Computational Journal of Mathematical and Statistical Sciences
x(x), -
DOI:10.21608/cjmss.2025.373562.1146
https://cjmss.journals.ekb.eg/

A Discrete Expansion of the Lindley Distribution: Mathematical and Statistical Characterizations with Estimation Techniques, Simulation, and Goodness-of-Fit Analysis

Diksha Das[1], Mohamed F. Abouelenein[2], Bhanita Das[1], Partha Jyoti Hazarika[3], Mahmoud El-Morshedy[4,*], Noura Roushdy[2], Mohamed S. Eliwa[5]

[1] Department of Statistics, North-Eastern Hill University, 793022, 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, Dibrugarh University, Assam 786004, India.
[4] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia.
[5] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

* **Correspondence:** m.elmorshedy@psau.edu.sa

Abstract: The objective of this paper is to introduce the discretized two-parameter Lindley (D2PL) distribution, a novel discrete probability model that extends the classical Lindley distribution into the discrete domain. This distribution features two parameters, providing greater modeling flexibility and encompassing existing discrete models, such as the one-parameter discrete Lindley and geometric distributions. The paper thoroughly characterizes the D2PL distribution, deriving several key properties essential for reliability modeling. Additional analyses include infinite divisibility, log-convexity, and classical moment measures such as raw moments, dispersion index, skewness, and kurtosis, offering insights into the distribution's shape and tail behavior. The probability mass function of the D2PL distribution can exhibit uni-modal and decreasing forms, making it useful for asymmetric count data. Its hazard rate function can model various failure rate patterns, accommodating both under-dispersed and over-dispersed count data. Parameter estimation is conducted through maximum likelihood and method of moments, with Monte Carlo simulations verifying the efficiency and reliability of the estimators. The model's robustness is further demonstrated through applications on real-world count datasets, showing superior goodness of fit over established discrete distributions, highlighting its effectiveness for complex discrete data.

Keywords: Survival-based discretization technique; Lindley distribution; Failure analysis; Simulation; Parameter estimation; Data analysis.
Mathematics Subject Classification: 60E05, 68R01, 62E10, 62F10.
Received: 17 May 2025; Revised: 6 August 2025; Accepted: 12 August 2025; Online: 22 August 2025.

1. Introduction

In scientific research, we often encounter situations where the variable of interest is continuous, but it is more practical to measure it in a discrete form. For life testing or reliability experiments, numerous continuous models are available in the literature. However, it can be impractical or inconvenient to measure the lifetime of a device on a continuous scale. For instance, when evaluating the reliability of an on/off-switching device, the lifetime of the switch is determined by the number of times it is operated, which is essentially a discrete random variable. The reliability of an airplane tire is typically assessed based on the number of landings it can withstand. The lifespan of an electric circuit is often measured by the frequency of breakdowns within a month, and the effectiveness of a particular drug is determined by the number of days it remains viable before its expiration date. In survival analysis, one might want to record the number of days that a patient has survived since starting therapy or the number of days from remission to relapse. The survival time of a patient with a brain hemorrhage is determined by the length of time they are observed.

In these scenarios, while a variable seems continuous, the actual lifetimes are recorded discretely, thus, they are considered discrete random variables. These instances illustrate that although continuous lifetime models might not always be measured on a continuous scale, they can often be approached as discrete random variables. Over the past two decades, typical discrete distributions like geometric, Poisson, and negative binomial have been applied to model lifetime data. However, Alamatsaz et al. [4] asserted that these established discrete distributions do not always provide the best fit for both count and time data. For instance, the Poisson distribution is commonly used to model counts but is not suitable for modeling time. Additionally, binomial and negative binomial distributions are not widely regarded as effective models for reliability, failure times, counts, and similar data. This is partly because they are not defined over the entire set of non-negative integers. Therefore, it is essential to identify more suitable discrete lifetime distributions that are specifically tailored to various types of discrete data, such as reliability and failure times. These contexts necessitate creating discrete forms of current continuous distributions. In recent decades, there has been substantial interest from researchers in converting continuous probability distributions into discrete forms. This has resulted in numerous established discretized versions of existing continuous distributions. Investigating the discretization of statistical models is vital for managing discrete lifetime data and count data across various fields, including biological and medical sciences, physical sciences, engineering, agriculture, and more. Diverse techniques exist in academic literature for discretizing continuous distributions. Chakraborty [8] offers a thorough review of numerous discrete derivation strategies for continuous distributions. The survival function approach ranks among the most widely adopted methods for this purpose. It utilizes the survival function associated with the continuous model. Nakagawa and Osaki [30] pioneered this technique to formulate the discrete Weibull distribution. In reliability theory, lifetime models are classified based on their survival functions and other reliability characteristics. For example, we have the increasing failure rate (IFR) and decreasing failure rate (DFR) classes, as well as the increasing failure rate average (IFRA) and decreasing failure rate average (DFRA) classes. There are also the increasing mean residual lifetime (IMRL) and decreasing mean residual lifetime (DMRL) classes (see Kemp [24]). When a continuous distribution is discretized using the survival function approach, it retains the same functional form as the original survival function. As a result, many important properties related to reliability theory are preserved in the discrete form. Therefore, discretizing a continuous lifetime

model provides a straightforward method for deriving a corresponding discrete lifetime model.

Several prominent continuous distributions have been discretized using this survival-function-based method. For instance, Nakagawa and Osaki [30] introduced a discrete form of the Weibull distribution. Roy applied a similar approach to develop discrete analogues of the normal and Rayleigh distributions in [34] and [35], respectively. Krishna and Pundir explored discrete versions of the Burr and Pareto distributions [26], while Chakraborty and Chakravarty proposed a discrete form of the gamma distribution [9]. Chakraborty [8] provided a comprehensive list of the various discretized distributions available in the literature. Recently, Olivera et al. [31] proposed the discrete power Lindley model as an effective distribution for both count and failure time datasets. The discrete generalized inverse Weibull model was studied by Para and Jan [32] in relation to medical science data. Additionally, Chakraborty et al. [10] proposed the discrete Gumbel model, which is suitable for various skewed datasets. Almetwally et al. [5] discussed overview of discrete distributions in modelling COVID-19 data sets. Gillariose et al. [19] explored the discrete Weibull Marshall–Olkin family of distributions. Finally, in the context of daily new COVID-19 cases from various countries, Ibrahim and Almetwally [22] reported the discrete Marshall-Olkin Lomax model. Abd EL-Hady et al. [2] discussed discrete exponentiated generalized family of distributions More recently, Das et al. [14] proposed a new right-skewed distribution known as the discrete generalized Gompertz distribution. In another study, Das and Das [13] developed the discretized Fréchet–Weibull distribution, which is suitable for modeling datasets with both over-dispersion and under-dispersion. Aljohani et al. [3] discussed the discrete Marshall-Olkin length-biased exponential distribution in the context of analyzing COVID-19 data from China and Pakistan. Additionally, Balubaid et al. [6] developed the Discrete Weibull Exponential distribution, which features both symmetric and asymmetric shapes in its probability mass function. Barbiero and Hitaj [7] studied the discrete half-logistic distribution, particularly its applications in the insurance sector. Finally, Hadi and Khudhair [20] examined the discrete Fréchet distribution concerning the duration of stay for a hundred stroke patients in a hospital.

Lindley [27, 28] introduced a one-parameter continuous distribution known as the Lindley distribution. This distribution has been shown to be a mixture of the exponential distribution with parameter θ and the gamma distribution with parameters 2 and θ, with a mixing proportion given by $p = \frac{\theta}{\theta+1}$. Despite its potential, the relevance of the Lindley distribution as a model for lifetime data has been largely overlooked in the literature. The popularity of the exponential distribution continued to dominate the modeling of lifetime data until Ghitany et al. [17] discussed various mathematical properties of the Lindley distribution. They demonstrated that the probability distribution function (pdf) of the continuous Lindley distribution serves as a better model for waiting times and survival times compared to the exponential distribution. A number of researchers have developed advanced variations of the Lindley distribution. In recent studies, Chesneau et al. [11] presented the inverted modified Lindley distribution. In a subsequent investigation, Chesneau et al. [12] explored the modified Lindley distribution, while Gillariose et al. [18] introduced the Marshall-Olkin Modified Lindley distribution, among other variants.

Notably, the Lindley distribution has only an increasing hazard rate function. In 2013, Shanker et al. introduced the two-parameter Lindley (TPL) distribution in their study [36]. This distribution is useful for modeling waiting times and survival data. The TPL distribution can exhibit both decreasing and increasing hazard rate functions. Additionally, the mean of the distribution is always greater than the mode, which indicates that the distribution is positively skewed. It also exhibits both over-dispersion

and under-dispersion. Moreover, the TPL distribution encompasses the one-parameter Lindley distribution as a special case. This distribution can be represented as a mixture of exponential (θ) and gamma $(2, \theta)$ distributions, with a mixing proportion given by $p = \frac{\theta}{\alpha+\theta}$. The pdf of the TPL distribution is defined as follows:

$$f_X(x) = \frac{\theta^2}{\theta + \alpha}(1 + \alpha x)e^{-\theta x}\,; \qquad x > 0, \tag{1.1}$$

where the parameters $\theta > 0,\ \alpha > -\theta$, and the survival function (sf) is given by

$$S_X(x) = \frac{\theta + \alpha + \alpha\theta x}{\theta + \alpha}e^{-\theta x}\,; \qquad x > 0,\ \theta > 0,\ \alpha > -\theta. \tag{1.2}$$

The discrete version of the one-parameter Lindley distribution was introduced by Emilio and Enrique [16]. This model serves as a viable alternative to the Poisson distribution for analyzing automobile claim frequency data. The discrete Lindley distribution is characterized by its unimodal shape, leptokurtic nature, and positive skewness. However, one significant limitation of this distribution is its inability to effectively model under-dispersed datasets. This limitation has led us to develop a discretized version of the two-parameter Lindley distribution, known as the discretized two-parameter Lindley (D2PL) distribution. Thus, the development of this new model is based on several key points: The D2PL model possesses several important features that enhance its practical utility. It offers explicit forms for its probability mass function, reliability function, and failure rate, making the distribution both accessible and computationally efficient for various applications. One of the key advantages of the D2PL distribution is its flexibility and notability in handling dispersion data; it can effectively discuss and evaluate dataset exhibiting either overdispersion or underdispersion, which is often encountered in real scenarios. Further, the D2PL model is capable of representing different forms of the hazard rate function, including the well-known bathtub shape, as well as strictly increasing or decreasing patterns. This adaptability makes it a suitable candidate for reliability analysis and life data modeling. Moreover, the D2PL model is particularly well-suited for asymmetric data, especially when the dataset is right-skewed or shows various degrees of peakedness or tail behavior. Such flexibility allows for more accurate modeling of non-symmetric discrete phenomena. Empirical comparisons with other discrete distributions consistently highlight the superior performance of the D2PL distribution, particularly in terms of goodness-of-fit across various kinds of dataset. This strong performance underscores its reliability and relevance for researchers and practitioners working with dependent or complex discrete dataset.

The remainder of this paper is organized to provide a comprehensive understanding of the proposed D2PL distribution and its applications. Section 2 begins with the formal derivation of the D2PL distribution, followed by an in-depth exploration of its key structural characteristics, notable special cases, and graphical representations to illustrate its flexibility. This section also includes the procedure for generating random samples from the distribution, which is essential for both theoretical exploration and simulation studies. In Section 3, we extend the foundational results by presenting additional mathematical insights that deepen our understanding of the distribution's behavior. Section 4 is devoted to deriving various distributional properties, such as moments, reliability measures, and tail behavior, which are critical for statistical modeling and interpretation. Estimation techniques are the focus of Section 5, where we examine both maximum likelihood estimation and the method of moments, discussing their implementation and theoretical underpinnings. To evaluate the performance of these estimators, a detailed simulation study is carried out in Section 6, providing empirical evidence of

their accuracy and robustness across different parameter settings. Section 7 illustrates the practical relevance of the D2PL distribution by applying it to two real-life count datasets. Here, we compare its fitting capabilities with those of alternative models, demonstrating its superiority in capturing complex data patterns. Finally, Section 8 summarizes the main findings of the study and suggests avenues for future research.

2. Discretized two-parameter Lindley distribution

The discretized two-parameter Lindley distribution has been derived by using the discretization approach, after the parameterization, $\lambda = e^{-\theta}$ in the survival function of the continuous two-parameter Lindley distribution as defined in Eq. (1.2). Then the probability mass function (pmf) of the D2PL distribution is derived as follows

$$\begin{aligned} P[Y=y] &= \frac{e^{-\theta y}}{\theta+\alpha}\Big[\{\alpha+\theta(1+\alpha y)\}(1-e^{-\theta})-\alpha\theta e^{-\theta}\Big] \\ &= \frac{\lambda^y}{\alpha-log\lambda}\Big[\alpha\lambda\, log\lambda+(1-\lambda)\{\alpha-log\lambda^{(1+\alpha y)}\}\Big], \end{aligned} \tag{2.1}$$

where the parameter $\lambda = e^{-\theta}$, such that $0 < \lambda < 1$ and $\alpha > log\lambda$. Further, to prove that the pmf equation's is a valid pmf, it must satisfy the condition $\sum_{y=0}^{\infty} P[Y=y] = 1$, where

$$\begin{aligned} \sum_{y=0}^{\infty} P[Y=y] &= \sum_{y=0}^{\infty} \frac{\lambda^y}{\alpha-log\lambda}\Big[\alpha\lambda\, log\lambda+(1-\lambda)\{\alpha-log\lambda^{(1+\alpha y)}\}\Big] \\ &= \frac{1}{\alpha-log\lambda}\left[\{\alpha\lambda\, log\lambda+(1-\lambda)\alpha-(1-\lambda)\, log\lambda\}\Big(\sum_{y=0}^{\infty}\lambda^y\Big)-\alpha(1-\lambda)\, log\lambda\Big(\sum_{y=0}^{\infty} y\lambda^y\Big)\right] \\ &= \frac{1}{\alpha-log\lambda}\left[\{\alpha\lambda\, log\lambda+(1-\lambda)\alpha-(1-\lambda)\, log\lambda\}\Big(\frac{1}{1-\lambda}\Big)-\alpha(1-\lambda)\, log\lambda\Big(\frac{\lambda}{(1-\lambda)^2}\Big)\right] \\ &= \frac{1}{\alpha-log\lambda}\left[\frac{\alpha\lambda\, log\lambda}{1-\lambda}+\alpha-log\lambda-\frac{\alpha\lambda\, log\lambda}{1-\lambda}\right] \\ &= 1. \end{aligned}$$

Hence, the form of the pmf derived in Eq. (2.1) is a valid pmf. The cumulative distribution function (cdf) of D2PL distribution is obtained as

$$\begin{aligned} F_Y(y) &= P(Y\le y)=\sum_{k=0}^{y} P[Y=k] \\ &= \sum_{k=0}^{y}\frac{\lambda^y}{\alpha-log\lambda}[\alpha\lambda\, log\lambda+(1-\lambda)\{\alpha-log\lambda^{(1+\alpha y)}\}] \\ &= \frac{\alpha}{\alpha-log\lambda}[1-\lambda^{y+1}+\{(2+y)\lambda^{y+1}-1\}\, log\lambda] \qquad ; y\in \mathbf{Z}_+. \end{aligned} \tag{2.2}$$

The sf of the D2PL distribution can be formulated as

$$S_Y(y) = \frac{\alpha\lambda^{y+1}\{1-(2+y)log\lambda\}+(\alpha-1)\, log\lambda}{\alpha-log\lambda} \qquad ; y\in \mathbf{Z}_+. \tag{2.3}$$

The hazard rate function (hrf) of the D2PL distribution is given by

$$h_Y(y) = \frac{\lambda^y[\alpha\lambda\, log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}]}{\alpha\lambda^{y+1}\{1-(2+y)\, log\lambda\} + (\alpha-1)\, log\lambda} \quad ; y \in \mathbf{Z}_+. \tag{2.4}$$

The proposed D2PL distribution reduces to the following mentioned distributions under specific parameter values:

1. For $\alpha = 1$, D2PL(λ, α) reduces to discrete Lindley(λ) (see Emilio and Enrique [16]).
2. For $\alpha = 0$ and $\lambda = 1 - p$, D2PL(λ, α) reduces to Geometric(p), where $0 < p < 1$.

2.1. Graphical representation

The possible shapes of the pmf and hrf of D2PL distribution for different values of the parameter (λ, α) are presented in Figure 1 and Figure 2, respectively.

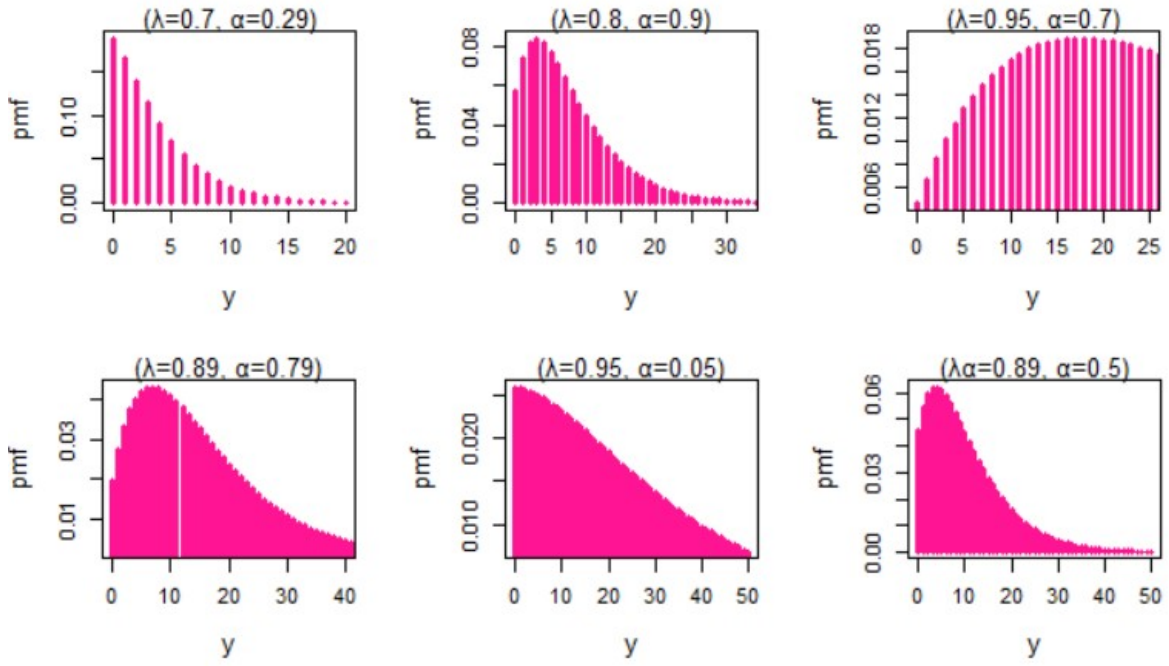

Figure 1. The pmf plots of the D2PL distribution for different parameter values of λ and α.

In Figure 1, we can see that the pmf of the D2PL for various values of the parameter (λ, α) has a long right tail. This indicates that the proposed distribution is positively skewed. It is also apparent that the shape of the pmf of this distribution can be decreasing, or unimodal. In Figure 2, it can be noted the possible shapes of the hrf of the D2PL model for different values of the parameter (λ, α). The hrf function of the D2PL model can be increasing, decreasing or uni-modal-shaped, depending on the choice of parameter values. It is noteworthy that, in all cases, the value remains below 1. A key characteristic of the proposed distribution is its inverse bathtub-shaped hazard function, along with its capability to capture both increasing and decreasing hazard rate patterns features that are rarely observed in count distributions.

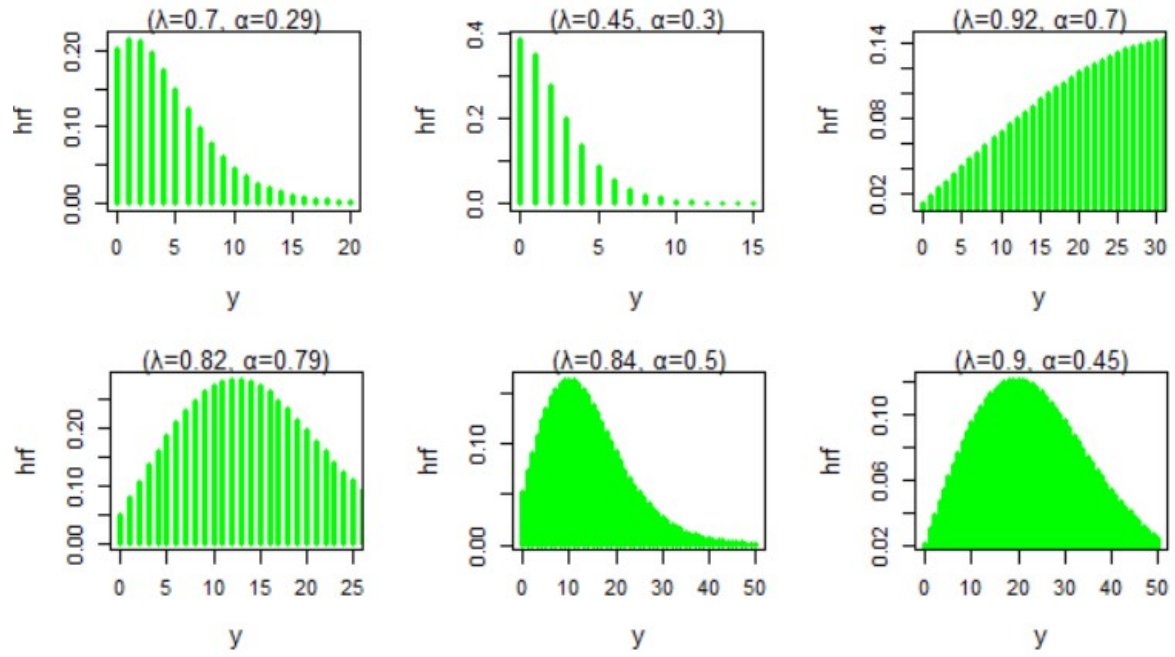

Figure 2. The hrf plots of the D2PL distribution for different parameter values of λ and α.

2.2. Random number generation from the D2PL distribution

The pdf of the continuous TPL distribution can be expressed as a combination of exponential and gamma distributions. To generate a random sample from the D2PL distribution, we use the concept that the discrete version of the continuous TPL distribution corresponds to the D2PL distribution. This leads to the following algorithm:

Step 1: Generate a sample "u" of size "n" from Uniform (0,1) and specify the initial values of the parameters λ and α.

Step 2: Compute $\frac{\theta}{\theta+\alpha}$, where $\theta = -log\ \lambda$.

Step 3: For $(i = 1, 2, ..., n)$, if $u_i \leq \frac{\theta}{\theta+\alpha}$, generate $X_i \sim$ Exponential (θ), otherwise generate $X_i \sim$ Gamma $(2, \theta)$, where $\theta = -log\ \lambda$.

Step 4: Now obtain Y_i, using the form $Y_i = \lfloor X_i \rfloor$, where $\lfloor . \rfloor$ is the floor function.

3. Supplementary results

3.1. Infinite divisibility of D2PL distribution

The mathematical concept of infinite divisibility reports a fundamental structural characteristic of probability models. As listed by Steutel and Van Harn [37], a model can be tested for infinite divisibility by verifying either of the two specific conditions listed below.

Lemma 3.1. *For a discrete probability model p_y, a necessary requirement for it to be infinitely divisible is that $p_0 > 0$.*

Lemma 3.2. *If a discrete model sequence p_j, where $j \in \mathbb{Z}_+$, is infinitely divisible, then each term must satisfy $p_j \leq e^{-1}$ for all $j \in \mathbb{N}$.*

Theorem 3.3. *The D2PL model with parameters (λ, α) generally exhibits infinite divisibility.*

Proof. To generate this, we first verify the condition given in Lemma 3.1. Specifically, we demonstrate that

$$p_0 = \mathbb{P}[Y = 0\,;\, \lambda, \alpha] > 0, \quad \text{for all} \quad 0 < \lambda < 1, \alpha > log\lambda.$$

For example, assume $y = 0$, $\lambda = 0.8$, and $\alpha = 0.34$, we find $p_0 = 0.0922 > 0$, satisfying the requirement. Next, referring to Lemma 3.2, we need to obtain that

$$p_j = \mathbb{P}[Y = j\,;\, \lambda, \alpha] < e^{-1}, \quad \text{for all} \quad j \in \mathbb{N}, \quad \text{and} \quad 0 < \lambda < 1, \alpha > log\lambda.$$

As an illustrative case, let $y = 1$, $\lambda = 0.6$, and $\alpha = 0.25$. The resulting probability is $p_1 = 0.2993$, which is clearly less than $e^{-1} \approx 0.3679$. So, the condition is satisfied. □

Figure 1 clearly shows that there can be many combinations of parameter values satisfying the conditions in Lemma 3.1 and Lemma 3.2, indicating that the D2PL distribution is infinite divisible.

3.2. Monotonic property of D2PL distribution

To prove that D2PL distribution is log-convex, it is sufficient to show that

$$\{P[Y = y; \lambda, \alpha]\}^2 \;\leq\; P[Y = y + 1; \lambda, \alpha]\,.\,P[Y = y - 1; \lambda, \alpha],$$

where $y \in \mathbf{Z}_+$ and the parameters $0 < \lambda < 1$ and $\alpha > log\ \lambda$, otherwise it is log-concave. For this, if we can show that the ratio

$$\frac{P[Y = y + 1; \lambda, \alpha]}{P[Y = y; \lambda, \alpha]} = \frac{\lambda^{y+1}[\alpha\lambda\ log\lambda + (1 - \lambda)\{\alpha - log\lambda^{(1+\alpha\ \overline{y+1})}\}]}{\lambda^{y}[\alpha\lambda\ log\lambda + (1 - \lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}]} \qquad ; y \in \mathbf{Z}_+\ ,$$

is an increasing function of y, it implies that the model is log-convex (otherwise it is log-concave). For this distribution, when considering different sets of combinations for the parameters λ and α, the ratio shows both decreasing and increasing functions of y. Therefore, depending on the values of the parameters considered, the pmf of the D2PL distribution behaves as either log-convex or log-concave. When the pmf of a discrete random variable is log-convex (or log-concave), it is proven that the hazard function is non-increasing (or increasing). As a result, the D2PL distribution has a decreasing failure rate (DFR), and an increasing failure rate (IFR) distribution. This information is also represented in Figure 2.

3.3. Distribution of order statistics

Order statistics are a fundamental tool in inference and non-parametric statistics. In this subsection, we present results related to order statistics for the D2PL distribution. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample drawn from the D2PL distribution, and let $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ denote the corresponding order statistics. The cdf of the r^{th} order statistic from the D2PL distribution can be expressed as follows:

$$F_{Y_{(r)}}(y_{(r)}) \;=\; \sum_{r=i}^{n} \binom{n}{r} [F_Y(y_{(r)}; \lambda, \alpha)]^r [1 - F_Y(y_{(r)}; \lambda, \alpha)]^{n-r}$$

$$= \left(\frac{1}{\alpha - log\lambda}\right)^n \sum_{r=i}^{n} \binom{n}{r} \alpha^r \left[1 - \lambda^{y_{(r)}+1} + \{(2 + y_{(r)})\, \lambda^{y_{(r)}+1} - 1\}\, log\lambda\right]^r$$
$$\times \left[\alpha\lambda^{y_{(r)}+1}\{1 - (2 + y_{(r)})\, log\lambda\} + (\alpha - 1)\, log\lambda\right]^{n-r}. \quad (3.1)$$

Using the binomial expansion for $[1 - F_Y(y_{(r)}; \lambda, \alpha)]^{n-r}$, we get

$$[1 - F_Y(y_{(r)}; \lambda, \alpha)]^{n-r} = \sum_{j=0}^{n-r} \binom{n-r}{j} (-1)^j [F_Y(y_{(r)}; \lambda, \alpha)]^j.$$

Therefore,

$$F_{Y_{(r)}}(y_{(r)}) = \sum_{r=i}^{n} \sum_{j=0}^{n-r} \binom{n}{r} \binom{n-r}{j} (-1)^j [F_Y(y_{(r)}; \lambda, \alpha)]^{r+j}$$
$$= \sum_{r=i}^{n} \sum_{j=0}^{n-r} \binom{n}{r} \binom{n-r}{j} (-1)^j \left\{\frac{\alpha\, [1 - \lambda^{y_{(r)}+1} + \{(2 + y_{(r)})\lambda^{y_{(r)}+1} - 1\}\, log\lambda]}{\alpha - log\lambda}\right\}^{r+j}. \quad (3.2)$$

And the corresponding pmf of the r^{th} order statistics of D2PL can be expressed by

$$P[Y_{(r)} = y_{(r)}] = \binom{n}{r} [F_Y(y_{(r)}; \lambda, \alpha)]^{r-1} \times [1 - F_Y(y_{(r)}; \lambda, \alpha)]^{n-r} \times P[Y = y_{(r)}; \lambda, \alpha]$$
$$= \left(\frac{1}{\alpha - log\lambda}\right)^n \binom{n}{r} \alpha^{r-1}\, \lambda^{y_{(r)}} \left[1 - \lambda^{y_{(r)}+1} + \{(2 + y_{(r)})\lambda^{y_{(r)}+1} - 1\}\, log\lambda\right]^{r-1}$$
$$\times \left[\alpha\lambda^{y_{(r)}+1}\{1 - (2 + y_{(r)})log\lambda\} + (\alpha - 1)\, log\lambda\right]^{n-r}$$
$$\times \left[\alpha\lambda\, log\lambda + (1 - \lambda)\{\alpha - log\lambda^{(1+\alpha y_{(r)})}\}\right]. \quad (3.3)$$

The pmf of the smallest order statistics $Y_{(1)} = min(Y_1, Y_2, \ldots, Y_n)$ of D2PL distribution is given by

$$P[Y_{(1)} = y_{(1)}] = \frac{n\, \lambda^{y_{(1)}}}{(\alpha - log\lambda)^n} \left[\alpha\lambda^{y_{(1)}+1}\{1 - (2 + y_{(1)})\, log\lambda\} + (\alpha - 1)\, log\lambda\right]^{n-1}$$
$$\times \left[\alpha\lambda\, log\lambda + (1 - \lambda)\{\alpha - log\lambda^{(1+\alpha y_{(1)})}\}\right]. \quad (3.4)$$

The pmf of the largest order statistics $Y_{(n)} = max(Y_1, Y_2, \ldots, Y_n)$ of D2PL distribution is given by

$$P[Y_{(n)} = y_{(n)}] = \frac{\alpha^{n-1}\, \lambda^{y_{(n)}}}{(\alpha - log\lambda)^n} \left[1 - \lambda^{y_{(n)}+1} + \{(2 + y_{(n)})\lambda^{y_{(n)}+1} - 1\}\, log\lambda\right]^{n-1}$$
$$\times \left[\alpha\lambda\, log\lambda + (1 - \lambda)\{\alpha - log\lambda^{(1+\alpha y_{(n)})}\}\right]. \quad (3.5)$$

4. Distributional properties of D2PL distribution

4.1. Moment generation function

The moment generating function (mgf) of the D2PL model is obtained as

$$M_Y(t) = \sum_{y=0}^{\infty} e^{ty} P[Y = y]$$

$$= \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\ log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}], \tag{4.1}$$

where $0 < \lambda < 1, \alpha > log\lambda; \quad y \in \mathbf{Z}_+$.

Theorem 4.1. *The r^{th} moment about the origin of the D2PL distribution is of the form*

$$\mu_r^{'} = \sum_{y=0}^{\infty} y^r \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\ log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}], \tag{4.2}$$

where $0 < \lambda < 1, \alpha > log\lambda; \quad y \in \mathbf{Z}_+$.

Proof. Differentiating the mgf of the D2PL distribution as in Eq.(4.1), we get

$$M_Y^{(r)}(t) = \frac{d^r}{dt^r} M_Y(t) = \sum_{y=0}^{\infty} y^r e^{ty} \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\ log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}].$$

Then, the r^{th} moment about origin of D2PL distribution is given by

$$\mu_r^{'} = M_Y^{(r)}(t)\Big|_{t=0} = \sum_{y=0}^{\infty} y^r \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\ log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}].$$

□

4.1.1. Moments

Using $r = 1, 2, 3$ and 4, in Eq.(4.2) after some tedious computations the first four moments about the origin of the D2PL distribution are obtained as follows:

$$\mu_1^{'} = \frac{\lambda[\alpha(1-\lambda) - (1+\alpha-\lambda)\ log\lambda]}{(\alpha - log\lambda)\ (1-\lambda)^2}, \tag{4.3}$$

$$\mu_2^{'} = \frac{\lambda[\alpha(1-\lambda^2) - \{(\alpha+1) + 3\alpha\lambda - \lambda^2\}\ log\lambda]}{(\alpha - log\lambda)\ (1-\lambda)^3}, \tag{4.4}$$

$$\mu_3^{'} = \frac{\lambda[\alpha(1+3\lambda-3\lambda^2-\lambda^3) - \{(1+\alpha) + (10\alpha+3)\lambda + (7\alpha-3)\lambda^2 - \lambda^3\}\ log\lambda]}{(\alpha - log\lambda)\ (1-\lambda)^4}, \tag{4.5}$$

$$\mu_4^{'} = \lambda\Big[\alpha(1+10\lambda-10\lambda^3-\lambda^4) - \{(1+\alpha) + (25\alpha+10)\lambda + 55\alpha\lambda^2 + (15\alpha-8)\lambda^3 - (10\alpha+1)\lambda^4\}\ log\lambda\Big] \Big/ (\alpha - log\lambda)\ (1-\lambda)^5. \tag{4.6}$$

Clearly, the moments about origin of the D2PL distribution exist, and after some complex calculations, their closed forms can be derived. Using the above forms of the moments about the origin, one can easily obtain the moments about mean of the D2PL distribution. Thus, the variance of the D2PL distribution is obtained as

$$\sigma^2 = \mu_2^{'} - \mu_1^{'2} = \frac{\lambda[(\alpha - log\lambda)(1-\lambda)A_1 - \lambda\ \{A_2\}^2]}{(\alpha - log\lambda)^2\ (1-\lambda)^4}, \tag{4.7}$$

where $A_1 = \alpha(1-\lambda^2) - \{(\alpha+1) + 3\alpha\lambda - \lambda^2\}\, log\lambda$ and $A_2 = \alpha(1-\lambda) - (1+\alpha-\lambda)\, log\lambda$.

From Eq. (4.3) and (4.7) it is clear that as the values of λ and α increase, the mean and variance also increase. This relationship is demonstrated in Figure 3.

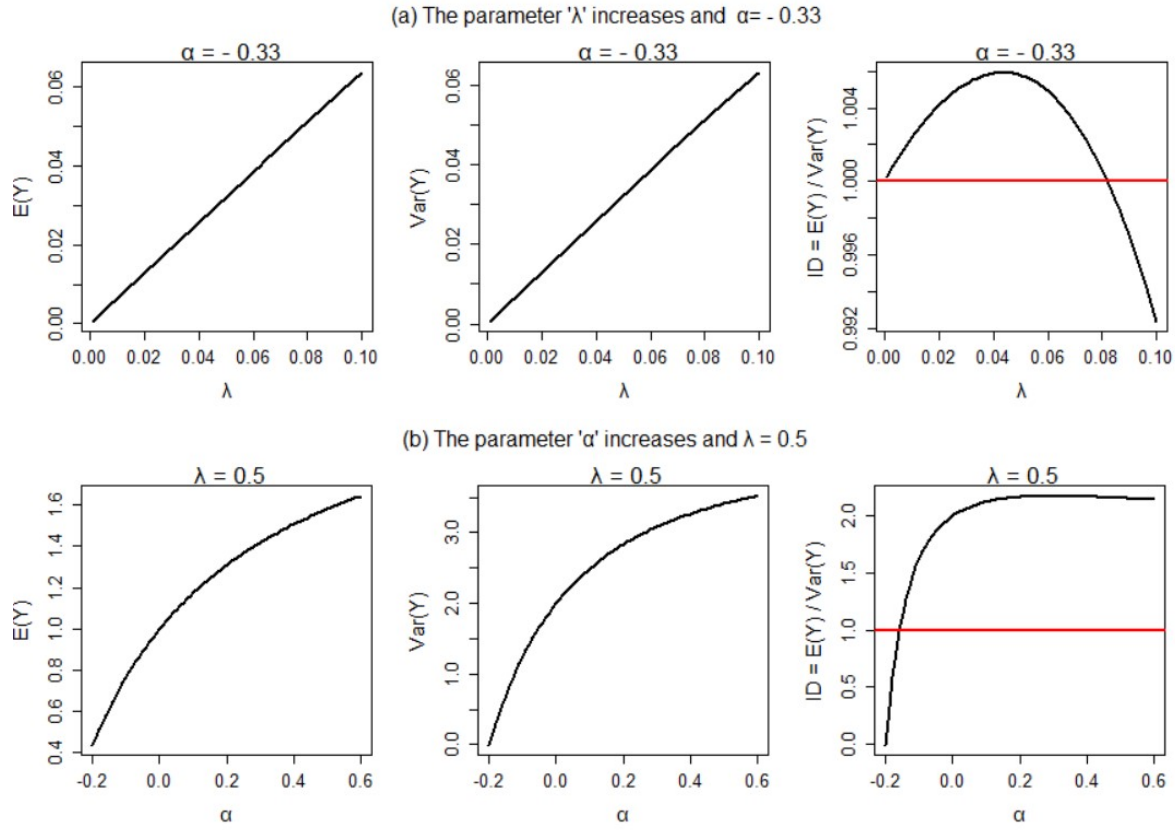

Figure 3. The impact of λ and α on mean, variance and index of dispersion for the D2PL distribution.

Figure 3 depicts how changing the parameters λ and α affects the mean and variance, as well as the index of dispersion. Figure 3(a) shows the impact of changing the parameter λ while keeping the parameter α fixed. In contrast, Figure 3(b) illustrates the impact of changing the parameter α while keeping the parameter λ fixed. Moreover, Figure 3 (a) and (b), also illustrate that the index of dispersion can be both greater and less than one, indicating that the D2PL distribution is both over-dispersed and under-dispersed.

4.2. Probability generating function

The probability generating function (pgf) of the D2PL model is obtained as

$$\begin{aligned} G_Y(t) &= \sum_{y=0}^{\infty} t^y P[Y=y] \\ &= \sum_{y=0}^{\infty} \frac{(\lambda t)^y}{\alpha - log\lambda}[\alpha\lambda\, log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}], \end{aligned} \tag{4.8}$$

where $0 < \lambda < 1, \alpha > log\lambda; \quad y \in \mathbf{Z}_+$.

Theorem 4.2. *The r^{th} factorial moment of D2PL distribution is of the form*

$$\mu_{(r)} = \sum_{y=0}^{\infty} y_{(r-1)} \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\, log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}], \tag{4.9}$$

where $y_{(r-1)} = y\,(y-1)\,(y-2)...(y-r+1)$.

Proof. Differentiating the mgf of the D2PL distribution as in Eq.(4.8), we get

$$G_Y^{(r)}(t) = \frac{d^r}{dt^r} G_Y(t) = \sum_{y=0}^{\infty} y\,(y-1)\,(y-2)...(y-r+1)\, t^{y-r} \times \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\, log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}].$$

Then, the r^{th} factorial moment of D2PL distribution is given by

$$\mu_{(r)} = G_Y^{(r)}(t)\big|_{t=1} = \sum_{y=0}^{\infty} y_{(r-1)} \frac{\lambda^y}{\alpha - log\lambda}[\alpha\lambda\, log\lambda + (1-\lambda)\{\alpha - log\lambda^{(1+\alpha y)}\}],$$

where $y_{(r-1)} = y\,(y-1)\,(y-2)...(y-r+1)$. □

4.3. Numerical computation

In this subsection, we will calculate the values of mean (μ), variance (σ^2), index of dispersion (ID), skewness (S_k), and kurtosis (K_r) for the D2PL distribution, considering different combinations of the parameters λ and α. The skewness and kurtosis are obtained using the formula:

$$S_k = \frac{\mu_3}{{\mu_2}^{3/2}} \qquad \text{and} \qquad K_r = \frac{\mu_4}{{\mu_2}^2}.$$

In Tables 1 the values are calculated for the fixed value of α, while λ is increasing. And in Table 2 the values are calculated for the fixed value of λ, while α increases. These numerical results are obtained using the R software.

Table 1. Descriptive statistics of the D2PL model as "λ" increases.

λ	$\alpha = -0.3$					λ	$\alpha = -0.4$				
	μ	σ^2	ID	S_k	K_r		μ	σ^2	ID	S_k	K_r
0.07	0.048	0.049	0.979	4.637	30.197	0.06	0.032	0.031	1.032	5.182	30.674
0.10	0.069	0.069	1.000	3.810	22.148	0.068	0.036	0.035	1.029	4.763	25.280
0.12	0.082	0.083	0.988	3.389	18.111	0.075	0.039	0.038	1.026	4.426	20.902
0.15	0.103	0.102	1.009	2.814	12.241	0.086	0.045	0.042	1.071	3.925	14.184
0.17	0.116	0.113	1.027	2.409	7.692	0.091	0.048	0.044	1.091	3.701	11.052
0.19	0.129	0.123	1.049	1.939	1.882	0.100	0.051	0.047	1.085	3.288	5.056

Table 2. Descriptive statistics of the D2PL model as "α" increases.

α	$\lambda = 0.15$					α	$\lambda = 0.4$				
	μ	σ^2	ID	S_k	K_r		μ	σ^2	ID	S_k	K_r
-0.34	0.091	0.084	1.083	2.325	1.752	-0.05	0.608	0.971	0.626	2.215	12.627
-0.30	0.103	0.102	1.009	2.814	12.241	0.3	0.918	1.634	0.561	1.967	8.877
-0.24	0.119	0.127	0.937	3.069	17.772	1.4	1.282	2.168	0.591	1.605	4.750
-0.15	0.143	0.161	0.888	3.108	19.451	2.5	1.412	2.294	0.616	1.509	3.269
0.06	0.189	0.224	0.844	2.904	18.201	3.2	1.458	2.332	0.625	1.480	2.709
0.12	0.199	0.239	0.833	2.840	17.710	4.5	1.513	2.369	0.639	1.448	2.021

From the above Table 1 and Table 2, the following observations can be made: The D2PL model demonstrates strong adaptability in handling a variety of dataset characteristics. It is particularly effective for datasets exhibiting positive skewness, making it well-suited for asymmetric models. Furthermore, the D2PL model accommodates data with different kurtosis forms, including both leptokurtic and platykurtic distributions, offering flexibility in capturing various tail behaviors. Another notable property of the D2PL model is that its mean and variance tend to increase as the parameters λ and α grow, allowing dynamic adjustment of the scale and spread of the dataset. Moreover, the D2PL distribution proves useful in applications involving both over-dispersed ($ID > 1$) and under-dispersed ($ID < 1$) data, enhancing its applicability in real discrete dataset analysis.

5. Parameter Estimation

5.1. Maximum likelihood estimation

In this section, the estimation of the D2PL parameters has been discussed using the maximum likelihood estimation (MLE) method. Let $y_1, y_2, y_3, ..., y_n$ be a random sample from the D2PL distribution, then the corresponding likelihood function is given as

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

The log-likelihood function is given as

$$\begin{aligned} log\ L(y) &= \sum_{i=1}^{n} log\ P[Y = y_i] \\ &= \sum_{i=1}^{n} \Big\{ y_i\ log\lambda + log\left[\alpha\lambda\ log\lambda + (1 - \lambda\{\alpha - (1 + \alpha\ y_i)\ log\lambda\})\right] - log\left[\alpha - log\lambda\right] \Big\} \\ &= \sum_{i=1}^{n} y_i\ log\lambda + \sum_{i=1}^{n} log\left[\alpha\lambda\ log\lambda + (1 - \lambda\{\alpha - (1 + \alpha\ y_i)\ log\lambda\})\right] - n\ log\left[\alpha - log\lambda\right] \end{aligned} \quad (5.1)$$

Differentiating Eq.5.1 with respect to λ and α, we get

$$\frac{\partial logL(y)}{\partial\lambda} = \frac{1}{\lambda}\Big[n\bar{y} + \frac{n}{\alpha - log\lambda} + \sum_{i=1}^{n} \frac{\{1 + \alpha(1 + y_i)\}\ \lambda\ log\lambda + (\lambda - 1)(1 + \alpha y_i)}{b(y_i)}\Big], \quad (5.2)$$

$$\frac{\partial logL(y)}{\partial\alpha} = \sum_{i=1}^{n} \frac{\{\lambda + (1 - \lambda)y_i\}\ log\lambda - (1 - \lambda)}{b(y_i)} - \frac{n}{\alpha - log\lambda}, \quad (5.3)$$

where $b(y_i) = [\alpha\lambda\, log\lambda + (1 - \lambda\{\alpha - (1 + \alpha\, y_i)\, log\lambda\})]$. To estimate the parameter vector $\theta = (\lambda, \alpha)^T$, we solve the nonlinear equations (5.2) and (5.3) by equating them to zero and applying the Newton–Raphson iterative technique. This yields the maximum likelihood estimates $\hat{\theta} = (\hat{\lambda}, \hat{\alpha})^T$. For constructing confidence intervals for the estimated parameters, the Fisher information matrix $\mathcal{I}(\theta)$ is essential. This matrix is obtained by computing the second-order partial derivatives of the log-likelihood function with respect to the model parameters. Since the exact analytical form of $\mathcal{I}(\theta)$ may be complex to derive, it is often approximated numerically for practical implementation.

$$\mathcal{I}(\hat{\theta}) = \begin{bmatrix} -\frac{\partial^2 L}{\partial \lambda^2}\Big|_{(\hat{\lambda},\hat{\alpha})} & -\frac{\partial^2 L}{\partial \lambda \partial \alpha}\Big|_{(\hat{\lambda},\hat{\alpha})} \\ -\frac{\partial^2 L}{\partial \alpha \partial \lambda}\Big|_{(\hat{\lambda},\hat{\alpha})} & -\frac{\partial^2 L}{\partial \alpha^2}\Big|_{(\hat{\lambda},\hat{\alpha})} \end{bmatrix}.$$

5.2. Moment method of estimation

Under the moment method of estimation (MME), the parameters λ and α can be obtained by solving the equations

$$\frac{\lambda[\alpha(1-\lambda) - (1 + \alpha - \lambda)\, log\lambda]}{(\alpha - log\lambda)\,(1-\lambda)^2} = m_1,$$

and

$$\frac{\lambda[\alpha(1-\lambda^2) - \{(\alpha+1) + 3\alpha\lambda - \lambda^2\}\, log\lambda]}{(\alpha - log\lambda)\,(1-\lambda)^3} = m_2,$$

where m_1 and m_2 are the mean of the observed sample and the second raw moment observed, respectively. Alternatively, moment estimates can be obtained following the method suggested by Khan et al. [25]. According to this method, the parameters λ and α can be estimated by minimizing the equation

$$\left[\frac{\lambda[\alpha(1-\lambda) - (1 + \alpha - \lambda)\, log\lambda]}{(\alpha - log\lambda)\,(1-\lambda)^2} - m_1\right]^2 + \left[\frac{\lambda[\alpha(1-\lambda^2) - \{(\alpha+1) + 3\alpha\lambda - \lambda^2\}\, log\lambda]}{(\alpha - log\lambda)\,(1-\lambda)^3} - m_2\right]^2.$$

6. Simulation study

In this section, the performance of the maximum likelihood estimators and the method of moments estimators is evaluated through a simulation study conducted using the statistical software R (*optimal* package) based on data generated in Section 2.2. In this study, we generated 1000 samples of sizes 50, 80, 120, and 150 from a D2PL distribution with known parameter values of λ and α. Four different schemes were considered: (i) $\lambda = 0.5$, $\alpha = 0.02$, (ii) $\lambda = 0.8$, $\alpha = 0.5$, $(iii)\lambda = 0.2$, $\alpha = -0.3$ and (iv) $\lambda = 0.07$, $\alpha = 1.5$. In this subsection, the average values, bias, and mean squared error (MSE) of the parameter estimators (λ, α) for the D2PL distribution are estimated. Additionally, the average width of the 95% confidence interval estimators (AW) and the coverage probabilities (CP) for the maximum likelihood estimators and the method of moments estimators are calculated. The results of the simulation analysis are shown in Table 3, 4 and 5. It is important to note that all the values in these tables represent the average estimates for 1000 samples.

Table 3. The average estimates, average biases, variance, MSEs, AWs, and CPs for scheme I.

		Scheme I: $\lambda = 0.5, \alpha = 0.02$			
		MLE		**MME**	
	Size "n"	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$
Average Estimate	50	0.5421	0.0152	0.4503	0.0341
	100	0.5332	0.0163	0.4712	0.0272
	150	0.5201	0.0171	0.4781	0.0242
	200	0.5040	0.0188	0.4845	0.0211
Average Bias	50	0.0421	-0.0048	-0.0497	0.0141
	100	0.0332	-0.0037	-0.0288	0.0072
	150	0.0201	-0.0029	-0.0219	0.0042
	200	0.0040	-0.0012	-0.0146	0.0011
MSE	50	0.1262	0.1322	1.1342	0.8677
	100	0.0947	0.1041	0.5022	0.6214
	150	0.0751	0.0642	0.0835	0.1217
	200	0.0421	0.0313	0.0311	0.0402
AW	50	0.3212	0.2005	0.1525	0.2247
	100	0.1736	0.1621	0.1214	0.1014
	150	0.1423	0.1304	0.1003	0.0762
	200	0.0511	0.1132	0.0622	0.0415
CP	50	0.835	0.871	0.849	0.902
	100	0.847	0.893	0.887	0.890
	150	0.862	0.911	0.879	0.895
	200	0.908	0.907	0.896	0.901

Table 4. The average estimates, average biases, variance, MSEs, AWs, and CPs for scheme II.

		Scheme II: $\lambda = 0.8, \alpha = 0.5$			
		MLE		**MME**	
	Size "n"	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$
Average Estimate	50	0.7204	0.4052	0.9043	0.5512
	100	0.7428	0.4321	0.8803	0.5403
	150	0.7715	0.4615	0.8421	0.5205
	200	0.7955	0.4845	0.8133	0.5032
Average Bias	50	-0.0796	-0.0948	0.1043	0.0512
	100	-0.0572	-0.0679	0.0803	0.0403
	150	-0.0285	-0.0385	0.0421	0.0205
	200	-0.0045	-0.0155	0.0133	0.0032
MSE	50	1.1052	0.8053	1.0035	0.8733
	100	1.0403	0.5007	0.8526	0.5114
	150	0.7271	0.1121	0.4241	0.2201
	200	0.3435	0.0322	0.1013	0.0541
AW	50	0.8952	0.5855	0.8825	0.8715
	100	0.6853	0.4432	0.7641	0.7045
	150	0.5274	0.2508	0.4371	0.5012
	200	0.2175	0.1141	0.2134	0.1221
CP	50	0.861	0.801	0.895	0.883
	100	0.893	0.845	0.901	0.892
	150	0.908	0.863	0.914	0.903
	200	0.905	0.871	0.907	0.912

The results of the simulation analysis presented in Tables 3 and 4 include both maximum likelihood estimation (MLE) and method of moments estimation procedures. In contrast, Table 5 focuses solely on the analysis using MLE. A look at the tables 3 and 4 reveals that in both schemes I and II, the MLE

Table 5. The average estimates, average biases, MSEs, AWs, and CPs for scheme III and scheme IV.

		Scheme III		**Scheme IV**	
		$\lambda = 0.2\ ,\ \alpha = -0.3$		$\lambda = 0.07\ ,\ \alpha = 1.5$	
	Size "n"	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\alpha}$
Average Estimate	50	0.35121	-0.45103	0.05209	1.72004
	100	0.30432	-0.38004	0.06009	1.62212
	150	0.22008	-0.32214	0.06822	1.55233
	200	0.20241	-0.29045	0.06934	1.51102
Average Bias	50	0.15121	-0.15103	-0.01791	0.22005
	100	0.10432	-0.08004	-0.00991	0.12212
	150	0.02008	-0.02214	-0.00178	0.05233
	200	0.00241	-0.00955	-0.00066	0.01102
MSE	50	0.15313	0.10044	0.14105	0.22012
	100	0.05032	0.03021	0.01517	0.14117
	150	0.02062	0.01202	0.00443	0.04331
	200	0.00413	0.00231	0.00101	0.01021
AW	50	0.72211	0.58814	0.60512	0.37723
	100	0.41165	0.32204	0.41152	0.12064
	150	0.18562	0.10521	0.20117	0.06274
	200	0.05411	0.01442	0.06271	0.00431
CP	50	0.811	0.799	0.866	0.869
	100	0.835	0.865	0.916	0.928
	150	0.884	0.909	0.938	0.961
	200	0.917	0.928	0.947	0.908

is the best method of estimation in comparison to the method of moments. However, the tables 3, 4, and 5 demonstrate that the biases and mean squared errors (MSEs) both drop to zero as the sample size n grows. This illustrates the MLEs' impartiality and constancy. Furthermore, the average width of the confidence interval estimators for MLEs decreases with increasing sample size, a trend that is constant across all tables. Furthermore, the coverage probability is never less than 0.799, even though it occasionally surpasses the nominal value of 0.95. Generally speaking, the greater the sample size, the higher the coverage probability.

7. Applications: Evaluation and decision-making of the quality of fit

In this section, we demonstrate the flexibility of the proposed D2PL distribution using two real-life datasets. The fitting of the D2PL distribution is compared with seven other competitive discrete distributions: discrete inverted Nadarajah-Haghighi (DINH), discrete Burr type II (DBX-II), discrete inverse Weibull (DIW), discrete Rayleigh (DR), discrete inverse Rayleigh (DIR), Poisson (Poi), and discrete Pareto (DPa). For each dataset, we evaluate the D2PL distribution against the other distributions using specific criteria. These criteria include the log-likelihood (log L), Chi-square (χ^2) statistic and its corresponding p-value, as well as various information criteria such as the Akaike Information Criterion (AIC), the corrected Akaike Information Criterion (AICC), the Hannan–Quinn Information Criterion (HQIC), and the Bayesian Information Criterion (BIC). We compute these selection criteria for all of the compared distributions to determine which one fits the data best. The distribution that exhibits the lowest values for AIC, AICC, HQIC, and BIC, the highest log-likelihood value, and the greatest p-value is deemed the most suitable for the given data.

7.1. Application I

The first dataset pertains to mammalian cytogenetic dosimetry lesions observed in rabbit lymphoblasts following exposure to streptonigrin (NSC-45383), a potent antitumor antibiotic known for its ability to induce DNA damage. In this study, the administered dose was set at -60 μg/kg, a concentration carefully chosen to evaluate the cytogenetic effects of streptonigrin on chromosomal integrity. The experimental design involved the analysis of chromosomal aberrations, including breaks, gaps, and rearrangements, which are critical indicators of genotoxic stress. These lesions were quantified to assess the extent of DNA damage and repair mechanisms in the lymphoblasts, providing insights into the compound's mutagenic potential. To thoroughly analyze the dataset and gain a deeper understanding of its underlying patterns, non-parametric plots were generated. These plots are particularly useful for visualizing the distribution and behavior of the data without making any assumptions about its underlying statistical properties. The resulting visualizations, as depicted in Figure 4, provide a clear and intuitive representation of key trends, outliers, and relationships within the data.

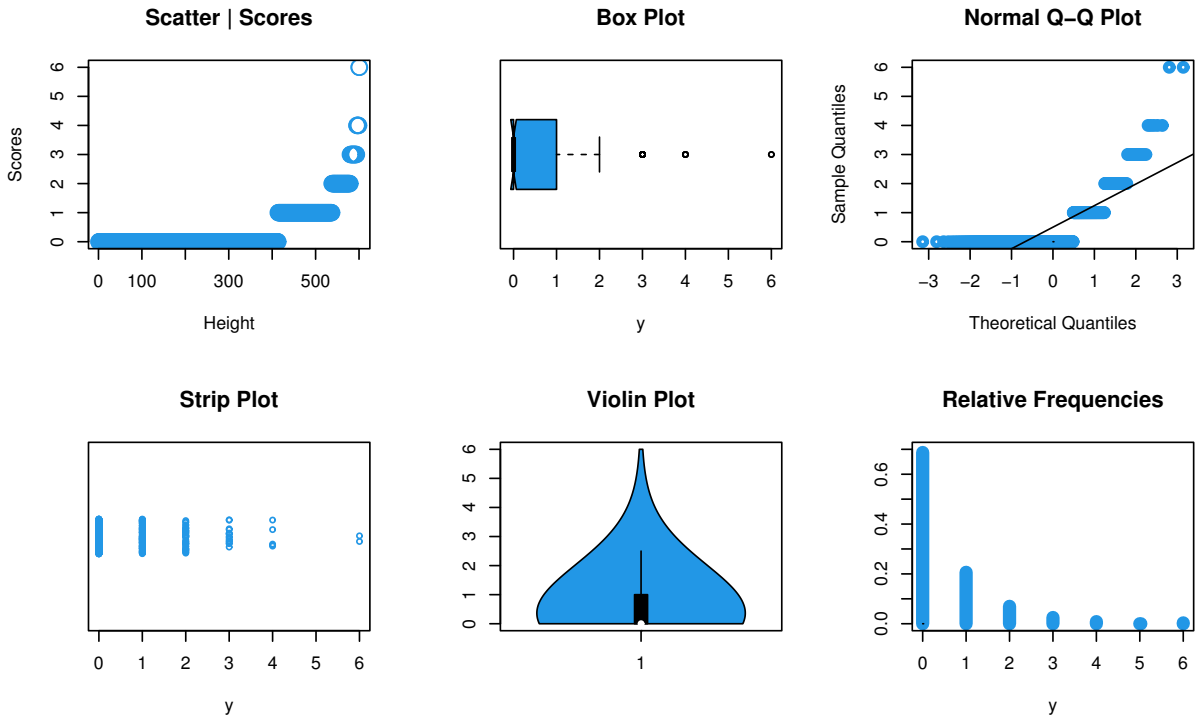

Figure 4. Plots based on non-parametric methods for dataset I.

The MLEs for the models tested, along with the results of the goodness-of-fit tests (GOFT), are presented in Table 6. Figure 5 provides an explanation and validation of the results presented in Table 5. To highlight the unique characteristics of each estimate, profile log-likelihood graphs and contour graphs were created, as illustrated in Figure 6.

Table 6. The GOFT for dataset I.

X	OF	D2PL	DB-XII	DINH	DIW	DR	DIR	Poi	DPa
0	413	407.68	412.71	414.58	411.60	273.43	411.86	374.05	447.06
1	124	131.14	132.73	91.62	136.59	274.53	134.96	177.38	84.54
2	42	42.18	33.25	31.37	29.13	50.49	29.47	42.06	29.96
3	15	13.57	11.29	15.79	10.39	2.515	10.68	6.65	13.99
4	5	4.36	4.79	9.50	4.82	0.04	5.01	0.79	7.66
5	0	1.40	2.36	6.34	2.61	0.00	2.74	0.07	4.64
6	2	0.67	3.87	31.80	5.86	0.00	6.28	0.00	13.15
Total		601	601	601	601	601	601	601	601
$-l$		556.516	560.131	595.708	564.056	700.860	564.095	582.677	580.141
MLE_α		-5.574×10^{-7}	0.187	23.1217	0.684	0.545	0.685	0.474	0.140
MLE_λ		0.3217	1.653	0.0137	2.041	–	–	–	–
AIC		1117.032	1124.262	1195.416	1132.112	1403.720	1130.190	1167.354	1162.282
CAIC		1117.052	1124.282	1195.436	1132.132	1403.727	1130.197	1167.361	1162.289
HQIC		1120.456	1127.686	1198.840	1135.536	1405.432	1131.902	1169.066	1163.994
BIC		1125.829	1133.059	1204.213	1140.909	1408.119	1134.588	1171.753	1166.681
χ^2		0.659	5.563	51.489	12.234	156.05	11.861	58.562	39.711
d.f		2	2	4	3	1	4	3	4
P.value		0.719	0.062	< 0.001	0.007	< 0.001	0.018	< 0.001	< 0.001

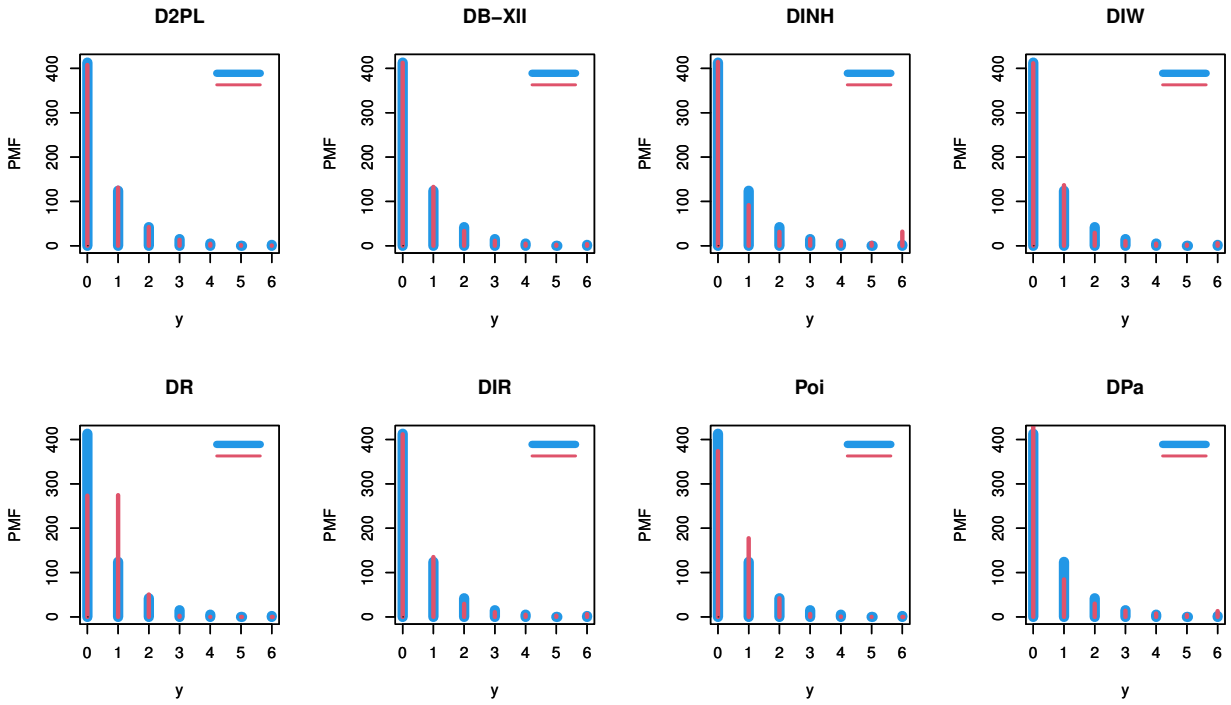

Figure 5. The estimated probability mass functions for dataset I.

It is evident from Table 6 that the AIC, AICC, HQIC, and BIC values are the smallest for the D2PL distribution while the χ^2 value is the lowest. Additionally, the p-value for D2PL is the greatest. These results indicate that the D2PL distribution provides a better fit to data set I compared to the seven other competitive distributions.

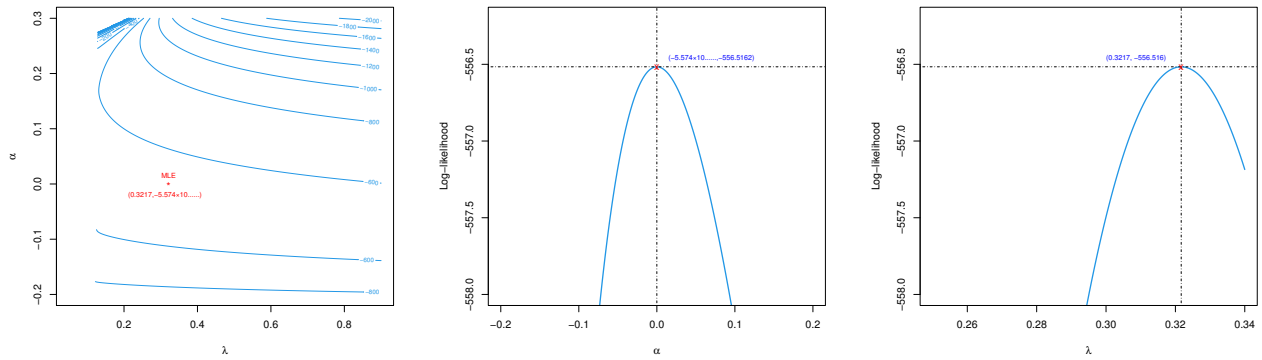

Figure 6. The contour plot and profile log-likelihood of the D2Pl for data set I.

7.2. Application II

The dataset provided captures the frequency of computer breakdowns recorded over a span of 128 consecutive weeks of operation, as documented by Hand et al. [21]. This longitudinal dataset offers valuable insights into the reliability and performance of the computer systems under study, highlighting patterns and trends in system failures over time. By analyzing the count of breakdowns, researchers can identify potential periods of instability, assess the effectiveness of maintenance protocols, and evaluate the overall durability of the hardware. The extended timeframe of 128 weeks ensures a robust sample size, allowing for a comprehensive examination of both short-term fluctuations and long-term trends in system performance. Such datasets are particularly useful for predictive modeling, enabling the development of strategies to minimize downtime and optimize system efficiency. The dataset not only underscores the importance of continuous monitoring in maintaining system integrity but also contributes to the broader understanding of failure dynamics in complex computing environments. Non-parametric plots are well-suited for the preliminary visualization of this dataset, and the associated outcomes are illustrated in Figure 7.

Table 7 displays the observed frequency, expected frequency, the MLEs for the parameter(s), -L, AIC, AICC, HQIC, BIC, and Chi-square test values, along with their respective P-values, for all competing distributions applied to dataset II. Figure 8 offers a comprehensive explanation and confirmation of the findings outlined in Table 7. To emphasize the distinct features of each estimate, profile log-likelihood curves and contour plots were generated, as depicted in Figure 9.

Among all the tested models, the D2PL model demonstrates superior performance as the most optimal distribution for fitting the dataset II. The estimated probability mass functions and the log-likelihood profiles of the estimators for dataset II are presented in Figures 8 and 9, respectively.

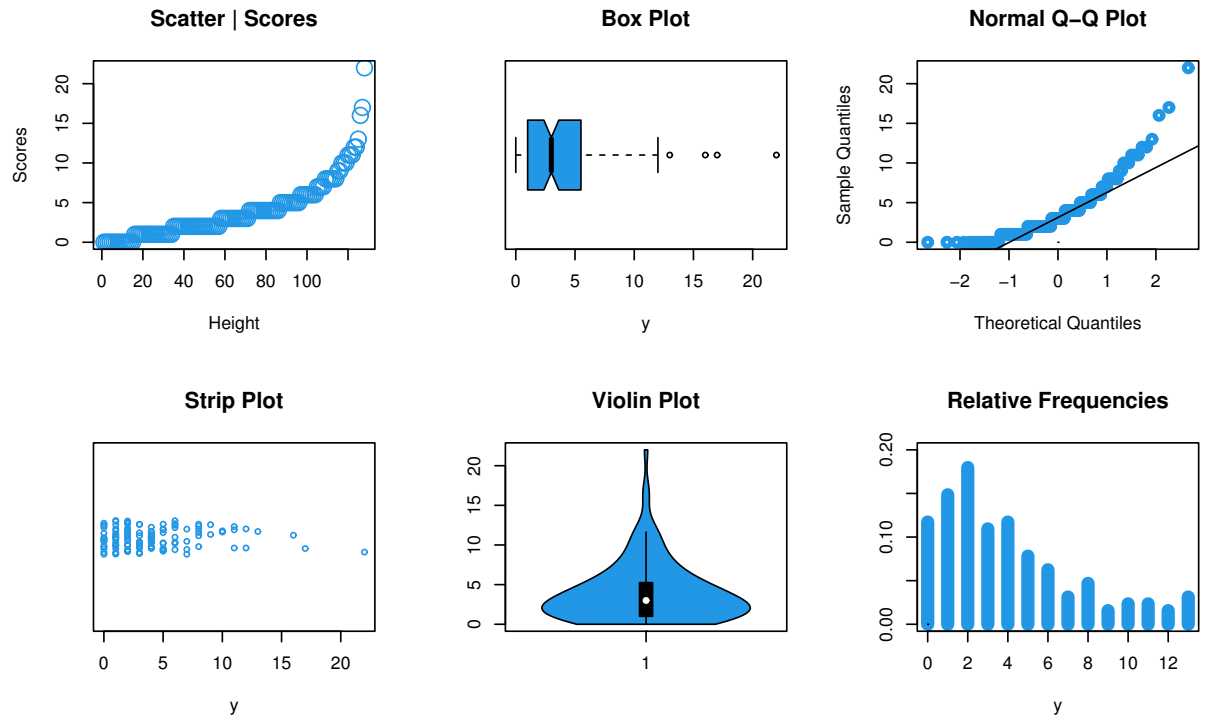

Figure 7. Plots based on non-parametric methods for dataset II.

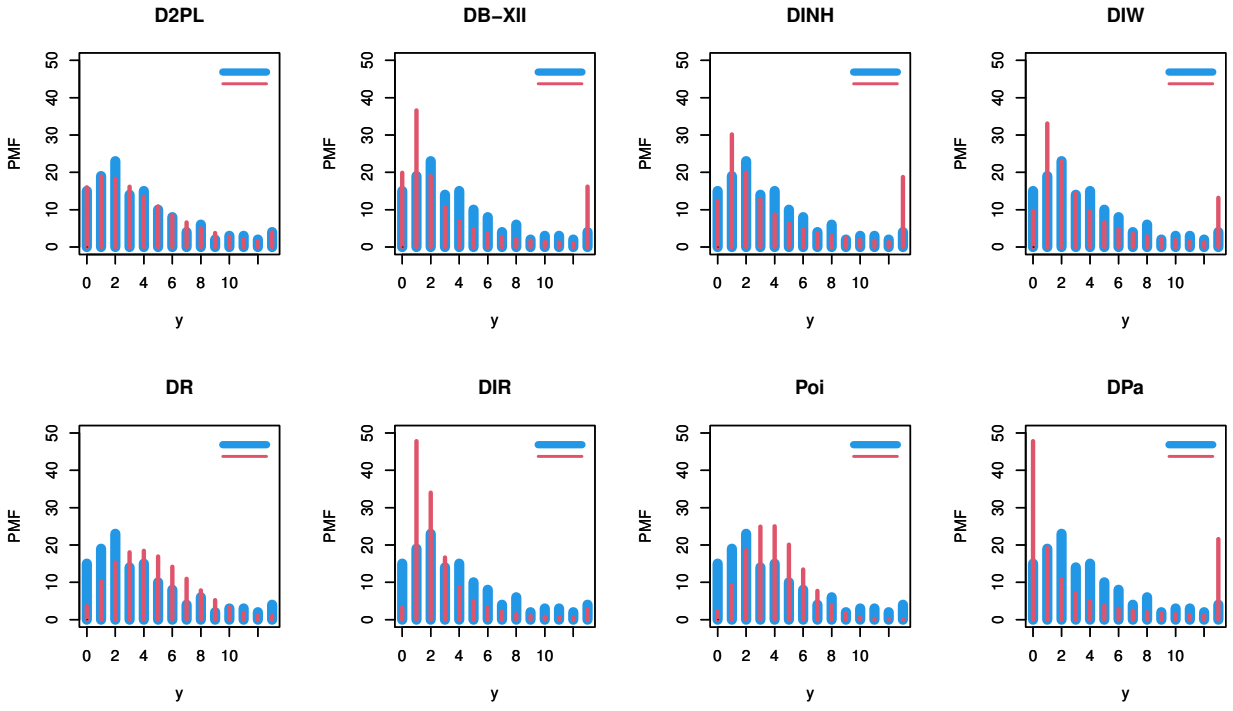

Figure 8. The estimated probability mass functions for dataset II.

Table 7. The GOFT for dataset II.

X	OF	D2PL	DB-XII	DINH	DIW	DR	DIR	Poi	DPa
0	15	16.01	19.89	12.40	9.69	3.64	3.24	2.31	47.81
1	19	18.93	36.59	30.16	33.09	10.30	47.81	9.27	19.19
2	23	18.32	19.05	19.94	23.10	15.31	34.02	18.61	10.76
3	14	16.15	10.69	12.78	14.48	18.04	16.65	24.91	7.02
4	15	13.49	6.83	8.72	9.54	18.44	8.77	25.01	5.00
5	10	10.87	4.76	6.29	6.64	16.92	5.08	20.09	3.77
6	8	8.54	3.52	4.73	4.84	14.17	3.17	13.44	2.97
7	4	6.59	2.72	3.69	3.65	10.94	2.11	7.71	2.40
8	6	5.01	2.17	2.95	2.85	7.84	1.47	3.87	1.99
9	2	3.76	1.77	2.41	2.27	5.23	1.06	1.74	1.69
10	3	2.80	1.48	2.01	1.85	3.26	0.79	0.69	1.45
11	3	2.07	1.25	1.69	1.53	1.89	0.60	0.25	1.26
12	2	1.52	1.08	1.46	1.28	1.04	0.47	0.08	1.11
+13	4	3.94	16.19	18.76	13.18	0.99	2.76	0.01	21.59
Total	128	128	128	128	128	128	128	128	128
$-l$		316.625	342.581	331.931	330.446	347.148	356.525	384.974	369.766
MLE_α		1.1348	0.785	1.244	0.076	0.972	0.025	4.016	0.509
MLE_λ		0.679	3.309	1.634	1.235	–	–	–	–
AIC		637.250	689.162	667.862	664.892	696.296	715.050	771.948	741.532
CAIC		637.346	689.258	667.958	664.988	696.327	715.082	771.979	741.564
HQIC		639.568	691.479	670.17	667.209	697.454	716.209	773.107	742.691
BIC		642.954	694.866	673.566	670.596	699.148	717.902	774.800	744.384
χ^2		3.690	40.183	20.199	18.709	49.897	50.533	88.995	94.971
d.f		8	5	6	6	8	5	6	6
P.value		0.884	< 0.001	0.003	0.005	< 0.001	< 0.001	< 0.001	< 0.001

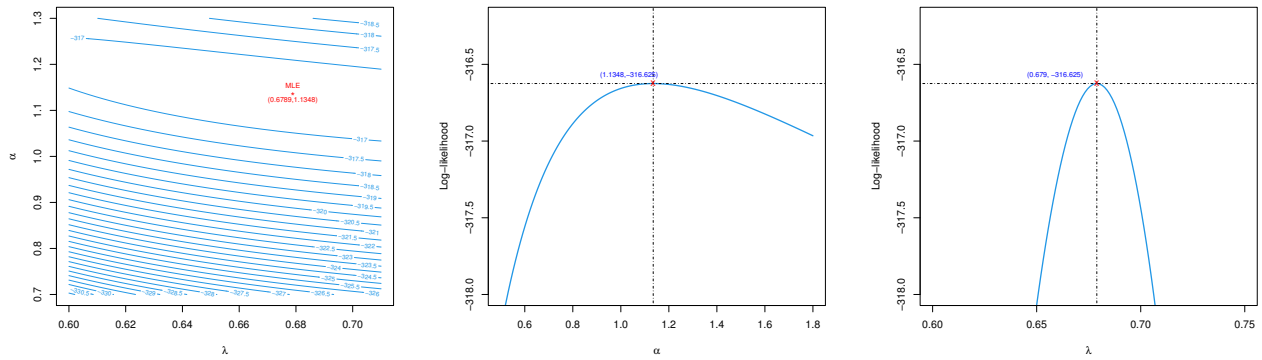

Figure 9. The contour plot and profile log-likelihood of the D2Pl for data set II.

8. Conclusions and future research

This study introduced a novel discrete distribution defined on the set of positive integers by discretizing the continuous two-parameter Lindley distribution. The newly formulated distribution, the discretized two-parameter Lindley (D2PL) distribution, was systematically developed using the survival function approach. The cumulative distribution function, survival function, and hazard

rate function for the D2PL distribution were derived in explicit closed forms. Additionally, key statistical properties of the distribution were explored. Closed-form expressions for non-central moments were provided, allowing for the straightforward computation of central and factorial moments. The distribution demonstrated unimodal and positively skewed shapes, and numerical evaluations indicated its ability to capture both platykurtic and leptokurtic behaviors. Depending on the parameter values, the hazard rate function of the D2PL distribution exhibited increasing, decreasing, or upside-down bathtub shapes, making it highly adaptable for real-world applications. This flexibility enabled the model to effectively describe both overdispersed and underdispersed datasets. Moreover, the proposed distribution encompasses the discrete Lindley with one parameter and the Geometric distributions as special cases. One practical advantage of the D2PL distribution is the simplicity of generating random numbers from it. Parameter estimation was conducted using the maximum likelihood method and the method of moments. However, the study ensures the efficiency and reliability of the maximum likelihood estimates. The proposed model was evaluated by applying it to two real-world datasets, where it demonstrated a better fit than existing discrete distributions. Goodness-of-fit measures confirmed that the data samples were effectively represented by the D2PL distribution. These results highlight the potential of the D2PL distribution as a strong alternative to current discrete probability models for analyzing count data. As future work, a multivariate extension will be explored both mathematically and through the development of a prediction model.

Funding: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2502).
Conflict of Interest: The authors declare no conflicts of interest.
Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

References

1. Aarset, M. V. (1987). How to identify a bathtub hazard rate. IEEE Transactions on Reliability, 36(1):106–108.
2. Abd EL-Hady, A. E., Hegazy, M. A., & EL-Helbawy, A. A. (2023). Discrete exponentiated generalized family of distributions. Computational Journal of Mathematical and Statistical Sciences, 2(2), 303-327.
3. Aljohani, H. M., Muhammad, A. H., Zafar, J., Almetwally, E. M., Alghamdi, A. S., Hussam, E., and Muse, A. H. (2023). Analysis of covid-19 data using discrete marshall–olkinin length biased exponential: Bayesian and frequentist approach. Scientific Reports, 13:1–18.
4. Alamatsaz, M. H., Dey, S., Dey, T. and Harandi, S. S. (2016). Discrete Generalized Rayleigh Distribution. Pakistan Journal of Statistics, 32(1):1-20.
5. Almetwally, E. M., Dey, S., & Nadarajah, S. (2023). An overview of discrete distributions in modelling COVID-19 data sets. Sankhya A, 85(2), 1403-1430.
6. Balubaid, A., Klakattawi, H., and Alsulami, D. (2024). On the Discretization of the Weibull-G Family of Distributions: Properties, Parameter Estimates, and Applications of a New Discrete Distribution. Symmetry, 16(11):1–16.

7. Barbiero, A. and Hitaj, A. (2024). Discrete half-logistic distributions with applications in reliability and risk analysis. Annals of Operations Research, 340(1):27–57.

8. Chakraborty, S. (2015). Generating Discrete Analogues of Continuous Probability Distributions- A Survey of Methods and Constructions. Journal of Statistical Distributions and Applications, 2(6):1-15.

9. Chakraborty, S. and Chakravarty, D. (2012). Discrete Gamma Distributions: Properties and Parameter Estimations. Communications in Statistics - Theory and Methods, 41(18):3301–3324.

10. Chakraborty, S., Chakravarty, D., Josmar, M., and B.Wesley (2021). A discrete analog of Gumbel distribution: properties, parameter estimation and application. Journal of Applied Statistics, 48(4):712–737.

11. Chesneau, C., Tomy, L., Gillariose, J., and Jamal, F. (2020). The Inverted Modified Lindley Distribution. Journal of Statistical Theory and Practice, 14(46):1-17.

12. Chesneau, C., Tomy, L., and Gillariose, J. (2021). A new modified Lindley distribution with properties and applications. Journal of Statistics and Management Systems, 24(7):1383–1403.

13. Das, D. and Das, B. (2023). Discretized Fréchet–Weibull distribution: Properties and application. Journal of the Indian Society for Probability and Statistics, 24(1):243–282.

14. Das, D., Das, B., and Hazarika, P. J. (2023). Discretized version of generalized Gompertz distribution with application to real-life data. International Journal of Agricultural and Statistical Sciences, 19(1):441–456.

15. El-Morshedy, M., Eliwa, M. S., and Nagy, H. (2019). A new two-parameter exponentiated discrete Lindley distribution: properties, estimation and applications. Journal of Applied Statistics, 47(2):354–375.

16. Emilio, G. D. and Enrique, C. O. (2011). The discrete Lindley distribution: properties and applications. Journal of Statistical Computation and Simulation, 84(11):1405–1416.

17. Ghitany, M. E., Atieh, B., and Nadarajah, S. (2008). Lindley distribution and its applications. Mathematics and Computers in Simulation, 78(4):493–506.

18. Gillariose, J., Tomy, L., Jamal, F., and Chesneau, C. (2020). The Marshall-Olkin Modified Lindley Distribution: Properties and Applications. Journal of Reliability and Statistical Studies, 13(1):177–198.

19. Gillariose, J., Balogun, O.S., Almetwally, E.M., Khan Sherwani, R.A., Jamal, F., and Joseph, J. (2021). On the Discrete Weibull Marshall–Olkin Family of Distributions: Properties, Characterizations, and Applications. Axioms, 10(4):1-16.

20. Hadi, N. S. and Khudhair, P. J. (2024). On discrete Fréchet distribution: Estimation and application. The Journal Of Administration & Economics, 49(142):189–199.

21. Hand, D., Daly, F., Lunn, A., McConway, K., and Ostrowski, E. (1994). Handbook of Small Data Sets. Chapman & Hall, London.

22. Ibrahim, G. and Almetwally, E. (2021). Discrete Marshall–Olkin Lomax Distribution Application of COVID-19. Biomedical Journal of Scientific and Technical Research, 32(5):25381-25390.

23. Johnson, N., Kemp, A., and Kotz, S. (2005). Univariate discrete distributions. Wiley, New York.

24. Kemp, A. W. (2004). Classes of discrete lifetime distributions, Communications in Statistics-Theory and Methods, 33 (12):3069-3093.

25. Khan, M. S. A., Khalique, A., and Aboummoh, A. M. (1989). On estimating parameters in a discrete Weibull distribution. IEEE Transactions on Reliability, 38(3):348–350.

26. Krishna, H. and Pundir, P. (2009). Discrete Burr and discrete Pareto distributions. Statistical Methodology, 6(2):177–188.

27. Lindley, D. V. (1958). Fiducial distributions and Bayes's theorem. Journal of the Royal Statistical Society, Series B, 20(1):102–107.

28. Lindley, D. V. (1965). Introduction to Probability and Statistics from Bayesian Viewpoint. Cambridge University Press, New York.

29. Makcutek, J. (2008). A Generalization of the Geometric distribution and its application in quantitative linguistics. Romanian Reports in Physics, 60(30):501–509.

30. Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. IEEE Transactions on Reliability, 24(5):300–301.

31. Oliveira, R. P. S., Mazucheli, J., Santos, M. L. A., and Barco, K. V. P. (2018). A discrete analogue of the continuous power Lindley distribution and its applications. Revista Brasileira De Biometria, 36(3):649–667.

32. Para, B. A. and Jan, T. R. (2019). On three parameters discrete generalized inverse Weibull distribution: properties and applications. Annals of Data Science, 6(3):549–570.

33. Phyo, I. (1973). Use of a chain binomial in the epidemiology of caries. Journal of Dental Research, 52(4):750–752.

34. Roy, D. (2003). The discrete normal distribution. Communications in Statistics - Theory and Methods, 32(10):1871–1883.

35. Roy, D. (2004). Discrete Rayleigh distribution. IEEE Transactions on Reliability, 53(2):255–260.

36. Shanker, R., Sharma, S., and Shanker, R. (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. Applied Mathematics, 4(2):363–368.

37. Steutel, F. and Van Harn, K. (2004). Infinite divisibility of probability distributions on the real line. Pure and Applied Mathematics. Marcel Dekker.