{\sqrt{\pi}}{2^{3/2}}-\frac{\pi}{8}$
Normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{(-\frac{(x-\mu)^2}{\sigma^2})}$, $-\infty<x<\infty$	$\frac{2-\sqrt{3}}{16\pi\sigma^2\sqrt{3}}$
Laplace distribution	$\frac{1}{2}e^{-\lvert x\rvert}$ $-\infty<x<\infty$	$\frac{1}{24}$
Logistic distribution	$\frac{e^{-x}}{(1+e^{-x})^2}$, $-\infty<x<\infty$	$\frac{1}{8}$
Cauchy distribution	$\frac{1}{\pi(1+x^2)}$, $-\infty<x<\infty$	$\frac{1}{8\pi}-\frac{1}{16\pi^2}$

$VJ(U_n^Y)$, where $X_{i:n}$ is the ith order statistic in a random sample of size n. Vaselabadi et al. (2021) showed that varextropy is location-invariant but not scale-invariant, that is, if $Y=aX+b$, where $a>0$ and $-\infty<b<\infty$, then $VJ(Y)=\frac{1}{a^2}VJ(X)$.

We have the following result for varextropy of order statistics of symmetric distribution.

Lemma 1. *Let $X_1,X_2,\ldots,X_n$ be random sample from continuous distribution with symmetric around a finite μ with sample size n. Then*

$$VJ(X_{i:n})=VJ(X_{n-i+1:n}),\quad 1\leq i\leq n.$$

Proof. The result follows by location-invariant property of varextropy. □

3 Weighted varextropy

Applications of weighted distributions include distribution theory, dependability, probability, ecology, biostatistics, and applied statistics. Two random variables can have the same extropy as well as the same varextropy in some situations. For example, consider random variables X and Y with pdfs, respectively:

$$f_X(x)=\begin{cases}2x, & 0<x<1,\\ 0, & \text{otherwise},\end{cases}\qquad f_Y(x)=\begin{cases}2(1-x), & 0<x<1,\\ 0, & \text{otherwise}.\end{cases}$$

We get $J(X)=J(Y)=-2/3$, $VJ(X)=VJ(Y)=1/18$, but $VJ^w(X)=1/12$ and $VJ^w(Y)=1/180$. So here, weighted varextropy can also play a role as a measure of uncertainty. Gupta and Chaudhary (2023) defined general weighted extropy with nonnegative weight $w(x)$ as

$$J^w(X)=-\frac{1}{2}\int_0^\infty w(x)f_X^2(x)\,dx.$$

Table 2: Expression for $VJ^x(X)$.

Distribution	pdf	$VJ^x(X)$
Uniform	$\frac{1}{b-a}, \quad a<x<b$	$\frac{1}{4}\left[\frac{b^3-a^3}{3(b-a)^3}-\frac{(b^2-a^2)^2}{4(b-a)^4}\right]$
Exponential	$\lambda e^{-\lambda x}, \quad x\geq 0, \lambda>0$	$\frac{5}{1728}$
Weibull distribution	$2xe^{-x^2}, \quad x>0$	$\frac{1}{4}\left(\frac{1}{3^{3/2}}-1\right)$
Normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{\left(-\frac{(x-\mu)^2}{\sigma^2}\right)}, \quad -\infty<x<\infty$	$\frac{1}{4}\left(\frac{1}{(2\pi\sigma^2)^{3/2}}\left(\mu^2+\frac{1}{6\sigma^2}\right)-\mu^2\right)^2$
Laplace distribution	$\frac{1}{2}e^{-\lvert x\rvert} \quad -\infty<x<\infty$	$\frac{1}{216}$
Logistic distribution	$\frac{e^{-x}}{(1+e^{-x})^2}, \quad -\infty<x<\infty$	$\frac{1}{8}$
Cauchy distribution	$\frac{1}{\pi(1+x^2)}, \quad -\infty<x<\infty$	$\frac{1}{16\pi^2}$

We define the general weighted varextropy of a discrete probability distribution $P=\{p_1,\ldots,p_n\}$ with $X=\{x_1,x_2,\ldots,x_n\}$ and weights $w=\{w_1,w_2,\ldots,w_n\}$ as

$$VJ^w(P)=\sum_{i=1}^{n}w_i^2(1-p_i)\left(\ln((1-p_i))\right)^2-\left(\sum_{i=1}^{n}w_i(1-p_i)\ln((1-p_i))\right)^2.$$

When $w_i=x_i, \ \forall\ i=1,2,\ldots,n$, then the weighted varextropy is given as

$$VJ^x(P)=\sum_{i=1}^{n}x_i^2(1-p_i)\left(\ln((1-p_i))\right)^2-\left(\sum_{i=1}^{n}x_i(1-p_i)\ln((1-p_i))\right)^2.$$

We define general weighted varextropy for an absolutely continuous random variable as

$$\begin{aligned}VJ^w(X)&=\mathrm{Var}\left(-\frac{1}{2}w(X)f_X(X)\right)\\&=\frac{1}{4}\left[E(w^2(X)f^2(X))-(E(w(X)f_X(X)))^2\right]\\&=\frac{1}{4}\left[\int_{S_X}w^2(x)f^3(x)\,dx-\left(\int_{S_X}w(x)f^2(x)\,dx\right)^2\right].\end{aligned}$$

When $w(x)=x$, then weighted varextropy is given as

$$VJ^x(X)=\frac{1}{4}\left[\int_{S_X}x^2f^3(x)\,dx-\left(\int_{S_X}xf^2(x)\,dx\right)^2\right].$$

The weighted varextropy $VJ^x(X)$ for some standard distributions are given in Table 2.

4 A characterization of uniform distribution

In many practical problems, the goodness-of-fit test may be reduced to the problem of testing uniformity. Since the varextropy of X is the variance of $-\frac{1}{2}f_X(x)$, the varextropy is nonnegative for any random variable X. Among all distributions with support on $[0,1]$, the uniform distribution has the maximum extropy. An important property of the uniform distribution is that it obtains the minimum varextropy among all distributions having support on $[0,1]$ (see, Qiu and Jia (2018)).

The characterization provided in Theorem 1 is significant because it establishes a clear and definitive criterion for identifying a uniform distribution based on varextropy. By showing that a random variable X has zero varextropy if and only if it is uniformly distributed over $[0,1]$, it offers a direct and precise method for testing uniformity without needing complex parametric assumptions. This improves upon existing characterizations by linking uniformity to an easily computable quantity, varextropy, which is grounded in variance, making it more practical for statistical analysis. Previous methods might have relied on more complex or indirect approaches, but the varextropy-based test is simple, theoretically sound, and offers a direct comparison for uniformity. This approach fills a gap by providing a nonparametric and computationally feasible solution to uniformity testing, making it a more accessible tool in both theoretical and applied statistics.

The characterization in Theorem 1 makes a few key assumptions. First, it assumes that the random variable X is continuous and has support on the interval $[0,1]$. This is crucial because the result specifically applies to distributions confined to this interval, such as the uniform distribution. Second, the characterization assumes that the pdf $f_X(x)$ is well-defined and continuous over this support. This ensures that the varextropy formula, which relies on the second moment of the pdf, can be computed without encountering issues related to discontinuities or undefined behavior. Additionally, the proof assumes that $f_X(x)$ integrates to 1 over $[0,1]$, which is a fundamental property of any valid probability density function. These assumptions are necessary to guarantee the correctness and applicability of the characterization, ensuring it is valid for continuous distributions on the unit interval and can be used as a reliable test for uniformity. Noughabi and Noughabi (2023) applied varentropy to test for uniformity. They showed that the varentropy of X is zero if and only if X follows the standard uniform distribution, and they used their proposed varentropy estimators as test statistics for conducting goodness-of-fit tests for uniformity. Following result is a characterization of the uniform distribution using varextropy (see (Chaudhary and Gupta, 2024, Theorem 11)).

Theorem 1. *Let X be a continuous random variable with support on $[0,1]$. Then $VJ(X) = 0$ if and only if X has a uniform distribution on the interval $[0,1]$.*

Proof. Let random variable X have a uniform distribution on the interval $[0,1]$; then $f_X(x) = 1$ for $0 \le x \le 1$, and

$$VJ(X) = \frac{1}{4}\int_0^1 f^3(x)\,dx - \frac{1}{4}\left[\int_0^1 f^2(x)\,dx\right]^2 = 0.$$

Conversely, $VJ(X) = 0$ implies $\mathrm{Var}(f_X(X)) = 0$, that is, $f_X(x) = c$. Since

$$\int_0^1 f_X(x)\,dx = 1, \quad \text{therefore} \quad f_X(x) = 1, \quad 0 \le x \le 1.$$

Hence, the proof is complete. □

5 Nonparametric estimators

Suppose that $X_{1:n}, X_{2:n}, X_{3:n}, \ldots, X_{n:n}$ are order statistics of random sample $X_1, X_2, \ldots, X_n$ from cdf F. Then, then empirical distribution function of cdf F is given by

$$F_n(x) = \begin{cases} 0, & x < X_{1:n} \\ \frac{i}{n}, & X_{i:n} \le x < X_{i+1:n}, \quad i = 1,2,\ldots,n-1, \\ 1, & x \ge X_{n:n}. \end{cases}$$

Noughabi and Noughabi (2024) provided various estimators of $VJ(X)$. $VJ(X)$ can be expressed as

$$VJ(X) = \frac{1}{4}\left[\int_0^1 \left(\frac{d}{dp}(F^{-1}(p))\right)^{-2} dp - \left(\int_0^1 \left(\frac{d}{dp}(F^{-1}(p))\right)^{-1} dp\right)^2\right].$$

Following the idea of Vasicek (1976), Noughabi and Noughabi (2024) proposed the estimator Δ for $VJ(X)$ as

$$\Delta = \frac{1}{4n}\sum_{i=1}^{n}\left(\frac{2m/n}{X_{i+m:n} - X_{i-m:n}}\right)^2 - \frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2m/n}{X_{i+m:n} - X_{i-m:n}}\right)\right)^2.$$

Here, the window size m is a positive integer less than or equal to $\frac{n}{2}$. If $i+m > n$, then we consider $X_{i+m:n} = X_{n:n}$, and if $i-m < 1$, then we consider $X_{i-m:n} = X_{1:n}$. The proposed estimator for varextropy, Δ, calculates the weighted variance of order statistics based on sample data, using a window size parameter m. It is defined as

$$\Delta = \frac{1}{4n}\sum_{i=1}^{n}\left(\frac{2m/n}{X_{i+m:n} - X_{i-m:n}}\right)^2 - \frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2m/n}{X_{i+m:n} - X_{i-m:n}}\right)\right)^2.$$

This estimator is consistent, meaning it converges to the true value of varextropy as the sample size increases and is flexible for a range of distributions. Its primary advantages include its practical applicability for goodness-of-fit tests, such as testing uniformity, and its ability to offer consistent results for large datasets. However, it is sensitive to the choice of the window size m, and for large samples, it can become computationally intensive. Additionally, for small sample sizes, the estimator may not be highly accurate, and its distribution under the null hypothesis requires Monte Carlo simulations to determine critical values. Despite these limitations, the estimator improves upon existing methods by directly utilizing order statistics and providing a reliable approach to test uniformity.

The proposed estimator, Δ, for varextropy offers several advantages when compared to other estimators used in statistical tests for uniformity. One key feature is its use of order statistics, which captures more nuanced information about the distribution of data, particularly in non-parametric contexts. Compared to traditional estimators like the sample variance or methods based on moments, Δ incorporates weighted variations in the sample, making it more sensitive to the underlying distribution, especially for detecting deviations from uniformity.

When compared to earlier estimators like the ones proposed by Noughabi and Noughabi (2023), which are based on varentropy for uniformity testing, the proposed estimator has a distinct advantage in terms of flexibility and consistency. While their estimator is also consistent, it is based on a more complex approach, potentially requiring additional assumptions about the distribution shape. The Δ estimator, on the other hand, relies on empirical distributions and requires fewer assumptions about the underlying data, making it more adaptable.

However, one limitation of Δ is its reliance on the window size parameter m, which requires tuning and may impact its performance in smaller datasets. Other methods, like those using bootstrap resampling techniques, can offer an alternative, providing robust estimates without relying on window size. Overall, the proposed estimator provides a more robust and flexible approach than many existing alternatives, particularly when testing for uniformity in real-world data.

6 A test of uniformity

In this section, we introduce a statistical test for uniformity based on the concept of varextropy, specifically using the estimator Δ proposed by Noughabi and Noughabi (2024). It has been established that the varextropy of a random variable X is zero if and only if X follows a standard uniform distribution. Using this property, we can utilize the proposed varextropy estimators as test statistics for conducting goodness-of-fit tests to determine whether a given sample follows a uniform distribution. The hypothesis of interest is framed as follows:

- Null hypothesis (H_0): The random variable X is uniformly distributed.
- Alternative hypothesis (H_1): The random variable X is not uniformly distributed.

We propose using Δ, an estimator of the varextropy $VJ(X)$, as the test statistic. The estimator Δ is consistent, meaning that as the sample size n increases, Δ converges in probability to the true value of $VJ(X)$. Under the null hypothesis H_0, if X follows a uniform distribution, Δ converges in probability to zero. On the other hand, if X is not uniformly distributed (under H_1), Δ converges to a nonzero value. This distinction allows us to use large values of Δ as evidence of nonuniformity. Therefore, we reject the null hypothesis when Δ exceeds a certain threshold.

Since the distribution of the test statistic Δ under the null hypothesis is too complex to derive analytically, we use Monte Carlo simulation to empirically determine the critical values and power of the test. The critical region for the test is defined as $\Delta \geq C_{1-\alpha}$, where $C_{1-\alpha}$ is the critical value corresponding to the significance level α. For a given sample size n and significance level α, we compute $C_{1-\alpha}$ using a Monte Carlo simulation. This approach allows us to determine the appropriate threshold for rejecting the null hypothesis based on simulated data from the uniform distribution.

Table 3: Critical values at significance level $\alpha = 0.05$.

$m\backslash n$	10	20	30	40	50	80	100
2	4.7570	3.1388	2.2838	1.9478	1.6402	1.1355	1.0451
3	1.4909	1.2126	0.7925	0.6502	0.5729	0.4235	0.3559
4	0.7064	0.6089	0.4724	0.3841	0.3434	0.2541	0.2121
5	0.4074	0.4252	0.3525	0.2881	0.2396	0.1800	0.1541
9		0.1551	0.1703	0.1542	0.1399	0.1039	0.0947
14			0.0869	0.0973	0.1006	0.0829	0.0731
19				0.0637	0.0722	0.0722	0.0665
24					0.0528	0.0639	0.0619
30						0.0514	0.0546
39						0.0380	0.0436
49							0.0336

6.1 Critical points

We define a function to calculate the value of Δ. A sample of size n is generated from the $U(0,1)$ distribution, and the test statistic is computed for the sample data. After 10,000 replications, the $(1-\alpha)^{\text{th}}$ quantile of the test statistics is determined as the critical value at significance level α. Critical values for $\alpha = 0.05$ are given in Table 3 for different values of m and n.

To derive the critical values of the proposed test statistic $\widehat{\Delta}$, we employ a Monte Carlo simulation approach due to the analytical intractability of its sampling distribution under the null hypothesis. Specifically, we generate 10,000 independent random samples of size n from the standard uniform distribution $U(0,1)$, which represents the null hypothesis H_0. For each simulated sample, we compute the value of the test statistic $\widehat{\Delta}$ using the nonparametric estimator that involves a fixed window size m. After obtaining 10,000 such values of $\widehat{\Delta}$, we determine the empirical $(1-\alpha)$-th quantile to serve as the critical value $C_{1-\alpha}$ at a given significance level α. These critical values are summarized in Table 3 for various combinations of n and m, thus providing practical benchmarks for implementation.

6.2 Power of test

We used the following procedure to estimate the power of the test. For each sample size n, we generate 10,000 random samples of size n from the alternative distribution. The test statistic is then computed for each sample. The power of the test at a significance level α is estimated as the proportion of these 10,000 samples that fall within the corresponding critical region.

The estimated power of the test is obtained as the proportion of samples for which the test statistic exceeds the critical value, leading to the rejection of H_0. This empirical procedure provides a consistent and practical method for evaluating the effectiveness of the test. The results, presented in Table 4, demonstrate that the proposed test performs well in detecting deviations from uniformity and exhibits higher power compared to existing tests for standard alternatives like the Beta(1,2) distribution. The pdf of the beta distribution with parameters a

Table 4: Power at significance level $\alpha = 0.05$.

$m\backslash n$	10	20	30	40	50	80	100
2	0.0817	0.0966	0.0960	0.1038	0.1137	0.1492	0.1746
3	0.1197	0.1343	0.1558	0.1689	0.1836	0.2522	0.2924
4	0.1546	0.1882	0.2082	0.2302	0.2570	0.3536	0.4467
5	0.1786	0.2370	0.2621	0.2874	0.3359	0.4451	0.5483
9		0.3962	0.4360	0.4654	0.5072	0.6540	0.7211
14			0.5991	0.6221	0.6566	0.7380	0.8187
19				0.7752	0.7734	0.8311	0.8700
24					0.8875	0.8829	0.9128
30						0.9397	0.9459
39						0.9883	0.9850
49							0.9983

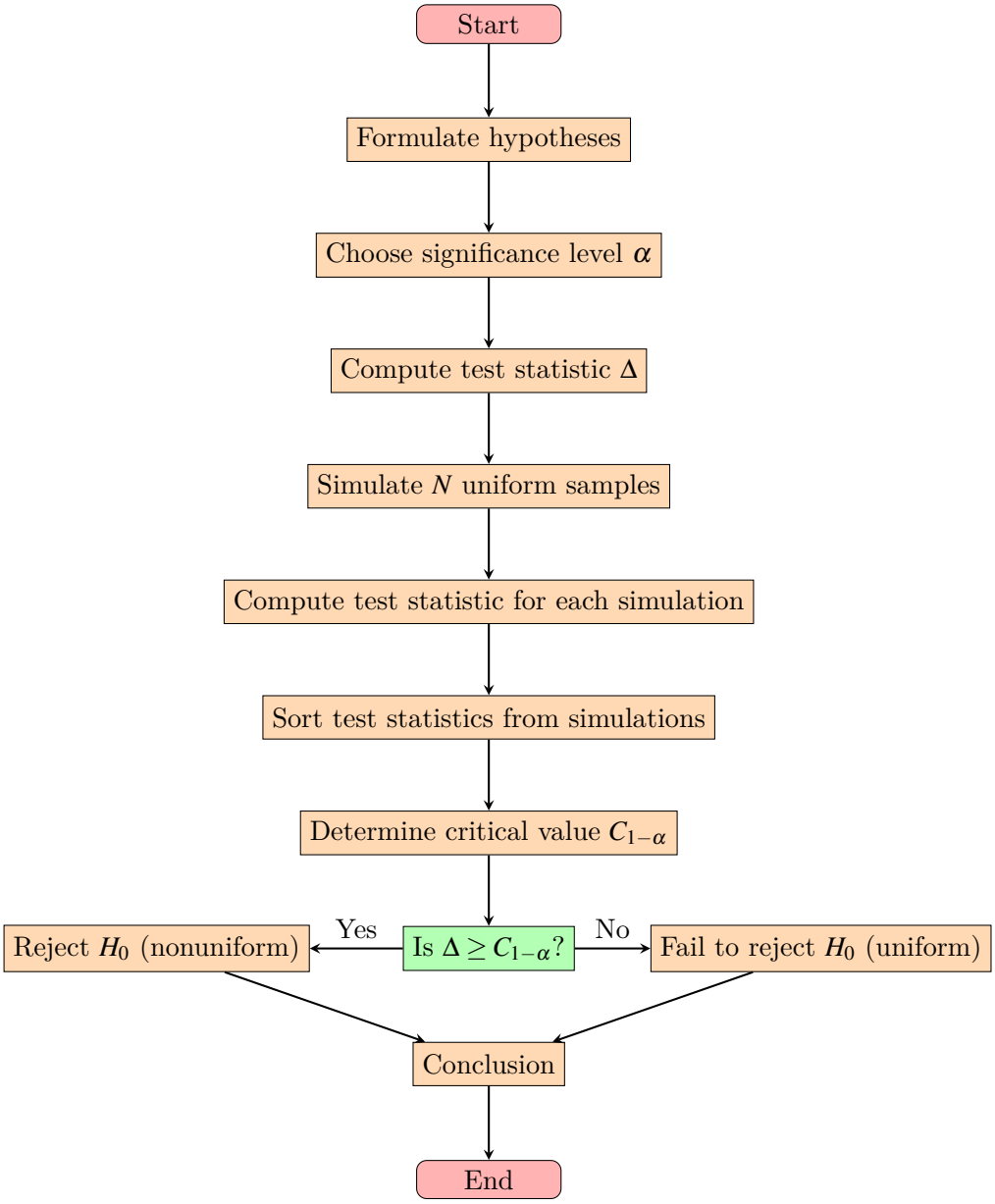
and b is given by

$$f_x(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, \quad 0 < x < 1,$$

where $B(a,b)$ is the complete beta function.

To estimate the power of the test, we again utilize a Monte Carlo simulation, this time under the alternative hypothesis H_1. We generate 10,000 random samples of size n from a nonuniform distribution, such as the Beta(1,2) distribution, which is a common alternative to $U(0,1)$. For each sample, the value of $\widehat{\Delta}$ is calculated and compared to the corresponding critical value derived under H_0.

The estimated power against the alternative Beta(1,2) distribution is given in Table 4 at the significance level $\alpha = 0.05$. Our test performs well in detecting nonuniform data. Note that Beta(1,1) is identically distributed with $U(0,1)$. The power of this test against the alternative Beta(1,1) is approximately α, so the test achieves its level of significance. The power of our test is higher than the power of the test proposed by Noughabi and Noughabi (2023) for the common alternative Beta(1,2).



6.3 Application to real data

The dataset used in this application comes from a real-world environmental study involving vinyl chloride concentrations. Vinyl chloride is a toxic substance, and understanding its distribution is critical for assessing environmental risks and regulatory compliance. In particular, the dataset represents a sample of vinyl chloride measurements that have been transformed to fit a uniform distribution using the probability integral transformation, as proposed by Xiong et al. (2022). This transformation is commonly used to standardize nonuniform data so that it can be tested

for uniformity.

The choice of this dataset is directly linked to our theoretical findings in the previous sections, particularly the application of the test of uniformity based on the varextropy estimator $\widehat{\Delta}$. As we shown earlier, the test statistic $\widehat{\Delta}$ can effectively detect whether a dataset conforms to a uniform distribution, which is important for validating the uniformity of the transformed data. Given that uniformity is a key assumption in many statistical procedures, it is essential to verify whether the transformation of the vinyl chloride concentrations actually results in data that adheres to the uniform distribution.

For this dataset, we computed the value of the proposed test statistic $\widehat{\Delta}$ using a window size $m = 16$ and sample size $n = 34$. The computed statistic was found to be $\widehat{\Delta} = 0.0329$. The critical value for the test at a significance level of $\alpha = 0.05$ was obtained through Monte Carlo simulation, resulting in a critical value of 0.0733 for $m = 16$ and $n = 34$. Since the test statistic $\widehat{\Delta} = 0.0329$ is less than the critical value of 0.0733, the observed statistic lies within the acceptance region.

This outcome suggests that the transformed data conforms to the uniform distribution, and therefore, we fail to reject the null hypothesis of uniformity. In other words, our proposed test successfully verifies that the transformation applied to the vinyl chloride data indeed resulted in a uniform distribution, aligning with the expectations of the transformation method.

Dataset 1: 0.0518, 0.0518, 0.1009, 0.1009, 0.1917, 0.1917, 0.1917, 0.2336, 0.2336, 0.2336, 0.2733, 0.2733, 0.3467, 0.3805, 0.3805, 0.4126, 0.4431, 0.4719, 0.4719, 0.4993, 0.6162, 0.6550, 0.6550, 0.7059, 0.7211, 0.7356, 0.7623, 0.7863, 0.8178, 0.8810, 0.9337, 0.9404, 0.9732, 0.9858.

The dataset represents vinyl chloride concentrations transformed into a uniform distribution using the probability integral transformation (Xiong et al., 2022). The value of the test statistic $\widehat{\Delta}$ is 0.0329 when the window size $m = 16$ and the sample size $n = 34$. The critical point is 0.0733 at the 5% level of significance, based on Monte Carlo simulations for $m = 16$ and $n = 34$. Since the estimated value of the test statistic lies in the acceptance region, our test based on $\widehat{\Delta}$ fails to reject the null hypothesis. Therefore, the test verifies that the data is fitted to a uniform distribution.

The results of our uniformity test based on the varextropy estimator $\widehat{\Delta}$ indicate that the transformed vinyl chloride data fits well with a uniform distribution. The calculated test statistic ($\widehat{\Delta} = 0.0329$) was smaller than the critical value (0.0733) at the 5% significance level, suggesting that we failed to reject the null hypothesis of uniformity.

However, there are several limitations and potential biases in our analysis. First, the critical values were derived using Monte Carlo simulations, which, while accurate, are approximations and depend on the number of replications used (10,000 in this case). Furthermore, the performance of the test is influenced by the sample size, and our results may not generalize well to smaller or larger samples. Another limitation is the assumption that the data under the null hypothesis is perfectly uniform, which may not always hold in practice, especially with real-world data where small deviations from uniformity can occur. Additionally, the window size used in the calculation of $\widehat{\Delta}$ could impact the test's power and its sensitivity to nonuniformity. While the test performed well in detecting significant departures from uniformity in this case, its ability to detect subtle differences might be limited. Moreover, the Monte Carlo method, while effective, can introduce bias if the number of replications is not large enough or if the underlying assumptions about the test statistic are inaccurate.

Lastly, the choice of dataset, in this case, vinyl chloride concentrations, may not be representative of other datasets, and the conclusions drawn here might not be directly applicable to different contexts. Despite these limitations, the test demonstrates its utility in assessing uniformity, and future research can refine its performance and extend its application to other distributions.

7 Conclusion

The results of our study demonstrate that the proposed test based on the varextropy estimator $\widehat{\Delta}$ is an effective tool for assessing the uniformity of datasets. We applied the test to real-world vinyl chloride concentration data that had been transformed to fit a uniform distribution using the probability integral transformation. The test statistic $\widehat{\Delta} = 0.0329$ was found to be smaller than the critical value of 0.0733 at a 5% significance level, leading to the conclusion that the transformed data conforms to a uniform distribution. This outcome is consistent with our theoretical expectation that $\widehat{\Delta}$ should be small when the data follows a uniform distribution.

Moreover, the test showed strong performance in detecting deviations from uniformity when applied to simulated data from a Beta(1,2) distribution, a common alternative to the uniform distribution. As expected, the test's power increased with sample size, and the critical values, derived through Monte Carlo simulations, provided a reliable framework for determining decision thresholds for uniformity testing at various levels of significance. These results demonstrate the robustness and practicality of the proposed test, especially in the context of assessing uniformity in real-world datasets.

However, the study also highlights certain limitations. The Monte Carlo simulation approach, while effective, relies on approximations that depend on the number of replications used, and the performance of the test can be influenced by the sample size and the choice of window size. Additionally, the test assumes that the null hypothesis represents perfectly uniform data, which may not always hold in practice, particularly when small deviations from uniformity exist in real-world data. These factors suggest that while the test performs well under the given conditions, its generalizability to other datasets and scenarios may require further investigation.

Future research could address several areas for improvement. First, exploring the test's performance with smaller sample sizes and more diverse datasets would provide insights into its robustness and lead to better calibration of critical values. Investigating the test's sensitivity to a wider range of nonuniform distributions, beyond Beta(1,2), could help evaluate its applicability to different data patterns. Enhancing the Monte Carlo simulation process through more efficient sampling techniques or parallel processing could improve both computational speed and scalability. Additionally, examining the test's robustness to different distributional assumptions, such as normal or skewed distributions, would further validate its flexibility.

Comparing the varextropy-based test with other established goodness-of-fit tests, such as the Kolmogorov–Smirnov or Anderson-Darling tests, could provide valuable insights into its relative strengths and weaknesses. Expanding the test's application to multivariate or time-series data could broaden its utility in domains such as finance, ecology, and other fields that deal with complex data structures. Furthermore, implementing dynamic window size selection methods might enhance the accuracy and adaptability of the test across various datasets.

Lastly, applying the test to additional real-world datasets from diverse fields would help assess its practical applicability and identify domain-specific challenges or opportunities for refinement. Addressing these areas in future research could improve the accuracy, efficiency, and broad applicability of the proposed uniformity test, making it a valuable tool for detecting uniformity across a variety of statistical and applied contexts

Acknowledgement

The authors are grateful to the editors and anonymous referees for their insightful comments and valuable suggestions, which have significantly improved the quality of this paper.

References

Arıkan, E. (2016). Varentropy decreases under the polar transform. *IEEE Transactions on Information Theory*, **62**, 3390–3400.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1998). *Records*. John Wiley & Sons, New York.

Balakrishnan, N., Buono, F., and Longobardi, M. (2020). On weighted extropies. *Communications in Statistics - Theory and Methods*, **51**, 6250–6267.

Belzunce, F., Guillamón, A., Navarro, J., and Ruiz, J. M. (2001). Kernel estimation of residual entropy. *Communications in Statistics - Theory and Methods*, **30**, 1243–1255.

Chacko, M., and Grace, A. (2024). Varextropy of k-record values and its applications. *Communications in Statistics - Theory and Methods*, 1–24.

Chaudhary, S. K., and Gupta, N. (2024). Extropy and varextropy estimators with applications. *arXiv preprint* arXiv:2401.13065.

De Crescenzo, A., Paolillo, L., and Suárez-Llorens, A. (2025). Stochastic comparisons, differential entropy and varentropy for distributions induced by probability density functions. *Metrika*, **88**, 43–59.

Dudewicz, E. J., and van der Meulen, E. C. (1981). Entropy-based tests of uniformity. *Journal of the American Statistical Association*, **76**, 967–974.

Goodarzi, F. (2022). Results on conditional variance in parallel system and lower bounds for varextropy.*Communications in Statistics - Theory and Methods*, **53**, 1590–1610.

Goodarzi, F. (2024). Results on conditional variance in parallel system and lower bounds for varextropy. *Communications in Statistics - Theory and Methods*, **53**, 1590–1610.

Gupta, N., and Chaudhary, S. K. (2023). Some characterizations of continuous symmetric distributions based on the extropy of record values. *Statistical Papers*, 1–18.

Hazeb, R., Bayoud, H., and Raqab, M. (2021). Kernel and CDF-based estimation of extropy and entropy from progressively type-II censoring with application for goodness of fit problems. *Stochastics and Quality Control*, **36**(1), 73–83.

Lad, F., Sanfilippo, G., and Agro, G. (2015). Extropy: Complementary dual of entropy. *Statistical Science*, **30**, 40–58.

Maadani, S., Mohtashami Borzadaran, G. R., and Rezaei Roknabadi, A. H. (2022). Varentropy of order statistics and some stochastic comparisons. *Communications in Statistics - Theory and Methods*, **51**, 6447–6460.

Noughabi, H. A., and Noughabi, M. S. (2023). Varentropy estimators with applications in testing uniformity. *Journal of Statistical Computation and Simulation*, **93**, 2582–2599.

Noughabi, H. A., and Noughabi, M. S. (2024). On the estimation of varextropy. *Statistics*, **58**, 827–841.

Qiu, G. (2017). The extropy of order statistics and record values. *Statistics and Probability Letters*, **120**, 52–60.

Qiu, G., and Jia, K. (2018). Extropy estimators with applications in testing uniformity. *Journal of Nonparametric Statistics*, **30**, 182–196.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423.

Vaselabadi, N. M., Tahmasebi, S., Kazemi, M. R., and Buono, F. (2021). Results on varextropy measure of random variables. *Entropy*, **23**, 356.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **38**, 54–59.

Xiong, P., Zhuang, W., and Qiu, G. (2022). Testing exponentiality based on the extropy of record values. *Journal of Applied Statistics*, **49**, 782–802.

Zaid, E. O. A., Attwa, R. A. E. W., and Radwan, T. (2022). Some measures information for generalized and q-generalized extreme values and its properties. *Fractals*, **30**, 2240246.

End of SMPS Journal – Volume 2, Issue 1

For manuscript submission and future issues, visit:
www.smps.um.ac.ir