EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2025, Vol. 18, Issue 3, Article Number 6382
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

Length-Biased Discretized Fréchet-Weibull Probability Distribution: Mathematical Theory, Simulation Analysis, and Goodness-of-Fit Evaluation Using Real-Life Data

Diksha Das[1], Mohamed F. Abouelenein [2], Bhanita Das [1], Mohamed S. Eliwa[3,4,*]

[1] *Department of Statistics, North-Eastern Hill University, Shillong, Meghalaya, India*
[2] *Department of Insurance and Risk Management, College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Riyadh, Saudi Arabia*
[3] *Department of Statistics and Operations Research, College of Science, Qassim University, Saudi Arabia*
[4] *Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt*

Abstract. As data becomes more complex and abundant, the need for innovative statistical tools grows. This study introduces a new modeling framework based on a length-biased variant of the discretized Fréchet-Weibull distribution, designed to handle intricate data structures often encountered in practical applications. The paper explores the mathematical foundation of this distribution and derives its probability functions, providing an intuitive understanding of its behavior across various parameters. It highlights the flexibility of the model to capture a range of hazard rate forms, including increasing, decreasing, upside-down bathtub-shaped, and unimodal functions, making it suitable for diverse failure-time and count-data scenarios. A key feature is the model's ability to address overdispersion and underdispersion, as well as asymmetric data behaviors. It is effective in scenarios with excess zeros, common in fields like insurance and epidemiology. The characteristics of the length-biased discretized Fréchet-Weibull distribution are examined using conditional expectations and reverse hazard rate functions. Parameter estimation is performed using maximum likelihood estimation, with a simulation study assessing the reliability and efficiency of the estimators. The practical value of the model is demonstrated through applications to two real datasets, showing a superior fit compared to several established alternatives.

2020 Mathematics Subject Classifications: 62E99, 62E15

Key Words and Phrases: Length-biased modeling approach, Failure analysis, Conditional expectation, Infinite divisibility property, Simulation, Statistics and numerical data

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v18i3.6382

Email addresses: mseliwa@mans.edu.eg (M. S. Eliwa)

1. Introduction

In statistical investigations, the choice and application of appropriate sampling techniques mark the initial and critical step of data collection. These techniques form the backbone of any empirical study, as they determine the representativeness and reliability of the collected data. Sampling plays a pivotal role in ensuring valid conclusions are drawn about the population under study. Despite its importance, the specific method employed for selecting samples is frequently underreported or neglected. Such oversight can result in the omission of rare occurrences that hold significant inferential value. For instance, the presence of a child with albinism in a family, a notably uncommon phenomenon, serves as a prototypical example of rare events that can be overlooked if the sampling process fails to account for them adequately. Identifying and incorporating such rare events into the analytical framework is essential, particularly when the objective is to make accurate generalizations about the entire population, including these low-probability cases. Without such due diligence, conclusions drawn may not reflect the true nature of the population. In practical research settings involving human populations, flora, fauna, aquatic species, and insects, it is often infeasible to guarantee equal selection probability for all sampling units due to the absence of a complete and well-defined sampling frame. This limitation introduces bias into the observations since some units inherently possess a higher likelihood of selection than others. Consequently, biased data are a common occurrence in scientific investigations. To mitigate this issue, sampling methods such as probability proportional to size (PPS) have been developed. PPS is particularly effective in constructing weighted distributions by modifying the original probability structure based on the selection mechanism. By appropriately adjusting the probability function of the variable of interest, PPS improves the efficiency and accuracy of statistical estimators.

When dealing with populations where the likelihood of recording specific events varies, it becomes necessary to model this inequality using a weight function. This function expresses how the frequency of observed events deviates from their actual probability. Fisher [1] was among the pioneers who recognized how ascertainment procedures could influence frequency estimation, thereby introducing the concept of weighted probability distributions. This foundational work was later generalized by Rao [2], laying the groundwork for the formal development of weighted distributions. These distributions are especially useful when classical models fail to accommodate real-world irregularities, such as the non-observability of certain outcomes, data degradation, or unequal selection probabilities inherent in the sampling scheme. A special case of weighted distributions arises when the probability of observing a non-negative random variable is directly proportional to its magnitude. This yields a class of models known as length-biased distributions, first introduced by Cox [3]. These distributions have since found extensive applications across various disciplines, including econometrics, environmental science, biomedical research, demography, and forestry. Considerable research has been conducted to derive length-biased forms of well-known continuous distributions. For example, Patil and Rao [4] developed length-biased counterparts of log-normal, gamma, beta, and Pareto distributions. Gove [5] reviewed the practical applications of such models, particularly focusing

on the Weibull distribution in forestry. Subsequent contributions include length-biased Weibull by Pandya et al. [6], length-biased beta by Mir et al. [7], length-biased exponentiated inverted Weibull by Seenoi et al. [8], and length-biased weighted Lomax by Ahmad et al. [9]. Recently, Chaito and Khamkong [10] introduced the length-biased Weibull–Rayleigh model for hydrological data analysis. Despite the rich body of work on continuous length-biased distributions, there remains a scarcity of research concerning their discrete counterparts.

As Patil and Rao [11] observed, "Although situations involving weighted distributions seem to occur frequently in various fields, the underlying concept of weighted distributions as a major stochastic concept does not seem to have been widely recognized." This statement remains valid today. Although numerous continuous weighted models have been extensively studied, relatively little attention has been paid to discrete models, indicating an unexplored avenue with substantial potential for research and application. Motivated by these observations, the present study introduces a novel length-biased distribution tailored specifically for discrete datasets. This model is derived from the recently proposed discretized Fréchet-Weibull distribution by Das and Das [12], and is henceforth referred to as the length-biased discretized Fréchet-Weibull (LBDFW) distribution. Weighted or length-biased distributions offer a systematic method to adjust the sampling bias associated with unascertained events. They refine the observed data, resulting in more reliable statistical inference. In parallel, generalized distributions offer improved flexibility and better adaptability to diverse datasets compared to their baseline forms. The proposed LBDFW distribution, being a length-biased generalization, is capable of modeling data sets with both overdispersion and underdispersion efficiently. Furthermore, the lack of existing discrete length-biased models makes this contribution a valuable addition to the statistical modeling literature.

While the length-biased approach has been successfully employed in developing discrete probability models, many of the existing models exhibit limited adaptability to complex data features, particularly when applied to real-world datasets characterized by irregular dispersion patterns, asymmetry, and excess zeros. Traditional length-biased discrete models often lack the structural flexibility to simultaneously accommodate both overdispersion and underdispersion, which are common in count data arising in fields such as epidemiology, insurance, reliability engineering, and social sciences. A significant limitation observed in previous length-biased discrete models lies in their inability to represent a wide variety of hazard rate shapes. Most such models are constrained to monotonic hazard functions (increasing or decreasing), which severely restricts their application in domains requiring more complex hazard behaviors such as bathtub-shaped or unimodal failure rates. This inflexibility often leads to poor model fit and suboptimal predictive performance in practical scenarios. Additionally, many earlier models constructed via the length-biased mechanism are derived from standard or classical distributions, which may not inherently possess sufficient shape parameters to regulate skewness and tail behavior effectively. As a result, these models struggle to capture asymmetric distributions or data exhibiting leptokurtic or platykurtic tendencies, thereby limiting their capacity to reflect real-world heterogeneity. Furthermore, parameter identifiability and computational chal-

lenges often arise in these models, especially when the weighting scheme is imposed without adequate theoretical support or empirical justification. In such cases, the length-biased transformation may amplify noise or measurement error in the data, leading to biased inference. In light of these limitations, the LDFW distribution is proposed as a comprehensive and flexible alternative. By extending a generalized parent model the discretized Fréchet-Weibull distribution through a length-biased transformation, the LBDFW model overcomes the rigidity of previous formulations. It accommodates diverse dispersion patterns, tail behaviors, and hazard shapes, while maintaining tractability for estimation and interpretation. Its ability to adapt to both common and extreme data scenarios, including the presence of zero inflation, asymmetry, and nonstandard hazard dynamics, establishes the LBDFW model as a motivated and superior alternative in the class of length-biased discrete models.

The remainder of the paper is structured as follows. First, the necessary preliminaries related to length-biased distributions are discussed. This includes the foundational definitions, key concepts, and the rationale for incorporating length-biasing into discrete probability models, particularly within the context of count data analysis. Next, a new model is introduced, called the LBDFW distribution. Its derivation is outlined starting from the continuous Fréchet-Weibull form, followed by an appropriate discretization approach that integrates the length-biased transformation. In addition, graphical representations of the probability mass and cumulative distribution functions are provided to illustrate the behavior of the model under different parameter configurations. The characterization of the proposed LBDFW model is then explored using tools such as conditional expectation and the reversed hazard rate function, shedding light on its underlying stochastic structure. This is followed by a set of numerical calculations aimed at analyzing the behavior of the model parameters under various scenarios, highlighting the flexibility of the model and the range of shapes. Parameter estimation is carried out using the method of maximum likelihood. The likelihood function is constructed, and the corresponding estimating equations are derived and numerically solved. To assess the performance of these estimators, a comprehensive simulation study is conducted across multiple scenarios involving different sample sizes and parameter combinations, focusing on empirical bias and mean squared error. The practical effectiveness of the LBDFW model is demonstrated by applying it to two real-world count datasets. Comparative analysis is performed against existing models, namely, the discretized Fréchet-Weibull, weighted discretized Fréchet-Weibull, discrete generalized inverse Weibull, and discrete generalized Weibull distributions, using several goodness-of-fit and model selection criteria. The paper concludes by summarizing the key findings, highlighting the strengths and potential limitations of the proposed model, and offering recommendations for future research directions.

2. Mathematical Theory and Graphical Representation of the LBDFW Distribution

Let Y be a non-negative count random variable with probability mass function (pmf) $P[Y = y]$, where $y \in \mathbf{Z}_+ = \{0, 1, 2, ...\}$ and Y_W be the corresponding weighted random

variable. Let $w(y)$ be a non-negative weight function on $\mathbf{Z}_+$ having a finite expectation

$$E[w(y)] = \sum_y w(y) \times P[Y = y] < \infty.$$

Then, following the notations in Kokonendji and Casany [13], the pmf of the weighted random variable Y_W is given by

$$P^W[Y_W = y] = \frac{w(y) \times P[Y = y]}{E[w(y)]} \quad ; \quad y \in \mathbf{Z}_+. \tag{2.1}$$

Depending upon the choice of the weight function $w(y)$, we have different weighted models. The two most prevalent types of weighted distributions are the length-biased and the size-biased distributions. A distribution is said to be length-biased when the probability of detecting a positive-valued random variable is proportionate to the variable's value. On the other hand, the weighted distributions where the observations from a sample are recorded with probability proportional to some measure of unit size, are called size-biased distributions. It is noteworthy that a particular case of size-biased distribution is length-biased distribution. Focusing on the discrete random variable Y, according to Castillo and Casany [14], let Y_S denote the weighted random variable associated with the weight function $w(y) = y^c$. The resulting distribution is referred to as size-biased of order c. When $w(y) = y$ (i.e., $c = 1$), the distribution of the random variable Y_L is known as the length-biased distribution. The pmf of the length-biased distribution is given by:

$$P^L[Y_L = y] = \frac{y \times P[Y = y]}{E[y]} \quad ; \quad y \in \mathbf{Z}_+, \tag{2.2}$$

where $E[y] = \sum_y y.P[Y = y] < \infty$. Using the survival function approach of discretization, recently Das and Das [12] developed a new discrete distribution called discretized Fréchet-Weibull (DFW) model. Its hazard rate function exhibits increasing, decreasing, and up-side down bathtub shapes, which are seldom observed in count distributions. Further, this distribution can be modeled with both positively and negatively skewed data. Let ξ be the parameter vector of DFW, defined as $\xi = (\alpha, \beta, m, k)^T$, such that $\xi \in (R_+ \times R_+ \times R_+ \times R_+)$. The pmf of the random variable $Y \sim \text{DFW}(\alpha, \beta, m, k)$, is given by

$$P[Y = y; \xi] = exp\Big\{ -\beta^\alpha \Big(\frac{m}{y+1}\Big)^{\alpha k} \Big\} - exp\Big\{ -\beta^\alpha \Big(\frac{m}{y}\Big)^{\alpha k} \Big\}, \tag{2.3}$$

where $y \in \mathbf{Z}_+$ and the parameters $(\alpha, \beta, m, k) > 0$. Considering the pmf of $\text{DFW}(\alpha, \beta, m, k)$ as in Eq.(2.3), the denominator in Eq.(2.2) can be written as

$$\begin{aligned} E[y] &= \sum_{y=0}^{\infty} y \times P[Y = y; \xi] \\ &= \sum_{y=0}^{\infty} y \times exp\Big\{ -\beta^\alpha \Big(\frac{m}{y+1}\Big)^{\alpha k} \Big\} - exp\Big\{ -\beta^\alpha \Big(\frac{m}{y}\Big)^{\alpha k} \Big\} \end{aligned}$$

$$= \sum_{y=0}^{\infty} q(y;\xi) = Q(y;\xi), \tag{2.4}$$

where $y \in \mathbf{Z}_+$ and the parameter vector $\xi = (\alpha, \beta, m, k)$. Further,

$$q(y;\xi) = y \times \left[exp\left\{ -\beta^{\alpha}\left(\frac{m}{y+1}\right)^{\alpha k}\right\} - exp\left\{ -\beta^{\alpha}\left(\frac{m}{y}\right)^{\alpha k}\right\} \right], \tag{2.5}$$

and the summation of the term $q(y;\xi)$ over the entire range of $\mathbf{Z}_+$ cannot be written in closed form and is termed as $Q(s;\xi)$. Let Y_L denote the length-biased version of the random variable $Y \sim \text{DFW}(\alpha, \beta, m, k)$, named as LBDFW model, then its pmf is given by

$$\begin{aligned} P^L[Y_L = y] &= \frac{y \times P[Y = y;\xi]}{\sum_y y \times P[Y = y;\xi]} \quad ; \quad y \in \mathbf{Z}_+ \\ &= \frac{q(y;\xi)}{Q(y;\xi)} = P(y;\xi), \end{aligned} \tag{2.6}$$

where $y \in \mathbf{Z}_+$ and the parameter vector $\xi = (\alpha, \beta, m, k)$. And the terms $q(y;\xi)$ and $Q(y;\xi)$ is as defined in Eq.(2.5) and Eq.(2.4) respectively. For LBDFW distribution also, α and k are the shape parameters and β and m are the scale parameters. The cumulative distribution function (cdf) of the LBDFW distribution is given by

$$\begin{aligned} F^L(y) = P^L[Y_L \le y] &= \sum_{s=0}^{y} P^L[Y_L = s] \\ &= \sum_{s=0}^{y} \frac{q(s;\xi)}{Q(s;\xi)} \\ &= \frac{\sum_{s=0}^{y} q(s;\xi)}{Q(s;\xi)} = \frac{V(y;\xi)}{Q(y;\xi)} = C(y;\xi), \end{aligned} \tag{2.7}$$

where $s \in \mathbf{Z}_+$. Further, the term $Q(s;\xi)$ is as defined in Eq.(2.4). The survival function (sf) of LBDFW distribution is given by

$$\begin{aligned} S^L(y) = P^L[Y_L \ge y] &= 1 - F^L(y) \\ &= 1 - C(y;\xi) \\ &= 1 - \frac{V(y;\xi)}{Q(y;\xi)} = S(y;\xi), \end{aligned} \tag{2.8}$$

where $y \in \mathbf{Z}_+$. Moreover, the terms $Q(y;\xi)$, $q(y;\xi)$ and $V(y;\xi)$ are as defined in Eq.(2.4) , Eq.(2.5) and Eq.(2.7), respectively. The hazard rate function (hrf) of LBDFW distribution is given by

$$H^L(y) = \frac{P^L[Y_L = y]}{S^L(y)} = \frac{P(y;\xi)}{S(y;\xi)} = H(y;\xi),$$

where $y \in \mathbf{Z}_+$. The terms $P(y;\xi)$ and $S(y;\xi)$ are as defined in Eq.(2.6) and Eq.(2.8), respectively.

Figure 1 illustrates the probability mass function of the LBDFW distribution for different parameter combinations. Figure 2 visually demonstrates the effect of increasing the shape parameters (α, k) and scale parameters (β, k) on the pmf of the LBDFW distribution. Additionally, in Figure 3, the possible shapes of the hazard rate function for the LBDFW distribution are portrayed.

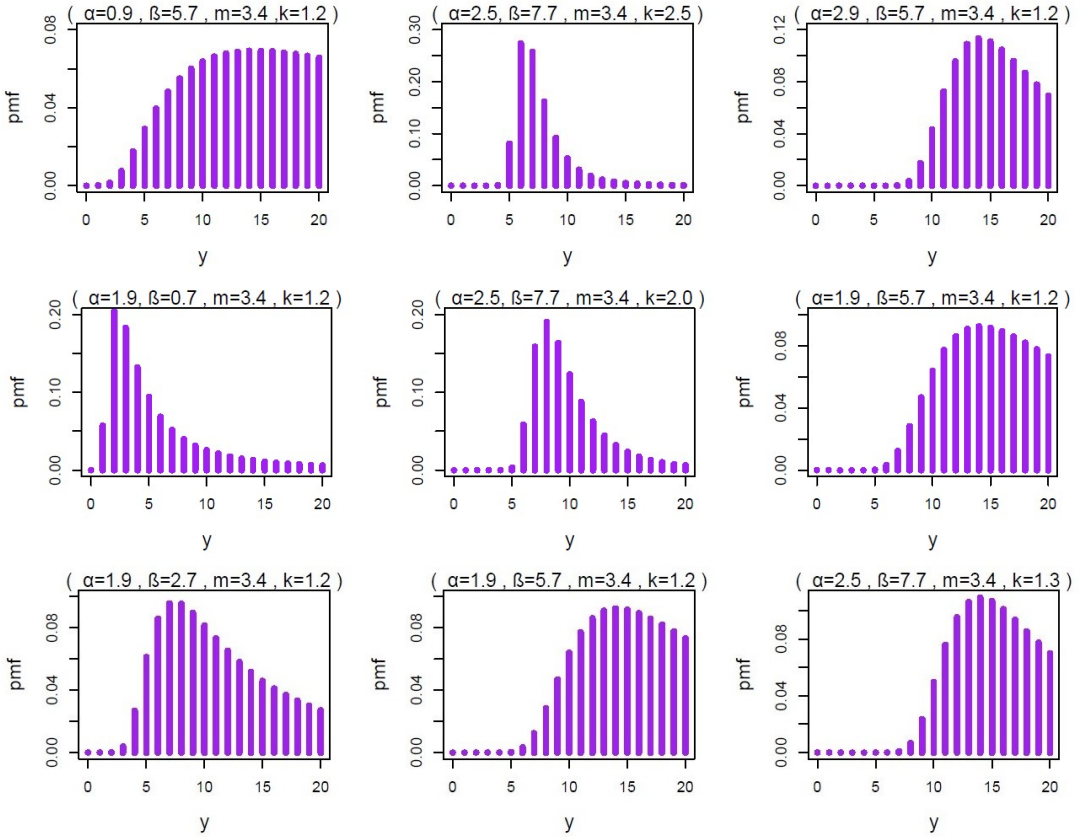

Figure 1: The pmf plot of LBDFW distribution for different values the parameter (α, β, m, k).

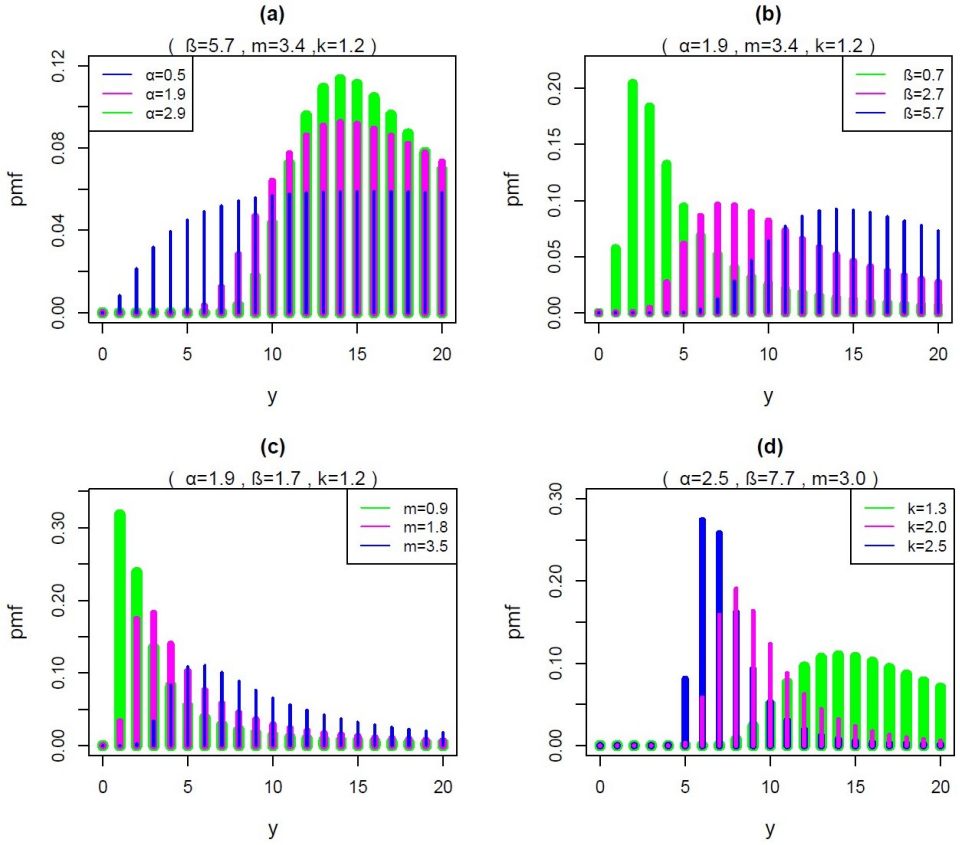

Figure 2: Impact of shape and scale parameters in the pmf plot of LBDFW distribution.

Figure 1 portrays the various forms of the LBDFW model, which can be unimodal and exhibit right skewness. Observations from Figure 2 (a) and (d) reveal that an increase in the shape parameters (α and k) leads to a higher peakedness in the pmf of LBDFW. Conversely, Figure 2 (b) and (c) show that as the scale parameters (β and m) grow, the shape of the pmf of LBDFW becomes flatter. In Figure 3, the hrf of LBDFW distribution is plotted for various parameters (α, β, m, k). Figures 3 depicts that, depending on the selected values of (α, β, m, k), the LBDFW model can display increasing, increasing-decreasing-increasing, or up-side-down bathtub shaped hazard rate functions.

3. Statistical Properties

3.1. Monotonic property

Determination of log-convex or log-concave behavior of a distribution, enables us to infer about the monotonic property of the failure rate. Also, it is known that, if a discrete distribution is log-convex then the hazard rate function will decrease, and if log-concave then the hazard rate function will increase. To prove that the distribution is log-convex,

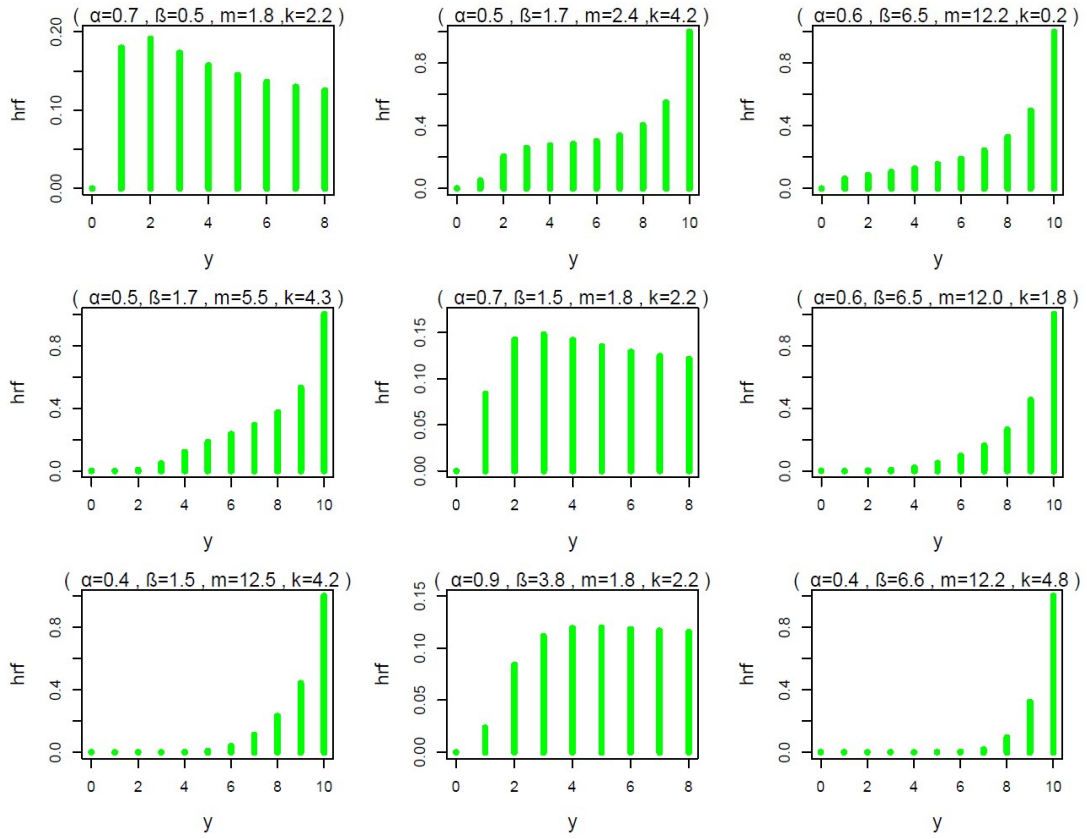

Figure 3: The hrf plot of LBDFW distribution for different values the parameter (α, β, m, k).

we need to show that

$$\{P^L(y;\xi)\}^2 \leq P^L(y+1;\xi)\ .\ P^L(y-1;\xi),$$

for $y \in \mathbf{Z}_+$ and $\xi = (\alpha, \beta, m, k)$. For this, it is sufficient if we can show that the ratio

$$\frac{P^L(y+1;\xi)}{P^L(y;\xi)} = \frac{q(y+1;\xi)}{q(y;\xi)},$$

for $y \in \mathbf{Z}_+$ and $\xi = (\alpha, \beta, m, k)$, is an increasing function of y, then the distribution is log-convex, otherwise, it is log-concave. For the proposed distribution, for different choices of the parameter values the ratio is a decreasing function of y and consequently the LBDFW distribution is log-concave. Hence, LBDFW model has an increasing failure rate distribution.

3.2. Infinite divisibility

According to Steutel and Harn [15], the following lemma enables us to determine the infinite divisibility property of a distribution, stated as

Lemma 3.1. *A necessary condition for infinite divisibility of a discrete distribution p_y is that $p_0 > 0$.*

It is seen from the pmf plot of LBDFW model as shown in Figure 1, that there can be no such combinations of the parameter values, that do satisfy the condition in Lemma (3.1). Symbolically, in the case of LBDFW model, for any combination choice of the parameter values $(\alpha, \beta, m, k), p_0 = 0$. Hence LBDFW model is not infinite divisible.

3.3. Moment generating function

The moment generating function (mgf) is a powerful tool in probability theory, used to derive moments (like the mean and variance) of a random variable. In the discrete case, the mgf is defined for a discrete random variable Y that takes values in a countable set (e.g., integers) and has a pmf. The mgf of the LBDFW model can be formulated as

$$\begin{aligned} M_{Y_L}(t) &= E[e^{ty}] \\ &= \sum_{y=0}^{\infty} e^{ty}\ P^L[Y_L = y;\xi] \\ &= \frac{1}{Q(y;\xi)} \sum_{y=1}^{\infty} e^{ty}\ q(y;\xi), \end{aligned} \tag{3.1}$$

where the terms $Q(y;\xi)$ and $q(y;\xi)$ are as defined in Eq.(2.4) and Eq.(2.5) respectively and the parameter vector $\xi = (\alpha, \beta, m, k)$. Now, differentiating the mgf of LBDFW model as in Eq.(3.1), for 'r' times w.r.t 't', we get

$$\begin{aligned} M^{(r)}_{Y_L}(t) &= \frac{d^r}{dt^r} M_{Y_L}(t) \\ &= \frac{1}{Q(y;\xi)} \sum_{y=1}^{\infty} y^r\ e^{ty}\ q(y;\xi). \end{aligned} \tag{3.2}$$

Thus, the r^{th} moment of LBDFW model is given by

$$\begin{aligned} E[Y_L^r] &= M^{(r)}_{Y_L}(t)\Big|_{t=0} \\ &= \frac{1}{Q(y;\xi)} \sum_{y=1}^{\infty} y^{r+1}\ p(y;\xi), \end{aligned} \tag{3.3}$$

where the terms $Q(y;\xi)$ is as defined in Eq.(2.4) and the term $p(y;\xi) = P[Y = y;\xi]$, as defined in Eq.(2.3).

3.4. Probability generating function

The probability generating function (pgf) and the mfg are both tools in probability theory used to summarize information about a random variable, particularly its distribution and moments; however, they differ in definition, domain, and application, especially

for discrete random variables with the pgf designed for discrete non-negative integers to encode probabilities through powers of (.), while the mgf applies to both discrete and continuous variables to encode moments through powers of (.), focusing more on moment structure. The pgf of the LBDFW model can be expressed as

$$\begin{aligned} G_{Y_L}(t) &= E[t^y] \\ &= \sum_{y=0}^{\infty} t^y \; P^L[Y_L = y;\xi] \\ &= \frac{1}{Q(y;\xi)} \sum_{y=1}^{\infty} t^y \; q(y;\xi), \end{aligned} \tag{3.4}$$

where the terms $Q(y;\xi)$ and $q(y;\xi)$ are as defined in Eq.(2.4) and Eq.(2.5) respectively and the parameter vector $\xi = (\alpha, \beta, m, k)$. Now, differentiating the pgf of LBDFW model as in Eq.(3.1), for 'r' times w.r.t 't', we get

$$\begin{aligned} G^{(r)}_{Y_L}(t) &= \frac{d^r}{dt^r} G_{Y_L}(t) \\ &= \frac{1}{Q(y;\xi)} \sum_{y=1}^{\infty} y_{(r-1)} \; t^{y-r} \; q(y;\xi), \end{aligned} \tag{3.5}$$

where $y_{(r-1)} = y(y-1)(y-2)....(y-r-2)(y-r-1)$. Thus, the r^{th} factorial moment of LBDFW model is given by

$$\begin{aligned} E[Y_L^{(r)}] &= G^{(r)}_{Y_L}(t)\Big|_{t=1} \\ &= \frac{1}{Q(y;\xi)} \sum_{y=1}^{\infty} y_{(r-1)} \; q(y;\xi), \end{aligned} \tag{3.6}$$

where $y_{(r-1)} = y(y-1)(y-2)....(y-r-2)(y-r-1)$ and the terms $Q(y;\xi)$ and $q(y;\xi)$ are as defined in Eq.(2.4) and Eq.(2.5), respectively.

3.5. Descriptive statistical analysis and numerical computations

This section presents an analysis of the behavior of several key statistical measures namely, the mean (μ), median (Med), variance (σ^2), skewness (S_k), kurtosis (K_r), and index of dispersion (ID) for the LBDFW model under varying values of the parameters α, β, m, and k. The numerical computations were executed using R software, and the corresponding results are summarized in Table 1.

From the observations in Table 1, several important conclusions can be drawn regarding the performance of the proposed model across different types of count data. First, the model proves to be particularly suitable for datasets that exhibit positive skewness, which is commonly encountered in situations where events are rare, and most counts are

Table 1: Descriptive statistics of the LBDFW model.

α	β	m	k	μ	Med	σ^2	S_k	K_r	ID
2.5				24.21	25.06	20.06667	1.35324	4.05788	0.82886
3.9				15.32	18.23	17.34444	1.35265	3.88372	1.13214
4.9	5.7	3.4	1.2	14.53	15.16	10.76667	0.42404	1.80627	0.74100
5.5				13.34	13.95	2.900000	0.18012	1.89628	0.21739
7.9				12.75	14.68	2.100000	0.07684	1.89276	0.16471
	0.7			1.30	1.57	0.47475	2.14142	8.46582	0.36519
	1.7			4.17	5.41	1.37485	2.42512	13.39448	0.32969
5.7	2.7	3.0	2.5	7.21	7.92	3.39989	1.89471	8.13549	0.47155
	3.7			9.56	12.14	5.42242	0.99227	3.73424	0.56719
	5.7			14.46	15.30	14.39232	2.86742	15.27046	0.99532
		0.8		0.32	0.65	0.42182	2.46918	12.38623	1.31818
		1.8		2.37	3.04	2.01323	1.58753	5.76545	0.84947
1.9	3.7	2.5	2.2	4.00	4.83	2.94949	1.26260	5.42316	0.73737
		3.5		5.96	7.08	7.02869	1.54696	5.30689	1.17931
		4.7		8.73	9.15	11.20919	0.86267	7.71215	1.28399
			2.8	5.22	7.20	1.64808	1.42627	5.32961	0.31572
			3.3	4.34	6.31	0.77212	1.25840	6.01170	0.17791
2.5	7.7	3.0	5.4	3.16	4.82	0.33778	3.38924	7.53870	0.10689
			6.8	2.76	3.09	0.26505	0.27711	2.76294	0.09603
			7.2	2.61	3.59	0.20091	0.10986	2.06060	0.11529

concentrated at lower values. Additionally, the proposed model is versatile in handling both platykurtic distributions (where kurtosis < 3), characterized by lighter tails, and leptokurtic distributions (where kurtosis > 3), which feature heavier tails and a sharper peak at the mean. This adaptability makes the model applicable to a wide variety of empirical count data exhibiting different tail behaviors. Furthermore, it is observed that as the shape parameters α and k increase, both the mean and variance decrease. This behavior indicates that higher values of these parameters result in more concentrated distributions, often representing data with lower variability and smaller expected counts. In contrast, an increase in the scale parameters β and m leads to an increase in both the mean and variance, suggesting that larger scale parameters expand the distribution, leading to higher expected counts and greater dispersion, which is typically seen in overdispersed count data. Finally, the proposed model demonstrates its flexibility by being suitable for modeling count data sets with varying levels of dispersion. It can effectively handle both overdispersed data (where the index of dispersion $ID > 1$), commonly found in fields such as insurance claims and ecological data, as well as underdispersed data (where $ID < 1$), which is often observed in more tightly controlled or regulated systems. These characteristics highlight the model's capability to capture the diverse range of behaviors seen in real-world count data.

4. Characterizations of the LBDFW Distribution

Characterizing distributions is a fundamental aspect of statistical research, especially for those working in applied fields. Researchers often seek to determine whether a given model accurately reflects the characteristics of a specific distribution, as this directly in-

fluences the validity and applicability of their findings. To make this determination, they rely on the characterizations of the distribution, which provide conditions and criteria that confirm whether the data conforms to the assumed distribution. This section is dedicated to exploring specific characterizations of the LBDFW model. These characterizations are based on two main approaches: first, by considering an appropriate function of the random variable, particularly through conditional expectation, and second, by examining the distribution in terms of the reverse hazard function. Both of these methods offer valuable insights into the properties of the model, enhancing its utility in real-world applications and allowing researchers to assess how well it fits various data scenarios.

4.1. Characterizations based on conditional expectation

In this subsection, we present our first characterization of LBDFW model in terms of the conditional expectation of certain functions of the random variable. The choice of the function depends on the form of the pmf.

Theorem 1. *Let* $Y : \Omega \to \mathbb{N}^* \ (\mathbb{N} \cup \{0\})$ *be a random variable. The pmf of* Y *is Eq.*(2.6) *if and only if*

$$E\left\{\left[C\left(Y+1;\xi\right)+C\left(Y;\xi\right)\right] \mid Y \leq k\right\} = C\left(k+1;\xi\right). \tag{4.1}$$

Proof: If Y has pmf as given in Eq.(2.6), then for $k \in \mathbb{N}$, the left-hand side of Eq.(4.1), using telescoping sum formula, will be

$$\left(F\left(k\right)\right)^{-1} \sum_{y=0}^{k} \left[C\left(y+1;\xi\right)^2 - C\left(y;\xi\right)^2\right] = C\left(k+1;\xi\right)^{-1} C\left(k+1;\xi\right)^2 = C\left(k+1;\xi\right).$$

Conversely, if Eq.(4.1) holds, then

$$\sum_{y=0}^{k} \left\{\left[C\left(y+1;\xi\right)+C\left(y;\xi\right)\right] f\left(y\right)\right\} = F\left(k\right) C\left(k+1;\xi\right) \tag{4.2}$$

From Eq.(4.2), we also have

$$\begin{aligned} \sum_{y=0}^{k-1} \left\{\left[C\left(y+1;\xi\right)+C\left(y;\xi\right)\right] f\left(y\right)\right\} &= F\left(k-1\right) C\left(k;\xi\right) \\ &= \left(F\left(k\right)-f\left(k\right)\right) C\left(k;\xi\right). \end{aligned} \tag{4.3}$$

Now, subtracting Eq.(4.3) from Eq.(4.2), yields

$$\left\{C\left(k+1;\xi\right)+C\left(k;\xi\right)\right\} f\left(k\right) = F\left(k\right)\left\{C\left(k+1;\xi\right)-C\left(k;\xi\right)\right\} + f\left(k\right) C\left(k;\xi\right).$$

From the above equality, we have

$$\frac{f\left(k\right)}{F\left(k\right)} = \frac{C\left(k+1;\xi\right)-C\left(k;\xi\right)}{C\left(k+1;\xi\right)} = 1 - \frac{C\left(k;\xi\right)}{C\left(k+1;\xi\right)},$$

which is the reverse hazard function corresponding to the pmf in Eq.(2.6), so Y has pmf Eq.(2.6). The application of characterizations based on conditional expectation for the LBDFW model provides valuable insights into the behavior of the distribution. By utilizing the conditional expectation, researchers can derive important properties of the model, such as the relationship between the random variable and its parameters under specific conditions. This approach helps to establish the underlying structure of the distribution, offering a clear understanding of how the distribution behaves as the model's parameters vary. In practice, these characterizations can be used to evaluate the fit of the LBDFW model to real-world data, making it a useful tool in various applied fields where understanding the conditional behavior of the data is crucial. Additionally, conditional expectation plays a key role in determining the expected outcomes in situations where information about prior events or conditions is available, further enhancing the model's applicability.

4.2. Characterizations of distributions based on reverse hazard function

This subsection deals with the characterization of LBDFW model in terms of the reverse hazard function.

Theorem 2. *Let $Y : \Omega \to \mathbb{N}^*$ be a random variable. The pmf of Y is Eq.(2.6), if and only if its reverse hazard function, r_F , satisfies the difference equation*

$$r_F(k+1) - r_F(k) = \frac{C(k;\xi)}{C(k+1;\xi)} - \frac{C(k+1;\xi)}{C(k+2;\xi)}; \qquad k \in \mathbb{N}, \tag{4.4}$$

with the initial condition $r_F(0) = 1$.

Proof: If Y has pmf Eq.(2.6), then clearly Eq.(4.4) holds. Now, if Eq.(4.4) holds, then for every $y \in \mathbb{N}$, we have

$$\begin{aligned}\sum_{k=0}^{y-1} \{r_F(k+1) - r_F(k)\} &= \sum_{k=00}^{y-1} \left\{ \frac{C(k;\xi)}{C(k+1;\xi)} - \frac{C(k+1;\xi)}{C(k+2;\xi)} \right\} \\ &= -\frac{C(y;\xi)}{C(y+1;\xi)},\end{aligned}$$

or

$$r_F(y) - r_F(0) = -\frac{C(y;\xi)}{C(y+1;\xi)},$$

or, in view of the initial condition

$$r_F(y) = 1 - \frac{C(y;\xi)}{C(y+1;\xi)}; \quad y \in \mathbb{N}^*,$$

which is the reverse hazard function corresponding to the pmf Eq.(2.6). The characterization of the LBDFW model through the reversed hazard rate function offers an important

analytical tool for understanding the distribution's behavior from a failure-time or reliability perspective. The reversed hazard rate function, which is defined as the ratio of the pmf to the cdf, is particularly useful in scenarios where interest lies in past lifetimes or the time until an event has already occurred. This approach is applicable in fields such as reliability engineering, biomedical studies, and survival analysis, where backward-looking analyses are critical. In the context of the LBDFW model, utilizing the reversed hazard rate function allows researchers to formulate conditions under which a random variable belongs to the LBDFW family. It also provides a mechanism for deriving uniqueness theorems and establishing relationships between the model's parameters and its reliability structure. Such characterizations are valuable in practice for verifying the adequacy of the LBDFW model in modeling systems or processes that exhibit specific patterns of early failures or sudden breakdowns. Hence, the reversed hazard rate function serves not only as a theoretical foundation for distributional characterization but also as a practical tool for data validation and model.

5. A Comprehensive Approach to Maximum Likelihood Estimation for Count Data Structures

This section presents the estimation of the parameters of the LBDFW model using the maximum likelihood estimation method. Let $y_1, y_2, y_3, \ldots, y_n$ be a random sample drawn from the LBDFW distribution. The corresponding likelihood function is given by:

$$L(y) = \prod_{i=1}^{n} P^L[Y_L = y_i].$$

The log-likelihood function is given as

$$\begin{aligned} log\ L(y) &= \sum_{i=1}^{n} log\ \frac{q(y_i;\xi)}{Q(y;\xi)} \\ &= \sum_{i=1}^{n} log\ P[Y=y_i] + \sum_{i=1}^{n} log\ y_i - n\ log\ Q(y;\xi) \\ &= \sum_{i=1}^{n} log\left[exp\left\{ -\beta^{\alpha}\left(\frac{m}{y_i+1}\right)^{\alpha k}\right\} - exp\left\{ -\beta^{\alpha}\left(\frac{m}{y_i}\right)^{\alpha k}\right\}\right] + \sum_{i=1}^{n} log\ y_i \\ &\quad -n\ log\ Q(y;\xi). \end{aligned} \tag{5.1}$$

It is clear from the Eq.(5.1), that the second and third terms are constants and so the differentiation will only depend on the first term. Now differentiating Eq.(5.1) with respect to α, β, m and k, we get

$$\frac{\partial\ log\ L(y)}{\partial \alpha} = \sum_{i=1}^{n} \frac{B(y_i+1)\ C(y_i+1) - B(y_i)\ C(y_i)}{A(y_i)}, \tag{5.2}$$

$$\frac{\partial\ log\ L(y)}{\partial\beta} = \sum_{i=1}^{n} \frac{B(y_i+1)\ D(y_i+1) - B(y_i)\ D(y_i)}{A(y_i)}, \tag{5.3}$$

$$\frac{\partial\ log\ L(y)}{\partial m} = \sum_{i=1}^{n} \frac{B(y_i+1)\ E(y_i+1) - B(y_i)\ E(y_i)}{A(y_i)}, \tag{5.4}$$

$$\frac{\partial\ log\ L(y)}{\partial k} = \sum_{i=1}^{n} \frac{B(y_i+1)\ F(y_i+1) - B(y_i)\ F(y_i)}{A(y_i)}, \tag{5.5}$$

where
$A(y_i) = exp\Big\{ -\beta^{\alpha}\Big(\frac{m}{y_i+1}\Big)^{\alpha k}\Big\} - exp\Big\{ -\beta^{\alpha}\Big(\frac{m}{y_i}\Big)^{\alpha k}\Big\} = P[Y=y_i]$, $B(y_i) = exp\Big\{ -\beta^{\alpha}\Big(\frac{m}{y_i}\Big)^{\alpha k}\Big\}$,
$C(y_i) = \Big\{ -\alpha\beta\Big(\frac{m}{y_i}\Big)^{k}\Big\}^{\alpha-1}$, $D(y_i) = -\alpha\beta^{\alpha-1}\Big(\frac{m}{y_i}\Big)^{\alpha k}$, $E(y_i) = -\alpha k\beta^{\alpha}m^{\alpha k-1}y^{-\alpha k}$ and
$F(y_i) = -\alpha^2\beta^{\alpha}k\Big(\frac{m}{y_i}\Big)^{\alpha k-1}$.

Solving the non-linear equations, Eq.(5.2), (5.3), (5.4) and (5.5) by equating them to zero, we obtain the maximum likelihood estimators (mle) of $\hat{\xi} = (\hat{\alpha}, \hat{\beta}, \hat{m}, \hat{k})^T$ for the parameter vector $\xi = (\alpha, \beta, m, k)^T$. These equations do not have explicit solutions and so they are obtained by using the statistical package *optim* in R programming. Now applying the usual large sample approximation, the mle $\hat{\xi}$ can be treated as being approximately multivariate normal with variance-covariance matrix equal to the inverse of the expected Fisher's information matrix, i.e.

$$\sqrt{n}(\hat{\xi} - \xi) \longrightarrow N(0\ ,\ n\ \mathcal{I}^{-1}(\xi)),$$

where $\mathcal{I}^{-1}(\xi)$ is the limiting variance-covariance matrix of $\hat{\xi}$. The elements of 4×4 Fisher's Information matrix are as given below

$$\mathcal{I}(\hat{\xi}) = \begin{bmatrix} I_{11} & I_{12} & I_{13} & I_{14} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ I_{31} & I_{32} & I_{33} & I_{34} \\ I_{41} & I_{42} & I_{43} & I_{44} \end{bmatrix},$$

where the elements of $\mathcal{I}(\hat{\xi})$ are as defined below
$I_{11} = -\frac{\partial^2\ log\ L(y)}{\partial\alpha^2}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$, $I_{22} = -\frac{\partial^2\ log\ L(y)}{\partial\beta^2}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$,
$I_{33} = -\frac{\partial^2\ log\ L(y)}{\partial m^2}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$, $I_{44} = -\frac{\partial^2\ log\ L(y)}{\partial k^2}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$,
$I_{12} = I_{21} = -\frac{\partial^2\ log\ L(y)}{\partial\alpha\beta}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$, $I_{13} = I_{31} = -\frac{\partial^2\ log\ L(y)}{\partial\alpha m}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$,
$I_{14} = I_{41} = -\frac{\partial^2\ log\ L(y)}{\partial\alpha k}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$, $I_{23} = I_{32} = -\frac{\partial^2\ log\ L(y)}{\partial\beta m}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$,
$I_{24} = I_{42} = -\frac{\partial^2\ log\ L(y)}{\partial\beta k}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$, $I_{34} = I_{43} = -\frac{\partial^2\ log\ L(y)}{\partial mk}\Big|_{(\hat{\alpha},\hat{\beta},\hat{m},\hat{k})}$.
This can be used to obtain the $100(1-\eta)\%$ asymptotic confidence interval of the parameter

vector $\xi = (\alpha, \beta, m, k)^T$. Let, ξ_j be the j^{th} $(j = 1, 2, 3, 4)$ parameter of LBDFW model, then the $100(1-\eta)\%$ asymptotic confidence interval of ξ_j is given by

$$(\widehat{\xi_j} - z_{\frac{\eta}{2}}\sqrt{\mathcal{I}_{j,j}} \quad , \quad \widehat{\xi_j} + z_{\frac{\eta}{2}}\sqrt{\mathcal{I}_{j,j}}),$$

where $\mathcal{I}_{j,j}$ is the j^{th} diagonal element of $\mathcal{I}^{-1}(\hat{\xi})$, for $j = 1, 2, 3, 4$ and $z_{\frac{\eta}{2}}$ is the upper $\frac{\eta}{2}$ point of standard normal distribution.

6. Monte Carlo Simulation Study of Estimator Properties

To assess the performance and reliability of the maximum likelihood estimators (MLEs) for the parameters of the LBDFW distribution, a comprehensive simulation study is conducted. This study is designed to evaluate how well the MLEs perform under varying sample sizes and parameter configurations. The `fitdistr` function from the statistical software **R** is utilized for this purpose, allowing for efficient numerical estimation of parameters through optimization routines that maximize the log-likelihood function. The MLEs are particularly useful in frequentist inference but also serve as critical components in initializing Markov Chain Monte Carlo (MCMC) algorithms. In this simulation framework, 1000 independent random samples are generated from the LBDFW distribution for each of the sample sizes $n = 50, 100, 150$, and 200. The simulations are performed under two different combinations of true parameter values, selected to represent distinct shapes and scaling behaviors of the LBDFW model:

(i) Scheme I: $\alpha = 0.5,\ \beta = 0.7,\ m = 2.0,\ k = 0.8,$

(ii) Scheme II: $\alpha = 1.5,\ \beta = 5.5,\ m = 2.2,\ k = 0.4.$

For each simulated dataset, the MLEs of the parameters α, β, m, and k are computed. The performance of these estimators is evaluated using two statistical measures. The first is the empirical bias, which quantifies the average deviation of the estimated parameter from its true value across the replications. The second measure is the mean squared error (MSE), which incorporates both the variance of the estimator and the square of its bias, providing a comprehensive metric to assess the accuracy and reliability of the estimators. These metrics are computed across the 1000 replications using the following formulas:

$$\text{Bias}(\theta) = \frac{1}{1000}\sum_{s=1}^{1000}(\hat{\theta}_s - \theta), \qquad \text{MSE}(\theta) = \frac{1}{1000}\sum_{s=1}^{1000}(\hat{\theta}_s - \theta)^2,$$

where $\hat{\theta}_s$ denotes the estimate of the parameter θ from the s-th sample. Beyond the classical evaluation using MLE, the results offer valuable insights for broader estimation frameworks. The MLEs derived in this study can serve as effective initial values in iterative algorithms, thereby improving convergence and computational efficiency particularly in scenarios involving complex or multi-modal parameter spaces. A summary of the simulation outcomes, including average estimates, empirical biases, and mean squared errors (MSEs) for each parameter under the two specified scenarios, is provided in Table 2.

Table 2: The average estimate, biases, and MSEs for case I and case II.

		Scheme I				**Scheme II**			
		α=0.5, β=0.7, m=2.0, k=0.8				α=1.5, β=5.5, m=2.2, k=0.4			
	n	$\hat{\alpha}$	$\hat{\beta}$	$\hat{m}$	$\hat{k}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{m}$	$\hat{k}$
Avg Est	50	0.5405	0.8079	1.8494	0.6358	1.9521	5.5807	2.5181	0.6212
	100	0.5142	0.7865	2.5233	1.6214	1.9055	5.5302	2.4117	0.5327
	150	0.5077	0.7571	2.3217	1.3322	1.8553	5.5211	2.3005	0.5074
	200	0.5042	0.7145	2.0727	0.9041	1.8102	5.5007	2.2017	0.4424
Bias	50	0.0404	0.1079	-0.1506	-0.1642	0.1521	0.0807	0.3181	0.2212
	100	0.0142	0.0865	0.5233	0.8214	0.1055	0.0302	0.2117	0.1327
	150	0.0077	0.0571	0.3217	0.5322	0.0553	0.0211	0.1005	0.1074
	200	0.0043	0.0145	0.0727	0.1041	0.0102	0.0007	0.0017	0.0424
MSE	50	0.0361	0.2347	0.0495	0.0293	0.4127	0.5138	0.0419	0.1338
	100	0.0278	0.1229	0.0197	0.0149	0.2341	0.3346	0.0316	0.0726
	150	0.0204	0.0527	0.0111	0.0672	0.1015	0.1742	0.0214	0.0487
	200	0.0142	0.0315	0.0104	0.0218	0.0331	0.1014	0.0056	0.0204

These results illustrate the performance of the MLEs as the sample size increases. Specifically, as the sample size grows, the estimators exhibit a reduction in both bias and MSE, indicating a convergence toward the true parameter values. This observed trend highlights the consistency and efficiency of the MLEs, reinforcing their suitability for classical statistical inference. Such improvements in estimator accuracy with larger sample sizes underscore the reliability of MLEs in providing precise and unbiased parameter estimates, which further supports their application in real-world data analysis.

7. Analysis of Fit: Evaluating Model Performance

In this section, the proposed LBDFW model is applied to two real-world count datasets and its performance is compared against four existing discrete distributions: the discretized Fréchet-Weibull (DFW) model introduced by Das and Das [12], the weighted Fréchet-Weibull (WDFW) model by Das and Das [16], the discrete generalized Weibull (DGW) distribution proposed by Para and Jan [17], and the discrete generalized inverse Weibull (DGIW) distribution developed by Para and Jan [18]. The assessment of model adequacy and comparative performance is primarily based on the discrepancy coefficient (C), which is computed as the ratio of the chi-square test statistic (χ^2) to the sample size N, i.e., $C = \frac{\chi^2}{N}$. This measure was originally proposed by Makcutek [19] as an effective tool for model evaluation. A smaller value of C indicates a better fit of the model to the data. This interpretation is supported by previous studies such as Grzybek [20], Nekoukhou et al. [21], and Das and Das [16]. In addition to the discrepancy coefficient, other standard goodness-of-fit metrics are employed for comprehensive evaluation, including the negative log-likelihood (–log l) and the p-value associated with the χ^2 goodness-of-fit test. These multiple indicators allow for a robust comparison of the models' performance in capturing the structure and variability in the data. All parameter estimates for the fitted models are obtained via the maximum likelihood method using the *fitdistr* function available in the R programming environment. This approach ensures the reliability and consistency

of parameter estimation, thereby supporting the conclusions drawn from the comparative model analysis.

7.1. Application I: Microcalanus nauplii dataset

The first dataset represents the counts of *Microcalanus nauplii* in marine plankton samples, with a total of 150 observations. This dataset is sourced from the work of Bliss and Fisher (1953) and is presented in Table 3.

Table 3: The counts of Microcalanus nauplii in samples of marine plankton.

Y	0	1	2	3	4	5	6	7	8	9	10
Frequency	0	2	4	3	5	8	16	13	12	13	15
Y	11	12	13	14	15	16	17	18	19	20	21
Frequency	15	9	9	7	4	4	6	2	0	2	1

A key characteristic of this dataset is the absence of any observations with a count of 0, i.e., the frequency corresponding to the variable value $Y = 0$ is zero. This is an important point because it indicates that the data may follow a *length-biased* distribution. In a length-biased distribution, events are not randomly sampled, but rather, the likelihood of selecting an event is proportional to its "length" or magnitude. In practical terms, this often means that larger values are more likely to be observed, as they correspond to longer "lifetimes" or larger quantities in the context of sampling. Therefore, for length-biased distributions, it is typical to observe that $p_0 = 0$ (the probability of the smallest value, $Y = 0$, is zero). This feature makes the dataset a potential candidate for modeling using a length-biased distribution, particularly the LBDFW model.

Additionally, the dataset exhibits characteristics of *overdispersion*, which is another important aspect to consider when selecting an appropriate model. Overdispersion occurs when the variance of the data exceeds its mean. In this case, the variance is 17.36, which is significantly higher than the mean of 9.60. This overdispersion suggests that the data do not follow a simple Poisson or binomial distribution, both of which would require the mean and variance to be approximately equal. The LBDFW model, which accounts for varying degrees of dispersion, is suitable for modeling overdispersed data like this. Furthermore, the dataset is *platykurtic*, with a kurtosis value of 2.79, which is less than 3 (the kurtosis value for a normal distribution). Platykurtic distributions have lighter tails than the normal distribution, indicating that extreme values are less frequent than they would be in a leptokurtic distribution (which has heavy tails). The LBDFW model is also appropriate for platykurtic data, as it is capable of handling distributions with different tail behaviors, further making it a fitting choice for this dataset.

To summarize, the *Microcalanus nauplii* count data exhibits the essential features that align with the properties of the LBDFW model: it is length-biased (with zero frequency for $Y = 0$), overdispersed (variance exceeds the mean), and platykurtic (with a kurtosis less than 3). These characteristics make it an ideal candidate for modeling using the LBDFW distribution, which accounts for both the overdispersion and the non-normal shape of the data.

On the other hand, an essential diagnostic tool for understanding the underlying behavior of the hazard rate function (hrf) of a dataset is the *Total Time on Test* (TTT) transform plot. According to Aarset [22], the shape of the scaled TTT transform provides valuable insights into the nature of the hazard rate. Specifically, if the scaled TTT plot is strictly concave, it implies that the distribution has an increasing hazard rate (IHR); conversely, a strictly convex TTT plot indicates a decreasing hazard rate (DHR). These conclusions are not merely suggestive but are considered necessary and sufficient conditions under regularity assumptions. Figure 4 displays both the empirical and scaled TTT plots along with the estimated hazard rate function for Dataset I.

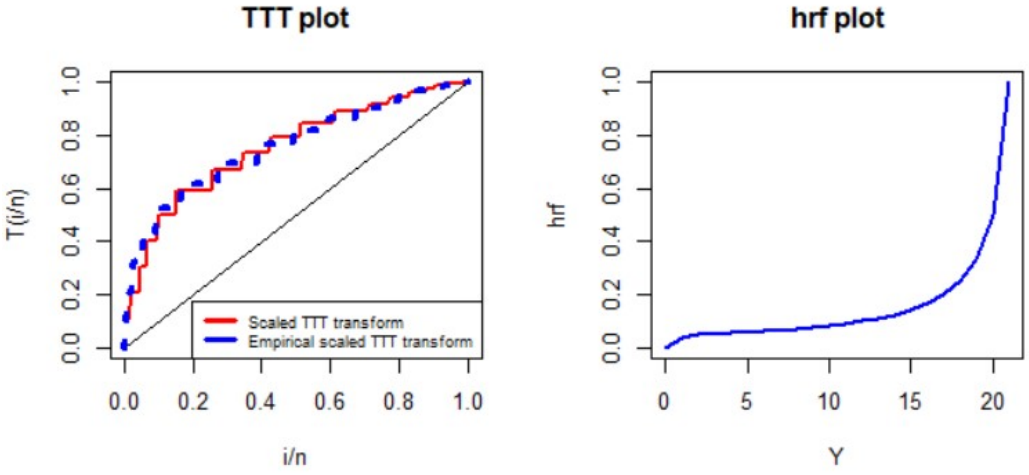

Figure 4: Empirical and fitted scaled TTT-transform plots and hrf plot for the Dataset I.

A clear concave pattern is observed in the empirical and scaled TTT curves. This pattern strongly indicates that the hazard rate is increasing over time, meaning that the likelihood of failure or occurrence of an event becomes higher as time progresses. Such a trend in the hazard rate is characteristic of aging systems or biological processes where the risk escalates with time, and it is crucial to adopt a distribution that captures this dynamic behavior. The LBDFW distribution, by construction, accommodates increasing hazard rates through its flexible parameterization. Therefore, the observed concavity in the TTT plot offers strong empirical justification for the appropriateness of the LBDFW model in fitting Dataset I. This supports the selection of the LBDFW model as a robust candidate for modeling time-to-event or count data that exhibits an increasing risk profile.

In Table 4, the MLEs of the model parameters are reported along with their corresponding standard errors (SEs) and 95% confidence intervals for all models fitted to Dataset I. These results provide insights into the precision and reliability of the parameter estimates across the considered models. Furthermore, Table 5 presents a comprehensive summary of the goodness-of-fit statistics. These measures facilitate a comparative evaluation of the models and help identify the one that best captures the underlying structure of the data. As presented in Table 5, for Dataset I, the LBDFW distribution yields the highest p-value among all competing models, along with the lowest value of the Cramér-von Mises statistic

(C). These results indicate that the LBDFW model provides a superior fit to the data in comparison to the other four considered distributions: WDFW, DFW, DGIW, and DGW. Further visual support for this conclusion is provided in Figure 5, where the observed versus estimated frequency plots for all five models are depicted. The LBDFW distribution demonstrates close alignment with the observed data across the entire range. Additionally, Figure 6 displays the cdf, quantile-quantile (Q-Q), and probability-probability (P-P) plots for Dataset I under the LBDFWD model. These plots serve as graphical diagnostic tools that assess how well the theoretical distribution matches the empirical data. The strong agreement seen in these plots further corroborates the superior performance of the LBDFW model, as supported by the numerical goodness-of-fit measures in Table 5.

Table 4: The MLEs and 95% confidence intervals for Dataset I.

Distribution	**Estimates**	**SE**	95% **CI**
DGW	$\hat{\alpha} = 2.6167$	$\hat{\alpha} = 0.1672$	(2.2890 , 2.9444)
	$\hat{\beta} = 3.5951$	$\hat{\beta} = 4.0617$	(-4.3658 , 11.5560)
	$\hat{\theta} = 0.2492$	$\hat{\theta} = 0.1449$	(-0.0348 , 0.5332)
DGIW	$\hat{\alpha} = 3.7597$	$\hat{\alpha} = 0.1004$	(3.5629 , 3.9565)
	$\hat{\beta} = 4.1752$	$\hat{\beta} = 1.2340$	(1.7566 , 6.5938)
	$\hat{\theta} = 0.3425$	$\hat{\theta} = 1.7677$	(-3.1222 , 3.8072)
DFW	$\hat{\alpha} = 0.1664$	$\hat{\alpha} = 0.6042$	(-1.0178 , 1.3506)
	$\hat{\beta} = 8.4490$	$\hat{\beta} = 7.0008$	(-5.2726 , 22.1706)
	$\hat{m} = 10.5188$	$\hat{m} = 9.9157$	(-8.9160 , 29.9536)
	$\hat{k} = 14.5766$	$\hat{k} = 8.4153$	(-1.9174 , 31.0706)
WDFW	$\hat{\alpha} = 0.7007$	$\hat{\alpha} = 1.5877$	(-2.4112 , 3.8126)
	$\hat{\beta} = 6.0632$	$\hat{\beta} = 1.9325$	(2.2755 , 9.8509)
	$\hat{m} = 9.5495$	$\hat{m} = 1.8228$	(5.9768 , 13.1222)
	$\hat{k} = 0.3089$	$\hat{k} = 1.6457$	(-2.9167 , 3.5345)
LBDFW	$\hat{\alpha} = 0.5311$	$\hat{\alpha} = 0.7143$	(-0.8689 , 1.9311)
	$\hat{\beta} = 0.0115$	$\hat{\beta} = 1.4641$	(-2.8581 , 2.8811)
	$\hat{m} = 0.3855$	$\hat{m} = 1.4981$	(-2.5508 , 3.3218)
	$\hat{k} = 0.1854$	$\hat{k} = 0.5078$	(-0.8099 , 1.1807)

Table 5: Summary of the goodness-of-fit criteria for Dataset I.

X	**OBSERVED FREQUENCY**	**EXPECTED FREQUENCY**				
		DFW	DGIW	DGW	WDFW	LBDFWD
0	0	1.35690	1.31135	1.30793	0.87026	0.00000
1	2	1.64341	1.44566	3.53676	2.48425	1.33946
2	4	2.15580	2.51402	4.98729	5.68172	2.05602
3	3	2.77937	3.15271	6.84038	8.53768	4.27637
4	5	4.17540	5.47476	9.68080	11.30281	6.55209
5	8	6.06240	7.53582	13.51861	13.04596	10.37112
6	16	7.26042	9.39889	17.99567	14.79084	16.36445
7	13	7.71637	12.79333	16.61990	15.54618	16.04524
8	12	9.88273	14.05578	15.12964	14.31499	15.71974
9	13	11.78055	14.21134	12.58317	12.09787	14.59121
10	15	12.38363	13.45702	9.56968	10.09440	12.26137
11	15	12.07360	11.72581	8.97964	8.50375	10.23122
12	9	11.25050	9.90723	6.24950	7.52497	8.57198
13	9	10.20428	8.83889	5.81230	6.35707	7.64339
14	7	9.11190	7.23774	4.97685	5.19915	6.20132
15	4	8.06855	6.01129	3.64316	4.05033	5.11563
16	4	7.11835	5.09169	2.47773	2.90984	4.08875
17	6	6.27621	4.65679	1.67329	2.47697	3.06274
18	2	5.54159	3.88119	1.54194	1.65110	2.03761
19	0	4.90656	3.02065	1.12539	1.53166	1.41331
20	2	4.36025	2.38055	1.18747	0.71814	1.08984
21	1	3.89123	1.89749	0.56290	0.31007	0.96714
N	150	150	150	150	150	150
- log l		792.4160	767.2430	784.7434	752.5601	751.6274
χ^2		34.8988	28.9782	20.1133	14.6939	10.4474
Df		13	14	11	12	11
p-value		0.0009	0.0105	0.0438	0.2586	**0.4906**
C		0.2327	0.1932	0.1341	0.0979	**0.0696**

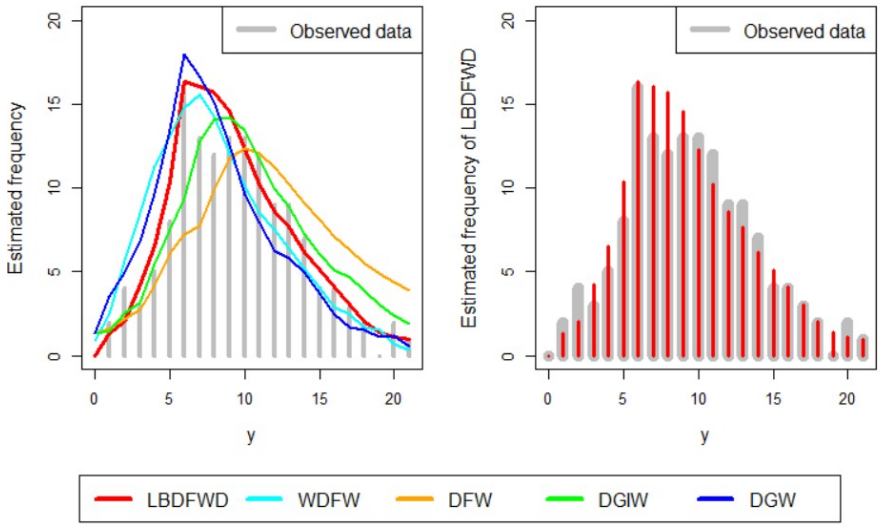

Figure 5: Comparison of observed and estimated frequency distributions for Dataset I.

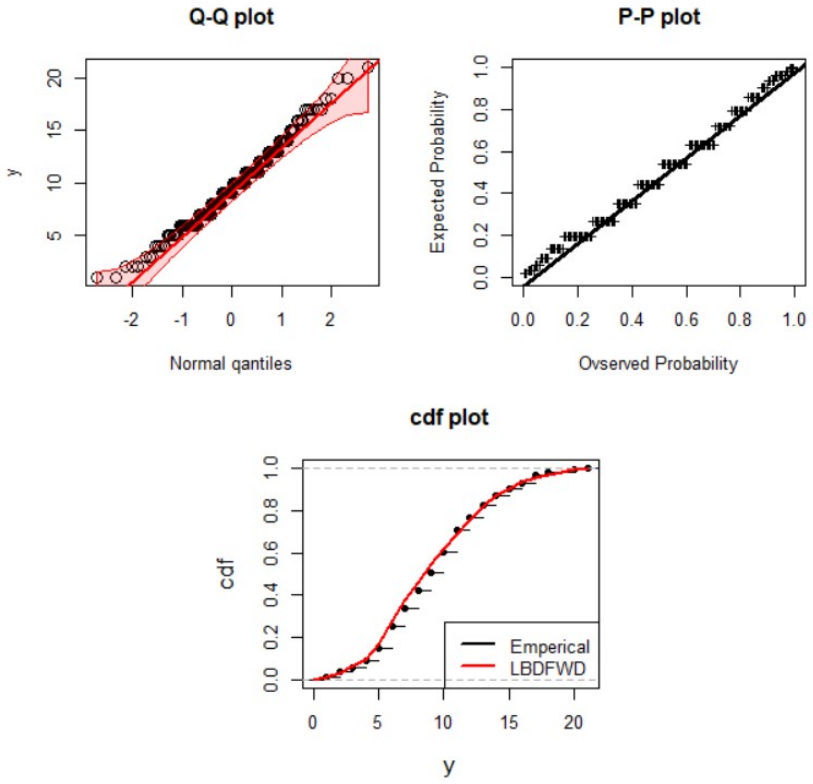

Figure 6: The Q-Q, P-P and cdf plots for Dataset I using LBDFW model.

7.2. Application II: Rifle shots dataset

The second dataset, originally presented in Déniz et al. [1], consists of the number of successful hits from ten rifle shots aimed at each of 100 targets. This yields a total of 100 observations and is displayed in Table 6.

Table 6: The results of ten shots fired a rifle at each of 100 targets.

Y	0	1	2	3	4	5	6	7	8	9	10	Total
Frequency	0	2	4	10	22	26	18	12	4	2	0	100

An important initial observation is the absence of zeros in the data, i.e., the frequency for the count value $Y = 0$ is zero. This characteristic aligns well with the assumptions of a length-biased distribution, where the probability mass at zero is typically zero or negligible. From a descriptive statistical perspective, the dataset has a mean of 5.00 and a variance of 2.67, indicating a clear case of underdispersion, where the spread of the data is less than expected under a Poisson framework. Furthermore, the kurtosis value of 2.96, being less than the normal value of 3, indicates a platykurtic distribution one with lighter tails than the normal distribution. These statistical features suggest that traditional discrete distributions, such as the Poisson or negative binomial, may not adequately capture the shape

and dispersion pattern present in this dataset. Given these distributional characteristics underdispersion, platykurtic behavior, and the absence of zero counts, the LBDFW model emerges as a particularly suitable candidate for modeling this dataset. The flexibility of the LBDFW model in accommodating light-tailed distributions and capturing reduced variability enhances its applicability to data of this nature. Further statistical fitting and goodness-of-fit assessments reinforce the model's suitability, as will be demonstrated in subsequent sections.

For the second dataset, a detailed analysis is carried out to assess the suitability of the proposed model using both graphical and numerical diagnostic tools. Figure 7 presents the empirical and fitted scaled TTT transform plots alongside the estimated hrf. The observed concavity in both empirical and fitted scaled TTT curves suggests that the underlying hazard rate is increasing. According to the result by Aarset [22], such a concave shape strongly indicates an increasing hazard rate, which aligns well with the characteristics of the LBDFW model. Therefore, the LBDFW model is a reasonable and statistically supported choice for modeling this dataset.

The parameter estimates obtained via the MLE method, along with their corresponding SEs and 95% confidence intervals (CIs), are summarized in Table 7 for all competing fitted models. This tabular presentation facilitates the comparison of estimation precision and interval coverage across models. Furthermore, Table 8 reports the goodness-of-fit statistics for each model, including the $-\log \ell$, the p-value of the χ^2, and C criterion. These metrics collectively provide comprehensive insight into how well each model captures the observed data. Notably, the LBDFW model exhibits superior fit characteristics across multiple indicators, reinforcing its flexibility and robustness in handling underdispersed and platykurtic count data, as observed in Dataset II.

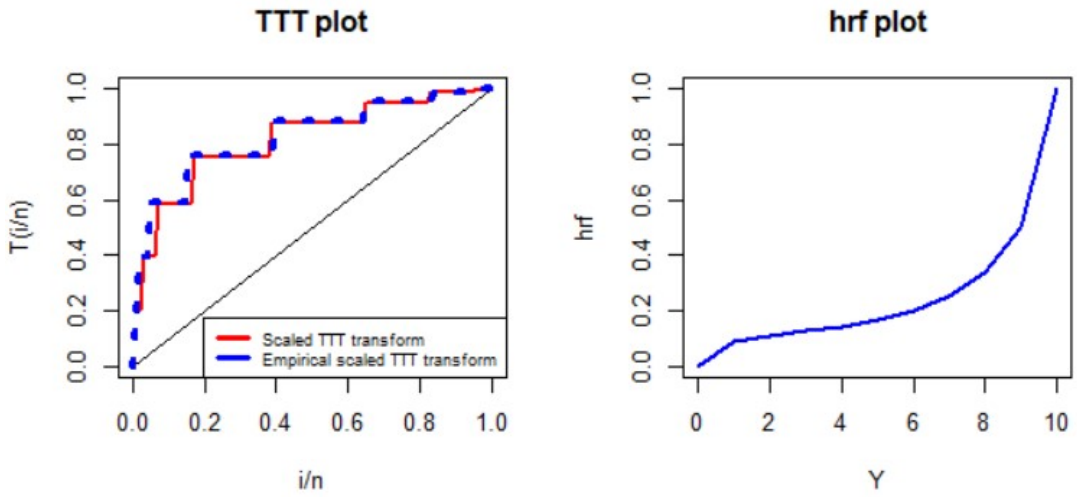

Figure 7: Empirical and fitted scaled TTT-transform plots and hrf plot for the dataset II.

Table 7: The MLEs and 95% confidence intervals for Dataset II.

Distribution	**Estimates**	**SE**	95% **CI**
DGW	$\hat{\alpha} = 5.7988$	$\hat{\alpha} = 0.3029$	(5.2051 , 6.3925)
	$\hat{\beta} = 4.4631$	$\hat{\beta} = 2.1567$	(0.2360 , 8.6902)
	$\hat{\theta} = 0.7488$	$\hat{\theta} = 0.6874$	(-0.5985 , 2.0961)
DGIW	$\hat{\alpha} = 2.8750$	$\hat{\alpha} = 0.1914$	(2.4999 , 3.2501)
	$\hat{\beta} = 2.7365$	$\hat{\beta} = 2.1761$	(-1.5287 , 7.0017)
	$\hat{\theta} = 0.1157$	$\hat{\theta} = 2.0966$	(-3.9936 , 4.2250)
DFW	$\hat{\alpha} = 2.7354$	$\hat{\alpha} = 5.6777$	(-8.3929 , 13.8637)
	$\hat{\beta} = 3.7154$	$\hat{\beta} = 2.0182$	(-0.2403 , 7.6711)
	$\hat{m} = 2.3143$	$\hat{m} = 1.8381$	(-1.2884 , 5.9170)
	$\hat{k} = 2.5415$	$\hat{k} = 1.9332$	(-1.2476 , 6.3306)
WDFW	$\hat{\alpha} = 0.0664$	$\hat{\alpha} = 2.8493$	(-5.5182 , 5.6510)
	$\hat{\beta} = 2.4692$	$\hat{\beta} = 9.3324$	(-15.8223 , 20.7607)
	$\hat{m} = 11.1764$	$\hat{m} = 3.0994$	(5.1016 , 17.2512)
	$\hat{k} = 1.3679$	$\hat{k} = 0.8770$	(-0.3510 , 3.0868)
LBDFWD	$\hat{\alpha} = 0.5771$	$\hat{\alpha} = 4.8544$	(-8.9375 , 10.0917)
	$\hat{\beta} = 0.0050$	$\hat{\beta} = 3.4300$	(-6.7178 , 6.7278)
	$\hat{m} = 0.3124$	$\hat{m} = 1.3341$	(-2.3024 , 2.9272)
	$\hat{k} = 0.2389$	$\hat{k} = 2.6820$	(-5.0178 , 5.4956)

Table 8: Summary of the goodness-of-fit criteria for Dataset II.

X	**OBSERVED FREQUENCY**	**EXPECTED FREQUENCY**				
		WDFW	DFW	DGIW	DGW	LBDFWD
0	0	20.92571	0.00000	2.57227	1.28007	0.00000
1	2	15.72648	8.78675	7.84169	5.97291	1.80404
2	4	12.70339	14.07008	11.29771	10.49810	3.61697
3	10	10.82606	16.59386	13.68973	16.55380	10.28112
4	22	9.91835	17.62054	14.78582	20.37221	23.27671
5	26	7.95892	16.20030	15.02092	19.19780	24.21558
6	18	6.95038	13.11319	13.82923	12.54988	18.13423
7	12	5.90002	8.09352	9.24622	6.92631	11.04672
8	4	4.81552	3.90680	6.15143	3.56770	4.95902
9	2	2.70390	0.92583	3.49832	2.02982	1.87372
10	0	1.57126	0.68917	2.06666	1.05140	0.79188
N	100	100	100	100	100	100
- log l		460.5225	463.2380	462.2228	465.0553	460.5295
χ^2		120.3101	24.9363	24.3823	18.5098	1.5678
Df		7	6	6	6	4
p-value		< 0.0001	0.0004	0.0004	0.0051	**0.8146**
C		1.2031	0.2494	0.2438	0.1851	**0.0157**

Figure 8 displays the observed frequency distribution of Dataset II alongside the estimated frequencies obtained from the LBDFW, WDFW, DFW, DGIW, and DGW models. These graphical comparisons allow for a visual assessment of each model's ability to replicate the underlying data structure. Additionally, Figure 9 presents the cdf, Q–Q, and P–P plots for Dataset II. These diagnostic plots offer further evidence supporting the goodness-of-fit results summarized in Table 8, highlighting the superior performance of the LBDFW

model in capturing the empirical distribution of the data.

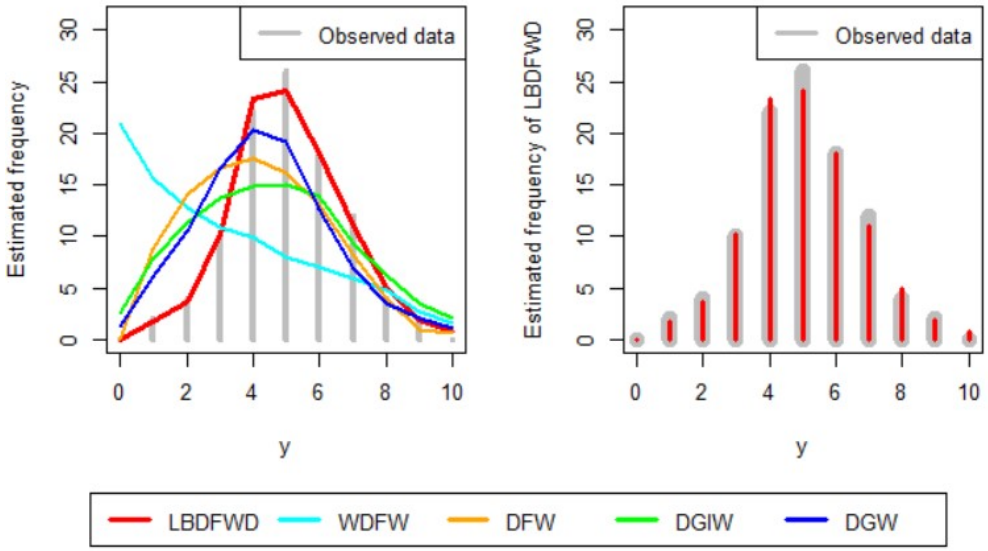

Figure 8: Comparison of observed and estimated frequency distributions for Dataset II.

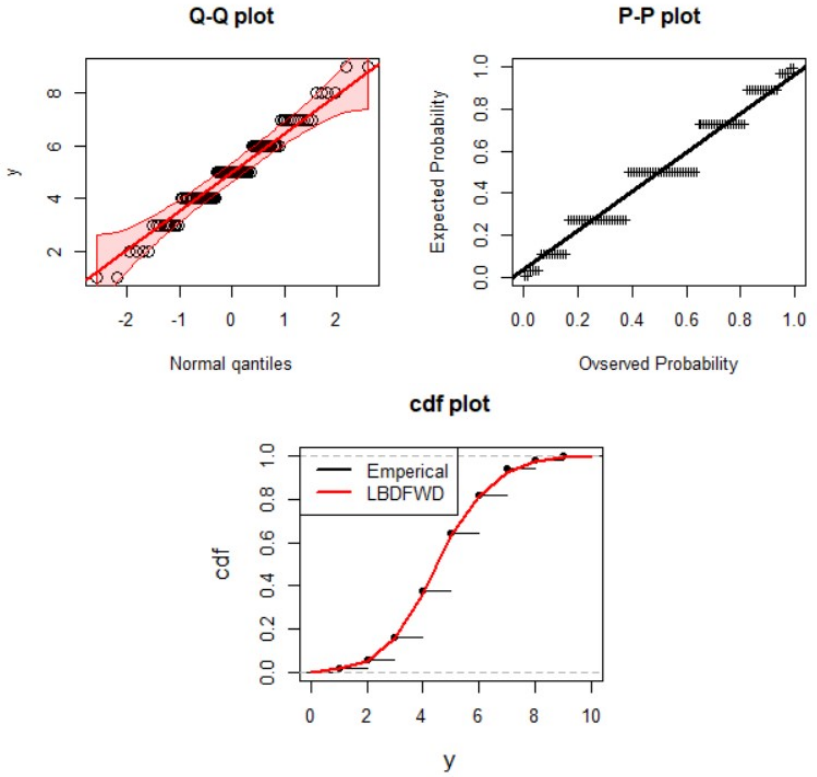

Figure 9: The Q-Q, P-P and cdf plots for Dataset II using LBDFW model.

8. Conclusions

This study presented a new statistical model called the length-biased discretized Fréchet-Weibull (LBDFW) distribution. It was developed to meet the increasing demand for flexible tools that can handle complex and diverse types of data. The LBDFW model is designed to capture key features often found in real datasets, such as overdispersion, underdispersion, zero-inflation, and skewed tails. By analyzing its probability and hazard functions, the model showed strong adaptability, accurately representing a wide range of hazard shapes including increasing, decreasing, bathtub, and unimodal forms. Parameters of the model were estimated using the maximum likelihood method. A simulation study confirmed the estimators were efficient and reliable under different conditions, supporting the model's use in standard statistical inference. The model was also tested on real-world count data, where it showed a better fit compared to several well-known discrete models, both in statistical tests and visual comparisons. Looking ahead, the LBDFW model offers promising directions for further research. It can be extended using Bayesian methods like MCMC to improve inference in more complex scenarios. There's also potential for developing multivariate versions to capture relationships across multiple variables or adapting the model for regression analysis with covariates. Additionally, applying the model to censored data, such as in progressive or hybrid censoring schemes, could be explored. From a practical perspective, the LBDFW model holds value in areas like reliability engineering, epidemiology, environmental science, and actuarial work especially when dealing with skewed or heavy-tailed count data.

Data Availability Statement

The data sets are available in the paper.

Conflicts of Interest

The authors declare no conflict of interests.

References

[1] R. A. Fisher. The effects of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6:13–25, 1934.

[2] C. R. Rao. On discrete distributions arising out of methods of ascertainment. *Sankhya*, 27(2):311–324, 1965.

[3] D. Cox. *Renewal Theory*. Methuen's monographs on applied probability and statistics, London; New York, 1962. John Wiley & Sons.

[4] G. P. Patil and C. R. Rao. Weighted distribution and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34:179–189, 1978.

[5] J. H. Gove. Estimation and application of size-biased distributions in forestry. In *Modeling Forest Systems: Workshop on the Interface Between Reality, Modeling and the Parameter Estimation Processes*, pages 201–212, 2002.

[6] M. Pandya, S. Pandya, and P. Andharia. Bayes estimation of weibull length biased distribution. *Asian Journal of Current Engineering and Maths*, 2(1):44–49, 2013.

[7] K. Mir, A. Ahmed, and J. A. Reshi. Structural properties of length biased beta distribution of first kind. *American Journal of Engineering Research*, 2(2):1–6, 2013.

[8] P. Seenoi, T. Supapakorn, and W. Bodhisuwan. The length-biased exponentiated inverted weibull distribution. *International Journal of Pure and Applied Mathematics*, 92(2):191–206, 2014.

[9] A. Ahmad, S. Ahmad, and A. Ahmed. Length-biased weighted lomax distribution: Statistical properties and application. *Pakistan Journal of Statistics and Operation Research*, 12(2):245–255, 2016.

[10] T. Chaito and M. Khamkong. The length–biased weibull–rayleigh distribution for application to hydrological data. *Lobachevskii Journal of Mathematics*, 42(13):3253–3263, 2021.

[11] G. P. Patil and C. R. Rao. *Weighted Distributions: A Survey of Their Application*. North Holland, Amsterdam, The Netherlands, 1977.

[12] D. Das and B. Das. Discretized fréchet–weibull distribution: Properties and application. *Journal of the Indian Society for Probability and Statistics*, 24(1):1–40, 2023.

[13] C. C. Kokonendji and M. P. Casany. A note on weighted count distributions. *Journal of Statistical Theory and Applications*, 11(4):337–352, 2012.

[14] J. D. Castillo and M. P. Casany. Weighted poisson distributions for overdispersion and underdispersion situations. *Annals of the Institute of Statistical Mathematics*, 50(3):567–585, 1998.

[15] F. Steutel and K. van Harn. *Infinite Divisibility of Probability Distributions on the Real Line*. Pure and Applied Mathematics. Marcel Dekker Inc., 2004.

[16] D. Das and B. Das. On weighted discretized fréchet-weibull distribution with application to real-life data. *Applied Mathematics & Information Sciences*, 7(5):791–806, 2023.

[17] B. A. Para and T. R. Jan. Discrete generalized weibull distribution: Properties and applications in medical sciences. *Pakistan Journal of Statistics*, 33(5):337–354, 2017.

[18] B. A. Para and T. R. Jan. On three parameters discrete generalized inverse weibull distribution: Properties and applications. *Annals of Data Science*, 6(3):549–570, 2019.

[19] J. Makcutek. A generalization of the geometric distribution and its application in quantitative linguistics. *Romanian Reports in Physics*, 60(3):501–509, 2008.

[20] P. Grzybek. On the systematic and system-based study of grapheme frequencies: A re-analysis of german letter frequencies. *Glottometrics*, 15:82–91, 2007.

[21] V. Nekoukhou, M. H. Alamatsaz, and H. Bidram. A discrete analog of the generalized exponential distribution. *Communications in Statistics - Theory and Methods*, 41(11):2000–2013, 2012.

[22] M. V. Aarset. How to identify a bathtub hazard rate. *IEEE Transactions on Relia-*

bility, 36(1):106–108, 1987.