AIMS Mathematics, 9(4): 9394–9418.
DOI: 10.3934/math.2024458
Received: 27 December 2023
Revised: 03 February 2024
Accepted: 20 February 2024
Published: 07 March 2024

http://www.aimspress.com/journal/Math

Research article

A discrete extension of the Burr-Hatke distribution: Generalized hypergeometric functions, different inference techniques, simulation ranking with modeling and analysis of sustainable count data

Khaled M. Alqahtani[1], **Mahmoud El-Morshedy**[1,4], **Hend S. Shahen**[2,4] **and Mohamed S. Eliwa**[3,4,*]

[1] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[2] Department of Mathematics, Misr Institute for Computer Science, Egypt
[3] Department of Statistics and Operation Research, College of Science, Qassim University, Buraydah 51482, Saudi Arabia
[4] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

* **Correspondence:** Email: m.eliwa@qu.edu.sa.

Abstract: The intertwining relationship between sustainability and discrete probability distributions found its significance in decision-making processes and risk assessment frameworks. Count data modeling and its practical applications have gained attention in numerous research studies. This investigation focused on a particular discrete distribution characterized by a single parameter obtained through the survival discretization method. Statistical attributes of this distribution were accurately explicated using generalized hypergeometric functions. The unveiled characteristics highlighted its suitability for analyzing data displaying "right-skewed" asymmetry and possessing extended "heavy" tails. Its failure rate function effectively addressed scenarios marked by a consistent decrease in rates. Furthermore, it proved to be a valuable tool for probabilistic modeling of over-dispersed data. The study introduced various estimation methods such as maximum product of spacings, Anderson-Darling, right-tail Anderson-Darling, maximum likelihood, least-squares, weighted least-squares, percentile, and Cramer-Von-Mises, offering comprehensive explanations. A ranking simulation study was conducted to evaluate the performance of these estimators, employing ranking techniques to identify the most effective estimator across different sample sizes. Finally, real-world sustainability engineering and medical datasets were analyzed to demonstrate the significance and application of the newly introduced model.

Keywords: statistical model; survival discretization technique; generalized hypergeometric function; estimation methods; computer simulation; sustainable extreme count data
Mathematics Subject Classification: 62E99, 62E15

1. Introduction

In the analysis of real-world sustainability data, it is common to utilize continuous random distributions like the Burr-Hatke exponential (BHE) distribution. However, there are instances where the measurement of lifetimes is discrete, such as recording survival time in months or weeks. In such cases, employing a discrete random variable is more suitable. Additionally, practical problems in engineering and applied sciences often involve count phenomena, like the number of earthquakes in a year, accidents at a location, doctor visits, or insurance claims. Despite the availability of various established discrete models, there is a continued need for more flexible distributions that can effectively capture the diverse characteristics of sustainability datasets. This includes factors like asymmetry, under or over-dispersion, and variations in the failure rate function. Recognizing the significance of discrete probability models in our previous survey, we have developed and extensively explored a discrete probability distribution. This new model serves as the discrete counterpart to the BHE distribution. The BHE distribution has gained widespread utility in reliability analysis, survival modeling, and risk assessment due to its versatility in capturing diverse data patterns. Known for its flexibility in modeling right-skewed and heavy-tailed data, the BHE model is well-suited for characterizing a broad spectrum of real-world phenomena. Its adaptability extends to applications in survival analysis, providing a valuable tool for researchers to effectively model complex datasets and gain a deeper understanding of the underlying mechanisms governing observed events. For additional information and in-depth details about the BHE distribution, please refer to the citation labeled as [1]. If the expression for the survival function (SF) and probability density function (PDF) of a random variable X conforms to the following, it is recognized as adhering to the BHE distribution

$$S(x;\lambda) = \frac{e^{-\lambda x}}{1+\lambda x};\ \lambda > 0,\ x > 0, \tag{1.1}$$

and

$$g(x;\lambda) = \lambda e^{-\lambda x}\frac{2+\lambda x}{(1+\lambda x)^2};\ \lambda > 0,\ x > 0, \tag{1.2}$$

respectively, where $\lambda > 0$ is a scale parameter. In accordance with survival discretization techniques, one can derive a discrete BHE (DBHE) distribution. Survival discretization techniques are a set of statistical methods used to transform continuous probability distributions, such as the BHE distribution, into discrete versions suitable for practical applications. These techniques are particularly valuable when dealing with real-world data, which is often recorded in discrete units or intervals. By means of this process, the probability mass function can be obtained as

$$\Pr(X = x;.) = S(x;.) - S(x+1;.);\ x = 0,1,2,3,.... \tag{1.3}$$

Several discrete distributions have been suggested and examined, utilizing the discrete survival function and other techniques as a foundation, including: Discrete Burr-Hatke [2], discrete linear exponential [3], discrete Pareto [4], discrete inverse Rayleigh [5], discrete inverse Weibull [6], discrete Lindley [7], new discrete extended Weibull [8], discrete generalized geometric [9], discrete Gompertz [10], discrete generalized exponential type II [11], an overview of discrete models for fitting COVID-19 datasets [12], discrete Ramos-Louzada [13], discrete generalized Rayleigh [14], and discrete Marshall-Olkinin [15], as well as the references cited within.

The structure of this article is as follows: In Section 2, we introduce the DBHE distribution, developed through the survival discretization approach. Section 3 explores a range of statistical properties. Section 4 delves into the estimation of distribution parameters using various methods. In Section 5, we present a comprehensive simulation study based on ranking techniques. Section 6 demonstrates the versatility of the DBHE distribution by analyzing different datasets. Finally, Section 7 offers concluding remarks summarizing the findings presented in this paper.

2. The structural characteristics of the DBHE distribution

Using Eqs (1.1) and (1.3), the SF for the DBHE distribution is expressed as

$$S\left(x;\beta\right)=\frac{\beta^{x+1}}{1-(x+1)\ln\beta};\ x\in\mathbb{N}_0, \tag{2.1}$$

where $0<\beta=e^{-\lambda}<1$ and $\mathbb{N}_0=0,1,2,3,....$ The behavior of the SF is described by

$$S\left(x;\beta\right)=\begin{cases}\frac{\beta}{1-\ln\beta};\ x=0\\ 1;\quad \beta\to 1.\end{cases} \tag{2.2}$$

The associated cumulative distribution function (CDF) and probability mass function (PMF) for (2.1) can be formulated as follows:

$$F(x;\beta)=1-\frac{\beta^{x+1}}{1-(x+1)\ln\beta};\ x\in\mathbb{N}_{0,} \tag{2.3}$$

and

$$\Pr\left(X=x;\beta\right)=\beta^{x}\left[\frac{1}{1-x\ln\beta}-\frac{\beta}{1-(x+1)\ln\beta}\right];\ x\in\mathbb{N}_{0,} \tag{2.4}$$

respectively, where β controls the shape of the distribution. The behavior of the PMF is given by

$$\Pr\left(X=x;\beta\right)=\begin{cases}1-\frac{\beta}{1-\ln\beta};\ x=0\\ 0;\qquad\quad \beta\to 1.\end{cases} \tag{2.5}$$

Figure 1 displays the PMF plots for different values of the parameter β.

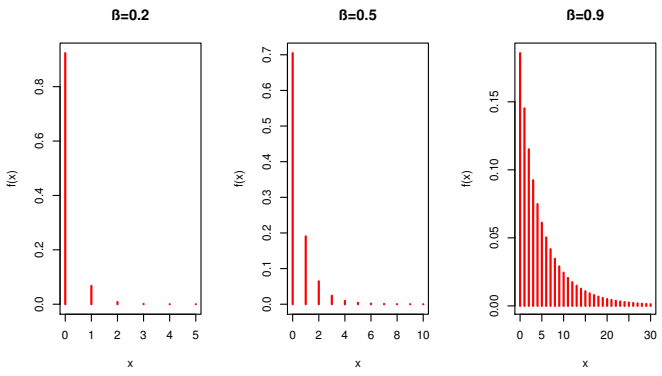

Figure 1. The PMF plots of the DBHE model.

It is worth noting that the PMF is highly effective for modeling unimodal-shaped data. Furthermore, it can also be applied to analyze asymmetric "positively-skewed" data, showcasing its versatility in capturing various data patterns. The hazard rate function (HRF) can be formulated as

$$h(x;\beta) = 1 - \frac{\beta(1 - x\ln\beta)}{1 - (x+1)\ln\beta};\ x \in \mathbb{N}_0. \tag{2.6}$$

The reversed hazard rate function (RHRF) is expressed as follows:

$$r(x;\beta) = \frac{\beta^x\left[\frac{1}{1-x\ln\beta} - \frac{\beta}{1-(x+1)\ln\beta}\right]}{1 - \frac{\beta^{x+1}}{1-(x+1)\ln\beta}};\ x \in \mathbb{N}_0. \tag{2.7}$$

The hazard rate is a measure of an item's death rate at a specific age x and is a component of the broader hazard function equation. This equation evaluates the probability that an item, having survived up to a certain time t, will continue to endure beyond that point. In essence, it quantifies the likelihood that an item surviving one moment will persist to the next one. The hazard rate is particularly relevant to non-repairable items and is sometimes referred to as the failure rate. Its significance extends to the design of secure systems in various domains such as commerce, engineering, finance, insurance, and regulatory industries. It can be expressed as a ratio of probability density to its corresponding survival function. Conversely, the reversed hazard rate of a random life is defined as the ratio between the life probability density and its distribution function. This concept holds significance in the analysis of censored data and finds applications in fields such as forensic sciences. Figure 2 illustrates the HRF and RHRF plots for varying values of the parameter β.

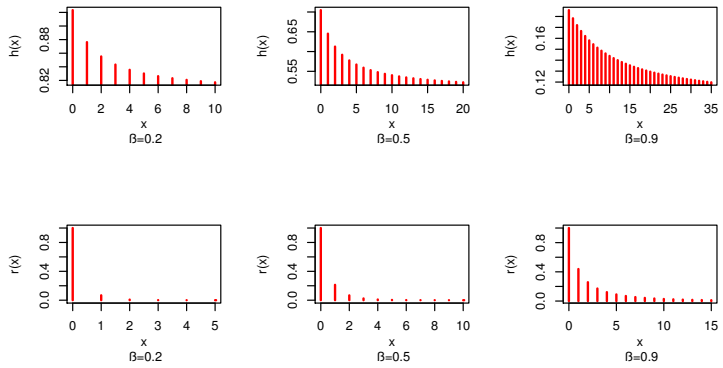

Figure 2. The HRF and RHRF of the DBHE distribution.

The observation of decreasing HRF and RHRF carries significant implications across multiple disciplines. This includes reliability engineering, where it signifies a decrease in system failure rates over time, healthcare, where it indicates improving survival probabilities, finance, where it suggests decreasing default probabilities, environmental sciences, where it hints at slowed environmental degradation, manufacturing, where it implies improved product quality, and public policy, where it informs safety measures and disaster preparedness, highlighting the importance of statistical analysis and hazard rate modeling for informed decision-making and process optimization in risk assessment and reliability domains.

3. Some statistical features

3.1. Statistical moments and associated notions

The moment generating function (MGF) and cumulant generating function (CGF) are essential tools in probability theory and statistics, offering valuable insights and advantages in various aspects of statistical analysis and probability modeling. Consider X as a random variable conforming to the DBHE distribution. The MGF, denoted as $\Pi_X(t)$, and the CGF, denoted as $K_X(t)$, can be represented in terms of generalized hypergeometric functions as follows:

$$\begin{aligned}\Pi_X(t;\beta) &= \sum_{x=0}^{\infty} e^{tx} \Pr(X = x;\beta) \\ &= (1 - \frac{\beta}{1-\ln\beta})\, hypergeom([1,\lambda_1,\lambda_2],[\lambda_3,\lambda_4], e^t\beta),\end{aligned} \tag{3.1}$$

and

$$\begin{aligned}K_X(t;\beta) &= \ln(\Pi_X(t;\beta)) \\ &= \ln\left(1 - \frac{\beta}{1-\ln\beta}\right) + \ln\left(hypergeom([1,\lambda_1,\lambda_2],[\lambda_3,\lambda_4], e^t\beta)\right),\end{aligned} \tag{3.2}$$

where $\lambda_1 = \frac{-1}{\ln\beta}$, $\lambda_2 = \frac{(\beta-2)\ln\beta+1-\beta}{(-1+\beta)\ln\beta}$, $\lambda_3 = \frac{-1+2\ln\beta}{\ln\beta}$, and $\lambda_4 = \frac{1-\ln\beta-\beta}{(-1+\beta)\ln\beta}$. The equation represented by (3.1) can be derived using the Maple software, utilizing the *hypergeom*(.) function, which is a generalized hypergeometric function. This mathematical function finds applications across diverse fields such as complex analysis, differential equations, and statistical mechanics. Renowned for its role as a solution to the hypergeometric differential equation, it is extensively employed in expressing solutions to problems characterized by symmetry, particularly those featuring spherical or cylindrical symmetry. The initial four moments of the DBHE distribution can be formulated as follows:

$$E(X) = A\, hypergeom([2,B,C],[D,E],\beta), \tag{3.3}$$

$$E\left(X^2\right) = A\, hypergeom([2,2,B,C],[1,D,E],\beta), \tag{3.4}$$

$$E\left(X^3\right) = A\, hypergeom([2,2,2,B,C],[1,1,D,E],\beta), \tag{3.5}$$

and

$$E\left(X^4\right) = A\, hypergeom([2,2,2,2,B,C],[1,1,1,D,E],\beta), \tag{3.6}$$

where $A = \frac{\beta\left[(-2+\beta)\ln\beta+1-\beta\right]}{1+2(\ln\beta)^2-3\ln\beta}$, $B = \frac{-1+\ln\beta}{\ln\beta}$, $C = \frac{(-3+2\beta)\ln\beta+1-\beta}{(-1+\beta)\ln\beta}$, $D = \frac{-1+3\ln\beta}{\ln\beta}$, $E = \frac{(-2+\beta)\ln\beta+1-\beta}{(-1+\beta)\ln\beta}$. Let $n = [n_1, n_2, ...]$, $p = nops(n)$, $d = [d_1, d_2, ...]$, and $q = nops(d)$. The $hypergeom(n,d,z)$ calling sequence is the generalized hypergeometric function $F(n,d,z)$. This function is frequently denoted by $pFq(n,d,z)$. For the variable z, the $pFq(n,d,z)$ can be formulated as

$$pFq(n,d,z) = \sum_{k=0}^{\infty} \frac{z^n.a(n_i,k)}{k!.b(d_j,k)},$$

where

$$a(n_i, k) = \prod_{i=1}^{p} pochhammer(n_i, k) \text{ and } b(d_j, k) = \prod_{j=1}^{q} pochhammer(d_j, k).$$

The Pochhammer symbol can be listed as

$$pochhammer(z, n) = z(z + 1)...(z + n - 1).$$

For additional information, please refer to the Maple software's library. Using Eqs 3.3–3.6, the variance, skewness and kurtosis can be derived as

$$var(X) = E(X^2) - [E(X)]^2, \tag{3.7}$$

$$skewness(X) = \frac{E(X^3) - 3E(X^2)E(X) + 2[E(X)]^3}{[Var(X)]^{3/2}}, \tag{3.8}$$

and

$$kurtosis(X) = \frac{E(X^4) - 4E(X)E(X^3) + 6E(X^2)[E(X)]^2 - 3[E(X)]^4}{[Var(X)]^2}. \tag{3.9}$$

Table 1 provides a compilation of numerical descriptive measures that serve as valuable tools for gaining insights into the attributes of the DBHE distribution. These measures aid researchers and analysts in comprehending aspects like central tendency, variability, shape, and other critical properties. The choice of which measures to emphasize may vary depending on the specific analysis and application.

Table 1. Numerical descriptors for characterizing the DBHE distribution.

Measure $\downarrow \beta \longrightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Mean	0.0322	0.0876	0.1702	0.2906	0.4697	0.7499	1.2305	2.2094	5.1776
Var	0.0350	0.1062	0.2299	0.4458	0.8399	1.6249	3.4397	8.9575	40.5734
Skewness	6.3536	4.3745	3.5841	3.1582	2.9003	2.7370	2.6342	2.5731	2.5420
Kurtosis	49.4831	27.1804	20.3660	17.1898	15.4389	14.4014	13.7775	13.4179	13.2383

Based on the information in Table 1, it's evident that as β approaches 1, the mean and variance of the DBHE distribution exhibit an increase, whereas the skewness and kurtosis experience a decrease. Moreover, the presented model demonstrates its capability to effectively model distributions that are positively skewed and leptokurtic in nature. Leptokurtic is a statistical term used to describe a distribution that has heavier tails and a sharper peak (higher kurtosis) compared to a normal distribution. This indicates that the distribution has more extreme values or outliers than a normal distribution, leading to a higher concentration of data points in the center and in the tails. In simple terms, a leptokurtic distribution has a more peaked and less spread-out shape than a normal distribution.

3.2. Dispersion index and variation coefficient

The index of dispersion (IOD) quantifies the absolute spread of data, while the coefficient of variation (COV) gauges the relative spread. Both metrics are valuable across diverse fields like

epidemiology, finance, and quality control, where understanding data variability is crucial for decision-making. An IOD below 1 suggests underdispersion, indicating data points cluster closely around the mean. In contrast, values exceeding 1 signal overdispersion, revealing greater variability than expected by the assumed model. An IOD of 1 suggests a random distribution where spread is proportional to the mean. When interpreting the COV, a low COV indicates minor relative variability compared to the mean, while a high COV suggests significant relative variability. These measures offer essential insights for effective analysis and decision-making in various domains. Consider X as a random variable conforming to the DBHE distribution, then the IOD and the COV can be formulated as

$$\mathrm{IOD}\left(X;\beta\right)=\frac{hypergeom\left([2,2,B,C],[1,D,E],\beta\right)}{hypergeom\left([2,B,C],[D,E],\beta\right)}-A\;hypergeom\left([2,B,C],[D,E],\beta\right), \tag{3.10}$$

and

$$\mathrm{COV}\left(X;\beta\right)=\sqrt{\frac{hypergeom\left([2,2,B,C],[1,D,E],\beta\right)}{A\;\left(hypergeom\left([2,B,C],[D,E],\beta\right)\right)^{2}}-1}. \tag{3.11}$$

The statistics for the DBHE distribution, including the IOD and COV can be reported in Table 2.

Table 2. The IOD and COV of the DBHE distribution.

Measure $\downarrow \beta \longrightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
IOD	1.0964	1.2097	1.3505	1.5342	1.7881	2.1664	2.7955	4.0542	7.8363
COV	5.8351	3.7128	2.8165	2.2978	1.9512	1.6997	1.5073	1.3546	1.2302

Based on the information in Table 2, it's evident that as β approaches 1, the IOD increases while the COV decreases. Additionally, the proposed model is best suited for modeling data with overdispersion characteristics.

3.3. Order statistics and L-moment statistics

Consider a scenario where we have a set of n random variables, denoted as $X_1, X_2, ..., X_n$, which are arranged in nondecreasing order and expressed as $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$. In the context of order statistics, it's important to note that there are no constraints placed on whether these X_is are independent or identically distributed. However, many well-established results pertaining to order statistics are derived under the classical assumption that the X_is are independent and identically distributed (iid). The CDF of the ith order statistic is expressed as follows:

$$\begin{aligned} F_{i:n}\left(x;\beta\right) &= \sum_{k=i}^{n}\binom{n}{k}\left[F_i\left(x;\beta\right)\right]^k\left[1-F_i\left(x;\beta\right)\right]^{n-k} \\ &= \sum_{k=i}^{n}\sum_{j=0}^{n-k}\Phi_m^{(n,k)}\left[F_i\left(x;\beta\right)\right]^{k+j}, \end{aligned} \tag{3.12}$$

where $\Phi_m^{(n,k)}=(-1)^j\binom{n}{k}\binom{n-k}{j}$. Moreover, the associated PMF of the ith order statistic is given by

$$f_{i:n}(x;\beta) = F_{i:n}(x;\beta) - F_{i:n}(x-1;\beta)$$
$$= \sum_{k=i}^{n}\sum_{j=0}^{n-k} \Phi_m^{(n,k)} \left[f_i(x;\beta)\right]^{k+j}. \tag{3.13}$$

Thus, the rth moments of $X_{i:n}$ can be expressed as

$$\mathbf{E}\left(X_{i:n}^r\right) = \sum_{x=0}^{\infty}\sum_{k=i}^{n}\sum_{j=0}^{n-k} \Psi_m^{(n,k)} x^r \left[f_i(x;\beta)\right]^{k+j}. \tag{3.14}$$

L-moments are statistical summary measures for probability distributions, introduced by [11]. They share similarities with ordinary moments but are calculated using linear functions applied to the ordered data values. The L-moment of a random variable X is expressed as follows:

$$\lambda_\delta = \frac{1}{\delta}\sum_{i-0}^{\delta-1} (-1)^i \binom{\delta-1}{i} \mathbf{E}\left(X_{\delta-i:\delta}\right). \tag{3.15}$$

Using (3.15), several statistical measures based on L-moment statistics can be computed, including: mean $=\lambda_1$, coefficient of skewness $=\frac{\lambda_3}{\lambda_2}$, and coefficient of kurtosis.$=\frac{\lambda_4}{\lambda_2}$. In summary, order statistics help organize and analyze data by arranging it in a specific order, while L-moment statistics provide robust and efficient tools for estimating distribution parameters and understanding the shape and characteristics of a distribution. Higher-order L-moments provide information about the shape and tail characteristics of the distribution. Both concepts play important roles in various statistical applications, particularly when dealing with nonparametric or nonstandard distributions.

4. Various estimation approaches

4.1. Maximum product of spacings estimator (MPSE)

In this section, we delve into the estimation of DBHE parameter through the MPSE method, utilizing a complete sample. Consider a random sample $X_1, X_2, \ldots, X_n$ drawn from the DBHE distribution. For $j = 1, 2, \ldots, m+1$, let

$$W_j(\beta) = F\left(x_{(j)}|\beta\right) - F\left(x_{(j-1)}|\beta\right),$$

be the uniform spacings of a random sample from the DBHE model, where $F\left(x_{(0)}|\beta\right) = 0$, $F\left(x_{(m+1)}|\beta\right) = 1$ and $\sum_{j=1}^{m+1} W_j(\beta) = 1$. The MPSE of β, say $\widehat{\beta}_{MPS}$, can be derived by maximizing the geometric mean of the spacings

$$V(\beta) = \left[\prod_{j=1}^{m+1} W_j(\beta)\right]^{\frac{1}{m+1}}, \tag{4.1}$$

with respect to the parameter β.

4.2. Anderson-Darling and right-tail Anderson-Darling estimators

Assume a random sample $X_1, X_2, \ldots, X_n$ drawn from the DBHE model. The Anderson-Darling estimator (ADE) is another type of minimum distance estimator. The ADE of the DBHE parameter, say $\widehat{\beta}_{AD}$, is derived by minimizing

$$AD(\beta) = -m - \frac{1}{m}\sum_{j=1}^{m}(2j-1)\left[\log F\left(x_{(j)}|\beta\right) + \log\left(1 - F\left(x_{(j)}|\beta\right)\right)\right]. \tag{4.2}$$

Concerning the parameter β, the model is subject to optimization, while the right-tail Anderson-Darling estimator (RADE) of the model parameter is achieved through minimization

$$RAD(\beta) = \frac{m}{2} - 2\sum_{j=1}^{m} F\left(x_{(j:m)}|\beta\right) - \frac{1}{m}\sum_{j=1}^{m}(2j-1)\left[\log\left(1 - F\left(x_{(m+1-j:m)}|\beta\right)\right)\right], \tag{4.3}$$

with respect to the parameter β.

4.3. Maximum likelihood estimation (MLE)

Consider a random sample $X_1, X_2, \ldots, X_n$ drawn from the DBHE model. The log-likelihood function (L) for the DBHE distribution can be represented as follows:

$$L(\underline{x}|\beta) = \ln\beta\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\ln\left[\left(\frac{1}{1 - x_i\ln\beta} - \frac{\beta}{1-(x_i+1)\ln\beta}\right)\right]. \tag{4.4}$$

Taking the derivative of the log-likelihood with respect to β and equating it to zero, we obtain

$$\frac{\partial L(\underline{x}|\beta)}{\partial\beta} = \frac{1}{\beta}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\frac{\frac{x_i}{\beta}(1 - x_i\ln\beta)^{-2} - (x_i+1)(1-(x_i+1)\ln\beta)^{-2} - (1-(x_i+1)\ln\beta)^{-1}}{(1 - x_i\ln\beta)^{-1} - \beta(1-(x_i+1)\ln\beta)^{-1}}. \tag{4.5}$$

Finding an analytical solution for this equation is not possible. Therefore, it requires the application of a numerical iterative method, like the Newton-Raphson method, within the R software, or other optimization techniques.

4.4. Weighted (least-squares) estimators

Consider a random sample from the DBHE model, with order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(m)}$. The least-squares estimator (LSE) of the DBHE parameter, denoted as $\widehat{\beta}_{LS}$, can be obtained by solving the nonlinear equation defined as follows:

$$\sum_{j=1}^{m}\left[F\left(x_{(j)}|\beta\right) - \frac{j}{m+1}\right]\Delta_\beta\left(x_{(j)}|\beta\right) = 0, \tag{4.6}$$

with respect to the parameter β, where

$$\Delta_\beta\left(x_{(i)}|\beta\right) = \frac{\partial}{\partial\beta}F\left(x_{(j)}|\beta\right). \tag{4.7}$$

Note that the solution of $\Delta_\beta\left(x_{(j)}|\beta\right)$ can be obtained numerically. The weighted LSE (WLSE), say $\widehat{\beta}_{WLS}$, can be derived by solving the nonlinear equation defined by

$$\sum_{j=1}^{m} \frac{(m+1)^2(m+2)}{j(m-j+1)}\left[F\left(x_{(j)}|\beta\right) - \frac{j}{m+1}\right]\Delta_\beta\left(x_{(j)}|\beta\right) = 0, \tag{4.8}$$

with respect to the parameter β.

4.5. Cramer-Von-Mises estimator (CVME)

The CVME arises as the disparity between the estimated CDF and the empirical CDF. Estimating the CVME of the DBHE parameter involves solving the non-linear equation defined as follows:

$$\sum_{j=1}^{m}\left[F\left(x_{(j)}|\beta\right) - \frac{2j-1}{2m}\right]\Delta_\beta\left(x_{(j)}|\alpha,\beta\right) = 0, \tag{4.9}$$

with respect to the parameter β, where $\Delta_\beta\left(x_{(j)}|\alpha,\beta\right)$ is defined in Eq (4.7).

4.6. Percentile estimator (PCE)

Consider $z_j = j/(m+1)$ to be an unbiased estimator of $F\left(x_{(j)}|\beta\right)$. Hence, the PCE of the parameter β, denoted by $\widehat{\beta}_{PC}$, can be reported by minimizing

$$P(\beta) = \sum_{j=1}^{m}\left(x_{(j)} - D\left(z_j\right)\right)^2,$$

with respect to the parameter β where $D\left(z_j\right) = F^{-1}\left(x_{(j)}|\beta\right)$ is the quantile function of the DBHE model.

5. Simulation ranking techniques: different estimators

In this segment, we assess the effectiveness of MPSE, ADE, MLE, LSE, RADE, PCE, CVME, and WLSE concerning the sample size 'n', and utilizing the R software with DEHB parameters. The process of generating a random variable X from the DEHB distribution begins by generating the value Y from the continuous distribution. Subsequently, the obtained Y value undergoes discretization to produce X, where X is defined as the greatest integer less than or equal to Y. To replicate this, we perform Markov Chain Monte Carlo (MCMC) simulations using various schemes. The assessment is carried out through a simulation study:

(1) Generate $N = 10000$ samples of various sizes "$n_i; i = 1,2,3,4,5$" from the DBHE model as follows:

- Scheme I: $\beta = 0.2 \mid n_1 = 50,\ n_2 = 150,\ n_3 = 300,\ n_4 = 700,\ n_5 = 1000.$
- Scheme II: $\beta = 0.4 \mid n_1 = 50,\ n_2 = 150,\ n_3 = 300,\ n_4 = 700,\ n_5 = 1000.$
- Scheme III: $\beta = 0.7 \mid n_1 = 50,\ n_2 = 150,\ n_3 = 300,\ n_4 = 700,\ n_5 = 1000.$
- Scheme III: $\beta = 0.9 \mid n_1 = 50,\ n_2 = 150,\ n_3 = 300,\ n_4 = 700,\ n_5 = 1000.$

(2) Compute the MPSE, ADE, MLE, LSE, RADE, PCE, CVME, and WLSE for the 10000 samples, say $\widehat{\beta_k}$ for $k = 1, 2, ..., 10000$.

(3) Caculate the bias, mean squared errors (MSE), and mean relative errors (MRE) for $N = 10000$ samples as

$$|\text{Bias}(\beta)| = \frac{1}{N}\sum_{k=1}^{N}\left|\widehat{\beta_k} - \beta_k\right|, \ \text{MSE}(\beta) = \frac{1}{N}\sum_{k=1}^{N}(\widehat{\beta_k} - \beta_k)^2, \ \text{MRE}(\beta) = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|\widehat{\beta_k} - \beta_k\right|}{\beta_k}.$$

The MSE measures the average squared difference between predicted and actual values, with a lower MSE indicating closer predictions to actual values. On the other hand, MRE expresses the average relative difference as a percentage, offering insights into accuracy and normalization across varying data magnitudes. MSE emphasizes precision by squaring errors, while MRE considers the relative magnitude of errors. MSE can be sensitive to outliers, while MRE, in percentage terms, may be less influenced. Despite MSE being less interpretable due to squared units, MRE, as a percentage, provides a standardized measure of error. The choice between MSE and MRE depends on data characteristics and the desired focus on precision or accuracy in predictions.

(4) The empirical results of simulation are reported in the Tables 3–7.

Table 3. Simulation outcomes for Scheme I.

n	Est.	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
50	|Bias|	0.34973_3	0.29618_1	0.46524_6	0.50175_7	0.37808_4	0.52773_8	0.34260_2	0.42618_5
	MSE	0.45538_2	0.44273_1	0.52777_6	0.54179_7	0.47531_4	0.58460_8	0.46015_3	0.50378_5
	MRE	0.15179_2	0.14756_1	0.17592_6	0.18060_7	0.15844_4	0.19487_8	0.15338_3	0.16793_5
Sum of Ranks		7_2	3_1	18_6	21_7	12_4	24_8	8_3	15_5
150	|Bias|	0.10324_2	0.10021_1	0.14097_7	0.14016_6	0.11351_3	0.19009_8	0.11474_4	0.13246_5
	MSE	0.25401_1	0.25596_2	0.29967_7	0.29734_6	0.26894_4	0.35137_8	0.26808_3	0.28586_5
	MRE	0.08467_1	0.08532_2	0.09989_7	0.09911_6	0.08965_4	0.11712_8	0.08936_3	0.09529_5
Sum of Ranks		4_1	5_2	21_7	18_6	11_4	24_8	10_3	15_5
300	|Bias|	0.04940_1	0.05065_2	0.07209_7	0.06848_6	0.05531_4	0.09601_8	0.05449_3	0.06583_5
	MSE	0.17905_1	0.18159_2	0.21475_7	0.20673_6	0.18875_4	0.24535_8	0.18464_3	0.20397_5
	MRE	0.05968_1	0.06053_2	0.07158_7	0.06891_6	0.06292_4	0.08178_8	0.06155_3	0.06799_5
Sum of Ranks		3_1	6_2	21_7	18_6	12_4	24_8	9_3	15_5
500	|Bias|	0.02782_1	0.02783_2	0.04208_6	0.04210_7	0.03083_3	0.05671_8	0.03135_4	0.03926_5
	MSE	0.13190_1	0.13342_2	0.16247_6	0.16354_7	0.14135_4	0.19151_8	0.13993_3	0.15853_5
	MRE	0.04397_1	0.04447_2	0.05416_6	0.05451_7	0.04712_4	0.06384_8	0.04664_3	0.05284_5
Sum of Ranks		3_1	6_2	18_6	21_7	11_4	24_8	10_3	15_5
700	|Bias|	0.02318_2	0.02030_1	0.02991_7	0.02949_6	0.02425_4	0.04188_8	0.02400_3	0.02938_5
	MSE	0.12310_3	0.11298_1	0.13755_7	0.13668_5	0.12323_4	0.16443_8	0.12302_2	0.13679_6
	MRE	0.04103_3	0.03766_1	0.04585_7	0.04556_5	0.04108_4	0.05481_8	0.04101_2	0.04560_6
Sum of Ranks		8_3	3_1	21_7	16_5	12_4	24_8	7_2	17_6
1000	|Bias|	0.01456_2	0.01404_1	0.01980_7	0.01926_5	0.01649_3	0.02909_8	0.01655_4	0.01951_6
	MSE	0.09578_2	0.09016_1	0.11259_7	0.11084_5	0.10259_3	0.13486_8	0.10420_4	0.11105_6
	MRE	0.03193_2	0.03005_1	0.03753_7	0.03695_5	0.03420_3	0.04495_8	0.03473_4	0.03702_6
Sum of Ranks		6_2	3_1	21_7	15_5	9_3	24_8	12_4	18_6

Table 4. Simulation outcomes for Scheme II.

n	Est.	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
50	|Bias|	0.80628_4	0.47446_1	1.11787_6	1.25880_8	0.76985_3	1.01214_5	0.74891_2	1.20883_7
	MSE	0.61326_2	0.55957_1	0.71997_5	0.77203_7	0.64010_4	0.78679_8	0.63223_3	0.76958_6
	MRE	0.20442_2	0.18652_1	0.23999_5	0.25734_7	0.21337_4	0.26226_8	0.21074_3	0.25653_6
Sum of Ranks		$8_{2.5}$	3_1	16_5	22_8	11_4	21_7	$8_{2.5}$	19_6
150	|Bias|	0.17246_2	0.15364_1	0.25285_5	0.26010_6	0.19976_4	0.38913_8	0.18798_3	0.29936_7
	MSE	0.32265_2	0.31331_1	0.39058_5	0.39261_6	0.35063_4	0.50374_8	0.33573_3	0.41663_7
	MRE	0.10755_2	0.10444_1	0.13019_5	0.13087_6	0.11688_4	0.16791_8	0.11191_3	0.13888_7
Sum of Ranks		6_2	3_1	15_5	18_6	12_4	24_8	9_3	21_7
300	|Bias|	0.07844_2	0.07387_1	0.12350_6	0.12035_5	0.09051_4	0.20140_8	0.08908_3	0.13782_7
	MSE	0.22091_2	0.21636_1	0.27815_6	0.26876_5	0.23424_3	0.36016_8	0.23593_4	0.29216_7
	MRE	0.07364_2	0.07212_1	0.09272_6	0.08959_5	0.07808_3	0.12005_8	0.07864_4	0.09739_7
Sum of Ranks		6_2	3_1	18_6	15_5	10_3	24_8	11_4	21_7
500	|Bias|	0.04588_2	0.04408_1	0.07387_6	0.07206_5	0.05217_4	0.12520_8	0.05047_3	0.08070_7
	MSE	0.16877_2	0.16455_1	0.21315_6	0.21211_5	0.18004_4	0.28765_8	0.17703_3	0.22255_7
	MRE	0.05626_2	0.05485_1	0.07105_6	0.07070_5	0.06001_4	0.09588_8	0.05901_3	0.07418_7
Sum of Ranks		6_2	3_1	18_6	15_5	12_4	24_8	9_3	21_7
700	|Bias|	0.03622_2	0.03134_1	0.05102_6	0.05053_5	0.03916_4	0.09451_8	0.03833_3	0.05715_7
	MSE	0.15228_2	0.13628_1	0.17847_5	0.17955_6	0.15678_4	0.24824_8	0.15529_3	0.18863_7
	MRE	0.05076_2	0.04543_1	0.05949_5	0.05985_6	0.05226_4	0.08275_8	0.05176_3	0.06288_7
Sum of Ranks		6_2	3_1	16_3	17_4	12_4	24_8	9_3	21_7
1000	|Bias|	0.02397_2	0.02164_1	0.03412_6	0.03304_5	0.02594_3	0.06449_8	0.02656_4	0.03988_7
	MSE	0.12222_2	0.10724_1	0.14774_6	0.14451_5	0.12861_3	0.20268_8	0.13117_4	0.15738_7
	MRE	0.04074_2	0.03575_1	0.04925_6	0.04817_5	0.04287_3	0.06756_8	0.04372_4	0.05246_7
Sum of Ranks		6_2	3_1	18_6	15_5	9_3	24_8	12_4	21_7

Table 5. Simulation outcomes for Scheme III.

n	Est.	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
50	|Bias|	0.27631_3	0.23145_1	0.36734_6	0.39714_7	0.29833_4	0.41941_8	0.27006_2	0.33011_5
	MSE	0.40371_2	0.39195_1	0.46807_6	0.48101_7	0.42139_4	0.52144_8	0.40801_3	0.44260_5
	MRE	0.16148_2	0.15678_1	0.18723_6	0.19241_7	0.16855_4	0.20858_8	0.16320_3	0.17704_5
Sum of Ranks		7_2	3_1	18_6	21_7	12_4	24_8	8_3	15_5
150	|Bias|	0.08084_2	0.07844_1	0.11095_7	0.11035_6	0.08925_3	0.15236_8	0.09025_4	0.10242_5
	MSE	0.22456_1	0.22672_2	0.26574_7	0.26364_6	0.23837_4	0.31444_8	0.23757_3	0.25125_5
	MRE	0.08983_1	0.09069_2	0.10629_7	0.10546_6	0.09535_4	0.12578_8	0.09503_3	0.10050_5
Sum of Ranks		4_1	5_2	21_7	18_6	11_4	24_8	10_3	15_5
300	|Bias|	0.03881_1	0.03974_2	0.05669_7	0.05389_6	0.04344_4	0.07691_8	0.04280_3	0.05081_5
	MSE	0.15865_1	0.16115_2	0.19036_7	0.18333_6	0.16723_4	0.21977_8	0.16358_3	0.17917_5
	MRE	0.06346_1	0.06446_2	0.07614_7	0.07333_6	0.06689_4	0.08791_8	0.06543_3	0.07167_5
Sum of Ranks		3_1	6_2	21_7	18_6	12_4	24_8	9_3	15_5
500	|Bias|	0.02183_1	0.02203_2	0.03309_6	0.03312_7	0.02419_3	0.04548_8	0.02461_4	0.03029_5
	MSE	0.11687_1	0.11937_2	0.14403_6	0.14503_7	0.12517_4	0.17171_8	0.12395_3	0.13921_5
	MRE	0.04675_1	0.04775_2	0.05761_6	0.05801_7	0.05007_4	0.06869_8	0.04958_3	0.05568_5
Sum of Ranks		3_1	6_2	18_6	21_7	11_4	24_8	10_3	15_5
700	|Bias|	0.01817_2	0.01604_1	0.02350_7	0.02320_6	0.01906_4	0.03369_8	0.01884_3	0.02268_5
	MSE	0.10892_2	0.10094_1	0.12192_7	0.12121_6	0.10921_4	0.14760_8	0.10900_3	0.12018_5
	MRE	0.04357_2	0.04038_1	0.04877_7	0.04849_6	0.04369_4	0.05904_8	0.04360_3	0.04807_5
Sum of Ranks		6_2	3_1	21_7	18_6	12_4	24_8	9_3	15_5
1000	|Bias|	0.01140_2	0.01125_1	0.01557_7	0.01513_6	0.01296_3	0.02338_8	0.01300_4	0.01506_5
	MSE	0.08473_2	0.08189_1	0.09979_7	0.09823_6	0.09094_3	0.12102_8	0.09233_4	0.09754_5
	MRE	0.03389_2	0.03276_1	0.03992_7	0.03929_6	0.03638_3	0.04841_8	0.03693_4	0.03902_5
Sum of Ranks		6_2	3_1	21_7	18_6	9_3	24_8	12_4	15_5

Table 6. Simulation outcomes for Scheme IV.

n	Est.	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
50	|Bias|	0.48453_4	0.30092_1	0.67474_5	0.75898_8	0.47083_3	0.67704_6	0.46383_2	0.72447_7
	MSE	0.48386_2	0.44610_1	0.56657_5	0.60566_7	0.50502_4	0.64649_8	0.50089_3	0.60264_6
	MRE	0.19354_2	0.17844_1	0.22663_5	0.24227_7	0.20201_4	0.25860_8	0.20035_3	0.24105_6
Sum of Ranks		$8_{2.5}$	3_1	15_5	$22_{7.5}$	11_4	$22_{7.5}$	$8_{2.5}$	19_6
150	|Bias|	0.10762_2	0.09761_1	0.15730_5	0.16135_6	0.12518_4	0.26601_8	0.11732_3	0.18488_7
	MSE	0.25540_2	0.25070_1	0.30892_5	0.30980_6	0.27802_4	0.41701_8	0.26621_3	0.32859_7
	MRE	0.10216_2	0.10028_1	0.12357_5	0.12392_6	0.11121_4	0.16680_8	0.10648_3	0.13144_7
Sum of Ranks		6_2	3_1	15_5	18_6	12_4	24_8	9_3	21_7
300	|Bias|	0.04951_2	0.04680_1	0.07715_6	0.07498_5	0.05662_4	0.13726_8	0.05582_3	0.08559_7
	MSE	0.17549_2	0.17350_1	0.22003_6	0.21250_5	0.18559_3	0.29748_8	0.18721_4	0.23061_7
	MRE	0.07019_2	0.06940_1	0.08801_6	0.08500_5	0.07424_3	0.11899_8	0.07488_4	0.09224_7
Sum of Ranks		6_2	3_1	18_6	15_5	10_3	24_8	11_4	21_7
500	|Bias|	0.02901_2	0.02830_1	0.04622_6	0.04503_5	0.03289_4	0.08573_8	0.03173_3	0.05024_7
	MSE	0.13427_2	0.13411_1	0.16877_6	0.16768_5	0.14296_4	0.23787_8	0.14040_3	0.17576_7
	MRE	0.05371_2	0.05364_1	0.06751_6	0.06707_5	0.05718_4	0.09515_8	0.05616_3	0.07030_7
Sum of Ranks		6_2	3_1	18_6	15_5	12_4	24_8	9_3	21_7
700	|Bias|	0.02282_2	0.02017_1	0.03196_6	0.03163_5	0.02461_4	0.06517_8	0.02409_3	0.03541_7
	MSE	0.12079_2	0.11176_1	0.14134_5	0.14209_6	0.12438_4	0.20631_8	0.12322_3	0.14866_7
	MRE	0.04832_2	0.04471_1	0.05654_5	0.05684_6	0.04975_4	0.08252_8	0.04929_3	0.05946_7
Sum of Ranks		6_2	3_1	16_5	17_6	12_4	24_8	9_3	21_7
1000	|Bias|	0.01517_2	0.01416_1	0.02140_6	0.02073_5	0.01632_3	0.04440_8	0.01674_4	0.02484_7
	MSE	0.09729_2	0.09046_1	0.11701_6	0.11438_5	0.10205_3	0.16827_8	0.10413_4	0.12430_7
	MRE	0.03892_2	0.03618_1	0.04680_6	0.04575_5	0.04082_3	0.06731_8	0.04165_4	0.04972_7
Sum of Ranks		6_2	3_1	18_6	15_5	9_3	24_8	12_4	21_7

Table 7. Ranking of estimation methods based on simulation results.

	n	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
Schema I	50	2	1	6	7	4	8	3	5
	150	1	2	7	6	4	8	3	5
	300	1	2	7	6	4	8	3	5
	500	1	2	6	7	4	8	3	5
	700	3	1	7	5	4	8	2	5
	1000	2	1	7	5	3	8	4	5
Schema II	50	2.5	1	5	8	4	7	2.5	6
	150	2	1	5	6	4	8	3	7
	300	2	1	6	5	3	8	4	7
	500	2	1	6	5	4	8	3	7
	700	2	1	3	4	4	8	3	7
	1000	2	1	6	5	3	8	4	7
Schema III	50	2	1	6	7	4	8	3	5
	150	1	2	7	6	4	8	3	5
	300	1	2	7	6	4	8	3	5
	500	1	2	6	7	4	8	3	5
	700	2	1	7	6	4	8	3	5
	1000	2	1	7	6	3	8	4	5
Schema IV	50	2.5	1	5	7.5	4	7.5	2.5	6
	150	2	1	5	6	4	8	3	7
	300	2	1	6	5	3	8	4	7
	500	2	1	6	5	4	8	3	7
	700	2	1	5	6	4	8	3	7
	1000	2	1	6	5	3	8	4	7
Sum of Ranks		44	30	144	141.5	90	190.5	76	142
Overall Rank		2	1	7	5	4	8	3	6

From Tables 3 to 7, it is evident that as the sample size 'n' increases, the bias of the parameter β tends to decrease toward zero. Similarly, both the MSE and MRE of the DBHE parameter also decrease toward zero with increasing sample size 'n'. These findings indicate the consistent performance of the derived estimators. Furthermore, all estimation methods demonstrate good performance across different sample sizes, with Table 7 highlighting that the ADE method performs the best.

6. Sustainability data analysis: goodness-of-fit

In this section, we will delve into the significance of the proposed distribution by analyzing various datasets from different domains. We will evaluate how well the DBHE distribution fits these datasets in comparison to several other competing distributions, including the discrete Pareto (DP), discrete Rayleigh (DR), discrete inverse Rayleigh (DIR), discrete Burr-Hatke (DBH), Poisson (Poi), and discrete Burr-XII (DB-XII) distributions. To assess the goodness-of-fit (GOF), we will employ various criteria, which encompass the negative log-likelihood ($-L$), Akaike information criterion (AIC), Bayesian information criterion (BIC), corrected Akaike information criterion (CAIC), Hannan-Quinn information criterion (HQIC), and the Kolmogorov-Smirnov (KS) test, along with its associated P-value. In the interpretation of AIC, CAIC, BIC, and HQIC, lower values indicate a better balance between model fit and simplicity. Consequently, the model with the lowest AIC, CAIC, BIC, and HQIC is considered the most suitable among the available options. BIC imposes a stricter penalty on complex models in comparison to AIC and CAIC, displaying a more conservative preference for selecting simpler models, especially in scenarios with smaller sample sizes. On the other hand, since there is a limited number of frequencies for each observation in datasets I, II, and IV the Pearson's Chi-square statistic cannot be employed for an inference test. Therefore, the KS measure is adequate in this case.

6.1. Dataset I: Failure times

The first dataset pertains to the time until failure of 15 electron components during an accelerated life test (refer to [16]). To explore the characteristics of dataset I, we have created nonparametric plots, which include box plots, normal quantile-quantile (Q-Q) plots, violin plots, and strip plots. For additional details and visual representations, please refer to Figure 3.

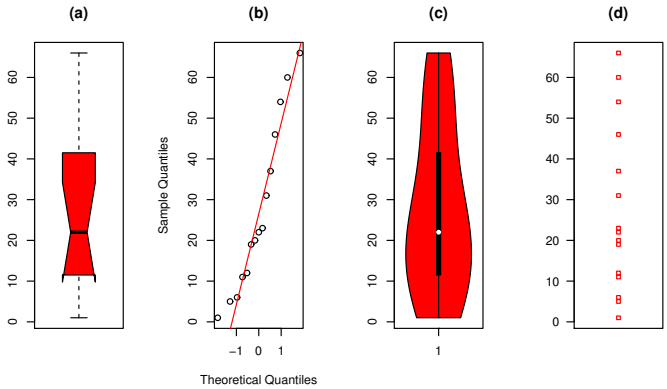

Figure 3. Nonparametric plots for dataset I.

The MLEs along with their respective SE, C.I for the parameter(s), and GOF test results for this dataset can be found in Tables 8 and 9. Notably, the values of $-L$, AIC, BIC, CAIC, HQIC, and KS are all lower, and the P-value is higher for the DBHE distribution in comparison to the values obtained for the other models. As a result, based on this analysis of the real dataset, it appears that the proposed distribution is a highly competitive model.

Figure 4 depicts the probability-probability (P-P) plot for dataset I, while Figure 5 showcases the estimated CDFs and the profile of the L for the parameter β in dataset I. Figure 5 reinforces our empirical findings, supporting the conclusion that the DBHE distribution is a more suitable fit for analyzing this data. Additionally, it highlights that the estimator for β is indeed unique.

Table 10 provides a compilation of various estimation methods applied to dataset I within the framework of the proposed model.

The analysis revealed that all estimation methods perform satisfactorily for data fitting, with the WLSE approach emerging as the most effective among them.

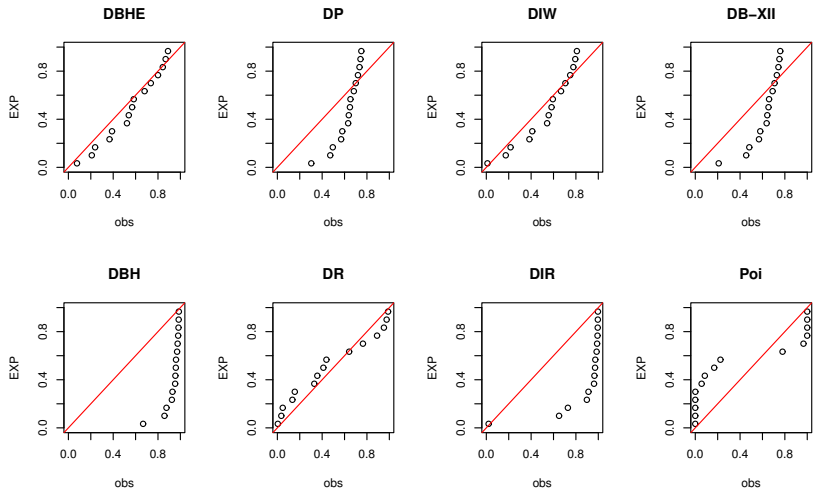

Figure 4. The P-P plot for dataset I.

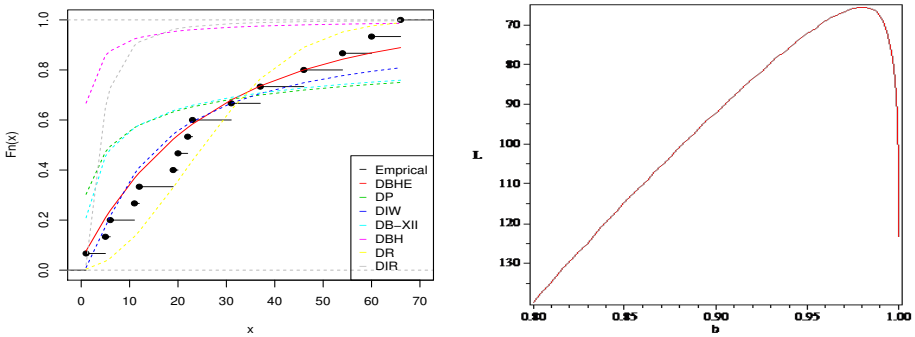

Figure 5. The estimated CDFs (left panel) and L profile of $\widehat{\beta}$ (right panel) for dataset I.

Table 8. The MLEs, standard error (SE), and confidence interval (C.I) for dataset I.

	β			α		
Model ↓ Parameter⟶	MLE	SE	C.I	MLE	SE	CI
DBHE	0.9801	0.0057	[0.9692, 0.9915]	–	–	–
DR	0.9991	2.581×10^{-4}	[0.9980, 0.9993]	–	–	–
DIR	1.801×10^{-7}	0.0552	[0, 0.1075]	–	–	–
DBH	0.9992	0.0076	[0.9843, 1.0142]	–	–	–
DPa	0.7201	0.0611	[0.6004, 0.8398]	–	–	–
Poi	27.5332	1.3553	[24.8781, 30.1892]	–	–	–
DIW	2.212×10^{-4}	7.751×10^{-4}	[0, 0.0013]	0.8752	0.1642	[0.5542, 1.1964]
DB-XII	0.9756	0.0512	[0.8743, 1]	13.3676	27.7857	[0, 67.8244]

Table 9. The GOF test for dataset I.

Statistic ↓ Parameter⟶	DBHE	DR	DIR	DBH	DPa	Poi	DIW	DB-XII
$-L$	65.5581	66.3943	89.0961	91.3684	77.4023	151.2064	68.7037	75.7245
AIC	133.1174	134.7880	180.192	184.7368	156.8047	304.4129	141.4063	155.4483
CAIC	133.4247	135.0961	180.4994	185.0445	157.1124	304.7206	142.4068	156.4480
BIC	133.8256	135.4967	180.8990	185.4448	157.5127	305.1209	142.8223	156.8645
HQIC	133.1094	134.7814	180.1841	184.7292	156.7971	304.4053	141.3919	155.4334
KS	0.1896	0.2161	0.6984	0.7917	0.4051	0.3812	0.2092	0.3887
P-value	0.5886	0.4330	< 0.0001	< 0.0001	0.0094	0.0258	0.4827	0.0152

Table 10. Various estimators for dataset I.

Method →	MLE	MPSE	LSE	CVME	WLSE	PCE	ADE	RADE
β	0.9801	0.9818	0.9836	0.9834	0.9828	0.9772	0.9831	0.9818
KS	0.1896	0.1609	0.1569	0.1542	0.1452	0.1932	0.1487	0.1597
P-Value	0.5886	0.7756	0.8004	0.8166	0.8664	0.5144	0.8479	0.7833

6.2. Dataset II: Leukemia remission times

This dataset pertains to leukemia remission times, measured in weeks, for a total of 20 patients, as described in [17], utilizing the concept of discretization. In order to delve into the characteristics of dataset II, we have generated nonparametric plots, including box plots, normal Q-Q plots, violin plots, and strip plots. For more comprehensive information and visual representations, please consult Figure 6.

The MLEs along with their corresponding SE, C.I for the parameter(s), and the results of the GOF tests for this dataset are provided in Tables 11 and 12. Importantly, it's noteworthy that the values of $-L$, AIC, BIC, CAIC, HQIC, and KS all exhibit lower values, while the P-value is higher when considering the DBHE distribution in comparison to the values obtained for the other models. Consequently, based

on this comprehensive analysis of the real dataset, it is evident that the proposed distribution stands out as a highly competitive model.

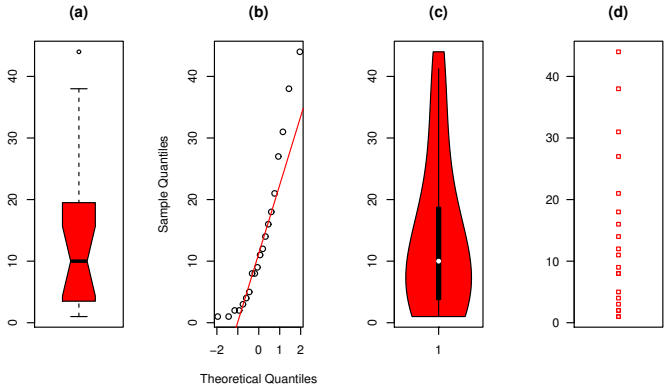

Figure 6. Nonparametric plots for dataset II.

Table 11. The MLEs, SE, and C.I for dataset II.

	β			α		
Model ↓ Parameter ⟶	MLE	SE	C.I	MLE	SE	C.I
DBHE	0.9603	0.0097	[0.9412, 0.9794]	–	–	–
DR	0.9971	0.0007	[0.9961, 0.9982]	–	–	–
DIR	3.374×10^{-7}			–	–	–
DBH	0.9972	0.0124	[0.9734, 1.0213]	–	–	–
DPa	0.6552	0.0619	[0.5342, 0.7770]	–	–	–
Poi	13.7545	0.8292	[12.1267, 15.3887]	–	–	–
DIW	0.0039	0.0072	[0, 0.0184]	1.0073	0.1751	[0.6640, 1.3501]
DB-XII	0.9943	0.0113	[0.9765, 1.0132]	158.3545	35.4094	[0, 3395.9312]

Table 12. The GOF test for dataset II.

Statistic↓ Parameter⟶	DBHE	DR	DIR	DBH	DPa	Poi	DIW	DB-XII
-L	73.5159	79.3092	85.0865	94.6355	84.5822	145.4324	74.7965	79.9804
AIC	149.0318	160.6175	172.1711	191.2695	171.1659	292.8652	153.5932	163.9614
CAIC	149.2541	160.8401	172.3944	191.4917	171.3876	293.0870	154.2997	164.6671
BIC	150.0275	161.6136	173.1672	192.2652	172.1613	293.862	155.5851	165.9527
HQIC	149.2262	160.8124	172.3665	191.4639	171.3596	293.0598	153.9824	164.3511
KS	0.1471	0.2541	0.4822	0.6691	0.3721	0.3799	0.1966	0.2913
P-value	0.7800	0.1323	< 0.0001	< 0.0001	0.008	0.006	0.4221	0.0671

Figure 7 illustrates the P-P plot for dataset II, while Figure 8 presents the estimated CDFs and the profile of the L for the parameter β in dataset II. Figure 8 reaffirms our empirical observations, providing

further support for the suitability of the DBHE distribution in analyzing this dataset. Furthermore, it underscores the uniqueness of the estimator for β.

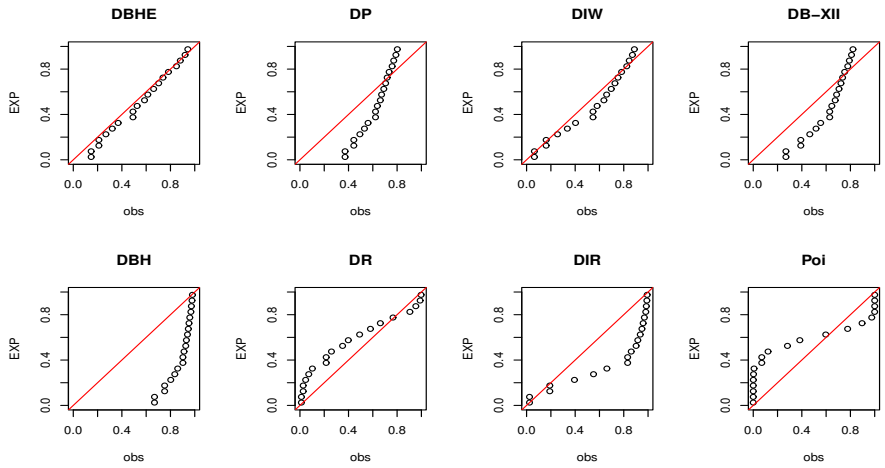

Figure 7. The P-P plot for dataset II.

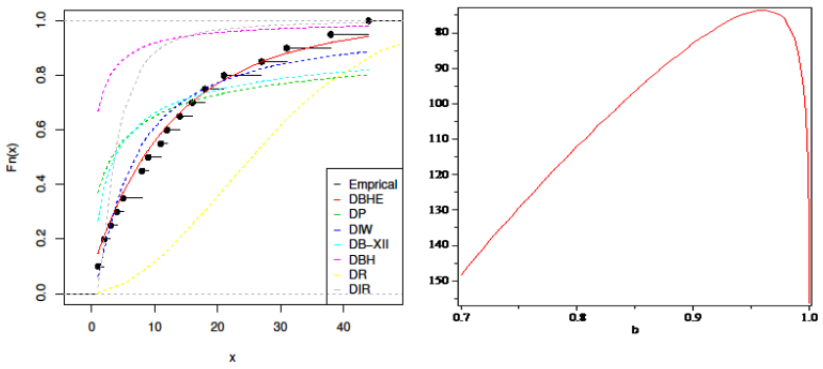

Figure 8. The estimated CDFs (left panel) and L profile of $\widehat{\beta}$ (right panel) for dataset II.

Table 13 presents an overview of diverse estimation techniques applied to dataset II under the proposed model framework.

Table 13. Various estimators for dataset II.

	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
β	0.9686	0.9703	0.9652	0.9710	0.9681	0.9641	0.9708	0.9661
KS	0.1233	0.1119	0.1550	0.1146	0.1280	0.1642	0.1128	0.1462
P-Value	0.9307	0.9670	0.7531	0.9600	0.9106	0.6839	0.9649	0.8106

The examination indicated that all estimation methods demonstrate satisfactory performance in terms of fitting the data, with the ADE approach emerging as the most effective among the available methods.

6.3. *Dataset III: Carious teeth*

The third dataset pertains to the count of carious teeth among the four deciduous molars. Detailed information regarding this dataset can be referenced in the work of Krishna and Pundir, as cited in [4]. In order to investigate the attributes of dataset III, we have generated nonparametric plots, which encompass box plots, normal Q-Q plots, violin plots, and strip plots. For more comprehensive information and visual representations, see consult Figure 9.

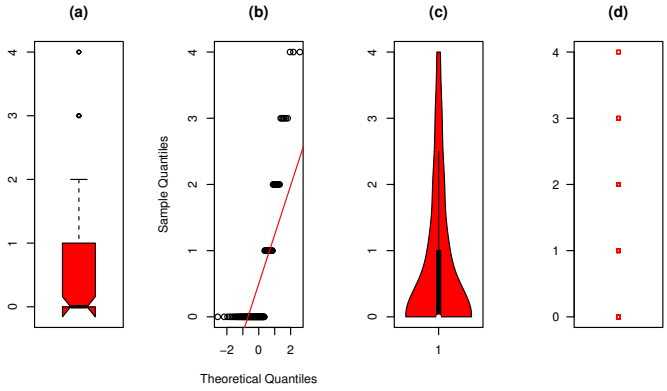

Figure 9. Nonparametric plots for dataset III.

The MLEs along with their corresponding SE, C.I for the parameter(s), and the results of the GOF tests for this dataset are available in Tables 14–16. Remarkably, it is evident that the DBHE distribution shows lower values for the chi-squared (χ^2) statistic while yielding higher p-values in comparison to the values obtained for the other models. As a result, this comprehensive analysis of the real dataset strongly suggests that the proposed distribution is a highly competitive model.

Table 14. The MLEs, SE and C.I for dataset III.

	β			α		
Model ↓ Parameter ⟶	MLE	SE	C.I	MLE	SE	C.I
DBHE	0.5767	0.0372	[0.5042, 0.6495]	–	–	–
DR	0.6651	0.0290	[0.6081, 0.7225]	–	–	–
DIR	0.6259	0.0491	[0.5292, 0.7214]	–	–	–
Geo	0.5988	0.0379	[0.5242, 0.6738]	–	–	–
DPa	0.1842	0.0325	[0.1207, 0.2479]	–	–	–
Poi	0.6700	0.0819	[0.5096, 0.8304]	–	–	–
PoiLi	1.9982	0.2636	[1.4812, 2.5146]	–	–	–
DLi	1.2942	0.1042	[1.0901, 1.4987]	–	–	–
DLogL	0.7455	0.1016	[0.5462, 0.9449]	1.7682	0.2671	[1.2440, 2.2921]
DIW	0.6338	0.0492	[0.5375, 7293]	1.5764	0.2515	[1.0843, 2.0676]
DW	0.3745	0.0496	[0.2782, 0.4706]	0.8951	0.1192	[0.6627, 1.1282]
EDLi	0.3791	0.0651	[0.2527, 0.5063]	0.5437	0.1587	[0.2343, 0.8529]
DLi-II	0.4012	0.2695	[0, 0.9281]	0.4782	0.5293	[0, 1.5147]
GGeo	0.4676	0.0892	[0.2932, 0.6414]	0.6784	0.3027	[0.0863, 1.2705]
DGE-II	0.4681	0.0728	[0.3270, 0.6092]	0.7181	0.2062	[0.3146, 1.1222]
DLFR	0.4013	0.0560	[0.2912, 0.5115]	1.0000	0.0449	[0.9132, 1]

Table 15. The GOF test for dataset III.

X	Ob. Fr.	DBHE	DR	DIR	Geo	DPa	DLi	PoiLi	Poi
0	64	62.8037	33.5000	62.5034	59.8802	69.0678	57.1253	37.5183	51.1709
1	17	21.3654	46.9437	26.4176	24.0238	15.3611	26.8834	25.0582	34.2845
2	10	8.5966	17.0130	5.9918	9.6383	6.0031	10.4459	15.6336	11.4853
3	6	3.7795	2.3970	2.1903	3.8669	3.0100	3.7068	9.3877	2.5650
$\geq$ 4	3	3.4548	0.6463	2.9126	2.5908	6.5579	1.8385	12.4902	0.4943
Total	100	100	100	100	100	100	100	100	100
χ^2		1.5748	48.2769	9.0561	3.3515	3.2416	6.6322	30.8894	13.2954
df		2	1	2	2	2	2	2	1
P-value		0.4550	< 0.001	0.0113	0.188	0.199	0.0362	< 0.001	< 0.001

Table 16. The GOF test for dataset III part II.

		Expected Frequences (Ex. Fr.)							
X	Ob. Fr.	DLogL	DW	DIW	GGeo	EDLi	DLi-II	DGE-II	DLFR
0	64	62.7253	62.6000	63.3000	62.7335	63.5850	59.8817	63.5630	59.9011
1	17	22.4187	21.3414	22.4805	21.3633	19.7546	24.0262	20.1733	24.0136
2	10	7.0053	8.8439	6.4429	8.7638	9.0954	9.6448	8.7926	9.6362
3	6	2.9774	3.8811	2.7621	3.8645	4.1898	3.8710	4.0029	3.8667
≥ 4	3	4.8734	3.3337	5.0143	3.2749	3.3752	2.5928	3.4682	2.6084
Total	100	100	100	100	100	100	100	100	100
χ^2		2.78403	1.50736	3.5001	1.5760	0.7490	3.3470	0.9809	3.3401
df		1	1	1	1	1	1	1	1
$P-value$		0.0952	0.2195	0.06137	0.2094	0.3868	0.0672	0.3219	0.0685

Figure 10 illustrates the observed and expected PMFs for dataset III. Figure 11 displays the L profile of the DBHE model parameters for dataset III, and it's noteworthy that the estimators are distinct and singular.

Table 17 offers a consolidated overview of diverse estimation techniques employed for dataset III within the context of the proposed model.

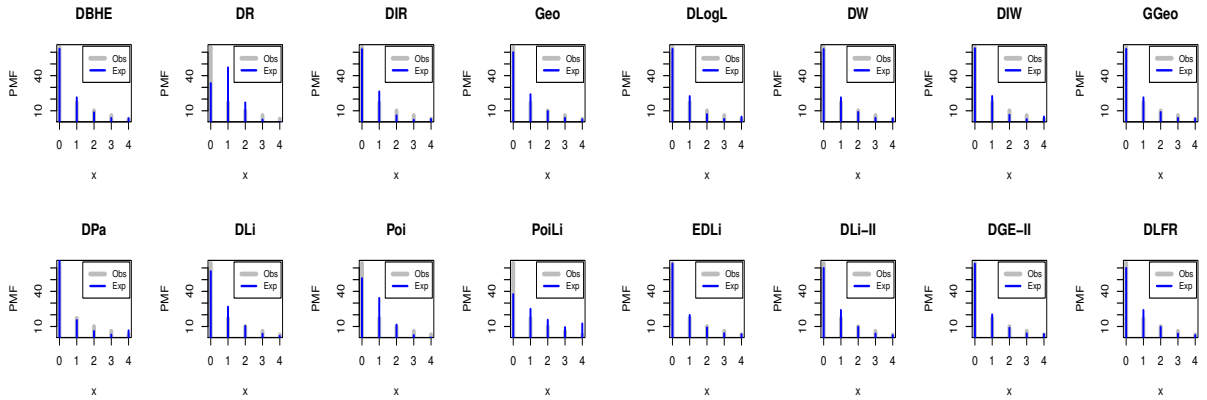

Figure 10. The observed and expected PMFs for dataset III.

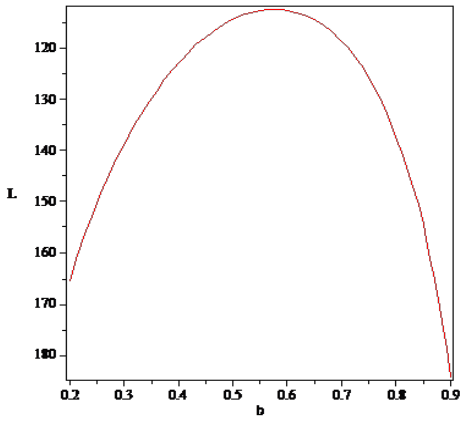

Figure 11. The L profile of $\widehat{\beta}$ for dataset III.

Table 17. Various estimators for dataset III.

	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
β	0.8543	0.8654	0.7990	0.8687	0.8512	0.7656	0.8663	0.8086
KS	0.2620	0.2817	0.3593	0.2902	0.2669	0.4173	0.2842	0.3411
P-Value	0.8056	0.7347	0.4361	0.7021	0.7885	0.3501	0.7255	0.5030

The analysis has shown that all the estimation methods perform well in terms of fitting the data, with the MPSE approach being the most effective among them.

6.4. Dataset IV: COVID-19 pandemic

The forth dataset comprises the number of deaths attributed to coronavirus in the Punjab region during the period from March 24, 2020, to April 30, 2020. The dataset is as follows: 1, 2, 3, 5, 5, 6, 9, 9, 11, 11, 11, 12, 15, 15, 16, 17, 18, 19, 21, 23, 24, 28, 34, 36, 37, 41, 42, 45, 51, 58, 65, 73, 81, 83, 91, 100, 103, 106. To examine the attributes of dataset IV, we have generated non-parametric plots, encompassing box plots, normal Q-Q plots, violin plots, and strip plots. For more in-depth information and visual representations, please consult Figure 12.

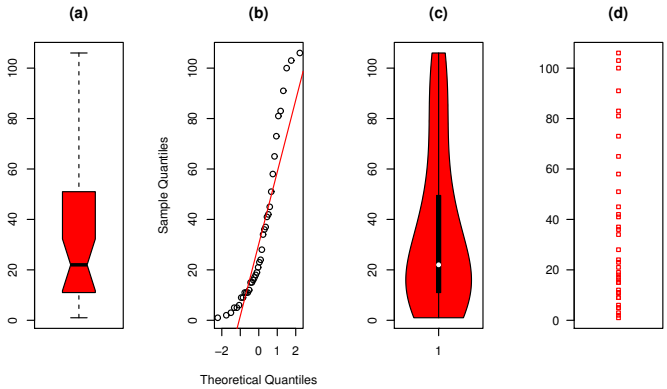

Figure 12. Nonparametric plots for dataset III.

The MLEs along with their corresponding SE, C.I for the parameter(s), and the results of the GOF tests for this dataset can be located in Tables 18 and 19. Significantly, it is evident that for the DBHE distribution, the values of $-L$, AIC, BIC, CAIC, HQIC, and KS all show lower values, while the P-value is higher when compared to the values obtained for the other models. Consequently, based on this comprehensive analysis of the real dataset, it is clear that the proposed distribution emerges as a highly competitive model.

Table 18. The MLEs, SE, and C.I for dataset IV.

	β			α		
Model ↓ Parameter⟶	MLE	SE	C.I	MLE	SE	C.I
DBHE	0.9838	0.0029	[0.9781, 0.9891]	–	–	–
DR	0.9996	0.00007	[0.9994, 0.9997]	–	–	–
DIR	1.634×10^{-10}	–	–	–	–	–
DBH	0.9996	0.0035	[0.9927, 1.0064]	–	–	–
DPa	0.7298	0.0373	[0.6567, 0.8031]	–	–	–
Poi	34.9211	0.9586	[33.0423, 36.7999]	–	–	–
DIW	0.00005	0.0001	[0, 0.0003]	0.8969	0.1070	[0.6874, 1.1067]
DB-XII	0.9960	0.0041	[0.9892, 1.0028]	79.5877	82.3391	[0, 2153.0236]

Table 19. The GOF test for dataset IV.

Statistic	DBHE	DR	DIR	DBH	DPa	Poi	DIW	DB-XII
$-L$	174.1947	186.7001	226.3555	241.3062	202.5788	594.7516	179.1153	198.7273
AIC	350.3893	375.4005	454.7092	484.6124	407.1552	1191.5021	362.2356	401.4544
CAIC	350.5005	375.5113	454.8201	484.7235	407.2676	1191.6130	362.5713	401.7976
BIC	352.0269	377.0386	456.3476	486.2534	408.7931	1193.1432	365.5042	404.7292
HQIC	350.9723	375.9832	455.2923	485.1951	407.7384	1192.0851	363.3955	402.6197
KS	0.1124	0.3089	0.6442	0.7786	0.3793	0.5193	0.1388	0.3667
P-value	0.7227	0.00142	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.4564	< 0.0001

Figure 13 presents the P-P plot for dataset IV, whereas Figure 14 exhibits the estimated CDFs and the profile of the L for the parameter β in dataset IV. Figure 14 further reinforces our empirical observations, providing additional support for the appropriateness of the DBHE distribution in analyzing this dataset. Additionally, it emphasizes the uniqueness of the estimator for β.

Table 20 offers a comprehensive compilation of various estimation techniques applied to dataset IV within the context of the proposed model framework.

Table 20. Various estimators for dataset IV.

	MPSE	ADE	MLE	LSE	RADE	PCE	CVME	WLSE
β	0.9862	0.9871	0.9855	0.9874	0.9865	0.9848	0.9873	0.9852
KS	0.1027	0.1046	0.1165	0.1091	0.0974	0.1328	0.1079	0.1208
P-Value	0.8426	0.8267	0.7180	0.7875	0.8829	0.6224	0.7972	0.6771

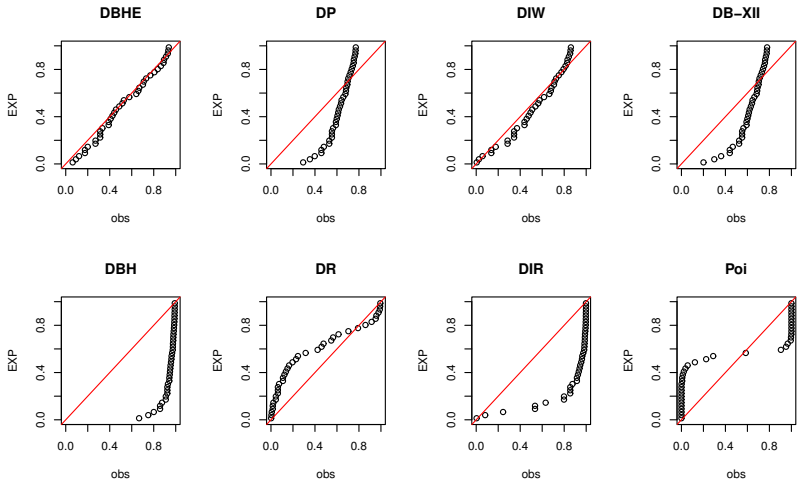

Figure 13. The P-P plot for dataset IV.

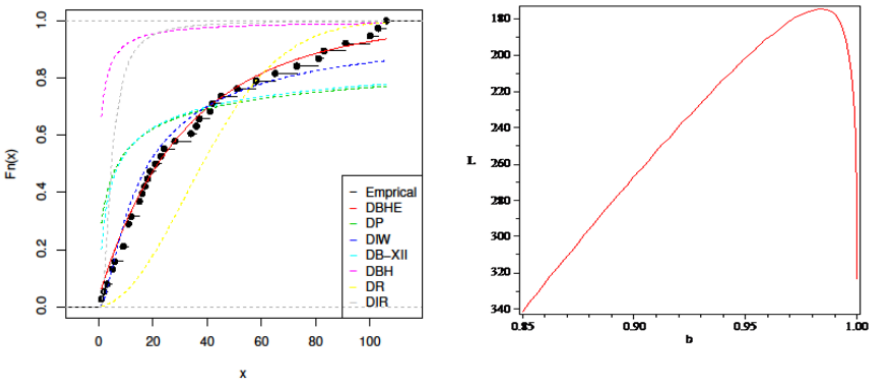

Figure 14. The estimated CDFs (left panel) and L profile of $\widehat{\beta}$ (right panel) for dataset IV.

The examination indicated that all estimation techniques adequately achieve data fitting, with the WLSE method standing out as the most efficient among them.

7. Concluding remarks and future work

This article centers on a discrete distribution with one parameter, developed using the survival discretization approach, referred to as the DBHE distribution. The statistical properties of the DBHE model have been derived and expressed in terms of generalized hypergeometric functions. It has been established that the DBHE model is particularly suitable for modeling right-skewed datasets characterized by leptokurtic shapes. The presented discrete distribution can serve as a valuable statistical tool for modeling a decreasing HRF in the presence of outlier observations. The DBHE parameter has been estimated using various approaches, including MPSE, ADE, MLE, LSE, RADE, PCE, CVME, and WLSE. Simulation studies conducted across different sample sizes, revealed that all these techniques are effective in estimating the DBHE parameter, with the ADE approach

performing best. Furthermore, the study includes the analysis of four real datasets to demonstrate the effectiveness of the DBHE distribution. It was observed that the DBHE distribution outperforms all other competing distributions across all aspects of the analysis. Looking ahead, the article hints at future directions, including the proposal and detailed discussion of bivariate extensions of the DBHE models, as well as the exploration of regression models and the integer-valued autoregressive of order one process along with their applications.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2023/01/27231).

Conflict of interest

The authors declare no conflicts of interest.

References

1. A. S. Yadav, E. Altun, H. M. Yousof, Burr-Hatke exponential distribution: A decreasing failure rate model, statistical inference and applications, *Ann. Data. Sci*, **8** (2021), 241–260. https://doi.org/10.1007/s40745-019-00213-8
2. M. El-Morshedy, M. S. Eliwa, E. Altun, Discrete Burr-Hatke distribution with properties, estimation methods and regression model, *IEEE Access*, **8** (2020), 74359–74370. https://doi.org/10.1109/ACCESS.2020.2988431
3. M. El-Morshedy, A discrete linear-exponential model: Synthesis and analysis with inference to model extreme count data, *Axioms*, **11** (2022), 531. https://doi.org/10.3390/axioms11100531
4. H. Krishna, P. S. Pundir, Discrete Burr and discrete Pareto distributions, *Statist. Methodol.*, **6** (2009), 177–188. https://doi.org/10.1016/j.stamet.2008.07.001
5. T. Hussain, M. Ahmad, Discrete inverse Rayleigh distribution, *Pakistan J. Statist.*, **30** (2014), 203.
6. M. A. Jazi, C. D. Lai, M. H. Alamatsaz, A discrete inverse Weibull distribution and estimation of its parameters, *Statist. Methodol.*, **7** (2010), 121–132. https://doi.org/10.1016/j.stamet.2009.11.001
7. E. Gómez-Déniz, E. Calderín-Ojeda, The discrete Lindley distribution: properties and applications, *J. Statist. Comput. Simul.*, **81** (2011), 1405–1416. https://doi.org/10.1080/00949655.2010.487825
8. J. M. Jia, Z. Z. Yan, X. Y. Peng, A new discrete extended Weibull distribution, *IEEE Access*, **7** (2019), 175474–175486. https://doi.org/10.1109/ACCESS.2019.2957788
9. E. Gómez-Déniz, Another generalization of the geometric distribution, *Test*, **19** (2010), 399–415. https://doi.org/10.1007/s11749-009-0169-3

10. M. A. Hegazy, R. E. Abd El-Kader, A. A. El-Helbawy, G. R. Al-Dayian, Bayesian estimation and prediction of discrete Gompertz distribution, *J. Adv. Math. Comput. Sci.*, **36** (2021), 1–21.
11. V. Nekoukhou, M. H. Alamatsaz, H. Bidram, Discrete generalized exponential distribution of a second type, *Statistics*, **47** (2013), 876–887. https://doi.org/10.1080/02331888.2011.633707
12. E. M. Almetwally, S. Dey, S. Nadarajah, An overview of discrete distributions in modelling COVID-19 data sets, *Sankhya A*, **85** (2023), 1403–1430. https://doi.org/10.1007/s13171-022-00291-6
13. A. S. Eldeeb, M. Ahsan-ul-Haq, M. S. Eliwa, A discrete Ramos-Louzada distribution for asymmetric and over-dispersed data with leptokurtic-shaped: Properties and various estimation techniques with inference, *AIMS Math.*, **7** (2022), 1726–1741. https://doi.org/10.3934/math.2022099
14. H. Haj Ahmad, D. A. Ramadan, E. M. Almetwally, Evaluating the discrete generalized Rayleigh distribution: Statistical inferences and applications to real data analysis, *Mathematics*, **12** (2024), 183. https://doi.org/10.3390/math12020183
15. H. M. Aljohani, M. Ahsan-ul-Haq, J. Zafar, E. M. Almetwally, A. S. Alghamdi, E. Hussam, et al., Analysis of COVID-19 data using discrete Marshall-Olkinin length biased exponential: Bayesian and frequentist approach, *Sci. Rep.*, **13** (2023), 12243. https://doi.org/10.1038/s41598-023-39183-6
16. J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, Hoboken: John Wiley & Sons, 2011.
17. P. Damien, S. Walker, A Bayesian non-parametric comparison of two treatments, *Scand. J. Statist.*, **29** (2002), 51–56. https://doi.org/10.1111/1467-9469.00891

AIMS Press