Mathematics

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

AIMS Mathematics, 10(5): 11382–11410.
DOI: 10.3934/math.2025518
Received: 08 March 2025
Revised: 21 April 2025
Accepted: 13 May 2025
Published: 19 May 2025

Research article

Zero-inflated discrete Lindley distribution: Statistical and reliability properties, estimation techniques, and goodness-of-fit analysis

Wael W. Mohammed[1], **Kalpasree Sharma**[2], **Partha Jyoti Hazarika**[2], **G. G. Hamedani**[3], **Mohamed S. Eliwa**[4,*] **and Mahmoud El-Morshedy**[5,6]

[1] Department of Mathematics, College of Science, University of Ha'il, Ha'il 2440, Saudi Arabia
[2] Department of Statistics, Dibrugarh University, Dibrugarh 786004, India
[3] Department of Mathematical and Statistical Sciences, Marquette University, Wisconsin, USA
[4] Department of Statistics and Operations Research, College of Science, Qassim University, Saudi Arabia
[5] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[6] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

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

Abstract: This study introduced a two-parameter zero-inflated discrete random variable distribution designed to model failure profiles in zero-inflated, dispersed datasets, commonly found in biological engineering and reliability analysis. The proposed distribution combined traditional count models, such as Poisson, Lindley, or negative binomial, with a probability mass at zero, providing a robust framework for addressing excess zeros and the underlying dispersion of data. The mathematical foundation of the distribution was derived with an emphasis on its statistical and reliability properties. The probability mass function was applicable to datasets with asymmetric dispersion and varying kurtosis structures. In addition, the hazard rate function was used to analyze failure rate behaviors, capturing patterns such as increasing, decreasing, and bathtub-shaped failure rates, often encountered in real-world datasets. Also, characterization of the proposed distribution was explored based on conditional expectation and the hazard rate function. Parameter estimation techniques were proposed, alongside computational simulations, to identify the most consistent estimators for data modeling. The goodness of fit of the proposed model was rigorously evaluated by comparing it with existing count models, demonstrating its superior ability to model zero-inflated, overdispersed data. Finally, the practical application of the new distribution was demonstrated using real-life biological engineering datasets, highlighting its effectiveness and flexibility in modeling complex zero-inflated data across various failure profiles and reliability contexts.

Keywords: statistical model; dispersion effects; conditional expectation; failure analysis; estimation methods; simulation; data analysis
Mathematics Subject Classification: 62E15, 62E99

1. Introduction

Poisson distribution serves as the ideal model for the analysis of count data, with its mean and variance being equal. However, researchers very often come across data depicting a surplus number of zeros, ones, twos, etc. This inflated data may sometimes lead to overdispersion, and a failure to take note of these excessive counts may result in biased parameter estimates and inappropriate statistical inference. This phenomenon has led to the development of models that could provide a good fit for inflated datasets. In practice, one often comes across zero -inflated datasets. Zero inflation means that the number of zero counts in the given data set exceeds the threshold limit of being accepted by conventional discrete parametric family. Count data with excess zeros are frequently encountered in various fields of study, such as ecology, medicine, public health, and insurance, leading to the widespread application of zero-inflated models.

Neyman [1] and Feller [2] were among the first to introduce the theory of zero inflation to address the issue of excess zeros. Cohen [3] and Yoneda [4] extended these initial studies to zero-inflated Poisson (ZIP) models. Later, Lambert [5] studied the ZIP regression model using the expectation maximization approach with an example of manufacturing defects. Gupta et al. [6] introduced zero-inflated modified power series distributions (IMPSD) and explored their structural properties and maximum likelihood estimates. Lin and Tsai [7] proposed a model that accommodates both excessive zeros and ones, naming it the zero-one inflated Poisson (ZOIP) model. Arora and Chaganty [8] investigated the distributional properties of zero- and k-inflated Poisson regression models, estimating the parameters using the expectation-minimization (EM) algorithm. Sun et al. [9] introduced the zero-one-two-inflated Poisson (ZOTIP) distribution, which encompasses the ZIP and ZOIP distributions as special cases, and developed key distributional properties. Melkersson and Rooth [10] proposed a zero-two-inflated Poisson distribution, and Begum et al. [11] extended the model to a zero-two-three inflated Poisson (ZTTIP) distribution, applying it to model complete female fertility data. Saboori and Doostparast [12] expanded the horizon of inflated distributions by proposing a zero to k inflated Poisson regression model capable of accommodating any degree of inflation points and thus making the entire family of Poisson-based inflated distributions accessible in a single model. Famoye and Singh [13] proposed a zero-inflated generalized Poisson (ZIGP) regression model to model domestic violence data with too many zeros. Hall [14] developed a zero-inflated binomial (ZIB) model, also incorporating random effects to add flexibility to the model. Rahman et al. [15] introduced a one-inflated binomial distribution (OIBD) and discussed its application. Two- inflated binomial distribution was used by Singh et al. [16] for modeling the pattern of sex composition of children in the state of Uttar Pradesh (India). A characterization of the ZIB model is presented by Nanjundan and Pasha [17]. Heilbron [18] proposed the zero-inflated negative binomial (ZINB) regression models and the statistical inferences were studied by Garay et al. [19]. Suresh et al. [20] characterized the ZINB distribution through a linear differential equation satisfied by its probability generating function. Furthermore, Alshkaki [21] proposed a zero-one inflated negative binomial (ZOINB) distribution. Also, very recently, Serra and Polestico [22] discussed a zero- and k-inflated negative binomial (ZkINB) distribution which is a mixture of a multinomial logistic and negative binomial distribution. Johnson et al. [23] and Iwunor [24] proposed the zero-inflated geometric (ZIG) distribution and derived the parameter estimators. Barriga and Louzada [25] proposed the zero-inflated Conway–Maxwell-Poisson (ZICOM) distribution. Lemonte et al. [26] introduced a new inflated model known as the zero-inflated Bell regression model. Rivas and Campos [27] proposed a model termed as zero inflated Waring distribution. Ospina and Ferrari [28] developed inflated beta distributions. Ferreira and Mazucheli [29] proposed the zero, one and zero-and-one-inflated new unit-Lindley distributions as natural extensions of the new unit-Lindley distribution to model continuous responses measured at the following intervals [0, 1), (0, 1] and [0, 1].

In statistical literature, we find a number of traditional count distributions. However, in the recent past, an emergence of discrete analogue of continuous distributions has occurred. One such distribution is the discrete Lindley distribution (Bakouch et al. [30]) which is the discrete analogue of continuous Lindley distribution (Lindley [31]). The discrete Lindley distribution is based on a single-parameter, making it a viable alternative to Poisson. The key advantage of this distribution is its over-dispersed nature, which makes it more flexible as compared to traditional count models for modeling actuarial data, and also in reliabilty and failure time analysis. Unlike many other discrete distributions, the discrete Lindley distribution is capable of accurately modeling both times and counts, even though it is defined by just one parameter. Thus, taking a cue from the positive aspects of this ditribution, here, we propose to develop a zero inflated model for the count dataset built on the discrete Lindley distribution reported by Bakouch et al. [30].

1.1. Limitations of existing models compared to the zero-inflated discrete Lindley distribution

Traditional zero-inflated count models have been instrumental in handling excess zeros across various domains. However, they often fall short when dealing with complex data marked by overdispersion, asymmetry, structural zeros, and non-standard hazard behaviors. Their rigid distributional assumptions limit flexibility in capturing skewness, kurtosis, and tail behavior, while their typically monotonic hazard functions hinder modeling intricate risk patterns. These models are also sensitive to outliers and structural zeros, leading to biased estimates and poor fits in the presence of rare but influential observations.

The zero-inflated discrete Lindley (ZIDL) distribution addresses these challenges by offering enhanced flexibility in modeling various dispersion levels, skewness, and tail behaviors. It supports a wide range of shapes of the risk rate, including increasing, decreasing and bathtub forms, which makes it especially useful in survival and reliability studies. With closed-form expressions for key measures and robustness to data irregularities, ZIDL ensures stable and accurate parameter estimation. Moreover, ZIDL's adaptable structure supports integration with covariate modeling, hybrid censoring, and both frequentist and Bayesian frameworks, broadening its utility across scientific disciplines. In sum, ZIDL provides a unified, robust framework for modeling complex zero-inflated count data beyond the reach of traditional models.

1.2. Key motivations

The proposed novel statistical distribution offers a flexible and robust framework for modeling complex data. It effectively handles asymmetry, varying kurtosis, and both overdispersion and equidispersion common challenges in real-world datasets. Its ability to model diverse hazard rate shapes, such as increasing, decreasing, and bathtub-shaped patterns, makes it ideal for applications in reliability and survival analysis. The model is also resilient to outliers and provides closed-form expressions for key statistical properties, supporting both theoretical and computational efficiency. Its adaptability allows for seamless integration into analytical frameworks, enhancing decision making across fields such as engineering, healthcare, and finance.

1.3. Organization of the paper

Section 2 introduces the ZIDL and develops its mathematical formulation. This section defines the distribution and derives its probability mass function (pmf), highlighting its flexibility in modeling datasets with zero-inflation and various dispersion characteristics. It serves as the foundation for understanding the key features of the proposed distribution. Section 3 explores the distributional properties of the ZIDL, including the derivation of generating functions, moments (such as mean, variance, and higher-order moments), and skewness. Additionally, reliability properties associated with this distribution are discussed, providing a comprehensive overview of its behavior under different parameter settings. These properties are essential for understanding the distribution's performance in real-world applications, particularly in fields such as survival analysis and reliability engineering. Section 4 focuses on the characterization of the ZIDL. Detailed mathematical characterizations are provided, including theorems that define the conditions under which the ZIDL can be effectively applied. These characterizations play a crucial role in demonstrating the robustness and applicability of the distribution, particularly in complex data scenarios involving asymmetric distributions and varying failure rates. Section 5 investigates various estimation methods for the parameters of the ZIDL. Techniques such as maximum likelihood estimation (MLE), method of moments estimation (MoE), and proportional estimation (ProE) are presented. Each method is discussed in detail, with strengths and weaknesses highlighted, offering a comprehensive guide for selecting the most suitable estimation technique based on the dataset at hand. Section 6 is dedicated to a simulation study, testing the consistency and efficiency of the proposed estimation methods. By generating simulated data with known parameters, the performance of the estimators is evaluated in terms of bias, mean squared error (MSE), and consistency across different sample sizes. The results from this study are crucial for assessing the reliability of the proposed methods in practical applications. Section 7 applies the ZIDL to two real-life datasets to demonstrate its practical applicability. These datasets are carefully selected to illustrate the strengths of the distribution in modeling real-world phenomena, such as overdispersion and zero-inflation, commonly encountered in fields like economics, healthcare, and engineering. Finally, Section 8 concludes the paper by summarizing the key findings and discussing potential future directions for research. The conclusion reflects on the strengths of the proposed distribution, its applications, and the implications of the results for various fields of study. The significance of the ZIDL as a versatile tool for statistical analysis and decision-making is emphasized.

2. Zero-inflated discrete Lindley distribution: Mathematical Genesis

A random variable X is said to follow the discrete Lindley distribution [30] with parameter θ if it takes only nonnegative values, and its PMF is given by:

$$Pr[X = x] = \frac{p^x}{1+\theta}\{\theta(1-2p) + (1-p)(1+\theta x)\}, \tag{2.1}$$

where $p = exp(-\theta)$, for $\theta > 0$ and $x = 0, 1, 2, ...$ The cumulative distribution function and the survival function are given respectively as

$$F(x;\theta, p) = 1 - \frac{1+\theta+\theta x}{1+\theta} p^x; \quad x = 0, 1, 2, 3, ... \tag{2.2}$$

$$S(x;\theta, p) = \frac{1+\theta+\theta x}{1+\theta} p^x; \quad x = 0, 1, 2, 3... \tag{2.3}$$

Let U be a Bernoulli random variable with probability of success $(1-\pi)$, X be a random variable from the discrete Lindley distribution, and let Y be defined as follows

$$Y = \begin{cases} 0, & U = 0, \\ X, & U = 1. \end{cases}$$

Then, the random variable Y is said to follow ZIDL distribution and its PMF is given as

$$Pr(Y = y;\theta, \pi) = \begin{cases} \pi + \frac{(1-\pi)}{1+\theta}\{\theta(1-2e^{-\theta}) + (1-e^{-\theta})\}, & y = 0, \\ (1-\pi)\frac{e^{-\theta y}}{1+\theta}\{\theta(1-2e^{-\theta}) + (1-e^{-\theta})(1+\theta y)\}, & y \geq 1, \end{cases} \tag{2.4}$$

where π is the inflation parameter $(0 < \pi < 1)$, 0 is the inflation point and $\theta > 0$. For the rest of the article, the variable Y will be denoted by the notation $Y \sim ZIDL(\theta, \pi)$. The corresponding cumulative distribution function (CDF) is given by

$$F(y;\theta, \pi) = 1 + \frac{e^{-\theta(y+1)}(\pi-1)\{1+(2+y)\theta\}}{1+\theta}; \quad y = 0, 1, 2, 3, ... \tag{2.5}$$

Particular cases:

(1) When $\pi \to 0$, $ZIDL(\theta, \pi)$ reduces to discrete Lindley distribution with parameter θ.
(2) When $\pi \to 1$, $ZIDL(\theta, \pi)$ approaches a constant value π.

Figure 1 represents the PMFs under a specific parameter value. It should be noted that the PMF of the ZIDL model can be used to analyze heavy-tailed right skewed data.

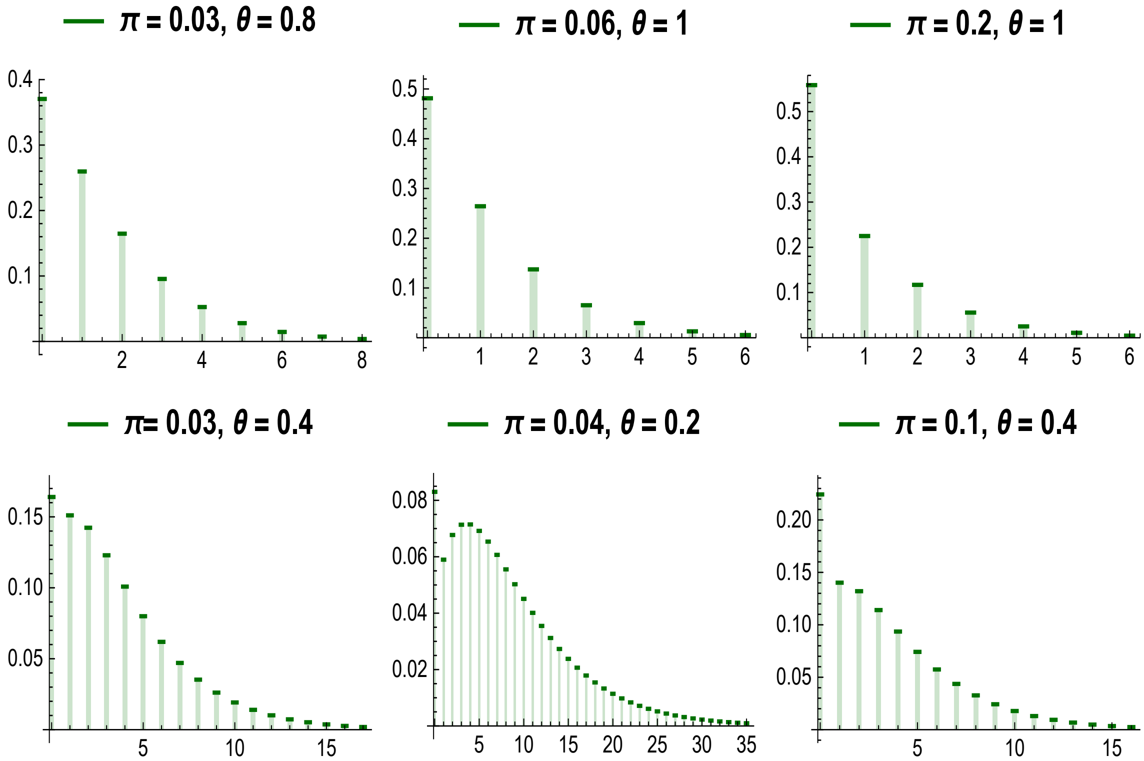

Figure 1. Plot of PMF of $ZIDL(\theta, \pi)$ for different values of θ and π.

Remark 2.1. *The survival function (SF) of $Y \sim ZIDL(\theta, \pi)$ can be formulated as*

$$S(y; \theta, \pi) = P(Y \geq y) = \frac{e^{-\theta(y+1)}\{1 + (2 + y)\theta\}(1 - \pi)}{1 + \theta}; \quad y = 0, 1, 2, 3, ... \tag{2.6}$$

Remark 2.2. *Let $y_1, y_2, y_3, ..., y_n$ be a random sample from the $ZIDL(\theta, \pi)$ distribution. Define "z" as the number of Y_i's taking the value 0. Then, Eq (2.4) can be expressed as*

$$Pr(Y = y_i) = \Phi_1^z \Phi_2^{1-z}, \tag{2.7}$$

where

$$\Phi_1 = \pi + \frac{(1 - \pi)}{1 + \theta}\{\theta(1 - 2e^{-\theta}) + (1 - e^{-\theta})\}, \quad \textit{and}$$

$$\Phi_2 = (1 - \pi)\frac{e^{-\theta y}}{1 + \theta}\{\theta(1 - 2e^{-\theta}) + (1 - e^{-\theta})(1 + \theta y)\}.$$

The hazard rate function (HRF) for $ZIDL(\theta, \pi)$ can be reported as

$$h(y; \theta, \pi) = \frac{Pr(Y = y)}{Pr(Y \geq y)} = \frac{(1 + \theta)\Phi_1^z \Phi_2^{1-z}}{(1 - \pi)e^{-\theta(y+1)}\{1 + (2 + y)\theta\}}; \quad y = 0, 1, 2, 3, ... \tag{2.8}$$

Figure 2 represents some of the possible shapes of the HRF. It can be seen that the failure rate of $ZIDL(\theta, \pi)$ displays increasing, decreasing and bathtub shapes, which makes the distribution flexible enough to use in reliability analysis.

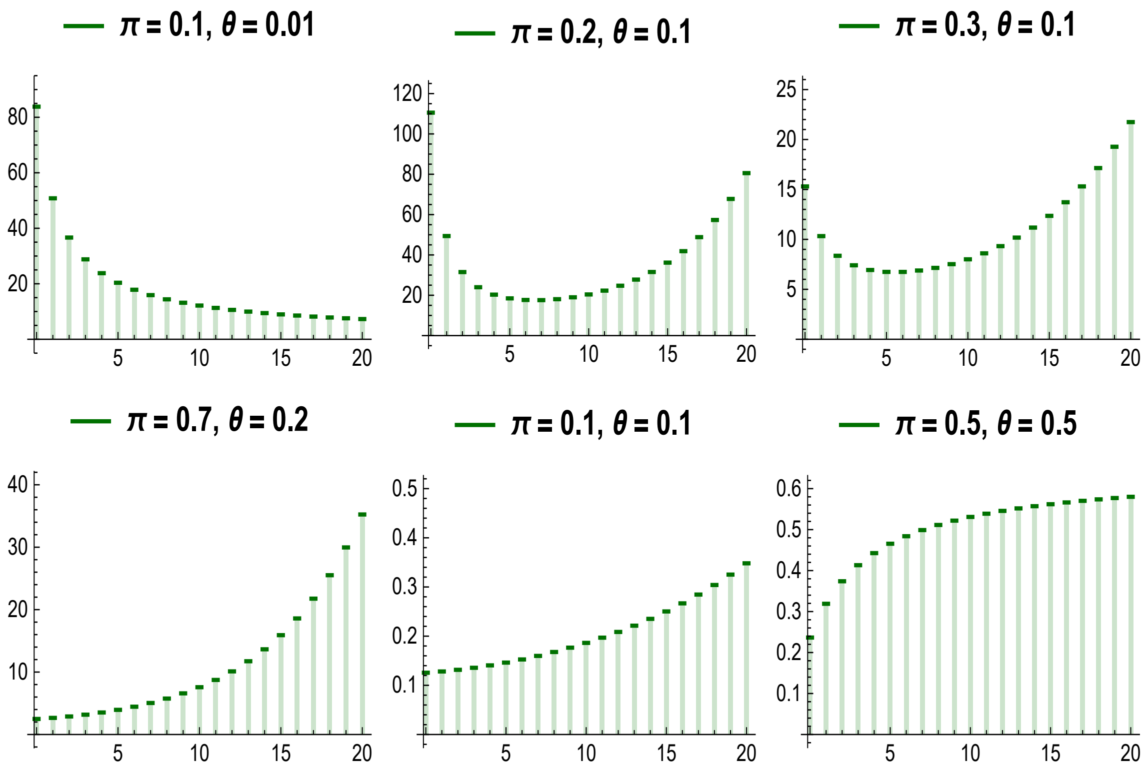

Figure 2. Plots of HRF of $ZIDL(\theta, \pi)$.

3. Statistics and reliability properties

In the study of discrete random zero-inflated models, various statistical and reliability properties such as moments, index of dispersion, skewness, kurtosis, mean residual lifetime, mean in active time, and insurance premiums are essential for a deeper understanding of the system's behavior and risk profile. These properties are particularly useful for analyzing data with excess zeros and dispersion, which are commonly encountered in fields like engineering, medicine, and insurance. The moments (mean, variance) provide crucial insights into the central tendency and variability of the data, while the index of dispersion reveals whether the data exhibits overdispersion. Skewness and kurtosis offer information on the symmetry and tail behavior of the distribution, which helps in identifying outliers or extreme events. The mean residual lifetime and mean in active time are key reliability measures that indicate the expected future behavior and performance of a system or process. Additionally, from an insurance perspective, these properties are directly tied to risk assessment, influencing the calculation of insurance premiums to reflect the potential for extreme losses or claims. Understanding these properties in zero-inflated models enables more accurate modeling of failure profiles, leading to better decision-making in diverse applications. This section serves as an introduction to discuss these important statistical and reliability properties, laying the groundwork for their application in complex, real-world datasets. Maple software has been used to derive closed-form expressions for these properties.

3.1. Probability generating function

The probability generating function (PGF) plays a fundamental role in the study of discrete distributions. While the PMF provides direct probabilities and the mean gives a measure of central tendency, the PGF offers a compact representation from which all moments of the distribution can be derived systematically. It also facilitates the study of important theoretical properties, such as dispersion, skewness, and recurrence relations. Moreover, the PGF is particularly useful in the context of branching processes, queueing theory, and reliability modeling, where it allows for simplified derivations and transformations. In addition to its theoretical value, including the PGF strengthens the mathematical characterization of the ZIDL model and supports potential extensions or applications in stochastic modeling frameworks.

Theorem 3.1. *If $Y \sim ZIDL(\theta, \pi)$, then the probability generating function is given as*

$$\begin{aligned} G_y(s) = & \frac{1}{(e^{\theta} - s)^2(1+\theta)}[e^{2\theta}(1+\theta) + (1+\pi(s-1))s(1+\theta) \\ & - e^{\theta}(1+s+2\theta+\pi(s-1)(1+2\theta))]. \end{aligned} \tag{3.1}$$

Proof.

$$\begin{aligned} G_y(s) &= E(s^y) \\ &= \sum_{y=0}^{\infty} s^y Pr(Y=y) \\ &= p(0) + \sum_{y=1}^{\infty} s^y p(y) \\ &= \frac{1}{(e^{\theta} - s)^2(1+\theta)}[e^{2\theta}(1+\theta) + (1+\pi(s-1))s(1+\theta) - e^{\theta}(1+s+2\theta+\pi(s-1)(1+2\theta))]. \end{aligned}$$

□

3.2. Moment generating function

Theorem 3.2. *If $Y \sim ZIDL(\theta, \pi)$, then the moment generating function is obtained as*

$$M_Y(t) = 1 + \frac{e^{-\theta}(\pi-1)(1+2\theta)}{1+\theta} + \frac{e^{t}(\pi-1)(1+3\theta+e^{t}(1+\theta)-e^{t-\theta}(1+2\theta)-e^{\theta}(1+2\theta))}{(e^{t}-e^{\theta})^2(1+\theta)}. \tag{3.2}$$

Proof.

$$\begin{aligned} M_Y(t) &= E(e^{ty}) \\ &= \sum_{y=0}^{\infty} e^{ty} Pr(Y=y) \\ &= 1 + \frac{e^{-\theta}(\pi-1)(1+2\theta)}{1+\theta} + \frac{e^{t}(\pi-1)(1+3\theta+e^{t}(1+\theta)-e^{t-\theta}(1+2\theta)-e^{\theta}(1+2\theta))}{(e^{t}-e^{\theta})^2(1+\theta)}. \end{aligned}$$

□

Remark 3.1. *Replacing t by it in Eq (3.2), we get the characteristic function given as*

$$\Phi_Y(t) = 1 + \frac{e^{-\theta}(\pi - 1)(1 + 2\theta)}{1 + \theta} + \frac{e^{it}(\pi - 1)(1 + 3\theta + e^{it}(1 + \theta) - e^{it-\theta}(1 + 2\theta) - e^{\theta}(1 + 2\theta))}{(e^{it} - e^{\theta})^2(1 + \theta)}. \quad (3.3)$$

3.3. Moments

The rth order raw moment of $Y \sim ZIDL(\theta, \pi)$ can be obtained using the general expression as given below

$$\begin{aligned} \mu_r^{'} &= \sum_{y=0}^{\infty} y^r Pr(Y = y) \\ &= \frac{(1 - \pi)}{1 + \theta} \sum_{y=1}^{\infty} y^r e^{-\theta y}(\theta(1 - 2e^{-\theta}) + (1 - e^{-\theta})(1 + \theta y). \end{aligned} \quad (3.4)$$

The explicit expressions of the first four moments are listed below

$$\mu_1^{'} = \frac{(1 - \pi)(e^{\theta}(1 + 2\theta) - \theta - 1)}{(1 + \theta)(e^{\theta} - 1)^2}, \quad (3.5)$$

$$\mu_2^{'} = \frac{(1 - \pi)(e^{2\theta}(1 + 2\theta) + 3e^{\theta}\theta - \theta - 1)}{(1 + \theta)(e^{\theta} - 1)^3}, \quad (3.6)$$

$$\mu_3^{'} = \frac{(1 - \pi)(e^{3\theta}(1 + 2\theta) + e^{2\theta}(3 + 13\theta) + e^{\theta}(4\theta - 3) - \theta - 1)}{(1 + \theta)(e^{\theta} - 1)^4}, \quad (3.7)$$

$$\mu_4^{'} = \frac{(1 - \pi)(e^{4\theta}(1 + 2\theta) + e^{3\theta}(10 + 35\theta) + 55e^{2\theta}\theta + e^{\theta}(5\theta - 10) - \theta - 1)}{(1 + \theta)(e^{\theta} - 1)^5}. \quad (3.8)$$

Using the above expressions, the first four central moments can be derived as

$$\mu_1 = 0, \quad (3.9)$$

$$\begin{aligned} \mu_2 = &\frac{(1 - \pi)}{(e^{\theta} - 1)^4(1 + \theta)^2}((\pi - 1)(1 + \theta - e^{\theta}(1 + 2\theta))^2 \\ &+ (e^{\theta} - 1)(1 + \theta)(e^{2\theta}(1 + 2\theta) + 3e^{\theta}\theta - \theta - 1)), \end{aligned} \quad (3.10)$$

$$\begin{aligned} \mu_3 = &\frac{(1 - \pi)}{(e^{\theta} - 1)^6(1 + \theta)^3}(2(\pi - 1)^2(e^{\theta}(1 + 2\theta) - \theta - 1)^3 \\ &+ 3(e^{\theta} - 1)(\pi - 1)(1 + \theta)(e^{\theta}(1 + 2\theta))(e^{2\theta}(1 + 2\theta) + 3e^{\theta}\theta - \theta - 1) \\ &+ (e^{\theta} - 1)^2(1 + \theta)^2(e^{3\theta}(1 + 2\theta) + e^{2\theta}(3 + 13\theta) + e^{\theta}(4\theta - 3) - \theta - 1)), \end{aligned} \quad (3.11)$$

$$\begin{aligned} \mu_4 = &\frac{(1 - \pi)}{(e^{\theta} - 1)^8(1 + \theta)^4}(3(\pi - 1)^3(1 + \theta - e^{\theta}(1 + 2\theta))^4 \\ &+ 6(e^{\theta} - 1)(\pi - 1)^2(1 + \theta)(1 + \theta - e^{\theta}(1 + 2\theta))^2(e^{2\theta}(1 + 2\theta) + 3e^{\theta}\theta\theta - 1) \\ &+ (e^{\theta} - 1)^3(1 + \theta)^3(e^{4\theta}(1 + 2\theta) + 5e^{3\theta}(2 + 7\theta) + 55e^{2\theta}\theta + 5e^{\theta}(\theta - 2) - \theta - 1) \\ &+ 4(e^{\theta} - 1)^2(\pi - 1)(1 + \theta)^2(e^{\theta}(1 + 2\theta) - \theta - 1)(e^{3\theta}(1 + 2\theta) \\ &+ e^{2\theta}(3 + 13\theta) + e^{\theta}(4\theta - 3) - \theta - 1)). \end{aligned} \quad (3.12)$$

According to the rth moment, the skewness and kurtosis coefficient can be, respectively, formulated as follows:

$$\beta_1 = \frac{a_1 + a_2 + a_3}{(1-\pi)\{(\pi-1)(1+\theta-a_4)^2 + (e^\theta - 1)(1+\theta)(-1-\theta+3e^\theta\theta+e^\theta a_4\}^3}, \tag{3.13}$$

where

$$\begin{aligned} a_1 &= 2(\pi-1)^2(-1-\theta+e^\theta(1+2\theta))^3, \\ a_2 &= 3(e^\theta-1)(\pi-1)(1+\theta)(-1-\theta+e^\theta(1+2\theta))(-1-\theta+3e^\theta\theta+e^{2\theta}(1+2\theta)), \\ a_3 &= (e^\theta-1)^2(1+\theta)^2(-1-\theta+e^{3\theta}(1+2\theta)+e^\theta(4\theta-3)+e^{2\theta}(3+13\theta))^2, \\ a_4 &= e^\theta(1+2\theta), \end{aligned}$$

and

$$\beta_2 = \frac{b_1 + b_2 + b_3 + b_4}{(1-\pi)((\pi-1)(1+\theta-b_5)^2 + (e^\theta-1)(1+\theta)(-1-\theta+3e^\theta\theta+e^\theta b_5))^2}, \tag{3.14}$$

where

$$\begin{aligned} b_1 &= 3(\pi-1)^3(1+\theta-e^\theta(1+2\theta))^4, \\ b_2 &= 6(e^\theta-1)(\pi-1)^2(1+\theta)(1+\theta-e^\theta(1+2\theta))^2(-1-\theta+3e^\theta\theta+e^{2\theta}(1+2\theta)), \\ b_3 &= (e^\theta-1)^3(1+\theta)^3(-1+5e^\theta(\theta-2)-\theta+55e^{2\theta}\theta+e^{4\theta}(1+2\theta)+5e^{3\theta}(2+7\theta)), \\ b_4 &= 4(-1+e^\theta)^2(\pi-1)(1+\theta)^2(-1-\theta+b_5)(-1-\theta+e^{2\theta}b_5+e^\theta(4\theta-3)e^{2\theta}(3+13\theta)), \\ b_5 &= e^\theta(1+2\theta). \end{aligned}$$

Figures 3 and 4 show the plots of skewness and kurtosis coefficient based on the value of different parameters. It should be noted that $ZIDL(\theta, \pi)$ can be applied to analyze positively skewed data in leptokurtic or platykurtic form.

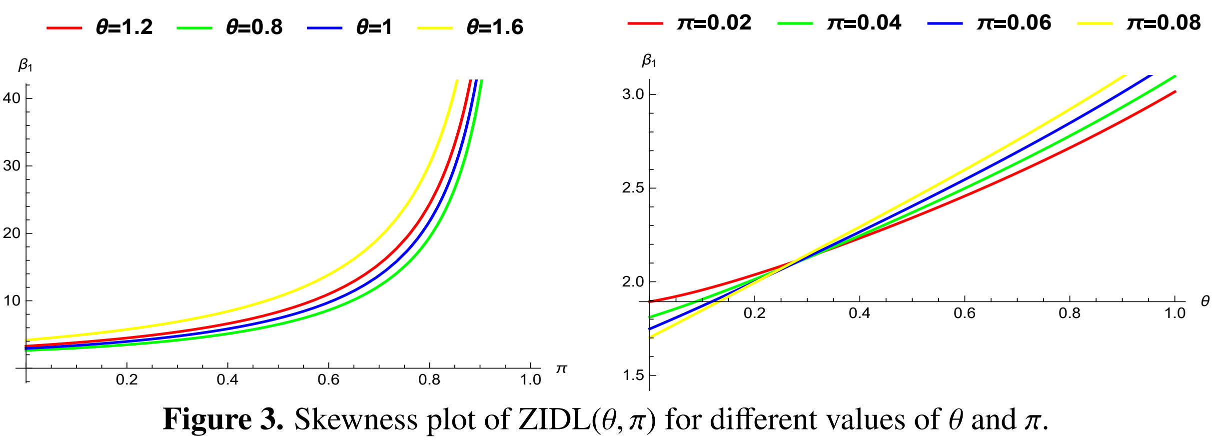

Figure 3. Skewness plot of ZIDL(θ, π) for different values of θ and π.

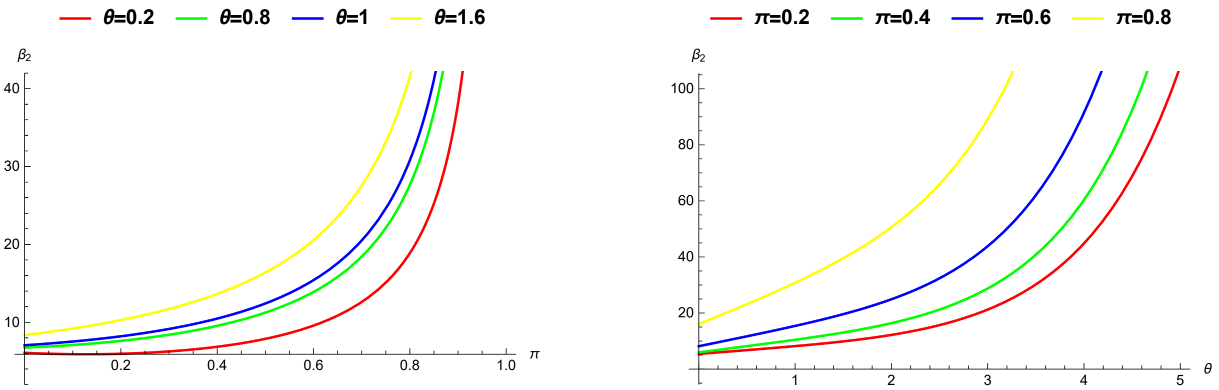

Figure 4. Kurtosis of ZIDL(θ, π) for different values of θ and π.

3.4. *Index of dispersion*

The index of dispersion gives an idea if a distribution is ideal for modeling an over-dispersed, under-dispersed or equi-dispersed dataset. Let I_y denote index of dispersion of the distribution of the random variable Y. An over-dispersed dataset has $I_y > 1$, an under-dispersed dataset has $I_y < 1$, and an equi-dispersed dataset has $I_y = 1$.

The index of dispersion is given by

$$I_y = 1 + \frac{2 + 2\theta}{1 - e^{\theta} + \theta - 2e^{\theta}\theta} + \frac{\theta(\pi - 1)}{(e^{\theta} - 1)^2} + \frac{3 + \pi + 2\theta(1 + \theta)}{(1 + \theta)(e^{\theta} - 1)}. \tag{3.15}$$

From Figure 5, we can easily see that the $ZIDL(\theta, \pi)$ distribution can accommodate both the equi-dispersed and over-dispersed dataset.

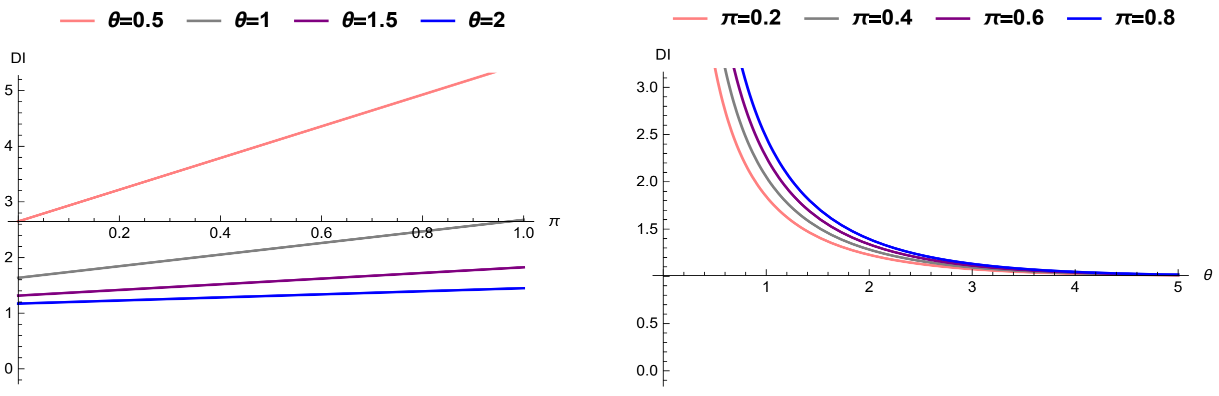

Figure 5. Index of dispersion of $ZIDL(\theta, \pi)$ for different values of θ and π.

3.5. *Mean active time and its residual coefficient of variation*

The mean active time (MAT), which estimates the anticipated amount of time that a system or component stays functional before breakdown, is an essential metric in reliability and survival study. This measure is essential to evaluate the robustness and longevity of systems and products in a variety of industries, including manufacturing, healthcare, and engineering. Remaining life of a system after it has lasted a particular period of time can be more reliably determined by calculating the residual coefficient of variation (RCOV), which is the ratio of the standard deviation to the mean of the residual life. Considering $F(.)$ as the CDF of an element with a finite first moment, where Y denotes the random variable associated with $F(.)$, in the discrete context, the MAT, say $\Upsilon(k;\theta,\pi)$, is defined as follows:

$$\Upsilon(k;\theta,\pi) = E(Y-k|Y\geq k);\quad k=1,2,3,....$$

For the $ZIDL(\theta,\pi)$ random variable, the MAT can be listed as

$$\Upsilon(k;\theta,\pi) = \frac{(\theta k+2\theta+1)e^{\theta}-\theta k-\theta-1}{(\theta k+\theta+1)(e^{\theta}-1)^2};\ k=1,2,3,.... \tag{3.16}$$

The HRF and MAT function are related by

$$h(k;\theta,\pi) = 1-\frac{[(\theta k+2\theta+1)e^{\theta}-\theta k-\theta-1]\varphi(k;\theta,\pi)}{(\theta[k+1]+\theta+1)(e^{\theta}-1)^2+(\theta[k+1]+2\theta+1)e^{\theta}-\theta[k+1]-\theta-1},$$

where

$$\varphi(k;\theta,\pi) = \frac{(\theta[k+1]+\theta+1)(e^{\theta}-1)^2}{(\theta k+\theta+1)(e^{\theta}-1)^2}.$$

The HRF, the SF and MAT are related by

$$S(k;\theta,\pi) = \prod_{0\leq i\leq k}\frac{[(\theta k+2\theta+1)e^{\theta}-\theta k-\theta-1]\varphi(k;\theta,\pi)}{(\theta[k+1]+\theta+1)(e^{\theta}-1)^2+(\theta[k+1]+2\theta+1)e^{\theta}-\theta[k+1]-\theta-1},$$

where $\Upsilon(0;\theta,\pi) = E(Y)$. The function for the variance residual life (VRL), denoted as $\Upsilon_{vrl}(k)$, is defined by

$$\begin{aligned}\Upsilon_{vrl}(k;\theta,\pi) &= E(Y^2|Y\geq k)-[E(Y|Y\geq k)]^2\\ &= -\frac{(\pi-1)e^{-\theta k}[Ae^{2\theta}+Be^{\theta}+k^2\theta+k-\theta-1]}{(1+\theta)(e^{\theta}-1)(1-2e^{\theta}+e^{2\theta})}-(2k-1)\Upsilon(k;\theta,\pi)-[\Upsilon(k;\theta,\pi)]^2,\end{aligned} \tag{3.17}$$

where

$$A = k^2\theta+2\theta k+k \text{ and } B = -2k^2\theta-2\theta k-2k+3\theta+1.$$

The random variable Y exhibits increasing (decreasing) VRL if

$$\Upsilon_{vrl}(k+1;\theta,\pi)\leq(\geq)\Upsilon(k;\theta,\pi)\left[1+\Upsilon(k+1;\theta,\pi)\right].$$

The RCOV, denoted as $\Xi(k;\theta,\pi)$, can be explicitly derived as:

$$\Xi(k;\theta,\pi) = \sqrt{\Upsilon_{VRL}(k;\theta,\pi)}/\Upsilon(k;\theta,\pi);\ k=1,2,3,....$$

The HRF, MAT, and VRL of the $ZIDL(\theta,\pi)$ model are interconnected in the following manner

$$\Upsilon_{vrl}(k+1;\theta,\pi)-\Upsilon_{vrl}(k;\theta,\pi) = h(k;\theta,\pi)[\Upsilon_{vrl}(k+1;\theta,\pi)-\Upsilon(k;\theta,\pi)(1+\Upsilon(k+1;\theta,\pi))].$$

Moreover, the HRF, MAT, VRL, and RCOV of the $ZIDL(\theta,\pi)$ distribution are interconnected as

$$\Upsilon_{vrl}(k+1;\theta,\pi)-\Upsilon_{vrl}(k;\theta,\pi) \quad = \quad h(k;\theta,\pi)[\Upsilon(k+1;\theta,\pi)]^2\times\Big\{[\Xi(k+1;\theta,\pi)]^2$$

$$-\frac{\Upsilon(k;\theta,\pi)(1+\Upsilon(k+1;\theta,\pi))}{[\Upsilon(k+1;\theta,\pi)]^2}\Bigg\}. \tag{3.18}$$

When combined, HRF, MAT, VRL, and RCOV help determine the stability and dependability of systems, optimize maintenance plans, improve product designs, and facilitate better resource allocation and risk management decision-making. By taking these steps, engineers and analysts may create strong plans that reduce downtime and increase the operational lifespan of vital systems.

3.6. Insurance premiums

Two essential techniques for figuring out insurance premiums are the expected value principle (ExVP) and the exponential premium principle (EPP). The ExVP is a simple approach that does not take loss fluctuation into consideration. It computes the premium as the total of the expected loss plus a loading factor to cover profit and administrative expenses. On the other hand, the EPP uses an exponential function of the loss distribution to account for the insurer's risk aversion. This approach uses a parameter to account for risk aversion when calculating the premium, which is based on the expected value of the exponential loss. The EPP offers a premium that more accurately represents the insurer's preferred level of risk and the possibility of catastrophic losses, albeit being more complicated. The ExVP can be expressed as

$$\mathrm{ExVP}(\varpi;.) = (1+\varpi)\sum_{z=0}^{\infty} z\,\Pr(Z=z;.),$$

where $\mathrm{ExVP}(\varpi;.)$ is the insurance premium, ϖ is the risk loading factor, and the term $(1+\varpi)$ represents the risk loading. The ExVP of the ZIDL(θ,π) distribution, say, ExVP(θ,π), is characterized by the equation

$$\begin{aligned}\mathrm{ExVP}(\varpi;\theta,\pi) &= (1+\varpi)\sum_{z=0}^{\infty} z\,\Pr(Z=z;\theta,\pi)\\ &= \frac{(1+\varpi)(1-\pi)(e^{\theta}(1+2\theta)-\theta-1)}{(1+\theta)(e^{\theta}-1)^2}.\end{aligned} \tag{3.19}$$

The EPP is calculated by solving for EPP in the equation

$$B(s-\mathrm{EPP}(\varpi;.)) = E(s-Z),$$

where s denotes the wealth of an individual and $B(z) = -e^{-\varpi z}$ represents the exponential utility function. For the ZIDL(θ,π) distribution, the EPP, say, EPP(θ,π), can be listed as

$$\begin{aligned}\mathrm{EPP}(\varpi;\theta,\pi) =& \frac{1}{\varpi(1+\theta)}(1+\theta+(1+2\theta)(\pi-1)e^{-\theta}\\ &+\frac{e^{\varpi}(\pi-1)(1+3\theta+e^{\varpi}(1+\theta)-e^{\varpi-\theta}(1+2\theta)-e^{\theta}(1+2\theta))}{(e^{\varpi}-e^{\theta})^2}).\end{aligned} \tag{3.20}$$

The ExVP and EPP are frequently used in a variety of insurance sectors, including health insurance, liability insurance, and property insurance (e.g., coverage for high-risk activities, natural catastrophes, and pre-existing medical problems). In these industries, to guarantee sufficient risk coverage, premiums are frequently determined using intricate actuarial models that include the epp.

4. Characterization under conditional expectation and HRF

Characterizations of distributions are necessary and sufficient conditions for any statistical phenomenon. Characterizations of distributions are important to many researchers in the applied fields. An investigator will be vitally interested to know if their model fits the requirements of a particular distribution. To this end, one will depend on the characterizations of this distribution which provide conditions under which the underlying distribution is indeed that particular distribution. This section is devoted to certain characterizations of $ZIDL(\theta, \pi)$ in two directions: (*i*) based on an appropriate function of the random variable; and (*ii*) in terms of the hazard function.

4.1. Characterizations based on conditional expectation

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

Theorem 4.1. *Let* $Y : \Omega \to \mathbb{N}^* \ (\mathbb{N} \cup \{0\})$ *be a random variable. The PMF of Y is Eq* (2.4) *if and only if*

$$E\left\{\left[\left(\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta Y)\right)^{-1}\right] \mid Y>k\right\} = \frac{1}{\left(1-e^{-\theta}\right)\left(1+(2+k)\,\theta\right)}. \tag{4.1}$$

Proof. If Y has PMF in Eq (2.4), then for $k \in \mathbb{N}$, the left-hand side of Eq (4.1), using the infinite geometric sum formula, will be

$$\begin{aligned}(1-F(k))^{-1}\sum_{y=k+1}^{\infty}\left(\frac{1-\pi}{1+\theta}\right)e^{-\theta y} &= \left(\frac{1+\theta}{1-\pi}\right)\left(\frac{e^{\theta(k+1)}}{1+(2+k)\,\theta}\right)\left(\frac{1-\pi}{1+\theta}\right)\left(\frac{e^{-\theta(k+1)}}{1-e^{-\theta}}\right)\\ &= \frac{1}{\left(1-e^{-\theta}\right)\left(1+(2+k)\,\theta\right)}.\end{aligned}$$

Conversely, if Eq (4.1) holds, then

$$\begin{aligned}&\sum_{y=k+1}^{\infty}\left\{\left[\left(\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta y)\right)^{-1}\right]f(y)\right\}\\ &= (1-F(k))\left(\frac{1}{\left(1-e^{-\theta}\right)\left(1+(2+k)\,\theta\right)}\right)\\ &= (1-F(k+1)+f(k+1))\left(\frac{1}{\left(1-e^{-\theta}\right)\left(1+(2+k)\,\theta\right)}\right).\end{aligned} \tag{4.2}$$

From Eq (4.2), we also have

$$\sum_{y=k+2}^{\infty}\left\{\left[\left(\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta y)\right)^{-1}\right]f(y)\right\} = (1-F(k+1))\left(\frac{1}{\left(1-e^{-\theta}\right)\left(1+(3+k)\,\theta\right)}\right). \tag{4.3}$$

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

$$\left\{\left(\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta\,(k+1))\right)^{-1}\right\}f(k+1)$$

$$= (1 - F(k+1)) \left\{ \begin{array}{c} \left(\frac{1}{(1-e^{-\theta})(1+(2+k)\theta)} \right) - \\ \left(\frac{1}{(1-e^{-\theta})(1+(3+k)\theta)} \right) \end{array} \right\} + f(k+1) \left(\frac{1}{(1-e^{-\theta})(1+(2+k)\theta)} \right).$$

From the above equality, we have

$$\frac{f(k+1)}{1-F(k+1)} = \frac{\left(\frac{1}{(1-e^{-\theta})(1+(2+k)\theta)} \right) - \left(\frac{1}{(1-e^{-\theta})(1+(3+k)\theta)} \right)}{\left(\frac{1}{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(k+1))} \right) - \left(\frac{1}{(1-e^{-\theta})(1+(2+k)\theta)} \right)},$$

and after some computations

$$\frac{f(k+1)}{1-F(k+1)} = \frac{\theta\left(1-2e^{-\theta}\right) + \left(1-e^{-\theta}\right)(1+\theta(k+1))}{e^{-\theta}(1+(3+k)\theta)},$$

which is the HRF corresponding to the PMF in Eq (2.4). □

Remark 4.1. *Theorem 3 is a characterization of pmf based on the conditional expectation of a function of the random variable Y, which is chosen based on the nature of the given pmf. In general, it would be as follows:*

$$E[\psi(Y) \mid Y > k] = \phi(k), \quad k \in N^*.$$

One of the suitably chosen $\psi(Y)$, to make $\phi(k)$ as simple as possible, in this case is

$$\psi(Y) = \left(\theta\left(1-2e^{-\theta}\right) + \left(1-e^{-\theta}\right)(1+\theta Y)\right)^{-1}.$$

4.2. Characterizations of distributions based on HRF

This subsection deals with the characterization of the $ZIDL(\theta, \pi)$ distribution in terms of the HRF.

Theorem 4.2. *Let $Y : \Omega \to \mathbb{N}^*$ be a random variable. The PMF of Y is Eq (2.4) if and only if its HRF satisfies the difference equation*

$$h_F(k+1) - h_F(k) = \left(\frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(k+1))}{e^{-\theta}(1+(3+k)\theta)} \right) - \left(\frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(k))}{e^{-\theta}(1+(2+k)\theta)} \right), \quad (4.4)$$

where $k \in \mathbb{N}$, with the initial condition $h_F(1) = \frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta)}{e^{-\theta}(1+3\theta)}$.

Proof. If Y has PMF Eq (2.4), then clearly Eq (4.4) holds. Now, if Eq (4.4) holds, then for every $y \in \mathbb{N}$, we have

$$\sum_{k=1}^{y-1} \{h_F(k+1) - h_F(k)\} = \sum_{k=1}^{y-1} \left\{ \begin{array}{c} \left(\frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(k+1))}{e^{-\theta}(1+(3+k)\theta)} \right) - \\ \left(\frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(k))}{e^{-\theta}(1+(2+k)\theta)} \right) \end{array} \right\}$$

$$= \left(\frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(y))}{e^{-\theta}(1+(2+y)\theta)} \right) - \frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)}{e^{-\theta}(1+2\theta)},$$

or

$$h_F(y) - h_F(1) = \left(\frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta(y))}{e^{-\theta}(1+(2+y)\theta)} \right) - \frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta)}{e^{-\theta}(1+3\theta)},$$

or, in view of the initial condition

$$h_F(y) = \frac{\theta\left(1-2e^{-\theta}\right)+\left(1-e^{-\theta}\right)(1+\theta y)}{e^{-\theta}(1+(2+y)\theta)}, \quad y \in \mathbb{N},$$

which is the hazard function corresponding to the PMF in Eq (2.4). □

5. Estimation methods

5.1. Maximum likelihood method

Let $Y \sim ZIDL(\theta, \pi)$ and let us take a random sample of size n, say, $y_1, y_2, y_3, ...y_n$, from this distribution.

$$A_i = \begin{cases} 1, & y_i = 0, \\ 0, & otherwise. \end{cases}$$

Then, the PMF of $ZIDL(\theta, \pi)$ can be written as

$$\begin{aligned} Pr(Y = y_i) =& [\pi + \frac{(1-\pi)}{1+\theta}\{\theta(1-2e^{-\theta}) + (1-e^{-\theta})\}]^{A_i}[(1-\pi)\frac{e^{-\theta y_i}}{1+\theta}\{\theta(1-2e^{-\theta}) \\ &+ (1-e^{-\theta})(1+\theta y_i)\}]^{1-A_i}. \end{aligned} \tag{5.1}$$

The likelihood function, say $L = L(\theta, \pi; y_1, y_2, ...y_n)$, can be formulated as

$$\begin{aligned} L =& [\pi + \frac{(1-\pi)}{1+\theta}\{\theta(1-2e^{-\theta}) + (1-e^{-\theta})\}]^{n_0}\prod_{i=1}^{n}[(1-\pi)\frac{e^{-\theta y_i}}{1+\theta}\{\theta(1-2e^{-\theta}) \\ &+ (1-e^{-\theta})(1+\theta y_i)\}]^{a_i}, \end{aligned} \tag{5.2}$$

where $n_0 = \sum_{i=1}^{n} A_i$, represents the number of zeros in the sample and $(1 - A_i) = a_i$. Therefore, the log-likelihood function of the parameters θ and π can be derived as

$$\begin{aligned} \log L =& n_0 \log[\pi + \frac{(1-\pi)}{1+\theta}\{\theta(1-2e^{-\theta}) + (1-e^{-\theta})\}] - \theta\sum_{i=1}^{n} a_i y_i \\ &+ (n-n_0)\log(1-\pi) - (n-n_0)\log(1+\theta) \\ &+ \sum_{i=1}^{n}\log\{\theta(1-2e^{-\theta}) + (1-e^{-\theta})(1+\theta y_i)\}. \end{aligned} \tag{5.3}$$

Differentiating Eq (5.3) with respect to parameters π and θ, we get the score functions as-

$$\frac{\delta}{\delta\pi}\log L = n_0[\pi - 1 + \frac{e^{\theta}(1+\theta)}{1+2\theta}]^{-1} - \frac{(n-n_0)}{(1-\pi)}, \tag{5.4}$$

$$\frac{\delta}{\delta\theta}\log L = n_0 c_1 - \sum_{i=1}^{n} a_i y_i - \frac{(n-n_0)}{(1+\theta)} + \sum_{i=1}^{n} a_i c_2, \tag{5.5}$$

where

$$c_1 = \frac{(1-\pi)(3+2\theta)\theta}{(1+\theta)\{e^{\theta}(1+\theta)+(\pi-1)(1+2\theta)\}},$$
$$c_2 = \frac{e^{\theta}+2\theta-1+(e^{\theta}+\theta-1)y_i}{e^{\theta}(1+\theta)-2\theta-1+(e^{\theta}-1)\theta y_i}.$$

To derive the asymptotic confidence intervals for both parameters, the information matrix is needed. The second-order partial derivatives of the log-likelihood are provided as follows:

$$\frac{\delta^2}{\delta\pi^2}\log L = -n_0\{\pi-1+\frac{e^{\theta}(1+\theta)}{1+2\theta}\}^{-2} - (n-n_0)(1-\pi)^2,$$
$$\frac{\delta^2}{\delta\theta^2}\log L = n_0 c_3 + (n-n_0)(1+\theta)^{-2} + \sum_{i=1}^{n} a_i c_4,$$
$$\frac{\delta^2}{\delta\theta\delta\pi}\log L = \frac{-n_0 e^{\theta}\theta(3+2\theta)}{\{e^{\theta}(1+\theta)+(\pi-1)(1+2\theta)\}^2},$$

where

$$c_3 = \frac{(\pi-1)\{(1-\pi)(3+4\theta)+e^{\theta}(1+\theta)(-3+\theta(2+\theta)(1+2\theta))\}}{(1+\theta)^2\{e^{\theta}(1+\theta)+(\pi-1)(1+2\theta)\}^2},$$
$$c_4 = \frac{(2+y_i)^2+e^{\theta}\{(\theta-1)(5+2\theta)+y_i(-6+\theta(2+3\theta)+(\theta^2-2)y_i)-e^{2\theta}(1+y_i)^2\}}{\{e^{\theta}(1+\theta)+(\pi-1)(1+2\theta)\}^2}.$$

The Fisher's information matrix for (π,θ) can be expressed as

$$I = \begin{bmatrix} -E(\frac{\delta^2}{\delta\pi^2}\log L) & -E(\frac{\delta^2}{\delta\theta\delta\pi}\log L) \\ -E(\frac{\delta^2}{\delta\theta\delta\pi}\log L) & -E(\frac{\delta^2}{\delta\theta^2}\log L) \end{bmatrix}.$$

This can be approximated by

$$\hat{I} = \begin{bmatrix} -\frac{\delta^2}{\delta\pi^2}\log L & -\frac{\delta^2}{\delta\theta\delta\pi}\log L \\ -\frac{\delta^2}{\delta\theta\delta\pi}\log L & -\frac{\delta^2}{\delta\theta^2}\log L \end{bmatrix}_{(\pi,\theta)=(\hat{\pi}_{ML},\hat{\theta}_{ML})}.$$

Under some general regularity conditions, for large n, $\sqrt{n}(\hat{\pi}_{ML}-\pi, \hat{\theta}_{ML}-\theta)$ is bivariate normal with mean vector $(0,0)$ and the dispersion matrix

$$\hat{I}^{-1} = \frac{1}{I_{11}I_{22}-I_{12}I_{21}}\begin{bmatrix} I_{22} & -I_{12} \\ -I_{21} & I_{11} \end{bmatrix} = \begin{bmatrix} J_{11} & -J_{12} \\ -J_{21} & J_{22} \end{bmatrix}.$$

Thus, the asymptotic $(1-\alpha)\times 100\%$ confidence interval for π and θ are given, respectively, as

$$(\hat{\pi}_{ML} - Z_{\alpha/2}\sqrt{J_{11}},\ \ \hat{\pi}_{ML} + Z_{\alpha/2}\sqrt{J_{11}}) \text{ and } (\hat{\theta}_{ML} - Z_{\alpha/2}\sqrt{J_{11}},\ \ \hat{\theta}_{ML} + Z_{\alpha/2}\sqrt{J_{11}}).$$

5.2. Proportion estimation method

Given a random sample of size n, denoted as $Y_1, Y_2, ..., Y_n$, drawn from the ZIDL(θ, π) distribution, the estimation of the unknown parameters θ and π can be obtained using the ProE method by solving the following system of equations:

$$1 + \frac{(\pi - 1)(1 + 2\theta)}{(1 + \theta)} e^{-\theta} - \frac{O}{n} = 0$$

and

$$1 + \frac{(\pi - 1)(1 + 3\theta)}{(1 + \theta)} e^{-2\theta} - \frac{O + V}{n} = 0,$$

where O and V are the number of zeros and ones in the sample.

5.3. Moments estimation method

Suppose we have a random sample of size n, represented as $Y_1, Y_2, ..., Y_n$, drawn from the ZIDL(θ, π) distribution. To estimate the unknown parameters θ and π by using the MoE method, we can determine their values by solving the following system of equations:

$$\frac{(1 - \pi)(e^{\theta}(1 + 2\theta) - \theta - 1)}{(1 + \theta)(e^{\theta} - 1)^2} - \overline{Y} = 0,$$

and

$$\frac{(1 - \pi)\left[(\pi - 1)(1 + \theta - e^{\theta}(1 + 2\theta))^2 + (e^{\theta} - 1)(1 + \theta)(e^{2\theta}(1 + 2\theta) + 3e^{\theta}\theta - \theta - 1)\right]}{(e^{\theta} - 1)^4(1 + \theta)^2} - S^2 = 0,$$

where $\overline{Y}$ and S^2 are the mean and variance of the sample.

6. Simulation study

In this section, the Markov chain Monte Carlo simulation technique is used to evaluate the performance of MLE, ProE, and MoE. The $ZIDL(\theta, \pi)$ distribution is analyzed in the simulation using the R function optim() with the argument approach = “L-BFGS-B”. A total of $N = 10000$ samples are created, divided into six distinct sizes ($n = 20, 50, 100, 200, 300, 500$). These samples are taken from the $ZIDL(\theta, \pi)$ distribution at various $ZIDL(\theta, \pi)$ parameter values, as shown in Tables 1 and 2. The simulation study demonstrates that the bias and MSE consistently decrease as the sample size increases, confirming the consistency of the MLE, MoE, and ProE methods. This trend is clearly illustrated through both graphical representations (Figures 6 and 7) and numerical results (Tables 1 and 2). The behavior of the estimators under different sample sizes, particularly at the fixed parameter values, highlights their robustness and efficiency. Among the three, the MLE method shows particularly strong performance, establishing its reliability for accurately estimating the parameters of the proposed distribution. These findings collectively underscore the practical value and statistical soundness of the proposed estimation techniques.

Table 1. Simulation results at $\theta = 1.5$ and $\pi = 0.9$.

		MLE		ProE		MoE	
Parameter	n	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE
θ	20	0.204661	0.177394	0.237451	0.184962	0.243645	0.197493
	50	0.162393	0.133270	0.180042	0.153746	0.196384	0.164837
	100	0.120484	0.083891	0.143745	0.124847	0.139681	0.106452
	200	0.029470	0.007386	0.094861	0.034810	0.074382	0.010348
	300	0.000246	0.000083	0.001472	0.000648	0.000809	0.000249
	500	0.000009	0.000000	0.000084	0.000008	0.000028	0.000007
π	20	0.374581	0.294763	0.422048	0.327534	0.404681	0.304681
	50	0.319042	0.210374	0.349471	0.243638	0.358261	0.253846
	100	0.204765	0.143783	0.239387	0.166396	0.220847	0.143854
	200	0.130845	0.084652	0.164937	0.113876	0.153947	0.103771
	300	0.007814	0.000543	0.063864	0.007458	0.032841	0.002374
	500	0.000047	0.000002	0.000892	0.000032	0.000192	0.000009

Table 2. Simulation results at $\theta = 2.5$ and $\pi = 0.9$.

		MLE		ProE		MoE	
Parameter	n	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE
θ	20	0.410473	0.284671	0.442738	0.343745	0.439471	0.320473
	50	0.337365	0.204763	0.374946	0.255381	0.369461	0.242836
	100	0.204731	0.139478	0.254375	0.164932	0.230768	0.149874
	200	0.120487	0.079371	0.143843	0.110474	0.153874	0.120484
	300	0.024978	0.000439	0.104735	0.008473	0.117451	0.019475
	500	0.000847	0.000027	0.009461	0.000649	0.006093	0.000248
π	20	0.294749	0.243851	0.344384	0.288497	0.312539	0.255484
	50	0.255494	0.188439	0.274379	0.213535	0.264836	0.199473
	100	0.219479	0.133854	0.243635	0.163851	0.220947	0.144376
	200	0.143846	0.054731	0.144386	0.084734	0.146383	0.086357
	300	0.084975	0.000745	0.094782	0.003647	0.103984	0.007453
	500	0.000534	0.000008	0.000884	0.000017	0.000947	0.000057

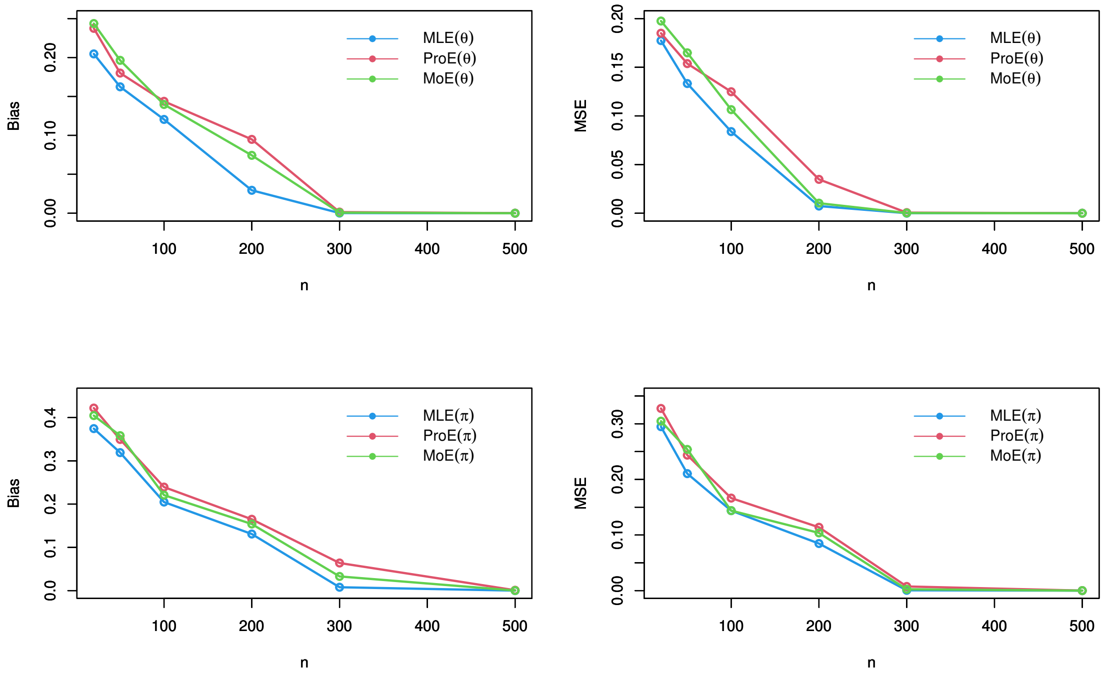

Figure 6. The simulation visualization plots corresponding to $\theta = 1.5$ and $\pi = 0.9$.

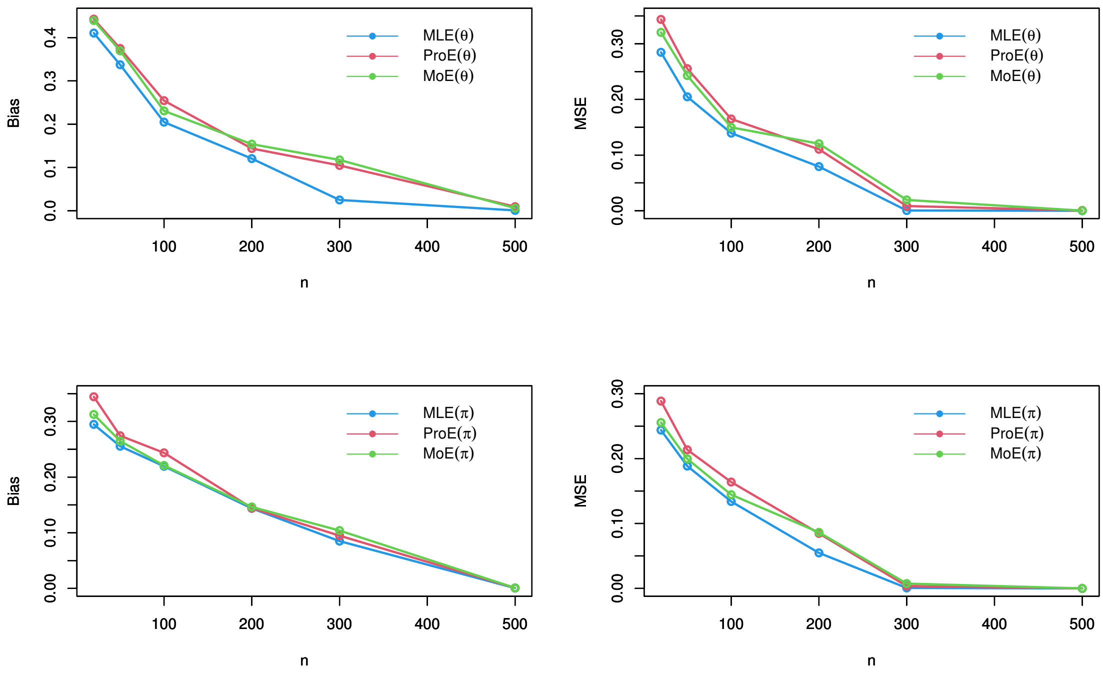

Figure 7. The simulation visualization plots corresponding to $\theta = 2.5$ and $\pi = 0.9$.

7. Biological engineering data analysis

In this section, the credibility of the $ZIDL(\theta, \pi)$ distribution is put forward by taking into consideration two real life datasets. We compare our proposed distribution with some of the well-known and some of the novel discrete distributions used to model dispersed count data such as the ZIP distribution, ZIB distribution, the Poisson-geometric distribution (POIG) distribution (Nandi et al. [32]), the discrete Lindley (DL) distribution (Bakouch et al. [30]), the discrete modified Lindley distribution (DML) (Tomy et al. [33]), the Poisson modified Lindley distribution (POIML) (Chesneau et al. [34]), and the discrete gamma Lindley distribution (DGL) distribution (El-Morshedy et al. [35]). The expected frequencies for each of these models corresponding to the observed frequencies (OF) of the given dataset are calculated. The comparison of the fitted distributions involves the evaluation of certain criteria, specifically, the negative log-likelihood ($-L$) and the Chi-square ($\varkappa^2$) test along with its associated P-value.

7.1. Dataset I

The first dataset provides information on infection produced by the parasite Trypanosoma murmanensis in cod. The response variable is the number of parasites found in the cods (Intensity). The dataset was extracted from the package countreg of the R software [36]. Plotting nonparametric information graphs allows one to discuss the behavior of the dataset (see, Figure 8). The data was observed to be skewed to the right with outlier observations. In addition, the shape of the hazard rate is decreasing.

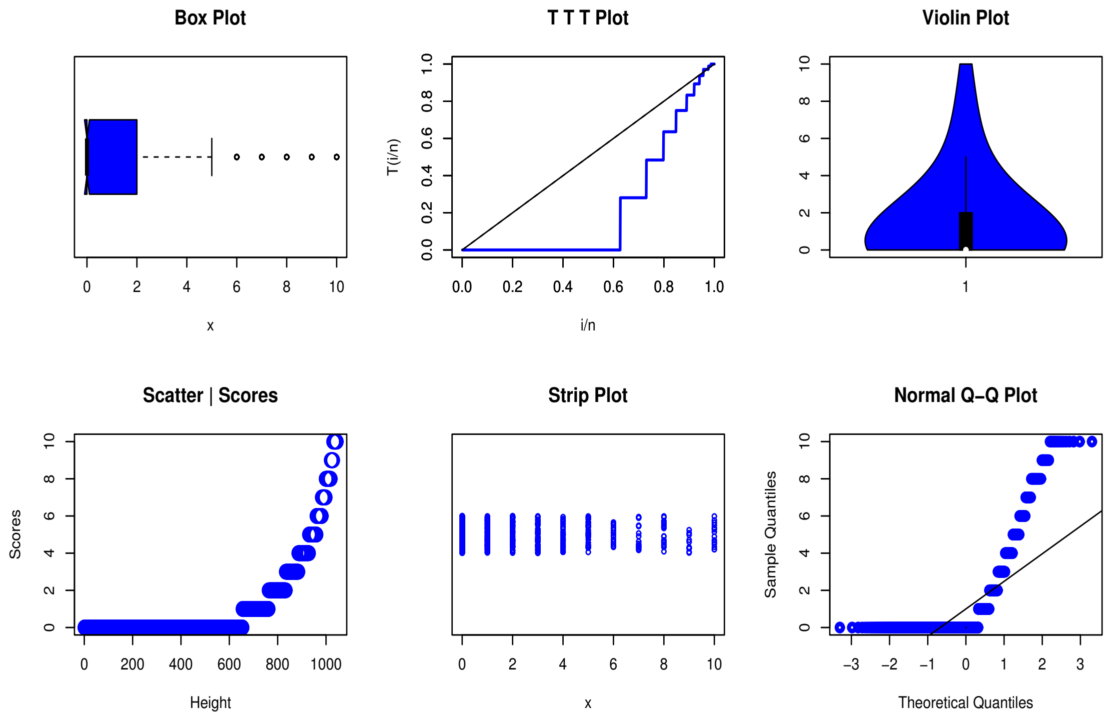

Figure 8. Nonparametric visualization plots for dataset I.

The $ZIDL(\theta, \pi)$ model's goodness-of-fit test and a few other well-known competitive models are included in Table 3. Based on significance level 0.05, it was found that the $ZIDL(\theta, \pi)$ performs the best for this set of data out of all examined models. Figure 9 supports our empirical results. To prove the unique property for each estimator, the contour plots and log-likelihood profiles are plotted in Figure 10.

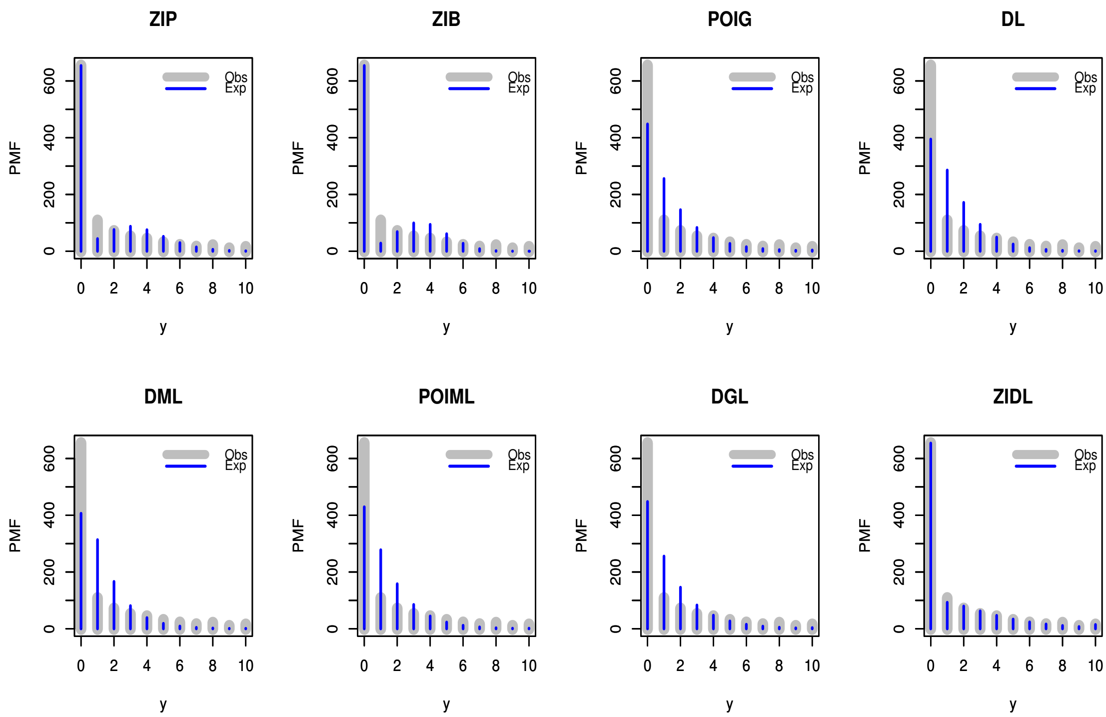

Figure 9. The estimated PMF for the dataset I.

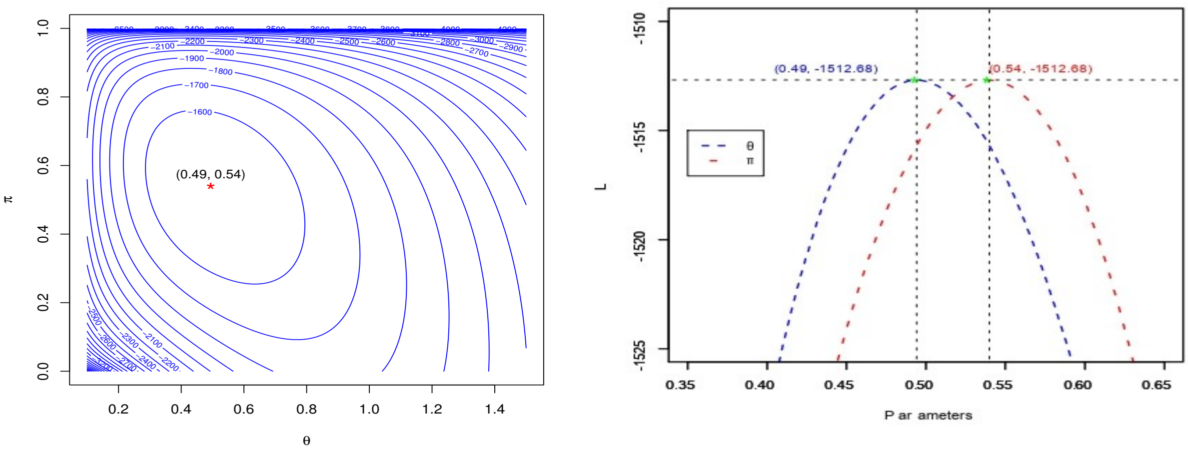

Figure 10. Contour plots (left panel) and log-likelihood profiles (right panel) for dataset I.

Table 3. The goodness-of-fit test for dataset I.

y	OF	ZIP	ZIB	POIG	DL	DML	POIML	DGL	ZIDL
0	654	654.01	653.98	448.04	394.74	406.48	429.39	447.98	653.96
1	108	44.15	28.19	255.76	285.64	313.72	278.10	255.75	92.94
2	71	76.14	68.76	145.99	171.68	166.49	158.38	146.01	79.76
3	52	87.53	99.36	83.34	94.41	81.12	85.29	83.36	62.72
4	44	75.46	94.22	47.57	49.27	38.94	44.75	47.59	46.85
5	31	52.05	61.27	27.15	24.84	18.84	23.21	27.17	33.82
6	22	29.92	27.66	15.50	12.22	9.23	11.99	15.51	23.82
7	16	14.74	8.56	8.85	5.90	4.57	6.19	8.86	16.48
8	21	6.36	1.74	5.05	2.81	2.28	3.21	5.06	11.25
9	10	2.44	0.20	2.88	1.32	1.15	1.67	2.89	7.59
10	15	1.21	0.01	3.83	1.15	1.18	1.82	3.82	14.78
Total	1044	1044	1044	1044	1044	1044	1044	1044	1044
$-L$		1512.69	1752.97	1661.84	1762.18	1779.26	1717.24	1661.88	1512.68
MLEs		$\hat{\pi}$=0.61 $\hat{\lambda}$=3.45	$\hat{\pi}$=0.62 $\hat{p}$=3.45	$\hat{\lambda}$¡0.0001 $\hat{\theta}$=0.43	$\hat{\theta}$=0.85	$\hat{p}$=0.51	$\hat{\theta}$=0.89	$\hat{\alpha}$=0.36 $\hat{\eta}$=0.57	$\hat{\pi}$=0.54 $\hat{\theta}$=0.49
χ^2		260.6	542.97	339.74	699.77	680.93	539.79	339.78	15.03
d.f		6	5	7	7	6	7	7	8
p.value		¡0.0001	¡0.0001	¡0.0001	¡0.0001	¡0.0001	¡0.0001	¡0.0001	0.06

Table 4 shows different estimates based on the first dataset. It has been observed that the MLE approach is the best among all the methods tested.

Table 4. Different estimators for dataset I.

Technique	$\hat{\pi}$	$\hat{\theta}$	χ^2	d.f	P.value
MLE	0.54	0.49	15.03	8	0.06
ProE	0.52	0.56	34.16	8	< 0.0001
MoE	0.53	0.50	16.11	8	0.04

7.2. Dataset II

The second dataset is related to mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin (NSC-45383), exposure -60μ g/kg (Shanker et al. [37]). It is possible to examine the dataset's behavior by plotting nonparametric information graphs (see Figure 11). There were outlier observations and a rightward skew in the data. Furthermore, the hazard rate is exhibiting a decreasing form.

Table 5 contains the goodness-of-fit test results for the $ZIDL(\theta, \pi)$ model along with a few other popular competing models. Out of all the models that were looked at, it was discovered that the $ZIDL(\theta, \pi)$ performs the best for this collection of data based on significance level 0.05. Our empirical results are corroborated by Figure 12. To demonstrate the unique property of each estimator, contour plots and log likelihood profiles are plotted in Figure 13.

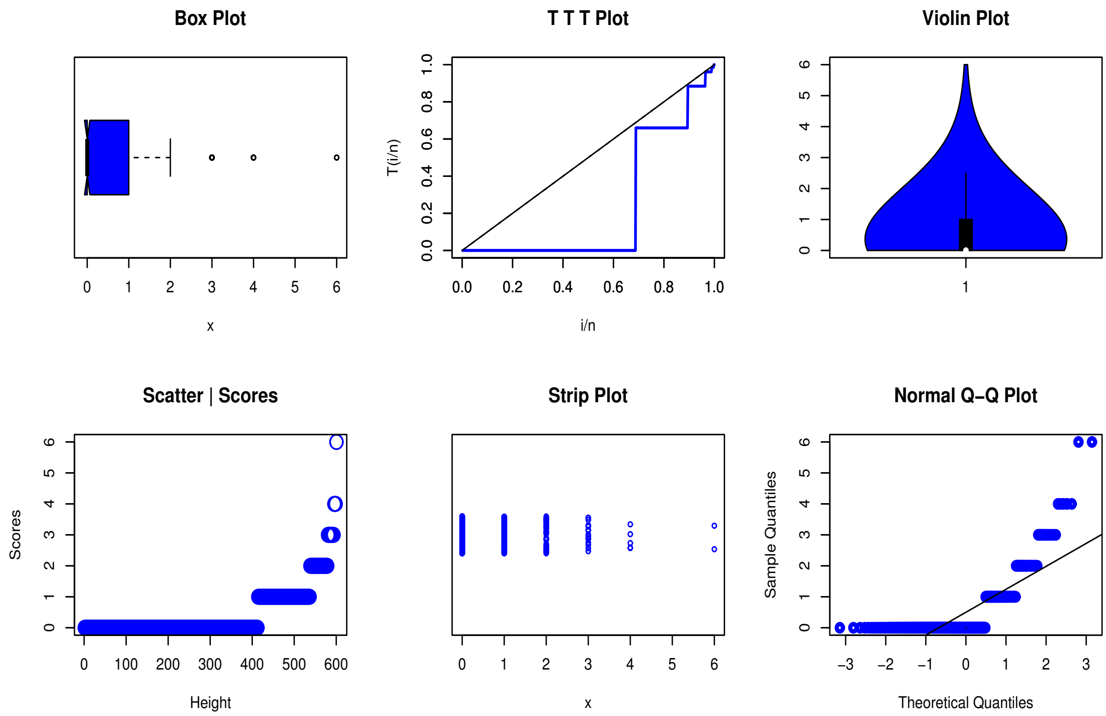

Figure 11. Nonparametric visualization plots for dataset II.

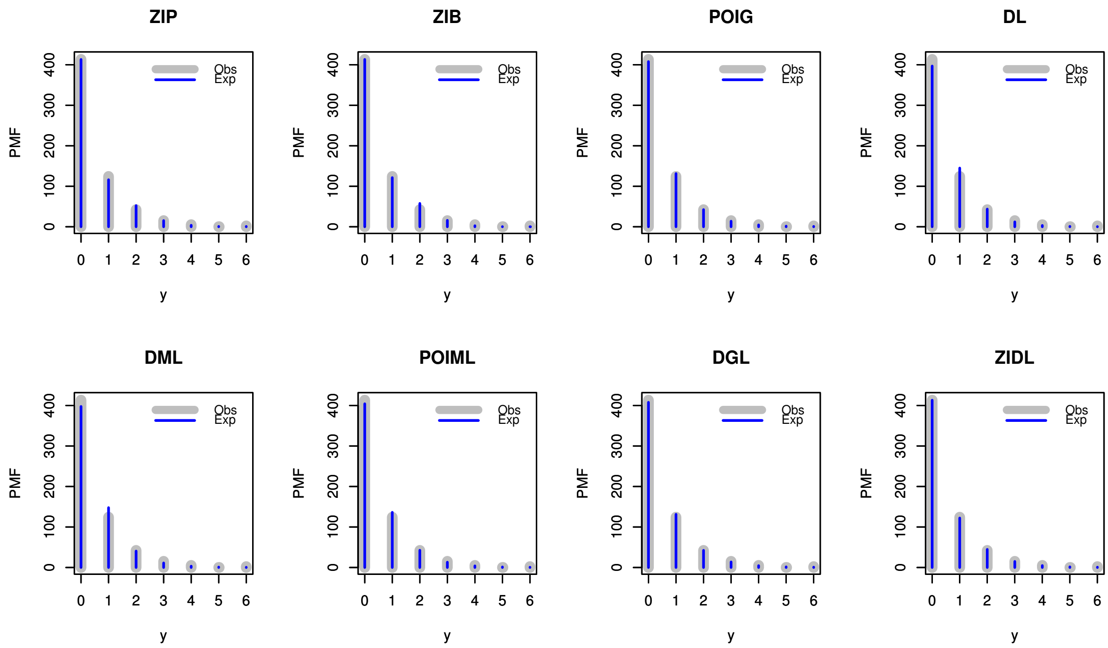

Figure 12. The estimated PMF for the dataset II.

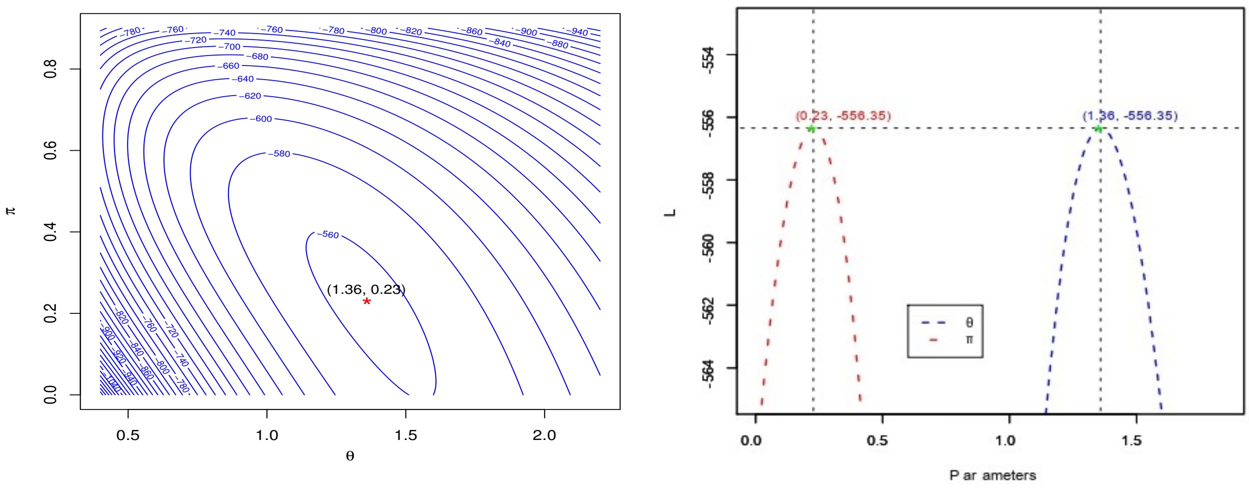

Figure 13. Contour plots (left panel) and log-likelihood profiles (right panel) for dataset II.

Table 5. The goodness-of-fit test for dataset II.

y	OF	ZIP	ZIB	POIG	DL	DML	POIML	DGL	ZIDL
0	413	412.99	413.02	407.67	396.71	397.43	404.07	407.68	412.97
1	124	115.98	121.16	131.14	144.79	147.61	136.09	131.14	122.05
2	42	52.14	57.48	42.18	43.48	40.42	42.44	42.18	44.48
3	15	15.14	15.70	13.57	11.91	11.11	12.87	13.57	14.79
4	5	3.51	2.42	4.36	3.09	3.15	3.87	4.36	4.66
5	0	0.63	0.20	1.40	0.77	0.91	1.16	1.40	1.42
6	2	0.61	0.00	0.68	0.25	0.37	0.50	0.67	0.63
Total	601	601	601	601	601	601	601	601	601
$-L$		559.58	566.68	556.52	559.65	559.86	557.25	556.52	556.35
MLEs		$\hat{\pi}$=0.47	$\hat{\pi}$=0.54	$\hat{\lambda}$¡0.0001	$\hat{\theta}$=1.55	$\hat{p}$=0.29	$\hat{\theta}$=2.28	$\hat{\alpha}$=0.53	$\hat{\pi}$=0.23
		$\hat{\lambda}$=0.89	$\hat{p}$=0.17	$\hat{\theta}$=0.67				$\hat{\eta}$=0.32	$\hat{\theta}$=1.36
χ^2		2.75	6.15	0.66	5.93	7.13	2.02	0.66	0.186
d.f		1	1	2	2	2	3	2	2
p.value		0.10	0.01	0.72	0.05	0.03	0.57	0.72	0.91

Various estimates derived from the dataset II are displayed in Table 6. Among all the evaluated strategies, the MLE strategy has been found to be the most effective.

Table 6. Various estimators for dataset II.

Strategy	$\hat{\pi}$	$\hat{\theta}$	χ^2	d.f	P.value
MLE	0.23	1.36	0.19	2	0.91
ProE	0.21	1.39	0.26	2	0.88
MoE	0.26	1.33	0.42	2	0.81

8. Conclusions and future work

In this article, a new zero-inflated probability model was proposed for a discrete random variable, and its distributional properties, including statistics, reliability and insurance properties, were extracted and studied in detail. The PMF of the proposed model proved to be effective in modeling and analyzing asymmetric data with varying forms of kurtosis, including leptokurtic and platykurtic. Additionally, the proposed model was capable of handling different types of dispersion, such as underdispersion and overdispersion. The reliability properties of the proposed model were explored, highlighting its relevance in the insurance field. A characterization of the proposed model was derived, based on the conditional distribution and the HRF, underscoring its significance in various applications. It was reported that the HRF of the new model could effectively model increasing, decreasing, and bathtub profiles. Various estimation techniques were used: MLE, MoE, and ProE to estimate the parameters of the proposed model. A simulation study was conducted to assess the bias and variance of the estimators under various schemes, revealing that all estimation methods performed well, with the maximum likelihood technique yielding superior results. The practical applicability of the proposed model was demonstrated using two real-life datasets, where the reported distribution provided a better fit to over-dispersed data compared to other count probability models. Future research could expand this work by exploring other inflated points, such as inflated one, inflated zero, and inflated twice. Furthermore, the development of a regression model for the proposed model and the study of its properties and applications remain potential areas for future investigation, although these were not addressed in this study due to structural complexity.

Author contributions

Wael W. Mohammed: Methodology, resources, supervision, visualization; Kalpasree Sharma: Conceptualization, formal analysis, software, visualization, writing – review & editing; Partha Jyoti Hazarika: Data curation, methodology, resources, supervision, validation; G. G. Hamedani: Conceptualization, investigation, validation, writing – review & editing; Mohamed S. Eliwa: Conceptualization, data curation, formal analysis, investigation, software, writing – review & editing; Mahmoud El-Morshedy: Formal analysis, investigation, methodology, software, validation, visualization. All authors have read and agreed to the published version of the manuscript

Use of Generative-AI tools declaration

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

Acknowledgments

The researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025).

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflict of interest

The authors declare no conflicts of interest.

References

1. J. Neyman, On a new class of "contagious" distributions, applicable in entomology and bacteriology, *Ann. Math. Statist.*, **10** (1939), 35–57. https://doi.org/10.1214/aoms/1177732245
2. W. Feller, On a general class of "contagious" distributions, *Ann. Math. Statist.*, **14** (1943), 389–400.
3. A. C. Cohen, Estimating the parameters of a modified poisson distribution, *J. Amer. Statist. Assoc.*, **55** (1960), 139–143.
4. K. Yoneda, Estimations in some modified poisson distributions, *Yokohama Math. J.*, **10** (1962), 73–96.
5. D. Lambert, Zero-inflated poisson regression, with an application to defects in manufacturing, *Technometrics*, **34** (1992), 1–14.
6. P. L. Gupta, R. C. Gupta, R. C. Tripathi, Inflated modified power series distributions with applications, *Comm. Statist. Theory Methods*, **24** (1995), 2355–2374. https://doi.org/10.1080/03610929508831621
7. T. H. Lin, M. H. Tsai, Modeling health survey data with excessive zero and k responses, *Stat. Med.*, **32** (2013), 1572–1583. https://doi.org/10.1002/sim.5650
8. M. Arora, N. R. Chaganty, EM estimation for zero-and k-inflated poisson regression model, *Computation*, **9** (2021), 94. https://doi.org/10.3390/computation9090094
9. Y. Sun, S. Zhao, G. L. Tian, M. L. Tang, T. Li, Likelihood-based methods for the zero-one-two inflated poisson model with applications to biomedicine, *J. Stat. Comput. Simul.*, **93** (2023), 956–982. https://doi.org/10.1080/00949655.2021.1970162
10. M. Melkersson, D. O. Rooth, Modeling female fertility using inflated count data models, *J. Popul. Econ.*, **13** (2000), 189–203. https://doi.org/10.1007/s001480050133
11. M. Begum, A. Mallick, N. Pal, A generalized inflated poisson distribution with application to modeling fertility data, *Thail. Stat.*, **12** (2014), 135–159.
12. H. Saboori, M. Doostparast, Zero to k inflated poisson regression models with applications, *J. Stat. Theory Appl.*, **22** (2023), 366–392. https://doi.org/10.1007/s44199-023-00067-3
13. F. Famoye, K. P. Singh, Zero-inflated generalized poisson regression model with an application to domestic violence data, *J. Data Sci.*, **4** (2006), 117–130. https://doi.org/10.6339/JDS.2006.04(1).257

14. D. B. Hall, Zero-inflated poisson and binomial regression with random effects: A case study, *Biometrics*, **56** (2000), 1030–1039. https://doi.org/10.1111/j.0006-341X.2000.01030.x

15. T. Rahman, P. J. Hazarika, M. P. Barman, One inflated binomial distribution and its real-life applications, *Indian J. Sci. Tech.*, **14** (2021), 1839–1854. http://dx.doi.org/10.17485/IJST/v14i22.404

16. B. P. Singh, S. Maheshwari, P. K. Gupta, A probability model for sex composition of children in the presence of son preference, *Demography India*, **44** (2015), 50–57.

17. G. Nanjundan, S. Pasha, A characterization of zero-inflated binomial model, *Int. J. Math. Comp. Res.*, **3** (2015), 1188–1190.

18. D. C. Heilbron, Zero-altered and other regression models for count data with added zeros, *Biom. J.*, **36** (1994), 531–547. https://doi.org/10.1002/bimj.4710360505

19. A. M. Garay, E. M. Hashimoto, E. M. M. Ortega, V. H. Lachos, On estimation and influence diagnostics for zero-inflated negative binomial regression models, *Comput. Statist. Data Anal.*, **55** (2011), 1304–1318. https://doi.org/10.1016/j.csda.2010.09.019

20. R. Suresh, G. Nanjundan, S. Nagesh, S. Pasha, On a characterization of zero-inflated negative binomial distribution, *Open J. Stat.*, **5** (2015), 511–513. http://dx.doi.org/10.4236/ojs.2015.56053

21. R. S. A. Alshkaki, Estimation of the parameters of the zero-one inflated power series distribution, *Bull. Math. Stat. Res.*, **4** (2016), 45–54.

22. I. J. A. Serra, D. L. L. Polestico, On the zero and k-inflated negative binomial distribution with applications, *Adv. Appl. Stat.*, **88** (2023), 1–23.

23. N. L. Johnson, A. W. Kemp, S. Kotz, *Univariate discrete distributions*, John Wiley & Sons, Inc., 2005. http://dx.doi.org/10.1002/0471715816

24. C. C. Iwunor, Estimation of parameters of the inflated geometric distribution for rural out-migration, *Genus*, **51** (1995), 253–260.

25. G. D. C. Barriga, F. Louzada, The zero-inflated conway–maxwell–poisson distribution: Bayesian inference, regression modeling and influence diagnostic, *Stat. Methodol.*, **21** (2014), 23–34. https://doi.org/10.1016/j.stamet.2013.11.003

26. A. J. Lemonte, G. Moreno-Arenas, F. Castellares, Zero-inflated bell regression models for count data, *J. Appl. Stat.*, **47** (2019), 265–286. https://doi.org/10.1080/02664763.2019.1636940

27. L. Rivas, F. Campos, Zero inflated waring distribution, *Comm. Statist. Simulation Comput.*, **52** (2023), 3676–3691. https://doi.org/10.1080/03610918.2021.1944638

28. R. Ospina, S. L. Ferrari, Inflated beta distributions, *Stat. papers*, **51** (2010), 111–126. https://doi.org/10.1007/s00362-008-0125-4

29. A. B. Ferreira, J. Mazucheli, The zero-inflated, one and zero-and-one-inflated new unit-Lindley distributions, *Braz. J. Biom.*, **40** (2022), 291–396. https://doi.org/10.28951/bjb.v40i3.571

30. H. S. Bakouch, M. A. Jazi, S. Nadarajah, A new discrete distribution, *Statistics*, **48** (2014), 200–240. https://doi.org/10.1080/02331888.2012.716677

31. D. V. Lindley, Fiducial distributions and bayes' theorem, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **20** (1958), 102–107.

32. A. Nandi, S. Chakraborty, A. Biswas, A new over-dispersed count model based on poisson-geometric convolution, *Comm. Statist. Simulation Comput.*, 2024, 1–28. https://doi.org/10.1080/03610918.2024.2329997

33. L. Tomy, G. Veena, C. Chesneau, Exponentiated discrete modified Lindley distribution and its applications in the healthcare sector, In: *G Families of probability distributions*, CRC Press, 2023, 279–287.

34. C. Chesneau, L. Tomy, G. Veena, The poisson-modified lindley distribution, *Appl. Math. E-Notes*, **22** (2022), 18–31.

35. M. El-Morshedy, E. Altun, M. S. Eliwa, A new statistical approach to model the counts of novel coronavirus cases, *Math. Sci.*, **16** (2022), 37–50. https://doi.org/10.1007/s40096-021-00390-9

36. A. Zeileis, C. Kleiber, countreg: Count data regression, *R package version 0.2-1*. Aviliable from: `https://rdrr.io/rforge/countreg/`

37. R. Shanker, H. Fesshaye, S. Selvaraj, A. Yemane, On zero-truncation of poisson and poisson-lindley distributions and their applications, *Biom. Biostat. Int. J.*, **2** (2015), 168–181. https://doi.org/10.15406/bbij.2015.02.00045