Filomat 33:17 (2019), 5589–5610
https://doi.org/10.2298/FIL1917589A

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Bivariate Discrete Nadarajah and Haghighi Distribution: Properties and Different Methods of Estimation

Sajid Ali[a], Muhammad Shafqat[a], Ismail Shah[a], Sanku Dey[b]

[a] *Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan*
[b] *Department of Statistics, St. Anthony's College, Shillong-793001, Meghalaya, India*

Abstract. The exponential distribution is commonly used to model different phenomena in statistics and reliability engineering. A new extension of exponential distribution known as the Nadarajah and Haghighi [An extension of the exponential distribution, *Statistics: A Journal of Theoretical and Applied Statistics* 45 (2011) 543-558.] distribution was introduced in the literature to accommodate the inflation of zero in the data. In practice, however, discrete data are easy to collect as compared to continuous data. Discrete bivariate distributions play important roles in modeling bivariate lifetime count data. Thus focusing on the utility of discrete data, this study presents a new bivariate discrete Nadarajah and Haghighi distribution. We discuss some basic properties of the proposed distribution and study seven different methods of estimation for the unknown parameters to assess the performance of the proposed bivariate discrete model. Two data sets are also analyzed to demonstrate how the proposed model may work in practice. Results show that the proposed model is very flexible and performs better than some of the existing models.

1. Introduction

Weibull and gamma distributions are commonly used to analyze monotonic hazard rates. The Weibull distribution has an advantage of having closed form survival function over the gamma distribution and is suitable to model constant, increasing and decreasing failure rates. Although Weibull and gamma distributions are generalized forms of the exponential model, the popularity of the exponential distribution is still undoubted. This distribution has also a special relationship with the Poisson point process. In fact, this is time-between-events distribution that is used to characterize the point process. There are many studies that extend the existing probability models, see, for instance, [6, 9, 11, 12, 14, 16, 17, 19, 21, 26, 29], and references cited therein.

Researchers in many fields encounter discrete data, for example, the marks obtained by students in an examination, the number of goals scored by a football team, the number of cycles prior to the first failure when devices work in cycles, etc. Although discrete data is as important as continuous, studies on discrete distributions are less common than continuous distributions. To bridge this gap, this study

2010 *Mathematics Subject Classification*. Primary 62E15; Secondary 62F10, 62H10, 62H15
Keywords. Bivariate Discrete Nadarajah and Haghighi Distribution, Method of maximum likelihood estimation, Method of least squares, Method of percentile estimation, Anderson Darling estimation method, Nadarajah and Haghighi Distribution
Received: 16 December 2018; Revised: 07 August 2019; Accepted: 07 September 2019
Communicated by Aleksandar Nastić
The first author research is supported by University Research Fund provided by Quaid-i-Azam University, Islamabad, Pakistan.
Email addresses: sajidali.qau@hotmail.com (Sajid Ali), shafqatmna@gmail.com (Muhammad Shafqat), ishah@qau.edu.pk (Ismail Shah), sankud66@gmail.com (Sanku Dey)

aims to introduce a new bivariate discrete probability model from the extension of exponential distribution introduced by Nadarajah and Haghighi[20] and to derive some of its basic properties. The survival function of the Nadarajah and Haghighi (NH) distribution is:

$$S(t) = \exp(1 - (1 + \lambda t)^{\alpha}) \tag{1}$$

where $\alpha, \lambda > 0$ are the shape and the rate parameters, respectively, of the NH distribution. The resulting cumulative distribution function (CDF), probability density function (PDF) and hazard rate function (HRF) are given as follows:

$$F(t) = 1 - \exp\left(1 - (1 + \lambda t)^{\alpha}\right), \tag{2}$$

$$f(t) = \alpha\lambda(1 + \lambda t)^{\alpha-1} \exp\left(1 - (1 + \lambda t)^{\alpha}\right) \tag{3}$$

$$h(t) = \alpha\lambda(1 + \lambda t)^{\alpha-1} \tag{4}$$

The hazard rate function (HRF) of NH distribution has a closed form (Equation 4) as in the case of Weibull distribution and the generalized exponential distribution. Equation (3) reduces to the exponential distribution for $\alpha = 1$, and has an interesting property of having zero mode yet allowing increasing, decreasing and constant HRFs.

Discretization of continuous distribution can be done by using different methodologies. For instance, Nekoukhou et al. [22] used a technique to convert a continuous distribution into discrete analogue. To this end, for any continuous distribution on $\mathfrak{R}^{+} = [0, +\infty)$ with f(t), one can construct a discrete distribution supported on the set of integers, $N_0 = 0, 1, 2, \cdots$ whose probability mass function (pmf) is of the form

$$P_t = P(T = t) = s(t) - s(t + 1), \qquad t = 0, 1, 2, 3, \cdots . \tag{5}$$

where $s(t)$ is the survival function of $f(t)$. Substituting Equation (1) in Equation (5), the pmf of the resulting discrete NH distribution is given by

$$P(T = t) = \exp(1)(\theta^{(\frac{1}{\lambda}+t)^{\alpha}} - \theta^{(\frac{1}{\lambda}+t+1)^{\alpha}}), \qquad t = 0, 1, 2, 3, \cdots \tag{6}$$

where $\theta = \exp(-\lambda^{\alpha})$ and $0 < \theta < 1$. The new distribution is named as the discrete NH (DNH) distribution denoted by the DNH$(\alpha, \lambda, \theta)$ with parameters $\alpha, \lambda > 0$ and $0 < \theta < 1$.

Generally, bivariate distributions are better suited for practical use of both, simple and compact statistical expressions. In real life, many phenomena occur in two dimensions whereas in statistics, models that describe such phenomena are called the bivariate probability distributions. In many different real world applications, bivariate distributions have been applied to model dependent random quantities. As continuous random variables are commonly encountered in practice, discrete random quantities can also be observed in several different practical experiments. For example, in lifetime analysis and modeling, the failure of the items is generally collected and reported hourly, daily, weekly, and so forth. Similarly, the number of goals scored by two competing teams or the number of insurance claims for two different causes is purely discrete in nature. In the literature, there are many bivariate distributions, e.g., Sarhan and Balakrishnan[27] developed a bivariate distribution using the generalized exponential (GE) distribution. Gupta and Kundu[10] introduced a four-parameter bivariate generalized exponential (BVGE) distribution with GE marginals. Another bivariate extension of the GE was introduced by Nekoukhou and Kundu[23]. Later, Ashour et al.[2] obtained joint moments and the moment generating function for the bivariate generalized exponential distribution in closed form. Freund[8] introduced a bivariate extension of the exponential distribution. Block and Basu[3] used an absolutely continuous bivariate extension of the exponential distribution and discussed its properties. Some more recent contributions can be seen in [5, 13–15, 18, 24, 25], and references cited therein. Before proceeding further, it is worth mentioning that the cited literature does not account the inflation of zero. Therefore, the unique focus of this article is to introduce a bivariate discrete distribution suitable for accommodating the inflation of zero.

The first objective of this paper is to propose a new discrete bivariate Nadarajah and Haghighi (BDNH) distribution and to derive some of its basic properties. The second objective is to consider seven different methods of estimation for the unknown parameters of the proposed bivariate discrete Nadarajah Haghighi (BDNH) distribution. In particular, we compare the maximum likelihood estimators, least-squares estimators, weighted least-squares estimators, maximum product of spacings estimators, Cramér-von Mises estimators, Anderson-Darling estimators, and Right-tail Anderson-Darling estimators. Further, we study the behavior of these estimators for different sample sizes and for different parameter values. Sinces it is difficult to compare theoretically the performances of the different estimators, we perform extensive simulation studies to compare the performances of the different estimation methods based on bias and root mean squared error (RMSE).

The rest of the study is arranged as follows. Section 2 discusses some important features and properties of the proposed model, including, probability mass function (PMF), cumulative distribution function (CDF), conditional probability mass function and cumulative function, Survival function (SF) and stochastic ordering. Estimation of the parameters of the bivariate discrete Nadarajah Haghighi (BDNH) distribution based on seven different methods is discussed in Section 3. To assess the performance of different estimation methods, a simulation study is conducted in Section 4. Two applications of the BDNH distribution and a comparison to the bivariate discrete Weibull (BDW) and bivariate discrete generalized exponential (BDGE) distributions are presented in Section 5. At the end, some concluding remarks are given in Section 6.

2. BDNH distribution and its properties

Suppose $U_1 \sim DNH(\alpha_1, \lambda, \theta)$, $U_2 \sim DNH(\alpha_2, \lambda, \theta)$ and $U_3 \sim DNH(\alpha_3, \lambda, \theta)$ and they are independent. If $T_1 = \max(U_1, U_2)$ and $T_2 = \max(U_2, U_3)$, then we say that the bivariate vector (T_1, T_2) has a BDNH distribution with parameters vector $\Psi = (\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)^T$. From now on, we will denote this discrete bivariate distribution by $BDNH(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)$.

If $(T_1, T_2) \sim BDNH(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)$, then CDF of (T_1, T_2) for $t_1 \in N_0, t_2 \in N_0$ and $z = min\{t_1, t_2\}$ is

$$\begin{aligned} F_{T_1,T_2}(t_1, t_2) &= (1 - \exp(1)\theta^{(1+t_1)^{\alpha_1}})(1 - \exp(1)\theta^{(1+t_2)^{\alpha_2}})(1 - \exp(1)\theta^{(1+Z)^{\alpha_3}}) \\ &= F_{DNH}(t_1; \alpha_1, \lambda, \theta)F_{DNH}(t_2; \alpha_2, \lambda, \theta)F_{DNH}(z; \alpha_3, \lambda, \theta) \end{aligned} \tag{7}$$

$$= \begin{cases} F_{DNH}(t_1; \alpha_1, \lambda, \theta)F_{DNH}(t_2; \alpha_2, \lambda, \theta)F_{DNH}(t_1; \alpha_3, \lambda, \theta) & \text{if } t_1 < t_2 \\ F_{DNH}(t_1; \alpha_1, \lambda, \theta)F_{DNH}(t_2; \alpha_2, \lambda, \theta)F_{DNH}(t_2; \alpha_3, \lambda, \theta) & \text{if } t_2 < t_1 \\ F_{DNH}(t; \alpha_1, \lambda, \theta)F_{DNH}(t; \alpha_2, \lambda, \theta)F_{DNH}(t; \alpha_3, \lambda, \theta) & \text{if } t_1 = t_2 = t \end{cases}$$

The corresponding joint PMF of (T_1, T_2) for $t_1, t_2 \in N_0$ is given by

$$f_{t_1,t_2}(t_1, t_2) = \begin{cases} f_1(t_1, t_2) & \text{if } 0 \le t_1 < t_2 \\ f_2(t_1, t_2) & \text{if } 0 \le t_2 < t_1 \\ f_0(t_1, t_2) & \text{if } 0 \le t_1 = t_2 = t \end{cases}$$

Note that the expression $f_1(t_1, t_2)$, $f_2(t_1, t_2)$ and $f_3(t_1, t_2)$ for $t_1, t_2 \in N_0$ can easily be obtained by using the relation

$$f_{T_1,T_2}(t_1, t_2) = F_{T_1,T_2}(t_1, t_2) - F_{T_1,T_2}(t_1 - 1, t_2) - F_{T_1,T_2}(t_1, t_2 - 1) + F_{T_1,T_2}(t_1 - 1, t_2 - 1) \tag{8}$$

If $(T_1, T_2) \sim BDNH(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)$, then the joint survival function (SF) of the vector (T_1, T_2) can also be expressed in a compact form using the following relationship

$$S_{T_1,T_2}(t_1, t_2) = 1 - F_{T_1}(t_1) - F_{T_2}(t_2) + F_{T_1,T_2}(t_1, t_2). \tag{9}$$

The conditional PMF and CDF of the BDNH distribution are given as the conditional distribution of T_1 given T_2. If $(T_1, T_2) \sim BDNH(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)$, then the conditional PMF of T_1 given $T_2 = t_2$, say, $f_{T_1|T_2=t_2}(t_1|t_2)$ is given by

$$f_{T_1|T_2=t_2}(t_1|t_2) = \begin{cases} f_1(t_1|t_2) & \text{if } 0 \le t_1 < t_2 \\ f_2(t_1|t_2) & \text{if } 0 \le t_2 < t_1 \\ f_0(t|t) & \text{if } 0 \le t_1 = t_2 = t \end{cases}$$

where

$$f_i(t_1|t_2) = \frac{f_i(t_1, t_2)}{f_{DNH}(t_2, \alpha_2, \lambda, \theta) f_{DNH}(t_2, \alpha_3, \lambda, \theta)} \tag{10}$$

and

$$f_0(t|t) = \frac{f_0(t)}{f_{DNH}(t, \alpha_2, \lambda, \theta) f_{DNH}(t, \alpha_3, \lambda, \theta)} \tag{11}$$

The conditional CDF of T_1 given $T_2 = t_2$, say, $F_{T_1|T_2=t_2}(t_1)$ is given by
$F_{T_1|T_2=t_2}(t_1) = P(T_1 \le t_1 | T_2 = t_2)$

$$= \begin{cases} \frac{F_{DNH}(t_1;\alpha_1,\alpha_3,\lambda,\theta) f_{DNH}(t_2;\alpha_2,\lambda,\theta)}{f_{DNH}(t_2;\alpha_2,\alpha_3)} & \text{if } 0 \le t_1 < t_2 \\ F_{DNH}(t_1; \alpha_1, \lambda, \theta) & \text{if } 0 \le t_2 < t_1 \\ \frac{F_{DNH}(t;\alpha_1+\alpha_2+\alpha_3,\lambda,\theta) - F_{DNH}(t;\alpha_1) F_{DNH}(t-1;\alpha_2+\alpha_3)}{f_{DNH}(t;\alpha_2+\alpha_3,\lambda,\theta)} & \text{if } 0 \le t_1 = t_2 = t. \end{cases}$$

Suppose $(T_1, T_2) \sim BDNH(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)$, then T_2 is left-tail decreasing in T_1, and T_2 is stochastically increasing in T_1. Let (T_1, T_2) be a pair of random variables, then (i) T_2 is said to be left-tail decreasing in T_1 if and only if $P(T_2 \le t_2 | T_1 \le t_1)$ is a non-increasing function of t_1 for every t_2, and (ii) T_2 is said to be stochastically increasing in T_1 if and only if $P(T_2 \le t_2 | T_1 = t_1)$ is a non-increasing function of t_1 for every t_2. Suppose (T_1, T_2) is a pair of discrete random variables having support on $N_0 \times N_0$, then we define a total positivity of order two property (TP_2) when the joint PMF $f(t_1, t_2)$ satisfies

$$f(t_{11}, t_{21}) f(t_{12}, t_{22},) \ge f(t_{12}, t_{21}) f(t_{11}, t_{22}) \tag{12}$$

3. Parameter Estimation

This section considers different methods of parameter estimation for $BDNH(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta)$ assuming a bivariate sample vector of n observations.

3.1. Maximum Likelihood Estimation Method

This method is extensively used in statistics to estimate parameters [4] and has very nice properties. Assuming a bivariate random vector $D = \{(t_{11}, t_{21}), \cdots, (t_{1n}, t_{2n})\}$ of size n from the BDNH distribution, we are interested in to estimate the parameters $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ using the information of the observed bivariate sample vector D. To proceed further, we use the following notations:

$$l_1 = \{i : t_{1i} < t_{2i}\}, \qquad l_2 = \{i : t_{1i} > t_{2i}\}, \qquad l_0 = \{i : t_{1i} = t_{2i} = t_i\}, \tag{13}$$

and $n_1 = |l_1|$, $n_2 = |l_2|$. Then, the log-likelihood function can be written as

$$l(\alpha_1, \alpha_2, \alpha_3, \lambda, \theta | D) = l_1 \cup l_2 \cup l_0 \tag{14}$$

where

$$
\begin{aligned}
l_1 = n_1\log(\alpha_1+\alpha_2+\alpha_3) + (\alpha_1-1)\sum_{i=1}^{n_1}\log(1+t_{2i}) + \sum_{i=1}^{n_1}(1+t_{2i})^{\alpha_2}\log(\theta)\\
+n_1\log((\log(\theta))^2) + (\alpha_2+\alpha_3-2)\sum_{i=1}^{n_1}\log(1+t_{1i})\\
+2\sum_{i=1}^{n_1}(1+t_{1i})^{\alpha_1}\log(\theta) + 2\sum_{i=1}^{n_1}(1+t_{1i})^{\alpha_3}\log(\theta)
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
l_2 = n_2\log(\alpha_1+\alpha_2+\alpha_3) + (\alpha_1-1)\sum_{i=1}^{n_2}\log(1+t_{1i}) + \sum_{i=1}^{n_2}(1+t_{1i})^{\alpha_1}\log(\theta)\\
+n_2\log((\log(\theta))^2) + (\alpha_2+\alpha_3-2)\sum_{i=1}^{n_2}\log(1+t_{2i})\\
+2\sum_{i=1}^{n_2}(1+t_{2i})^{\alpha_3}\log(\theta) + 2\sum_{i=1}^{n_2}(1+t_{2i})^{\alpha_2}\log(\theta)
\end{aligned} \tag{16}
$$

$$
\begin{aligned}
l_0 = n_0\log(\alpha_1+\alpha_2+\alpha_3) + 3n_0\log(\log(\frac{1}{\theta})) + 3\sum_{i=1}^{n_0}(1+t_i)^{\alpha_2}\log(\theta)\\
+3\sum_{i=1}^{n_0}(1+t_i)^{\alpha_1}\log(\theta) + 2\sum_{i=1}^{n_0}(1+t_i)^{\alpha_3}\log(\theta)\\
+(\alpha_1+\alpha_2+\alpha_3-3)\sum_{i=1}^{n_0}\log(1+t_i) - 2n_0
\end{aligned} \tag{17}
$$

The maximum likelihood estimators of the unknown parameters can be obtained by maximizing the Equation (14) with respect to the unknown parameters. To this end, the parameters can be obtained by solving five non-linear equations simultaneously with the help of an iterative procedure like, the Newton-Raphson method.

3.2. Least Squares Estimation Method

The method of least squares estimate parameters by minimizing the squared discrepancies between the observed data and their expected values [28]. Under this method, the estimators of the unknown parameters $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ of BDNH distribution can be obtained by minimizing

$$
\sum_{i=1}^{n}\left[F_j(T_{(i)}) - \frac{i}{n_j+1}\right]^2, \qquad j = 0, 1, 2 \tag{18}
$$

with respect to unknown parameters $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ.

Suppose $F_1(T_{(i)})$ denote the distribution function of the ordered random variables $T_{(1)} < T_{(2)} < \cdots < T_{(n_1)}$, where $\{T_1, T_2, \cdots, T_{n_1}\}$ is a random sample of size n_1 from distribution function $F(.)$ when $t_1 < t_2$, $F_2(T_{(i)})$ denote the distribution function of the ordered random variables $T_{(1)} < T_{(2)} < \cdots < T_{(n_2)}$, where $\{T_1, T_2, \cdots, T_{n_2}\}$ is a random sample of size n_2 from distribution function $F(.)$ when $t_2 < t_1$ and $F_0(T_{(i)})$ denote the distribution function of the ordered random variables $T_{(1)} < T_{(2)} < \cdots < T_{(n_0)}$, where $\{T_1, T_2, \cdots, T_{n_0}\}$ is a random sample of size n_0 from distribution function $F(.)$ when $t_1 = t_2 = t$. In this settings, the least squares estimators of $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ, say $\hat{\alpha}_{1LSE}$, $\hat{\alpha}_{2LSE}$, $\hat{\alpha}_{3LSE}$ $\hat{\lambda}_{LSE}$ and $\hat{\theta}_{LSE}$ respectively, can be obtained by minimizing

$$\sum_{i=1}^{n_1}\left[F_1(T_{(i)}) - \frac{i}{n_1+1}\right]^2 I_{t_1<t_2} + \sum_{i=1}^{n_2}\left[F_2(T_{(i)}) - \frac{i}{n_2+1}\right]^2 I_{t_2<t_1} + \sum_{i=1}^{n_0}\left[F_0(T_{(i)}) - \frac{i}{n_0+1}\right]^2 I_{t_1=t_2} \tag{19}$$

The weighted least squares estimators of the unknown parameters can be obtained by minimizing

$$\sum_{i=1}^{n_1} w_i\left[F_1(T_{(i)}) - \frac{i}{n_1+1}\right]^2_{I_{t_1<t_2}} + \sum_{i=1}^{n_2} w_i\left[F_2(T_{(i)}) - \frac{i}{n_2+1}\right]^2_{I_{t_2<t_1}} + \sum_{i=1}^{n_0} w_i\left[F_0(T_{(i)}) - \frac{i}{n_0+1}\right]^2_{I_{t_1=t_2}} \tag{20}$$

where, $I(\cdot)$ is the indicator function and

$$F_1(T_{(i)}) = (1-\exp(1)\theta^{(1+t_{1(i)})^{\alpha_1}})(1-\exp(1)\theta^{(1+t_{2(i)})^{\alpha_2}})(1-\exp(1)\theta^{(1+t_{1(i)})^{\alpha_3}}) \qquad \text{if } 0 \le t_1 < t_2 \tag{21}$$

$$F_2(T_{(i)}) = (1-\exp(1)\theta^{(1+t_{1(i)})^{\alpha_1}})(1-\exp(1)\theta^{(1+t_{2(i)})^{\alpha_2}})(1-\exp(1)\theta^{(1+t_{2(i)})^{\alpha_3}}) \qquad \text{if } 0 \le t_2 < t_1 \tag{22}$$

$$F_0(T_{(i)}) = (1-\exp(1)\theta^{(1+t_{(i)})^{\alpha_1}})(1-\exp(1)\theta^{(1+t_{(i)})^{\alpha_2}})(1-\exp(1)\theta^{(1+t_{(i)})^{\alpha_3}}) \qquad \text{if } 0 \le t_1 = t_1 = t \tag{23}$$

with respect to $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ. The weights w_i are equal to $\frac{1}{V(T_i)} = \frac{(n+1)^2(n+2)}{i(n-i+1)}$.

3.3. Maximum Product of Spacings Method

The maximum product of spacings (MPS) method is an alternate to MLEs for the estimation of parameters. With this method, first we define the uniform spacings of a random sample from the BDNH distribution as follows:

$$D(t_1, t_2; \alpha_1, \alpha_2, \alpha_3, \lambda, \theta) = D_{1i}(t_1; \alpha_1, \lambda, \theta) I_{t_1<t_2} + D_{2i}(t_2; \alpha_2, \lambda, \theta) I_{t_2<t_1} + D_{0i}(t_0; \alpha_3, \lambda, \theta) I_{t_2=t_1} \tag{24}$$

where, $I(\cdot)$ is an indicator function and

$$D_{1i}(t_1; \alpha_1, \lambda, \theta) = F(t_{1i:n_1}|\alpha_1, \lambda, \theta) - F(t_{1i-1:n_1}|\alpha_1, \lambda, \theta), \qquad \text{if } 0 \le t_1 < t_2 \tag{25}$$

$$D_{2i}(t_2; \alpha_2, \lambda, \theta) = F(t_{2i:n_2}|\alpha_2, \lambda, \theta) - F(t_{2i-1:n_2}|\alpha_2, \lambda, \theta), \qquad \text{if } 0 \le t_2 < t_1 \tag{26}$$

$$D_{0i}(t_0; \alpha_3, \lambda, \theta) = F(t_{0i:n_0}|\alpha_3, \lambda, \theta) - F(t_{0i-1:n_0}|\alpha_3, \lambda, \theta), \qquad \text{if } 0 \le t_1 = t_2 = t \tag{27}$$

where $i = 1, 2, \cdots, n$, $F(t_{j0}|\alpha_1, \alpha_2, \alpha_3, \lambda, \theta) = 0$ and $F(t_{jn}|\alpha_1, \alpha_2, \alpha_3, \lambda, \theta) = 1$, $j = 0, 1, 2$. The MPS estimators $\hat{\alpha}_{1MPS}$, $\hat{\alpha}_{2MPS}$, $\hat{\alpha}_{3MPS}$, $\hat{\lambda}_{MPS}$ and $\hat{\theta}_{MPS}$, of the parameters $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ are obtained by maximizing with respect to $\alpha_1, \alpha_2, \alpha_3, \lambda$ and θ.

3.4. Cramer-Von-Mises Method

The Cramer-Von-Mises type minimum distance estimator has empirical evidence in the literature that the bias of the estimator is smaller than the other minimum distance type estimators. Thus, the Cramer-Von-Mises estimates for BDNH distribution are $\hat{\alpha}_{1CME}$, $\hat{\alpha}_{2CME}$, $\hat{\alpha}_{3CME}$, $\hat{\lambda}_{CME}$, and $\hat{\theta}_{CME}$ of the parameters and are obtained by minimizing the following function with respect to parameters.

$$C(t_1, t_2; \alpha_1, \alpha_2, \alpha_3\lambda, \theta) = C_1(t_1; \alpha_1, \lambda, \theta) I_{t_1<t_2} + C_2(t_2; \alpha_2, \lambda, \theta) I_{t_2<t_1} + C_0(t; \alpha_3, \lambda, \theta) I_{t_2=t_1} \tag{28}$$

where, $I(\cdot)$ is an indicator function and

$$C_1(t_1;\alpha_1,\lambda,\theta) = \frac{1}{12n_1} + \sum_{i=1}^{n_1}\left(F(t_{1i:n_1}|\alpha_1,\lambda,\theta) - \frac{2i-1}{2n_1}\right)^2 \tag{29}$$

$$C_2(t_1;\alpha_2,\lambda,\theta) = \frac{1}{12n_2} + \sum_{i=1}^{n_2}\left(F(t_{2i:n_2}|\alpha_2,\lambda,\theta) - \frac{2i-1}{2n_2}\right)^2 \tag{30}$$

$$C_0(t;\alpha_3,\lambda,\theta) = \frac{1}{12n_0} + \sum_{i=1}^{n_0}\left(F(t_{i:n_1}|\alpha_3,\lambda,\theta) - \frac{2i-1}{2n_0}\right)^2 \tag{31}$$

The estimators can be obtained by solving the above non-linear equations.

3.5. Anderson-Darling and Right-Tail Anderson-Darling Method

The Anderson-Darling (AD) test is used to detect departure of sample distribution from the assumed theoretical distribution. Furthermore, the AD test converges to asymptote very quickly. The Anderson-Darling estimators $\hat{\alpha}_{1ADE}$, $\hat{\alpha}_{2ADE}$, $\hat{\alpha}_{3ADE}$, $\hat{\lambda}_{ADE}$ and $\hat{\theta}_{ADE}$ of the parameters $\alpha_1,\alpha_2,\alpha_3,\lambda$ and θ are obtained by minimizing the following function with respect to $\alpha_1,\alpha_2,\alpha_3,\lambda$ and θ.

$$A(t_1,t_2;\alpha_1,\alpha_2,\alpha_3\lambda,\theta) = A_1(t_1;\alpha_1,\lambda,\theta)I_{t_1<t_2} + A_2(t_2;\alpha_2,\lambda,\theta)I_{t_2<t_1} + A_0(t;\alpha_3,\lambda,\theta)I_{t_1=t_2} \tag{32}$$

where, $I(\cdot)$ is an indicator function and

$$A_1(t_1;\alpha_1,\lambda,\theta) = -n_1 - \frac{1}{n_1}\sum_{i=1}^{n_1}(2i-1)\{\log F(t_{1i:n_1}|\alpha_1,\lambda,\theta) + \log\bar{F}(t_{1n_1+1-i:n_1}|\alpha_1,\lambda,\theta)\} \tag{33}$$

$$A_2(t_2;\alpha_2,\lambda,\theta) = -n_2 - \frac{1}{n_2}\sum_{i=1}^{n_2}(2i-1)\{\log F(t_{2i:n_2}|\alpha_2,\lambda,\theta) + \log\bar{F}(t_{1n_2+1-i:n_2}|\alpha_2,\lambda,\theta)\} \tag{34}$$

$$A_0(t;\alpha_3,\lambda,\theta) = -n_0 - \frac{1}{n_0}\sum_{i=1}^{n_0}(2i-1)\{\log F(t_{i:n_0}|\alpha_3,\lambda,\theta) + \log\bar{F}(t_{1n_0+1-i:n_0}|\alpha_3,\lambda,\theta)\} \tag{35}$$

These estimators can also be obtained by solving the non-linear equations iteratively.

The Right-tail Anderson-Darling (RAD) estimators $\hat{\alpha}_{1RADE}$, $\hat{\alpha}_{2RAD}$, $\hat{\alpha}_{3RAD}$, $\hat{\lambda}_{RAD}$ and $\hat{\theta}_{RAD}$ of the parameters $\alpha_1,\alpha_2,\alpha_3,\lambda$ and θ are obtained by minimizing the following function with respect to the respective parameters.

$$R(t_1,t_2;\alpha_1,\alpha_2,\alpha_3\lambda,\theta) = R_1(t_1;\alpha_1,\lambda,\theta)I_{t_1<t_2} + R_2(t_2;\alpha_2,\lambda,\theta)I_{t_2<t_1} + R_0(t;\alpha_3,\lambda,\theta)I_{t_1=t_2} \tag{36}$$

where, $I(\cdot)$ is an indicator function and

$$R_1(t_1;\alpha_1,\lambda,\theta) = \frac{n_1}{2} - 2\sum_{i=1}^{n_1}F(t_{1i:n_1}|\alpha_1,\lambda,\theta) - \frac{1}{n_1}\sum_{i=1}^{n_1}(2i-1)\log\bar{F}(t_{1n_1+1-i:n_1}|\alpha_1,\lambda,\theta) \tag{37}$$

$$R_2(t_2;\alpha_2,\lambda,\theta) = \frac{n_2}{2} - 2\sum_{i=1}^{n_2}F(t_{2i:n_2}|\alpha_2,\lambda,\theta) - \frac{1}{n_2}\sum_{i=1}^{n_2}(2i-1)\log\bar{F}(t_{2n_2+1-i:n_2}|\alpha_2,\lambda,\theta) \tag{38}$$

$$R_0(t;\alpha_3,\lambda,\theta) = \frac{n_0}{2} - 2\sum_{i=1}^{n_0}F(t_{i:n_0}|\alpha_3,\lambda,\theta) - \frac{1}{n_0}\sum_{i=1}^{n_0}(2i-1)\log\bar{F}(t_{n_0+1-i:n_0}|\alpha_3,\lambda,\theta) \tag{39}$$

4. Monte Carlo Simulation Study

This section assesses the performance of different estimation methods used to estimate the unknown parameters of the proposed distribution. To this end, we conduct a simulation study to compare the different estimation methods stated in the previous section. Simulation study provides a powerful technique for answering a broad set of questions and flexible framework to answer some specific questions. We evaluate the performance of the maximum likelihood estimation, the least squared estimation, weighted least squared estimation, maximum product spacing, Cramér-Von-Mises, Anderson-Darling, and Right-tail Anderson-Darling estimation given by Equations (14), (18), (20), (24), (28), (32) and (36), respectively.

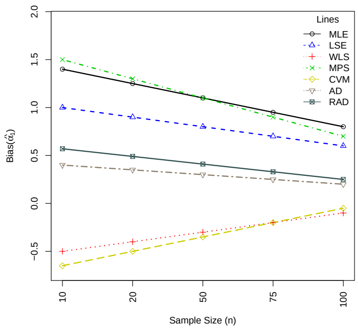
To assess the performance, we compared different estimation methods on the basis of biases, RMSE, overall and sum of ranks assuming $n = \{10, 20, 50, 75, 100\}$. The results of simulation study are depicted in Figures (1) to (9). In terms of performance, it is observed that Anderson-Darling estimators are the best as these estimators produce the least biases with least RMSE for most of the configurations considered in our studies. The next best method is the Right-Tail Anderson-Darling, followed by least square estimators. The Cramér-Von-Mises method ranked 4th, while weighted least square estimators ranked 5th, method of maximum product spacing ranked 6th among the seven methods of estimation.

It is worth mentioning that to calculate biases and RMSEs for different estimators, we first obtained the MLEs and then these estimates were used as the initial values for the other estimation methods. To obtain the MLEs, we have selected initial values very closed to the nominal values and to confirm the accuracy of the Monte Carlo simulations, we also re-run the computational code by changing the starting values for different parameter settings. It is observed that if the initial values are closed to the nominal value, the computational time is minimum as compared to the case where initial values were far away from the nominal values. However, it is also observed that despite the time of calculation, the final estimates were almost the same. In other words, to verify whether global maxima has been attained by the computer code for obtaining the estimates, different starting values have been used.

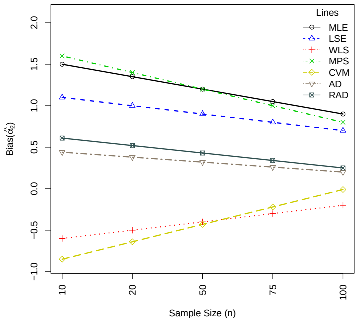
Assuming $\alpha_1 = 1, \alpha_2 = 1.5, \alpha_3 = 1.2, \lambda_1 = 1, \lambda_2 = 0.5, \lambda_3 = 1.5$ and $\theta = 0.4$, the results of biases, RMSEs, and the sum of ranks are depicted in Figures (1-3). Figure (1) shows the biases of the parameters across different sample sizes. Overall, the biases decrease by increasing sample size. In Figure (2), the RMSE of the parameters are presented. Figure (2) also shows the relation between RMSE and sample size, i.e., the RMSE decreases with increasing the sample size.

Assuming $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.2, \lambda_1 = 1.5, \lambda_2 = 0.5, \lambda_3 = 2$ and $\theta = 0.4$, the results of biases, RMSEs and ranks are depicted in Figures (4-6). Figure (4) shows the biases of the parameters across different estimation methods and sample sizes. Overall, the biases decrease by increasing the sample size which is similar to the results shown in Figure (1). To be more specific, consider Figure (4) where the red line indicates the method of WLS while blue, black and green lines indicate the LSE, MLE and MPS, respectively. These lines show the biases of the different methods of estimation on the basis of different sample sizes. In Figure (4) and Figure (5), the RMSE is depicted and it is noticed that it decreases with increasing the sample size.

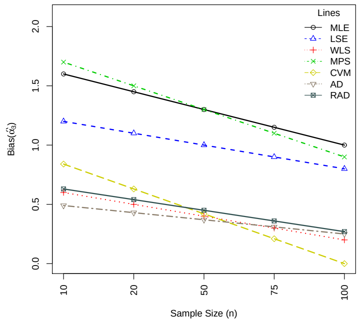
Next, we calculate the biases, RMSEs, and ranks assuming $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.2, \lambda = 1.5$ and $\theta = 0.6$ for different methods of estimation and the results are plotted in Figures (7-9). In particular, Figure (7) shows the simulation results for biases across different estimation methods and sample sizes. It can be seen from the figures that overall biases decrease by increasing sample size. In Figure (8) RMSE is shown while Figure (9) depicts overall and sum of ranks for the above parameters configuration. Figure (8) also shows the relation between RMSE and sample size, i.e., the RMSE decrease by increasing the sample size.



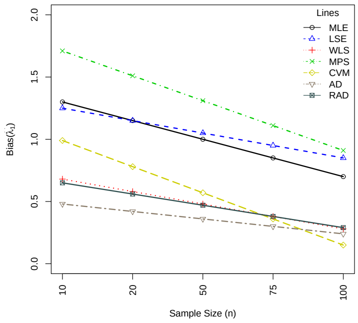
(a) $Bias(\hat{\alpha}_1)$



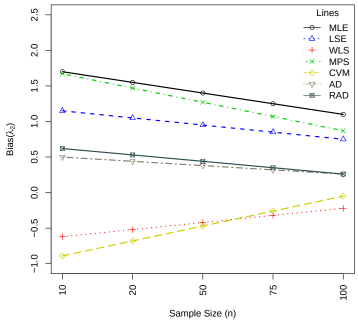
(b) $Bias(\hat{\alpha}_2)$



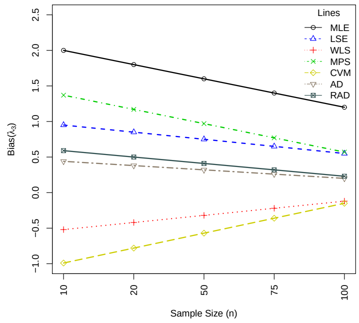
(c) $Bias(\hat{\alpha}_3)$



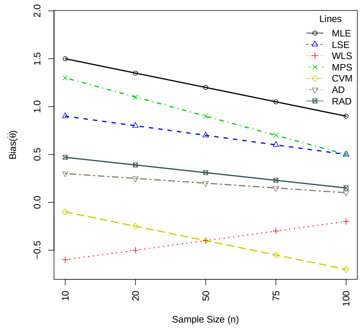
(d) $Bias(\hat{\lambda}_1)$



(e) $Bias(\hat{\lambda}_2)$

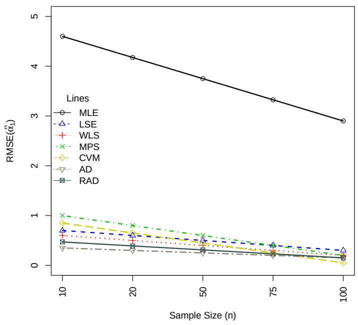


(f) $Bias(\hat{\lambda}_3)$

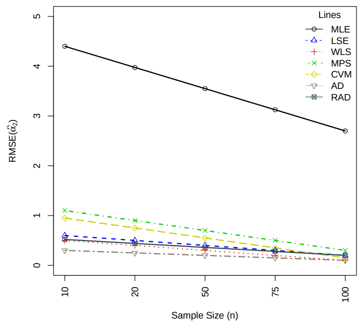


(g) $Bias(\hat{\theta})$

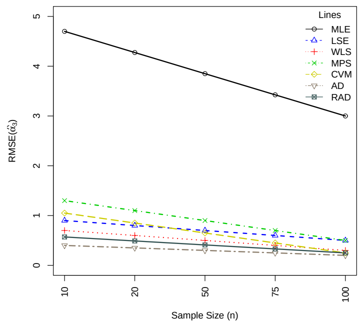
Figure 1: Bias($\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$), Bias($\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$) and Bias($\hat{\theta}$) for $n = 10, 20, 50, 75, 100$, $\alpha_1 = 1, \alpha_2 = 1.5, \alpha_3 = 1.2, \lambda_1 = 1, \lambda_2, = 0.5, \lambda_3 = 1.5$ and $\theta = 0.4$



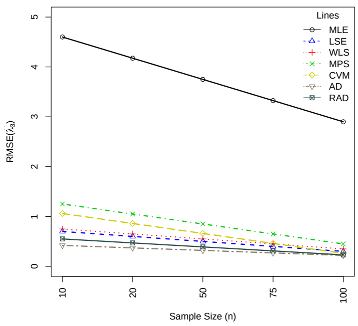
(a) $RMSE(\hat{\alpha_1})$



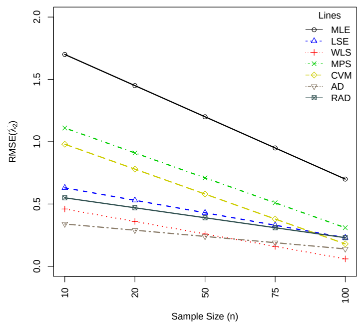
(b) $RMSE(\hat{\alpha_2})$



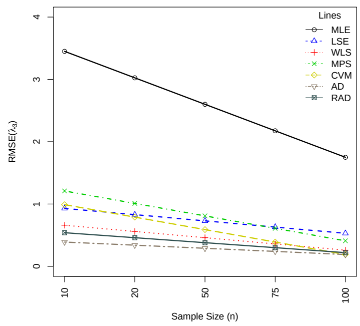
(c) $RMSE(\hat{\alpha_3})$



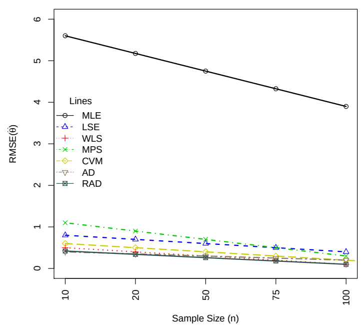
(d) $RMSE(\hat{\lambda}_1)$



(e) $RMSE(\hat{\lambda}_2)$

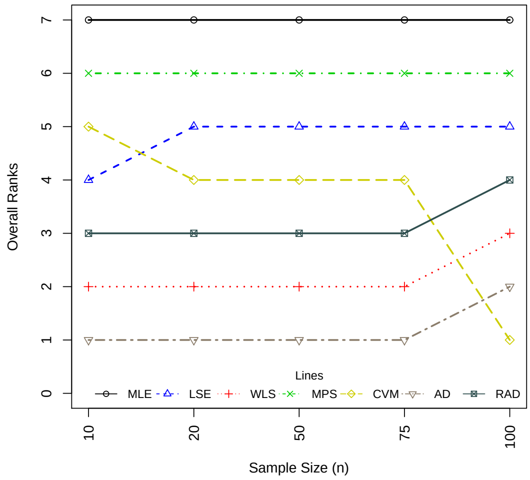


(f) $RMSE(\hat{\lambda}_3)$

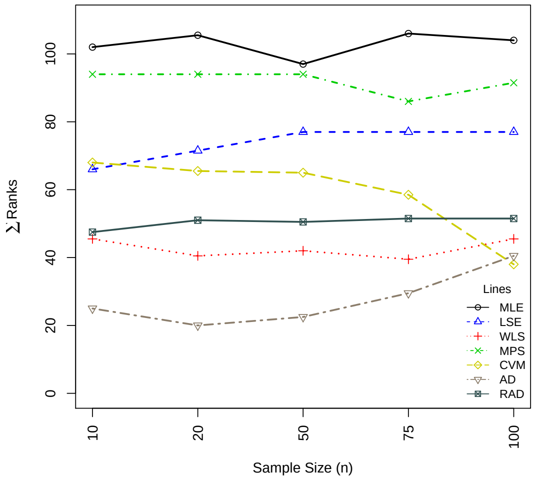


(g) $RMSE(\hat{\theta})$

Figure 2: RMSE($\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}$), RMSE($\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$) and RMSE($\hat{\theta}$) for $n = 10, 20, 50, 75, 100$, $\alpha_1 = 1, \alpha_2 = 1.5, \alpha_3 = 1.2, \lambda_1 = 1, \lambda_2, = 0.5, \lambda_3 = 1.5$ and $\theta = 0.4$

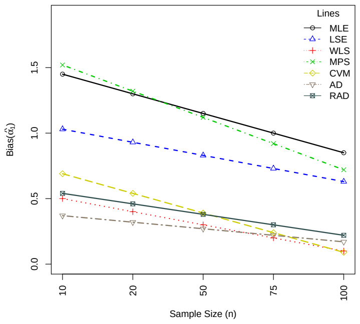


(a) Overall ranks

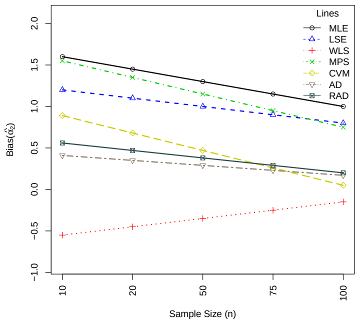


(b) Sum of Ranks

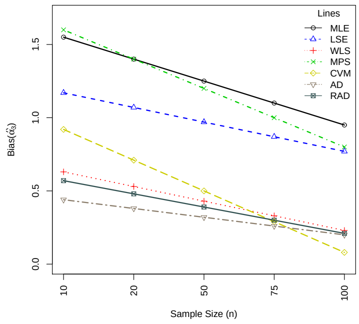
Figure 3: Overall Ranks, Sum of ranks versus $n = 10, 20, 50, 75, 100$ for $\alpha_1 = 1, \alpha_2 = 1.5, \alpha_3 = 1.2, \lambda_1 = 1, \lambda_2, = 0.5, \lambda_3 = 1.5$ and $\theta = 0.4$



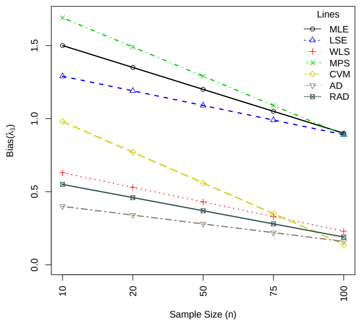
(a) $Bias(\hat{\alpha}_1)$



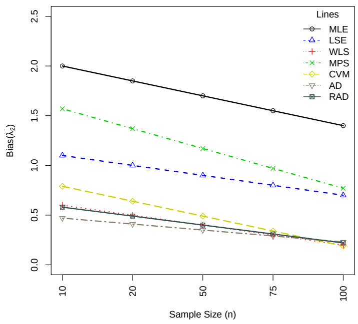
(b) $Bias(\hat{\alpha}_2)$



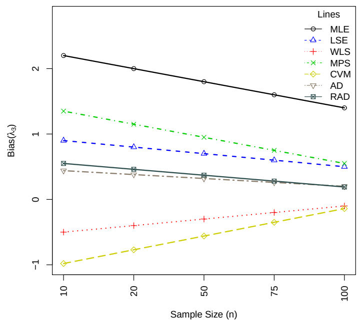
(c) $Bias(\hat{\alpha}_3)$



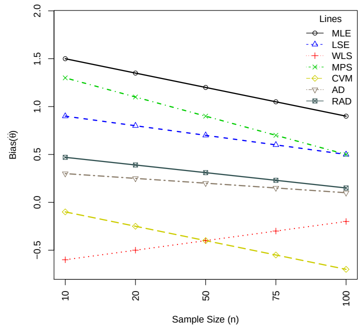
(d) $Bias(\hat{\lambda}_1)$



(e) Bias($\hat{\lambda}_2$)

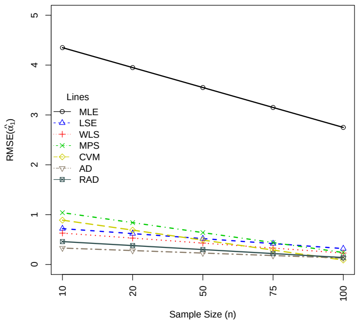


(f) Bias($\hat{\lambda}_3$)

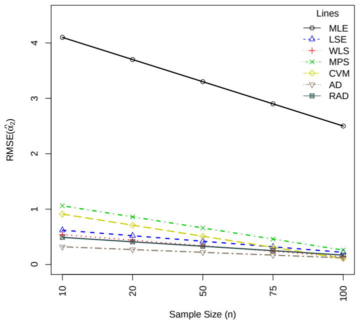


(g) $Bias(\hat{\theta})$

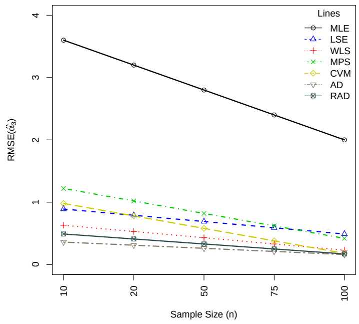
Figure 4: Bias($\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}$), Bias($\hat{\lambda_1}, \hat{\lambda_2}, \hat{\lambda_3}$) and Bias($\hat{\theta}$) versus $n = 10, 20, 50, 75, 100$ for $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.2, \lambda_1 = 1.5, \lambda_2, = 0.5, \lambda_3 = 2$ and $\theta = 0.4$



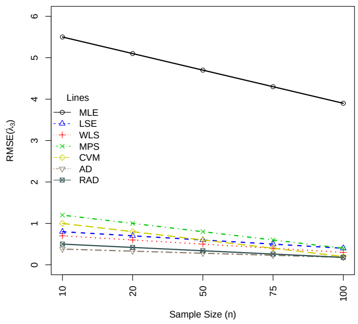
(a) RMSE($\hat{\alpha}_1$)



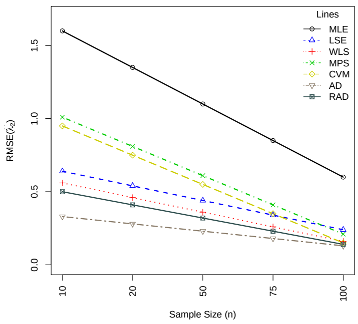
(b) RMSE($\hat{\alpha}_2$)



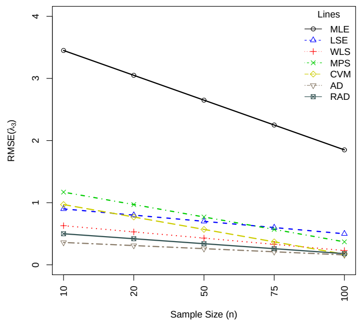
(c) RMSE($\hat{\alpha}_3$)



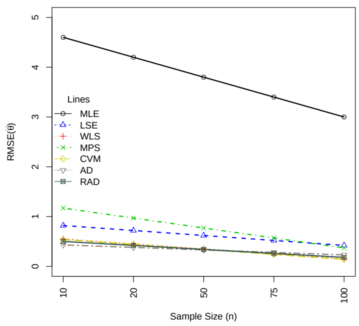
(d) RMSE($\hat{\lambda}_1$)



(e) RMSE($\hat{\lambda}_2$)

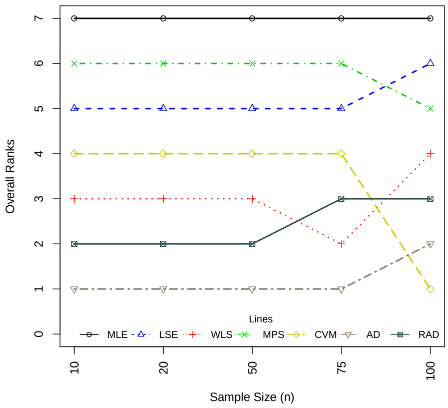


(f) RMSE($\hat{\lambda}_3$)

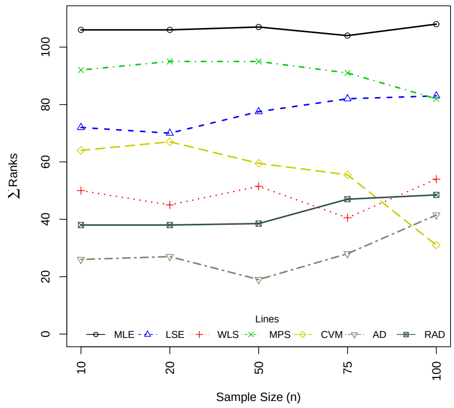


(g) RMSE($\hat{\theta}$)

Figure 5: RMSE($\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$), RMSE($\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$) and RMSE($\hat{\theta}$) for $n = 10, 20, 50, 75, 100$. $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.2, \lambda_1 = 1.5, \lambda_2, = 0.5, \lambda_3 = 2$ and $\theta = 0.4$

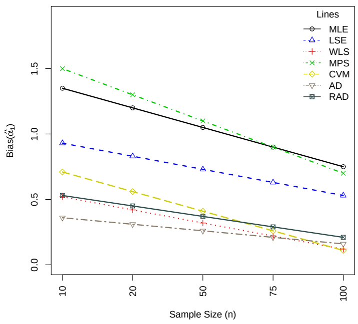


(a) Overall Ranks

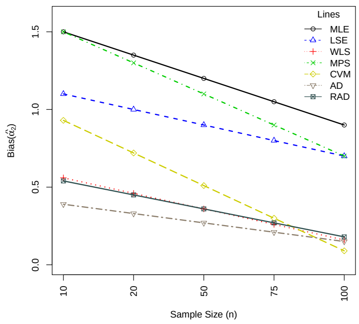


(b) Sum of Ranks

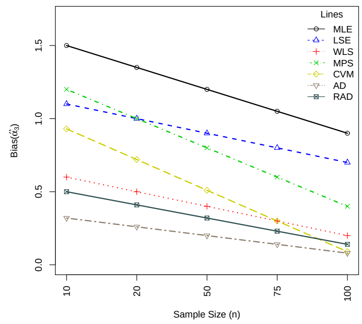
Figure 6: Overall Ranks, Sum of ranks for $n = 10, 20, 50, 75, 100$.
$\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.2, \lambda_1 = 1.5, \lambda_2, = 0.5, \lambda_3 = 2$ and $\theta = 0.4$



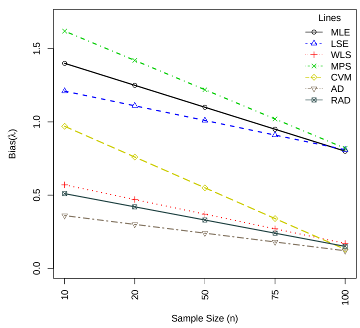
(a) $Bias(\hat{\alpha_1})$



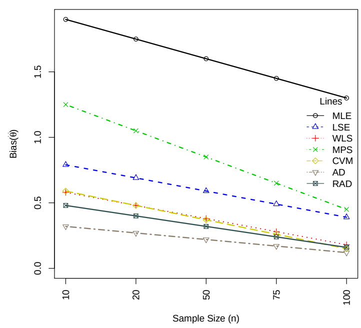
(b) $Bias(\hat{\alpha_2})$



(c) $Bias(\hat{\alpha_3})$

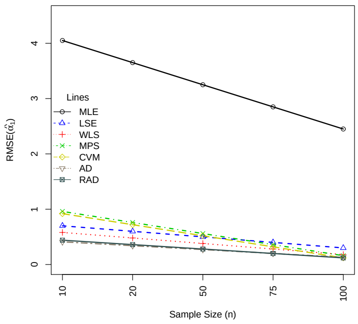


(d) $Bias(\hat{\lambda})$

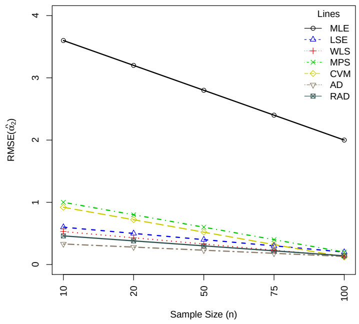


(e) Bias($\hat{\theta}$)

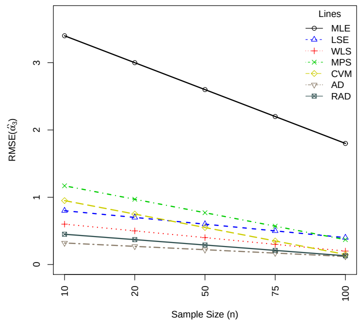
Figure 7: Bias($\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}$), Bias($\hat{\lambda}$) and Bias($\hat{\theta}$) versus $n = 10, 20, 50, 75, 100$. $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.5, \lambda = 1.5$ and $\theta = 0.6$



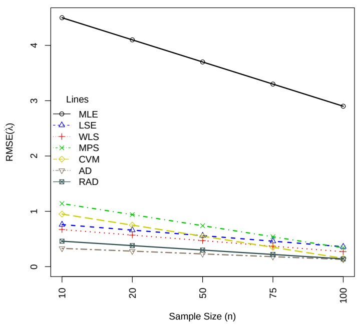
(a) RMSE($\hat{\alpha_1}$)



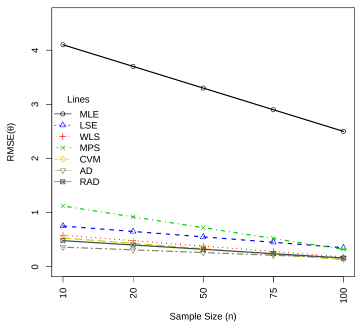
(b) RMSE($\hat{\alpha_2}$)



(c) RMSE($\hat{\alpha_3}$)

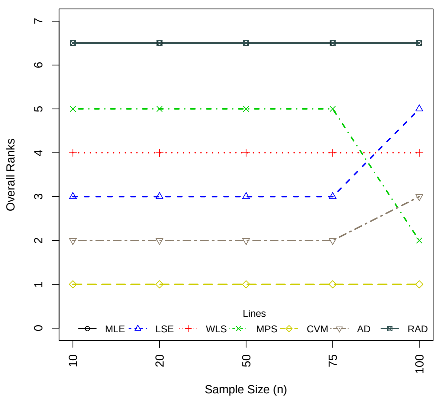


(d) RMSE($\hat{\lambda}$)

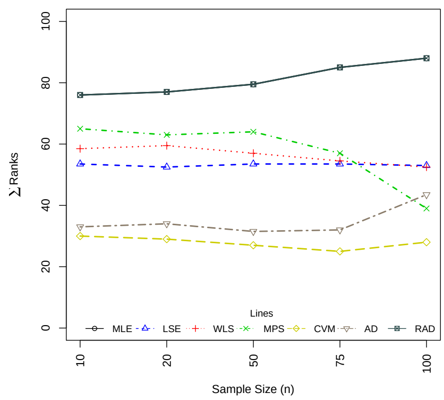


(e) RMSE($\hat{\theta}$)

Figure 8: RMSE($\hat{\alpha_1}, \hat{\alpha_2}, \hat{\alpha_3}$), RMSE($\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$) and RMSE($\hat{\theta}$) for $n = 10, 20, 50, 75, 100$. $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.5, \lambda = 1.5$ and $\theta = 0.6$



(a) Overall Ranks



(b) Sum of Ranks

Figure 9: Overall Ranks, Sum of ranks for $n = 10, 20, 50, 75, 100$. $\alpha_1 = 1.5, \alpha_2 = 2, \alpha_3 = 1.5, \lambda = 1.5$ and $\theta = 0.6$

5. Real Data Analysis

Here, the analysis of two data sets is presented to show how the BDNH can be applied in practice. Further, a comparison with two existing competing discrete bivariate distributions is also given in this section. First, we check whether the considered data sets actually come from the BDNH distribution or not by model selection tests and compare the fits with the bivariate discrete Weibull (BDW) and bivariate discrete generalized exponential. In order to compare the distributions, we consider Akaike Information Criterion (AIC) [1], Bayesian information criterion (BIC) and log-likelihood. To estimate the parameters using real data sets, we set the initial values very close to the estimated parameter values of the BDW [14]. The reason of considering BDW is the closeness between BDW and BDNH distributions. After estimating the parameters, we considered them as the true values and then by taking parametric bootstrap [7] samples of size (n), where n is the size of data. We have calculated the Bias and RMSE of the estimates. The steps to calculate the Bias and RMSE are as follows.

1. Let $X_1^*, X_2^*, \cdots, X_n^*$ denote the bootstrap samples of size n from BDNH($\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\lambda}, \hat{\theta}$).
2. Compute the MLEs denoted by $\hat{\alpha}_1^*, \hat{\alpha}_2^*, \hat{\alpha}_3^*, \hat{\lambda}^*, \hat{\theta}^*$
3. Using the results of Step 2 as the initial values, compute the estimators for other considered methods.
4. Repeat Steps 1-3 R times to obtain a set of bootstrap samples to estimate the desired parameters, say, $\{\hat{\Theta}_i^*, i = 1, 2, \cdots, R\}$.
5. In the case of an arbitrary parameter $\hat{\Theta}$, the estimated bias and RMSE are computed as:

$$\text{Bias}(\hat{\Theta}^*) = \frac{1}{R}\sum_{i=1}^{R}(\hat{\Theta}_i^* - \hat{\Theta}), \qquad \text{RMSE}(\hat{\Theta}^*) = \sqrt{\frac{1}{R}\sum_{i=1}^{R}(\hat{\Theta}_i^* - \hat{\Theta})^2} \tag{40}$$

where $\hat{\Theta}_i^*$ is the estimated parameter value from the i-th bootstrap sample, generated from BDNH($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\lambda}$, $\hat{\theta}$) and $\hat{\Theta}$ as the true value.

5.1. Football Data

The first data set to be analyzed represents the Italian Series A football match score played between two Italian football giants, ACF Fiorentina (T_1) and Juventus (T_2) during the period 1996 to 2011 [14]. The data set is given in Table (1).

Table 1: UEFA Champion's League data

Obs.	ACF Firontina (T_1)	Juventus (T_2)	Obs.	ACF Firontina (T_1)	Juventus (T_2)
1	1	2	2	0	0
3	1	1	4	2	2
5	1	1	6	0	1
7	1	1	8	3	2
9	1	1	10	2	1
11	1	2	12	3	3
13	0	1	14	1	2
15	1	1	16	1	3
17	3	3	18	0	1
19	1	1	20	1	1
21	1	0	22	3	0
23	1	2	24	1	1
25	0	1	26	0	1

On football data set, we fit the bivariate discrete Nadarajah and Haghighi (BDNH) distribution, bivariate discrete Weibull (BDW) distribution, and bivariate generalized exponential (BDGE) distribution and the results are tabulated in Table (2).

To select the most suitable estimation method for the football data set, we have calculated the Bias and RMSEs of different parameters of the model assuming different methods of estimation discussed previously. From the results tabulated in Table (2), it is noticed that the AD method is the best parameter estimation method, as it has the lowest bias and RMSE. Furthermore, the proposed model is the best fitted as compared to other assumed distributions, because BDNH has the lowest AIC and BIC values. Furthermore, the AD is the best method for the parameter estimation of the BDW distribution while the CVM for the BDGE distribution. It is also noticed that the BDW distribution fits better to the football data than the BDGE distribution.

Table 2: Estimate of the parameters of BDNH, BDW and BDGE Distributions for the football data.

Est.	MLE	LSE	WLS	MPS	CVM	AD	RAD
			BDNH ($\lambda \neq 1$)				
Bias($\hat{\alpha_1}$)	0.231	0.252	0.222	0.235	0.231	0.229	0.236
RMSE($\hat{\alpha_1}$)	0.825	0.765	0.736	0.725	0.722	0.719	0.725
Bias($\hat{\alpha_2}$)	0.220	0.191	0.195	0.191	0.186	0.185	0.182
RMSE($\hat{\alpha_2}$)	0.864	0.242	0.231	0.220	0.218	0.219	0.228
Bias($\hat{\alpha_3}$)	1.090	0.900	0.916	0.863	0.843	0.841	0.837
RMSE($\hat{\alpha_3}$)	1.362	0.313	0.310	0.311	0.310	0.308	0.380
Bias($\hat{\lambda}_1$)	0.960	1.061	1.161	1.088	1.077	1.074	1.075
RMSE($\hat{\lambda}_1$)	0.552	0.450	0.439	0.440	0.437	0.436	0.454
Bias($\hat{\lambda}_2$)	0.671	0.652	0.671	0.649	0.637	0.635	0.633
RMSE($\hat{\lambda}_2$)	0.233	0.184	0.180	0.188	0.186	0.185	0.197
Bias($\hat{\lambda}_3$)	0.251	0.202	0.211	0.208	0.210	0.208	0.204
RMSE($\hat{\lambda}_3$)	0.356	0.300	0.310	0.320	0.318	0.318	0.380
Bias($\hat{\theta}$)	0.344	0.335	0.350	0.336	0.337	0.336	0.337
RMSE($\hat{\theta}$)	1.633	0.652	0.632	0.624	0.622	0.620	0.641
AIC	655.632	650.631	651.623	652.563	650.321	650.319	652.391
BIC	654.241	651.251	651.861	653.091	650.411	650.409	653.400
Log-like	-432.637	-431.357	-430.372	-429.896	-425.367	-432.873	-432.764
			BDNH ($\lambda = 1$)				
Bias($\hat{\alpha_1}$)	0.783	0.864	0.861	0.728	0.758	0.775	0.774
RMSE($\hat{\alpha_1}$)	0.227	1.197	1.195	1.190	1.220	1.217	1.218
Bias($\hat{\alpha_2}$)	1.633	0.835	0.781	0.687	0.665	0.675	0.678
RMSE($\hat{\alpha_2}$)	1.735	0.701	0.690	0.682	0.677	0.675	0.679
Bias($\hat{\alpha_3}$)	1.486	0.766	0.701	0.847	0.817	0.806	0.817
RMSE($\hat{\alpha_3}$)	1.736	0.700	0.703	0.695	0.678	0.677	0.686
Bias($\hat{\theta}$)	0.399	0.407	0.401	0.389	0.378	0.388	0.385
RMSE($\hat{\theta}$)	1.659	0.642	0.640	0.638	0.628	0.619	0.625
AIC	685.642	679.662	679.891	676.746	674.746	673.762	674.753
BIC	678.561	681.580	681.738	679.568	679.568	678.577	677.587
Log-like	-389.738	-390.161	-390.193	-389.536	-389.536	-385.567	-387.573
			BDW				
Bias($\hat{\alpha}$)	0.979	0.901	0.913	0.925	0.901	0.893	0.913
RMSE($\hat{\alpha}$)	0.811	0.810	0.807	0.806	0.805	0.806	0.842
Bias($\hat{p_0}$)	0.001	0.002	0.003	0.002	0.003	0.002	0.004
RMSE($\hat{p_0}$)	0.001	0.000	0.000	0.000	0.000	0.000	0.001
Bias($\hat{p_1}$)	0.247	0.247	0.231	0.210	0.208	0.207	0.212

RMSE($\hat{p}_1$)	0.051	0.051	0.050	0.048	0.045	0.042	0.046
Bias($\hat{p}_2$)	0.049	0.050	0.053	0.057	0.054	0.053	0.057
RMSE($\hat{p}_2$)	0.009	0.008	0.008	0.007	0.006	0.005	0.009
AIC	710.616	709.461	708.536	709.531	708.372	708.677	711.785
BIC	713.637	711.352	710.336	711.006	709.262	709.265	712.671
Log-like	-430.342	-429.234	-424.563	-425.635	-423.466	-423.487	-424.437
			BDGE				
Bias($\hat{\alpha}_1$)	0.283	1.273	1.272	1.275	1.277	1.276	1.278
RMSE($\hat{\alpha}_1$)	0.845	0.842	0.846	0.847	0.801	0.800	0.821
Bias($\hat{\alpha}_2$)	0.771	0.568	0.565	0.545	0.542	0.556	0.558
RMSE($\hat{\alpha}_2$)	0.536	0.537	0.534	0.532	0.528	0.529	0.532
Bias($\hat{\alpha}_3$)	0.036	1.046	0.043	0.041	0.039	0.040	0.042
RMSE($\hat{\alpha}_3$)	0.647	0.648	0.644	0.644	0.610	0.609	0.612
Bias($\hat{p}$)	0.341	0.336	0.338	0.337	0.336	0.339	0.337
RMSE($\hat{p}$)	0.746	0.743	0.740	0.738	0.732	0.731	0.734
AIC	712.646	711.687	710.896	710.736	709.124	709.147	712.536
BIC	714.638	713.647	712.637	712.746	710.362	710.367	712.377
Log-like	-397.647	-390.648	-391.654	-392.637	-390.237	-390.234	-395.355

In Table (3) we tabulated the sum of ranks and overall ranks to evaluate the performance of different estimation methods. One can notice that for the BDNH and BDW, the AD is the most appropriate method of estimation while CVM for the BDGE. It is also noticed that the BDGE distribution fits better to the nasal drainage severity data than the BDW distribution.

Table 3: Sum of ranks and overall ranks for the football data set

	MLE	LSE	WLS	MPS	CVM	AD	RAD
			BDNH ($\lambda \neq 1$)				
$\sum$Ranks	74	56	59	55	42	29	54
Overall ranks	7	5	6	4	2	1	3
			BDNH ($\lambda = 1$)				
$\sum$Ranks	45	42	38	28	22	20	26
Overall ranks	7	6	5	4	2	1	3
			BDW				
$\sum$Ranks	32	28	28	24	17	12	34
Overall ranks	6	4	4	3	2	1	7
			BDGE				
$\sum$Ranks	39	40	38	31	17	22	31
Overall ranks	6	7	5	3	1	2	3

5.2. Nasal Drainage Severity Score

The second data set represents the efficacy of steam inhalation in the treatment of common cold symptoms of recent onset [14]. We analyze the data for the first two days, which are presented in Table (4).

Table 4: Nasal drainage severity score for 30 patients

Obs.	Day 1 (T_1)	DAy 2 (T_2)	Obs.	Day 1 (T_1)	Day 2 (T_2)
1	1	1	2	0	0
3	1	1	4	1	1
5	0	2	6	2	0
7	2	2	8	1	1
9	3	2	10	2	2
11	1	0	12	2	3
13	1	3	14	2	1
15	2	3	16	2	1
17	1	1	18	2	2
19	3	1	20	1	1
21	2	1	22	2	2
23	1	1	24	2	2
25	2	0	26	1	1
27	0	1	28	1	1
29	1	1	30	3	3

On Nasal drainage severity score data set [14], we fit the BDNH distribution, BDW distribution, and BDGE distribution and the results are listed in Table (5).

To assess the fitness of the proposed and other related models for this data set and to evaluate the performance of the different estimation methods, we calculate the bias, RMSE, AIC, BIC and log-likelihood reported in Table (5). It is observed from the table that the proposed model is the best as compared to other models, as BDNH has the lowest AIC and BIC compared to BDW and BDGE distributions.

Table 5: Estimate of the parameters of BDNH, BDW and BDGE distributions for the Nasal Drainage Severity Score data

Est.	MLE	LSE	WLS	MPS	CVM	AD	RAD
			BDNH ($\lambda \neq 1$)				
Bias($\hat{\alpha}_1$)	0.361	0.401	0.410	0.409	0.407	0.405	0.411
RMSE($\hat{\alpha}_1$)	0.639	0.631	0.629	0.627	0.625	0.624	0.627
Bias($\hat{\alpha}_2$)	0.632	0.645	0.640	0.637	0.638	0.634	0.632
RMSE($\hat{\alpha}_2$)	0.652	0.650	0.647	0.647	0.645	0.648	0.651
Bias($\hat{\alpha}_3$)	0.083	0.035	0.029	0.026	0.024	0.021	0.023
RMSE($\hat{\alpha}_3$)	0.037	0.039	0.037	0.038	0.036	0.037	0.039
Bias($\hat{\lambda}_1$)	0.775	0.538	0.536	0.532	0.530	0.527	0.526
RMSE($\hat{\lambda}_1$)	0.546	0.538	0.536	0.536	0.534	0.539	0.541
Bias($\hat{\lambda}_2$)	0.863	0.765	0.764	0.759	0.756	0.758	0.756
RMSE($\hat{\lambda}_2$)	0.663	0.657	0.655	0.653	0.650	0.653	0.652
Bias($\hat{\lambda}_3$)	0.585	0.573	0.570	0.567	0.564	0.565	0.563
RMSE($\hat{\lambda}_3$)	0.434	0.430	0.429	0.424	0.421	0.422	0.424
Bias($\hat{\theta}$)	0.537	0.539	0.536	0.534	0.535	0.533	0.532
RMSE($\hat{\theta}$)	0.356	0.353	0.349	0.347	0.345	0.346	0.347
AIC	633.522	632.637	631.837	630.478	629.543	628.987	629.564
BIC	634.637	6733.638	632.837	631.568	630.642	629.742	629.987
Log-like	-389.672	-388.892	-387.368	-386.674	-385.656	-386.748	-386.865

			BDNH ($\lambda = 1$)				
Bias($\hat{\alpha}_1$)	0.056	0.055	0.057	0.055	0.054	0.053	0.055
RMSE($\hat{\alpha}_1$)	0.017	0.016	0.014	0.012	0.011	0.01	0.011
Bias($\hat{\alpha}_2$)	0.002	0.004	0.005	0.004	0.003	0.002	0.004
RMSE($\hat{\alpha}_2$)	0.000	0.001	0.000	0.002	0.001	0.001	0.002
Bias($\hat{\alpha}_3$)	0.054	0.058	0.059	0.057	0.055	0.053	0.052
RMSE($\hat{\alpha}_3$)	0.015	0.014	0.012	0.010	0.009	0.008	0.010
Bias($\hat{\theta}$)	2.356	2.348	2.351	2.348	2.347	2.348	2.350
RMSE($\hat{\theta}$)	0.852	0.850	0.848	0.847	0.845	0.839	0.840
AIC	713.638	712.874	711.091	711.001	710.01	709.847	710.043
BIC	714.452	713.485	712.647	712.234	711.526	710.098	710.798
Log-like	-352.526	-351.647	-350.672	-350.125	-350.000	-349.847	-350.035
			BDW				
Bias($\hat{\alpha}$)	0.657	0.650	0.649	0.652	0.649	0.662	0.660
RMSE($\hat{\alpha}$)	0.579	0.564	0.556	0.553	0.551	0.553	0.552
Bias($\hat{p_0}$)	0.069	0.071	0.073	0.075	0.074	0.077	0.075
RMSE($\hat{p_0}$)	0.018	0.016	0.013	0.012	0.011	0.010	0.012
Bias($\hat{p_1}$)	0.003	0.002	0.004	0.007	0.008	0.009	0.007
RMSE($\hat{p_1}$)	0.001	0.000	0.001	0.002	0.003	0.004	0.006
Bias($\hat{p_2}$)	0.069	0.07	0.069	0.072	0.073	0.07	0.065
RMSE($\hat{p_2}$)	0.016	0.018	0.017	0.014	0.013	0.011	0.012
AIC	710.617	709.678	708.676	707.728	706.537	705.978	707.784
BIC	713.637	712.891	711.637	710.618	709.547	707.467	707.986
Log-like	-330.342	-329.897	-329.019	-328.26	-327.732	-326.684	-327.647
			BDGE				
Bias($\hat{\alpha}_1$)	0.876	0.870	0.862	0.861	0.842	0.837	0.826
RMSE($\hat{\alpha}_1$)	0.638	0.611	0.516	0.506	0.496	0.495	0.512
Bias($\hat{\alpha}_2$)	0.632	0.628	0.626	0.624	0.582	0.567	0.568
RMSE($\hat{\alpha}_2$)	0.472	0.470	0.447	0.439	0.432	0.431	0.435
Bias($\hat{\alpha}_3$)	0.078	0.075	0.072	0.071	0.067	0.065	0.065
RMSE($\hat{\alpha}_3$)	0.021	0.020	0.019	0.018	0.016	0.017	0.018
Bias($\hat{p}$)	0.134	0.132	0.129	0.127	0.126	0.123	0.113
RMSE($\hat{p}$)	0.114	0.113	0.112	0.111	0.109	0.108	0.107
AIC	710.564	709.638	709.627	708.637	706.676	707.766	708.627
BIC	712.535	711.627	710.736	709.637	707.623	707.736	708.003
Log-like	-334.764	-332.627	-331.627	-330.672	-328.697	-328.376	-329.627

In Table (6), we have tabulated the sum of ranks and overall ranks to evaluate the performance of different estimation methods. The results show that for BDNH, depending on the value of λ, CVM and AD are the best estimation methods. For the BDW, the WLS is the most considered appropriate methods of estimation while AD for the BDGE.

Table 6: Sum of ranks and overall ranks for the Nasal Drainage Severity Score data set

	MLE	LSE	WLS	MPS	CVM	AD	RAD
			BDNH ($\lambda \neq 1$)				
$\sum$Ranks	80	79	63	49	30	36	43
Overall ranks	7	6	5	4	1	2	3
			BDNH ($\lambda = 1$)				
$\sum$Ranks	39	36	43	31	20	12	26
Overall ranks	6	5	7	4	2	1	3

			BDW				
$\sum$Ranks	31	30	27	33	29	36	30
Overall ranks	5	3	1	6	2	7	3
			BDGE				
$\sum$Ranks	56	47	41	30	20	12	16
Overall ranks	7	6	5	4	3	1	2

5.3. Model Comparison

Previously, different estimation methods are explored to select the most appropriate estimation method for the BDW distribution. This section provides a summary of model comparisons of the BDNH, BDW, and BDGE distributions using the real data sets. To this end, we compare the AIC, BIC, and log-likelihood of these models. From Table-7, one can say that the BDNH distribution may be used to analyze the data sets more appropriately than the BDW and BDGE distributions.

Table 7: A Comparison of BDNH, BDW, and BDGE for the Real Data Sets

Data	Model	AIC	BIC	Log-likelihood
Football	**BDNH**	**650.319**	**650.409**	**-432.873**
	BDW	708.677	709.265	-423.487
	BDGE	709.147	710.367	-390.234
Nasal	**BDNH**	**629.543**	**630.642**	**-385.656**
	BDW	708.676	711.637	-329.019
	BDGE	707.766	707.736	-328.376

6. Conclusion

This study introduced a new bivariate discrete distribution, the BDNH distribution, along with some of its statistical properties. To evaluate the performance of the newly proposed bivariate discrete model, a simulation study is conducted assuming seven different methods of estimation, namely, maximum likelihood, the least squares, weighted least squares, maximum product spacing, Cramér-Von Mises, Anderson-Darling, and right tailed Anderson-Darling. The results show that the performance of our proposed model is quite satisfactory. To show practical applications of the proposed distribution, two real world data sets are considered and different accuracy measures, the bias, RMSE, AIC, BIC and log-likelihood are calculated. The results are compared with existing bivariate discrete models namely, bivariate discrete Weibull distribution and bivariate discrete generalized exponential distribution. It is observed that the new distribution is quite flexible to fit different practical applications. We also noticed that the method of maximum likelihood estimator is not performing upto expectations and therefore, biased corrected maximum likelihood estimators may be studied in the future. Further, one can extend this study by investigating other properties of the DNH and BDNH distributions that have not been discussed in this study, such as hypothesis testing and parameter estimation through Bayesian approach.

Acknowledgements

The work of first author is supported by the University Research Fund (URF) provided by the Quaid-i-Azam University, Islamabad, Pakistan. Further, the authors also acknowledge the feedback of anonymous reviewers to improve the presentation of the work.

References

[1] H. Akaike, Likelihood of a model and information criteria, Journal of Econometrics 16(1) (1981) 3–14.
[2] S. K. Ashour, A. A. Essam, H. Z. Muhammed, Moment generating function of the bivariate generalized exponential distribution, Applied Mathematical Sciences 3(59) (2009) 2911–2918.
[3] H. W. Block, A. P. Basu, A continuous bivariate exponential extension, Journal of the American Statistical Association 69(348) (1974) 1031–1037.
[4] G. Casella, R.L. Berger, Statistical Inference, (2 edition) Duxbury Pacific Grove, CA, 2002.
[5] I. Dewan, S. Nandi, An EM algorithm for the estimation of parameters of bivariate generalized exponential distribution under random left censoring, Journal of Statistical Computation and Simulation 83(9) (2013) 1648–1660.
[6] S. Dey, V.K. Sharma, M. Mesfioui, A new extension of Weibull distribution with application to lifetime data, Annals of Data Science 4(1) (2017) 31–61.
[7] B. Efron, The Jackknife, the Bootstrap, and Other Resampling Plans, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics 1982.
[8] J.E. Freund, A bivariate extension of the exponential distribution, Journal of the American Statistical Association 56(296) (1961) 971–977.
[9] H.W. Gómez, O. Venegas, Erratum to: a new family of slash-distributions with elliptical contours, Statistics & Probability Letters 78(14) (2008) 2273–2274.
[10] R.D. Gupta, D. Kundu, Bivariate generalized exponential distribution, Journal of Multivariate Analysis 41(2) (2009) 173–188.
[11] Y.A. Iriarte, N.O. Castillo, H. Bolfarine, H.W. Gómez, Modified slashed- Rayleigh distribution, Communications in Statistics-Theory and Methods 47(13) (2018) 3220–3233.
[12] Y.A. Iriarte, F. Vilca, H. Varela, H.W. Góomez, Slashed generalized Rayleigh distribution, Communications in Statistics-Theory and Methods 46(10) (2017) 4686–4699.
[13] D. Kundu, V. Nekoukhou, Univariate and bivariate geometric discrete generalized exponential distributions, Journal of Statistical Theory and Practice 12 (2018) 595–614.
[14] D. Kundu, V. Nekoukhou, On bivariate discrete Weibull distribution, Communications in Statistics-Theory and Methods 48(14) (2019) 3464–3481.
[15] H. Lee, J.H. Cha, On two general classes of discrete bivariate distributions, The American Statistician 69(3) (2015) 221–230.
[16] F. Louzada, V. Marchi, M. Roman, The exponentiated exponential-geometric distribution: a distribution with decreasing, increasing and unimodal failure rate, Statistics: A Journal of Theoretical and Applied Statistics 48(1) (2014) 167–181.
[17] A.W. Marshall, I. Olkin, A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika 84(3) (1997) 641–652.
[18] M.E. Morshedy, A. Khalil, Bivariate discrete exponentiated Weibull distribution: Properties and applications, arXiv preprint arXiv:1805.05199, 2018.
[19] S.M.T.K. Mostafaee, M. Mahdizadeh, A.J. Lemonte, The Marshall-Olkin extended generalized Rayleigh distribution: Properties and applications, Communications in Statistics-Theory and Methods 46(2) (2017) 653–671.
[20] S. Nadarajah, F. Haghighi, An extension of the exponential distribution, Statistics: A Journal of Theoretical and Applied Statistics 45(6) (2011) 543–558.
[21] S. Nadarajah, S. Kotz, The beta exponential distribution, Reliability Engineering & System Safety 91(6) (2006) 689–697.
[22] V. Nekoukhou, M.H. Alamatsaz, H. Bidram, Discrete generalized exponential distribution of a second type, Statistics: A Journal of Theoretical and Applied Statistics 47(4) (2013) 876–887.
[23] V. Nekoukhou, D. Kundu, Bivariate discrete generalized exponential distribution, Statistics: A Journal of Theoretical and Applied Statistics 51(5) (2017) 1143–1158.
[24] E. Omey, L.D. Minkova, Bivariate geometric distributions, Journal of the Bulgarian Academy 67(9) (2014) 1201–1210.
[25] J. Reyes, H.W. Gómez, H. Bolfarine, Modified slash distribution, Statistics: A Journal of Theoretical and Applied Statistics 47(5) (2013) 929–941.
[26] M.M. Ristić, D. Kundu, Generalized exponential geometric extreme distribution, Journal of Statistical Theory and Practice 10(1) (2016) 179–201.
[27] A.M. Sarhan, N. Balakrishnan, A new class of bivariate distributions and its mixture, Journal of Multivariate Analysis 98(7) (2007) 1508–1527.
[28] J.J. Swain, S. Venkatraman, J.R. Wilson, Least-squares estimation of distribution functions in Johnson's translation system, Journal of Statistical Computation and Simulation 29(4) (1988) 271–297.
[29] M.H. Tahir, S. Nadarajah, Parameter induction in continuous univariate distributions: Well-established g families, Anais da Academia Brasileira de Ciências 87(2) (2015) 539–568.