Ferdowsi University of Mashhad

Print ISSN: 3115-9370
Online ISSN: 3060-7876

SMPS

Stochastic Models in Probability and Statistics

Vol. 3, No. 1, 2026

In the Name of God

Stochastic Models in Probability and Statistics (SMPS)

Volume 3, Issue 1 July 2026

ISSN-Print: 3115-9370 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

©2026 All rights reserved. Stochastic Models in Probability and Statistics

Editorial Board

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

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

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

Associate Editor Members

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

- **Maria Longobardi**
  Università di Napoli Federico II, Italy
  Email: maria.longobardi@unina.it
- **Gholamreza Mohtashami Borzadaran**
  Ferdowsi University of Mashhad, Iran
  Email: grmohtashami@um.ac.ir
- **Asok Nanda**
  IISER, India
  Email: asok@iiserkol.ac.in
- **Jorge Navarro**
  University of Murcia, Spain
  Email: jorgenav@um.es
- **Hon Keung Tony Ng**
  Bentley University, USA
  Email: tng@bentley.edu
- **Sangun Park**
  Yonsei University, South Korea
  Email: sangun@yonsei.ac.kr
- **Franco Pellerey**
  Politecnico di Torino, Italy
  Email: pellerey@polito.it
- **Mohammad Z. Raqab**
  University of Jordan
  Email: mraqab@ju.edu.jo
- **Alfonso Suarez Llorens**
  Universidad de Cádiz, Spain
  Email: suarez@uca.es
- **Peng Zhao**
  Jiangsu Normal University, China
  Email: zhaop@jsnu.edu.cn

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

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

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

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

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

Some ratio type estimators for estimating the population mean in case of missing data **1-11**
Ajeet Kumar Singh and Himanshi Kumawat

Multimatrix variate distributions **13-27**
José A. Díaz-García and Francisco J. Caro-Lopera

Exploring an extension of the Rayleigh distribution: statistical estimation and practical application **29-44**
Harsh Tripathi and Varun Agiwal

Quasi-identifiability of multipath change-point models with general link functions **45-52**
Masoud Majidizadeh

Efficient pseudo Rao-Blackwellized estimator in ranked sampling **53-70**
Bardia Panahbehagh and Mahdi Salehi

Wiener index and saturated nodes of random exponential recursive ternary trees **71-84**
Mehri Javanian

Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 1–11, Research Paper
https://smps.um.ac.ir, ISSN: 3115-9370

Some ratio type estimators for estimating the population mean in case of missing data

Ajeet Kumar Singh* **and Himanshi Kumawat**

Department of Statistics, University of Rajasthan, Jaipur, India

Email(s): ajeetvns.singh@gmail.com, kumawathimanshi301@gmail.com

Abstract. This paper investigates ratio-type estimators for estimating the population mean in the presence of missing data. To address the problem of missing observations, a class of generalized estimators is proposed. The bias and mean squared error of the proposed estimators are derived under optimal conditions using first-order approximations. The performance of the proposed estimators is then evaluated and compared with that of the classical mean estimator, the ratio estimator, and the compromised estimator using several empirical data sets. The results show that the proposed estimators outperform the existing estimators in terms of efficiency.

Keywords: Bias, Imputation, Mean squared error, Percent relative efficiency.

Mathematics Subject Classification (2020): 62D05, 62D10, 62F10.

1 Introduction and preliminaries

In survey sampling, missing data may arise due to non-response, attrition, or errors in data collection. Such missingness poses a serious challenge, as it can lead to biased estimates, undermine the validity of statistical inference, and reduce the efficiency of estimators. Common approaches for handling missing data include imputation, weighting adjustments, and model-based methods. Among these approaches, ratio-type estimators have been widely studied for estimating the population mean because of their ability to efficiently incorporate auxiliary information. These estimators exploit the relationship between the study variable and an auxiliary variable, often yielding more precise estimates than those obtained under simple random sampling. However, in the presence of missing data, classical ratio estimators must be modified to account for different missingness mechanisms, namely missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). In this study, we focus on ratio-type estimators under missing data and investigate their performance across various missingness scenarios.

Let a finite population consist of N distinct units with associated values (x_i, y_i). The population means

*Corresponding author
Received: 22 September 2025 / Revised: 1 November 2025 / Accepted: 13 February 2026

https://doi.org/10.22067/smps.2025.95580.1052

The Author(s). This is an open access article under CC BY 4.0.

are defined as

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad \text{and} \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

where Y denotes the study variable and X is an auxiliary variable used for estimation. A sample S of size n is drawn from the finite population $P = 1, 2, \ldots, N$ using simple random sampling without replacement (SRSWOR) to estimate the population mean.

The sample S consists of r responding units ($r < n$), forming the response set R, and $n - r$ non-responding units, forming the complementary set R^c. The variable Y is the primary variable of interest, while X is an auxiliary variable correlated with Y. For each unit $i \in R$, y_i denotes the observed value of the study variable, whereas for each unit $i \in R^c$, the corresponding y_i values are missing and therefore require imputation.

Kalton et al. (1981) addressed issues of non-response and imputation in survey sampling, while Meeden (2000) proposed a theoretical framework for imputation. Lee et al. (1994) introduced a ratio method of imputation, and Singh and Horn (2000) proposed an improved technique known as the compromised method of imputation. Subsequently, Singh and Deo (2003) developed a modified estimator to handle missing data more efficiently.

A widely recognized mean imputation method, along with its sample mean and variance, can be expressed as $\bar{y}_m = \bar{y}_r$ and

$$V(\bar{y}_m) = \delta_1 C_Y^2 \bar{Y}^2, \tag{1}$$

respectively, where $\delta_1 = \frac{1}{r} - \frac{1}{N}$. It is also well recognized that, under the ratio method of imputation, the ratio estimator of the population mean is

$$\bar{y}_R = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}.$$

The bias and mean squared error (MSE) of $\bar{y}_R$, in the presence of missing data, can be expressed as follows:

$$B(\bar{y}_R) = \delta_1 \bar{Y} \left(C_X^2 - \rho_{YX} C_Y C_X\right)$$

and

$$MSE(\bar{y}_R) = \bar{Y}^2(\delta_1 C_Y^2 + \delta_4(C_X^2 - 2\rho_{YX} C_Y C_X)), \tag{2}$$

respectively, $\delta_4 = \frac{1}{r} - \frac{1}{n}$, ρ_{XY} denotes the correlation coefficient between X and Y, and C_X and C_Y represent the coefficients of variation of X and Y, respectively. Singh and Horn (2000) proposed the compromised method of imputation, defined as

$$\bar{y}_{\text{comp}} = \alpha \bar{y}_r + (1 - \alpha)\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}, \tag{3}$$

where α is a constant. The value of α that minimizes the MSE is given by $\alpha = 1 - \rho_{XY} C_Y / C_X$. The corresponding bias and MSE are given by

$$B(\bar{y}_{\text{comp}}) = \delta_4 \bar{Y} \left(1 - \rho_{YX} \frac{C_Y}{C_X}\right) \rho_{YX} C_Y C_X$$

and

$$MSE(\bar{y}_{\text{comp}}) = \bar{Y}^2(\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2), \tag{4}$$

respectively. Pandey et al. (2015) and Singh and Nath (2018) proposed the following estimators

$$\bar{y}_{Re} = \bar{y}_r \exp\left(\frac{\bar{x}_1 - \bar{x}_r}{\bar{x}_1 + \bar{x}_r}\right),$$

and

$$\bar{y}_{dr} = \bar{y}_r \left(\frac{n_1 \bar{x}_1 - n\bar{x}_r}{(n_1 - n)\bar{x}_r}\right)^{\beta}.$$

The bias and MSE of $\bar{y}_{Re}$ are given by

$$B(\bar{y}_{Re}) = \delta_5 \bar{Y}\left(\frac{3}{8}C_X^2 - \frac{1}{2}\rho_{YX}C_Y C_X\right),$$

$$\text{MSE}(\bar{y}_{Re}) = \bar{Y}^2\left(\delta_1 C_Y^2 + \delta_5\left(\frac{C_X^2}{4} - \rho_{YX}C_Y C_X\right)\right), \tag{5}$$

where $\delta_5 = \frac{1}{r} - \frac{1}{n_1}$. Similarly, for $\bar{y}_{dr}$, the bias and MSE are

$$B(\bar{y}_{dr}) = -\delta_5 \frac{\bar{Y}}{2}\left(g_1 C_X + \rho_{YX}C_Y\right)\rho_{YX}C_Y,$$

and

$$\text{MSE}(\bar{y}_{dr}) = \bar{Y}^2\left(\delta_1 C_Y^2 - \delta_5 \rho_{YX}^2 C_Y^2\right), \tag{6}$$

respectively, where $g_1 = \frac{n}{n_1 - n}$.

In the remainder of this paper, new estimators are proposed in Section 2. Their statistical properties and performance in terms of MSE are investigated. Section 3 presents numerical results and graphical analysis, while concluding remarks are provided in Section 4.

2 Proposed estimators and their properties

Motivated by Bahl and Tuteja (1991) and Singh et al. (2014, 2015a,b), we propose a new class of ratio-type estimators for estimating the population mean under SRSWOR.

$$T_{1P} = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)\exp\left[\alpha\left(\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right)\right], \tag{7}$$

$$T_{2P} = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\alpha}\exp\left[\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right], \tag{8}$$

$$T_{3P} = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\alpha}\exp\left[\beta\left(\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right)\right], \tag{9}$$

where α and β are constants chosen to minimize the MSE of the estimators T_{1P}, T_{2P}, and T_{3P}.
Using the transformations $\bar{y}_r = \bar{Y}(1+e_1)$ and $\bar{x}_r = \bar{X}(1+e_2)$, such that $|e_i| < 1, i = 1,2$. Expanding T_{1P}, T_{2P}, and T_{3P} in terms of e_1 and e_2, retaining terms up to the first order of approximation, the expressions in (7), (8) and (9) can be written as

$$T_{1P} = \bar{Y} + \bar{Y}\left(e_1 + e_2^2 - e_2\left(1 + \frac{\alpha}{2}\right) - e_1 e_2\left(1 + \frac{\alpha}{2}\right) + \alpha\left(1 + \frac{\alpha}{2}\right)\frac{e_2^2}{4} + \frac{\alpha}{2}e_2^2\right), \tag{10}$$

$$T_{2P} = \bar{Y} + \bar{Y}\left(e_1 + \frac{3}{8}e_2^2 - e_2\left(\alpha + \frac{1}{2}\right) - e_1 e_2\left(\alpha + \frac{1}{2}\right) + \alpha\left(1 + \frac{\alpha}{2}\right)e_2^2\right), \tag{11}$$

and

$$T_{3P} = \bar{Y} + \bar{Y}\left(e_1 - \left(\alpha + \frac{\beta}{2}\right)\left(e_2 + e_1 e_2 - \frac{e_2^2}{2}\right) + \left(\alpha + \frac{\beta}{2}\right)^2\frac{e_2^2}{2}\right), \tag{12}$$

respectively. We recall that the following relations hold for e_1 and e_2:

$$E(e_1) = E(e_2) = 0,\ E(e_1^2) = \delta_1 C_Y^2,\ E(e_2^2) = \delta_1 C_X^2 \ \text{ and } \ E(e_1 e_2) = \delta_1 \rho_{YX} C_Y C_X. \tag{13}$$

Using (10)–(13), we derive the bias and MSE of the estimators T_{1P}, T_{2P}, and T_{3P}. The results are presented below.

(i) Bias of proposed estimators. Using (10) and (13), we obtain the bias of T_{1P} up to the first-order approximation, which is given by

$$\begin{aligned}
Bias(T_{1P}) &= E(T_{1P}-\bar{Y}) \\
&= E\left(\bar{Y}\left[e_1+e_2^2-e_2\left(1+\frac{\alpha}{2}\right)-e_1e_2\left(1+\frac{\alpha}{2}\right)+\alpha\left(1+\frac{\alpha}{2}\right)\frac{e_2^2}{4}+\frac{\alpha}{2}e_2^2\right]\right) \\
&= \bar{Y}\left[E(e_2^2)-E(e_1e_2)\left(1+\frac{\alpha}{2}\right)+\alpha\left(1+\frac{\alpha}{2}\right)\frac{E(e_2^2)}{4}+\frac{\alpha}{2}E(e_2^2)\right] \\
&= \bar{Y}\left[\delta_1C_X^2-\delta_1\rho_{YX}C_XC_Y\left(1+\frac{\alpha}{2}\right)+\alpha\left(1+\frac{\alpha}{2}\right)\frac{\delta_1C_X^2}{4}+\frac{\alpha}{2}\delta_1C_X^2\right] \\
&= \bar{Y}\delta_1\left[(C_X^2-\rho_{YX}C_YC_X)+\frac{\alpha}{2}\left(\frac{3}{2}C_X^2-\rho_{YX}C_YC_X\right)+\frac{\alpha^2}{8}C_X^2\right].
\end{aligned}$$

Similarly, using (11), (12) and (13), we obtain the bias of T_{2P} and T_{3P} up to the first-order approximation, which are given by

$$Bias(T_{2P}) = \bar{Y}\delta_1\left(\frac{3}{8}C_X^2-\frac{1}{2}\rho_{YX}C_YC_X+\alpha(C_X^2-\rho_{YX}C_YC_X)+\frac{\alpha^2}{2}C_X^2\right)$$

and

$$Bias(T_{3P}) = \bar{Y}\delta_1\left[\left(\alpha+\frac{\beta}{2}\right)\left(\frac{C_X^2}{2}\left(\alpha+\frac{\beta}{2}+1\right)-\rho_{YX}C_YC_X\right)\right],$$

respectively.

(ii) MSE of proposed estimators. Using (10) and (13), we obtain the MSE of T_{1P} up to the first-order approximation, which is given by

$$\begin{aligned}
MSE(T_{1P};\alpha) &= E\left(T_{1P}-\bar{Y}\right)^2 \\
&= E\left(\bar{Y}^2\left[e_1+e_2^2-e_2\left(1+\frac{\alpha}{2}\right)-e_1e_2\left(1+\frac{\alpha}{2}\right)+\alpha\left(1+\frac{\alpha}{2}\right)\frac{e_2^2}{4}+\frac{\alpha}{2}e_2^2\right]^2\right) \\
&= E\left(\bar{Y}^2\left[e_1^2+e_2^2+\frac{\alpha^2}{4}-2e_1e_2+\alpha e_2^2-\alpha e_1e_2\right]\right) \\
&= \bar{Y}^2\delta_1\left(C_Y^2+C_X^2-\rho_{YX}C_YC_X+\alpha\left(C_X^2-\rho_{YX}C_YC_X\right)+\frac{\alpha^2}{4}C_X^2\right). \qquad (14)
\end{aligned}$$

Similarly, using (11), (12) and (13), we obtain the MSE of T_{2P} and T_{3P} up to the first-order approximation, which are given by

$$MSE\left(T_{2P};\alpha\right) = \bar{Y}^2\delta_1\left(C_Y^2+\frac{C_X^2}{4}-\rho_{YX}C_YC_X+\alpha\left(C_X^2-2\rho_{YX}C_YC_X\right)+\alpha^2C_X^2\right), \qquad (15)$$

and

$$MSE\left(T_{3P};\alpha,\beta\right) = \bar{Y}^2\delta_1\left(C_Y^2+\left(\alpha+\frac{\beta}{2}\right)^2C_X^2-2\left(\alpha+\frac{\beta}{2}\right)\rho_{YX}C_YC_X\right), \qquad (16)$$

respectively.
The MSEs in (14), (15) and (16) depend on the parameters α and β. Since these expressions are differentiable with respect to α and β, the values of α and β that minimize the MSEs can be obtained by setting the corresponding partial derivatives equal to zero.
From (14), we obtain

$$\frac{\partial}{\partial\alpha}\mathrm{MSE}(T_{1P};\alpha)=\bar{Y}^2\delta_1\left(C_X^2-\rho_{YX}C_YC_X+\frac{\alpha}{2}C_X^2\right).$$

Setting the above expression equal to zero yields

$$\alpha_{\mathrm{op}}=2\rho_{YX}\frac{C_Y}{C_X}-2.$$

Consequently,

$$\mathrm{MSE}(T_{1P};\alpha_{\mathrm{op}})=\bar{Y}^2\delta_1C_Y^2\left(1-\rho_{YX}^2\right). \tag{17}$$

Also, from (15), we have

$$\frac{\partial}{\partial\alpha}\mathrm{MSE}(T_{2P};\alpha)=\bar{Y}^2\delta_1\left(C_X^2-2\rho_{YX}C_YC_X+2\alpha C_X^2\right).$$

Equating this expression to zero gives $\alpha_{\mathrm{op}}=\rho_{YX}\frac{C_Y}{C_X}-1/2$, and hence

$$\mathrm{MSE}(T_{2P};\alpha_{\mathrm{op}})=\bar{Y}^2\delta_1C_Y^2\left(1-\rho_{YX}^2\right). \tag{18}$$

Similarly, from (16), to obtain the optimum condition of the proposed estimator T_{3p}, let $k=\alpha+\beta/2$. Then, we have

$$MSE(T_{3P};k)=\bar{Y}^2\delta_1\left(C_Y^2+k^2C_X^2-2k\rho_{YX}C_YC_X\right). \tag{19}$$

Differentiating (19) with respect to k and equating to zero, we get $k_{op}=\rho_{YX}C_Y/C_X$. Hence, the optimum condition is

$$\alpha+\beta/2=\rho_{YX}C_Y/C_X.$$

Thus, in this case, the minimum MSE is given by

$$\mathrm{MSE}\left(T_{3P};\alpha_{\mathrm{op}},\beta_{\mathrm{op}}\right)=\bar{Y}^2\delta_1C_Y^2\left(1-\rho_{YX}^2\right). \tag{20}$$

Remark 1. *As observed from* (17), (18), *and* (20), *the MSEs of the proposed estimators* (7), (8), *and* (9) *are identical when evaluated at the optimal values* α_{op} *and* β_{op}. *Henceforth, we denote the resulting optimal estimator by* T_P, *for example by* $\alpha+\beta/2=\rho_{YX}C_Y/C_X$ *the estimator in* (12) *reduced to*

$$T_P=\bar{Y}+\bar{Y}\left(e_1-(\rho_{YX}C_Y/C_X)\left(e_2+e_1e_2-e_2^2/2\right)+(\rho_{YX}C_Y/C_X)^2e_2^2/2\right),$$

and its MSE is given by

$$\mathrm{MSE}(T_P)=\bar{Y}^2\delta_1C_Y^2\left(1-\rho_{YX}^2\right).$$

The MSEs of all estimators are summarized in Table 1. From Table 1, the relative performance of the estimators can be readily compared in terms of their MSE.

Table 1: Performance of T_P relative to other estimators in terms of MSE.

Estimator	MSE	T_p is more efficient if
$\bar{y}_m$	(1)	$N-r>0$
$\bar{y}_R$	(2)	$\delta_1\rho_{YX}^2C_Y^2+\delta_4\left(C_X^2-2\rho_{YX}C_YC_X\right)>0$
$\bar{y}_{comp}$	(4)	$2nN-r(N-n)>0$
$\bar{y}_{Re}$	(5)	$\left(\delta_1\rho_{YX}^2C_Y^2+\delta_5\left(\frac{1}{4}C_X^2-\rho_{YX}C_YC_X\right)\right)>0$
$\bar{y}_{dr}$	(6)	$N-n_1>0$

Table 2: Descriptive data of seven populations.

Population	N	n	r	$\overline{Y}$	$\overline{X}$	$C^2{}_{\mathrm{Y}}$	$C^2{}_{\mathrm{X}}$	ρ_{YX}
I Cochran (1977)	34	11	9	4.92	2.59	1.0248	1.5175	0.7326
II Murthy (1967)	34	10	7	199.44	208.89	0.5673	0.5191	0.9801
III Anderson (2003)	25	9	7	183.84	185.72	0.0029	0.0027	0.7108
IV Shukla et al. (2011)	200	50	45	42.485	18.515	0.1102	0.1415	0.8652
V Mukhopadhyay (2000)	20	7	5	41.5	441.95	0.0555	0.0522	0.6521
VI Kadilar and Cingi (2008)	19	7	5	575	13573.68	2.2290	0.9101	0.88
VII Diana and Francesco (2010)	8011	400	250	28229.43	1.69	0.6193	0.2130	0.46

3 Numerical results and graphical analysis

To examine the efficiency of the proposed estimators, seven finite population data sets (from Cochran (1977), Murthy (1967), Anderson (2003), Shukla et al. (2011), Mukhopadhyay (2000), Kadilar and Cingi (2008) and Diana and Francesco (2010)) are considered. The corresponding descriptive data N, n, r, $\bar{Y}$, $\bar{X}$, C_{Y}, C_{X} and ρ_{YX} are presented in Table 2.

Numerical results for the bias, MSE, and percent relative efficiency (PRE) of the competing estimators are reported and discussed. In what follows, $\mathrm{PRE}(\hat{\theta}_1,\hat{\theta}_2)$ denotes the percent relative efficiency of $\hat{\theta}_2$ with respect to $\hat{\theta}_1$, defined as

$$\mathrm{PRE}(\hat{\theta}_1,\hat{\theta}_2)=100\,\frac{\mathrm{MSE}(\hat{\theta}_1)}{\mathrm{MSE}(\hat{\theta}_2)}.$$

The numerical results in Table 4 indicate that the proposed imputation strategy yields a lower MSE compared to the existing imputation strategies.

The MSE comparisons of the estimators $\bar{y}_m$, $\bar{y}_R$, $\bar{y}_{comp}$, $\bar{y}_{Re}$, $\bar{y}_{dr}$ and T_P for different non-response rates are presented in Table 5.

1. The PRE results presented in Tables 6, 7, and 8 clearly demonstrate that the proposed estimator T_p

Table 3: Bias values of the estimators.

Pop.	$\bar{y}_m$	$\bar{y}_R$	$\bar{y}_{COMP}$	$\bar{y}_{Re}$	$\bar{y}_{dr}$	T_P
I	0	0.24275	0.03614	0.0362	-0.2363	0.0854
II	0	-0.2888	-0.1118	-1.3845	-9.5980	-0.1246
III	0	0.0136	0.0031	0.0002	-0.0340	0.0050
IV	0	0.0244	0.0024	-0.0005	-0.0481	0.0093
V	0	0.1066	0.0273	0.0092	-0.1927	0.0358
VI	0	-29.089	-15.534	-17.903	-122.934	-20.031
VII	0	5.0246	1.5255	-0.3899	-7.8680	1.9706

Table 4: MSE values of the estimators.

Pop.	$\bar{y}_m$	$\bar{y}_R$	$\bar{y}_{COMP}$	$\bar{y}_{Re}$	$\bar{y}_{dr}$	T_P
I	2.02669	1.87525	1.75773	1.1776	1.1525	0.9389
II	2559.907	1631.473	1630.937	1002.079	448.613	100.8705
III	10.3697	8.9553	8.7527	7.1110	6.7314	5.1305
IV	3.4282	3.1288	3.0970	1.7046	1.4715	0.8619
V	14.3605	12.5869	12.0342	10.0211	9.91936	8.2539
VI	108607.5	78441.43	75995.09	71606.06	46347.47	24501.85
VII	1912690	1767868	1756029	1570428	1518632	1507965

Table 5: MSE values of estimators for different non-response rates.

Rate	Pop.	$\bar{y}_m$	$\bar{y}_R$	$\bar{y}_{COMP}$	$\bar{y}_{Re}$	$\bar{y}_{dr}$	T_P
5%	I	1.7511	1.6829	1.6300	1.0457	1.0248	0.8113
	II	1592.831	1592.831	1592.831	720.4471	410.5064	62.7639
	III	7.1692	7.1692	7.1692	5.3588	5.1479	3.5471
	IV	3.1518	3.0395	3.0276	1.6105	1.4020	0.7924
	V	8.8898	8.8898	8.8898	6.9139	6.7750	5.1096
	VI	66494.38	66494.38	66494.38	48874.65	36846.75	15001.13
	VII	1237271	1224567	1223529	1019142	986131.3	975464.2
10%	I	1.7511	1.6829	1.6300	1.0457	1.0248	0.8113
	II	1843.554	1602.849	1602.710	793.4627	420.3859	72.6434
	III	8.5694	7.9506	7.8620	6.1254	5.8407	4.2398
	IV	3.4282	3.1288	3.0971	1.7047	1.4715	0.8620
	V	11.1693	10.4303	10.1999	8.2877	8.0852	6.4197
	VI	84041.50	71472.31	70453.01	58346.07	40805.38	18959.76
	VII	1309431	1282612	1280420	1078040	1043022	1032355
15%	I	2.0267	1.8753	1.7577	1.1777	1.1525	0.9390
	II	2156.958	1615.372	1615.06	884.7321	432.7352	84.9928
	III	8.5694	7.9506	7.8620	6.1254	5.8407	4.2398
	IV	3.7442	3.2309	3.1765	1.8123	1.5510	0.9414
	V	11.1693	10.4303	10.1999	8.2877	8.0852	6.4197
	VI	84041.5	71472.31	70453.01	58346.07	40805.38	18959.76
	VII	1390080	1347486	1344004	1143867	1106606	1095939

Table 6: PRE($\bar{y}_m$,.) values of estimator.

Pop.	$\bar{y}_R$	$\bar{y}_{COMP}$	$\bar{y}_{Re}$	$\bar{y}_{dr}$	T_P
I	108.0756	115.3019	172.0946	175.8439	215.8442
II	156.9077	156.9592	255.4597	570.6269	2537.8141
III	115.7939	118.4746	145.8264	154.0497	202.1168
IV	109.5687	110.6917	201.1054	232.9688	397.7267
V	114.0903	119.3309	140.6372	144.7726	173.984
VI	138.4568	142.9138	151.6736	234.3332	443.2624
VII	108.1919	108.9213	121.7942	125.9482	126.8392

Table 7: PRE($\bar{y}_R$,.) values of estimator.

Pop.	$\bar{y}_{comp}$	$\bar{y}_{Re}$	$\bar{y}_{dr}$	T_P
I	106.6864	159.2354	162.7045	199.7159
II	100.0328	162.8088	363.6704	1617.3927
III	102.3151	125.9362	133.0378	174.5488
IV	101.0249	183.5428	212.6236	362.993
V	104.5933	123.2683	126.8929	152.4967
VI	103.2191	109.5458	169.2464	320.145
VII	100.6742	112.5723	116.4119	117.2353

Table 8: PRE($\bar{y}_{comp}$,.) values of estimator.

Pop.	$\bar{y}_{Re}$	$\bar{y}_{dr}$	T_P
I	149.2556	152.5073	187.1991
II	162.7554	363.5511	1616.862
III	123.0867	130.0276	170.5993
IV	181.6807	210.4664	359.3103
V	117.8548	121.3203	145.7996
VI	106.1294	163.9682	310.1607
VII	111.8185	115.6323	116.4503

consistently outperforms the sample mean, ratio, and compromised estimators across all populations considered.

2. The uniformly higher PRE values confirm that the proposed estimator achieves a substantial reduction in MSE and maintains strong performance across diverse population structures.

3. For all populations, the MSE increases as the non-response rate rises from 5% to 15% for each estimator under consideration. This pattern is expected, since higher non-response rates reduce the effective sample size and lead to increased variability.

4. However, the magnitude of the MSE increase differs notably among the estimators, highlighting variations in their robustness to non-response.

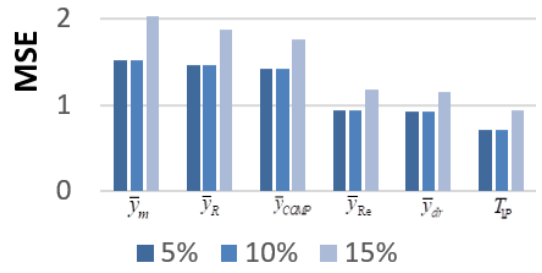


Figure 1: MSE comparison to non-response rate for population I.

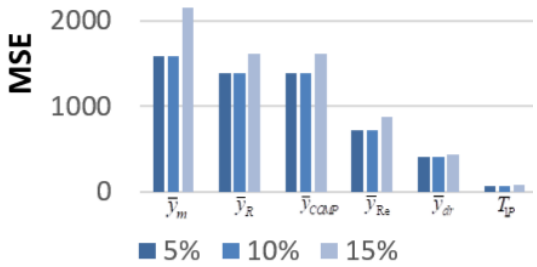


Figure 2: MSE comparison to non-response rate for population II.

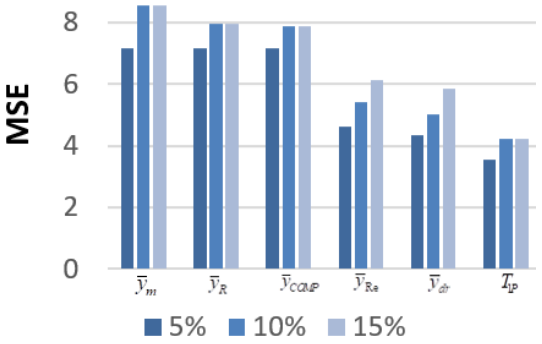


Figure 3: MSE comparison to non-response rate for population III.

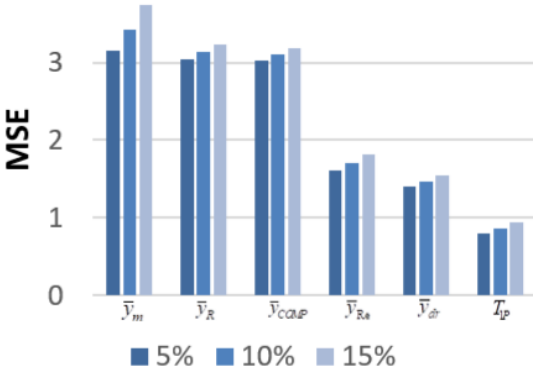


Figure 4: MSE comparison to non-response rate for population IV.

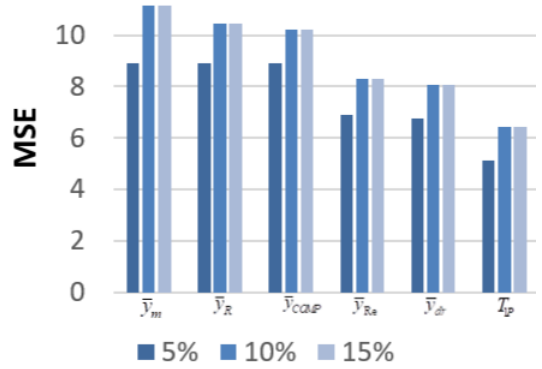


Figure 5: MSE comparison to non-response rate for population V.

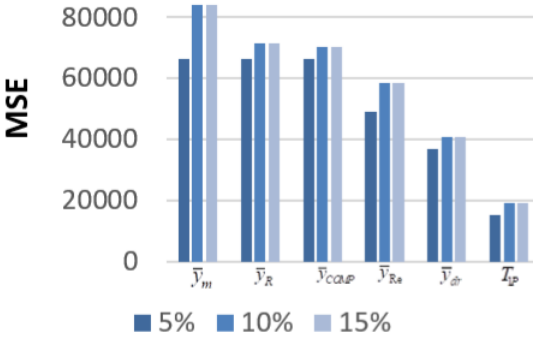


Figure 6: MSE comparison to non-response rate for population VI.

1. The numerical results and graphical analyses jointly confirm that the proposed estimator T_p consistently outperforms the mean, ratio, and compromised estimators in terms of efficiency.

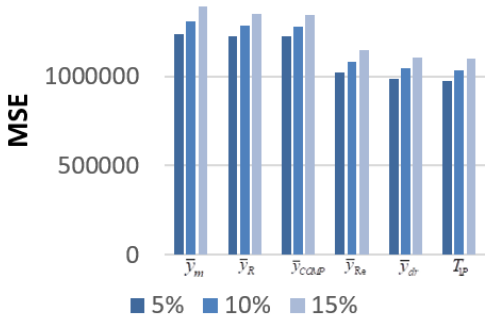


Figure 7: MSE comparison to non-response rate for population VII.

2. The estimator T_p not only achieves the minimum theoretical mean squared error (MSE) but also demonstrates superior empirical performance across varying non-response rates.

3. These findings provide strong evidence of the practical applicability of the proposed estimator in real-world survey settings, where non-response is inevitable.

4 Conclusions

This study proposes a generalized class of ratio-type estimators for estimating the population mean by effectively incorporating auxiliary information. The proposed class includes several existing estimators as special cases and allows for the determination of optimal parameter values based on the minimum MSE criterion. Under these optimal conditions, the proposed estimators attain identical MSEs.

Theoretical analysis establishes that the proposed estimator achieves a lower MSE than the traditional sample mean, ratio, and compromised estimators. These findings are further supported by an empirical investigation based on PRE measures, which consistently demonstrates the superior efficiency and robustness of the proposed estimator across different populations and varying non-response rates.

Overall, the results confirm the practical usefulness of the proposed methodology in survey sampling applications where auxiliary information is available and non-response is unavoidable. The analysis in this paper is restricted to first-order approximations under simple random sampling. Future research may extend the proposed framework to more complex sampling designs, such as stratified and multi-phase sampling, and examine its performance in the presence of non-response and measurement errors. In addition, the methodology may be adapted for estimating other population parameters, including the median, variance, proportion, and related characteristics.

5 Acknowledgements

We express our gratitude to the editors and reviewers for their thorough evaluation of the manuscript and for their valuable comments and suggestions.

References

Anderson, T. W. (2003). *An Introduction to the Multivariate Statistical Analysis.* 3rd Edition, Wiley, New York.

Bahl, S. Tuteja, R. K. (1991). Ratio and product type exponential estimator. *Information and Optimization Sciences*, Vol, XII, I, 159–163.

Cochran, W. (1977). *Sampling Techniques*. 3rd Edition. John Wiley and Sons, New York.

Diana, G. and Francesco, P. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics-Theory and Methods*, **39** 3245–3251.

Kadilar, C. and Cingi, H. (2008). Estimators for the population mean in the case of missing data. *Communications in Statistics-Theory and Methods*, **37**, 2226–2236.

Kalton, G., Kasprzyk, D., and Santos, R. (1981). Issues of nonresponse and imputation in the survey of income and program participation. *In Current topics in survey sampling* (pp. 455-480). Academic Press.

Lee, H., Rancourt, E. and Sarndal, C. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, **10**, 231–243.

Meeden, G. (2000). A decision theoretic approach to imputation in finite population sampling. *Journal of the American Statistical Association*, **95**, 586–595.

Mukhopadhyay, P. (2000). *Theory and Methods of Survey Sampling*, Prentice-Hall of India, New Delhi, India.

Murthy, M. N. (1967). *Sampling Theory and Methods*. 2nd Edition, Statistical Publishing Society, Calcutta.

Pandey, R., Thakur, N. S. and Yadav, K. (2015). Estimation of population mean using exponential ratio type imputation method under survey non-response. *Journal of the Indian Statistical Association*, **53**, 89–107.

Shukla, D., Thakur, N. S. and Thakur, D.S. (2011). Linear combination-based imputation method for missing data in sample. *International Journal of Modern Engineering Research*, **1**, 580–596.

Singh, B. K., and Nath, K. (2018). Some imputation methods in two-phase sampling scheme for estimation of population mean. *Research Reviews: Journal of Statistics (RRJoST)*, **7**, 1–16.

Singh, S. and Deo, B. (2003). Imputation by power transformation. *Statistical Papers*, **44**, 555–579.

Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. *Metrika*, **51**, 267–276.

Singh, A. K., Singh, P. and Singh, V. K. (2014). Exponential-type compromised imputation in survey sampling, *Journal of the Statistics Applications and Probability*, **3**, 211–217.

Singh, Priyanka, Singh, A. K**.** and Singh, V. K. (2015a). On the use of compromised imputation for missing data using factor-type estimators. *Journal of Statistics Applications and Probability Letters* 2(2), 1-9.

Singh, Priyanka, Singh, A. K. and Singh, V. K. (2015b). Estimation of population mean under non-response using various imputation methods for stratified population. *Journal of Advance Computing*, **4**, 112–122.

Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 13–27, Research Paper
https://smps.um.ac.ir, ISSN: 3115-9370

Multimatrix variate distributions

José A. Díaz-García[†,*] **and Francisco J. Caro-Lopera**[‡]

[†]*Facultad de Zootecnia y Ecología, Universidad Autónoma de Chihuahua,Periférico Francisco R. Almada Km 1, Zootecnia, Chihuahua, 33820, Chihuahua, México*
[‡]*Institute of Basic Sciences and Faculty of Engineering, University of Medellin, Carrera 87 No.30-65, Medellin, Antioquia, Colombia*

Email(s): jadiaz@cimat.mx, fjcaro@udemedellin.edu.co

Abstract. A new family of distributions indexed by the class of matrix-variate elliptically contoured distributions is proposed, extending several known bimatrix-variate models. These multimatrix-variate distributions open new perspectives in classical distribution theory, which is typically grounded in probabilistic independence assumptions and frequently relies on unverified fitting laws. Most of the multimatrix models derived here are invariant under the spherical family, a property that facilitates inference and reduces the need for prior specification of the underlying distributions. This invariance clarifies the statistical methodology and addresses certain limitations present in existing approaches, such as copula-based models. The paper also presents a variety of special cases, fundamental properties, and generalizations. The proposed joint distributions allow flexible combinations across scalars, vectors, and matrices, making them adaptable to complex modeling requirements. Moreover, they are computationally tractable, enabling a wide range of practical applications. In particular, we provide a comprehensive example in molecular docking for SARS-CoV-2, illustrating the analysis of matrix-dependent samples.

Keywords: Bimatrix variate, Matricvariate, Matrix variate, Matrix variate elliptically contoured distributions, Random matrices.

Mathematics Subject Classification (2020): 62H10, 62E15, 60E05.

1 Introduction

The modern world and its paradigms for understanding natural phenomena increasingly demand holistic solutions that involve the consensus of multiple sciences. The so-termed new way of doing science, the scientific computing, is based on the postulate that the phenomena of post modernity intertwine multiple dependencies between knowledge and variables. The paradox of the butterfly's flapping effect, for example, confirms that the independence of events in a modelable scientific phenomenon is increasingly difficult to accept in the scientific community. From the epistemological point of view, a construct of

*Corresponding author
Received: 19 August 2025 / Revised: 27 January 2026 / Accepted: 16 February 2026

https://doi.org/10.22067/smps.2026.94999.1050

The Author(s). This is an open access article under CC BY 4.0.

independent variables immensely distances reality from the model that seeks to explain it. Unless an expert manages to completely isolate its universe of explanatory variables of a phenomenon and has proven certainty of its independence, it is generally difficult from the usual theory to arrive at an explanation of the interrelationship in question. Chaos theory and the divergence of non-deterministic phenomena bring modern statistics closer to the elucidation of techniques increasingly in line with the divergent and integrated thinking of the expert. Initially, statistics moved away from interdisciplinary work with science, assuming theories based on probabilistic independence due to the great convenience in the computational solution. Such is the case of the well-known linear models, where the assumption of independent variables allows results as powerful, but difficult to achieve in reality, such as the Gauss-Markov theorem. But perhaps the example of greatest mathematical convenience and widely disseminated in statistics corresponds to the information from probabilistically independent samples that immensely facilitate a definition of tractable likelihood formed by the simple product of the probability densities of each variable. Experts, who deep down know of a high dependence on the samples of their experiments, have had no alternative but to use classical likelihood to approximate the highly dependent phenomena that emerge from their laboratories. A more robust modeling of the real world undoubtedly requires greater planning and statistical analysis of experiments in an interdisciplinary dialogue with the expert, who dictates the main postulates of the phenomena they are trying to decipher. In this sense, statistics does not command decisions and applications, but rather realistically guides the solution that the expert guides. Instead of establishing its classic postulates about independence, normality, estimability, linearity, it seeks to resolve the paradigms that emerge from modern modeling; methodological proposals that increasingly speak of the opposite: dependence, skewed or non-Gaussian distributions, functions of multiple local minima or maxima, nonlinearity and nonparametric statistics, etc. In response to these requirements, statistics has proposed separate theories of great complexity, but they are still not very popular and only exist in the literature that is also frequented by statisticians, not by users of other sciences. The best example speaks of normality, which, like the independence of variables, is almost never met. The normal model has been replaced by elliptical contour distributions, allowing probability distributions to be considered with heavier and lighter tails and a greater or lesser degree of kurtosis, among other significant characteristics. Although a major problem emerges when the underlying distribution is not known a priori, then the decision is left to adjustment criteria, without any explanation from expert science. In this context, invariances under complete families of distributions emerge of utmost importance, so that the results are applied regardless of knowledge of the base distribution.

In some experiments and phenomena where more than one random variable is studied, say X e Y, it is assumed that these variables are probabilistically independent with densities $dF_X(x)$ and $dG_Y(y)$. Then, if we are interested in the joint density of X and Y, this reduces to the elementary product $dF_X(x)dG_Y(y)$. However, as we have said, modern chaos theories and divergent modeling of complex holistic problems warn against this assumption of independence, in which case finding the joint density is not an easy task. For example, in a study of climatological data it is of interest to study the random variables associated with temperature and precipitation. For practical purposes these variables are considered independent, with normal and gamma distributions, respectively. But it is known that the temperature and precipitation are not probabilistically independent random variables. Then we open the question about a joint distribution of X and Y when the variables are not probabilistically independent. Furthermore, it would be desirable that given the joint density function of the probabilistically dependent variables X and Y, the corresponding marginal distributions of the random variables X and Y were the marginal distributions that are usually assumed under independence. For the univariate case, several bivariate distributions have been studied by different authors, the models include gamma-beta, gamma-gamma, beta-beta among others. They have found applications in problems of hydrology, finance and other areas disciplines, see Libby and Novick (1982), Chen and Novick (1984), Nadarajah (2007, 2013), Olkin and Liu (2003) and Sarabia et al. (2014) and references therein. In general, in these works, joint distributions of dependent random variables are constructed, making changes of variables on independent

random variables. From the matrix point of view bimatrix variate type distribution have been studied by several authors, see Olkin and Rubin (1964), Díaz-García, and Gutiérrez-Jáimez (2010a,b, 2011), and Bekker et al. (2011) among others. Ehlers (2011) presents a very complete exposition of the advances in matrix variate case up to that moment. The authors have studied this problem for scalar and vector cases in Díaz-García et al. (2022) and a matrix case in the multimatricvariate version in Díaz-García and Caro-Lopera (2022). Finally, a multimatrix theory that solves the previous problem but that cannot be computed would add to the hundreds of theoretical results in matrix analysis developed in the last 70 years. Then we expect that the joint distribution functions of several matrices are free of approximations and truncation of complex series of polynomials of matrix arguments.

The above discussion is placed in the present article by proposing a new method to establish computable joint density functions of two or more random matrices, which shall be termed *multimatrix variate distributions*, such that the corresponding marginal distributions are known. Section 1 establishes the notation to be used and some preliminary results on the calculation of Jacobians and some integrals of interest as well as the definition of the matrix variate elliptically contoured distributions. The main distributional results are collected in Section 2. Some related distributions of multimatrix variate distributions and particular generalisations are obtained in Section 3. Finally, Section 4 provides a dependent sample emergent from a problem of molecular docking into a new cavity of SARS-CoV-2. In particular, the joint distribution of the dependent sample is estimated by a simple optimisation routine.

2 Preliminary results

Matrix notations, matrix variate elliptically contoured distributions, and Jacobians are presented in this section. First, we start with notations and terminologies. $\mathbf{A} \in \Re^{n\times m}$ denotes a *matrix* with n rows and m columns; $\mathbf{A}' \in \Re^{m\times n}$ is the *transpose matrix*, and if $\mathbf{A} \in \Re^{n\times n}$ has an *inverse*, it shall be denoted by $\mathbf{A}^{-1} \in \Re^{n\times n}$. $\mathbf{A} \in \Re^{n\times n}$ is a *symmetric matrix* if $\mathbf{A} = \mathbf{A}'$. If all their eigenvalues are positive then $\mathbf{A}$ is a *positive definite matrix*, a fact denoted as $\mathbf{A} > \mathbf{0}$. An *identity matrix* shall be denoted by $\mathbf{I} \in \Re^{n\times n}$. To specify the size of the identity, we shall use $\mathbf{I}_n$. $\mathrm{tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A} \in \Re^{m\times m}$. If $\mathbf{A} \in \Re^{n\times m}$ then by $\mathrm{vec}(\mathbf{A})$ we mean the $mn \times 1$ vector formed by stacking the columns of $\mathbf{A}$ under each other; that is, if $\mathbf{A} = [\mathbf{a}_1\mathbf{a}_2\ldots\mathbf{a}_m]$, where $\mathbf{a}_j \in \Re^{n\times 1}$ for $j = 1,2,\ldots,m$

$$\mathrm{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}.$$

The *Frobenius norm* of a matrix $\mathbf{A}$ shall be denoted as $||\mathbf{A}||$. Typically the Frobenius norm is denoted as $||\mathbf{A}||_F$, to differentiate it from other matrix norms. Since we shall use only the Frobenius norm, it just be denoted as $||\mathbf{A}||$. It is defined by

$$||\mathbf{A}|| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})} = \sqrt{\mathrm{vec}'(\mathbf{A})\,\mathrm{vec}(\mathbf{A})}.$$

Finally, $\mathscr{V}_{m,n}$ denotes the *Stiefel manifold*, the space of all matrices $\mathbf{H}_1 \in \Re^{n\times m}$ $(n \geq m)$ with orthogonal columns, that is, $\mathscr{V}_{m,n} = \{\mathbf{H}_1 \in \Re^{n\times m}; \mathbf{H}_1'\mathbf{H}_1 = \mathbf{I}_m\}$. In addition, if $(\mathbf{H}_1' d\mathbf{H}_1)$ defines an *invariant measure on the Stiefel manifold* $\mathscr{V}_{m,n}$, from Theorem 2.1.15, p. 70 in Muirhead (2005),

$$\int_{\mathscr{V}_{m,n}} (\mathbf{H}_1' d\mathbf{H}_1) = \frac{2^m \pi^{mn/2}}{\Gamma_m[n/2]}, \tag{1}$$

Goodall and Mardia (1993) proposed the term *matrix multivariate* instead of *matrix variate*, commonly used in multivariate analysis literature, Gupta and Varga (1993), among many other authors. The term *matricvariate* was introducing by Dickey (1967) to distinguish between the two possible expressions of the density matrix multivariate T-distribution. A detail explanation about this topic is given in (Díaz-García, and Gutiérrez-Jáimez, 2006, see Section 3).

where $\Gamma_m[a]$ denotes the multivariate gamma function, see (Muirhead, 2005, Definiton 2.1.10, p. 61)

Next a result about Jacobians is derived.

Theorem 1. *Suppose that* $\mathbf{X} \in \Re^{n\times m}$ *and* $\mathbf{Y} \in \Re^{n\times m}$ *are matrices with mathematically independent elements.*

i) *Let* $\mathbf{Y} = (1-\operatorname{tr}\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}$. *Then*

$$(d\mathbf{Y}) = (1-\operatorname{tr}\mathbf{X}'\mathbf{X})^{-(nm/2+1)}(d\mathbf{X}).$$

ii) *If* $\mathbf{X} = (1+\operatorname{tr}\mathbf{Y}'\mathbf{Y})^{-1/2}\mathbf{Y}$, *we have*

$$(d\mathbf{X}) = (1+\operatorname{tr}\mathbf{Y}'\mathbf{Y})^{-(nm/2+1)}(d\mathbf{Y}).$$

Proof. **i)** The proof follows from Theorem 1 in Díaz-García et al. (2022), observing that $\mathbf{Y} = (1-\operatorname{tr}\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}$ can be written as $\mathbf{y} = (1-\mathbf{x}'\mathbf{x})^{-1/2}\mathbf{x}$, where $\mathbf{x} = \operatorname{vec}\mathbf{X} \in \Re^{mn}$ and $\mathbf{y} = \operatorname{vec}\mathbf{Y} \in \Re^{mn}$. **ii)**. Its proof is analogous to one given for **i)**. □

Now, let $\mathbf{V} \in \Re^{N\times m}$ random matrix with a matrix variate elliptically contoured distribution with respect to the Lebesgue measure $(d\mathbf{V})$. Therefore its density function is given by

$$dF_{\mathbf{V}}(\mathbf{V}) = |\boldsymbol{\Sigma}|^{-N/2}|\boldsymbol{\Theta}|^{-m/2}h\left\{\operatorname{tr}\left[(\mathbf{V}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{V}-\boldsymbol{\mu})\boldsymbol{\Theta}^{-1}\right]\right\}(d\mathbf{V}). \tag{2}$$

The location parameter is $\boldsymbol{\mu} \in \Re^{N\times m}$; and the scale parameters $\boldsymbol{\Sigma} \in \Re^{N\times N}$ and $\boldsymbol{\Theta} \in \Re^{m\times m}$, are positive definite matrices. The distribution shall be denoted by $\mathbf{V} \sim \mathscr{E}_{N\times m}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Theta};h)$, and indexed by the kernel function h: $\Re \to [0,\infty)$, where

$$\int_0^\infty u^{Nm/2-1}h(u)du < \infty.$$

When $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}_N$ and $\boldsymbol{\Theta} = \mathbf{I}_m$ a special case of a matrix variate elliptically contoured distribution appears, in this case it is said that $\mathbf{V}$ has a *matrix variate spherical distribution*.

Note that for constant $a \in \Re$, the substitution $v = u/a$ and $(du) = a(dv)$ in (Fang et al., 1990, Equation 2.21, p. 26) provides that

$$\int_{v>0} v^{Nm/2-1}h(av)(dv) = \frac{a^{-Nm/2}\Gamma_1[Nm/2]}{\pi^{Nm/2}}. \tag{3}$$

3 Multimatrix variate distributions

Assume that $\mathbf{X} \sim \mathscr{E}_{N\times m}(\mathbf{0},\mathbf{I}_N,\mathbf{I}_m;h)$, such that $n_0+n_1+\cdots+n_k = N$, and $\mathbf{X} = \left(\mathbf{X}_0',\mathbf{X}_1',\ldots,\mathbf{X}_k'\right)'$. Then (2) can be written as

$$dF_{\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k}(\mathbf{X}_1,\ldots,\mathbf{X}_k) = h[\operatorname{tr}(\mathbf{X}_0'\mathbf{X}_0+\mathbf{X}_1'\mathbf{X}_1+\cdots+\mathbf{X}_k'\mathbf{X}_k)]\bigwedge_{i=0}^{k}(d\mathbf{X}_i), \tag{4}$$

or in terms of the Frobenius norm, we can rewrite (4) as

$$dF_{\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k}(\mathbf{X}_1,\ldots,\mathbf{X}_k) = h\left(||\mathbf{X}_0||^2+||\mathbf{X}_1||^2+\cdots+||\mathbf{X}_k||^2\right)\bigwedge_{i=0}^{k}(d\mathbf{X}_i),$$

where $\mathbf{X}_i \in \Re^{n_i\times m}$, $n_i \geq m \geq 1$, $i = 0,1,\ldots,k$. Note that the random matrices $\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k$ are probabilistically dependent. Furthermore, only under a matrix variate normal distribution, the random matrices are independent, see Fang and Zhang (1990).

Analogously to (Díaz-García et al., 2022, Equation 4.2) but now defining $V_i = ||\mathbf{X}_i||^2$, $i = 0,1,\ldots,k$, it is obtained

$$dF_{V_0,\ldots,V_k}(v_0,\ldots,v_k) = \frac{\pi^{Nm/2}}{\prod_{i=0}^{k}\Gamma[n_i m/2]} h\left(\sum_{i=0}^{k} v_i\right) \prod_{i=0}^{k} v_i^{n_i m/2-1} \bigwedge_{i=0}^{k}(du_i),$$

where $N = n_0 + \cdots + n_k$, this distribution shall be termed *multivariate generalised gamma distribution* and its univariate case ($i = 0$) shall be termed *generalised gamma distribution*.

Theorem 2. *Suppose that* $\mathbf{X} = \left(\mathbf{X}_0', \ldots, \mathbf{X}_k'\right)'$ *has a matrix variate spherical distribution, with* $\mathbf{X}_i \in \Re^{n_i \times m}$, $n_i \geq m$, $i = 0,1,\ldots,k$. *Define* $V = ||\mathbf{X}_0||^2$. *Then, the joint density* $dF_{V,\mathbf{X}_1,\ldots,\mathbf{X}_k}(v,\mathbf{X}_1,\ldots,\mathbf{X}_k)$ *is given by*

$$\frac{\pi^{n_0 m/2}}{\Gamma_1[n_0 m/2]} h\left[v + \sum_{i=1}^{k} ||\mathbf{X}_i||^2\right] v^{n_0 m/2-1}(dv) \bigwedge_{i=1}^{k}(d\mathbf{X}_i), \tag{5}$$

where $V > 0$. *This distribution shall be termed* multimatrix variate generalised gamma-elliptical distribution.

Proof. We have that

$$dF_{\mathbf{X}_0,\mathbf{X}_1,\ldots,\mathbf{X}_k}(\mathbf{X}_1,\ldots,\mathbf{X}_k) = h\left(||\mathbf{X}_0||^2 + ||\mathbf{X}_1||^2 + \cdots + ||\mathbf{X}_k||^2\right) \bigwedge_{i=0}^{k}(d\mathbf{X}_i).$$

Define $V = ||\mathbf{X}_0||^2$ and from (Muirhead, 2005, Theorem 2.1.14, p. 65), we have that

$$(d\mathbf{X}_0) = 2^{-1} v^{n_0 m/2-1}(dv) \wedge (\mathbf{h}_1 d\mathbf{h}_1),$$

where $\mathbf{h}_1 \in \mathscr{V}_{1,n_0 m}$. Thus, the joint density of $V, \mathbf{h}_1, \mathbf{X}_1, \ldots, \mathbf{X}_k$ is

$$\frac{v^{n_0 m/2-1}}{2} h\left(v + ||\mathbf{X}_1||^2 + \cdots + ||\mathbf{X}_k||^2\right)(dv) \wedge (\mathbf{h}_1 d\mathbf{h}_1) \bigwedge_{i=1}^{k}(d\mathbf{X}_i).$$

Integration over $\mathbf{h}_1 \in \mathscr{V}_{1,n_0 m}$ by using (1) gives the required result. □

Theorem 3. *Assume that* $\mathbf{X} = \left(\mathbf{X}_0', \ldots, \mathbf{X}_k'\right)'$ *has a matrix variate spherical distribution, with* $\mathbf{X}_i \in \Re^{n_i \times m}$, $n_i \geq m$, $i = 0,1,\ldots,k$. *Define* $V = ||\mathbf{X}_0||^2$ *and* $\mathbf{T}_i = V^{-1/2}\mathbf{X}_i$, $i = 1,\ldots,k$. *The joint density* $dF_{V,\mathbf{T}_1,\ldots,\mathbf{T}_k}(v,\mathbf{T}_1,\ldots,\mathbf{T}_k)$ *is given by*

$$\frac{\pi^{n_0 m/2}}{\Gamma_1[n_0 m/2]} h\left[v\left(1 + \sum_{i=1}^{k} ||\mathbf{T}_i||^2\right)\right] v^{Nm/2-1}(dv) \bigwedge_{i=1}^{k}(d\mathbf{T}_i), \tag{6}$$

where $N = n_0 + n_1 + \cdots + n_k$, $V > 0$ *and* $\mathbf{T}_i \in \Re^{n_i \times m}$, $i = 1,\ldots,k$. *This distribution shall be termed* multimatrix variate generalised gamma-Pearson type VII distribution. *Moreover, the termed* multimatrix variate Pearson type VII *is the marginal density* $dF_{\mathbf{T}_1,\ldots,\mathbf{T}_k}(\mathbf{T}_1,\ldots,\mathbf{T}_k)$ *of* $\mathbf{T}_1,\ldots,\mathbf{T}_k$ *and is given by*

$$\frac{\Gamma_1[Nm/2]}{\pi^{(N-n_0)m/2}\Gamma_1[n_0 m/2]} \left(1 + \sum_{i=1}^{k} ||\mathbf{T}_i||^2\right)^{-Nm/2} \bigwedge_{i=1}^{k}(d\mathbf{T}_i), \tag{7}$$

where $\mathbf{T}_i \in \Re^{n_i \times m}$, $n_i \geq m$ *and* $N = n_0 + n_1 + \cdots + n_k$.

Proof. The density (6) follows from (5) by defining $\mathbf{T}_i = V^{-1/2}\mathbf{X}_i$, $i = 1,\ldots,k$. Hence by (Muirhead, 2005, Theorem 2.1.4, p. 57), we have

$$\bigwedge_{i=1}^{k}(d\mathbf{X}_i) = \prod_{i=1}^{k} v^{n_i m/2} \bigwedge_{i=1}^{k}(d\mathbf{T}_i).$$

And then the required result follows. Now, integrating (6) over $V > 0$, by using (3), provides the density (7). □

Theorem 4. *Suppose that* $\mathbf{X} = \left(\mathbf{X}_0', \ldots, \mathbf{X}_k'\right)'$ *has a matrix variate spherical distribution, with* $\mathbf{X}_i \in \Re^{n_i \times m}$, $n_i \geq m$, $i = 0,1,\ldots,k$. *Define* $V = ||\mathbf{X}_0||^2$ *and* $\mathbf{R}_i = \left(V + ||\mathbf{X}_i||^2\right)^{-1/2}\mathbf{X}_i$, $i = 1,\ldots,k$. *Then the joint density of* $V, \mathbf{R}_1, \ldots, \mathbf{R}_k$, *denoted as* $dF_{V,\mathbf{R}_1,\ldots,\mathbf{R}_k}(v, \mathbf{R}_1, \ldots, \mathbf{R}_k)$, *is given by*

$$\frac{\pi^{n_0 m/2}}{\Gamma_1[n_0 m/2]} v^{Nm/2-1} h\left[v\left(1 + \sum_{i=1}^{k} \frac{||\mathbf{R}_i||^2}{(1-||\mathbf{R}_i||^2)}\right)\right] \prod_{i=1}^{k}\left(1-||\mathbf{R}_i||^2\right)^{-n_i m/2-1} (dv) \bigwedge_{i=1}^{k}(d\mathbf{R}_i), \tag{8}$$

where $N = n_0 + n_1 + \cdots + n_k$, $V > 0$ *and* $\mathbf{R}_i \in \Re^{n_i \times m}$, *and* $||\mathbf{R}_i||^2 \leq 1$ $i = 1,\ldots,k$. *This distribution shall be termed* multimatrix variate generalised gamma-Pearson type II distribution. *Also, the marginal density* $dF_{\mathbf{R}_1,\ldots,\mathbf{R}_k}(\mathbf{R}_1, \ldots, \mathbf{R}_k)$ *is*

$$\frac{\Gamma_1[Nm/2]}{\pi^{(N-n_0)m/2}\Gamma_1[n_0 m/2]}\left[1 + \sum_{i=1}^{k} \frac{||\mathbf{R}_i||^2}{(1-||\mathbf{R}_i||^2)}\right]^{-Nm/2} \prod_{i=1}^{k}\left(1-||\mathbf{R}_i||^2\right)^{-n_i m/2-1} \bigwedge_{i=1}^{k}(d\mathbf{R}_i), \tag{9}$$

which shall be termed multimatrix variate Pearson type II distribution.

Proof. Consider the substitution $\mathbf{T}_i = (1-||\mathbf{R}_i||^2)^{-1/2}\mathbf{R}_i$, $i = 1,\ldots,k$ in (6), hence by Theorem 1,

$$\bigwedge_{i=1}^{k}(d\mathbf{T}_i) = \prod_{i=1}^{k}\left(1-||\mathbf{R}_i||^2\right)^{-n_i m/2-1} \bigwedge_{i=1}^{k}(d\mathbf{R}_i),$$

and the desired result follows. Now (9) follows by integration of (8) over $V > 0$ via (3). □

Theorem 5. *Consider the hypotheses of Theorem 3 and define* $\mathbf{F}_i = \mathbf{T}_i'\mathbf{T}_i > \mathbf{0}$, $i = 1,\ldots,k$. *Then the joint density* $dF_{V,\mathbf{F}_1,\ldots,\mathbf{F}_k}(v, \mathbf{F}_1, \ldots, \mathbf{F}_k)$ *is given by*

$$\frac{\pi^{Nm/2} v^{Nm/2-1}}{\Gamma_1[n_0 m/2]} \prod_{i=1}^{k}\left(\frac{|\mathbf{F}_i|^{(n_i-m-1)/2}}{\Gamma_m[n_i/2]}\right) h\left[v\left(1 + \sum_{i=1}^{k} \operatorname{tr}\mathbf{F}_i\right)\right](dv) \bigwedge_{i=1}^{k}(d\mathbf{F}_i). \tag{10}$$

This distribution shall be termed multimatrix variate generalised gamma-beta type II distribution. *The corresponding marginal density function* $dF_{\mathbf{F}_1,\ldots,\mathbf{F}_k}(\mathbf{F}_1, \ldots, \mathbf{F}_k)$ *is given by*

$$\frac{\Gamma_1[Nm/2]}{\Gamma_1[n_0 m/2]\prod_{i=1}^{k}\Gamma_m[n_i/2]} \prod_{i=1}^{k}|\mathbf{F}_i|^{(n_i-m-1)/2}\left(1 + \sum_{i=1}^{k} \operatorname{tr}\mathbf{F}_i\right)^{-Nm/2} \bigwedge_{i=1}^{k}(d\mathbf{F}_i), \tag{11}$$

and shall be termed multimatrix variate beta type II distribution.

Proof. Joint density functions (10) and (11) follow from substitutions $\mathbf{F}_i = \mathbf{T}_i'\mathbf{T}_i$, $i = 1,\ldots,k$ in (6) and (7), respectively. First, note that (Muirhead, 2005, Theorem 2.1.14, p. 66) provides

$$\bigwedge_{i=1}^{k}(d\mathbf{T}_i) = 2^{-mk}\prod_{i=1}^{k}|\mathbf{F}_i|^{(n_i-m-1)/2} \bigwedge_{i=1}^{k}(d\mathbf{F}_i) \bigwedge_{i=1}^{k}\left(\mathbf{H}_{1_i}'d\mathbf{H}_{1_i}\right),$$

where $\mathbf{H}_{1_i} \in \mathscr{V}_{n_i,m}$, $i = 1,\ldots,k$. Hence, integrating over $\mathbf{H}_{1_i} \in \mathscr{V}_{n_i,m}$ $i = 1,\ldots,k$ by (1) turns

$$\int_{\mathbf{H}_{1_1}} \cdots \int_{\mathbf{H}_{1_k}} \bigwedge_{i=1}^{k} (\mathbf{H}_{1_i}' d\mathbf{H}_{1_i}) = \frac{2^{mk}\pi^{(N-n_0)m/2}}{\prod_{i=1}^{k}\Gamma_m[n_i/2]},$$

and the required result is derived. □

A particular case of the density function (11) is proposed in (Muirhead, 2005, Problem 3.18, p.118).

Theorem 6. *Assuming that* $\mathbf{B}_i = \mathbf{R}_i'\mathbf{R}_i > \mathbf{0}$ *and* $\operatorname{tr}\mathbf{B}_i \leq 1$ *with* $i = 1,\ldots,k$, *in Theorem 4. Then the joint density* $dF_{V,\mathbf{B}_1,\ldots,\mathbf{B}_k}(v,\mathbf{B}_1,\ldots,\mathbf{B}_k)$ *is*

$$\frac{\pi^{Nm/2}\, v^{Nm/2-1}}{\Gamma_1[n_0m/2]} \prod_{i=1}^{k}\left(\frac{|\mathbf{B}_i|^{(ni-m-1)/2}}{\Gamma_m[n_i/2]}\right) h\left[v\left(1+\sum_{i=1}^{k}\frac{\operatorname{tr}\mathbf{B}_i}{(1-\operatorname{tr}\mathbf{B}_i)}\right)\right] \prod_{i=1}^{k}(1-\operatorname{tr}\mathbf{B}_i)^{-n_im/2-1}\,(dv)\bigwedge_{i=1}^{k}(d\mathbf{B}_i).$$

This distribution shall be termed multimatrix variate generalised gamma-beta type I distribution. *Moreover, the density function* $dF_{\mathbf{B}_1,\ldots,\mathbf{B}_k}(\mathbf{B}_1,\ldots,\mathbf{B}_k)$ *can written as*

$$\frac{\Gamma_1[Nm/2]}{\Gamma_1[n_0m/2]} \prod_{i=1}^{k}\left(\frac{|\mathbf{B}_i|^{(ni-m-1)/2}}{\Gamma_m[n_i/2]}\right)\left[1+\sum_{i=1}^{k}\frac{\operatorname{tr}\mathbf{B}_i}{(1-\operatorname{tr}\mathbf{B}_i)}\right]^{-Nm/2} \prod_{i=1}^{k}(1-\operatorname{tr}\mathbf{B}_i)^{-n_im/2-1}\bigwedge_{i=1}^{k}(d\mathbf{B}_i),$$

and this marginal distribution shall be called multimatrix variate beta type I distribution.

Proof. A similar procedure to the proof of Theorem 5 can be applied here. Just consider the joint density functions (8) and (9) under the change of variable $\mathbf{B}_i = \mathbf{R}_i'\mathbf{R}_i$ with $i = 1,\ldots,k$, and the derivation is complete. □

Finally, we have the following result.

Theorem 7. *Assume that* $\mathbf{X} = (\mathbf{X}_0',\ldots,\mathbf{X}_k')'$ *has a matrix variate spherical distribution, with* $\mathbf{X}_i \in \Re^{n_i\times m}$, $n_i \geq m$, $i = 0,1,\ldots,k$. *Define* $V = ||\mathbf{X}_0||^2$ *and* $\mathbf{W}_i = \mathbf{X}_i'\mathbf{X}_i > \mathbf{0}$, $i = 1,\ldots,k$.
Then, the joint density $dF_{V,\mathbf{W}_1,\ldots,\mathbf{W}_k}(v,\mathbf{W}_1,\ldots,\mathbf{W}_k)$ *is given by*

$$\frac{\pi^{Nm/2}v^{n_0m/2-1}}{\Gamma_1[n_0m/2]} \prod_{i=1}^{k}\left(\frac{|\mathbf{W}_i|^{(ni-m-1)/2}}{\Gamma_m[n_i/2]}\right) h\left[v+\sum_{i=1}^{k}\operatorname{tr}\mathbf{W}_i\right](dv)\bigwedge_{i=1}^{k}(d\mathbf{W}_i), \tag{12}$$

where $V > 0$. *This distribution shall be termed* multimatrix variate generalised gamma - generalised Wishart distribution.
Similarly, let $\mathbf{W}_i = \mathbf{X}_i'\mathbf{X}_i > \mathbf{0}$, $i = 1,\ldots,k$. *Therefore, the joint density* $dF_{\mathbf{W}_1,\ldots,\mathbf{W}_k}(\mathbf{W}_1,\ldots,\mathbf{W}_k)$ *is*

$$\pi^{(N-n_0)m/2}\left(\prod_{i=1}^{k}\frac{|\mathbf{W}_i|^{(ni-m-1)/2}}{\Gamma_m[n_i/2]}\right) h\left(\operatorname{tr}\sum_{i=1}^{k}\mathbf{W}_i\right)\bigwedge_{i=1}^{k}(d\mathbf{W}_i). \tag{13}$$

Proof. From Theorem 2 we have

$$\frac{\pi^{n_0m/2}}{\Gamma_1[n_0m/2]} h\left[v+\sum_{i=1}^{k}||\mathbf{X}_i||^2\right] v^{n_0m/2-1}(dv)\bigwedge_{i=1}^{k}(d\mathbf{X}_i).$$

Defining $\mathbf{W}_i = \mathbf{X}_i'\mathbf{X}_i$ with $i = 1,\ldots,k$ and proceeding as in the proof of Theorem 5, the result (12) is immediate. (13) is obtained in Díaz-García and Caro-Lopera (2022). □

Several multivariate density functions studied in the statistical literature are obtained as particular cases of the distributions proposed in this section. For example the distribution proposed in (Muirhead, 2005, Problem 3.18, p.118) and the distributions in (Gupta and Varga, 1993, Subsections 2.7.3.2 and 2.7.3.3, p. 54).

4 Some properties and generalization

This section focus on multimatrix variate distributions for two and three matrix arguments. Additionally, if the matrix arguments are non-singular, the corresponding inverse distributions are also obtained. Despite a laborious task, the proposed novel method can consider several marginal distributions. The content of the section is an useful combination for multiple possible uses in real phenomena guide by experts with the corresponding previous knowledge of the marginal distributions.

Theorem 8. *Assume that* $\mathbf{X}=(\mathbf{X}_0',\mathbf{X}_1',\mathbf{X}_2')'$ *has a matrix variate spherical distribution, with* $\mathbf{X}_i\in\Re^{n_i\times m}$, $n_i\geq m$, $i=0,1,2$. *Define* $V=||\mathbf{X}_0||^2$, $\mathbf{T}=V^{-1/2}\mathbf{X}_1$, *and* $\mathbf{R}=(V+||\mathbf{X}_2||)^{-1/2}\mathbf{X}_2$. *The joint density* $dF_{V,\mathbf{T},\mathbf{R}}(v,\mathbf{T},\mathbf{R})$ *is given by*

$$\frac{\pi^{n_0m/2}v^{Nm/2-1}}{\Gamma_1[n_0m/2]}h\left\{v\left[1+\left(1-||\mathbf{R}||^2\right)||\mathbf{T}||^2\right]\right\}\left(1-||\mathbf{R}||^2\right)^{(n_0+n_1)m/2-1}(dv)\wedge d\mathbf{T})\wedge d\mathbf{R}). \tag{14}$$

where $N=n_0+n_1+n_2$, $V>0$, $\mathbf{T}\in\Re^{n_1\times m}$, $\mathbf{R}\in\Re^{n_2\times m}$ *and* $||\mathbf{R}||^2\leq 1$. *This distribution shall be termed* threematrix variate generalized gamma-Pearson type VII - Pearson type II distribution. *Moreover, the termed* bimatrix variate Pearson type VII - Pearson type II distribution *is the marginal density* $dF_{\mathbf{T},\mathbf{R}}(\mathbf{T},\mathbf{R})$ *of* $\mathbf{T},\mathbf{R}$ *and is given by*

$$\frac{\Gamma_1[Nm/2]}{\pi^{(N-n_0)m/2}\Gamma_1[n_0m/2]}\left[\left(1+\left(1-||\mathbf{R}||^2\right)||\mathbf{T}||^2\right]^{-Nm/2}\left(1-||\mathbf{R}||^2\right)^{(n_0+n_1)m/2-1}(d\mathbf{T})\wedge d\mathbf{R}). \tag{15}$$

Proof. By Theorem 2

$$\frac{\pi^{n_0m/2}}{\Gamma_1[n_0m/2]}h\left[v+||\mathbf{X}_1||^2+||\mathbf{X}_2||^2\right]v^{n_0m/2-1}(dv)\wedge(d\mathbf{X}_1)\wedge(d\mathbf{X}_2),$$

where $V=||\mathbf{X}_0||^2>0$. Let $V_0=V+||\mathbf{X}_2||^2$, $\mathbf{T}=V^{-1/2}\mathbf{X}_1$, and $\mathbf{R}=(V+||\mathbf{X}_2||)^{-1/2}\mathbf{X}_2$. Hence $V=V_0-||\mathbf{X}_2||^2=V_0-||V_0\mathbf{R}_2||^2=V_0(1-||\mathbf{R}||^2)$, $\mathbf{T}=[V_0(1-||\mathbf{R}||^2)]^{-1/2}\mathbf{X}_1$, and $\mathbf{R}=V_0^{-1/2}\mathbf{X}_2$. Therefore, $(dv)=(dv_0)$ and

$$(dv)\wedge(d\mathbf{X}_1)\wedge(d\mathbf{X}_2)=v_0^{(n_1+n_2)m/2}(1-||\mathbf{R}||^2)^{n_1m/2}(dv_0)\wedge(d\mathbf{T})\wedge(d\mathbf{R}),$$

thus the desired result is obtained. The density (15) is achieved by integrating (14) over $V>0$ via (3). □

Corollary 1. *Under the hypotheses of Theorem 8, define* $\mathbf{F}=\mathbf{T}'\mathbf{T}>\mathbf{0}$ *and* $\mathbf{B}=\mathbf{R}'\mathbf{R}>\mathbf{0}$. *Then the joint density function* $dF_{v,\mathbf{F},\mathbf{B}}(V,\mathbf{F},\mathbf{B})$ *can be written as*

$$\frac{\pi^{Nm/2}v^{Nm/2-1}|\mathbf{F}|^{(n_1-m-1)/2}|\mathbf{B}|^{(n_2-m-1)/2}}{\Gamma_1[n_0m/2]\Gamma_m[n_1/2]\Gamma_m[n_2m/2]}h\left\{v\left[1+(1-\operatorname{tr}\mathbf{B})\operatorname{tr}\mathbf{F}\right]\right\}(1-\operatorname{tr}\mathbf{B})^{(n_0+n_1)m/2-1}(dv)\wedge d\mathbf{F})\wedge d\mathbf{B}). \tag{16}$$

This distribution shall be termed threematrix variate generalised gamma - beta type II - beta type I distribution. *And the corresponding marginal density function* $dF_{\mathbf{F},\mathbf{B}}(\mathbf{F},\mathbf{B})$ *is given by*

$$\frac{\Gamma_1[Nm/2]|\mathbf{F}|^{(n_1-m-1)/2}|\mathbf{B}|^{(n_2-m-1)/2}}{\Gamma_1[n_0m/2]\Gamma_m[n_1/2]\Gamma_m[n_2m/2]}\left[1+(1-\operatorname{tr}\mathbf{B})\operatorname{tr}\mathbf{F}\right]^{-Nm/2}(1-\operatorname{tr}\mathbf{B})^{(n_0+n_1)m/2-1}(d\mathbf{F})\wedge d\mathbf{B}), \tag{17}$$

a distribution that shall be called bimatrix variate beta type II - beta type I distribution.

Proof. Results (16) and (17) follow from (14) and (15), respectively, by substitutions $\mathbf{F}=\mathbf{T}'\mathbf{T}$ and $\mathbf{B}=\mathbf{R}'\mathbf{R}$ and the use of (Muirhead, 2005, Theorem 2.1.14, p. 66). Thus,

$$(d\mathbf{T})\wedge d\mathbf{R})=2^{-2m}|\mathbf{F}|^{(n_1-m-1)/2}|\mathbf{B}|^{(n_2-m-1)/2}(d\mathbf{F})\wedge d\mathbf{B})\bigwedge_{i=1}^{2}\left(\mathbf{H}_{1_i}'d\mathbf{H}_{1_i}\right).$$

Then, the demonstration is complete after a repeated implementation of (1). □

Note that different families of distributions can be obtained by replicating the procedure of Theorem 8. For example, if the change of variable on $\mathbf{X}_2$ is not carried out the corresponding distribution obtained can be termed *threematrix* variate generalized gamma-elliptical-Pearson type II distribution, and so on. Moreover, if $n_2 = 1$ is set in Theorem 8, the corresponding distribution can be termed threematrix variate generalised gamma-Pearson type VII-vector Pearson type II distribution, a fact that shows the versatility of the method proposed in this paper. In this case, a simultaneous model can involve a random variable V, a random matrix $\mathbf{T}$ and a random vector $\mathbf{r} = V^{-1/2}\mathbf{x}_2$.

In addition, when the multimatrix variate distributions involve non singular matrix arguments, we can find their inverse distribution (or if their arguments are rectangular or singular matrices, we can propose their corresponding generalise inverse distributions). For the distribution defined in Corollary 1, there are different possible multimatrix variate inverse distributions according to the selected argument combination of 1, 2 or 3 matrices. The following case exhibits the inverse when the three arguments are considered.

Theorem 9. *Assume the hypotheses of the Corollary 9 and define* $W = V^{-1}$, $\mathbf{A} = \mathbf{F}^{-1}$ *and* $\mathbf{U} = \mathbf{B}^{-1}$. *Then the joint density* $dF_{W,\mathbf{A},\mathbf{U}}(w,\mathbf{A},\mathbf{U})$ *can be written as*

$$\frac{\pi^{Nm/2}\nu^{Nm/2-1}|\mathbf{A}|^{-n_1/2}|\mathbf{U}|^{-n_2/2}}{\Gamma_1[n_0m/2]\Gamma_m[n_1/2]\Gamma_m[n_2m/2]}h\left\{w^{-1}\left[1+\left(1-\operatorname{tr}\mathbf{U}^{-1}\right)\operatorname{tr}\mathbf{A}^{-1}\right]\right\}\left(1-\operatorname{tr}\mathbf{U}^{-1}\right)^{(n_0+n_1)m/2-1}(dw)\wedge d\mathbf{A})\wedge d\mathbf{U}). \quad (18)$$

This distribution shall be called threematrix variate inverse generalised gamma-inverse beta type II - inverse beta type I distribution. *And the marginal density function* $dF_{\mathbf{A},\mathbf{U}}(\mathbf{A},\mathbf{U})$ *is given by*

$$\frac{\Gamma_1[Nm/2]|\mathbf{A}|^{-n_1/2}|\mathbf{U}|^{-n_2/2}}{\Gamma_1[n_0m/2]\Gamma_m[n_1/2]\Gamma_m[n_2m/2]}\left[1+\left(1-\operatorname{tr}\mathbf{U}^{-1}\right)\operatorname{tr}\mathbf{A}^{-1}\right]^{-Nm/2}\left(1-\operatorname{tr}\mathbf{U}^{-1}\right)^{(n_0+n_1)m/2-1}(d\mathbf{A})\wedge d\mathbf{U}). \quad (19)$$

A distribution that shall be termed bimatrix variate inverse beta type II - inverse beta type I distribution.

Proof. The density functions (18) and (19) are achieved from (16) and (17) by defining $W = V^{-1}$, $\mathbf{A} = \mathbf{F}^{-1}$ and $\mathbf{U} = \mathbf{B}^{-1}$ and by applying (Muirhead, 2005, Theorem 2.1.8, p. 59). □

Observe that the parameter domain of the multimatrix variate distributions can be extended to the complex or real fields. However their geometrical and/or statistical explication perhaps can be lost. These distributions are valid if we replace $n_i/2$ by a_i, $n_0m/2$ by a_0 and $Nm/2$ by a. Where the a'^s are complex numbers with positive real part. From a practical point of view for parametric estimation, this domain extension allows the use of nonlinear optimisation rather integer nonlinear optimisation, among other possibilities.

Moreover, note that each particular distribution is indexed by a kernel $h(\cdot)$ which defines itself another family of distributions. This opens a wide set of possible distributions available for an expert according to a preferred knowledge of a plausible $h(\cdot)$.

Finally, each distribution can be re-parametrized in order to reach no isotropy or general settings of its density function. In analogy with the normal case, the expressions obtained here appear in their standard form. However, perform the substitution $\mathbf{T}_i = \mathbf{A}_i^{-1}(\mathbf{S}_i - \boldsymbol{\mu}_i)\mathbf{B}_i^{-1}/\sqrt{r_i}$ in (7); where $\mathbf{A}_i \in \Re^{n_i\times n_i}$, $\mathbf{B}_i \in \Re^{m\times m}$ are non-singular matrices such that $\boldsymbol{\Sigma}_i = \mathbf{A}_i'\mathbf{A}_i > \mathbf{0}$ and $\boldsymbol{\Theta}_i = \mathbf{B}_i\mathbf{B}_i' > \mathbf{0}$, $\boldsymbol{\mu}_i \in \Re^{n_i\times m}$ and $r_i > 0$, $i = 1,\ldots,k$. Then by (Muirhead, 2005, Theorem 2.1.5, p. 58), we obtain

$$(d\mathbf{T}_i) = r_i^{-n_im/2}|\boldsymbol{\Sigma}_i|^{-m/2}|\boldsymbol{\Theta}_i|^{-n_i/2}(d\mathbf{S}_i).$$

And finally, the density $dF_{\mathbf{S}_1,\ldots,\mathbf{S}_k}(\mathbf{S}_1,\ldots,\mathbf{S}_k)$ can be written in the robust form

$$\frac{\Gamma_1[Nm/2]}{\pi^{(N-n_0)m/2}\Gamma_1[n_0m/2]\left(\prod_{i=1}^k r_i^{n_im/2}|\boldsymbol{\Sigma}_i|^{m/2}|\boldsymbol{\Theta}_i|^{n_i/2}\right)}\left[1+\operatorname{tr}\sum_{i=1}^k(\mathbf{S}_i-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{S}_i-\boldsymbol{\mu}_i)\boldsymbol{\Theta}_i^{-1}\right]^{-Nm/2}\bigwedge_{i=1}^k(d\mathbf{S}_i). \quad (20)$$

Observe that the particular values $k=1$, and $Nm/2=q$, $(N=n_0+n_1)$, turns (20) into the matrix variate Pearson type VII distribution, see (Gupta and Varga, 1993, Subsection 2.7.3.2, p. 54). And setting $q=(n_1m+r)/2$ in the preceding density, then the matrix variate T-distribution with r degrees of freedom is achieved. Furthermore, taking $r=1$ in the last density, then the matrix variate Cauchy distribution is obtained.

5 Application of molecular docking in SARS-CoV-2

Statistical applications supported by real data are usually difficult because the computation and underlying mathematical context force simplifications. Restrictions often related with inference based on independent sample information. It is extremely rare in nature to find time (space) probabilistic independent experiments. The most challenging problems in several areas are dependent in time and-or space. Some of them, such world pandemic situations are crucial and precise of urgent solutions. When a virus as SARS-CoV-2 arrives, a drug design is reacquired in a record time in order to avoid millions of deaths worldwide. A simplification of the docking problem involves two main steps: identify a target in the virus-protein, usually called pocket, and obtain a molecule (termed ligand) which must be place into an adequate site of the virus, for inhibiting its deathly activity. The target usually contains thousand of atoms with hundreds of possible unknown pockets were the small ligand can be placed. Discover the pocket and place the ligand is a colossal task for a human, however, the use of artificial intelligence can provide suitable sites to be considered by the biologist. Recently, a non conventional artificial intelligence method was provided in Ramirez et al. (2022) by a suitable recent discipline termed shape theory. The docking problem requires the use of certain invariance as translation, rotation and scaling, in order to avoid unnecessary computations. Instead of using the noisy Euclidean space latent in the classical artificial intelligence, the search is performed in quotient spaces with the demand symmetries. In this case, shape theory was applied in molecular docking for an automatic localization of ligand binding pockets in large proteins. The docking phenomena is modeled in Ramirez et al. (2022) by a physics chemistry punctuation function based on Lennard-Jones potential type $6-12$ and $6-10$. It includes the expected time dependence where the matrix representation at each state of the pocket-ligand goes into more stable and optimal low energy. Then the probabilistic dependent sample records n matrices in time ordered by a decreasing Lennard-Jones potential until the convergence is reached and the docking is stable. For this example, we have used the crystal structure of COVID-19 main protease an its inhibitor N3 due to Jin et al. (2020) and Liu et al. (2020). The protein consists of 2387 atoms and the ligand is constituted by 21 atoms. Once (Ramirez et al., 2022, Th. Sec. 3.) is applied, a plausible pocket is reached and the ligand is rotated rigidly into the cavity, until the energy is stabilized. In this case, $k=56$ locations of the ligand are dependently obtained inside a new cavity of 241 atoms. Figure 1 shows the dependent sample for this example, they constitutes 21×3 matrices $\mathbf{T}_i$, $i=1,\ldots,56$.

Hence the sample matrices $\mathbf{F}_i=\mathbf{T}_i'\mathbf{T}_i$ shall be the core for estimation of the parameters in the likelihood, corresponding to the joint distribution of $\mathbf{F}_i$ in equation (11).

In this example, we consider a parameter space extension, from $n_i/2, i=0,\ldots,56$ to corresponding positive reals $a_i, i=0,\ldots,56$. Then the optimisation problem is reduced to find the maximum likelihood estimators of a_0, and $a=a_1=\cdots=a_{56}$ in the following dependent sample joint distribution:

$$\frac{\Gamma_1[(a_0+ka)m]}{\Gamma_1[a_0m]\Gamma_m^k[a]}\prod_{i=1}^{k}|\mathbf{F}_i|^{a-\frac{m+1}{2}}\left(1+\sum_{i=1}^{k}\operatorname{tr}\mathbf{F}_i\right)^{-(a_0+ka)m}.$$

For computations we have used the free licensed software R. The optimisation routine is performed in the package Optimx. The following estimation is obtained consistently by several methods and different random seeds for the initial parameters:

$$a_0=0.34397, a=0.19735$$

Figure 1: Dependent sample of 56 movements of the rigid ligand N3 with 21 atoms inside a new cavity of 241 atoms found on SARS-CoV-2 main protease. For illustration, the ligand is oversized respect the cavity (gray); their 21 atoms are colored as red (C), green (N), blue (O). The animation shows the progressive optimisation of the molecular docking according to a decreasing Lennard-Jones potential. First a positive energy indicates that the coupling is repulsive, however it is diminishing until negative values of strong affinity are reached and stabilized into an attractive complex ligand-pocket.

This estimation is also important because the invariance of the likelihood, under the complete family of spherical distribution, avoids the difficult a priori knowledge of the underlying model. Also, no fitting test is required. Note that the usual estimation based on an independent sample can not be applied here because the molecular docking is reached by the calibration of the Lennard-Jones potential which was set by the expert as a decreasing function of the dependent locus in the ligand.

In order to provide the consistency of the method in multiple scenarios, we compute an exhaustive optimization under a large set of models based on different samples. We have shown the estimation of the model for the 56 time-dependent molecular docking; we now consider the influence of each time by removing it from the experiment. In that case 56 models are estimated and the extremely very similar results can be seen in cluster 1 (black) of Figure 2 and Table 1.

If we define w as the number of extractions in the complete set of 56 dependent-time molecular docking, then after considering the provided estimations for $w=0$ and $w=1$, we now study all the plausible $\frac{56!}{(56-2)!2!}=1540$ models for $w=2$. Again the results behave as we expect, the maximal log likelihood decreases because the sample information is reduced, but 99% confidence intervals for $\hat{a}_0$ and $\hat{a}$ perform properly. According to Table 1, the resolution of the 1540 points is so narrow that cluster 2 (red) of Figure 2 seems to be a point. The cluster location sustains the preceding two estimations, as we expect. For $w=3$ extraction, we have the large amount of $\frac{56!}{(56-3)!3!}=27720$ possible combinations; Table 1 and reduced cluster 3 (green) of Figure 2 mantains the tendency. Finally, for $w=4,5,\ldots,28$ the computations are colosal to be performed in the simple PC used for optimization, then we have studied all the ordered successive set of extractions from the following way. For $w=4$, $(1:3,4),(1:3,5),\ldots,(1:3,56),(1:2,4,56),(1:2,5,56),\ldots,(1:2,55:56),(1,4,55:56),(1,5,55:56),\ldots,(1,54:56),(4,54:56),(5,54:56),\ldots,(53:56)$. For $w=5$, $(1:4,5),(1:4,6),\ldots,(1:4,56),(1:3,5,56),(1:3,6,56),\ldots,(1:3,55,56),(1:2,5,55:56),(1:2,6,55:56),\ldots,(1:2,54:56),(1,5,54:56),(1,6,54:56),\ldots,(1,53:56),(5,53:56),(6,53:56),\ldots,(52:56)$, and so and on; the final set is given for $w=28$, as $(1:27,28),(1:27,29),\ldots,(1:27,56),(1:26,28,56),(1:26,29,56),\ldots,(1:26,55:56),(1:25,28,55:56),(1:$

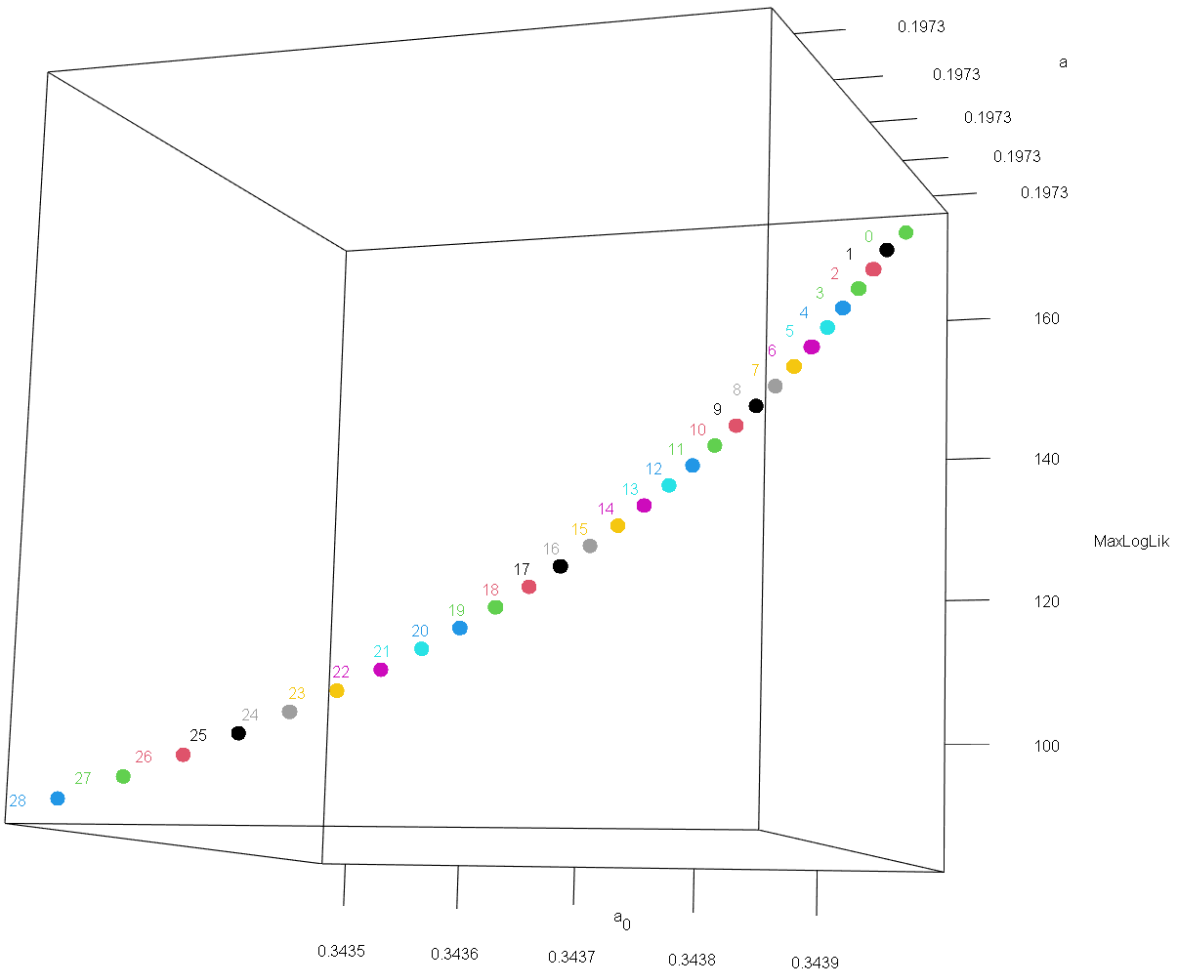


Figure 2: Evolution of maximal log likelihood information under several models with controlled substraction of dependent-time molecular docking. The models and estimations are described in Table 1. The full estimation for the complete set of 56 dependent-time is given in point 0 (black), with maximal log likelihood. The 56 possible models for a $w = 1$ subtraction are summarized in cluster 2 (red) which is over exceeded in size because the nitid resolution of the estimations. Clusters 2 (red) and 3 (green) represent the corresponding complete 1540 and 27720 models for extractions $w = 2$ and $w = 3$; again for displaying the estimation evolution the size is over exceeded, and they are confined in a very small space. The remaining extractions from $w = 4$ to $w = 28$ follows a particular but representative subset sequence given the prominent number of combination. They resulting expected monotone decreasing maximal log likelihood is provided and just verifies the gradual lost of meaningful sample information but the estimation of a_0 and a rounds the original estimation for the complete 56 dependent-time experiment.

Table 1: Exhaustive estimation for molecular docking under subtraction of w dependent-times samples. For each experiment from $w = 0$ to $w = 28$ extractions, the number M of studied models are shown with the corresponding 99% confidence intervals for estimations of "a_0", "a" and maximum log likelihood (MaxLogLik).

w	M	a_0	a	MaxLogLik
0	1	0.34396995340027808	0.19734727197157109	172.475
1	56	[0.3439581, 0.3439584]	[0.19734433, 0.19734435]	[169.3689293, 169.3689294]
2	1540	[0.343949, 0.343950]	[0.19734254, 0.19734257]	[166.2630596, 166.2630598]
3	27720	[0.343940, 0.343941]	[0.19734068, 0.19734070]	[163.1573540, 163.1573550]
4	206	[0.3439306, 0.3439312]	[0.19733875, 0.19733878]	[160.0518201, 160.0518202]
5	250	[0.3439209, 0.3439214]	[0.19733674, 0.19733676]	[156.9464627, 156.9464628]
6	291	[0.3439107, 0.3439115]	[0.19733466, 0.19733469]	[153.8412893, 153.8412895]
7	329	[0.3439001, 0.3439007]	[0.19733249, 0.19733251]	[150.7363072, 150.7363074]
8	364	[0.3438892, 0.3438894]	[0.19733023, 0.19733025]	[147.6315241, 147.6315243]
9	396	[0.3438777, 0.3438779]	[0.19732788, 0.19732790]	[144.5269480, 144.5269483]
10	425	[0.3438657, 0.3438662]	[0.19732543, 0.19732545]	[141.4225879, 141.4225882]
11	451	[0.3438532, 0.3438534]	[0.19732288, 0.19732289]	[138.3184529, 138.3184532]
12	474	[0.3438401, 0.3438403]	[0.19732021, 0.19732023]	[135.2145529, 135.2145531]
13	494	[0.3438264, 0.3438267]	[0.19731742, 0.19731743]	[132.1108983, 132.1108986]
14	511	[0.3438120, 0.3438125]	[0.19731450, 0.19731452]	[129.0075005, 129.0075008]
15	525	[0.3437971, 0.3437975]	[0.19731144, 0.19731146]	[125.9043716, 125.9043719]
16	536	[0.3437815, 0.3437817]	[0.19730824, 0.19730825]	[122.8015243, 122.8015246]
17	544	[0.3437648, 0.3437655]	[0.19730485, 0.19730488]	[119.6989727, 119.6989729]
18	549	[0.3437477, 0.3437479]	[0.19730132, 0.19730134]	[116.5967315, 116.5967318]
19	551	[0.3437292, 0.3437299]	[0.19729759, 0.19729764]	[113.4948170, 113.4948172]
20	550	[0.3437101, 0.3437107]	[0.19729367, 0.19729368]	[110.3932463, 110.3932466]
21	546	[0.3436900, 0.3436902]	[0.19728953, 0.19728954]	[107.2920384, 107.2920387]
22	539	[0.3436686, 0.3436688]	[0.19728514, 0.19728515]	[104.1912136, 104.1912139]
23	529	[0.3436459, 0.3436461]	[0.19728050, 0.19728051]	[101.0907940, 101.0907943]
24	516	[0.3436216, 0.3436223]	[0.19727558, 0.19727562]	[97.99080370, 97.99080400]
25	500	[0.3435963, 0.3435965]	[0.19727036, 0.19727037]	[94.89126900, 94.89126930]
26	481	[0.3435691, 0.3435693]	[0.19726479, 0.19726480]	[91.79221870, 91.79221900]
27	459	[0.3435401, 0.3435402]	[0.19725886, 0.19725887]	[88.69368450, 88.69368480]
28	434	[0.3435091, 0.3435093]	[0.19725252, 0.19725253]	[85.59570110, 85.59570150]

$25,29,55:56),\ldots,(1:25,54:56),\ldots,(1,28,31:56),(1,29,31:56),(1,30:56),(28,30:56),(29:56)$. The corresponding confidence intervals and narrow clusters follow the expected evolution given in Figure 2 and Table 1.

Finally, in the preceding example we have shown that the joint distribution functions of Sections 3 and 4 are easily computable and applied to real data. They provide an expected performance under multiple sub-samples in the rigorous dependent-time molecular docking setting. This is an important opportunity for several similar phenomena with multimatrix dependent samples.

6 Conclusions

This work introduces the class of *multimatrix variate distributions* by embedding previously isolated bimatrix-variate models within the general framework of elliptically contoured families. Most of the proposed distributions are invariant under the class of spherical models, an important property that eliminates the need for prior specification of the underlying joint distribution. The resulting models are computationally tractable, and several new properties and combinations involving three matrices are derived. These results address the problem of matrix-dependent samples by providing flexible constructions of joint distributions with desirable marginal distributions under independence. The versatility of the proposed framework allows scalars, vectors, and matrices to be incorporated simultaneously within a single joint distribution. Moreover, the computational feasibility of these models opens the way for further applications, including the evaluation of multiple probabilities on cones. Finally, the paper presents an application to molecular docking, specifically analyzing a known inhibitor within a newly identified cavity of the SARS-CoV-2 main protease.

Acknowledgements

The authors wish to thank the Editor and the anonymous reviewers for their constructive comments on the preliminary version of this paper.

References

Bekker, A., Roux, J. J. J., Ehlers, E. and Arashi, M. (2001). Bimatrix variate beta type IV distribution: relation to Wilks's statistics and bimatrix variate Kummer-beta type IV distribution. *Communication in Statistics-Theory and Methods*, **40**, 4165–4178.

Chen, J. J. and Novick, M. R. (1984). Bayesian analysis for binomial models with generalized beta prior distributions. *Journal of Educational Statistics*, **9**, 163–175.

Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2006). The distribution of the residual from general elliptical multivariate linear model. *Journal of Multivariate Analysis*, **97**, 1829–1841.

Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2010a). Bimatrix variate generalised beta distributions. *South African Statistical Journal*, **44**, 193–208.

Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2010b). Complex bimatrix variate generalised beta distributions. *Linear Algebra and its Applications*, **432**, 571–582.

Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2011). Noncentral bimatrix variate generalised beta distributions. *Metrika*, **73**, 317–333.

Díaz-García, J. A. and Caro-Lopera, F. J. (2022). Multimatricvariate distribution under elliptical models. *Journal of Statistical Planning and Inference*, **216**, 109-117.

Díaz-García, J. A., Caro-Lopera, F. J. and Pérez Ramírez, F. O. (2022). Multivector variate distributions: an application in Finance. *Sankhyā*, **84**, 534–555.

Dickey, J. M. (1967). Matricvariate generalizations of the multivariate t- distribution and the inverted multivariate t-distribution. *Annals of Mathematical Statistics*, **38**, 511–518.

Ehlers, R. (2011). *Bimatrix Variate Distributions of Wishart Ratios with Application*. Doctoral dissertation, Faculty of Natural & Agricultural Sciences University of Pretoria, Pretoria. http://hdl.handle.net/2263/31284.

Fang, K. T. and Zhang, Y. T. (1990). *Generalized Multivariate Analysis*. Science Press, Springer-Verlag, Beijing.

Fang, K. T., Zhang, Y. T. and Ng, K. W. (1990). *Symmetric Multivariate and realted distributions*. Springer-Science+Business Media, B. V., New Delhi.

Goodall, K. I. and Mardia, K. V. (1993). Multivariate aspects of shape theory. *Annals of Statistics*, **21**, 848–866.

Gupta, A. K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics*. Kluwer Academic Publishers, Dordrecht.

Jin, Z., Du, X., Xu, Y. et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, **582**, 289–293.

Libby, D. L. and Novick, M. R. (1982). Multivariate Generalized beta distributions with applications to utility assessment. *Journal of Educational Statistics*, **7**, 271–294.

Liu, X., Zhang, B., Jin, Z., Yang, H. and Rao, Z. (2020). The crystal structure of COVID-19 main protease in complex with an inhibitor N3. *Protein DataBank*, **10**, no. 10.2210.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York.

Nadarajah, S. (2007). A bivariate gamma model for drought. *Water Resources Research*, **43**, W08501.

Nadarajah, S. (2013). A bivariate distribution with gamma and beta marginals with application to drought data. *Journal of Applied Statistics*, **36**, 277–301.

Olkin, I. and Liu, R. (2003). A bivariate beta distribution. *Statistics and Probability Letters*, **62**, 407–412.

Olkin, I. and Rubin, H. (1964). Multivariate beta distributions and independence properties of Wishart distribution. *Annals of Mathematical Statistics*, **35**, 261–269. Correction 1966, 37(1), 297.

Ramírez-Velásquez, I., Bedoya-Calle, Á. H., Vélez, E., and Caro-Lopera, F. J. (2022). Shape theory applied to molecular docking and automatic localization of ligand binding pockets in large proteins. *ACS Omega*, **7**, 45991–46002.

Sarabia, J. M., Prieto, F. and Jordá, V. (2014). Bivariate beta-generated distributions with application to well-being data. *Journal of Statistical Distributions and Applications*, **1**, 15.

Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 29–44, Research Paper
https://smps.um.ac.ir, ISSN: 3115-9370

Exploring an extension of the Rayleigh distribution: statistical estimation and practical application

Harsh Tripathi[†,*] **and Varun Agiwal**[‡]

[†]*Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune, India*
[‡]*Indian Institute of Public Health, Hyderabaad, India*

Email(s): rsearchstat21@gmail.com, varunagiwal.stats@gmail.com

Abstract. In this paper, we introduce a new extension of the Rayleigh distribution, termed the Rayleigh-Exponential distribution (Ra-Ex-D). The proposed distribution is capable of modeling a wide range of data behaviors and demonstrates greater flexibility compared to classical distributions. Several statistical properties of the proposed distribution are derived, including survival characteristics, mean deviation, entropy, and conditional moments, along with the necessary proofs. Model parameters are estimated using both classical and Bayesian approaches under different prior assumptions. A simulation study is conducted to assess the performance of the estimation methods, demonstrating that the proposed procedures yield reliable and consistent estimates. Furthermore, the practical applicability of the model is illustrated through the analysis of a real-life dataset, where the proposed distribution exhibits superior fitting performance compared to existing models.

Keywords: Bayesian estimation; Classical estimation; Moments; Rayleigh distribution.

Mathematics Subject Classification (2020): 60E05, 62G30, 62P30.

1 Introduction

In recent decades, probability distributions have been widely used to solve various real-life problems. Many authors have developed new distributions based on transformations and families such as the alpha power transformation, Marshall–Olkin family, logarithmic family, exponentiated family,

*Corresponding author
Received: 17 May 2025 / Revised: 26 April 2026 / Accepted: 5 June 2026

https://doi.org/10.22067/smps.2026.93521.1047

The Author(s). This is an open access article under CC BY 4.0.

mixture of distributions, and the G-family, among others. These techniques typically add one or more parameters to a baseline model, thereby increasing the flexibility of the resulting distribution.

In recent years, several researchers have proposed probability distributions using transformation techniques and generalized families to further enhance modeling flexibility. Notable examples include the Marshall–Olkin additive Weibull distribution (Afify et al., 2018), the Marshall–Olkin Pareto distribution (Bdair and Haj Ahmad, 2021), and the Marshall–Olkin Gompertz distribution (Eghwerido et al., 2021), as well as broader developments in Marshall–Olkin-type models (Jose and Paul, 2018; Gonzalez-Hernández et al., 2021). Additionally, transformation-based approaches such as the alpha power transformation (Dey et al., 2019; Mahdavi and Kundu, 2017; Shukla et al., 2022) and transmuted families (Tripathi and Chitra, 2022; Tripathi and Mishra, 2022) have contributed significantly to generating flexible distributions.

Other important contributions include generalized and hybrid families like the TX family Klakattawi et al. (2022), exponentiated classes of distributions (Cordeiro and Lemonte, 2013; Yadav et al., 2021), quasi and extended Xgamma distributions (Sen et al., 2019; Tripathi et al., 2022), and the odd generalized N–H family (Ahmad et al., 2020). These developments are supported by general frameworks for generating continuous distributions (Lee et al., 2013), underscoring the growing interest in flexible modeling techniques.

Despite these advancements, challenges remain in effectively modeling complex lifetime data, particularly in capturing diverse hazard rate behaviors. Recent studies, such as Hassan et al. (2025) and Alsulami and Alghamdi (2025), have proposed extensions of the Weibull distribution to improve flexibility and fitting performance. However, there is still a need for more adaptable models that can simultaneously accommodate a wide range of hazard rate shapes while maintaining tractable statistical properties. This motivates the development of the proposed model. Hence, Elgarhy and Hassan (2019) introduced the Exponentiated Weibull–Weibull (EWW) distribution, which provides a flexible framework for modeling lifetime data. Several extensions of classical lifetime distributions have been developed in the literature to provide greater flexibility in modeling different shapes of probability density and hazard rate functions. Among these models, the Rayleigh–Exponential distribution (Ra-Ex-D) can be considered as a special case of the EWW distribution. We explored the literature and found that no one has made effort regarding the development of Ra-Ex-D.

The main objective of this article is to study the theoretical and inferential properties of the Ra-Ex-D. Specifically, we derive several important statistical properties of the proposed model and investigate different parameter estimation methods under both classical and Bayesian frameworks. In addition, a simulation study is conducted to examine the performance and consistency of the proposed estimators under different parameter combinations. Finally, the applicability of the proposed distribution is illustrated using real-life datasets.

The probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$ of Ra-Ex-D are given by

$$f(x) = 2\alpha\lambda e^{\lambda x}(e^{\lambda x}-1)e^{-\alpha(e^{\lambda x}-1)^2},\ x>0,\ \lambda>0,\ \alpha>0, \tag{1}$$

and

$$F(x) = 1 - e^{-\alpha(e^{\lambda x}-1)^2}, \tag{2}$$

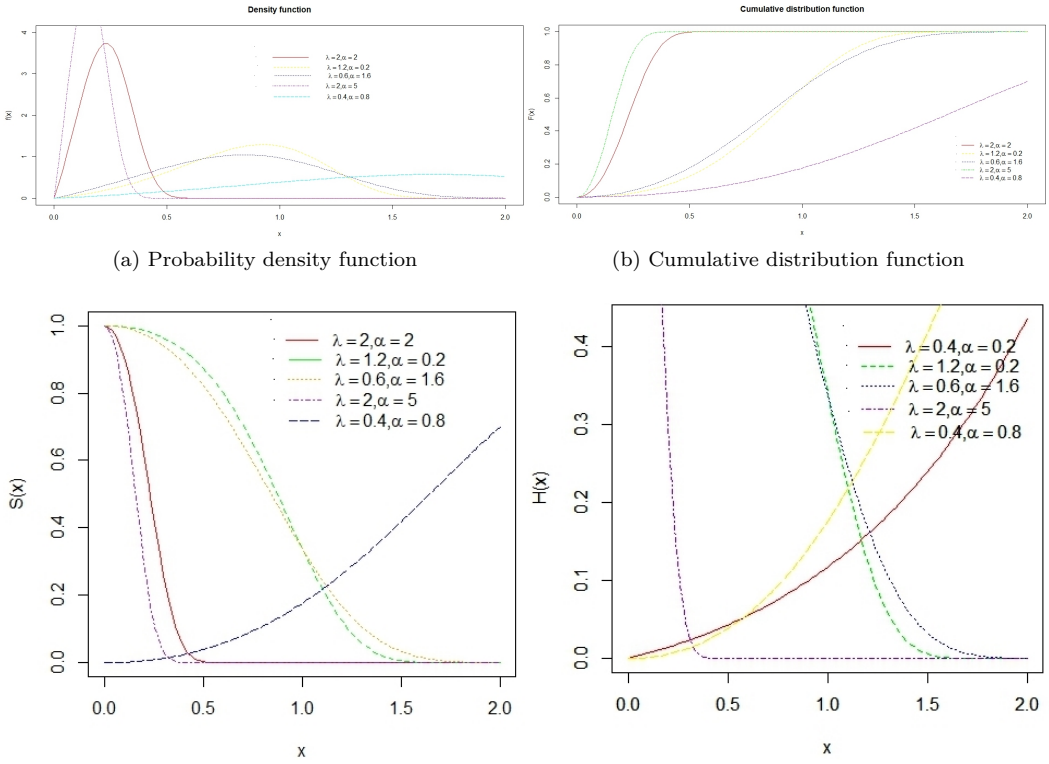
respectively. Survival function (SF) S(x) and hazard rate function (HRF) h(x) of proposed model are given below

$$S(x) = e^{-\alpha(e^{\lambda x}-1)^2}, \tag{3}$$

and

$$h(x) = 2\alpha\lambda e^{\lambda x}(e^{\lambda x} - 1), \tag{4}$$

respectively.
Figure 1(a) shows the pdf of the proposed distribution, given in (1), indicating that the proposed model is unimodal. The SF, given in (3), and the cdf, given in (2), are plotted in Figures 1(b) and 1(c), respectively. Figure 1(d) presents the HRF of the proposed distribution, as given in (4), which exhibits both increasing and decreasing hazard rates.



(a) Probability density function

(b) Cumulative distribution function

(c) Survival function

(d) Hazard rate function

Figure 1: Plots of the various functions of the Ra-Ex-distribution for different parameter values.

The remainder of the paper is organized as follows. Section 2 presents the statistical properties of the proposed distribution. Classical estimation methods are discussed in detail in Section 3. Bayesian estimation is outlined in Section 4. Section 5 presents a simulation study of the considered estimation methods. The application of the proposed probability distribution is discussed in Section 6. Concluding remarks on the Ra-Ex-D are provided in Section 7.

2 Statistical properties

In this section, we derive the statistical properties of the proposed Ra-Ex-D, viz., moments, conditional moments, mean deviation, and entropy.

2.1 Moments

Moments are an important property of a probability distribution, as they help measure skewness and kurtosis. Hence, moments are useful for describing the nature of a dataset. Let X be a random variable following the Ra-Ex-D, with pdf given in (1). Then, the r-th moment about the origin is given by

$$E(X^r) = 2\alpha\lambda \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{\lambda^i}{i!}\frac{(-1)^{j+k}}{j!}\alpha^j \binom{2j}{k} \frac{\Gamma(r+i+1)}{(\lambda k-2\lambda j)^{r+i+1}}(2^i-1), \quad r \geq 1.$$

2.2 Conditional moment

Assuming that the lifetime of the units under study follows our proposed model and exceeds a certain threshold x, one may be interested in the conditional moments. Moreover, these are useful for determining the mean deviation, as well as Bonferroni and Lorenz curves. Let X be a random variable following the Ra-Ex-D distribution as in (2); then the expression for the conditional moment is given by:

$$E(X^r|X>x) = \frac{2\alpha\lambda}{1-F(x)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{\lambda^i}{i!}\frac{(-1)^{j+k}}{j!}\alpha^j \binom{2j}{k} \frac{\Gamma(r+i+1,(\lambda k-2\lambda j)x)}{(\lambda k-2\lambda j)^{r+i+1}}(2^i-1).$$

2.3 Mean deviation

Mean deviation (MD) provides the scatterdness from mean. Mathematical expression of MD is given below

$$\begin{aligned}
\text{MD} &= \int_0^{\infty} |x-\mu| f(x)dx, \\
&= \int_0^{\mu} (\mu - x) f(x)dx + \int_{\mu}^{\infty} (x-\mu) f(x)dx \\
&= 2\mu F(\mu) - 2\mu + 2\int_{\mu}^{\infty} x f(x)dx,
\end{aligned}$$

where

$$\int_{\mu}^{\infty} x f(x)dx = 2\alpha\lambda \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{\lambda^i}{i!}\frac{(-1)^{j+k}}{j!}\alpha^j \binom{2j}{k} \frac{\Gamma(i+2,(\lambda k-2\lambda j)\mu)}{(\lambda k-2\lambda j)^{i+2}}(2^i-1)$$

and μ is the mean.

2.4 Entropy

Entropy measures the uncertainty or information content of a probability distribution and finds applications in various fields, including physics, statistics, reliability analysis, and kernel density

estimation. In the context of lifetime distributions, entropy provides insight into the variability and uncertainty associated with the underlying random variable.

The mathematical expression of the generalized entropy (GE) is given by

$$GE = \frac{\nu_\eta \mu^{-\eta} - 1}{\eta(\eta - 1)}; \quad \eta \neq 0, 1,$$

where $\nu_\eta = \int_0^\infty x^\eta f(x)\,dx$ represents the ηth moment of the distribution and μ denotes the mean. For the proposed Ra-Ex-D distribution, substituting the pdf given in (1) into the above expression, we have

$$\nu_\eta = 2\alpha\lambda \int_0^\infty x^\eta e^{\lambda x} u e^{-\alpha u^2}\,dx,$$

where $u = (e^{\lambda x} - 1)$. Substituting this expression for ν_η into the generalized entropy formula gives the entropy equation for the proposed distribution.

3 Classical estimation

In this section, we discuss the classical estimation of the parameters. To this end, we employ the following classical methods of estimation: maximum likelihood estimation (MLE), least squares estimation (LSE), weighted least squares estimation (WLSE), Cramér–von Mises estimation (CME), and maximum product spacing estimation (MPSE).

3.1 MLE

Suppose that $X_1, X_2, \ldots, X_n$, are independently and identically distributed (i.i.d.) random variables following the proposed distribution with the pdf given in (1). Then, the likelihood function is given by

$$L(\alpha, \lambda \mid x) = 2^n \alpha^n \lambda^n e^{\lambda \sum_{i=1}^n x_i} \prod_{i=1}^n \left(e^{\lambda x_i} - 1\right) e^{-\alpha \sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2},$$

and the corresponding log-likelihood function can be written as

$$\log L(\alpha, \lambda \mid x) = n(\log 2 + \log \alpha + \log \lambda) + \lambda \sum_{i=1}^n x_i + \sum_{i=1}^n \log\left(e^{\lambda x_i} - 1\right) - \alpha \sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2. \tag{5}$$

To obtain the maximum likelihood estimates of the parameters, we maximize the log-likelihood function with respect to α and λ. Differentiating equation (5) with respect to α and λ yields the log-likelihood equations:

$$\frac{\partial \log L(\alpha, \lambda \mid x)}{\partial \alpha} = \frac{1}{\alpha} - \sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2, \tag{6}$$

$$\frac{\partial \log L(\alpha, \lambda \mid x)}{\partial \lambda} = \frac{1}{\lambda} + \sum_{i=1}^n x_i + \sum_{i=1}^n \frac{x_i e^{\lambda x_i}}{\left(e^{\lambda x_i} - 1\right)} - 2\alpha \sum_{i=1}^n x_i e^{\lambda x_i} \left(e^{\lambda x_i} - 1\right). \tag{7}$$

From equation (6), the MLE of α for a given λ is obtained as

$$\hat{\alpha} = \frac{1}{\sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2}. \tag{8}$$

Substituting (8) into (7), the MLE of λ can be obtained by solving the following non-linear equation:

$$\frac{\partial \log L(\alpha,\lambda \mid x)}{\partial \lambda} = \frac{1}{\lambda} + \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i e^{\lambda x_i}}{\left(e^{\lambda x_i}-1\right)} - 2\frac{\sum_{i=1}^{n} x_i e^{\lambda x_i}\left(e^{\lambda x_i}-1\right)}{\sum_{i=1}^{n}\left(e^{\lambda x_i}-1\right)^2} = 0. \tag{9}$$

Since equation (9) is not in closed form, numerical techniques such as the Newton–Raphson method are employed to solve this equation and obtain the MLEs of α and λ.

3.2 MPSE

This method of estimation was proposed by Cheng and Amin (1983). Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x,\alpha,\lambda)$. Furthermore, let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the corresponding order statistics. Then the spacings D_i are defined as

$$D_1 = F(x_{(1)},\alpha,\lambda), \quad D_{n+1} = 1 - F(x_{(n)},\alpha,\lambda),$$

and

$$D_i = F(x_{(i)},\alpha,\lambda) - F(x_{(i-1)},\alpha,\lambda), \quad i = 2,3,\ldots,n,$$

where $\sum_{i=1}^{n+1} D_i = 1$. This method consists of finding the values of α and λ that maximize the geometric mean G of the spacings, i.e.,

$$G = \left\{\prod_{i=1}^{n+1} D_i\right\}^{\frac{1}{n+1}},$$

or equivalently,

$$\log(G) = \frac{1}{n+1}\sum_{i=1}^{n+1} \log D_i,$$

$$\log(G) = \frac{1}{n+1}\sum_{i=1}^{n+1} \log\left[F(x_{(i)},\alpha,\lambda) - F(x_{(i-1)},\alpha,\lambda)\right].$$

The basic concept behind this estimation method is that the differences between the cdf at neighboring points should be identically distributed. The MPSEs can be obtained by solving the following non-linear equations:

$$\frac{\partial \log(G)}{\partial \alpha} = \sum_{i=1}^{n+1} \frac{F'_\alpha(x_{(i)},\alpha,\lambda) - F'_\alpha(x_{(i-1)},\alpha,\lambda)}{F(x_{(i)},\alpha,\lambda) - F(x_{(i-1)},\alpha,\lambda)},$$

$$\frac{\partial \log(G)}{\partial \lambda} = \sum_{i=1}^{n+1} \frac{F'_\lambda(x_{(i)},\alpha,\lambda) - F'_\lambda(x_{(i-1)},\alpha,\lambda)}{F(x_{(i)},\alpha,\lambda) - F(x_{(i-1)},\alpha,\lambda)},$$

where

$$F'_\alpha(x,\alpha,\lambda) = \left(e^{\lambda x}-1\right)^2 e^{-\alpha\left(e^{\lambda x}-1\right)^2}$$

and

$$F'_\lambda(x,\alpha,\lambda) = 2\alpha\lambda e^{\lambda x}\left(e^{\lambda x}-1\right) e^{-\alpha\left(e^{\lambda x}-1\right)^2}.$$

3.3 LSE and WLSE

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the ordered sample of size n drawn from the pdf given in (1). Then the expectation of the empirical cdf is given by

$$E\left[F\left(X_{(i)}\right)\right] = \frac{i}{n+1}, \quad i = 1,2,3,\ldots,n.$$

The least squares estimates (LSEs) of α and λ are obtained by minimizing

$$LS(\alpha,\lambda) = \sum_{i=1}^{n}\left(F(x_{(i)},\alpha,\lambda) - \frac{i}{n+1}\right)^2 = \sum_{i=1}^{n}\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2 - \frac{i}{n+1}\right)^2,$$

where we have used the fact that $F(x,\alpha,\lambda) = 1-\alpha(e^{\lambda x}-1)^2$ for the proposed distribution. The LSEs $\hat{\alpha}$ and $\hat{\lambda}$ can be found by solving the following equations:

$$\frac{\partial LS(\alpha,\lambda)}{\partial \alpha} = 2\sum_{i=1}^{n} F'_{\alpha}(x,\alpha,\lambda)\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2 - \frac{i}{n+1}\right) = 0,$$

$$\frac{\partial LS(\alpha,\lambda)}{\partial \lambda} = 2\sum_{i=1}^{n} F'_{\lambda}(x,\alpha,\lambda)\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2 - \frac{i}{n+1}\right) = 0,$$

where $F'_{\alpha}(x,\alpha,\lambda)$ and $F'_{\lambda}(x,\alpha,\lambda)$ are as defined in the previous subsection. These equations can be solved numerically using methods such as the Newton–Raphson method.

The WLSEs of α and λ are obtained by minimizing, with respect to α and λ, the following function:

$$WLS(\alpha,\lambda) = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left(F(x_{(i)},\alpha,\lambda) - \frac{i}{n+1}\right)^2,$$

which can be expressed as

$$WLS(\alpha,\lambda) = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2 - \frac{i}{n+1}\right)^2.$$

Thus, the WLSEs are determined by solving the equations $\frac{\partial WLS(\alpha,\lambda)}{\partial \alpha} = 0$ and $\frac{\partial WLS(\alpha,\lambda)}{\partial \lambda} = 0$ simultaneously. These equations are analogous to those presented for the LSE, with the weight sequence incorporated appropriately.

3.4 CME

Another well-known estimation method is the Cramér–von Mises estimation (CME), introduced by Darling (2019). In the context of the proposed model, the CMEs of α and λ are obtained by minimizing the following function with respect to α and λ:

$$C(\alpha,\lambda) = \frac{1}{12n} + \sum_{i=1}^{n}\left(F(x_{(i)},\alpha,\lambda) - \frac{2i-1}{2n}\right)^2,$$

which can be expressed as

$$C(\alpha,\lambda) = \frac{1}{12n} + \sum_{i=1}^{n}\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2 - \frac{2i-1}{2n}\right)^2.$$

Thus, the CMEs are obtained by solving the following equations simultaneously: $\frac{\partial C(\alpha,\lambda)}{\partial \alpha} = 0$ and $\frac{\partial C(\alpha,\lambda)}{\partial \lambda} = 0$, where

$$\frac{\partial C(\alpha,\lambda)}{\partial \alpha} = 2\sum_{i=1}^{n} F'_{\alpha}(x,\alpha,\lambda)\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2-\frac{2i-1}{2n}\right),$$

$$\frac{\partial C(\alpha,\lambda)}{\partial \lambda} = 2\sum_{i=1}^{n} F'_{\lambda}(x,\alpha,\lambda)\left(1-\alpha\left(e^{\lambda x_{(i)}}-1\right)^2-\frac{2i-1}{2n}\right),$$

and $F'_{\alpha}(x,\alpha,\lambda)$ and $F'_{\lambda}(x,\alpha,\lambda)$ are as given above.

This section presented classical parameter estimation methods for the proposed distribution. The MLE involves solving nonlinear equations, which require iterative numerical optimization such as the Newton–Raphson method, leading to moderate computational complexity. The LSE and WLSE methods are computationally simpler but may be less efficient. Although more computationally intensive, Bayesian estimation was also implemented to provide more robust estimates, particularly for small sample sizes or complex likelihood surfaces.

4 Bayesian estimation

In the Bayesian framework, a probability model is fitted not only using the likelihood function of the observations but also incorporating prior beliefs about the unknown parameters in the form of prior distributions. This yields model results based on the combined information from the likelihood and the prior, summarized in the posterior distribution. For the proposed distribution, the likelihood function is

$$L(\alpha,\lambda \mid x) = 2^n \alpha^n \lambda^n e^{\lambda \sum_{i=1}^{n} x_i} \prod_{i=1}^{n}\left(e^{\lambda x_i}-1\right) e^{-\alpha \sum_{i=1}^{n}\left(e^{\lambda x_i}-1\right)^2}. \tag{10}$$

From a strict Bayesian perspective, it is not possible to definitively state that one prior is more relevant or better than another. Therefore, priors are broadly classified into two categories: informative and non-informative, each of which may be further subdivided. An informative prior is used when sufficient information about the parameters is available, meaning complete prior distributional information, such as conjugate priors or gamma priors, is at hand. Otherwise, a non-informative prior, which contains little or no prior information about the parameter(s), is considered (e.g., uniform prior or improper prior). Here, we employ the gamma prior distribution for our analysis because it offers greater flexibility, respects the positivity constraint, and exhibits conjugacy-like behavior that facilitates posterior computation. The forms of the priors are

$$\pi(\alpha) = \frac{q^p}{\Gamma(p)}\alpha^{p-1}\exp(-q\alpha); \quad p,q>0, \tag{11}$$

and

$$\pi(\lambda) = \frac{s^r}{\Gamma(r)}\lambda^{r-1}\exp(-s\lambda); \quad r,s>0, \tag{12}$$

where (p,q,r,s) are the hyperparameters. Note that if the hyperparameters are set to zero ($p=q=r=s=0$), the gamma prior reduces to a non-informative prior of the form

$$\pi(\alpha) = \frac{1}{\alpha}, \tag{13}$$

and

$$\pi(\lambda) = \frac{1}{\lambda}. \tag{14}$$

The joint posterior density function of the parameters given the observed random variables X is obtained as

$$\pi(\alpha, \lambda \mid x) \propto L(\alpha, \lambda \mid x)\, \pi(\alpha)\, \pi(\lambda).$$

Under the informative prior, the joint posterior density function is formed using equations (10)–(12)

$$\pi(\alpha, \lambda \mid x) \propto 2^n \alpha^{n+p-1} \lambda^{n+r-1} e^{\lambda\left(\sum_{i=1}^n x_i - s\right)} \prod_{i=1}^n \left(e^{\lambda x_i} - 1\right) e^{-\alpha\left(\sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2 + q\right)}.$$

Under the non-informative prior, the joint posterior density function is formed using equations (10), (13), and (14)

$$\pi(\alpha, \lambda \mid x) \propto 2^n \alpha^{n-1} \lambda^{n-1} e^{\lambda \sum_{i=1}^n x_i} \prod_{i=1}^n \left(e^{\lambda x_i} - 1\right) e^{-\alpha \sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2}, \tag{15}$$

In the Bayesian framework, a loss function is used to define the statistical risk (error) associated with parameter estimation. It is a function of the true and estimated values, and the goal is to select the estimator that minimizes the risk for each parameter. Here, we consider the squared error loss function (SELF) as a symmetric loss function to facilitate better understanding. The Bayes estimate under this loss function is the posterior mean, given by

$$L_{\text{SELF}}(\Theta, \hat{\Theta}) = \left(\hat{\Theta} - \Theta\right)^2.$$

Bayes estimators of any parametric function, say $\phi(\Theta)$, under SELF are defined as

$$\phi^{\star}_{\text{self}}(\Theta \mid \text{data}) = E_{\pi}\left[\phi(\Theta) \mid \text{data}\right] \propto \int_{\Theta} \phi(\Theta)\, \pi(\Theta \mid \text{data})\, d\Theta.$$

Due to the high-dimensional integration involved in the joint posterior distributions, the expression for the Bayes estimator cannot be solved analytically. Therefore, we adopt the most popular Markov Chain Monte Carlo (MCMC) methodology. In this methodology, a sequence of random samples is generated from the posterior distribution of interest. Here, we use hybrid Gibbs sampling methods to obtain the Bayes estimates based on the full conditional posterior distributions. Using thee joint posterior density function given in (15), the expressions for the conditional posterior pdfs of α and λ can be written as

$$\pi(\alpha \mid x, \lambda) \propto \alpha^{n+p-1} e^{-\alpha\left(\sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2 + q\right)},$$

and

$$\pi(\lambda \mid x, \alpha) \propto \lambda^{n+r-1} e^{\lambda\left(\sum_{i=1}^n x_i - s\right)} \prod_{i=1}^n \left(e^{\lambda x_i} - 1\right) e^{-\alpha \sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2}.$$

We can directly generate samples from the posterior distribution of α because its conditional posterior follows a gamma distribution with shape parameter $n+p$ and rate (or scale) parameter $\sum_{i=1}^n \left(e^{\lambda x_i} - 1\right)^2 + q$. However, we cannot directly generate samples from the posterior distribution of λ because it does not correspond to any well-known distributional form. Hence, the Metropolis–Hastings (MH) algorithm is used to produce the posterior samples. The steps of the hybrid Gibbs sampling approach are as follows:

Step 1. Start with an initial value $x^\star$ such that $h(x^\star) > 0$.

Step 2. Set $t = 1$.

Step 3. Propose a move y with conditional probability density given $x^\star$, denoted by $g(x^\star, y)$ (the candidate proposal density).

Step 4. Calculate the Hastings ratio:

$$\rho(x^\star, y) = \frac{h(y)\, g(y, x^\star)}{h(x^\star)\, g(x^\star, y)},$$

and if $g(x^\star, y) = g(y, x^\star)$ (symmetric proposal), the Hastings ratio reduces to

$$\rho(x^\star, y) = \frac{h(y)}{h(x^\star)}.$$

Step 5. Accept the proposed move y with probability $\gamma = \min[1, \rho(x^\star, y)]$ and reject with probability $1 - \gamma$.

Step 6. Repeat Steps 2–4 for $t = 10000$ iterations and record the sequence of parameter observations. Finally, obtain the Bayes estimates under SELF.

5 Simulation study

In this section, we present a numerical study to compare the behavior of the different estimation methods described above. Since closed-form solutions are not available, numerical optimization techniques are employed for classical estimators, while Bayesian estimates are obtained through posterior simulation. For this purpose, different combinations of parameter values and various sample sizes are considered to draw meaningful inferences. Random samples from the proposed distribution are generated using the inverse CDF method, which involves the following steps:

1. Set n, α, and λ.
2. Generate $U_l \sim U(0,1)$ for $l = 1, 2, \ldots, n$.
3. Compute $x = (x_1, x_2, \ldots, x_n)$ using U such that

$$x_l = F^{-1}(U_l) = \frac{1}{\lambda} \log \left[1 + \left(-\frac{\log(1 - U_l)}{\alpha} \right)^{\frac{1}{2}} \right].$$

4. Compute the estimates using the generated samples.

We consider sample sizes $n = (20, 50, 100)$ and four sets of parameter values:

$$(\alpha, \lambda) = \{(0.2, 0.5), (0.5, 1.5), (1, 1), (1.5, 1.2)\}$$

to evaluate the performance of the estimators. These values are selected to represent increasing, decreasing, and bathtub-shaped hazard functions. Bayesian estimation is implemented using hybrid Gibbs sampling and the Metropolis–Hastings algorithm with 10000 iterations, including a burn-in

period of 2000 iterations. All computations are performed using R-4.5.3. The hyperparameters of the gamma prior are chosen such that the prior means equal the true parameter values and the prior variances are equal to 1; these estimates (Bayes informative prior) are denoted as "Bayes IP". The non-informative prior estimates are denoted as "Bayes NIP". We obtain the parameter estimates under classical and Bayesian methods, reporting the average estimates (AV) and mean squared errors (MSE). For the different parameter combinations with varying n, Tables 1–4 record the AV and MSE.

Table 1: AV and MSE of the proposed model for classical and Bayesian estimates with $\alpha = 0.2$ and $\lambda = 0.5$.

P	n		MLE	MPSE	LSE	WLSE	CME	Bayes IP	Bayes NIP
$\alpha = 0.2$	20	AV	0.3669	0.3705	0.3830	0.3300	0.3515	0.2832	0.1809
		MSE	0.4402	0.4650	0.4717	0.3867	0.4136	0.0211	0.0477
	50	AV	0.2589	0.3429	0.3401	0.2744	0.3052	0.2501	0.1887
		MSE	0.2514	0.3311	0.3111	0.1016	0.1779	0.0053	0.0074
	100	AV	0.2271	0.3062	0.2768	0.2365	0.2454	0.1882	0.2126
		MSE	0.0640	0.0780	0.1101	0.0595	0.0680	0.0002	0.0008
$\lambda = 0.5$	20	AV	0.5874	0.4308	0.4671	0.4804	0.4908	0.5067	0.5519
		MSE	0.1273	0.0551	0.0529	0.0528	0.0557	0.0023	0.0059
	50	AV	0.5335	0.4583	0.4779	0.4881	0.4916	0.4692	0.5175
		MSE	0.0696	0.0451	0.0476	0.0467	0.0526	0.0016	0.0039
	100	AV	0.5148	0.4731	0.4863	0.4947	0.4975	0.4903	0.4915
		MSE	0.0558	0.0435	0.0456	0.0465	0.0489	0.0001	0.0002

Table 2: AV and MSE of the proposed model for classical and Bayesian estimates with $\alpha = 0.5$ and $\lambda = 1.5$.

P	n		MLE	MPSE	LSE	WLSE	CME	Bayes IP	Bayes NIP
$\alpha = 0.5$	20	AV	0.6023	0.6707	0.6858	0.5669	0.4580	0.5785	0.5820
		MSE	0.6237	0.6679	0.6780	0.6055	0.6307	0.0248	0.0396
	50	AV	0.5430	0.5929	0.5877	0.5168	0.4659	0.4486	0.4719
		MSE	0.5690	0.5970	0.5837	0.5688	0.5908	0.0095	0.0172
	100	AV	0.5290	0.4823	0.5463	0.5045	0.4905	0.5381	0.4951
		MSE	0.4758	0.5384	0.5682	0.4370	0.5075	0.0033	0.0050
$\lambda = 1.5$	20	AV	1.6401	1.4433	1.4351	1.4552	1.5199	1.4860	1.3840
		MSE	1.0252	0.7076	0.7598	0.7314	0.8281	0.0252	0.0280
	50	AV	1.6034	1.4673	1.4330	1.4596	1.5141	1.5497	1.4046
		MSE	0.8416	0.5783	0.6615	0.6030	0.7072	0.0157	0.0234
	100	AV	1.5576	1.4925	1.4544	1.4948	1.4955	1.4990	1.4618
		MSE	0.6565	0.5045	0.5766	0.5742	0.6017	0.0034	0.0092

From the tables, we observe that the MSEs for α based on WLSE are smaller than those of the other classical estimators for all sample sizes. MPSE performs better for λ when $\alpha < 1$, while for $\alpha \geq 1$, WLSE performs better compared to other classical estimators in terms of MSE. In comparison with Bayesian methods, informative prior estimators yield smaller MSEs than other

Table 3: AV and MSE of the proposed model for classical and Bayesian estimates with $\alpha = 1$ and $\lambda = 1$.

P	n		MLE	MPSE	LSE	WLSE	CME	Bayes IP	Bayes NIP
$\alpha = 1$	20	AV	1.4134	1.3021	0.9073	1.1235	0.9521	0.9242	1.1296
		MSE	0.6144	0.5981	0.6596	0.5626	0.6286	0.1357	0.1827
	50	AV	1.1459	1.1278	1.1960	1.1129	1.1716	1.0442	1.0863
		MSE	0.4195	0.4005	0.5543	0.3531	0.4251	0.0648	0.0139
	100	AV	1.1099	0.9210	1.0563	0.9722	0.9104	1.0012	1.0158
		MSE	0.2589	0.2318	0.4303	0.2086	0.3459	0.0131	0.0519
$\lambda = 1$	20	AV	1.2231	0.9109	1.0425	1.0371	1.0997	0.9318	1.1682
		MSE	0.4349	0.2228	0.1621	0.1362	0.2174	0.0210	0.0377
	50	AV	1.0743	0.8708	1.0177	1.0257	1.0650	1.0371	1.0601
		MSE	0.1533	0.1403	0.1292	0.0944	0.1366	0.0056	0.0064
	100	AV	1.0377	0.9121	1.0033	1.0162	1.0409	0.9591	0.9922
		MSE	0.0760	0.0792	0.0732	0.0575	0.0800	0.0028	0.0046

Table 4: AV and MSE of the proposed model for classical and Bayesian estimates with $\alpha = 1.5$ and $\lambda = 1.2$.

P	n		MLE	MPSE	LSE	WLSE	CME	Bayes IP	Bayes NIP
$\alpha = 1.5$	20	AV	1.6979	1.1828	1.1693	1.2095	1.0294	1.3572	1.3708
		MSE	0.6917	0.5852	0.6105	0.5712	0.6401	0.2525	0.3985
	50	AV	1.6211	1.3461	1.3180	1.3397	1.2232	1.4129	1.3758
		MSE	0.5796	0.5253	0.5520	0.5267	0.6079	0.1418	0.1728
	100	AV	1.5478	1.4653	1.4521	1.4617	1.4466	1.4822	1.4412
		MSE	0.4866	0.4398	4.8382	0.4577	0.4768	0.0202	0.0367
$\lambda = 1.2$	20	AV	1.5947	1.1616	1.3208	1.3114	1.4281	1.1335	1.0945
		MSE	0.4278	0.4465	0.3450	0.2964	0.4155	0.0187	0.0297
	50	AV	1.3246	1.0645	1.2457	1.2431	1.3250	1.1673	1.3005
		MSE	0.2834	0.3005	0.2324	0.2039	0.2548	0.0165	0.0272
	100	AV	1.2613	1.0861	1.2071	1.2336	1.2651	1.2241	1.2181
		MSE	0.1544	0.2076	0.1884	0.1437	0.1783	0.0121	0.0201

estimators for both parameters. We also see that the MSEs of classical estimators are larger than those of Bayesian estimators for all sample sizes and parameter combinations. Overall, the results indicate that Bayesian estimators consistently exhibit lower MSE compared to classical methods, particularly for small sample sizes. Among classical methods, MLE performs better than least squares-based estimators. It can also be observed that as the sample size increases, the difference in MSE between the two approaches decreases.

6 Real life example

In this section, we demonstrate the application of the proposed model to a real-world dataset. For this purpose, we consider a dataset representing the tensile strength (measured in GPa) of 69 carbon fibers tested under tension at a gauge length of 20 mm (Bader and Priest, 1982):

1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.997, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.140, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.570, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.880, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585, 3.858

A descriptive summary of the considered dataset is presented in Table 5. First, we assess whether the dataset fits the Ra-Ex-D distribution. For this purpose, we use Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), the Kolmogorov–Smirnov (K-S) statistic, and the corresponding p-value. A model with the lowest K-S value and the largest p-value indicates the best fit. For the proposed model, the values of the log-likelihood (L-L), AIC, BIC, K-S statistic, and p-value are reported in Table 6. From Table 6, we can readily verify that the Ra-Ex-D distribution fits the dataset better than the Lindley distribution (LD), Akash distribution (AKD), exponential distribution (ED), and generalized exponential distribution (GED). Figure 2 displays the histogram with fitted density, the empirical CDF plot, and the probability–probability (P-P) plot for the considered dataset. From Figure 2, we conclude that the Ra-Ex-D distribution provides a satisfactory fit. In certain scenarios, the Ra-Ex-D distribution may be preferred over other distributions for problem-solving. The estimated parameter values along with their standard deviations (SD) under different estimators are presented in Table 7.

Table 5: Descriptive summary for the considered dataset.

Minimum	Maximum	Q_1	Q_2	Q_3	Skewness	Kurtosis	Mean
1.312	3.858	2.098	2.478	2.773	0.102	3.2253	2.455

where Q_1, Q_2 and Q_3 are the first quartile, median and third quartile, respectively.

Table 6: Model fitting summary for the considered dataset.

Model	MLE	L-L	AIC	BIC	K-S	p-value
Ra-Ex-D	$\alpha = 0.0136$, $\lambda = 0.8407$	-55.4231	114.8463	119.3145	0.0905	0.6247
LD	$\lambda = 0.6535$	-119.3110	240.6220	242.8561	0.4005	<0.001
AKD	$\lambda = 0.9637$	-112.2662	226.5324	228.7665	0.3615	<0.001
ED	$\lambda = 0.4073$	-130.9789	263.9578	266.1919	0.4477	<0.001

Table 7: Classical and Bayes estimates for the parameters of the proposed model based on the real dataset.

Estimators		MLE	MPSE	LSE	WLSE	CME	Bayes NIP
α	Estimate	0.0136	0.0146	0.0158	0.0165	0.0151	0.0211
	SD	0.0118	0.0114	0.0109	0.0111	0.0106	0.0290
λ	Estimate	0.8407	0.9832	1.0801	1.0909	1.1126	0.7212
	SD	0.1306	0.1362	0.1598	0.1514	0.1451	0.1558

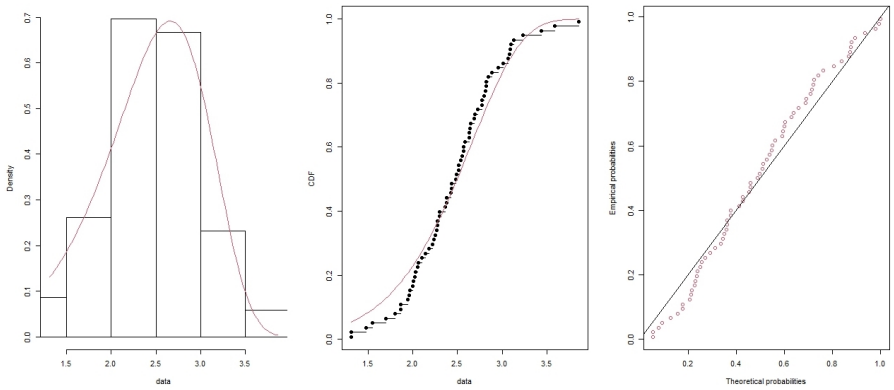


Figure 2: Histogram with fitted density, empirical CDF, and P-P plot.

7 Conclusions

In this paper, we developed a new probability distribution as an extended version of the Rayleigh distribution, termed the Ra-Ex-D. The main theoretical contribution of this work lies in deriving explicit mathematical expressions for its probability density function, cumulative distribution function, survival function, hazard rate function, and several important statistical characteristics, including moments and related measures. The shapes of the probability density function and hazard rate function of the proposed distribution are discussed through graphical illustrations.

The inferential framework of the model was developed under both classical and Bayesian paradigms, allowing for a comprehensive comparison of estimation procedures. The simulation results demonstrated that the estimators perform satisfactorily, with the mean squared error decreasing as the sample size increases, thereby confirming the consistency and reliability of the estimation methods. Overall, the Bayesian estimates are found to be better compared to the classical estimates. Finally, a real-life data example is provided to illustrate the applicability of the proposed model.

Acknowledgements

Availability of data and material: The link of the data set used in this study is included within the paper and also cited in the paper accordingly.

References

Afify, A. Z., Cordeiro, G. M., Yousof, H. M., Saboor, A., and Ortega, E. M. (2018). The Marshall-Olkin additive Weibull distribution with variable shapes for the hazard rate. *Hacettepe Journal of Mathematics and Statistics*, **47**, 365–381.

Ahmad, Z., Elgarhy, M., Hamedani, G. G., and Butt, N. S. (2020). Odd generalized NH generated family of distributions with application to exponential model. *Pakistan Journal of Statistics and Operation Research*, 53–71.

Alsulami, D., and Alghamdi, A. S. (2025). A new extended Weibull distribution: estimation methods and applications in engineering, physics, and medicine. *Mathematics*, **13**, 3262.

Bdair, O., and Haj Ahmad, H. (2021). Estimation of the Marshall-Olkin Pareto distribution parameters: comparative study. *Revista Investigacion Operacional*, **42**.

Bader, M. G., and Priest, A. M. (1982). Statistical aspects of fibre and bundle strength in hybrid composites. *Progress in Science and Engineering of Composites*, 1129–1136.

Cheng, R. C. H., and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B (Methodological)*, **45**, 394–403.

Cordeiro, G. M., and Lemonte, A. J. (2013). On the Marshall–Olkin extended weibull distribution. *Statistical Papers*, **54**, 333–353.

Dey, S., Ghosh, I., and Kumar, D. (2019). Alpha-power transformed Lindley distribution: properties and associated inference with application to earthquake data. *Annals of Data Science*, **6**, 623–650.

Darling, D. A. (1957). The kolmogorov-Smirnov, Cramer-von Mises tests. *The Annals of Mathematical Statistics*, **28**, 823–838.

Eghwerido, J. T., Ogbo, J. O., and Omotoye, A. E. (2021). The Marshall-Olkin Gompertz distribution: properties and applications. *Statistica*, **81**, 183–215.

Elgarhy, M., and Hassan, A. (2019). Exponentiated-Weibull distribution: statistical properties and applications. *Gazi University Journal of Science*, **32**, 616–635.

Jose, K., and Paul, A. L. B. I. N. (2018). Marshall-Olkin exponential power distribution and its generalization: theory and applications. *IAPQR Transactions*, **43**.

González-Hernández, I. J., Granillo-Macías, R., Rondero-Guerrero, C., and Simón-Marmolejo, I. (2021). Marshall-Olkin distributions: a bibliometric study. *Scientometrics*, **126**, 9005–9029.

Hassan, A. S., Morgan, Y. S., and Abd-Allah, M. (2025). Parameter estimation and data analysis using a novel extension of Weibull distribution. *The Egyptian Statistical Journal*, **69**, 131–154.

Klakattawi, H., Alsulami, D., Elaal, M. A., Dey, S., and Baharith, L. (2022). A new generalized family of distributions based on combining Marshal-Olkin transformation with TX family. *PloS one*, **17**, e0263673.

Lee, C., Famoye, F., and Alzaatreh, A. Y. (2013). Methods for generating families of univariate continuous distributions in the recent decades. *Wiley Interdisciplinary Reviews: Computational Statistics*, **5**, 219–238.

Mahdavi, A., and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics-Theory and Methods*, **46**, 6543–6557.

Shukla, S., Yadav, A. S., Rao, G. S., Saha, M., and Tripathi, H. (2022). Alpha power transformed Xgamma distribution and applications to reliability, survival and environmental data. *Journal of Scientific Research*, **66**, 330–346.

Sen, S., Afify, A. Z., Al-Mofleh, H., and Ahsanullah, M. (2019). The quasi xgamma-geometric distribution with application in medicine. *Filomat*, **33**, 5291–5330.

Tripathi, H., Yadav, A. S., Saha, M., and Shukla, S. (2022). A new flexible extension of Xgamma distribution and its application to COVID-19 Data. *Nepal Journal of Mathematical Sciences*, **3**, 11–30.

Tripathi, H., and Chitra. (2022). Transmuted extended xgamma-2 distribution and its statistical properties. *International Journal of Mechanical Engineering*, 7.

Tripathi, H., and Mishra, S. (2022). The transmuted inverse xgamma distribution and its statistical properties. *International Journal of Mechanical Engineering*, 7.

Yadav, A. S., Saha, M., Tripathi, H., and Kumar, S. (2021). The exponentiated XGamma distribution: a new monotone failure rate model and its applications to lifetime data. *Statistica*, **81**, 303–334.

Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 45–52, Research Paper
https://smps.um.ac.ir, ISSN: 3115-9370

Quasi-identifiability of multipath change-point models with general link functions

Masoud Majidizadeh*

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

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

Abstract. The nonsingularity of the Fisher information matrix is a fundamental requirement for standard asymptotic inference. The framework developed by Asgharian (2014) shows that, for identifiable and smooth models, the set of singular parameter values is negligible. This result was previously established for multipath change-point (MCP) models in which the change hazard is specified through a proportional odds ratio structure. In this work, we substantially broaden that setting by proving the quasi-identifiability of MCP models under a general class of strictly monotone link functions, encompassing the probit, logit, and complementary log-log models as special cases. Our main result demonstrates that, for this broad family of models, the injectivity of the link function is sufficient to guarantee quasi-identifiability under standard regularity conditions. Consequently, the general singularity theory applies directly, implying that the Fisher information matrix is nonsingular for almost all parameter values. This, in turn, justifies standard likelihood-based inference for a considerably wider and more flexible class of models.

Keywords: Fisher information matrix; Link function; Multipath change-point models; Quasi-identifiability; Singularities.

Mathematics Subject Classification (2020): 62F12, 62J12.

1 Introduction

The multipath change-point (MCP) model, introduced in various forms by authors such as Joseph and Wolfson (1993), provides a flexible framework for analyzing longitudinal data where subjects may experience a change in regime at different, unobserved times. Conceptually, it belongs to the broad family of finite mixture models, which are a standard tool for modeling latent heterogeneity in longitudinal and survival data (e.g., McLachlan et al. (2019); Titterington et al. (1985)). Statistical inference in such models, particularly for the effects of covariates on the change-point process, often relies on the asymptotic properties of the maximum likelihood estimator (MLE). A fundamental prerequisite for establishing these

*Corresponding author
Received: 24 October 2025 / Revised: 26 April 2026 / Accepted: 24 June 2026

https://doi.org/10.22067/smps.2026.96138.1055

The Author(s). This is an open access article under CC BY 4.0.

properties, such as asymptotic normality and efficiency, is the nonsingularity of the Fisher information matrix (FIM) at the true parameter value.

The MCP model has evolved considerably since its inception. The foundational framework and the properties of the MLE were first established in the seminal works of Joseph and Wolfson (1992) and Joseph and Wolfson (1993). This initial framework was subsequently extended to accommodate more complex data structures, such as correlated observations between paths (Joseph et al., 1996). A pivotal extension, which forms the basis for the current work, was the formal incorporation of covariates into the model by Asgharian and Wolfson (2001), who investigated the modeling challenges and established the consistency of the MLE. The model continues to be an active area of research, with recent advancements focusing on modern statistical challenges such as simultaneous variable selection in high-dimensional settings (Shohoudi et al., 2016). This progression highlights the increasing complexity of the model, underscoring the need for a rigorous understanding of the theoretical underpinnings required for valid inference, particularly in the presence of covariates.

While verifying non-singularity is often straightforward in simple, well-posed models (Wang, 1993), it becomes a formidable challenge in more complex and empirically vital frameworks. Two such areas are of particular interest: finite mixture models and change-point models. Finite mixture models, whose identifiability was first systematically studied by Teicher (1963), are notorious for exhibiting singularities when components overlap, leading to a loss of rank in the information matrix (Chen, 1995; Lindsay, 1995; Manole and Ho, 2020; McLachlan et al., 2019). For complex mixture models like the MCP, direct verification of FIM nonsingularity for all possible parameter values is often intractable due to the intricate dependencies within the likelihood function. Indeed, the potential for a singular information matrix is a well-known theoretical challenge in the general mixture model literature, often arising from component redundancy or overfitting (Kim and Lindsay, 2015). A powerful alternative approach was developed by Asgharian (2014), who established a general theory concerning the singularities of the FIM. The central result is that if a model satisfies certain smoothness conditions and, crucially, is identifiable, then the set of parameter values for which the FIM is singular constitutes a nowhere dense set of Lebesgue measure zero. This effectively means that singularities are rare and do not pose a practical obstacle to inference for almost all data-generating processes. To illustrate this theory, Asgharian (2014) successfully applied it to an MCP model where the hazard of change, $\pi_k(\mathbf{z})$, was specified using a proportional odds ratio structure, which is closely related to the logistic link function.

A natural question arises: do these powerful theoretical guarantees hold for other common specifications of the change hazard? This paper answers this question affirmatively by moving beyond a single link function and addressing a general class of models. Our primary contribution is to establish the quasi-identifiability of an MCP model that employs a general strictly monotone link function for the hazard, which includes the probit, logit, and complementary log-log as prominent examples. This general result is significant as it allows us to directly invoke the theorems of Asgharian (2014) to conclude that the FIM is nonsingular almost everywhere in the parameter space. This work therefore substantially broadens the class of identifiable MCP models for which standard likelihood-based inference is theoretically sound.

2 The MCP model with a general link function

We follow the general framework of the MCP model as described in Asgharian (2014). The full data for a single subject consists of observations $\mathbf{x} = (x_1, \ldots, x_m)$ and a vector of time-invariant covariates $\mathbf{z}$. The model is formulated as a finite mixture, where the density of the observations is given by

$$f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{m} \alpha_k(\mathbf{z}|\boldsymbol{\theta})\, g_{\mathbf{z},k}(\mathbf{x}|\boldsymbol{\gamma}). \tag{1}$$

Here, $g_{\mathbf{z},k}(\mathbf{x}|\boldsymbol{\gamma})$ is the conditional density of the observations given that the change occurs at time k, and $\alpha_k(\mathbf{z}|\boldsymbol{\theta}) = P(\tau = k|\mathbf{Z} = \mathbf{z})$ is the mixing weight. The weights are constructed from the discrete-time hazard of change, $\pi_k(\mathbf{z}|\boldsymbol{\theta}) = P(\tau = k|\tau \geq k, \mathbf{Z} = \mathbf{z})$, as

$$\alpha_k(\mathbf{z}|\boldsymbol{\theta}) = \pi_k(\mathbf{z}|\boldsymbol{\theta}) \prod_{j=1}^{k-1} (1 - \pi_j(\mathbf{z}|\boldsymbol{\theta})), \tag{2}$$

for $k = 1, \ldots, m-1$, with $\alpha_m(\mathbf{z}|\boldsymbol{\theta}) = \prod_{j=1}^{m-1}(1 - \pi_j(\mathbf{z}|\boldsymbol{\theta}))$ to ensure they sum to one.

The central innovation of this paper lies in generalizing the specification of the hazard function. We model the hazard using a structure analogous to a generalized linear model (GLM), employing a general link function, G. The hazard is defined as

$$\pi_k(\mathbf{z}|\boldsymbol{\theta}) = G\left(\mathbf{z}^\top \boldsymbol{\beta} + \rho_k\right), \quad \text{for } k = 1, \ldots, m-1, \tag{3}$$

where

- $G : \mathbb{R} \to (0,1)$ is a known link function, assumed to be twice continuously differentiable and strictly monotone;
- $\mathbf{z} \in \mathbb{R}^p$ is the vector of covariates;
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients;
- ρ_k represents the baseline hazard component at time k, forming the vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{m-1}) \in \mathbb{R}^{m-1}$.

The complete parameter vector for the model is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\gamma})$.

Remark 1 (The class of admissible link functions)**.** *The requirement that the link function G be strictly monotone (and thus injective) is the key property that facilitates identifiability. This general framework encompasses many of the most common models for binary outcomes, making our results widely applicable. Prominent examples include:*

- ***The probit model:*** *$G(x) = \Phi(x)$, where Φ is the standard normal CDF.*
- ***The logit model:*** *$G(x) = \frac{e^x}{1+e^x}$, which corresponds to the proportional odds ratio structure used in Asgharian (2014).*
- ***The complementary log-log (cloglog) model:*** *$G(x) = 1 - \exp(-\exp(x))$.*

By proving our results for a general G, we simultaneously establish them for all these specific models and countless others that satisfy the monotonicity condition.

3 Main result: quasi-identifiability

The primary theoretical hurdle to applying the singularity framework of Asgharian (2014) is to establish model identifiability. For models incorporating covariates, the appropriate concept is quasi-identifiability. We first formally define this concept, following the original work.

Definition 1 (Quasi-identifiability, Asgharian (2014))**.** *A model parameterized by $\boldsymbol{\theta} \in \Theta$ is said to be quasi-identifiable if for any two distinct parameter vectors $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$, the set of covariates $\mathbf{z}$ for which $f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}')$ for almost all $\mathbf{x}$ has measure zero with respect to the covariate distribution.*

To construct the proof of our main theorem in a clear, modular fashion, we first establish a simple lemma. This lemma confirms the intuitive fact that if the final mixing weights α_k are identical, then the initial hazard probabilities π_k that generated them must also have been identical.

Lemma 1 (Equivalence of weights and hazards)**.** *Let the mixing weights* $\{\alpha_k\}_{k=1}^m$ *be generated from the hazard probabilities* $\{\pi_k\}_{k=1}^{m-1}$ *as in equation* (2)*. Assume* $\pi_k \in (0,1)$ *for all* k*, as a hazard of 0 or 1 would imply a deterministic change process, collapsing the mixture structure. If two sets of hazards,* $\{\pi_k\}$ *and* $\{\pi_k^*\}$*, generate identical mixing weights* $\alpha_k = \alpha_k^*$ *for all* $k = 1,\ldots,m$*, then it must be that* $\pi_k = \pi_k^*$ *for all* $k = 1,\ldots,m-1$*.*

Proof. We proceed by induction on k. **Base Case (k=1):** By definition, $\alpha_1 = \pi_1$. Since $\alpha_1 = \alpha_1^*$, it immediately follows that $\pi_1 = \pi_1^*$.
Inductive Step: Assume that $\pi_j = \pi_j^*$ for all $j = 1,\ldots,k-1$. We need to show this implies $\pi_k = \pi_k^*$. The definition for α_k is $\alpha_k = \pi_k \prod_{j=1}^{k-1}(1-\pi_j)$. Since $\alpha_k = \alpha_k^*$, we have

$$\pi_k \prod_{j=1}^{k-1}(1-\pi_j) = \pi_k^* \prod_{j=1}^{k-1}(1-\pi_j^*).$$

By our inductive hypothesis, $\pi_j = \pi_j^*$ for all $j < k$, so the product terms $\prod(\cdot)$ on both sides are identical. Since $\pi_j < 1$, this product is non-zero. We can therefore divide both sides by this common term to obtain $\pi_k = \pi_k^*$. This completes the induction. □

We are now equipped to state and prove the main result of this paper.

Theorem 1 (General quasi-identifiability)**.** *Consider the MCP model defined by equations* (1) *and* (3)*, using any strictly monotone, twice continuously differentiable link function* G*. Assume that:*

(i) The family of component densities $\{g_{\mathbf{z},k}(\cdot|\gamma)\}_{k=1}^m$ *is identifiable with respect to* γ *and forms a linearly independent set.*

(ii) The covariate vector $\mathbf{z}$ *contains an intercept, and its distribution is not concentrated on a hyperplane.*

Then, the model is quasi-identifiable with respect to the full parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\rho}, \gamma)$*.*

Proof. Suppose for two parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, we have $f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}^*)$ for all covariates $\mathbf{z}$ in a set A of positive measure. From the model's mixture structure in (1), this implies

$$\sum_{k=1}^m \alpha_k(\mathbf{z}|\boldsymbol{\theta})\, g_{\mathbf{z},k}(\mathbf{x}|\gamma) = \sum_{k=1}^m \alpha_k(\mathbf{z}|\boldsymbol{\theta}^*)\, g_{\mathbf{z},k}(\mathbf{x}|\gamma^*).$$

By assumption (i), the component densities $\{g_{\mathbf{z},k}\}$ are linearly independent, which for a finite mixture implies the equality of both the component parameters and the mixing weights. Thus, for $\mathbf{z} \in A$, we must have $\gamma = \gamma^*$ and $\alpha_k(\mathbf{z}|\boldsymbol{\theta}) = \alpha_k(\mathbf{z}|\boldsymbol{\theta}^*)$ for all k.

By Lemma 1, the equality of the mixing weights implies the equality of the underlying hazard probabilities: $\pi_k(\mathbf{z}|\boldsymbol{\theta}) = \pi_k(\mathbf{z}|\boldsymbol{\theta}^*)$. Substituting our general model specification from (3) yields

$$G(\mathbf{z}^\top \boldsymbol{\beta} + \rho_k) = G(\mathbf{z}^\top \boldsymbol{\beta}^* + \rho_k^*).$$

The core assumption on our link function G is that it is **strictly monotone**, which guarantees it is **injective** (one-to-one). Therefore, the equality of the function's outputs implies the equality of its inputs:

$$\mathbf{z}^\top \boldsymbol{\beta} + \rho_k = \mathbf{z}^\top \boldsymbol{\beta}^* + \rho_k^*.$$

This can be rearranged to $\mathbf{z}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^*) = \rho_k^* - \rho_k$. This equation must hold for all $\mathbf{z} \in A$. If $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$, the left-hand side is a non-constant linear function of $\mathbf{z}$, while the right-hand side is a constant. By assumption (ii), the distribution of $\mathbf{z}$ is not concentrated on a hyperplane, meaning the set of $\mathbf{z}$ values for which a

non-constant linear function equals a constant must have measure zero. This contradicts our premise that the equality holds on a set A of positive measure.

The only way to resolve this contradiction is to conclude that $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Substituting this back into the linear equation immediately yields $\rho_k = \rho_k^*$. Combined with $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$, we have shown that $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, which completes the proof of quasi-identifiability. □

Example 1 (Importance of the theorem's assumptions)**.** To illustrate the necessity of the conditions in Theorem 1, consider a simple case with $m = 2$ (one potential change-point) and a single covariate z_1. For concreteness, we use a probit link, so the hazard is $\pi_1(\mathbf{z}) = \Phi(\beta_0 + \beta_1 z_1)$, where $\mathbf{z} = (1, z_1)^\top$. The model density is a two-component mixture

$$f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = \pi_1(\mathbf{z}) g_1(\mathbf{x}|\boldsymbol{\gamma}) + (1 - \pi_1(\mathbf{z})) g_2(\mathbf{x}|\boldsymbol{\gamma}).$$

Now, consider scenarios where the assumptions of the theorem are violated.

1. **Violation of covariate richness (Assumption ii):** Suppose the covariate z_1 is not sufficiently rich and can only take a single value, c. Then the linear predictor becomes a constant, $\eta = \beta_0 + \beta_1 c$, and so does the hazard, $\pi_1 = \Phi(\eta)$. In this case, countless combinations of β_0 and β_1 can produce the exact same value for η. For instance, $(\beta_0, \beta_1) = (\eta, 0)$ and $(\beta_0, \beta_1) = (\eta - c, 1)$ would yield the same hazard. The parameters β_0 and β_1 are not separately identifiable. This highlights why the covariates cannot be concentrated on a hyperplane (in this case, a single point).

2. **Violation of component distinguishability (Assumption i):** Suppose the component densities are identical, i.e., $g_1(\cdot|\boldsymbol{\gamma}) = g_2(\cdot|\boldsymbol{\gamma}) = g(\cdot|\boldsymbol{\gamma})$. The mixture model then collapses, as the mixing weights become irrelevant:
$$f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = \pi_1(\mathbf{z}) g(\mathbf{x}|\boldsymbol{\gamma}) + (1 - \pi_1(\mathbf{z})) g(\mathbf{x}|\boldsymbol{\gamma}) = g(\mathbf{x}|\boldsymbol{\gamma}).$$
The resulting density is entirely independent of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$. Consequently, these parameters are completely unidentifiable as they have no influence on the likelihood. This illustrates the necessity of linear independence for the component densities.

4 Implications for the information matrix

Having established the quasi-identifiability for a broad class of MCP models in Theorem 1, we are now positioned to directly apply the general singularity theory developed in Asgharian (2014). This theory connects model identifiability with the geometric and analytic properties of the set of singularities of the Fisher information matrix, $I(\boldsymbol{\theta})$. Our general result provides the key to unlock these powerful implications.

The following corollary is the principal outcome of our analysis. It confirms that for any model within our proposed general class, the set of parameter values where standard asymptotic theory might fail is negligible.

Corollary 1. *Consider the MCP model defined with any strictly monotone, twice continuously differentiable link function G as in Section 2. Assume that the conditions of Theorem 1 hold. Further, assume that the model satisfies the necessary smoothness and regularity conditions (e.g., A.1–A.5 and B.1–B.3 as specified in Asgharian (2014)). Let $\Lambda = \{\boldsymbol{\theta} \in \Theta : \det I(\boldsymbol{\theta}) = 0\}$ be the set of singularities of the Fisher information matrix. Then:*

1. *The set Λ is a nowhere dense subset of the parameter space Θ.*

2. *If, in addition, the component densities $g_{\mathbf{z},k}(\mathbf{x}|\boldsymbol{\gamma})$ are analytic functions of the parameters, the set Λ has Lebesgue measure zero.*

Proof. The proof is a direct application of the main results from Asgharian (2014). Theorem 1 of this paper establishes the quasi-identifiability of our general model, which is the central prerequisite for their Theorems 3 and 4. By assuming that the link function G is twice continuously differentiable and that the component densities satisfy standard smoothness conditions, the overall regularity conditions A.1–A.5 and B.1–B.3 of Asgharian (2014) are met. Specifically:

- Theorem 3 of Asgharian (2014) states that for any quasi-identifiable and sufficiently smooth model, the set of singularities Λ is nowhere dense.
- Theorem 4 of Asgharian (2014) strengthens this result, stating that if the model is also analytic, Λ has Lebesgue measure zero.

Since our general model satisfies the key conditions of quasi-identifiability and smoothness, the conclusions of the corollary follow directly by invoking their theorems. □

Remark 2 (Theoretical and practical implications)**.** *The conclusions of Corollary 1 are of profound theoretical importance. They provide a rigorous justification for the use of standard likelihood-based inferential procedures (e.g., Wald tests, likelihood ratio tests, and confidence intervals based on the inverse of the FIM) for a wide and versatile class of MCP models. The property of being nowhere dense implies that singularities are isolated, while having Lebesgue measure zero implies that the probability of randomly selecting a singular parameter is zero. Together, these results confirm that the pathologies associated with a singular FIM are confined to a negligible, thin set of parameter values.*

It is important, however, to distinguish this theoretical, asymptotic result from practical, finite-sample challenges. In applied settings, numerical issues akin to singularity can arise due to data structure, particularly the problem of complete or quasi-complete separation, where covariates perfectly predict the change-point time. In such cases, maximum likelihood estimates may diverge, and the observed Fisher information matrix can become ill-conditioned or numerically singular. Our results provide the theoretical foundation for asymptotic inference, but they do not eliminate the need for careful model diagnostics and an awareness of data-induced instabilities in finite samples.

5 Conclusion

In this paper, we addressed the fundamental issue of Fisher information matrix singularity for multipath change-point (MCP) models, building upon the theoretical framework of Asgharian (2014). Their work guaranteed the negligibility of singularities for identifiable models, a result they demonstrated for an MCP model with a proportional odds ratio (logistic) structure for the change hazard. We investigated whether these guarantees were an artifact of that specific choice or a more fundamental property of the modeling framework.

Our main contribution was to unify and generalize this result by establishing, under standard smoothness conditions, the quasi-identifiability of MCP models for a broad class of strictly monotone link functions. By focusing on the core mathematical property required for identifiability—the injectivity of the link function—our main theorem (Theorem 1) simultaneously covers the probit, logit, complementary log-log, and numerous other model specifications under a single, coherent proof.

The principal implication, presented in Corollary 1, is that for this entire class of models, the set of FIM singularities is both nowhere dense and of Lebesgue measure zero. This finding provides a rigorous and general theoretical justification for the application of standard likelihood-based inference to a versatile set of MCP models. Our work not only broadens the family of models for which inference is known to be theoretically sound but also reinforces the robustness of the singularity framework by demonstrating its wide applicability. While this asymptotic result provides a strong foundation, we also acknowledge that it must be distinguished from finite-sample challenges, where issues like complete data separation can still lead to numerical instability.

Future research could extend this line of inquiry in several directions, such as investigating the identifiability of models with more complex structures, including time-varying covariates or multiple change-points; exploring the challenges posed by non-monotone or data-driven link functions; or developing robust inferential methods that are less sensitive to the practical problem of data separation in finite samples. In essence, this work demonstrates that the validity of standard inference in these complex models hinges on a simple and verifiable property of the chosen link function: its monotonicity. This insight provides both theoretical assurance and practical guidance to researchers working with multipath change-point models.

Acknowledgements

I am deeply grateful to Professor Masoud Asgharian (McGill University) for his continuous invaluable guidance, which has significantly shaped both this work and my broader research endeavors. We would like to thank the Editor-in-Chief, Associate Editor and anonymous reviewers for their thorough evaluation of the manuscript and for their valuable comments and suggestions.

References

Asgharian, M. (2014). On the singularities of the information matrix and multipath change-point problems. *Theory of Probability and Its Applications*, **58**, 546–561.

Asgharian, M. and Wolfson, D. B. (2001). Covariates in multipath change-point problems: Modelling and consistency of the MLE. *Canadian Journal of Statistics*, **29**, 515–528.

Chen, J. (1995). Optimal rate of convergence for finite mixture models. *The Annals of Statistics*, **23**, 221–233.

Joseph, L., Vandal, A. C., and Wolfson, D. B. (1996). Estimation in the multipath change point problem for correlated data. *The Canadian Journal of Statistics*, **24**, 37–53.

Joseph, L. and Wolfson, D. B. (1992). Estimation in multi-path change-point problems. *Communications in Statistics-Theory and Methods*, **21**, 897–913.

Joseph, L. and Wolfson, D. B. (1993). Maximum likelihood estimation in the multi-path change-point problem. *Annals of the Institute of Statistical Mathematics*, **45**, 511–530.

Kim, D. and Lindsay, B. G. (2015). Empirical identifiability in finite mixture models. *Annals of the Institute of Statistical Mathematics*, **67**, 745–772.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics; Penn. State University.

Manole, T. and Ho, N. (2020). Uniform convergence rates for maximum likelihood estimation under two-component Gaussian mixture models. *arXiv preprint arXiv:2006.00704*.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, **6(**, 355–378.

Shohoudi, A., Khalili, A., Wolfson, D. B., and Asgharian, M. (2016). Simultaneous variable selection and de-coarsening in multi-path change-point models. *Journal of Multivariate Analysis*, **143**, 15–429.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, **34**, 1265–1269.

Titterington, D. M., Smith, A. F., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

Wang, X. (1993). Non-singularity of fisher information for autoregressive moying-average processes. *Journal of Time Series Analysis*, **14**, 547–548.

Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 53–70, Research Paper
https://smps.um.ac.ir, ISSN: 3115-9370

Efficient pseudo Rao-Blackwellized estimator in ranked sampling

Bardia Panahbehagh[†] **and Mahdi Salehi**[‡,*]

[†]*Department of Mathematics, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran*
[‡]*Department of Mathematics and Statistics, University of Neyshabur, Neyshabur, Iran*

Email(s): panahbehagh@khu.ac.ir, salehi.sms@neyshabur.ac.ir

Abstract. Ranked set sampling (RSS) and judgment post-stratification (JPS) are efficient sampling designs that rely on ranking units without requiring full measurement, typically through visual inspection or the use of auxiliary information. In their standard forms, including multi-ranker RSS, only a portion of the available ranking information is utilized, and the corresponding Rao–Blackwellized estimators are generally unavailable in closed form. In this paper, we propose two practical pseudo Rao–Blackwellization approaches that enhance RSS and JPS by incorporating ranking information from all selected sets. The resulting estimators admit closed-form expressions, make more complete use of the available ranking information for each measured unit, and mitigate the impact of inaccurate rankings from individual sets. We further extend the proposed framework to other ranked sampling designs, including median RSS and extreme RSS, and investigate the performance of the resulting estimators under both symmetric and asymmetric populations. Finally, we analyze two real data sets to illustrate the practical applicability of the proposed methods and to corroborate the findings from the simulation studies.

Keywords: Efficiency; Judgment; Median ranked set sampling; Post-stratification; Ranked set sampling; Rao-Blackwellization.

Mathematics Subject Classification (2020): 62G05, 62G30.

1 Introduction

Ranked set sampling (RSS) and judgment post-stratification (JPS), introduced by McIntyre (1952) and MacEachern et al. (2004), respectively, utilize ranking information to improve estimation accuracy. In these sampling schemes, ranking is typically performed without full measurement of the variable of interest; instead, it relies on visual inspection or expert judgment. Following its introduction for estimating pasture and forage yields, RSS was rigorously studied by Takahasi and Wakimoto (1968), and since then it

*Corresponding author
Received: 8 June 2026 / Revised: 1 July 2026 / Accepted: 6 July 2026

https://doi.org/10.22067/smps.2026.99259.1061

The Author(s). This is an open access article under CC BY 4.0.

has attracted considerable attention across various fields. For example, Yanagawa and Shirahata (1976) proposed a random allocation approach in RSS that coincides with JPS under an infinite population framework.

A substantial body of literature has compared RSS and JPS with simple random sampling (SRS). Notable contributions include Stokes and Sager (1988), Kaur et al. (1995), Chen et al. (2004), Salehi et al. (2016), and Dey et al. (2017). A key finding from these studies is that RSS is generally more efficient than, or at least as efficient as, SRS, depending on the quality of the ranking process. More recent developments have extended RSS to settings with missing data; for instance, Bhushan and Kumar (2024) proposed logarithmic imputation techniques under RSS to address non-response while preserving estimation efficiency.

MacEachern et al. (2004) introduced JPS as a post-stratification method applied to an SRS using auxiliary ranking information. While JPS typically improves upon SRS, it is generally less efficient than RSS for estimating population means. Frey and Ozturk (2011) showed that strata induced by ranking satisfy specific structural constraints not present in conventional stratification and proposed a Rao–Blackwellized estimator, albeit without a closed form. Ozturk (2012) further integrated ranking information from multiple rankers within JPS and RSS frameworks.

Subsequent work has explored theoretical and methodological extensions of JPS. Ozturk (2014) demonstrated that the efficiency of JPS-based quantile estimators lies between that of SRS and RSS in finite samples and approaches that of RSS as the sample size increases. Ozturk (2016) developed JPS designs for finite populations and proposed algorithms to approximate the Rao–Blackwellized mean estimator. Matthews and Wolfe (2017) introduced a unified framework encompassing both RSS and JPS methodologies. More recently, Ozturk and Bayramoglu Kavlak (2020) and Ozturk and Kravchuk (2021) proposed improved inference procedures by incorporating additional ranking information.

Despite its lower efficiency compared to RSS, JPS is often preferred in practice because it retains the structure of SRS, with ranking information used only at the estimation stage. This feature allows practitioners to apply standard inferential tools developed for SRS while still benefiting from auxiliary ranking information. Recent contributions, such as Ozturk (2023), continue to expand the applicability of rank-based methods in diverse experimental settings.

Both RSS and JPS involve multiple sets, each containing several units, raising an important question: how can the full ranking information across all sets be utilized optimally? Panahbehagh et al. (2018) addressed this issue by proposing a cost-efficient sampling strategy based on RSS. This work was extended by Panahbehagh and Bruggemann (2021) and Panahbehagh et al. (2021) to multivariate settings, where more than one unit can be selected from each set. As long as the number of selected units per rank remains smaller than the number of sets, the resulting design outperforms SRS.

In this paper, we propose a method that fully exploits the ranking information from all units across all sets, along with two closed-form estimators. In both RSS and JPS, ranking is typically based on easily observable auxiliary variables. While such variables provide substantial information, conventional approaches utilize only a limited portion of it. For example, Ozturk (2012) uses ranking information from a single set for each unit, whereas incorporating information from multiple sets can potentially yield greater efficiency.

The remainder of the paper is organized as follows. The next section briefly reviews rank-based sampling designs, including RSS and JPS. Section 2 introduces the proposed improving approaches. Section 3 presents the corresponding estimators of the population mean. A simulation study and further numerical explorations are considered in Section 4. Section 5 provides applications to two real datasets, and Section 6 concludes the paper.

1.1 A brief explanation of RSS and JPS

Suppose that the random variable X has probability density function (pdf) f_μ, with $E(X) = \mu$, $Var(X) = \sigma^2$, and $E(|X|^2) < \infty$. Our objective is to estimate the population mean μ by exploiting the ranking

information provided by one or more auxiliary covariates, denoted collectively by z.

To derive an RSS of size $n = cM$, an SRS of size M is selected from the population. None of the units are measured at this stage, but they are instead ranked by judgment without actual measurement. In this regard, visual inspection, expert opinion, or highly correlated concomitant variables may be used for judgment ranking. The unit, which seems to be the smallest one among the M units is measured, and it will be the first observation of the RSS, denoted by $X_{(1)11}$. Another SRS of size M is obtained, and the judgments are ranked similarly. The unit judgment ranked, which sounds to be the second smallest one, is measured as the second member of the RSS, denoted by $X_{(2)21}$. This procedure continues to up to the unit having the largest (M^{th}) judgment rank is measured and completes the last observation of the RSS, denoted by $X_{(M)M1}$. Finally, the sample

$$\left(X_{(1)11}, X_{(2)21}, \cdots, X_{(M)M1}\right)$$

is called a one-cycle RSS of size M. Repeating the above process $c-1$ times more, gives a c-cycle RSS containing $n = cM$ units, denoted by $\left\{X_{(i)ij},\ i = 1, \cdots, M,\ j = 1, \cdots, c\right\}$, where $X_{(i)ij}$ is the i^{th} judgment ranked unit from the i^{th} SRS in the j^{th} cycle.

JPS is another rank-based sampling that was proposed by MacEachern et al. (2004). In this design, an SRS of size n, $X_1, \cdots, X_n$, is taken from the population and actually measured. Then, a further SRS of size $M-1$ is obtained for each X_i. These extra units are used only as auxiliary ones for judgment ranking of X_i. Hence, we will finally have the pairs of (X_i, R_i), $i = 1, \cdots, n$, where R_i ($\in \{1, \cdots, M\}$) is the judgment rank of X_i among its corresponding SRS.

2 Pseudo Rao-Blackwellization of JPS and RSS

In this section we propose two methods to improve JPS and RSS by exploiting set information in a manner analogous to Rao–Blackwellization, but that yield closed-form, approximately unbiased estimators. The Improving JPS (IJPS) procedure for sample size n proceeds as follows. First, draw a simple random sample with replacement (SRS) of size n from f and fully measure its units, denoted $X_1, \ldots, X_n$. Independently, select n sets, each an SRS of size $M-1$, without full measurement, and use all sets to provide judgmental ranks for the measured sample units. Thus, for each unit X_k we obtain n ranks—one from each set—and form the pairs $(X_k, \mathbf{R}_k)$ for $k = 1, \ldots, n$, where $\mathbf{R}_k = (R_{k1}, R_{k2}, \ldots, R_{kn})$ and $1 \le R_{ki} \le M$ denotes the judgmental rank of X_k in the ith set (see Table 1).

Table 1: IJPS of size n ranking in M strata.

$\boldsymbol{R}$	X_1	X_2	$\cdots$	X_n
1^{st} set	R_{11}	R_{21}	$\cdots$	R_{n1}
2^{nd} set	R_{12}	R_{22}	$\cdots$	R_{n2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n^{th} set	R_{1n}	R_{2n}	$\cdots$	R_{nn}

For the improving RSS (IRSS) of size n, assume $n = cM$ where c is an integer. We implement a conventional RSS with a c-cycle of size M (see Table 2). Up to this point, we have a conventional RSS in which the final fully measured sample is given by

$$\{X_{(1)11}, X_{(2)21}, \ldots, X_{(M)M1}, X_{(1)12}, X_{(2)22}, \ldots, X_{(M)M2}, \ldots, X_{(1)1c}, X_{(2)2c}, \ldots, X_{(M)Mc}\},$$

where $X_{(i)ij}$ denotes the ith order statistic from the ith set in the jth cycle. We therefore have $n = cM$ sets, all of which can be utilized to extract additional information about the fully measured sample units. To

Table 2: IRSS of size n ranking in M strata. Each row is ordered. Underbraces indicate selected samples for full measurement.

		Ranks			
Cycle	Set	1	2	$\cdots$	M
1	1	$\underbrace{X_{(1)11}}$	$X_{(2)11}$	$\cdots$	$X_{(M)11}$
	2	$X_{(1)21}$	$\underbrace{X_{(2)21}}$	$\cdots$	$X_{(M)21}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	M	$X_{(1)M1}$	$X_{(2)M1}$	$\cdots$	$\underbrace{X_{(M)M1}}$
2	1	$\underbrace{X_{(1)12}}$	$X_{(2)12}$	$\cdots$	$X_{(M)12}$
	2	$X_{(1)22}$	$\underbrace{X_{(2)22}}$	$\cdots$	$X_{(M)22}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	M	$X_{(1)M2}$	$X_{(2)M2}$	$\cdots$	$\underbrace{X_{(M)M2}}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
c	1	$\underbrace{X_{(1)1c}}$	$X_{(2)1c}$	$\cdots$	$X_{(M)1c}$
	2	$X_{(1)2c}$	$\underbrace{X_{(2)2c}}$	$\cdots$	$X_{(M)2c}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	M	$X_{(1)Mc}$	$X_{(2)Mc}$	$\cdots$	$\underbrace{X_{(M)Mc}}$

this end, we construct a multi-part matrix (MPM), as illustrated in Table 2. The first part of this matrix consists of the data in Table 2 with the column corresponding to the smallest units ($X_{(1)ij}$) removed. The second part is obtained by removing the column corresponding to the second smallest units ($X_{(2)ij}$), and this process continues similarly. See Table 3 for the kth part of the MPM. Then, we use Part 1 to indicate n ranks for each of $X_{(1)1j}; j = 1,2,...,c$, Part 2 to indicate n ranks for each of $X_{(2)2j}; j = 1,2,...,c$, and so on. Therefore, for each fully measured unit, we obtain n ranks, which leads to a table analogous to Table 1, n observations, each with n ranks.

Then, we re-match each of $\{X_{(k)kj}; j = 1,2,...,c\}$, units (which have rank k in their original sets), in all the $n-1$ sets in Part k to create sets of size M and indicate $n-1$ additional ranks for each of $\{X_{(k)kj}; j = 1,2,...,c\}$. Therefore for each full measured unit we will have n ranks which leads to a table exactly like Table 1, n observations, each with n ranks. As a simple example consider $M = 3$ and $c = 2$, which leads to Table 4. Now each of $X_{(1)11}$ and $X_{(1)12}$ which will be fully measured, will receive 5 new ranks with each of the sets in Part 1 of Table 5. Please note that ranking them in their original sets leads to rank 1 for them. With this procedure, although we expect these units to receive rank 1, but the rank of them will be adjusted using the other sets, leading to an increase in the accuracy of our judgment in estimating the mean parameter.

In usual Rao-Blackwellization method, we would use more information involving all the possible combinations of all the units of the sets which will lead to complexity both in implementation and estimation. But based on the approaches proposed in this section, we will construct closed form of the estimators easily. Then the proposed methods utilize information inside the sets, more than usual approaches and less than Rao-Blackwellization, and that's why we name them pseudo Rao-Blackwellization approaches.

Table 3: Part k of Multi-part matrix of data to indicate ranks of full measured units.

Part k						
$X_{(1)11}$	$X_{(2)11}$	$\cdots$	$X_{(k-1)11}$	$X_{(k+1)11}$	$\cdots$	$X_{(M)11}$
$X_{(1)21}$	$X_{(2)21}$	$\cdots$	$X_{(k-1)21}$	$X_{(k+1)21}$	$\cdots$	$X_{(M)21}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_{(1)M1}$	$X_{(2)M1}$	$\cdots$	$X_{(k-1)M1}$	$X_{(k+1)M1}$	$\cdots$	$X_{(M)M1}$
$X_{(1)12}$	$X_{(2)12}$	$\cdots$	$X_{(k-1)12}$	$X_{(k+1)12}$	$\cdots$	$X_{(M)12}$
$X_{(1)22}$	$X_{(2)22}$	$\cdots$	$X_{(k-1)22}$	$X_{(k+1)22}$	$\cdots$	$X_{(M)22}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_{(1)M2}$	$X_{(2)M2}$	$\cdots$	$X_{(k-1)M2}$	$X_{(k+1)M2}$	$\cdots$	$X_{(M)M2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_{(1)1c}$	$X_{(2)1c}$	$\cdots$	$X_{(k-1)1c}$	$X_{(k+1)1c}$	$\cdots$	$X_{(M)1c}$
$X_{(1)2c}$	$X_{(2)2c}$	$\cdots$	$X_{(k-1)2c}$	$X_{(k+1)2c}$	$\cdots$	$X_{(M)2c}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_{(1)Mc}$	$X_{(2)Mc}$	$\cdots$	$X_{(k-1)Mc}$	$X_{(k+1)Mc}$	$\cdots$	$X_{(M)Mc}$

Table 4: An example of IRSS with $M = 3$ and $c = 2$.

		Ranks		
Cycle	Set	1	2	3
1	1	$\underbrace{X_{(1)11}}$	$X_{(2)11}$	$X_{(3)11}$
	2	$X_{(1)21}$	$\underbrace{X_{(2)21}}$	$X_{(3)21}$
	3	$X_{(1)31}$	$X_{(2)31}$	$\underbrace{X_{(3)31}}$
2	1	$\underbrace{X_{(1)12}}$	$X_{(2)12}$	$X_{(3)12}$
	2	$X_{(1)22}$	$\underbrace{X_{(2)22}}$	$X_{(3)22}$
	3	$X_{(1)32}$	$X_{(2)32}$	$\underbrace{X_{(3)32}}$

3 Estimators

JPS is more acceptable than RSS (because of working based on SRS) and its efficiency is asymptotically the same as RSS (MacEachern et al. (2004)). We therefore introduce an estimator for JPS in detail, and we will suffice to present only an estimator for RSS in this section.

Let $(X_k, \mathbf{R_k}); k = 1, 2, ..., n$, be an IJPS of size n drawn from a population with mean μ. Suppose further that the function $I^m_{R_{ki}}$ is an indicator variable which equals 1 if $R_{ki} = m$ and 0 otherwise. Now, the fraction of cases where X_k takes the rank m, denoted by ω_{km}, is obtained as

$$\omega_{km} = \frac{1}{n_m} \sum_{i=1}^{n} I^m_{R_{ki}}, \qquad n_m = \sum_{k=1}^{n} \sum_{i=1}^{n} I^m_{R_{ki}},$$

Table 5: The parts of the three-layer matrix associated with the example of Table 4.

Part 1		Part 2		Part 3	
$X_{(2)11}$	$X_{(3)11}$	$X_{(1)11}$	$X_{(3)11}$	$X_{(1)11}$	$X_{(2)11}$
$X_{(2)21}$	$X_{(3)21}$	$X_{(1)21}$	$X_{(3)21}$	$X_{(1)21}$	$X_{(2)21}$
$X_{(2)31}$	$X_{(3)31}$	$X_{(1)31}$	$X_{(3)31}$	$X_{(1)31}$	$X_{(2)31}$
$X_{(2)12}$	$X_{(3)12}$	$X_{(1)12}$	$X_{(3)12}$	$X_{(1)12}$	$X_{(2)12}$
$X_{(2)22}$	$X_{(3)22}$	$X_{(1)22}$	$X_{(3)22}$	$X_{(1)22}$	$X_{(2)22}$
$X_{(2)32}$	$X_{(3)32}$	$X_{(1)32}$	$X_{(3)32}$	$X_{(1)32}$	$X_{(2)32}$

where n_m counts the total number of observations of IJPS whose ranks equal m. Then, for estimating μ based on IJPS we propose the following estimator

$$\hat{\mu}_{\text{IJPS}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\mu}_{[m]}, \tag{1}$$

where

$$\hat{\mu}_{[m]} = \begin{cases} \sum_{k=1}^{n}\omega_{km}X_k, & n_m > 0, \\ 0, & n_m = 0. \end{cases} \tag{2}$$

Conjecture 1. Under the proposed IJPS mechanism, for any two distinct measured units X_k and $X_{k'}$, $k \neq k'$, the corresponding weighted observations satisfy

$$\text{Cov}\left(\omega_{k.}X_k, \omega_{k'.}X_{k'}\right) \leq 0.$$

Motivated by the fixed-sum property of the weights and by the numerical evidence from our simulation study, we state the following conjecture.

The intuition behind Conjecture 1 is as follows. The measured observations X_k and $X_{k'}$ are independent, while their weights $\omega_{k.}$ and $\omega_{k'.}$ are negatively related because they are constructed from the same ranking information and satisfy the fixed-sum relation

$$\sum_{k=1}^{n}\omega_{k.} = M$$

when all rank classes are non-empty. In particular, under exchangeability, this fixed-sum relation implies

$$\text{Cov}(\omega_{k.}, \omega_{k'.}) = -\frac{\text{Var}(\omega_{k.})}{n-1} \leq 0, \qquad k \neq k'.$$

Although the weights are rank-based and therefore related to the measured values, extensive simulations did not show any contradiction to Conjecture 1. Across the parent distributions, sample sizes, and set sizes considered in this study, the estimated covariance terms were consistently non-positive.

Lemma 1. *The random variable* $\hat{\mu}_{\text{IJPS}}$ *given by* (1) *is an asymptotically unbiased estimator of* μ. *More specifically, under the usual equal-probability rank-class argument,*

$$E\left(\hat{\mu}_{\text{IJPS}}\right) = \mu - \left(1 - \frac{1}{M}\right)^{n^2}\mu,$$

and hence

$$E\left(\hat{\mu}_{\mathrm{IJPS}}\right)-\mu=-\left(1-\frac{1}{M}\right)^{n^2}\mu\longrightarrow 0\qquad as\ n\to\infty.$$

Moreover, if Conjecture 1 holds, then

$$\hat{V}=\frac{n}{M^2(n-1)}\sum_{k=1}^{n}\left(\omega_{k.}X_k-\frac{1}{n}\sum_{k=1}^{n}\omega_{k.}X_k\right)^2 \tag{3}$$

is an empirical estimator of an upper bound for the variance of $\hat{\mu}_{\mathrm{IJPS}}$*, where*

$$\omega_{k.}=\sum_{m=1}^{M}\omega_{km}.$$

If Conjecture 1 is not invoked, the same quantity may be viewed as a first-order empirical variance estimator, obtained by estimating the individual variance component and ignoring the covariance component.

Proof. Let $\mu_{[m]}$ denote the mathematical expectation of an observation whose judgmental rank is m. Then, for a non-empty rank class,

$$E(\hat{\mu}_{[m]}\mid n_m>0) \quad = \quad \mu_{[m]}.$$

Therefore,

$$\begin{aligned}E(\hat{\mu}_{[m]}) &= \mu_{[m]}P(n_m>0)\\ &= \mu_{[m]}\left(1-\left(1-\frac{1}{M}\right)^{n^2}\right).\end{aligned}$$

Consequently,

$$\begin{aligned}E\left(\hat{\mu}_{\mathrm{IJPS}}\right) &= \frac{1}{M}\sum_{m=1}^{M}E(\hat{\mu}_{[m]})\\ &= \left(1-\left(1-\frac{1}{M}\right)^{n^2}\right)\frac{1}{M}\sum_{m=1}^{M}\mu_{[m]}\\ &= \left(1-\left(1-\frac{1}{M}\right)^{n^2}\right)\mu.\end{aligned}$$

Since

$$\left(1-\frac{1}{M}\right)^{n^2}\longrightarrow 0\qquad \text{as } n\to\infty,$$

the bias term vanishes, and hence $\hat{\mu}_{\mathrm{IJPS}}$ is asymptotically unbiased.

For the variance, using (1) and (2), we can write

$$\hat{\mu}_{\mathrm{IJPS}}=\frac{1}{M}\sum_{k=1}^{n}\omega_{k.}X_k.$$

Therefore,

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_{\mathrm{IJPS}}) &= \frac{1}{M^2}\mathrm{Var}\left(\sum_{k=1}^{n}\omega_{k.}X_k\right)\\ &= \frac{1}{M^2}\left[\sum_{k=1}^{n}\mathrm{Var}(\omega_{k.}X_k)+2\sum_{1\le k<k'\le n}\mathrm{Cov}(\omega_{k.}X_k,\omega_{k'.}X_{k'})\right].\end{aligned} \tag{4}$$

The first term in (4) is the individual variance component, whereas the second term is the covariance component. Under Conjecture 1, the covariance component is non-positive. Thus,

$$\mathrm{Var}(\hat{\mu}_{\mathrm{IJPS}}) \leq \frac{1}{M^2}\sum_{k=1}^{n}\mathrm{Var}(\omega_{k.}X_k).$$

By exchangeability,

$$\mathrm{Var}(\hat{\mu}_{\mathrm{IJPS}}) \leq \frac{n}{M^2}\mathrm{Var}(\omega_{k.}X_k).$$

Replacing $\mathrm{Var}(\omega_{k.}X_k)$ with its sample analogue gives

$$\hat{V} = \frac{n}{M^2(n-1)}\sum_{k=1}^{n}\left(\omega_{k.}X_k - \frac{1}{n}\sum_{k=1}^{n}\omega_{k.}X_k\right)^2.$$

If Conjecture 1 is not imposed, this expression still estimates the first-order individual variance component in (4). This completes the proof. □

Using almost the same approach proposed for the IJPS, one can obtain an alternative estimator based on an IRSS of size $n = cM$ explained in Table 2. More specifically, if we denote $W_{ijm} = \sum_{l=1}^{Mc} I(R_{ijl} = m)$, where R_{ijl} is the rank of $X_{(i)ij}$ in the l^{th} set, then the mentioned estimator is defined as

$$\hat{\mu}_{\mathrm{IRSS}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\mu}_{(m)}, \tag{5}$$

where

$$\hat{\mu}_{(m)} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{c}X_{(i)ij}W_{ijm}}{\sum_{i=1}^{M}\sum_{j=1}^{c}W_{ijm}}.$$

4 Further numerical studies

4.1 Examining various distributions

We conducted a simulation study under various distributions and defined efficiency as the ratio of the variance of the SRS estimator to the mean squared error (MSE) of the corresponding estimator. Specifically, let $\hat{\mu}$ be an estimator of μ obtained from one of the ranked sampling designs described above. We compute the relative efficiency as

$$RE = \frac{\frac{\sigma^2}{n}}{MSE(\hat{\mu})}, \tag{6}$$

where the numerator represents the variance of the sample mean under SRS of size n, and MSE denotes the Monte Carlo mean squared error. We considered a wide range of distributions, namely N(10,2), Uniform(0,1), Exp(1), Weibull(2), Beta(7,4), Gamma(2,1), Chisq(2), and Poisson(8), with $n = 12, 18, 24$ and $M = 3, 6$. The results are reported in Table 6. They show consistent efficiency gains for IJPS and IRSS compared with JPS and RSS, respectively, across all scenarios. Moreover, the efficiency gains tend to increase with respect to n and M in most cases, although the effect is more pronounced with respect to M. The improvements are also more substantial for symmetric distributions.

Results are shown in Table 6. They clearly show that there are efficiency gains for IJPS and IRSS over JPS and RSS, respectively, in all cases. Consequently, the results show that the gains in efficiency

increase as M and n increase; however, the increase in efficiency is more evident for M. Finally, the improvements are more significant for symmetric distributions.

To assess $\hat{V}$ in (3), we constructed 90% and 95% confidence intervals based on the tails of the standard normal distribution and the estimator $\hat{\mu}_{\text{IJPS}}$ in (1):

$$CI_{90} = \left(\hat{\mu}_{\text{IJPS}} - 1.64\sqrt{\hat{V}_{upper}},\quad \hat{\mu}_{\text{IJPS}} + 1.64\sqrt{\hat{V}_{upper}}\right)$$
$$CI_{95} = \left(\hat{\mu}_{\text{IJPS}} - 1.96\sqrt{\hat{V}_{upper}},\quad \hat{\mu}_{\text{IJPS}} + 1.96\sqrt{\hat{V}_{upper}}\right)$$

for $M = 3,6$ and $n = 24$. The coverage rates of these confidence intervals for different distributions are reported in Table 6, where C_{90} and C_{95} denote the coverage probabilities (CPs) of CI_{90} and CI_{95}, respectively. These CPs were computed as the proportion of intervals containing the population parameter μ over 2000 iterations. In most cases, the CP exceeded the nominal level, which is expected when an upper bound is used for the variance. At the same time, the upper bound has two advantages: it applies to all considered distributions and nearly guarantees the nominal coverage probability. Moreover, larger sample sizes led to more reliable coverage.

4.2 Proposing another improving approach

The approach proposed in Section 3 improves both JPS and RSS designs regardless of whether the parent distribution is symmetric or asymmetric. As stated previously and illustrated in Table 3 for the algorithm that improves RSS, the auxiliary units from the $n-1$ remaining sets whose judgment ranks coincide with those of a given fully measured unit are eliminated, and then the unbiased efficient estimator (5) is constructed. One may ask what happens if these auxiliary units are not eliminated. In that case we have $1 \le R_{ijl} \le M+1$, and the estimator (5) reduces to

$$\hat{\mu} = \frac{1}{M+1}\sum_{m=1}^{M+1}\hat{\mu}_{(m)}. \tag{7}$$

Similarly, for JPS, if the $n-1$ remaining sets contain their matching fully measured units, then the R_{ki}'s in Table 1 also lie in the interval $[1, M+1]$, and estimator (7) with $\hat{\mu}_{[m]}$ instead of $\hat{\mu}_{(m)}$ applies to the JPS design as well. We refer to the method described in Section 3 as the "First Improving Approach" and to the method presented here as the "Second Improving Approach."

4.3 Examining some more ranked sampling designs

In addition to this question, we are also interested in examining other ranked sampling designs, such as median RSS (MedRSS) and extreme RSS (ERSS). There are two main reasons for considering these variants of RSS: their simplicity and reduced ranking error (particularly in ERSS), as well as their strong performance under symmetric populations.

MedRSS was introduced by Muttlak (1997) as a simplified version of RSS. The following algorithm can be used to obtain a c-cycle MedRSS with set size M. If M is odd, observations are obtained by measuring the judgment medians from each SRS. For even set sizes, suppose that M simple random samples, each of size M, are drawn. Then, the units with the largest judgment ranks are measured from the first half of the samples, while the units with the smallest judgment ranks are measured from the second half. By repeating this procedure c times, the desired sample size $n = cM$ is obtained.

More specifically, the gathered observations are of the form

$$\left\{X_{\left(\frac{M+1}{2}\right)ij},\ i = 1,\cdots,M,\ j = 1,\cdots,c\right\},$$

Table 6: The values of the RE, the relative efficiency of the respective design to the SRS design, and the CP under different distributions (the simulation is carried out based on 2000 Monte Carlo replications).

		CP	RE				CP	RE			
n	M	f	JPS	IJPS	RSS	IRSS	f	JPS	IJPS	RSS	IRSS
12	3	$N(10,2)$	1.51	2.33	1.95	2.35	$Beta(7,4)$	1.42	2.25	1.97	2.36
	6		1.57	3.77	3.12	4.25		1.54	3.80	3.18	4.38
18	3		1.57	2.36	1.83	2.22		1.72	2.35	2.00	2.44
	6	C_{90}, C_{95}	1.83	4.1	3.08	4.36	C_{90}, C_{95}	1.89	4.21	3.2	4.54
24	3	0.92, 0.94	1.8	2.48	2.10	2.56	0.91, 0.93	1.70	2.55	1.99	2.42
	6	0.97, 0.98	2.22	4.33	3.30	4.79	0.97, 0.98	2.27	4.14	3.08	4.33
12	3	$Uniform(0,1)$	1.46	2.48	1.95	2.36	$Gamma(2,1)$	1.41	2.10	1.79	2.11
	6		1.61	4.15	3.39	4.53		1.48	3.26	2.89	4.06
18	3		1.64	2.42	1.98	2.47		1.49	2.09	1.71	2.08
	6	C_{90}, C_{95}	1.88	4.95	3.29	4.41	C_{90}, C_{95}	1.69	3.48	2.90	4.14
24	3	0.97, 0.98	1.84	2.68	2.02	2.55	0.96, 0.98	1.51	2.14	1.87	2.26
	6	0.96, 0.98	2.22	4.58	3.29	4.70	0.98, 0.99	1.98	3.59	2.83	3.93
12	3	$Exp(1)$	1.32	1.82	1.74	2.01	$Chisq(2)$	1.23	1.73	1.48	1.73
	6		1.41	2.91	2.57	3.59		1.22	2.52	2.13	2.92
18	3		1.36	1.89	1.70	2.01		1.56	1.84	1.52	1.77
	6	C_{90}, C_{95}	1.55	3.05	2.38	3.45	C_{90}, C_{95}	1.44	2.87	2.25	3.15
24	3	0.95, 0.97	1.49	1.92	1.62	1.93	0.94, 0.97	1.3	1.85	1.39	1.64
	6	0.98, 0.99	1.78	3.33	2.71	3.73	0.97, 0.99	1.74	2.77	2.41	3.37
12	3	$Weibull(2,1)$	1.38	2.32	1.88	2.25	$Poisson(8)$	1.32	1.94	1.7	1.97
	6		1.56	3.72	3.03	4.13		1.29	2.86	2.47	3.34
18	3		1.57	2.26	1.90	2.30		1.41	1.87	1.56	1.86
	6	C_{90}, C_{95}	1.78	4.43	3.07	4.25	C_{90}, C_{95}	1.68	3.20	2.57	3.53
24	3	0.96, 0.98	1.76	2.56	1.98	2.44	0.96, 0.97	1.51	1.96	1.63	1.93
	6	0.99, 0.99	2.04	4.02	3.02	4.43	0.97, 0.98	1.81	3.33	2.58	3.59
12	3	$\omega N(100,10)$	1.44	2.18	1.80	2.15	$\omega N(100,10)$	1.35	1.90	1.76	2.06
	6	$+(1-\omega)N(150,10)$,	1.59	3.98	3.47	4.51	$+(1-\omega)N(150,10)$,	1.43	2.89	3.03	4.17
18	3	$\omega \sim Ber(0.5)$	1.65	2.40	1.87	2.26	$\omega \sim Ber(0.2)$	1.56	2.03	1.84	2.17
	6	C_{90}, C_{95}	2.04	4.42	3.25	4.32	C_{90}, C_{95}	1.68	3.23	2.70	3.78
24	3	0.91, 0.93	1.68	2.44	2.00	2.41	0.92, 0.94	1.56	2.05	1.71	2.08
	6	0.97, 0.98	2.24	4.46	3.02	4.14	0.97, 0.98	1.87	3.28	2.82	3.91

when M is odd, and

$$\left\{X_{(\frac{M}{2})ij},\ i=1,\cdots,\frac{M}{2},\ j=1,\cdots,c\right\}\cup\left\{X_{(\frac{M}{2}+1)ij},\ i=\frac{M}{2}+1,\cdots,M,\ j=1,\cdots,c\right\},$$

when M is even. As mentioned earlier, if the required conditions of running the RSS are provided in a population, it would be logical to utilize it rather than SRS. But, if the set size M is not small, the ranking error would be non-negligible and this implies enhancing the variances of the estimators obtained based on the RSS. But then, inspecting the extreme values is not so difficult even in large sets. Therefore, Samawi et al. (1996) proposed ERSS which inspects only the extreme observations. Actually, in this sampling design, we select M random samples of size M and rank each sample separately with respect to a variable of interest by visual inspection. If the sample size M is even, we select from $\frac{M}{2}$ samples the smallest unit and from the $\frac{M}{2}$ remaining samples the largest unit for full measurement. If the sample size is odd, select from $\frac{M-1}{2}$ samples the smallest unit, from the other $\frac{M-1}{2}$ the largest unit and from one

sample the median of the sample for actual measurement. This cycle may be repeated c times in order to get $n = cM$ units. In summary, for an even M we have

$$\left\{X_{(1)ij},\ i=1,\cdots,\frac{M}{2},\ j=1,\cdots,c\right\}\cup\left\{X_{(M)ij},\ i=\frac{M}{2}+1,\cdots,M,\ j=1,\cdots,c\right\},$$

and for an odd M we have

$$\begin{aligned}&\left\{X_{(1)ij},\ i=1,\cdots,\frac{M-1}{2},\ j=1,\cdots,c\right\}\\&\cup\left\{X_{(M)ij},\ i=\frac{M+1}{2}+1,\cdots,M-1,\ j=1,\cdots,c\right\}\\&\cup\left\{X_{(\frac{M+1}{2})Mj},\ j=1,\cdots,c\right\}.\end{aligned}$$

Therefore in the next simulation scenario, we consider the population mean estimators on the basis of four ranked sampling schemes including RSS, JPS, MedRSS and ERSS of size $n = cM$, with $(c,M) = (3,4),(4,5),(4,15),(5,20),(5,30),(6,30)$, under the two approaches mentioned above as well as their existing conventional versions. In order to be closer, the standard normal distribution as well as the Gamma$(2,1)$ are selected as the parent distributions in order to involve both of symmetric and asymmetric populations. Two criteria are employed for the comparison purpose: the bias and the RE given by (6). Figures 1 and 2 depict the result. The following points are observed:

- The approaches of using full ranking information improve the ordinary methods of ranking. However, the first approach has a better performance than the second one according to the all criteria.

- Sampling design performance is heavily influenced by the parent distribution. For instance, the improved ERSS outperforms its competitors under normal distributions (representing the class of symmetric distributions). However, it becomes one of the least efficient designs, along with MedRSS, when the population follows an asymmetric distribution, such as the gamma distribution.

- Figures 1 and 2 confirm the asymptotic unbiasedness of the estimators in (1) and (5), which are obtained based on the improved JPS and improved JPS, respectively, under the first approach. However, the improved versions of ERSS and MedRSS produce unbiased estimators for μ only for populations with symmetric densities. This property was proved by Samawi et al. (1996) for the ordinary version of ERSS as well.

- The estimators obtained based on the second approach are clearly biased but still more efficient than their conventional counterparts.

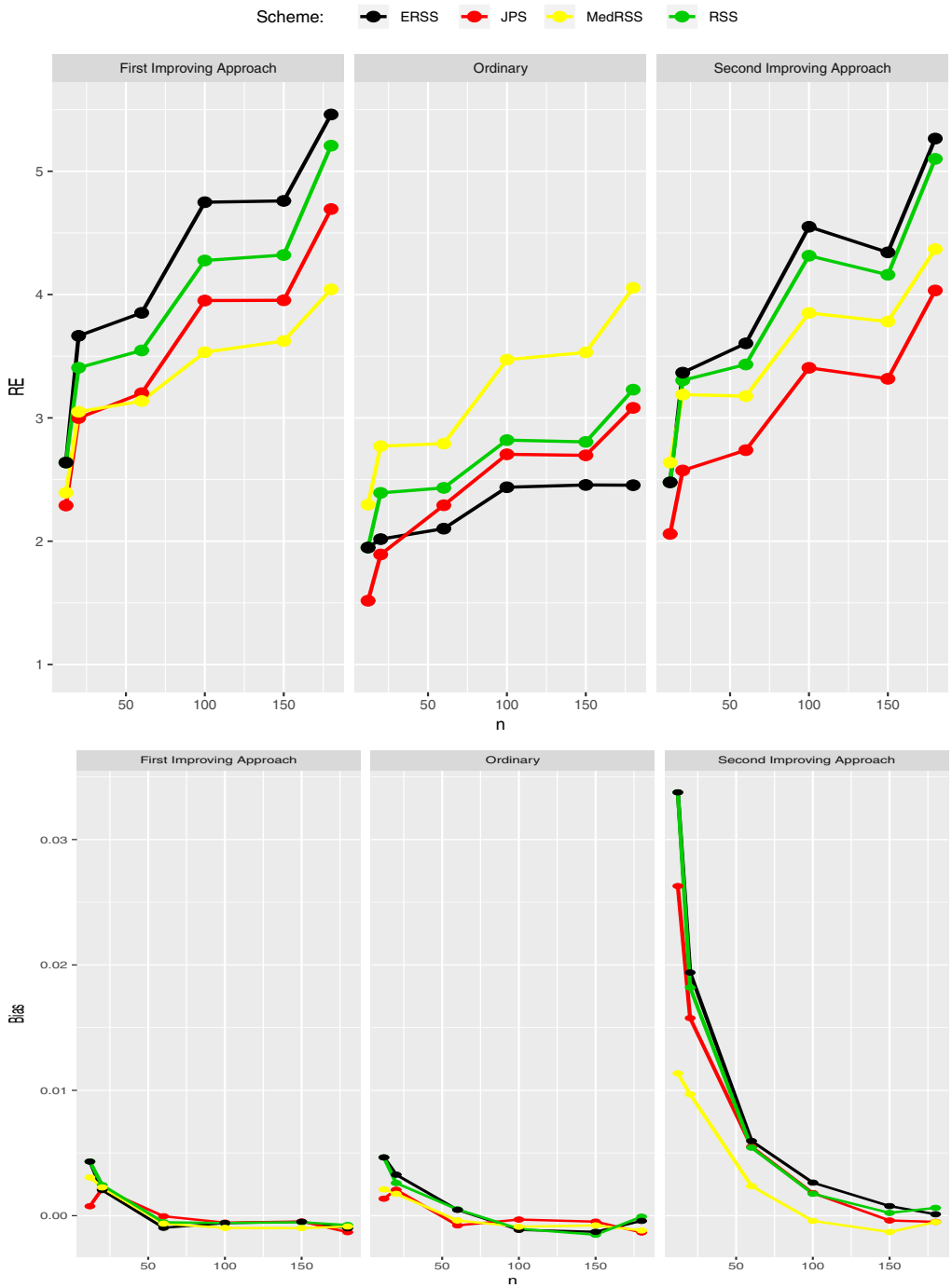


Figure 1: Comparison of the estimators obtained based on various sampling designs proposed via RE and bias criteria under the standard normal distribution.

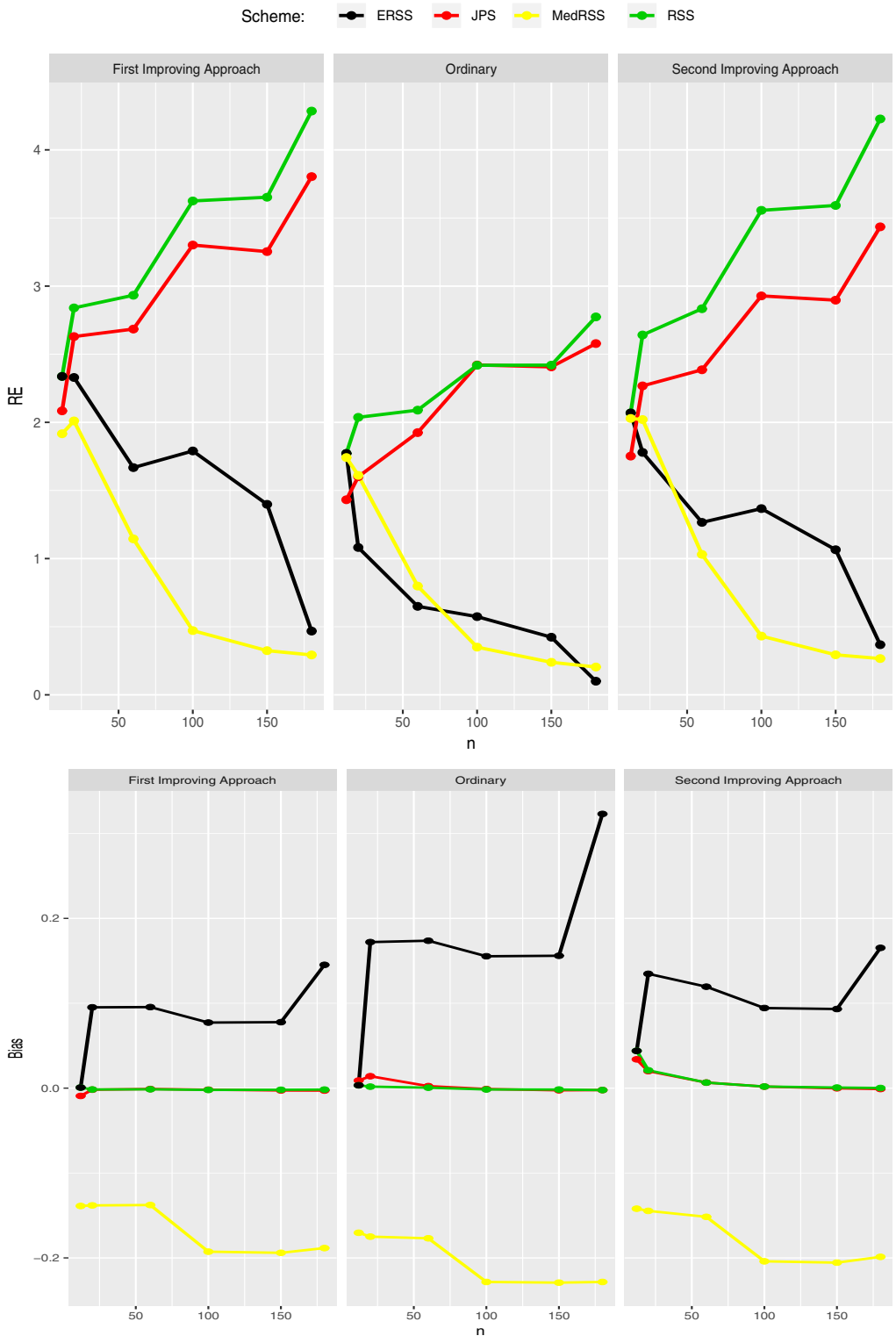


Figure 2: Comparison of the estimators obtained based on various sampling designs proposed via RE and bias criteria under the gamma distribution.

5 Real data analysis

Two real data sets are considered to illustrate the results in the previous sections. To check the chosen sample's sensitivity and obtain the efficiencies of the mean estimators derived based on various designs, the Monte Carlo technique with 2000 replications is employed. Moreover, the judgment ranking process is performed using a covariate variable highly correlated with the variable of interest. Then, the relative efficiencies are computed from (6). It is to be noted that here the variance that appeared in the numerator of (6) is substituted by the *MSE* of the estimator obtained based on the SRS.

5.1 Belgium municipalities data

We implemented the new method on Belgium Municipalities Data. The data set is available in the `R` package `sampling` under the name `belgianmunicipalities`. It contains different information about the Belgian population between 2001 and 2004 by municipalities. The data set contains 589 rows corresponding to each municipality of Belgium. We just considered two variables, the variable of interest x, which represents the taxable income in euros in 2001, and the auxiliary variable z, which represents the total population on July 1, 2004. As it is clearly observed from Table 7, the improved versions of the JPS and the RSS designs yield more efficient estimators. This fact confirms the results obtained in the preceding section.

Table 7: Belgium municipalities data. The numbers indicate efficiency of the respective design relative to SRS design.

n	M	JPS	IJPS	RSS	IRSS	n	M	JPS	IJPS	RSS	IRSS
12	3	1.09	1.24	1.25	1.36	10	5	1.02	1.68	1.36	1.72
	6	1.11	1.63	1.18	1.50		10	1.05	2.58	1.91	2.83
18	3	1.07	1.22	1.20	1.31	20	5	1.00	1.29	1.51	1.84
	6	1.08	1.59	1.60	2.01		10	1.05	2.14	1.81	2.58
24	3	1.05	1.24	1.20	1.30	50	5	1.33	1.37	1.31	1.62
	6	1.12	1.48	1.53	2.02		10	1.25	1.80	1.71	2.45
48	3	1.16	1.24	1.23	1.38	90	5	1.29	1.39	1.35	1.66
	6	1.32	1.56	1.47	1.86		10	1.49	1.79	1.45	1.96
24	4	1.16	1.48	1.30	1.55	42	7	1.10	1.50	1.52	1.92
	8	1.12	1.77	1.35	1.90		14	1.06	2.02	1.60	2.34
40	4	1.26	1.37	1.32	1.57	70	7	1.15	1.52	1.40	1.78
	8	1.20	1.64	1.49	1.96		14	1.36	2.00	1.67	2.33
56	4	1.14	1.39	1.11	1.36	84	7	1.33	1.56	1.51	1.90
	8	1.12	1.52	1.48	2.00		14	1.39	1.99	1.63	2.25
72	4	1.23	1.37	1.22	1.47	112	7	1.38	1.59	1.49	1.88
	8	1.25	1.55	1.38	1.82		14	1.43	1.90	1.67	2.16

5.2 Switzerland provinces data

For this part we used Standardized fertility measures and socio-economic indicators for 47 French-speaking provinces of Switzerland at about 1888. The data set is available in the `R` package `stats` under the name `swiss`. We selected two of them as the main and one as the auxiliary variables (see Figure 3):

x_1: Examination; draftees receiving highest mark on army examination,

x_2: Agriculture; of males involved in agriculture as an occupation,

z: Education; education beyond primary school for draftees.

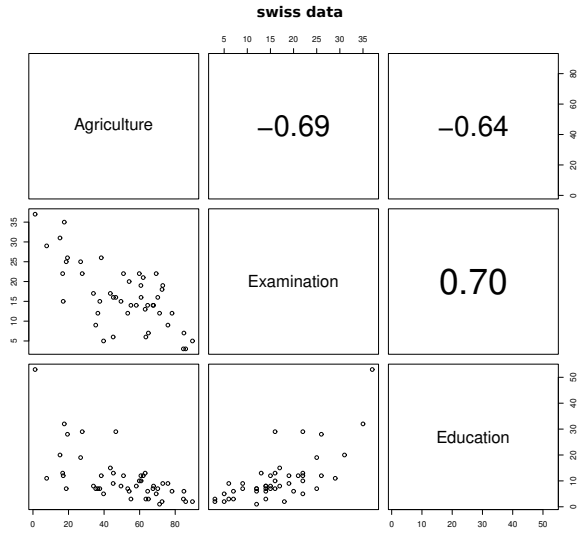


Figure 3: Relation between Switzerland provinces data.

Analogously to the simulation section, the analysis of the Switzerland provinces data in Table 8 suggests that IRSS and IJPS outperform the traditional RSS and JPS designs, respectively. Moreover, from Table 8 and Figure 3, it can be generally concluded that the improved estimators become more efficient as the judgment ranking becomes more precise (as we expected before). Also as an interesting point, as JPS has some bias and we used MSE to measure efficiency, in some cases JPS could not be as efficient as SRS. But we can see that the improved version was always more efficient than SRS.

Overall, the results in Tables 7 and 8 show that using more accurate ranking information can lead to noticeable gains in estimation efficiency in practical settings. For the Belgium municipalities data, the improved JPS and RSS methods consistently outperform their ordinary counterparts, with the gain becoming more highlighted as the set size $\boldsymbol{M}$ increases. For the Switzerland provinces data, the improved methods again provide more efficient estimates in most cases, especially when the auxiliary variable is strongly associated with the variable of interest. These findings indicate that the proposed improved designs are not only theoretically justified, but also practically useful for real data analysis, since they can generate more precise estimation without adding significant cost or complexity.

6 Conclusion

We have proposed simple and practical improvements to JPS and RSS that make use of the complete ranking information within each selected set, which is either already available or can be obtained at virtually no additional cost. The proposed methods are motivated by the Rao–Blackwellization principle and effectively exploit all available rank information, leading to more reliable rank judgments and, consequently, more efficient parameter estimation. They retain the simplicity of the original sampling designs while providing substantial gains in estimation efficiency. In particular, IJPS and IRSS offer a favorable

Table 8: Switzerland provinces data. The numbers indicate efficiency of the respective design relative to SRS design.

n	M	JPS	IJPS	RSS	IRSS	n	M	JPS	IJPS	RSS	IRSS
z = Education, x = Agriculture											
4	2	0.95	1.08	1.24	1.25	4		0.94	1.23	1.32	1.33
6		0.94	1.12	1.20	1.22			0.98	1.33	1.35	1.39
8		0.93	1.13	1.18	1.19			1.05	1.40	1.37	1.39
10		1.01	1.16	1.13	1.13			1.03	1.35	1.30	1.33
6	3	0.89	1.16	1.23	1.27	5		0.95	1.29	1.33	1.39
9		0.92	1.18	1.27	1.31			1.02	1.45	1.40	1.47
12		0.95	1.24	1.22	1.25			1.03	1.47	1.37	1.44
15		1.02	1.26	1.18	1.24			1.05	1.44	1.52	1.60
z = Education, x = Examination											
4	2	0.94	1.14	1.23	1.25	4		0.97	1.28	1.51	1.57
6		0.97	1.17	1.24	1.26			1.02	1.39	1.38	1.48
8		0.97	1.16	1.16	1.19			1.05	1.42	1.35	1.46
10		1.01	1.22	1.16	1.20			1.12	1.43	1.43	1.55
6	3	0.90	1.20	1.29	1.33	5		0.92	1.36	1.37	1.45
9		0.91	1.19	1.21	1.24			1.02	1.45	1.52	1.63
12		0.93	1.29	1.29	1.34			1.01	1.44	1.51	1.59
15		1.03	1.30	1.34	1.40			1.04	1.43	1.49	1.57

balance between improved precision and computational simplicity without introducing significant additional complexity. For symmetric or approximately symmetric populations, we recommend IERSS, which fully exploits the ranking information available in ERSS and generally provides the greatest efficiency gains.

Acknowledgements

The authors thank the reviewers for their valuable comments and suggestions, which have improved the quality of the paper.

References

Bhushan, S. and Kumar, A. (2024). Novel logarithmic imputation methods under ranked set sampling. *Journal of Statistics and Management Systems*, **27**, 235–257.

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

Dey, S., Salehi, M. and Ahmadi, J. (2017). Rayleigh distribution revisited via ranked set sampling. *Metron*, **75**, 69–85.

Frey, J. and Ozturk, O. (2011). Constrained estimation using judgment post-stratification. *Annals of the Institute of Statistical Mathematics*, **63**, 769–789.

Kaur, A., Patil, G.P. and Sinha, A. K. (1995). Ranked set sampling: An annotated bibliography. *Environmental and Ecological Statistics*, **2**, 25–54.

MacEachern, S.N., Stasny, E.A. and Wolfe, D. A. (2004). Judgment post-stratification with imprecise rankings. *Biometrics*, **60**, 207–215.

Matthews, M. and Wolfe, D. A. (2017). Unified ranked set sampling. *Statistics and Probability Letters*, **122**, 173–178.

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

Muttlak, H. A. (1997). Median ranked set sampling. *Journal of Applied Statistical Science*, **6**, 577–586.

Panahbehagh, B., Bruggemann, R., Parvardeh, A., Salehi, M. and Sabzalian, M. R. (2018). An unbalanced ranked-set sampling method to get more than one sample from each set. *Journal of Survey Statistics and Methodology*, **6**, 285–305.

Panahbehagh, B. and Bruggemann, R. (2021). Introduction into sampling theory, applying partial order concepts. In: Bruggemann, R., Carlsen, L., Beycan, T., Suter, C. and Maggino, F. (eds), *Measuring and Understanding Complex Phenomena*. Springer, Cham.

Panahbehagh, B., Bruggemann, R. and Salehi, M. (2021). Sampling of multiple variables based on partially ordered set theory. *Journal of the Iranian Statistical Society*, **20**, 307–331.

Samawi, H.M., Ahmed, M.S. and Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, **38**, 577–586.

Ozturk, O. (2014) Statistical inference for population quantiles and variance in judgment post-stratified samples. *Computational Statistics and Data Analysis*, **77**, 188–205.

Ozturk, O. and Bayramoglu Kavlak, K. K. (2020). Statistical inference using stratified judgment post-stratified samples from finite populations. *Environmental and Ecological Statistics*, **27**, 73–94.

Ozturk, O. and Kravchuk, O. (2021). Judgment post-stratified assessment combining ranking information from multiple sources, with a field phenotyping example. *Journal of Agricultural, Biological and Environmental Statistics*, **26**, 329–348.

Ozturk, O. (2012). Combining ranking information in judgment post-stratified and ranked set sampling designs. *Environmental and Ecological Statistics*, **19**, 73–93.

Ozturk, O. (2016). Statistical inference based on judgment post-stratified samples in finite population. *Survey Methodology*, **42**, 239–262.

Ozturk, O. (2023). Ranked set sampling: A review of its development and applications. *Journal of Applied Statistics*, **50**, 2785–2814.

Salehi, M., Ahmadi, J. and Dey, S. (2016). Comparison of two sampling schemes for generating record-breaking data from the proportional hazard rate models. *Communications in Statistics - Theory and Methods*, **45**, 3721–3733.

Stokes, S. and Sager, T. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *Journal of the American Statistical Association*, **83**, 374–381.

Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, **20**, 1–31.

Yanagawa, T. and Shirahata, S. (1976). Ranked set sampling theory with selective probability matrix. *Australian Journal of Statistics*, **18**, 45–52.

Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 71–84, Research Paper
https://smps.um.ac.ir, ISSN: 3115-9370

Wiener index and saturated nodes of random exponential recursive ternary trees

Mehri Javanian*

Department of Statistics, Faculty of Sciences, University of Zanjan, Iran

Email(s): javanian@znu.ac.ir

Abstract. An exponential recursive ternary tree (ERTT) is defined by the property that each node has three possible positions, referred to as external nodes, to which a child may be attached. At each growth step, every external node independently becomes a leaf with probability p, or remains external with probability $1-p$. In this note, we study two quantities associated with ERTTs: the external Wiener index, defined as the sum of distances over all pairs of external nodes, and the number of saturated nodes, i.e., nodes that have three children. For these trees, we derive the expected value and the second moment of both the number of saturated nodes and the Wiener index. Subsequently, applying the martingale convergence theorem, we establish almost sure convergence and convergence in quadratic mean for a suitably scaled external Wiener index, which reflects the typical distance between two randomly chosen external nodes. Finally, using the contraction method, we prove convergence in distribution for a scaled version of the number of saturated nodes.

Keywords: Contraction method; L^2-norm; Martingale convergence theorem; Wasserstein metric; Wiener index.

Mathematics Subject Classification (2020): 05C05, 60F25, 60F17.

1 Introduction and setting

A *random recursive tree* of size n is built sequentially: starting with a root, each new node attaches uniformly at random to one of the existing nodes Smythe and Mahmoud (1995). For only a few of variations of recursive trees, see Javanian and Vahidi-Asl (2003), Javanian and Vahidi-Asl (2006), Javanian and Vahidi-Asl (2012), Javanian (2013), Javanian (2017) and Zhang (2018). These trees have been widely studied and applied in various fields Chan et al. (2003), Najock and Heyde (1982), Moon (1974).

Since only one node is added at each step in the growth of a recursive tree, then these slow-growing models cannot be suitable for fast-growing phenomena such as the growing the Twitter and the spreading of infectious diseases. In Mahmoud (2022), it has been proposed a fast-growing recursive tree model, i.e. the *exponential recursive tree* (ERT), where every node independently attracts a new node with

*Corresponding author
Received: 30 May 2026 / Revised: 21 June 2026 / Accepted: 22 July 2026

https://doi.org/10.22067/smps.2026.99021.1060

The Author(s). This is an open access article under CC BY 4.0.

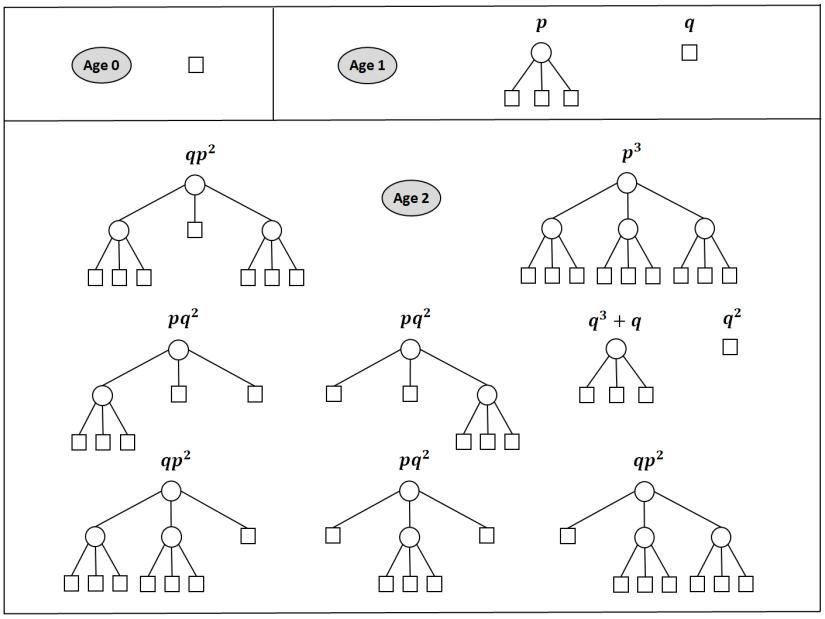


Figure 1: All exponential recursive ternary trees of index p and ages 0, 1 and 2. The probability of getting each tree appears above the root of that tree.

probability p, or fails to do at each step in the growth of the tree. The *size* (the number of nodes), the *leaves* (the nodes with no children), the *depth* of a node v (the distance, i.e. the number of edges, of node v from tree's root node) and the *total path length* (the sum of depth of all nodes) have been studied for ERTs by Aguech et al. (2021).

The ERT generalizes the recursive tree by allowing multiple nodes to grow simultaneously.

We use the term ERT following Mahmoud (2022). These trees are *recursive* in the sense that their construction is recursively defined: a tree of age n is built from independent copies of trees of age $n-1$ via a self-similar branching rule. This recursive structure is central to the distributional recurrences we derive in Sections 2–4. While the growth mechanism is parallel rather than sequential, the recursive nature of the tree is preserved.

In Ghasemi et al. (2023), an *exponential recursive k-ary tree* has been defined as an ERT in which every node has at most k children. The nodes that already exist in the tree are called *internal*. In order to identify the locations of possible insertions of a new node, an exponential recursive k-ary tree is extended by having $k-d$ *external* nodes joined as children of a node with d children. Internal and external nodes are shown by circles and squares, respectively (see, e.g., the trees in Figure 1).

An *exponential recursive k-ary tree of index p and age n* grows as follows: Initially (at age 0), there is an external node. For $n \geq 1$, at the nth step (at age n), every external node independently is changed into a leaf with probability $p \in (0,1)$, or not to be changed with probability $q := 1-p$. The parameter p is called the *index* of the ERT. The *age* of an ERT is the number of steps it takes to grow the tree. Figure 1 shows all exponential recursive 3-ary trees of ages 0, 1 and 2.

We assume $0 < p < 1$ throughout, which ensures non-degenerate stochastic growth. The boundary cases $p = 0$ (no growth) and $p = 1$ (deterministic complete tree) are trivial and can be obtained as limiting cases of our results.

The *Wiener index*, a distance-based topological index, is defined as the sum of distances between all nodes in a graph. For recursive trees, Neininger (2002) has obtained the expectations, the variances and

a bivariate limit law for the Wiener index and the total path length. It also has been obtained a strong limit theorem for the Wiener index of recursive trees in Grübel and Michailow (2015).

The *external Wiener index*, the sum of distances between all pairs of external nodes, provides a natural measure of the spread or compactness of the tree's boundary. This is particularly relevant for:

- Network analysis: In social networks modeled by exponential recursive trees, external nodes represent potential future connections or unoccupied slots. The external Wiener index captures the expected distance between these potential connection points, which is crucial for understanding information propagation, epidemic spreading, and network resilience.
- Population genetics and phylogenetics: External nodes in these trees correspond to extant species or individuals. The external Wiener index measures the genetic or phylogenetic distance among current population members, providing insights into diversity and evolutionary history.
- Algorithmic applications: In data structures such as digital search trees or trie structures, external nodes represent empty positions. The external Wiener index quantifies search or insertion costs, making it practically relevant for performance analysis.

By *saturated nodes*, we mean the nodes with 3 children in an exponential recursive ternary tree. A saturated node represents a full branching point. These nodes are of significant interest because:

- They indicate complete local structures, which are critical for understanding the tree's capacity and growth bottlenecks.
- In biological systems, saturated nodes may represent speciation events where all possible descendant lineages are present.
- In network models, saturated nodes correspond to fully connected local hubs, which affect robustness and clustering coefficients.

The limiting distribution of depth and the expectation and variance of *external path length* (the sum of depth of all external nodes) in recursive k-ary trees have been obtained in Javanian and Vahidi-Asl (2006), Javanian and Vahidi-Asl (2012) and Javanian (2014), respectively. The size, leaves, external path length and depth of a randomly chosen external node in exponential recursive k-ary tree have been investigated; see Ghasemi et al. (2023). The asymptotic behaviour of *profile* (number of nodes of the same depth) of exponential binary trees has been studied by Feng and Mahmoud (2018).

In this paper, for the sake of simplicity, we consider the exponential recursive k-ary tree for which $k=3$. The resulting tree is called *exponential recursive ternary tree* (ERTT). Our interest here is to study the external Wiener index W_n and number of saturated nodes Y_n in an ERTT of age n. We give the first two moments of W_n and Y_n. Then we show convergence in distribution and in L^2 of $\dfrac{W_n}{n(2p+1)^{2n}}$ that is a martingale with finite second moment. By the contraction method, we also characterize the limiting distribution of $\dfrac{Y_n}{(2p+1)^n}$ as the unique fixed point of a distributional equation.

We use the recursive structure of the ERTTs under consideration in order to setup some recurrences for Wiener index and the number of saturated nodes. After one step of the evolution of an ERTT, the initial external node either succeeds to recruit a leaf (event $\mathscr{R}$) or not. We denote the indicator function of the event $\mathscr{R}$ by $\mathbb{I}$. If the event $\mathscr{R}$ does not occur, then after $n-1$ steps, the initial external node of the tree fathers an ERTT $\tilde{T}_{n-1}$. Alternatively, If $\mathscr{R}$ occurs, the extended ERTT has an internal node with 3 external nodes. By time n, these 3 external nodes produce 3 independent ERTTs T'_{n-1}, T''_{n-1} and T'''_{n-1}. Now, we hook T'_{n-1}, T''_{n-1} and T'''_{n-1} to construct an ERTT of index p and age n. Let $(\tilde{X}_{n-1},\tilde{Z}_{n-1})$, (X'_{n-1},Z'_{n-1}), (X''_{n-1},Z''_{n-1}) and (X'''_{n-1},Z'''_{n-1}), be the pairs of the number of external nodes and the external

path length in $\tilde{T}_{n-1}$, T'_{n-1}, T''_{n-1} and T'''_{n-1}, respectively. Then for an ERTT of index p at age n with the number of saturated nodes Y_n and the Wiener index W_n, we have

$$X_n \stackrel{d}{=} (1-\mathbb{I})\tilde{X}_{n-1} + \mathbb{I}\Big(X'_{n-1} + X''_{n-1} + X'''_{n-1}\Big), \tag{1}$$

$$Z_n \stackrel{d}{=} (1-\mathbb{I})\tilde{Z}_{n-1} + \mathbb{I}\Big(Z'_{n-1} + Z''_{n-1} + Z'''_{n-1} + X'_{n-1} + X''_{n-1} + X'''_{n-1}\Big), \tag{2}$$

where the symbol $\stackrel{d}{=}$ denotes equality in distribution; the pairs $(\tilde{X}_{n-1}, \tilde{Z}_{n-1})$, (X'_{n-1}, Z'_{n-1}), (X''_{n-1}, Z''_{n-1}) and (X'''_{n-1}, Z'''_{n-1}) are independent copies of (X_{n-1}, Z_{n-1}) and independent of $\mathbb{I}$. Moreover, the almost sure and L^2 convergence of the scaled X_n and Z_n have been shown in Ghasemi et al. (2023),

$$\frac{X_n}{(2p+1)^n} \xrightarrow{\text{a.s.}} X_*, \qquad \frac{Z_n}{n(2p+1)^n} \xrightarrow{L^2 \ \& \ \text{a.s.}} X_*, \qquad \left(\mathbb{E}[X_*] = 1,\ \mathbb{E}[X_*^2] = \frac{3}{2p+1}\right), \tag{3}$$

as $n \to \infty$, for a limiting variable X_* characterized by inductively-constructed moments.

To setup the recurrences for the number of saturated nodes and Wiener index, let T_n be an ERTT of index p at age n, $(\tilde{Y}_{n-1}, \tilde{W}_{n-1})$, (Y'_{n-1}, W'_{n-1}), (Y''_{n-1}, W''_{n-1}) and (Y'''_{n-1}, W'''_{n-1}) be the pairs of the number of saturated nodes and the Wiener index in $\tilde{T}_{n-1}$, T'_{n-1}, T''_{n-1} and T'''_{n-1}, respectively. Let $\mathbb{J}_{n-1}$, $\mathbb{J}'_{n-1}$, $\mathbb{J}''_{n-1}$ and $\mathbb{J}'''_{n-1}$ be the indicators of the events that, after $n-1$ extra steps, the initial external node succeeds to recruit a child in T_{n-1}, T'_{n-1}, T''_{n-1} and T'''_{n-1}, respectively. The term $\mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1}$ appears because a new saturated node is created at the root exactly when all three subtrees have grown from their respective initial external nodes. The factor $\mathbb{I}$ indicates whether the initial external node successfully recruited a child. When this happens, the saturated nodes in the resulting tree are the sum of saturated nodes in the three subtrees plus one additional saturated node if all three subtrees are non-empty (i.e., all three $\mathbb{J}$'s equal 1). We will include the derivation:

$$Y_n \stackrel{d}{=} (1-\mathbb{I})\tilde{Y}_{n-1} + \mathbb{I}\Big(Y'_{n-1} + Y''_{n-1} + Y'''_{n-1} + \mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1}\Big). \tag{4}$$

Now, we explain the interaction terms such as $X'_{n-1}X''_{n-1}$, $Z'_{n-1}X''_{n-1}$, etc. These arise from distances between external nodes in different subtrees. For example, if an external node in subtree T'_{n-1} and an external node in subtree T''_{n-1} are considered, their distance is:

$$D(i,j) = D_{T'}(i) + 2 + D_{T''}(j).$$

Summing over all pairs gives the cross terms $Z'_{n-1}X''_{n-1} + Z''_{n-1}X'_{n-1} + 2X'_{n-1}X''_{n-1}$. The factor 2 appears in the constant term because the path goes up from the external node in T' to its root (distance $D_{T'}(i)$), then crosses through the common parent node (distance 2), and goes down to the external node in T'' (distance $D_{T''}(j)$). The $Z'X''$ and $Z''X'$ terms are the sums of $D_{T'}(i)$ over all pairs and $D_{T''}(j)$ over all pairs, respectively. Then we obtain

$$\begin{aligned} W_n \stackrel{d}{=} (1-\mathbb{I})\tilde{W}_{n-1} + \mathbb{I}\Big(&W'_{n-1} + W''_{n-1} + W'''_{n-1} + 2X'_{n-1}X''_{n-1} + 2X'_{n-1}X'''_{n-1} + 2X''_{n-1}X'''_{n-1} \\ &+ Z'_{n-1}X''_{n-1} + Z''_{n-1}X'_{n-1} + Z'_{n-1}X'''_{n-1} + Z'''_{n-1}X'_{n-1} + Z''_{n-1}X'''_{n-1} + Z'''_{n-1}X''_{n-1}\Big), \end{aligned} \tag{5}$$

where the triples $(Y'_{n-1}, \mathbb{J}'_{n-1}, W'_{n-1})$, $(Y''_{n-1}, \mathbb{J}''_{n-1}, W''_{n-1})$ and $(Y'''_{n-1}, \mathbb{J}'''_{n-1}, W'''_{n-1})$ are independent copies of $(Y_{n-1}, \mathbb{J}_{n-1}, W_{n-1})$ and independent of $\mathbb{I}$.

In Section 2, we calculate the first two moments of Y_n and W_n. If S_n be the size of be an ERTT of index p at age n, in Section 3 and Section 4, we use the martingale convergence theorem and contraction method to show the following two types of convergence for W_n and Y_n, respectively,

$$\frac{\mathbb{E}[Y_n]}{\mathbb{E}[S_n]} \le \frac{1}{3}, \qquad \frac{Y_n}{(2p+1)^n} \xrightarrow{\text{d}} \hat{Y} \quad \text{and} \qquad \frac{W_n}{n(2p+1)^{2n}} \xrightarrow{L^2 \ \& \ \text{a.s.}} \frac{3p}{2p+1}X_*^2,$$

where the random variable $\hat{Y}$ is the unique distributional fixed-point of a distributional recurrence.

Throughout the paper, we use a systematic notation for the three independent subtrees that arise when the root recruits a child: variables with a single prime (e.g., X'_{n-1}), double prime (e.g., X''_{n-1}), and triple prime (e.g., X'''_{n-1}) represent the three independent copies. Variables with a tilde (e.g., $\tilde{X}_{n-1}$) represent the 'no-recruitment' case. This convention is consistent across all parameters: X for external nodes, Z for external path length, W for Wiener index, Y for saturated nodes, and J for growth indicators. Thus, a reader encountering a symbol such as W''_{n-1} can immediately infer that it refers to the Wiener index of the second independent subtree T''_{n-1}.

The present work makes several novel contributions to the study of exponential recursive ternary trees. First, while Ghasemi et al. (2023) established convergence results for the external path length, we investigate the external Wiener index which provides a more complete description of the tree's global geometric structure. The Wiener index is of fundamental interest in chemical graph theory, network analysis, and phylogenetics, and our results give the first asymptotic characterization of this quantity for exponential recursive trees. Second, we introduce and analyze the number of saturated nodes (nodes with exactly 3 children), a parameter not previously studied in this context. Saturated nodes correspond to fully occupied branching points and have natural interpretations in population genetics (complete speciation events) and network science (local hubs). Third, our methodological contributions include: (i) a martingale construction involving second-order terms to prove almost sure and L^2 convergence of the scaled Wiener index; and (ii) an application of the contraction method to establish the limiting distribution of the number of saturated nodes. These results extend and complement those of Ghasemi et al. (2023), which focused on external path length and size, and they provide a foundation for further study of distance-based parameters in exponential recursive trees.

2 The first two moments

All algebraic simplifications in this section are obtained by direct substitution and summation of geometric series. No specialized numerical or algorithmic code is required; the computations follow from the recurrences stated explicitly in equations (1), (2), (4) and (5). We first need some results that the authors of Ghasemi et al. (2023) have found

$$\mathbb{E}[X_n] = (2p+1)^n, \quad \mathbb{E}[X_n^2] = 3(2p+1)^{2n-1} - 2(1-p)(2p+1)^{n-1} \quad \text{and} \quad \mathbb{E}[Z_n] = 3pn(2p+1)^{n-1}.$$

Theorem 1. *Let Y_n be the number of saturated nodes in an ERTT of age n and index p. We have*

$$\mathbb{E}[Y_n] = \frac{(p^3-p^2+p+2)(2p+1)^n}{2(4-p)(p^2-3p+5)} - \frac{1}{2} + (1-p)^n - \frac{3(1-p)^{2n}}{4-p} + \frac{(1-p)^{3n}}{p^2-3p+5}, \qquad n \geq 1,$$

$$\mathbb{E}[Y_n^2] = \frac{3(p^3-p^2+p+2)^2(2p+1)^{2n-1}}{4(4-p)^2(p^2-3p+5)^2} + \mathcal{O}\big((2p+1)^n\big), \qquad \text{as } n \to \infty.$$

Proof. Taking the expectation of the equation (4), one gets

$$\mathbb{E}[Y_n] = \big(2p+1\big)\mathbb{E}[Y_{n-1}] + p\big(1-(1-p)^{n-1}\big)^3,$$

with initial condition $\mathbb{E}[Y_0] = 0$. Iterating this recurrence we find

$$\mathbb{E}[Y_n] = \sum_{i=1}^{n} \big(2p+1\big)^{n-i} p\big(1-(1-p)^{i-1}\big)^3.$$

This sum yields the explicit solution as stated in the theorem.

Raise both sides of the distributional equation (4) to the second power and take expectations:

$$\mathbb{E}[Y_n^2] = \mathbb{E}\Big[(1-\mathbb{I})\tilde{Y}_{n-1}^2 + \mathbb{I}\Big(\big(Y'_{n-1}\big)^2 + \big(Y''_{n-1}\big)^2 + \big(Y'''_{n-1}\big)^2 + \mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1} + 2Y'_{n-1}Y''_{n-1} + 2Y'_{n-1}Y'''_{n-1} + 2Y''_{n-1}Y'''_{n-1} + 2\big(Y'_{n-1}\mathbb{J}'_{n-1}\big)\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1} + 2\big(Y''_{n-1}\mathbb{J}''_{n-1}\big)\mathbb{J}'_{n-1}\mathbb{J}'''_{n-1} + 2\big(Y'''_{n-1}\mathbb{J}'''_{n-1}\big)\mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\Big)\Big],$$

where $(Y'_{n-1},\mathbb{J}'_{n-1})$, $(Y''_{n-1},\mathbb{J}''_{n-1})$ and $(Y'''_{n-1},\mathbb{J}'''_{n-1})$ are independent copies of $(Y_{n-1},\mathbb{J}_{n-1})$ and independent of $\mathbb{I}$. Then

$$\mathbb{E}[Y_n^2] = (2p+1)\mathbb{E}\big[Y_{n-1}^2\big] + 6p\big(\mathbb{E}[Y_{n-1}]\big)^2 + 6p\big(1-(1-p)^{n-1}\big)^2\mathbb{E}[Y_{n-1}\mathbb{J}_{n-1}] + p\big(1-(1-p)^{n-1}\big)^3.$$

With the explicit substitution of $\mathbb{E}[Y_{n-1}]$ and by $Y_n\mathbb{J}_n = Y_n$, we have

$$\begin{aligned}
\mathbb{E}[Y_n^2] &= \sum_{i=1}^{n}(2p+1)^{n-i}p\Big(6\big(\mathbb{E}[Y_{i-1}]\big)^2 + 6\big(1-(1-p)^{i-1}\big)^2\mathbb{E}[Y_{i-1}] + \big(1-(1-p)^{i-1}\big)^3\Big)\\
&= \frac{3(p^3-p^2+p+2)^2(2p+1)^{2n-1}}{4(4-p)^2(p^2-3p+5)^2} + \frac{6(2+p)((2p+1)^n-(1-p)^{5n})}{(p-4)(p^2-3p+5)(p^4-5p^3+10p^2-10p+7)}\\
&\quad + \frac{3(p^3-p^2+p+2)(p^5-4p^4+5p^3+8p^2-21p+14)(2p+1)^n}{(p-2)(2p+1)(p-4)^2(p^2-3p+5)^2(p^2-3p+3)} - (1-p)^n + \frac{6(1-p)^{2n}}{4-p}\\
&\quad + \frac{(3p^7-18p^6+23p^5-12p^4+49p^3-284p^2+234p+194)(2p+1)^n}{2(2p+1)(p-4)^2(p^2-3p+5)^2} + \frac{1}{4} - \frac{5(1-p)^{3n}}{p^2-3p+5}\\
&\quad + \frac{18(1-p)((2p+1)^n-(1-p)^{4n})}{(p^3-4p^2+6p-6)(p-4)^2} - \frac{6((2p+1)^n-(1-p)^{6n})}{(p^5-6p^4+15p^3-20p^2+15p-8)(p^2-3p+5)^2}\\
&\quad + \frac{3(2+p)(p^3-p^2+p+2)(1-p)^{2n}(2p+1)^{n-1}}{(2-p)(p-4)^2(p^2-3p+3)(p^2-3p+5)} + \frac{6(p^3-p^2+p+2)(1-p)^{3n}(2p+1)^{n-1}}{(p-4)(p^2-3p+3)(p^2-3p+5)^2},
\end{aligned}$$

where the calculation has been done by MAPLE software (see the following Remark). □

Remark 1. *Regarding the MAPLE computations, we clarify the nature of our MAPLE usage. The computations in Theorems 1 and 2 involve:*

- *For Theorem 1: Substituting the explicit expression for $\mathbb{E}[Y_{i-1}]$ into the recurrence:*

  $$\mathbb{E}[Y_n^2] = \sum_{i=1}^{n}(2p+1)^{n-i}p\Big(6\big(\mathbb{E}[Y_{i-1}]\big)^2 + 6\big(1-(1-p)^{i-1}\big)^2\mathbb{E}[Y_{i-1}] + \big(1-(1-p)^{i-1}\big)^3\Big).$$

  This involves expanding polynomial expressions in $(2p+1)^i$, $(1-p)^i$, and their products, then summing geometric series.

- *For Theorem 2: Solving the recurrence (6) for $\mathbb{E}[W_n^2]$, which requires substituting the explicit solutions for the auxiliary moments $\mathbb{E}[Z_n^2]$, $\mathbb{E}[W_nX_n]$, $\mathbb{E}[W_nZ_n]$ and $\mathbb{E}[Z_nX_n]$, then simplifying the resulting expression.*

The MAPLE operations are straightforward: we used the `sum()` *command to evaluate finite sums and the* `simplify()` *command to combine like terms. There is no specialized algorithm or custom code to share, the computations consist of typing the sums as written in the paper and allowing MAPLE to perform the symbolic algebra. A representative example would be:*

Since the computations involve only basic symbolic summation and simplification, providing a separate MAPLE code file would not meaningfully enhance reproducibility beyond the explicit recurrences already given in the paper. The recurrences themselves are fully specified, and any computer algebra system can reproduce the results by following the same steps.

Let S_n be the size of be an ERTT of index p at age n, then by the result $\mathbb{E}[S_n] \sim \frac{(2p+1)^n}{2}$ in Ghasemi et al. (2023) and Theorem 1, we have

$$\frac{\mathbb{E}[Y_n]}{\mathbb{E}[S_n]} \leq \frac{1}{3}, \qquad \left(\because \quad \frac{p^3-p^2+p+2}{(4-p)(p^2-3p+5)} \leq \frac{1}{3} \right).$$

Theorem 2. *Let W_n be the wiener index in an ERTT of age n and index p. We have*

$$\mathbb{E}[W_n] = \frac{27p^2n^2(2p+1)^{2n-2}}{3np+2p+1} + \frac{9pn(2p+1)^{2n-1}}{2(3np+2p+1)} - \frac{3(2p+1)^{2n}}{2(3np+2p+1)} + \frac{3}{2}(2p+1)^{n-1}, \qquad n \geq 1,$$

$$\mathbb{E}[W_n^2] = \frac{324p^2(12p^2+31p+35)}{4(p+1)(4p^2+6p+3)} n^2(2p+1)^{4n-5} + \mathcal{O}\big(n(2p+1)^{4n}\big), \qquad \text{as } n \to \infty.$$

Proof. Take expectations both sides of the equation (5). By the identical distribution of W_n', W_n'', W_n''' and W_n; and the independence of $\mathbb{I}$, we get

$$\begin{aligned}\mathbb{E}[W_n] &= (2p+1)\mathbb{E}[W_{n-1}] + 6p\mathbb{E}[Z_{n-1}]\mathbb{E}[X_{n-1}] + 6p\big(\mathbb{E}[X_{n-1}]\big)^2 \\ &= (2p+1)\mathbb{E}[W_{n-1}] + 18p^2n(p+1)^{2n-3} + 6p(1-p)(p+1)^{2n-3},\end{aligned}$$

with $\mathbb{E}[W_0] = 0$. By iterating, we have

$$\mathbb{E}[W_n] = \sum_{i=1}^{n} (p+1)^{n-i}\Big(18p^2i(2p+1)^{2i-3} + 6p(1-p)(2p+1)^{2i-3}\Big),$$

that is obtained as claimed in the theorem.

For obtaining $\mathbb{E}[W_n^2]$, raise both sides of (5) to the second power and take expectations. The second moment recurrence takes the form,

$$\begin{aligned}\mathbb{E}[W_n^2] &= (2p+1)\mathbb{E}[W_{n-1}^2] + 6p\mathbb{E}[Z_{n-1}^2]\mathbb{E}[X_{n-1}^2] + 12p\big(\mathbb{E}[X_{n-1}^2]\big)^2 + 24p\mathbb{E}[Z_{n-1}X_{n-1}]\big(\mathbb{E}[X_{n-1}]\big)^2 \\ &\quad + 12p\mathbb{E}[X_{n-1}]\mathbb{E}[Z_{n-1}]\mathbb{E}[W_{n-1}] + 12p\mathbb{E}[W_{n-1}X_{n-1}]\mathbb{E}[Z_{n-1}] + 24p\mathbb{E}[W_{n-1}X_{n-1}]\mathbb{E}[X_{n-1}] \\ &\quad + 12p\mathbb{E}[W_{n-1}Z_{n-1}]\mathbb{E}[X_{n-1}] + 24p\mathbb{E}[Z_{n-1}X_{n-1}]\mathbb{E}[X_{n-1}^2] + 24p\mathbb{E}[X_{n-1}]\mathbb{E}[X_{n-1}^2]\mathbb{E}[Z_{n-1}] \\ &\quad + 6p\big(\mathbb{E}[Z_{n-1}X_{n-1}]\big)^2 + 12p\mathbb{E}[W_{n-1}]\big(\mathbb{E}[X_{n-1}]\big)^2 + 6p\big(\mathbb{E}[W_{n-1}]\big)^2 + 24p\mathbb{E}[X_{n-1}^2]\big(\mathbb{E}[X_{n-1}]\big)^2 \\ &\quad + 12p\mathbb{E}[X_{n-1}]\mathbb{E}[Z_{n-1}]\mathbb{E}[Z_{n-1}X_{n-1}] + 6p\mathbb{E}[Z_{n-1}^2]\big(\mathbb{E}[X_{n-1}]\big)^2 + 6p\mathbb{E}[X_{n-1}^2]\big(\mathbb{E}[Z_{n-1}]\big)^2. \qquad (6)\end{aligned}$$

In the recurrence (6), we need the expectations $\mathbb{E}[Z_n^2]$, $\mathbb{E}[W_nZ_n]$, $\mathbb{E}[W_nX_n]$ and $\mathbb{E}[Z_nX_n]$.

From (1), (2) and (5), we have the following recurrences, for $n \geq 1$,

$$\mathbb{E}[Z_nX_n] = (2p+1)\mathbb{E}[Z_{n-1}X_{n-1}] + 6p\mathbb{E}[X_{n-1}]\mathbb{E}[Z_{n-1}] + 3p\mathbb{E}[X_{n-1}^2] + 6p\big(\mathbb{E}[X_{n-1}]\big)^2,$$

$$\mathbb{E}[Z_n^2] = (2p+1)\mathbb{E}[Z_{n-1}^2] + 3p\mathbb{E}[X_{n-1}^2] + 6p\big(\mathbb{E}[Z_{n-1}]\big)^2 + 6p\big(\mathbb{E}[X_{n-1}]\big)^2 + 6p\mathbb{E}[Z_{n-1}X_{n-1}] + 12p\mathbb{E}[Z_{n-1}]\mathbb{E}[X_{n-1}],$$

$$\begin{aligned}\mathbb{E}[W_nX_n] &= (2p+1)\mathbb{E}[W_{n-1}X_{n-1}] + 6p\mathbb{E}[X_{n-1}]\mathbb{E}[Z_{n-1}X_{n-1}] + 6p\mathbb{E}[Z_{n-1}]\big(\mathbb{E}[X_{n-1}]\big)^2 + 6p\mathbb{E}[W_{n-1}]\mathbb{E}[X_{n-1}] \\ &\quad + 6p\mathbb{E}[Z_{n-1}]\mathbb{E}[X_{n-1}^2] + 12p\mathbb{E}[X_{n-1}]\mathbb{E}[X_{n-1}^2] + 6p\big(\mathbb{E}[X_{n-1}]\big)^3\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[W_nZ_n] &= (2p+1)\mathbb{E}[W_{n-1}Z_{n-1}] + 18p\mathbb{E}[X_{n-1}]\mathbb{E}[Z_{n-1}X_{n-1}] + 3p\mathbb{E}[W_{n-1}X_{n-1}] + 6p\big(\mathbb{E}[X_{n-1}]\big)^3 \\ &\quad + 6p\mathbb{E}[Z_{n-1}^2]\mathbb{E}[X_{n-1}] + 6p\mathbb{E}[Z_{n-1}]\mathbb{E}[W_{n-1}] + 6p\mathbb{E}[Z_{n-1}]\mathbb{E}[Z_{n-1}X_{n-1}] + 6p\mathbb{E}[X_{n-1}]\big(\mathbb{E}[Z_{n-1}]\big)^2 \\ &\quad + 6p\mathbb{E}[W_{n-1}]\mathbb{E}[X_{n-1}] + 6p\mathbb{E}[Z_{n-1}]\mathbb{E}[X_{n-1}^2] + 12p\mathbb{E}[X_{n-1}]\mathbb{E}[X_{n-1}^2] + 12p\mathbb{E}[Z_{n-1}]\big(\mathbb{E}[X_{n-1}]\big)^2,\end{aligned}$$

with initial conditions $\mathbb{E}[Z_0^2]=0$, $\mathbb{E}[W_0Z_0]=0$, $\mathbb{E}[W_0X_0]=0$ and $\mathbb{E}[Z_0X_0]=0$.

The above recurrence equations are of the form $y_n=(p+1)y_{n-1}+h_n$, with solution

$$y_n=\sum_{i=1}^{n}(p+1)^{n-i}h_i. \tag{7}$$

Therefore, their solutions are obtained as follows

$$\begin{aligned}
\mathbb{E}[Z_nX_n] &= 9pn(2p+1)^{2n-2}+3(1-p)(2p+1)^{2n-2}-6p(1-p)n(2p+1)^{n-2}-3(1-p)(2p+1)^{n-2},\\
\mathbb{E}[Z_n^2] &= 27p^2n^2(2p+1)^{2n-3}+8p(1-p)n(2p+1)^{2n-3}+\frac{3}{2}(1-p)(5-2p)(2p+1)^{2n-3}\\
&\quad -18(1-p)p^2n^2(2p+1)^{n-3}-6p(1-p)(4-p)n(2p+1)^{n-3}-\frac{3}{2}(1-p)(5-2p)(2p+1)^{n-3},\\
\mathbb{E}[W_nX_n] &= \frac{9p(p+5)}{p+1}n(2p+1)^{3n-3}-\frac{3(2p^3+22p^2+13p-1)}{2(p+1)^2}(2p+1)^{3n-3}-36p(1-p)n(2p+1)^{2n-3}\\
&\quad -\frac{3}{2}(8p^2-16p-1)(2p+1)^{2n-3}-\frac{3(1-p)(4p^2+3p+2)}{2(p+1)^2}(2p+1)^{n-2}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[W_nZ_n] &= \frac{27p^2(p+5)}{p+1}n^2(2p+1)^{3n-4}-\frac{9p(4p^3+38p^2+11p-17)}{2(p+2)^2}n(2p+1)^{3n-4}\\
&\quad -\frac{3(1-p)(4p^3+62p^2+17p-20)}{4(p+2)^2}(2p+1)^{3n-4}-108p^2(1-p)n^2(2p+1)^{2n-4}\\
&\quad -\frac{9}{2}p(20p^2-52p+23)n(2p+1)^{2n-4}-3(1-p)(10p^2-20p+1)(2p+1)^{2n-4}\\
&\quad -\frac{9p(1-p)(4p^2+3p+2)}{2(p+2)^2}n(2p+1)^{n-3}+\frac{3(1-p)(20p^3-8p^2-23p-16)}{4(p+2)^2}(2p+1)^{n-3}.
\end{aligned}$$

Again, according to (7), the solution to the recurrence (6) is

$$\begin{aligned}
\mathbb{E}[W_n^2] &= \frac{324p^2(12p^2+31p+35)}{4(p+1)(4p^2+6p+3)}n^2(2p+1)^{4n-5}\\
&\quad -\frac{27p(176p^6+880p^5+2056p^4+2222p^3+927p^2-66p-111)n(2p+1)^{4n-5}}{(p+1)^2(4p^2+6p+3)^2}+\mathscr{O}\big((2p+1)^{4n}\big),
\end{aligned}$$

as $n\to\infty$. Hence, we obtain the assertion. □

3 Almost sure convergence for the external Wiener index

In T_n, an ERTT of age n and index p, assume we numbered the external nodes of the tree arbitrarily with the numbers $1,2,\ldots,X_n$. Let $D_{i,j}$ be the distance between external nodes i and j, for $i,j\in\{1,2,\ldots,X_n\}$. The external Wiener index of T_n is

$$W_n=\sum_{\substack{i=1\\ i<j}}^{X_n}\sum_{j=1}^{X_n}D_{i,j}.$$

Theorem 3. *Let W_n be the external Wiener index in an ERTT of age n and index p. As $n\to\infty$, we have almost sure convergence and in L^2,*

$$\frac{W_n}{n(2p+1)^{2n}}\xrightarrow{L^2\ \&\ \text{a.s.}}\frac{3p}{2p+1}X_*^2.$$

Proof. Let $\mathbb{I}_k$ be a sequence of independent Bernoulli(p) random variables, defined on the same probability space as the trees. Consider $\{\mathbb{I}_k\}_{k=1}^{\infty}$ to be independent of the structure of the tree at any age, too. If each of two external nodes i and j or both succeed in attracting a new node, then by direct enumeration, the contribution of the external nodes i and j of T_{n-1} into W_n, is

$$9\mathbb{I}_i\mathbb{I}_j(D_{i,j}+2)D_{i,j}\mathbb{I}_{i<j}+3\mathbb{I}_i\big(1-\mathbb{I}_j\big)(D_{i,j}+1)\mathbb{I}_{i\neq j}+6\mathbb{I}_i-\mathbb{I}_i\mathbb{I}_jD_{i,j}\mathbb{I}_{i<j}-\mathbb{I}_i\big(1-\mathbb{I}_j\big)D_{i,j}$$

for $i,j\in\{1,2,\dots,X_{n-1}\}$.

Consequently, we write

$$\begin{aligned}W_n = W_{n-1} &- \sum_{\substack{i=1\\ i\neq j}}^{X_{n-1}}\sum_{j=1}^{X_{n-1}}\mathbb{I}_i\big(1-\mathbb{I}_j\big)D_{i,j} - \sum_{\substack{i=1\\ i<j}}^{X_{n-1}}\sum_{j=1}^{X_{n-1}}\mathbb{I}_i\mathbb{I}_jD_{i,j} + 6\sum_{i=1}^{X_{n-1}}\mathbb{I}_i + 3\sum_{\substack{i=1\\ i\neq j}}^{X_{n-1}}\sum_{j=1}^{X_{n-1}}\mathbb{I}_i\big(1-\mathbb{I}_j\big)(D_{i,j}+1)\\ &+9\sum_{\substack{i=1\\ i<j}}^{X_{n-1}}\sum_{j=1}^{X_{n-1}}\mathbb{I}_i\mathbb{I}_j(D_{i,j}+2).\end{aligned}$$

Let $\mathscr{F}_n$ be the sigma field generated by the first n steps of the evolution of T_n. Then

$$\begin{aligned}\mathbb{E}[W_n|\mathscr{F}_{n-1}] &= W_{n-1}+4p(1-p)W_{n-1}+8p^2W_{n-1}+6pX_{n-1}+3p(2p+1)X_{n-1}(X_{n-1}-1)\\ &= (2p+1)^2W_{n-1}+3p(2p+1)X_{n-1}^2-3p(2p-1)X_{n-1}.\end{aligned}\tag{8}$$

Let $M_n := \alpha_nW_n+\beta_nX_n^2+\gamma_nX_n$, for specified factors α_n, β_n and γ_n that transformed M_n a martingale. So, by (8) and a martingalization procedure, we have

$$\begin{aligned}\mathbb{E}[M_n|\mathscr{F}_{n-1}] &= \alpha_n(2p+1)^2W_{n-1}+3\alpha_np(2p+1)X_{n-1}^2-3\alpha_np(2p-1)X_{n-1}\\ &\quad+\beta_n\mathbb{E}\big[X_n^2\big|\mathscr{F}_{n-1}\big]+\gamma_n\mathbb{E}\big[X_n\big|\mathscr{F}_{n-1}\big]\\ &= \alpha_n(2p+1)^2W_{n-1}+3\alpha_np(2p+1)X_{n-1}^2-3\alpha_np(2p-1)X_{n-1}\\ &\quad+\beta_n(2p+1)^2X_{n-1}^2+4\beta_np(1-p)X_{n-1}+\gamma_n(2p+1)X_{n-1}\\ &= \alpha_n(2p+1)^2W_{n-1}+\big(3\alpha_np(2p+1)+\beta_n(2p+1)^2\big)X_{n-1}^2\\ &\quad+\big(\gamma_n(2p+1)+4\beta_np(1-p)-3\alpha_np(2p-1)\big)X_{n-1}\\ &= M_{n-1} = \alpha_{n-1}W_{n-1}+\beta_{n-1}X_{n-1}^2+\gamma_{n-1}X_{n-1}.\end{aligned}$$

This is possible, if we choose

$$\begin{aligned}\alpha_{n-1} &= \alpha_n(2p+1)^2,\\ \beta_{n-1} &= 3\alpha_np(2p+1)+\beta_n(2p+1)^2,\\ \gamma_{n-1} &= \gamma_n(2p+1)+4\beta_np(1-p)-3\alpha_np(2p-1).\end{aligned}$$

From these solvable recurrences, we obtain

$$\begin{aligned}\alpha_n &= \frac{\alpha_0}{(2p+1)^{2n}},\\ \beta_n &= \frac{\beta_0}{(2p+1)^{2n}}-\frac{3\alpha_0pn}{(2p+1)^{2n+1}},\\ \gamma_n &= \frac{3\alpha_0-4(1-p)\beta_0+2(2p+1)\gamma_0}{2(2p+1)^{n+1}}-\frac{6p(1-p)\alpha_0n}{(2p+1)^{2n+2}}-\frac{3\alpha_0-4(1-p)\beta_0}{2(2p+1)^{2n+1}}.\end{aligned}$$

For simplicity, we take $\alpha_0 = 1$ and $\beta_0 = 0 = \gamma_0$. Thus,

$$M_n = \frac{1}{(2p+1)^{2n}} W_n - \frac{3pn}{(2p+1)^{2n+1}} X_n^2 + \left(\frac{3}{2(2p+1)^{n+1}} - \frac{12p(1-p)n + 3(2p+1)}{2(2p+1)^{2n+2}} \right) X_n,$$

is a martingale with bounded moments

$$\begin{aligned}
\mathbb{E}[M_n] &= \frac{\mathbb{E}[W_n]}{(2p+1)^{2n}} - \frac{3pn\mathbb{E}[X_n^2]}{(2p+1)^{2n+1}} + \left(\frac{3}{2(2p+1)^{n+1}} - \frac{12p(1-p)n + 3(2p+1)}{2(2p+1)^{2n+2}} \right) \mathbb{E}[X_n] \\
&= \frac{27p^2n^2(2p+1)^{-2}}{3np+2p+1} + \frac{9pn(2p+1)^{-1}}{2(3np+2p+1)} - \frac{3}{2(3np+2p+1)} + \frac{3}{2(2p+1)^{n+1}} \\
&\quad - 3pn\left(\frac{3}{(2p+1)^2} - \frac{2(1-p)}{(2p+1)^{n+2}} \right) + \frac{3}{2(2p+1)} - \frac{12p(1-p)n + 3(2p+1)}{2(2p+1)^{n+2}} \\
&= 0,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[M_n^2] &= \frac{\mathbb{E}[W_n^2]}{(2p+1)^{4n}} + \frac{9p^2n^2\mathbb{E}[X_n^4]}{(2p+1)^{4n+2}} + \left(\frac{3}{2(2p+1)^{n+1}} - \frac{12p(1-p)n + 3(2p+1)}{2(2p+1)^{2n+2}} \right)^2 \mathbb{E}[X_n^2] \\
&\quad + \left(\frac{3}{(2p+1)^{3n+1}} - \frac{12p(1-p)n + 3(2p+1)}{(2p+1)^{4n+2}} \right) \left(\mathbb{E}[W_nX_n] - \frac{3pn\mathbb{E}[X_n^3]}{2p+1} \right) - \frac{6pn\mathbb{E}[W_nX_n^2]}{(2p+1)^{4n+1}} \\
&= \frac{1}{(2p+1)^{4n}} \mathbb{E}[W_n^2] + \frac{9p^2n^2}{(2p+1)^{4n+2}} \mathbb{E}[X_n^4] - \frac{6pn}{(2p+1)^{4n+1}} \mathbb{E}[W_nX_n^2] + \mathscr{O}(1) \\
&= \frac{1}{(2p+1)^{4n}} \left(\frac{324p^2(12p^2+31p+35)}{4(p+1)(4p^2+6p+3)} n^2(2p+1)^{4n-5} \right. \\
&\quad \left. - \frac{27p(176p^6+880p^5+2056p^4+2222p^3+927p^2-66p-111)\, n(2p+1)^{4n-5}}{(p+1)^2(4p^2+6p+3)^2} \right) \\
&\quad + \frac{9p^2n^2}{(2p+1)^{4n+2}} \left(\frac{9(12p^2+31p+35)(2p+1)^{4n-3}}{(p+1)(4p^2+6p+3)} \right) \\
&\quad - \frac{6pn}{(2p+1)^{4n+1}} \left(\frac{27p(12p^2+31p+35)n(2p+1)^{4n-4}}{(4p^2+6p+3)(p+1)} \right. \\
&\quad \left. - \frac{9(176p^6+880p^5+2056p^4+2222p^3+927p^2-66p-111)}{2(p+1)^2(4p^2+6p+3)^2} (2p+1)^{4n-4} \right) + \mathscr{O}(1) \\
&= 0 + \mathscr{O}(1) < \infty, \qquad \text{as } n \to \infty,
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}[X_n^3] &= \frac{3(p+5)(2p+1)^{3n-2}}{p+1} - 18(1-p)(2p+1)^{2n-2} + \frac{2(1-p)(2-p)(2p+1)^{n-1}}{p+1}, \\
\mathbb{E}[X_n^4] &= \frac{9(12p^2+31p+35)(2p+1)^{4n-3}}{(p+1)(4p^2+6p+3)} - \frac{36(p+5)(1-p)(2p+1)^{3n-3}}{p+1} \\
&\quad - \frac{12(1-p)(7p^2-6p-7)(2p+1)^{2n-3}}{p+1} - \frac{8(1-p)(2p^4-12p^3+7p^2+6p+3)(2p+1)^{n-2}}{(p+1)(4p^2+6p+3)},
\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[W_n X_n^2] = {} & \frac{27p(12p^2+31p+35)n(2p+1)^{4n-4}}{(4p^2+6p+3)(p+1)} \\ & - \frac{9(176p^6+880p^5+2056p^4+2222p^3+927p^2-66p-111)}{2(p+1)^2(4p^2+6p+3)^2}(2p+1)^{4n-4} \\ & + \frac{3(60p^4+300p^3-60p^2-300p)n(2p+1)^{3n-4}}{2(p+1)^2} + \mathscr{O}\big((2p+1)^{3n}\big), \quad \text{as } n\to\infty.\end{aligned}$$

So, by the martingale convergence theorem (Theorem 6.6.9 in Ash (1999)), M_n converges to some M_* almost surely and in L^2. Consequently,

$$\begin{aligned}\frac{M_n}{n} &= \frac{W_n}{n(2p+1)^{2n}} - \frac{3p}{2p+1}\left(\frac{X_n}{(2p+1)^n}\right)^2 + \left(\frac{3}{2(2p+1)} - \frac{12p(1-p)n+3(2p+1)}{2(2p+1)^{n+2}}\right)\frac{X_n}{n(2p+1)^n} \\ &\xrightarrow{L^2 \text{ \& a.s.}} 0.\end{aligned}$$

Hence, by (3), the result follows. □

Corollary 1. *Let D_n^* be the distance between two randomly chosen nodes in a random ERTT of age n and index p. Then we have*

$$\frac{\mathbb{E}[D_n^*]}{n} \longrightarrow \frac{6p}{2p+1}.$$

Proof. In T_n, a random ERTT of age n and index p, we have

$$\mathbb{E}[D_n^*|T_n] = \frac{2\sum_{i=1}^{X_n}\sum_{\substack{j=1\\ i<j}}^{X_n} D_{i,j}}{X_n(X_n-1)} = \frac{2W_n}{X_n(X_n-1)}.$$

Therefore, by (3) and Theorem 3, we have

$$\frac{\mathbb{E}[D_n^*|T_n]}{n} = \frac{2W_n/\left(n(2p+1)^{2n}\right)}{X_n(X_n-1)/\left((2p+1)^{2n}\right)} \xrightarrow{\text{a.s.}} \frac{6p}{2p+1}.$$

Since $W_n \le X_n(X_n-1)n$, then $\mathbb{E}[D_n^*|T_n]/n \le 2$. Consequently, by dominated convergence theorem,

$$\lim_{n\to\infty}\frac{\mathbb{E}[D_n^*]}{n} = \lim_{n\to\infty}\mathbb{E}\left[\frac{\mathbb{E}[D_n^*|T_n]}{n}\right] = \mathbb{E}\left[\lim_{n\to\infty}\frac{\mathbb{E}[D_n^*|T_n]}{n}\right].$$

This yields the result. □

4 L^2 convergence for the number of saturated nodes

In this section we apply the contraction method in Roesler and Rueschendorf (2001) or the multivariate contraction method in Neininger (2001) to state the limiting distribution of $\hat{Y}_n := \frac{Y_n}{(2p+1)^n}$.

We first need to set up some notation as follows: For a random variable X, we write $X \sim \lambda$ if the distribution of X is λ, i.e. the law $\mathscr{L}(X)$ of X is λ. The symbol $\|X\|_2 := (\mathbb{E}[|X|^2])^{1/2}$ denotes the usual

L_2-norm of X. The Wasserstein-metric ℓ_2 is defined on the space of probability distributions with finite second moments by

$$\ell_2(X,Y) := \ell_2(\mathscr{L}(X),\mathscr{L}(Y)) := \ell_2(\lambda,\nu) := \inf\{\|X-Y\|_2 : X\sim\lambda, Y\sim\nu\}.$$

By $\mathscr{M}_2$ the space of all probability distributions λ with mean $\dfrac{p^3-p^2+p+2}{2(4-p)(p^2-3p+5)}$ (as in Theorem 1) and finite second moment is denoted. The metric space $(\mathscr{M}_2,\ell_2)$ is complete and convergence in ℓ_2 is equivalent to convergence in distribution plus convergence of the second moments.

By equation (4) we have

$$\frac{Y_n}{(2p+1)^n} \overset{d}{=} \frac{\mathbb{I}}{2p+1}\left(\frac{Y'_{n-1}}{(2p+1)^{n-1}}+\frac{Y''_{n-1}}{(2p+1)^{n-1}}+\frac{Y'''_{n-1}}{(2p+1)^{n-1}}\right)+\frac{1-\mathbb{I}}{2p+1}\frac{\tilde{Y}_{n-1}}{(2p+1)^{n-1}}+\mathbb{I}\frac{\mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1}}{(2p+1)^n}. \quad (9)$$

Theorem 4. *Let Y_n be the saturated nodes of an ERTT of age n and index p. The normalized saturated nodes $\hat{Y}_n := \frac{Y_n}{(2p+1)^n}$ satisfies the distributional recursion*

$$\hat{Y}_n \overset{d}{=} \frac{\mathbb{I}}{2p+1}\left(\hat{Y}'_{n-1}+\hat{Y}''_{n-1}+\hat{Y}'''_{n-1}\right)+\frac{1-\mathbb{I}}{2p+1}\bar{Y}_{n-1}+\mathbb{I}\frac{\mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1}}{(2p+1)^n}, \quad (10)$$

where $\mathbb{J}'_{n-1}\overset{d}{=}\mathbb{J}''_{n-1}\overset{d}{=}\mathbb{J}'''_{n-1}\overset{d}{=}\text{Bernoulli}(1-(1-p)^{n-1})$, $\bar{Y}_n$, $(\hat{Y}'_n,\mathbb{J}'_{n-1})$, $(\hat{Y}''_n,\mathbb{J}''_{n-1})$, $(\hat{Y}'''_n,\mathbb{J}'''_{n-1})$ and $\mathbb{I}$ are independent, $\bar{Y}_n\overset{d}{=}\hat{Y}'_n\overset{d}{=}\hat{Y}''_n\overset{d}{=}\hat{Y}'''_n\overset{d}{=}\hat{Y}_n$, and $\mathbb{I}$ is a Bernoulli random variable with success probability p, then $\hat{Y}_n\xrightarrow{\text{d}}\hat{Y}$, as $n\to\infty$, where the random variable $\hat{Y}$ is the unique distributional fixed-point of

$$\hat{Y}\overset{d}{=}\frac{\mathbb{I}}{2p+1}\left(\hat{Y}'+\hat{Y}''+\hat{Y}'''\right)+\frac{1-\mathbb{I}}{2p+1}\hat{Y}, \quad (11)$$

with $\hat{Y}'$, $\hat{Y}''$ and $\hat{Y}'''$ are independent copies of $\hat{Y}$ and independent of $\mathbb{I}$.

Proof. The equation (10) is an immediate consequence of the equation (9), where $\bar{Y}_n := \frac{\tilde{Y}_n}{(2p+1)^n}$, $\hat{Y}'_n := \frac{Y'_n}{(2p+1)^n}$, $\hat{Y}''_n := \frac{Y''_n}{(2p+1)^n}$ and $\hat{Y}'''_n := \frac{Y'''_n}{(2p+1)^n}$. Using Chebyshev's inequality and $\mathbb{E}[\mathbb{J}''_{n-1}]=\mathbb{E}[\mathbb{J}'''_{n-1}]=1-(1-p)^{n-1}$, we have the convergence in probability $\frac{\mathbb{J}'_{n-1}\mathbb{J}''_{n-1}\mathbb{J}'''_{n-1}}{(2p+1)^n}\xrightarrow{\text{p}}0$. That is,

$$\hat{Y}_n\overset{d}{=}\frac{\mathbb{I}}{2p+1}\cdot\left(\hat{Y}'_{n-1}+\hat{Y}''_{n-1}+\hat{Y}'''_{n-1}\right)+\frac{1-\mathbb{I}}{2p+1}\cdot\hat{Y}_{n-1}+o_p(1), \quad (12)$$

where the notation $o_p(1)$ will denote an asymptotically negligible random variable that converges to 0 in probability. For purposes of convergence we can, and hence will, ignore the $o_p(1)$ term.

Consider the well-defined transformation $T:\mathscr{M}_2\to\mathscr{M}_2$,

$$T(\lambda) := \mathscr{L}\left(\frac{\mathbb{I}}{2p+1}\left(\hat{Y}'_\lambda+\hat{Y}''_\lambda+\hat{Y}'''_\lambda\right)+\frac{1-\mathbb{I}}{2p+1}\hat{Y}_\lambda\right), \quad (13)$$

where $\hat{Y}_\lambda\overset{d}{=}\hat{Y}$, $\hat{Y}'_\lambda\overset{d}{=}\hat{Y}'$, $\hat{Y}''_\lambda\overset{d}{=}\hat{Y}''$ and $\hat{Y}'''_\lambda\overset{d}{=}\hat{Y}'''$ have λ as distribution and are independent of $\mathbb{I}$.

At a first step, we have to prove that the transformation T has a unique fixed point with respect to the ℓ_2-metric. Let $\lambda,\nu\in\mathscr{M}_2$ be given. By (13), we have

$$T(\lambda)=\mathscr{L}\left(\frac{\mathbb{I}}{2p+1}\left(\hat{Y}'_\lambda+\hat{Y}''_\lambda+\hat{Y}'''_\lambda\right)+\frac{1-\mathbb{I}}{2p+1}\hat{Y}_\lambda\right),$$
$$T(\nu)=\mathscr{L}\left(\frac{\mathbb{I}}{2p+1}\left(\hat{Y}'_\nu+\hat{Y}''_\nu+\hat{Y}'''_\nu\right)+\frac{1-\mathbb{I}}{2p+1}\hat{Y}_\nu\right),$$

and

$$\begin{aligned}\ell_2^2(T(\lambda),T(\nu)) &\le 3\,\frac{\mathbb{E}[\mathbb{I}^2]}{(2p+1)^2}\,\mathbb{E}\big[\big|\hat{Y}_\lambda-\hat{Y}_\nu\big|\big]^2+\frac{\mathbb{E}\left[(1-\mathbb{I})^2\right]}{(2p+1)^2}\,\mathbb{E}\big[\big|\hat{Y}_\lambda-\hat{Y}_\nu\big|\big]^2\\ &=\frac{1}{2p+1}\,\mathbb{E}\big[\big|\hat{Y}_\lambda-\hat{Y}_\nu\big|\big]^2.\end{aligned}$$

Therefore, we have

$$\ell_2^2(T(\lambda),T(\nu)) \le \frac{1}{2p+1}\,\ell_2^2(\lambda,\nu).$$

Since $\frac{1}{2p+1}<1$, we deduce that T is a contraction with respect to the ℓ_2-metric. Thus, Banach's fixed point theorem provides existence and uniqueness of solutions of the fixed point equation $T(\lambda)=\lambda$. By (12), if $\hat{Y}_n$ converges in distribution to some random variable $\hat{Y}$, then $\hat{Y}$ satisfies (11). Consequently, the distribution of $\hat{Y}$ will be λ_0, the unique fixed point of T, i.e. $\mathscr{L}(\hat{Y})=T(\lambda_0)=\lambda_0$. Therefore, we have to prove $\lim_{n\to\infty}\ell_2(\hat{Y}_n,\hat{Y})=0$ to conclude $\hat{Y}_n \xrightarrow{\mathrm{d}} \hat{Y}$.

Since $\hat{Y}_n' \overset{d}{=} \hat{Y}_n'' \overset{d}{=} \hat{Y}_n''' \overset{d}{=} \hat{Y}_n$ and $\hat{Y}' \overset{d}{=} \hat{Y}'' \overset{d}{=} \hat{Y}''' \overset{d}{=} \hat{Y}$, we deduce

$$\begin{aligned}\lim_{n\to\infty}\ell_2^2(\hat{Y}_n,\hat{Y}) &\le \frac{p}{(2p+1)^2}\lim_{n\to\infty}\left(\left\|\hat{Y}_{n-1}'-\hat{Y}'\right\|_2^2+\left\|\hat{Y}_{n-1}''-\hat{Y}''\right\|_2^2+\left\|\hat{Y}_{n-1}'''-\hat{Y}'''\right\|_2^2\right)+\frac{1-p}{(2p+1)^2}\lim_{n\to\infty}\left\|\hat{Y}_{n-1}-\hat{Y}\right\|_2^2\\ &\le \frac{1}{2p+1}\lim_{n\to\infty}\left\|\hat{Y}_n-\hat{Y}\right\|_2.\end{aligned}$$

Therefore, we have

$$\lim_{n\to\infty}\ell_2^2(\hat{Y}_n,\hat{Y}) \le \frac{1}{2p+1}\lim_{n\to\infty}\ell_2^2(\hat{Y}_n,\hat{Y}).$$

This is true only if $\lim_{n\to\infty}\ell_2(\hat{Y}_n,\hat{Y})=0$. □

5 Conclusion

This work contributes to the growing body of literature on exponential recursive trees by providing a rigorous probabilistic analysis of two important parameters: the external Wiener index and the number of saturated nodes. The martingale approach and the contraction method, applied in concert, offer a robust framework for studying convergence phenomena in random recursive structures. While we have established convergence in distribution and L^2 for the scaled quantities, the explicit distribution of the limiting random variables X_* and $\hat{Y}$ remains unknown. Characterizing these distributions, perhaps through their moment generating functions or through recursive integral equations, would provide deeper insight into the asymptotic structure of ERTTs.

Acknowledgements

We would like to sincerely thank the reviewers for their insightful and constructive comments. Their suggestions have significantly strengthened the clarity and rigor of our manuscript. We have incorporated their feedback throughout the revised draft, and we are grateful for their time and expertise. The author also acknowledge the use of AI-assisted language tools for editing and refining the exposition. All technical and mathematical content is the sole responsibility of the authors.

References

Aguech, R., Bose, S., Mahmoud, H. and Zhang, Y. (2021). Some properties of exponential trees. *International Journal of Computer Mathematics*, **3**, 16–32.

Ash, R. B. (1999). *Probability and Measure Theory*. Second Edition, Academic Press, New York.

Chan, D. and Hughes, B., Leong, A. and Reed, W. (2003). Stochastically evolving networks. *Physical Review E*, **68**, 066124.

Feng, Y., and Mahmoud, H. (2018). Profile of random exponential binary trees. *Methodology and Computing in Applied Probability*, **20**, 575–587.

Grübel, R. and Michailow, I. (2015). Random recursive trees: a boundary theory approach. *Electronic Journal of Probability*, **20**, 1–22.

Ghasemi, M., Javanian, M. and Imany Nabiyyi, R. (2023). Note on the exponential recursive k-ary trees. *RAIRO-Theoretical Informatics and Applications*, **57**, 1–14.

Javanian, M. and Vahidi-Asl, M. Q. (2003). Note on the outdegree of a node in random recursive trees. *Journal of Applied Mathematics and Computing*, **13**, 99–103.

Javanian, M. and Vahidi-Asl, M. Q. (2006). Depth of nodes in random recursive k-ary trees. *Information Processing Letters*, **98**, 115–118.

Javanian, M. and Vahidi-Asl, M. Q. (2012). A strong law for the size of Yule m-oriented recursive trees. *Journal of Applied Mathematics, Statistics and Informatics*, **8**, 67–72.

Javanian, M. (2013). Limit distribution of the degrees in scaled attachment random recursive trees. *Bulletin of the Iranian Mathematical Society*, **39**, 1031–1036.

Javanian, M. (2014). On the external path length of random recursive k-ary trees. *Italian Journal of Pure and Applied Mathematics*, **31**, 21–26.

Javanian, M. (2017). On the size of paged recursive trees. *Discrete Mathematics, Algorithms and Applications*, **9**, 1–10.

Mahmoud, H. (2022). Profile of random exponential recursive trees. *Methodology and Computing in Applied Probability*, **24**, 259–275.

Moon, J. W. (1974). The distance between nodes in recursive trees. *London Mathematics Society Lecture Notes Series*, **13**, 125–132.

Najock, D. and Heyde, C. C. (1982). On the number of terminal vertices in certain random trees with an application to stemma construction in philology. *Journal of Applied Probability*, **19**, 675–680.

Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Structures and Algorithms*, **19**, 498–524.

Neininger, R. (2002). The Wiener index of random trees. *Combinatorics, Probability and Computing*, **11**, 587–597.

Roesler, U. and Rueschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica*, **29**, 3–33.

Smythe, R. T. and Mahmoud, H. (1995). A survey of recursive trees. *Theory of Probability and Mathematical Statistics*, **51**, 1–27.

Zhang, P. (2018). On several properties of plain-oriented recursive trees. arXiv:1706.02441v2.

End of SMPS Journal – Volume 3, Issue 1

For manuscript submission and future issues, visit:
www.smps.um.ac.ir

Stochastic Models in Probability and Statistics

CONTENTS Vol. 3, No. 1, 2026

Some ratio type estimators for estimating the population mean in case of missing data 1-11

Ajeet Kumar Singh and Himanshi Kumawat

Multimatrix variate distributions 13-27

José A. Díaz-García and Francisco J. Caro-Lopera

Exploring an extension of the Rayleigh distribution: statistical estimation and practical application 29-44

Harsh Tripathi and Varun Agiwal

Quasi-identifiability of multipath change-point models with general link functions 45-52

Masoud Majidizadeh

Efficient pseudo Rao-Blackwellized estimator in ranked sampling 53-70

Bardia Panahbehagh and Mahdi Salehi

Wiener index and saturated nodes of random exponential recursive ternary trees 71-84

Mehri Javanian

Web Site: https://smps.um.ac.ir
Email: smps@um.ac.ir
Print ISSN: 3115-9370
Online ISSN: 3060-7876