Pak.j.stat.oper.res. Vol.22 No.2 2026 pp 149-176 DOI: http://dx.doi.org/10.18187/pjsor.v22i2.4783

Pakistan Journal of Statistics and Operation Research

A Comprehensive Study of the Bernoulli-Transmuted Geometric Distribution: Theory and Applications

Anupama Nandi[1], Partha Jyoti Hazarika[2], Aniket Biswas[3], G. G. Hamedani[4], Morad Alizadeh[5], Mahmoud. El-Morshedy[6,7], Josmar Mazucheli[8], Mohamed S. Eliwa[9,10 *]

*Corresponding author

1. Department of Statistics, Dibrugarh University, India, anupamanandi0702@gmail.com
2. Department of Statistics, Dibrugarh University, India, parthajhazarika@gmail.com
3. Department of Mathematics, IIT Kharagpur, India, biswasaniket44@gmail.com
4. Department of Mathematical and Statistical Sciences, Marquette University, USA, gholamhoss.hamedani@marquette.edu
5. Department of Statistics, Persian Gulf University, Bushehr, Iran, m.alizadeh@pgu.ac.ir
6. Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia, m.elmorshedy@psau.edu.sa
7. Department of Statistics and Computer Sciences, Faculty of Science, Mansoura University, Mansoura 35516, Egypt
8. Department of Statistics, Universidade Estadual de Maringá, DEs, PR, Brazil
9. Department of Statistics and Operations Research, College of Science, Qassim University, Saudi Arabia, m.eliwa@qu.edu.sa
10. Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, Egypt.

Abstract

This paper presents the BerTG distribution, an innovative three-parameter discrete probability model developed by convolving a Bernoulli random variable with an independently distributed transmuted geometric random variable. The suggested distribution constitutes a significant and adaptable generalization that encompasses various established count distributions as specific instances, thereby offering a cohesive framework for modeling a range of count data formats. The BerTG distribution is notable for its exceptional ability to handle many types of dispersion, including as overdispersion, underdispersion, and equidispersion, commonly found in real-world count data, thereby overcoming a significant weakness of numerous conventional discrete models. A thorough examination of the distributional and structural characteristics of the BerTG model is conducted, including the probability mass function, cumulative distribution function, moments, moment generating function, factorial moments, probability generating function, and index of dispersion, among other aspects. Special emphasis is placed on reliability-theoretic attributes, encompassing the hazard rate function, survival function, reverse hazard rate function, and conditional expectation, which are meticulously generated and examined. Additionally, essential actuarial metrics, including the stop-loss premium, value-at-risk, and tail value-at-risk, are analysed to illustrate the model's appropriateness for risk-theoretic applications. Model parameters are estimated by maximum likelihood estimation, and the asymptotic properties of the resultant estimators are determined. A comprehensive simulation analysis is performed to assess the finite-sample performance of the estimators concerning bias, mean squared error, and consistency across diverse parameter configurations and sample sizes, thereby validating the reliability and accuracy of the estimation technique. The practical applicability of the BerTG distribution is evidenced through real-world data applications, wherein the model is applied to several empirical count datasets displaying diverse dispersion traits. Comparative analyses with various competing discrete distributions demonstrate that the BerTG model consistently attains superior goodness-of-fit performance, as indicated by standard model selection criteria such as the Akaike information criterion, Bayesian information criterion, and chi-square goodness-of-fit statistics. The results combined demonstrate that the BerTG distribution is a very competitive, manageable, and adaptable instrument for the statistical modelling of count data across several applicable fields.

Key words: Bernoulli distribution; Transmuted geometric distribution; Convolution; Dispersion; Maximum likelihood estimation; Reliability analysis; Actuarial measures; Simulation; Data analysis

Mathematical Subject Classification: 60E05; 62E15; 62F10.

1. Introduction

Over the last decade, there has been a notable upswing in interest surrounding the modeling of count data. The Poisson distribution stands out as the primary choice for modeling count datasets. The Poisson distribution is characterized by the mean and variance being identical. Consequently, it is not suitable for scenarios characterized by either over-dispersion or under-dispersion. To address the need for modeling both over and under-dispersed count data, various extensions and generalizations of the Poisson distribution have been developed in recent years. Key developments include hyper-Poisson (HP) distribution (Bardwell and Crow, 1964), double-Poisson distribution (Efron, 1986), weighted Poisson distribution (Del Castillo and Pérez-Casany, 1998), COM-Poisson distribution by Shmueli et al. (2005), weighted generalized Poisson distribution (Chakraborty, 2010), discrete weighted Poisson Lerch transcendental (DWPLT) distribution (El-Dawoody et al., 2023) and Zero-inflated Poisson generalized-Lindley (PGL) (Altun et al., 2023).

The gamma–Poisson mixture leads to the negative binomial distribution, introducing over-dispersion that the Poisson distribution cannot accommodate (Fisher et al. (1943)). To address over-dispersion in count data, several generalized forms of the geometric distribution have been developed (Chakraborty and Bhati (2016), Chakraborty and Gupta (2015), Gómez-Déniz (2010), Jain and Consul (1971), Makcutek (2008), and Tripathi et al. (1987)). An important generalization is the transmuted geometric distribution $TGD(\theta, \alpha)$ proposed by Chakraborty and Bhati (2016), which has been developed by applying the quadratic rank transmutation method (Shaw and Buckley, 2007). For $Y \sim TGD(\theta, \alpha)$ distribution, the probability mass function (pmf) of Y is

$$P(Y = y) = (1 - \alpha)\theta^y(1 - \theta) + \alpha(1 - \theta^2)\theta^{2y}, \quad y = 0, 1, 2... . \tag{1}$$

where, $0 < \theta < 1$ and $-1 < \alpha < 1$. It exhibits under-dispersion for $\alpha \in (-1, 0)$ and over-dispersion for $\alpha \in (0, 1)$. In contrast to count data distributions designed for over-dispersion, there are limited options in the literature to address under-dispersion.

An effective way to enhance the modeling flexibility of classical distributions is to consider convolutions of independent random variables. Bourguignon and Weiß (2017) introduced the BerG distribution by convolving Bernoulli and geometric random variables, while Bourguignon et al. (2022) proposed the BerPoi distribution based on Bernoulli–Poisson convolution. Related models developed using similar constructions include (Nandi et al., 2024a), (Nandi et al., 2024b) and (Nandi et al., 2025).

Building upon this simple yet powerful idea, we propose a novel count data model, the BerTG distribution, which addresses both over-dispersion and under-dispersion. This model is constructed by convolving two independent random variables: one following a Bernoulli distribution and the other a transmuted geometric distribution. This convolution enables the model to effectively capture both over-dispersed and under-dispersed count data. It extends important distributions like the transmuted geometric, geometric and BerG distributions. Ultimately, the BerTG model is proposed to provide a flexible and practical framework for modeling count data. It offers simple and closed-form expressions for key distributional properties and is well suited for data arising in real-world applications such as industrial reliability and mortality studies, where varying failure rate patterns are observed. The model is capable of handling over- and under-dispersion, different levels of skewness, and diverse kurtosis structures, and it consistently demonstrates superior performance compared to existing discrete models in the literature, particularly in the analysis of complete random data.

The remaining part of the article is organized as follows. Section 2 derives the BerTG distribution, and Section 3 presents its important statistical properties. Section 4 discusses characterizations of the proposed model, while Section 5 introduces its reliability and actuarial aspects. Parameter estimation using the maximum likelihood approach is described in Section 6, and the performance of the estimators is assessed through simulation experiments in Section 7. A BerTG-based count regression model is developed in Section 8, and Section 9 illustrates the model's practical usefulness using real data. Concluding remarks and future research directions are provided in Section 10.

2. Bernoulli-Transmuted Geometric Distribution

We obtain the BerTG distribution by convolving the Bernoulli distribution with the transmuted geometric distribution. Let, Y_1 follows the Bernoulli distribution $Ber(p)$ where $p \in (0,1)$ and Y_2 follow the transmuted geometric distribution $TGD(\theta, \alpha)$ where $\theta \in (0,1)$ and $\alpha \in (-1,1)$, respectively. Both the variables are restricted to non-negative integer values. Consider, $Y = Y_1 + Y_2$. Then, the pmf of $Y \sim BerTG(p, \theta, \alpha)$ is

$$p_Y(y) = \begin{cases} (1-p)(1-\theta)(1+\alpha\,\theta), & \text{if } \ y = 0 \\ \\ (1-\alpha)\,g_1(y,\theta,p) + \alpha\,g_2(y,\theta,p), & \text{if } y = 1,2,3,\ldots. \end{cases} \tag{2}$$

In (2), $g_i(y,\theta,p) = (1-\theta^i)\,\theta^{i(y-1)}(\theta^i(1-p)+p)$, $i = 1,2$. Defining $A_i = \begin{cases} 1 & \text{if } y = 0 \\ 0 & \text{otherwise} \end{cases}$, then for $y \in \{0,1,2\ldots\}$, the pmf in (2) can be written as

$$p_Y(y) = [(1-p)(1-\theta)(1+\alpha\,\theta)]^{A_i}\,[(1-\alpha)\,g_1(y,\theta,p) + \alpha\,g_2(y,\theta,p)]^{1-A_i}\,. \tag{3}$$

From the formulation of the $BerTG(p, \theta, \alpha)$ distribution, it is very convenient to obtain its mean and variance explicitly. Note that,

$$\begin{aligned} \mu = E(Y) &= p + \frac{\theta^2 + (1-\alpha)\theta}{1-\theta^2}, \\ \sigma^2 = V(Y) &= p(1-p) + \frac{\theta\left(1-\alpha+\theta\left(2+\theta(1-\alpha)-\alpha^2\right)\right)}{\left(1-\theta^2\right)^2}. \end{aligned} \tag{4}$$

Remark 1: The $BerTG(p, \theta, \alpha)$ distribution behaves like the BerG (Bourguignon and Weiß, 2017) distribution with parameter (p, θ) as $\alpha \to 0$, the transmuted geometric distribution Chakraborty and Bhati (2016) with parameter (θ, α) as $p \to 0$ and the geometric distribution with parameter θ as $p \to 0$ and $\alpha \to 0$.

The cumulative distribution function (cdf) of the random variable $Y \sim BerTG(p, \theta, \alpha)$ distribution is

$$F_Y(y) = (1-p)(1-\theta)(1+\alpha\,\theta) + Z_i\left[\sum_{u=1}^{y}\left((1-\alpha)\,g_1(u,\theta,p) + \alpha\,g_2(u,\theta,p)\right)\right], \tag{5}$$

where $Z_i = \begin{cases} 0 & \text{if } y = 0 \\ 1 & \text{otherwise} \end{cases}$. The corresponding survival function (sf) is

$$S_Y(y) = 1 - \left[(1-p)(1-\theta)(1+\alpha\,\theta) + Z_i \sum_{u=1}^{y-1} \left((1-\alpha)\,g_1(u,\theta,p) + \alpha\,g_2(u,\theta,p)\right)\right]. \tag{6}$$

The hazard rate function (hrf) is written as

$$h_Y(y) = \begin{cases} (1-p)(1-\theta)(1+\alpha\,\theta) & \text{if } \ y = 0 \\ \dfrac{(1-\alpha)\,g_1(y,\theta,p) + \alpha\,g_2(y,\theta,p)}{1-(1-p)(1-\theta)(1+\alpha\,\theta)} & \text{if } \ y = 1 \\ \dfrac{(1-\alpha)\,g_1(y,\theta,p) + \alpha\,g_2(y,\theta,p)}{1-\left[(1-p)(1-\theta)(1+\alpha\,\theta) + \sum_{u=1}^{y-1}\left((1-\alpha)\,g_1(u,\theta,p)+\alpha\,g_2(u,\theta,p)\right)\right]} & \text{if } \ y = 2,3,\ldots. \end{cases} \tag{7}$$

3. Properties of the BerTG distribution

In this section, we explore several important statistical properties of the proposed distribution. Some of the distributional properties studied here are the Recurrence relation, generating functions, moments related concepts, index of dispersion and coefficient of variation.

3.1. Recurrence relation

Probability recurrence relation helps in finding the subsequent term using the preceding term. It usually proves to be advantageous for calculating probabilities at different values. For a $BerTG(p,\theta,\alpha)$ distribution we have

$$p_Y(y+1) = p_Y(y)\left(\frac{\theta((1-\alpha)(\theta(1-p)+p) + \alpha(1+\theta)\theta^y(\theta^2(1-p)+p))}{(1-\alpha)(\theta(1-p)+p) + \alpha(1+\theta)\theta^{y-1}(\theta^2(1-p)+p)}\right), y = 1,2,\ldots. \tag{8}$$

which is the required recurrence relation.

3.2. Generating function

We use the notation G to denote a probability generating function (pgf) and use the notation of the corresponding random variable in the subscript. For $Y_1 \sim Ber(p)$ and $Y_2 \sim TGD(\theta,\alpha)$,

$$G_{Y_1}(s) = 1 - p(1-s) \quad \text{and} \quad G_{Y_2}(s) = \frac{(1-\theta)(1+\alpha\theta(1-s)-\theta^2 s)}{(1-\theta s)(1-\theta^2 s)}, \ |\theta^2 s| < 1.$$

Now by using the convolution property we obtain the pgf of $BerTG(p,\theta,\alpha)$ as

$$G_Y(s) = \frac{(1-p(1-s))(1-\theta)(1+\alpha\theta(1-s)-\theta^2 s)}{(1-\theta s)(1-\theta^2 s)}. \tag{9}$$

Similar methods are used to obtain the other generating functions, including the moment generating function $M_Y(t)$, characteristic function $\phi_Y(t)$ and cumulants generating function $K_Y(t)$. These are given below as

$$M_Y(t) = \frac{(1-p(1-e^t))(1-\theta)(1+\alpha\theta(1-e^t)-\theta^2 e^t)}{(1-\theta e^t)(1-\theta^2 e^t)}, \tag{10}$$

$$\phi_Y(t) = \frac{(1-p(1-e^{ti}))(1-\theta)(1+\alpha\theta(1-e^{it})-\theta^2 e^{it})}{(1-\theta e^{it})(1-\theta^2 e^{it})}, \tag{11}$$

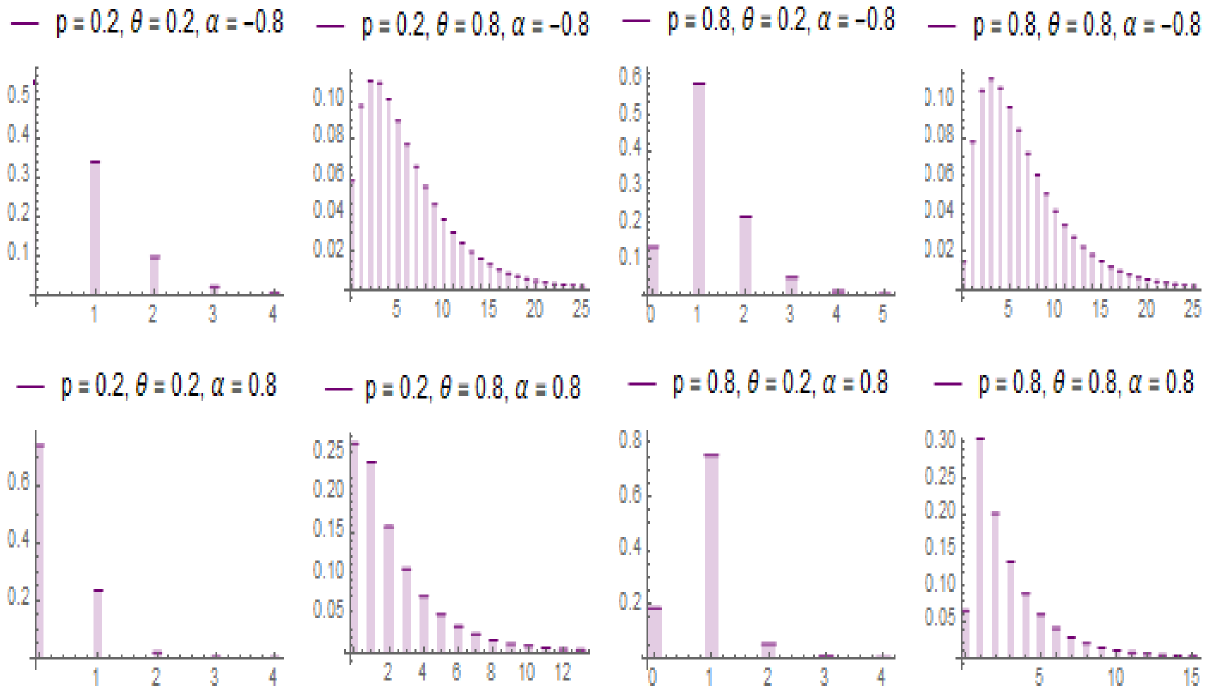

Figure 1: The pmf plots of $BerTG(p, \theta, \alpha)$ for different choices of p, θ and α.

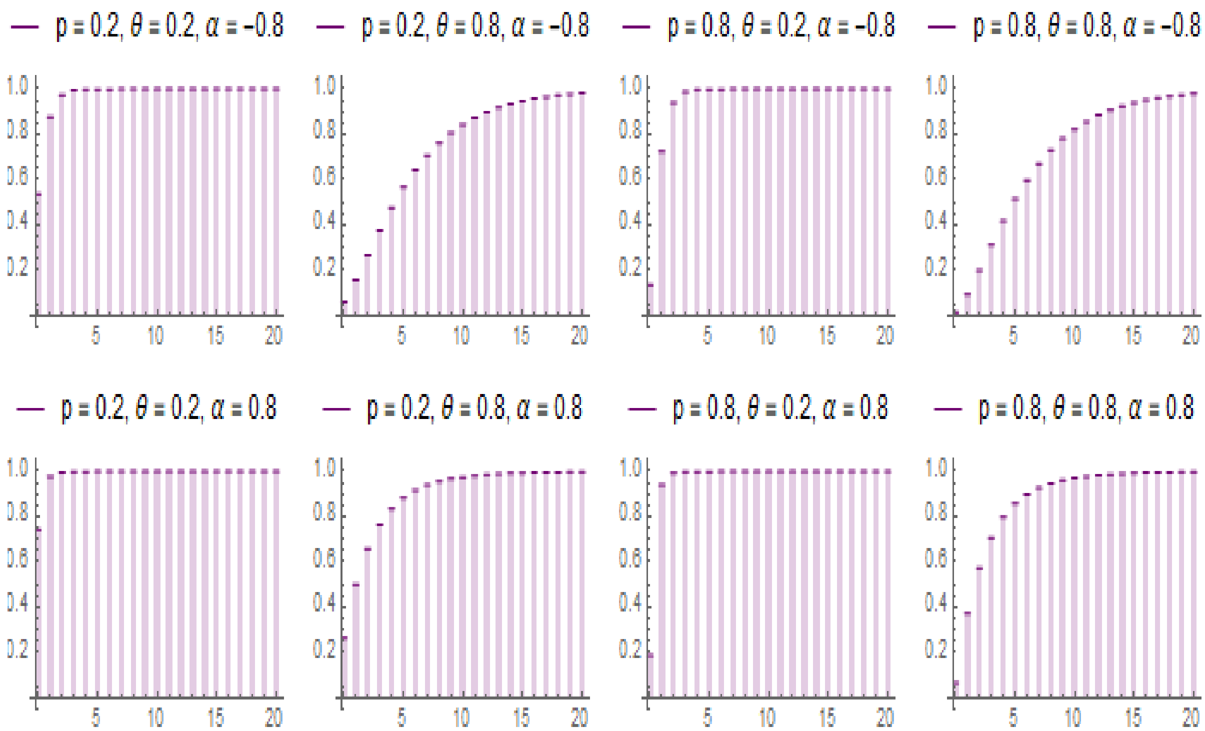

Figure 2: The cdf plots of $BerTG(p, \theta, \alpha)$ for different choices of p, θ and α.

and

$$K_Y(t) = \log\left[\frac{(1-p(1-e^t))(1-\theta)(1+\alpha\,\theta(1-e^t)-\theta^2 e^t)}{(1-\theta\,e^t)(1-\theta^2 e^t)}\right]. \tag{12}$$

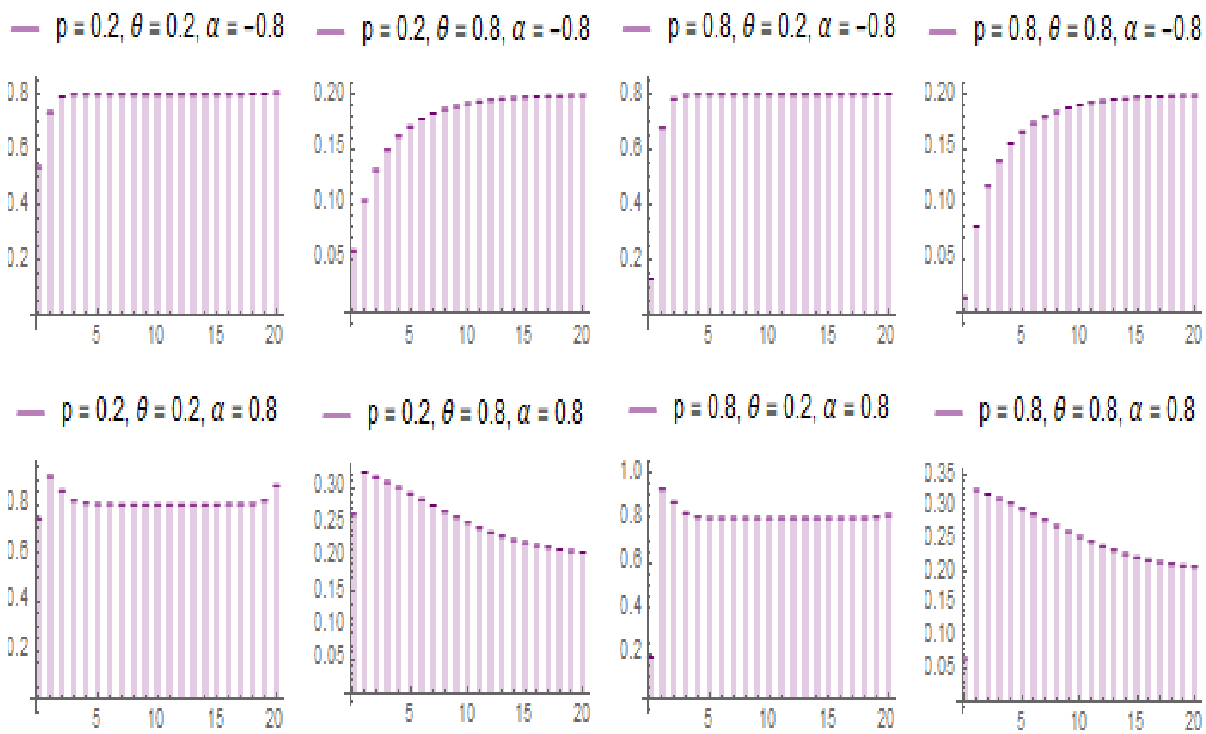

Figure 3: The hrf plots of $BerTG(p, \theta, \alpha)$ for different choices of p, θ and α.

3.3. Moments related concepts

Let μ'_k denote the raw moment of order k, that is $\mu'_k = E(Y^k)$. The k^{th} order raw moment of $Y \sim BerTG(p, \theta, \alpha)$ can be obtained using the general expressions of the raw moments of $Y_1 \sim Ber(p)$ and $Y_2 \sim TGD(\theta, \alpha)$ as follows.

$$E(Y^k) = E\left[\sum_{j=0}^{k} \binom{k}{j} {Y_1}^j {Y_2}^{k-j}\right]$$
$$= \sum_{j=0}^{k} \binom{k}{j} E({Y_1}^j) E({Y_2}^{k-j}).$$

Note that, $E({Y_1}^j) = p$ and $E(Y_{2(r)}) = (1-\alpha) r! \left(\frac{\theta}{1-\theta}\right)^r + \alpha r! \left(\frac{\theta^2}{1-\theta^2}\right)^r$, where $Y_{2(r)} = Y_2(Y_2-1)(Y_2-2)\ldots(Y_2-r+1)$.

$$E({Y_2}^{k-j}) = \sum_{r=0}^{k-j} \left\{ {k-j \atop r} \right\} E(Y_{2(r)})$$
$$= \sum_{r=0}^{k-j} \left\{ {k-j \atop r} \right\} (1-\alpha) r! \left(\frac{\theta}{1-\theta}\right)^r + \alpha r! \left(\frac{\theta^2}{1-\theta^2}\right)^r.$$

Consider $N_{k-j}(\theta, \alpha) = \sum_{r=0}^{k-j} \left\{ {k-j \atop r} \right\} (1-\alpha) r! \left(\frac{\theta}{1-\theta}\right)^r + \alpha r! \left(\frac{\theta^2}{1-\theta^2}\right)^r$. The general expressions for the raw moments of Y is as follows

$$\mu'_k = \sum_{j=0}^{k} \binom{k}{j} p\, N_{k-j}(\theta, \alpha). \tag{13}$$

The explicit expressions of the first four raw moments are

$$\mu_1' = p + \frac{\theta^2 + (1-\alpha)\theta}{1-\theta^2},$$

$$\mu_2' = \frac{1}{(1-\theta^2)^2}[\theta\left(-\alpha + \theta^3 - 3(\alpha-1)\theta^2 + (3-2\alpha)\theta + 1\right) + p\left(1-\theta^2\right)\left(\theta^2 - 2(\alpha-1)\theta + 1\right)],$$

$$\mu_3' = \frac{1}{(1-\theta^2)^3}[p\left(1-\theta^2\right)\left(\theta^4 + 6(1-\alpha)\theta^3 + (10-6\alpha)\theta^2 + 6(1-\alpha)\theta + 1\right) - \theta\left(\alpha - \theta^5 - 7(1-\alpha)\theta^4 + 4(3\alpha-4)\theta^3 - 16(1-\alpha)\theta^2 + (6\alpha-7)\theta - 1\right)],$$

$$\mu_4' = \frac{1}{(1-\theta^2)^4}[\theta(-\alpha + \theta^7 - 15(\alpha-1)\theta^6 + (61-50\alpha)\theta^5 - 115(\alpha-1)\theta^4 + (115-104\alpha)\theta^3 - 61(\alpha-1)\theta^2 + (15-14\alpha)\theta + 1) - p\left(\theta^2-1\right)(\theta^6 - 14(\alpha-1)\theta^5 + (47-36\alpha)\theta^4 - 68(\alpha-1)\theta^3 + (47-36\alpha)\theta^2 - 14(\alpha-1)\theta + 1)].$$

The first four central moments are

$$\mu_2 = p(1-p) + \frac{\theta\left(1-\alpha+\theta\left(2+\theta(1-\alpha)-\alpha^2\right)\right)}{(1-\theta^2)^2},$$

$$\mu_3 = p(p-1)(2p-1) - \frac{\theta\left(2\alpha^3\theta^2 + 3\alpha^2\theta\left(\theta^2+1\right) + \alpha\left(\theta^4+4\theta^2+1\right) - (\theta+1)^4\right)}{(1-\theta^2)^3},$$

$$\mu_4 = 3p^3(2-p) - \frac{2p^2\left(\theta\left(-3\alpha+\theta\left(-3\alpha^2-3\alpha\theta+\theta(2\theta+3)+2\right)+3\right)+2\right)}{(\theta^2-1)^2} + \frac{p\left(\theta\left(-6\alpha+\theta\left(-6\alpha^2-6\alpha\theta+\theta(\theta+6)+10\right)+6\right)+1\right)}{(\theta^2-1)^2} + \frac{\theta(\theta(\theta+7)+1)}{(\theta-1)^4} - \frac{\alpha\theta(\theta(\alpha+\theta)+1)\left(\theta\left(3(\alpha+2)+\theta\left(3\alpha^2+\theta^2+3(\alpha+2)\theta+22\right)\right)+1\right)}{(\theta^2-1)^4}.$$

The mean and variance of the $BerTG(p,\theta,\alpha)$ distribution correspond to the first raw and second central moments, respectively. While the skewness and kurtosis measures, denoted as $\mu_3/\mu_2^{3/2}$ and μ_4/μ_2^2, can be directly calculated from the central moments, we opt not to provide their detailed expressions due to their intricate structures. Instead, we present 3-D surface plots illustrating the skewness and kurtosis measures in Figure 4 and Figure 5, respectively.

Figure 4 reveals that the $BerTG(p,\theta,\alpha)$ distribution exhibits positive skewness. When α is held constant, it becomes evident that as $p \to 1$ and $\theta \to 0$, the skewness of $BerTG(p,\theta,\alpha)$ becomes zero. Interestingly, the skewness appears unaffected by changes in the parameter α suggesting that variations in α do not appreciably impact the skewness of the $BerTG(p,\theta,\alpha)$ distribution.

In Figure 5, the horizontal plane in each plot is consistently positioned at a value of 3, which never intersects the kurtosis surface. This implies that the $BerTG(p,\theta,\alpha)$ distribution is leptokurtic. Notably, all the parameter has a significant influence on the kurtosis. Examining the third row of the kurtosis plot reveals that, for a p value of 0.5 and as θ approaches 0, the curve intersects the horizontal plane, indicating a mesokurtic nature.

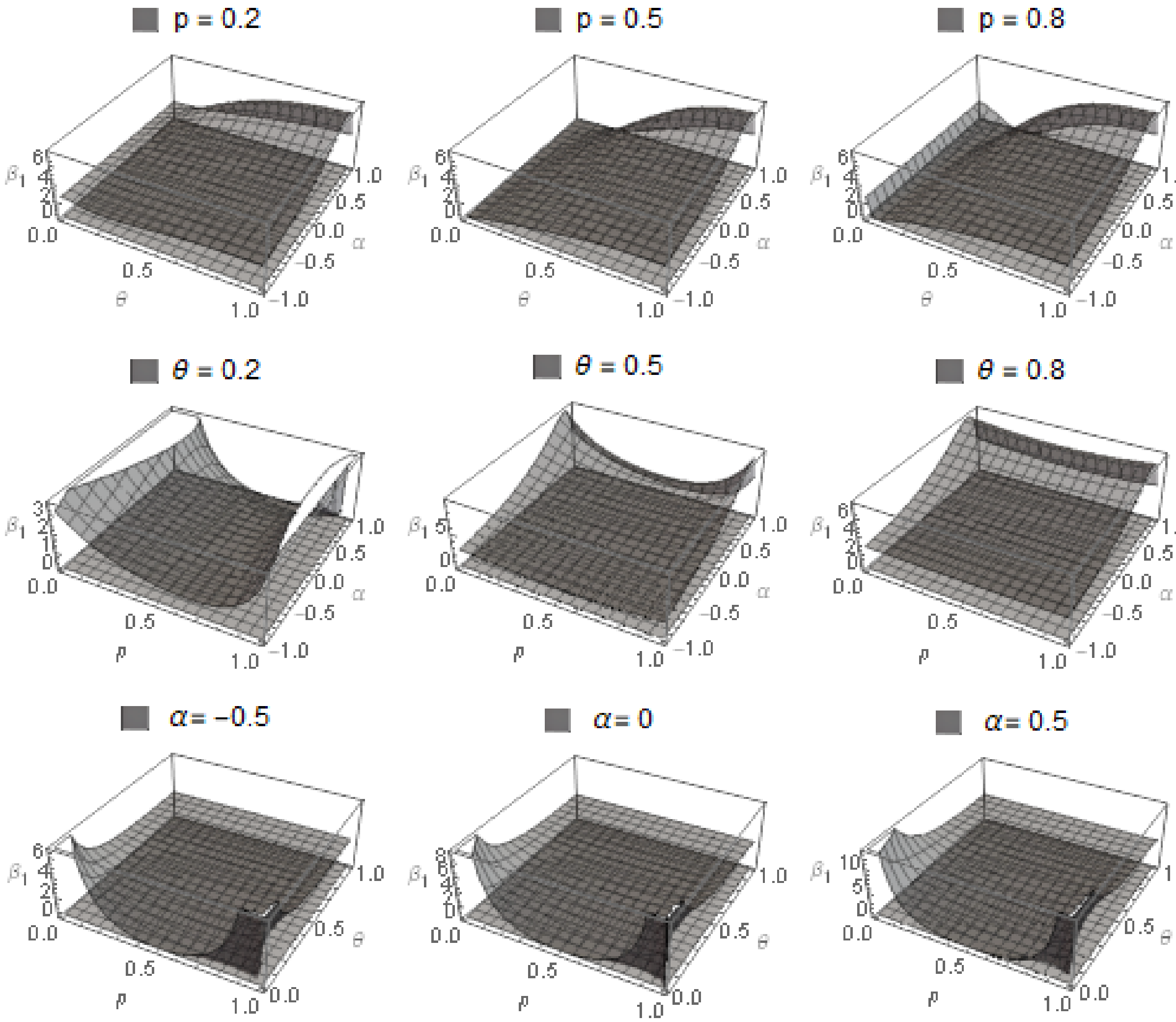

Figure 4: Skewness plots of $BerTG(p, \theta, \alpha)$ for different choices of p, θ, and α.

3.4. Dispersion index and coefficient of variation

The dispersion index (DI), represented as I_Y (Hoel (1943)), serves as a metric for quantifying the extent of dispersion within a distribution. It helps assess whether a given distribution is appropriate for representing a dataset that is either over-dispersed, under-dispersed, or equi-dispersed. When I_Y surpasses one, it indicates that the distribution of variable Y is suitable for modeling over-dispersion. Conversely, if I_Y is less than one, the distribution is apt for capturing under-dispersion. A distribution is considered equi-dispersed when I_Y equals one. The dispersion index is given by

$$I_Y = \frac{p(1-p)(1-\theta^2)^2 + \theta((1-\alpha) + \theta(2+\theta(1-\alpha)-\alpha^2))}{(1-\theta^2)(p(1-\theta^2)+\theta^2+\theta(1-\alpha))} \tag{14}$$

We present a 3-D Plot 6 depicting the index of dispersion for the $BerTG(p, \theta, \alpha)$ distribution, showcasing various combinations of (p, θ, α) to illustrate the dispersion characteristics. As evident from Figure 6, the $BerTG(p, \theta, \alpha)$ distribution exhibits both over-dispersion and under-dispersion characteristics. The dispersion index I_Y is greatly influenced by the choice of p and θ values. Over-dispersion is observed when $p \to 0$ and $\theta \to 1$, while under-dispersion is evident when $p \to 1$ and $\theta \to 0$. The influence of α on I_Y is observed to be marginal.

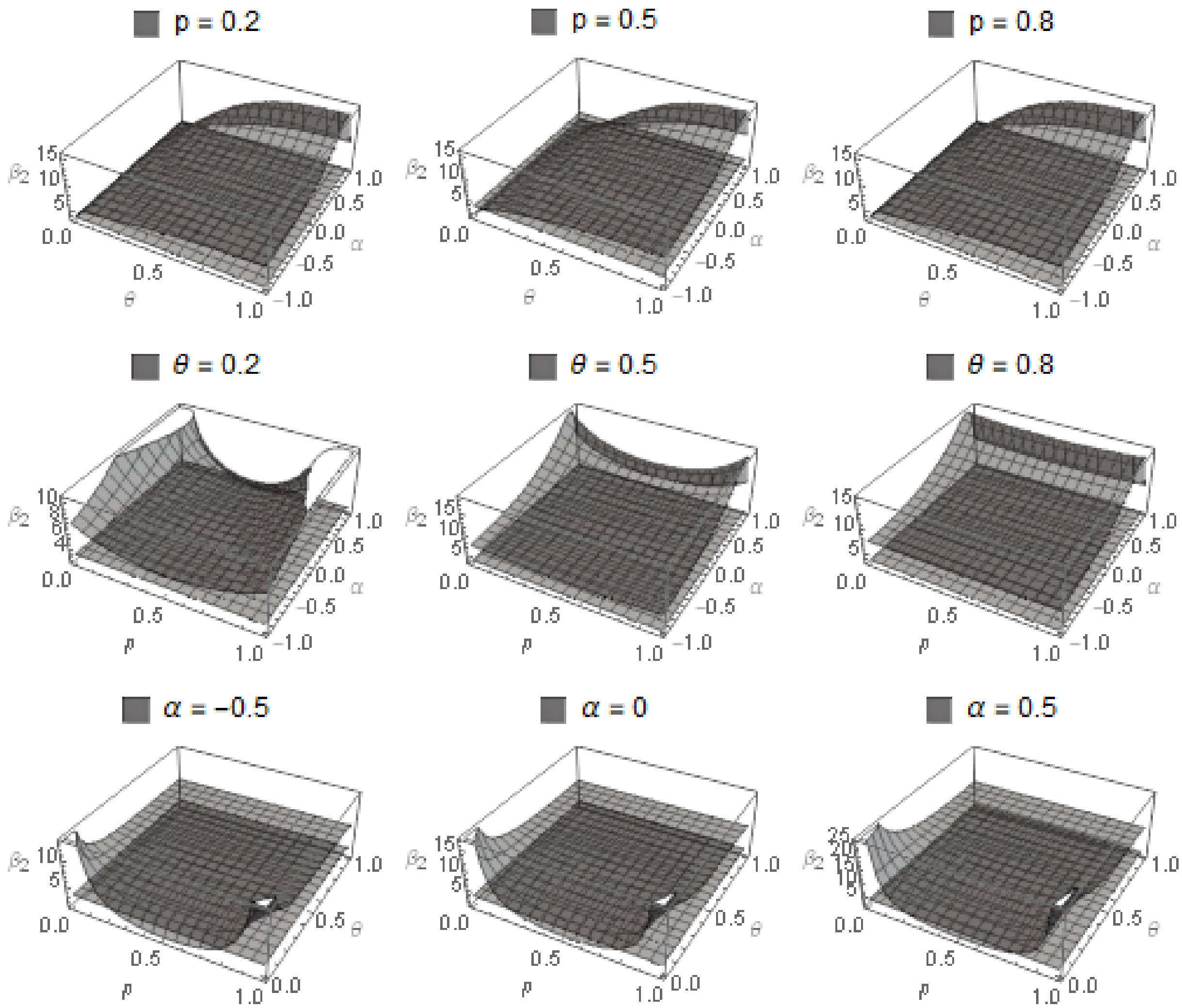

Figure 5: Kurtosis plots of $BerTG(p,\theta,\alpha)$ for different choices of p, θ, and α.

The coefficient of variation (cv) of $BerTG(p,\theta,\alpha)$ distribution is given by

$$cv = \frac{\sqrt{p(1-p)(1-\theta^2)^2 + \theta((1-\alpha)+\theta(2+\theta(1-\alpha)-\alpha^2))}}{p(1-\theta^2)+\theta^2+\theta(1-\alpha)} \times 100\%.$$

4. Characterization

Characterizations of distributions are of considerable interest to researchers in applied fields. Investigators are often keen to determine whether a given model satisfies the defining properties of a particular distribution. In this context, characterizations provide necessary conditions under which the underlying distribution uniquely corresponds to that specific distribution. Certain characterizations of BerTG distribution are presented in two directions: (i) based on the conditional expectation of an appropriate function of the random variable and (ii) in terms of the reverse hazard function.

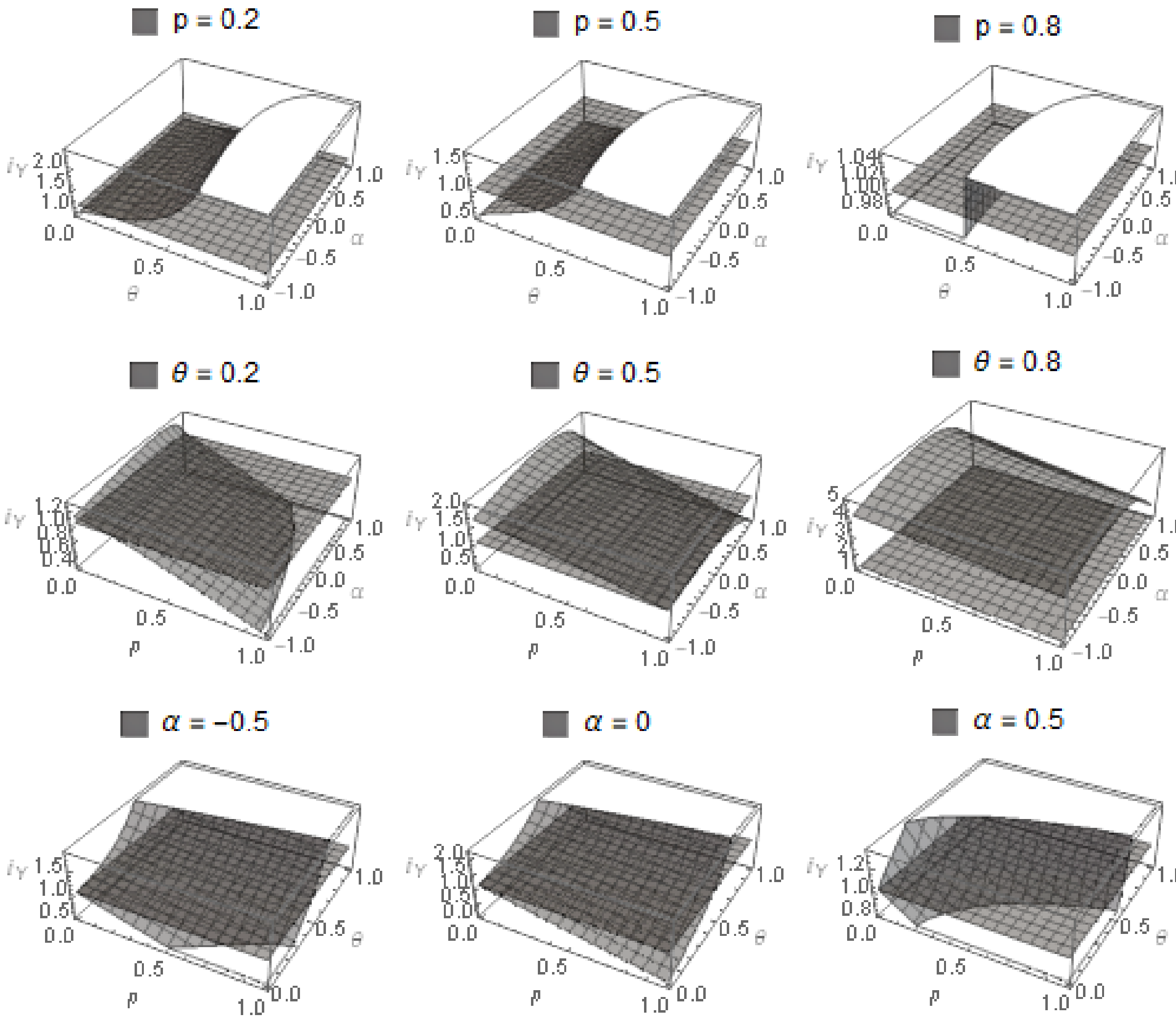

Figure 6: DI plots of $BerTG(p, \theta, \alpha)$ for various combination of p, θ, α.

4.1. Based on conditional expectation

In this subsection, we present our characterization of BerTG in terms of the conditional expectation of certain function of the random variable. The choice of the function depends on the form of the pmf. The pmf in (2) can be written as

$$p_Y(y) = \begin{cases} C_0 & \text{if} \;\; y = 0 \\ \theta^y Q\,(y) & \text{if} \;\; y \in \mathbb{N} \end{cases} \tag{15}$$

where $C_0 = (1-p)\,(1-\theta)\,(1+\alpha\theta)$, $Q\,(0) = 0$, $Q\,(y) = C_1 + C_2\,\theta^y$, $y \in \mathbb{N}$, $C_1 = (1-\alpha)\,(1-\theta)\,\theta^{-1}\,(\theta\,(1-p)+p)$ and $C_2 = \alpha\left(1-\theta^2\right)\theta^{-2}\left(\theta^2(1-p)+p\right)$.

In similar way the cdf can be written as $F(y) = \begin{cases} C_0 & \text{if} \;\; y = 0 \\ C_0 + \sum_{u=1}^{y} \theta^u Q\,(u) & \text{if} \;\; y \in \mathbb{N} \end{cases}$

Proposition 1. Y is a random variable with pmf given in (15) if and only if

$$E\left\{Q\left(y\right)^{-1} \mid Y \leq k\right\} = \left(\frac{1}{C_0 + \sum_{y=1}^{k} \theta^y Q\left(y\right)}\right)\left(\frac{1-\theta^{k+1}}{1-\theta}\right). \tag{16}$$

Proof. If Y has pmf (15), then for $k \in \mathbb{N}$, the left-hand side of (16), using finite geometric sum formula, will be

$$F\left(k\right)^{-1}\sum_{y=1}^{k}\theta^y = \left(\frac{1}{C_0 + \sum_{y=1}^{k} \theta^y Q\left(y\right)}\right)\left(\frac{1-\theta^{k+1}}{1-\theta}\right).$$

Conversely, if (16) holds, then

$$\sum_{y=1}^{k}\left\{Q\left(y\right)^{-1} p\left(y\right)\right\} = F\left(k\right)\left(\frac{1}{C_0 + \sum_{y=1}^{k} \theta^y Q\left(y\right)}\right)\left(\frac{1-\theta^{k+1}}{1-\theta}\right). \tag{17}$$

From (17), we also have

$$\begin{aligned}\sum_{y=1}^{k-1}\left\{Q\left(y\right)^{-1} p\left(y\right)\right\} &= F\left(k-1\right)\left(\frac{1}{C_0 + \sum_{y=1}^{k-1} \theta^y Q\left(y\right)}\right)\left(\frac{1-\theta^{k}}{1-\theta}\right)\\ &= \left(F\left(k\right) - p\left(k\right)\right)\left(\frac{1}{C_0 + \sum_{y=1}^{k-1} \theta^y Q\left(y\right)}\right)\left(\frac{1-\theta^{k}}{1-\theta}\right).\end{aligned} \tag{18}$$

Now, subtracting (18) from (17) , yields

$$\begin{aligned}&p\left(k\right)\left(Q\left(y\right)^{-1} - \frac{1-\theta^k}{(1-\theta)(C_0 + \sum_{y=1}^{k-1} \theta^y Q\left(y\right))}\right) =\\ &\quad F\left(k\right)\left(\frac{1-\theta^{k+1}}{(1-\theta)(C_0 + \sum_{y=1}^{k} \theta^y Q\left(y\right))} - \frac{1-\theta^k}{(1-\theta)(C_0 + \sum_{y=1}^{k-1} \theta^y Q\left(y\right))}\right).\end{aligned}$$

From the above equality, we have

$$\frac{p\left(k\right)}{F\left(k\right)} = \frac{\frac{1-\theta^{k+1}}{(1-\theta)(C_0+\sum_{y=1}^{k}\theta^y Q(y))} - \frac{1-\theta^k}{(1-\theta)(C_0+\sum_{y=1}^{k-1}\theta^y Q(y))}}{\frac{1}{Q(y)} - \frac{1-\theta^k}{(1-\theta)(C_0+\sum_{y=1}^{k-1}\theta^y Q(y))}}$$

$$= \begin{cases} 1 & \text{if } \ y = 0 \\ \frac{\theta^k Q(k)}{C_0+\sum_{y=1}^{k}\theta^y Q(y)} & \text{if } \ y \in \mathbb{N}. \end{cases}$$

which, after some computations, is the reverse hazard function corresponding to the pmf (15), so Y has pmf (15).

4.2. Based on reverse hazard function

Proposition 2. Y is a random variable with pmf given in (15) if and only if its reverse hazard rate function, r_F satisfies the difference equation

$$r_F\left(k+1\right)-r_F\left(k\right)=\frac{\theta^{k+1}Q\left(k+1\right)}{C_0+\sum_{y=1}^{k+1}\theta^{y}Q\left(y\right)}-\frac{\theta^{k}Q\left(k\right)}{C_0+\sum_{y=1}^{k}\theta^{y}Q\left(y\right)},\quad k\in\mathbb{N} \tag{19}$$

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

Proof. If Y has pmf (15) , then clearly (19) holds. Now, if (19) holds, then for every $y\in\mathbb{N}$, we have

$$\sum_{k=1}^{y-1}\left\{r_F\left(k+1\right)-r_F\left(k\right)\right\}=\sum_{k=1}^{y-1}\left(\frac{\theta^{k+1}Q\left(k\right)}{C_0+\sum_{y=1}^{k+1}\theta^{y}Q\left(y\right)}-\frac{\theta^{k}Q\left(k\right)}{C_0+\sum_{y=1}^{k}\theta^{y}Q\left(y\right)}\right),$$

or, using telescoping sum

$$r_F\left(y\right)-r_F\left(0\right)=\frac{\theta^{k}Q\left(y\right)}{C_0+\sum_{u=1}^{y}\theta^{u}Q\left(u\right)},$$

or in view of the initial condition

$$r_F(y)=\begin{cases}1, & \text{if } \ y=0\\ \frac{\theta^{k}Q(k)}{C_0+\sum_{y=1}^{k}\theta^{y}Q(y)}, & \text{if } \ y\in\mathbb{N}\end{cases}$$

which is the reverse hazard function corresponding to the pmf (15).

5. Reliability and Actuarial Properties

5.1. Residual coefficient of variation and mean variance based on past lifetimes

The residual coefficient of variation (rcov) and variance of mean past life (mpl) represent statistical tools utilized for examining the variability and temporal trends within data sets. Beginning with the residual coefficient of variation, it denotes a standardized metric expressing the dispersion relative to the mean, computed as the ratio of standard deviation to mean, thereby serving as a percentage. Specifically, the rcov isolates and quantifies the remaining variability in the data post-exclusion of identified factors like seasonal shifts or trends. This measure proves particularly advantageous in time series analysis and regression modeling for gauging the randomness or unpredictability of the residual variation. On the other hand, the variance of mean past life characterizes the variability in the means of previous data points over a specified duration, offering insights into the stability or volatility of the underlying phenomenon or process under investigation. Employed frequently in forecasting and predictive modeling, it assesses the reliability of historical data in forecasting future outcomes, with a lower variance indicating heightened consistency or stability and a higher variance suggesting increased unpredictability or fluctuation. These measures find extensive applications across diverse domains encompassing economics, finance, engineering, and social sciences. They facilitate endeavors such as market trend analysis, risk assessment, quality control, process optimization, and demographic trend examination, ultimately enhancing decision-making processes and fostering a deeper comprehension of intricate phenomena through insights into underlying patterns, variability, and trends over time. 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 mean residual life (mrl) $\varpi(k;p,\theta,\alpha)$ is defined as follows

$$\varpi(k;p,\theta,\alpha) = E(Y-k|Y \geq k); \quad k \in V, \tag{20}$$

where $V = \{1,2,3,...,l\}$ for $0 < l < \infty$. Consider Y be a BerTG random variable, then the mrl can be expressed as

$$\varpi(k;p,\theta,\alpha) = -\frac{-\alpha p\theta^4 + \alpha\theta^4 + p\theta^4 + \alpha p\theta^2 + \alpha\theta^3 - \theta^4 + \alpha p\theta^{k+4} - \alpha p\theta^{k+2} - p\theta^2 - \theta^3 - \alpha\theta^{k+4}}{(\theta^2-1)(\alpha p\theta^{k+2} - \alpha\theta^{k+2} - \alpha p\theta^2 - \alpha p\theta^k + \alpha p\theta + \alpha\theta^2 + p\theta^2 - p\theta - \theta^2)}; \ k \in V. \tag{21}$$

The function for the variance residual life (vrl), denoted as $\Upsilon_{vrl}(k)$, is defined

$$\begin{aligned}\Upsilon_{vrl}(k;p,\theta,\alpha) &= E(Y^2|Y \geq k) - [E(Y|Y \geq k)]^2 \\ &= \frac{2}{(\theta-1)^2 S(k-1;p,\theta,\alpha)}\Theta(k;p,\theta,\alpha) - (2k-1)\varpi(k;p,\theta,\alpha) - [\varpi(k;p,\theta,\alpha)]^2,\end{aligned}$$

where

$$\begin{aligned}\Theta(k;p,\theta,\alpha) &= (k\theta - k - \theta)[-\alpha p\theta^{k+1} + \alpha p\theta^k + \alpha\theta^{k+1} + p\theta^{k+1} - p\theta^k - \theta^{k+1}] \\ &\quad + \alpha(k\theta^2 - k - \theta^2)(p\theta^2 - p - \theta^2)\theta^{2k}.\end{aligned}$$

The random variable X exhibits increasing (decreasing) vrl if

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

The rcov, denoted as $\Xi(k;p,\theta,\alpha)$, can be explicitly derived as

$$\Xi(k;p,\theta,\alpha) = \sqrt{\Upsilon_{VRL}(k;p,\theta,\alpha)}/\varpi(k;p,\theta,\alpha); \ k \in V.$$

The hrf, mrl, and vrl of the BerTG model are interconnected in the following manner

$$\begin{aligned}\Upsilon_{vrl}(k+1;p,\theta,\alpha) - \Upsilon_{vrl}(k;p,\theta,\alpha) &= h(k;p,\theta,\alpha)[\Upsilon_{vrl}(k+1;p,\theta,\alpha) \\ &\quad - \varpi(k;p,\theta,\alpha)(1 + \varpi(k+1;p,\theta,\alpha))].\end{aligned}$$

Further, the hrf, mrl, vrl, and rcov of the BerTG distribution are interconnected as

$$\begin{aligned}\Upsilon_{vrl}(k+1;p,\theta,\alpha) - \Upsilon_{vrl}(k;p,\theta,\alpha) &= h(k;p,\theta,\alpha)[\varpi(k+1;p,\theta,\alpha)]^2 \times \Big\{[\Xi(k+1;p,\theta,\alpha)]^2 \\ &\quad - \frac{\varpi(k;p,\theta,\alpha)(1 + \varpi(k+1;p,\theta,\alpha))}{[\varpi(k+1;p,\theta,\alpha)]^2}\Big\}.\end{aligned}$$

Another significant concept within reliability theory pertains to what is known as the mpl. In the case of a discrete random variable, the mpl, denoted as $\varpi^*(k;p,\theta,\alpha)$, is defined as

$$\varpi^*(k;p,\theta,\alpha) = E(k-Y|Y \leq k); \quad k \in V.$$

For the BerTG distribution, the mpl can be expressed as

$$\varpi^*(k;p,\theta,\alpha) = \frac{\left\{\begin{array}{c}\alpha p\theta^{k+2} - \alpha\theta^{k+2} - p\theta^{k+2} - \alpha p\theta^{2(k+1)} + \alpha p\theta^{2k} - \alpha p\theta^k - k\theta^2 + p\theta^2 \\ -\alpha\theta^{k+1} + \theta^{k+2} + \alpha\theta^{2(k+1)} + p\theta^k + \alpha\theta - \theta^2 + \theta^{k+1} - k - \theta - p\end{array}\right\}}{(\theta^2-1)\left\{\begin{array}{c}\alpha p\theta^{k+1} - \alpha p\theta^{2(k+1)} + \alpha p\theta^{2k} - \alpha p\theta^k - \alpha\theta^{k+1} - p\theta^{k+1} \\ + \alpha\theta^{2(k+1)} + p\theta^k + \theta^{k+1} - 1\end{array}\right\}}.$$

Utilizing $\varpi^*(k; p, \theta, \alpha)$, the reversed hrf (rhrf) and cdf of the BerTG random variable distribution can be formulated as

$$r(k; p, \theta, \alpha) = \frac{1 - \varpi^*(k+1; p, \theta, \alpha) + \varpi^*(k; p, \theta, \alpha)}{\varpi^*(k; p, \theta, \alpha)};\ k \in V, \tag{22}$$

and

$$F(s; p, \theta, \alpha) = \left\{ \prod_{i=1}^{l} \left[\frac{\varpi^*(k; p, \theta, \alpha)}{\varpi^*(k+1; p, \theta, \alpha) - 1} \right] \right\}^{-1} \prod_{k=1}^{s} \left[\frac{\varpi^*(k; p, \theta, \alpha)}{\varpi^*(k+1; p, \theta, \alpha) - 1} \right];\ s \in V,$$

respectively. Further, the variance of mpl, say $\Upsilon^*_{vrl}(k)$, can be formulated as

$$\begin{aligned} \Upsilon^*_{vrl}(k; p, \theta, \alpha) &= E(Y^2 | Y \leq k) - [E(Y | Y \leq k)]^2 \\ &= (2k+1)\varpi^*(k; p, \theta, \alpha) - [\varpi^*(k; p, \theta, \alpha)]^2 \\ &\quad - \frac{2}{F(k; p, \theta, \alpha)} \left\{ \begin{array}{c} \frac{k(k+1)}{2} - \frac{\alpha p\theta - \alpha p - \alpha\theta - p\theta + p + \theta}{\theta^2 - 2\theta + 1} + \frac{\alpha(k\theta^2 - k - 1)\theta^{2k}[-p - \theta^2 + p\theta^2]}{\theta^4 - 2\theta^2 + 1} \\ + \frac{\alpha(p\theta^2 - \theta^2 - p)}{\theta^4 - 2\theta^2 + 1} + \frac{(k\theta - k - 1)\theta^k[-p - \alpha p\theta - \theta + \alpha p + p\theta + \alpha\theta]}{\theta^2 - 2\theta + 1} \end{array} \right\}. \end{aligned}$$

5.2. Premium principles

Premium principles are employed to calculate insurance premiums for various events, accounting for the associated risk levels. Over the years, several premium principles have been devised. This subsection presents some of them, assuming a loss distribution following the BerTG distribution. Within this subsection, let $\rho \geq 0$ denote the risk loading parameter.

5.2.1. Expected value principle

The expected value principle (evp) is a fundamental concept in insurance and risk management used to determine insurance premiums. It states that the premium charged for insurance coverage should be equal to the expected value of the losses, adjusted for a risk loading factor. The evp can be defined as

$$\text{evp}(\rho; .) = (1 + \rho)E(Y),$$

where evp$(\rho; .)$ is the insurance premium, ρ is the risk loading factor, and $E(Y)$ is the expected value of the losses (or the expected value of the distribution of losses). The term $(1 + \rho)$ "Risk Loading" represents the risk loading factor, which is added to the expected value of losses to cover various expenses and provide the insurer with a profit margin. The value of ρ depends on factors such as administrative costs, claims processing, underwriting, and the insurer's desired profit level. The evp of the BerTG distribution, say evp$(\rho; p, \theta, \alpha)$, is characterized by the equation

$$\text{evp}(\rho; p, \theta, \alpha) = (1 + \rho) \left(p + \frac{\theta(1 - \alpha) + \theta^2}{1 - \theta^2} \right).$$

The evp is a cornerstone of insurance pricing, utilized across various insurance types like property, liability, health, and life insurance. Insurers base premiums on anticipated losses, adjusting for risk loading to ensure fairness. This approach aligns premiums with the expected costs of coverage, enabling insurers to cover expenses and generate profit while ensuring policyholders pay appropriate premiums relative to the transferred risks. Economically, the evp encourages risk transfer, fostering efficiency through risk pooling and minimizing financial impacts of uncertainty. However, it assumes known and accurately estimated loss distributions, overlooking factors like moral hazard and market competition that can impact pricing and market behavior.

5.2.2. Exponential premium principle

The exponential premium principle (epp) is a fundamental concept in insurance pricing that relies on an exponential function to determine premiums. This principle recognizes that the probability of experiencing high-loss events tends to increase rapidly as the level of risk rises. The epp posits that insurance premiums should increase exponentially with the level of risk associated with the insured event. In other words, as the probability of severe losses grows, premiums should rise at an accelerating rate to adequately cover the potential costs of such events. The epp is derived by solving for epp in the equation

$$u(w - \mathrm{epp}(\rho; .)) = E(w - Y),$$

where w denotes the wealth of an individual and $u(y) = -e^{-\rho y}$ represents the exponential utility function. Consequently, the epp is obtained as

$$\mathrm{epp}(\rho; .) = \frac{1}{\rho} M_Y(\rho),$$

where $M_Y(\rho)$ is the mgf. The epp of the BerGT distribution, say $\mathrm{epp}(\rho; p, \theta, \alpha)$, is characterized as

$$\mathrm{epp}(\rho; .) = \frac{(1 - p(1 - e^{\rho}))(1 - \theta)(1 + \alpha\theta(1 - e^{\rho}) - \theta^2 e^{\rho})}{\rho(1 - \theta e^{\rho})(1 - \theta^2 e^{\rho})}.$$

This principle is commonly applied in various insurance sectors, including property insurance (e.g., coverage for natural disasters), liability insurance (e.g., coverage for high-risk activities), and health insurance (e.g., coverage for pre-existing medical conditions). In these sectors, premiums are often calculated based on complex actuarial models that incorporate the epp to ensure adequate risk coverage.

6. Parameter Estimation

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ denote a random sample of size n drawn from the $BerTG(p, \theta, \alpha)$ distribution, and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ denote a realization on $\mathbf{Y}$. The objective of this section is to estimate the parameters p, θ and α based on the available data $\mathbf{y}$ through maximum likelihood (ML) method. Using the pmf of $Y \sim BerTG(p, \theta, \alpha)$ in (3), the log-likelihood function of the parameters p, θ and α can easily be found as

$$l(p, \theta, \alpha; \boldsymbol{y}) = n_0 \log\{(1 - p)(1 - \theta)(1 + \alpha\,\theta)\} + (1 - A_i) \sum_{i=1}^{n_1} \log\{(1 - \alpha)\, g_1(y_i, p, \theta) + \alpha\, g_2(y_i, p, \theta)\} \tag{23}$$

where, $A_i = \begin{cases} 1, & \text{if } y_i = 0 \\ 0, & \text{otherwise} \end{cases}$, $n_0 = \sum_{i=1}^{n} I_0$, $n_1 = n - n_0$ and

$$g_i(y, \theta, p) = (1 - \theta^i)\, \theta^{i(y-1)} (\theta^i (1 - p) + p),\ i = 1, 2.$$

Let consider

$$\begin{aligned}
a_1(i) &= 1-\alpha+2\alpha\theta^{y_i+1}\\
a_2(i) &= \theta-\alpha\theta+2\alpha\theta^{y_i}+2\alpha\theta^{y_i+2}\\
a_3(i) &= \theta-\alpha\theta+2\alpha\theta^{y_i}+4\alpha\theta^{y_i+1}+2\alpha\theta^{y_i+2}\\
a_4(i) &= 1-\alpha+2\alpha\theta^{y_i}+2\alpha\theta^{y_i+1}\\
a_5(i) &= \theta-\alpha\theta+\alpha\theta^{y_i}+2\alpha\theta^{y_i+1}+\alpha\theta^{y_i+2}\\
a_6(i) &= 1-\alpha+\alpha\theta^{y_i}+\alpha\theta^{y_i+1}\\
b_1(i) &= \theta-\alpha\theta+3\alpha\theta^{y_i}+\alpha\theta^{y_i+4}\\
b_2(i) &= \theta(1-\theta)(3+\theta)(1-\alpha)+\alpha\theta^{y_i}(10-4\theta^2-6\theta^4)\\
b_3(i) &= (1+\theta)(1-\alpha)+2\alpha\theta^{y_i}+6\alpha\theta^{y_i+2}\\
b_4(i) &= \theta(1-\alpha)+4\alpha\theta^{y_i}(1+\theta)^2\\
b_5(i) &= (1-\alpha)+4\alpha\theta^{y_i}(1+\theta)\\
c(i) &= p(1-\theta)a_5(i)+\theta^2a_6(i)
\end{aligned}$$

To obtain the score functions, we differentiate (23) with respect to the parameters p, θ, and α.

$$\frac{\partial}{\partial p}l(p,\theta,\alpha;\boldsymbol{y}) = -\frac{n_0}{1-p}+\sum_{i=1}^{n_1}\frac{(1-A_i)\{(1-\alpha)(1-\theta)^2\theta^{y_i-1}+\alpha(1-\theta^2)^2\theta^{2(y_i-1)}\}}{(1-\alpha)\,g_1(y_i,p,\theta)+\alpha\,g_2(y_i,p,\theta)}, \tag{24}$$

$$\begin{aligned}\frac{\partial}{\partial \theta}l(p,\theta,\alpha;\boldsymbol{y}) = \frac{n_0(\alpha-2\alpha\theta-1)}{(1-\theta)(1+\alpha\theta)}-\sum_{i=1}^{n_1}(1-A_i)\\
\frac{a_1(i)\theta^3+a_2(i)p(1-\theta^2)-(a_3(i)p(1-\theta)+a_4(i)\theta^2)(1-\theta)y_i}{(1-\alpha)\,g_1(y_i,p,\theta)+\alpha\,g_2(y_i,p,\theta)},\end{aligned} \tag{25}$$

and $$\frac{\partial}{\partial \alpha}l(p,\theta,\alpha;\boldsymbol{y}) = \frac{n_0q}{1+\alpha\theta}-\sum_{i=1}^{n_1}\frac{(1-A_i)(g_1(y_i,p,\theta)-g_2(y_i,p,\theta))}{(1-\alpha)\,g_1(y_i,p,\theta)+\alpha\,g_2(y_i,p,\theta)}. \tag{26}$$

The maximum likelihood estimates are, in principle, obtained by simultaneously solving the equations formed by setting the right-hand sides of (24), (25), and (26) equal to zero. However, due to the structural complexity of these equations, explicit expressions for the maximum likelihood estimators are not tractable. Consequently, we directly maximize the log-likelihood function with respect to the parameters using appropriate numerical techniques. Specifically, the *optim* function in *R*, which efficiently implements the Nelder–Mead method, is employed for this purpose. Let $\widehat{p}_{ML}$, $\widehat{\theta}_{ML}$, and $\widehat{\alpha}_{ML}$ represent the maximum likelihood estimates (MLE) for p, θ, and α respectively. Now, our objective is to obtain asymptotic confidence intervals for all the parameters. For this purpose, we require the information matrix. The second-order partial derivative of the log-likelihood of the $BerTG(p, \theta, \alpha)$ are given below.

$$\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p^2} = -\frac{n_0}{(1-p)^2} - \sum_{i=1}^{n_1} \frac{(1-A_i)(1-\theta^2)a_5(i)^2}{\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2},$$

$$\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p \partial \theta} = -(1-A_i)\sum_{i=1}^{n_1} \frac{\theta[(1-\alpha)(\theta(1-\alpha)+2\alpha\theta^{y_i}(1+\theta+\theta^2)) + 2\alpha^2\theta^{2y_i}(1+\theta)^2 - \alpha(1-\alpha)(1-\theta^2)\theta^{y_i}y_i]}{\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2},$$

$$\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p \partial \alpha} = (1-A_i)\sum_{i=1}^{n_1} \frac{(1-\theta^2)\theta^{y_i+2}}{\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2},$$

$$\begin{aligned}\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta^2} = &-\frac{n_0(1+2\alpha\theta+\alpha^2(1-2\theta+2\theta^2))}{(1-\theta)^2(1+\alpha\theta)^2} + (1-A_i) \\ &\sum_{i=1}^{n_1}\Big[\frac{\left(\theta^3 a_1(i) + p(1-\theta^2)a_2(i) - (1-\theta)y_i(p(1-\theta)a_3(i)+\theta^2 a_4(i))\right)^2}{\theta^2(1-\theta)^2\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2} + \\ &\frac{2(p\, b_1(i) - \alpha\theta^{y_i+4}) - y_i(p\, b_2(i)+\theta^2 b_3(i)) + (1-\theta)y_i^2(p(1-\theta)b_4(i)+\theta^2 b_5(i))}{\theta^2(1-\theta)\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2}\Big],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta \partial \alpha} = &\frac{n_0}{(1+\alpha\theta)^2} + (1-A_i) \\ &\sum_{i=1}^{n_1} \frac{\theta^{y_i}(\theta^4 - p^2(1-\theta)^3(1+\theta) - 2p(\theta-\theta^3+\theta^4) + p(\theta+\theta^2-2\theta^3)y_i + (1+\theta)(\theta^3+p^2(1-\theta)^2(1+\theta)))}{\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \alpha^2} = &-\frac{n_0\theta^2}{(1+\alpha\theta)^2} - (1-A_i) \\ &\sum_{i=1}^{n_1} \frac{\left(\theta^2(\theta^{y_i+1}+\theta^{y_i}-1) + p(1-\theta)(\theta^{y_i+2}+2\theta^{y_i+1}+\theta^{y_i}-\theta)\right)^2}{\left(p(1-\theta)a_5(i)+\theta^2 a_6(i)\right)^2}.\end{aligned}$$

The Fisher's information matrix for (p, θ, α) is written as

$$I_Y(p,\theta,\alpha) = \begin{pmatrix} -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p^2}\right) & -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\theta}\right) & -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\alpha}\right) \\ -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\theta}\right) & -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta^2}\right) & -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta\partial\alpha}\right) \\ -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\alpha}\right) & -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta\partial\alpha}\right) & -E\left(\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \alpha^2}\right) \end{pmatrix}.$$

This can be approximated by

$$\widehat{I}_Y(p,\theta,\alpha) \approx \begin{pmatrix} -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p^2} & -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\theta} & -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\alpha} \\ -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\theta} & -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta^2} & -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta\partial\alpha} \\ -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial p\partial\alpha} & -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \theta\partial\alpha} & -\frac{\partial^2 l(p,\theta,\alpha;\boldsymbol{y})}{\partial \alpha^2} \end{pmatrix}_{(p,\theta,\alpha)=(\hat{p}_{ML},\hat{\theta}_{ML},\hat{\alpha}_{ML})}.$$

For a large value of n, the maximum likelihood estimators $\hat{p}_{ML}$, $\hat{\theta}_{ML}$, and $\hat{\alpha}_{ML}$ exhibit consistency and asymptotic normality with a mean vector of $(0,0,0)$ and a dispersion matrix denoted by $\hat{I}^{-1} = [dij]_{3\times3}$. Here, $dij = d(\hat{p}_{ML}, \hat{\theta}_{ML}, \hat{\alpha}_{ML}; y)$ represents the elements of the dispersion matrix $\hat{I}^{-1}$. The dispersion matrix $\hat{I}^{-1}$ includes the variances of $\hat{p}_{ML}$, $\hat{\theta}_{ML}$, and $\hat{\alpha}_{ML}$, denoted by d_{11}, d_{22}, and d_{33}, respectively. Furthermore, let z_γ represent the $(1-\gamma)$-th quantile of the standard normal distribution. The asymptotic $100 \times (1-\gamma)\%$ confidence interval for the parameters p, θ and α are given by

$$\left(\hat{p}_{ML} - z_{\gamma/2}\sqrt{d_{11}}\ ,\ \hat{p}_{ML} + z_{\gamma/2}\sqrt{d_{11}}\right), \left(\hat{\theta}_{ML} - z_{\gamma/2}\sqrt{d_{22}}\ ,\ \hat{\theta}_{ML} + z_{\gamma/2}\sqrt{d_{22}}\right)$$

and

$$\left(\hat{\alpha}_{ML} - z_{\gamma/2}\sqrt{d_{33}}\ ,\ \hat{\alpha}_{ML} + z_{\gamma/2}\sqrt{d_{33}}\right).$$

7. Simulation study

Here our aim is to examine the performance of maximum likelihood estimates for unknown parameters in finite sample settings. To generate a sample of size n, denoted as $\mathbf{y}$, from the $BerTG(p,\theta,\alpha)$ distribution, we independently generate $\mathbf{y_1}$ and $\mathbf{y_2}$ from $Ber(p)$ and $TGD(\theta,\alpha)$, respectively. It's important to note that both $\mathbf{y_1}$ and $\mathbf{y_2}$ consist of n independent components.

We generate $B = 10000$ samples of size n independently, and for each sample, we calculate the maximum likelihood estimates, storing them in different arrays. If $\hat{\eta}$ represents an estimate of the parameter η, then the bias of $\hat{\eta}$, denoted as $Bias(\hat{\eta})$, and the mean-squared error of $\hat{\eta}$, denoted as $MSE(\hat{\eta})$, are approximated as follows:

$$\frac{1}{B}\sum_{i=1}^{B}(\hat{\eta}_i - \hat{\eta}) \text{ and } \frac{1}{B}\sum_{i=1}^{B}(\hat{\eta}_i - \hat{\eta})^2,$$

respectively. Here $\hat{\eta}_i$ denotes the estimate of η for the $i-th$ generated sample.

Table 1: Performance of the maximum likelihood estimates.

p	θ	α	n	$Bias(\hat{p}_{ML})$	$MSE(\hat{p}_{ML})$	$Bias(\hat{\theta}_{ML})$	$MSE(\hat{\theta}_{ML})$	$Bias(\hat{\alpha}_{ML})$	$MSE(\hat{\alpha}_{ML})$
			100	0.0104	0.0095	0.0100	0.0100	0.1371	0.4009
0.2	0.2	-0.5	500	-0.0031	0.0041	0.0130	0.0029	0.0577	0.2553
			1000	-0.0057	0.0031	0.0087	0.0014	0.0189	0.1861
			100	-0.0407	0.0082	-0.0627	0.0164	-0.7822	1.1641
		0.5	500	-0.0219	0.0027	-0.0142	0.0062	-0.3441	0.5345
			1000	-0.0145	0.0015	-0.0044	0.0045	-0.2150	0.3347
			100	0.0341	0.0245	0.0158	0.0042	0.1424	0.2103
0.2	0.5	-0.5	500	0.0063	0.0118	0.0071	0.0008	0.0539	0.0709
			1000	-0.0045	0.0086	0.0043	0.0004	0.0212	0.0439
			100	-0.0475	0.0145	-0.0406	0.0099	-0.3140	0.3369
		0.5	500	-0.0267	0.0049	-0.0093	0.0042	-0.1077	0.1298

			1000	-0.0168	0.0022	-0.0024	0.0031	-0.0515	0.0747
			100	0.0311	0.0542	0.0041	0.0007	0.0430	0.1133
0.2	0.8	-0.5	500	-0.0169	0.0261	0.0009	0.0001	-0.0177	0.0336
			1000	-0.0204	0.0192	0.0003	0.0001	-0.0240	0.0228
			100	-0.0315	0.0236	-0.0251	0.0029	-0.2445	0.2276
		0.5	500	-0.0196	0.0068	-0.0063	0.0010	-0.0585	0.0673
			1000	-0.0131	0.0033	-0.0026	0.0006	-0.0260	0.0431
			100	-0.0208	0.0102	-0.0004	0.0085	0.0252	0.3353
0.5	0.2	-0.5	500	-0.0065	0.0022	0.0005	0.0020	-0.0199	0.1480
			1000	-0.0026	0.0009	0.0003	0.0009	-0.0235	0.0935
			100	-0.0129	0.0052	-0.0612	0.0148	-0.5916	0.8849
		0.5	500	-0.0031	0.0008	-0.0121	0.0050	-0.1952	0.2732
			1000	-0.0020	0.0004	-0.0020	0.0036	-0.0916	0.1311
			100	-0.0308	0.0250	0.0063	0.0038	0.0021	0.1922
0.5	0.5	-0.5	500	-0.0221	0.0098	0.0012	0.0007	-0.0279	0.0648
			1000	-0.0140	0.0055	-0.0002	0.0003	-0.0254	0.0366
			100	-0.0307	0.0121	-0.0492	0.0111	-0.3130	0.3568
		0.5	500	-0.0082	0.0017	-0.0116	0.0039	-0.0778	0.0872
			1000	-0.0051	0.0008	-0.0031	0.0027	-0.0308	0.0517
			100	-0.0518	0.0703	0.0015	0.0006	-0.0006	0.1034
0.5	0.8	-0.5	500	-0.0357	0.0285	0.0001	0.0001	-0.0170	0.0271
			1000	-0.0244	0.0149	0.0000	0.0001	-0.0131	0.0148
			100	-0.0420	0.0283	-0.0258	0.0029	-0.2266	0.2061
		0.5	500	-0.0096	0.0039	-0.0070	0.0010	-0.0599	0.0626
			1000	-0.0074	0.0019	-0.0026	0.0006	-0.0230	0.0413
			100	-0.0027	0.0028	-0.0003	0.0068	0.0088	0.2626
0.8	0.2	-0.5	500	-0.0004	0.0005	-0.0003	0.0013	-0.0278	0.0964
			1000	-0.0002	0.0003	-0.0002	0.0006	-0.0185	0.0541
			100	-0.0007	0.0020	-0.0489	0.0116	-0.4709	0.6847
		0.5	500	-0.0006	0.0004	-0.0085	0.0042	-0.1333	0.1730
			1000	-0.0005	0.0002	0.0001	0.0029	-0.0536	0.0825
			100	-0.0160	0.0100	0.0003	0.0032	-0.0086	0.1283
0.8	0.5	-0.5	500	-0.0032	0.0014	-0.0007	0.0006	-0.0104	0.0272
			1000	-0.0006	0.0006	-0.0001	0.0003	-0.0032	0.0130
			100	-0.0053	0.0034	-0.0457	0.0100	-0.2445	0.2352
		0.5	500	-0.0020	0.0006	-0.0117	0.0035	-0.0628	0.0662
			1000	-0.0014	0.0003	-0.0035	0.0024	-0.0258	0.0440
			100	-0.0355	0.0376	0.0080	0.0023	0.1085	0.2482
0.8	0.8	-0.5	500	-0.0067	0.0061	-0.0001	0.0001	-0.0019	0.0149
			1000	-0.0030	0.0027	0.0000	0.0001	-0.0013	0.0074
			100	-0.0113	0.0091	-0.0273	0.0031	-0.2149	0.1794
		0.5	500	-0.0033	0.0015	-0.0075	0.0009	-0.0601	0.0588
			1000	-0.0020	0.0007	-0.0029	0.0006	-0.0236	0.0390

From Table 1, it is apparent that the average biases and mean squared errors of maximum likelihood estimates for the parameters exhibit a decrease as the sample size increases. This declining trend supports the consistency of the maximum likelihood estimators.

8. BerTG regression model

Here we introduce a reparametrization of the $BerTG(p, \theta, \alpha)$ model, followed by the utilization of the generalized linear model (GLM) approach to establish a novel count regression model based on the proposed distribution, denoted as $BerTG_{GLM}$. To re-parameterize the $BerTG(p, \theta, \alpha)$ distribution, we express p in terms of μ as follows:

$$\begin{aligned} p &= \mu - \frac{\theta^2 + (1-\alpha)\theta}{1-\theta^2} \\ &= \mu - h(\theta, \alpha) \end{aligned}$$

where, $h(\theta, \alpha) = \frac{\theta^2+(1-\alpha)\theta}{1-\theta^2}$. This reparametrization is applied to the probability mass function (pmf) given in (3). The pmf of $Y \sim BerTG(\mu, \theta, \alpha)$ is given by

$$p_Y(y) = \{(1-\mu+h(\theta,\alpha))(1-\theta)(1+\alpha\,\theta)\}^{A_i}\{(1-\alpha)\,w_1(y)+\alpha\,w_2(y)\}^{1-A_i}. \quad (27)$$

In the above expression, $w_i(y) = (1-\theta^i)\,\theta^{i(y-1)}(\theta^i(1-\mu+h(\theta,\alpha)) + (\mu - h(\theta,\alpha)))$, $i = 1, 2$, and $A_i = \begin{cases} 1, & \text{if } y = 0 \\ 0, & \text{otherwise} \end{cases}$.

Consider an observed sample of size n, denoted by y_1, y_2, ..., y_n, drawn from the $BerTG(\mu, \theta, \alpha)$ model defined in (27).Each observation y_i represents a response variable associated with a given set of covariates $\boldsymbol{x_i'}$, for $i = 1, 2, ..., n$. We further assume that the mean of the response variable y_i is linked to the covariates through a log-link function, given by:

$$\mu_i = e^{\boldsymbol{x'_i\beta}}, \quad i = 1, 2, ..., n. \quad (28)$$

Here, the vector of covariates corresponding to the i-th observation, denoted as $\boldsymbol{x_i'}$, is given by $(1, x_{i1}, x_{i2}, ..., x_{ip})$ for $i = 1, 2, ..., n$. Additionally, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_p)'$ represents the vector of unknown regression coefficients. Replacing μ_i using (28) in (27), we write the pmf of $y_i|\boldsymbol{x_i'} \sim BerTG(\mu_i, \theta, \alpha)$ as

$$p_Y(y_i|\boldsymbol{x_i'}) = \{(1-e^{\boldsymbol{x'_i\beta}}+h(\theta,\alpha))(1-\theta)(1+\alpha\,\theta)\}^{A_i}\{(1-\alpha)\,w_1(y_i)+\alpha\,w_2(y_i)\}^{1-A_i}, \quad (29)$$

where $w_j(y_i) = (1-\theta^j)\,\theta^{j(y_i-1)}(\theta^j(1-e^{\boldsymbol{x'_i\beta}}+h(\theta,\alpha)) + (e^{\boldsymbol{x'_i\beta}} - h(\theta,\alpha)))$, $j = 1, 2$. Using (29), we obtain the log-likelihood function of $\boldsymbol{\delta} = (\boldsymbol{\beta},\, \theta,\, \alpha)'$ for given $y_1, y_2, ..., y_n$ and fixed $\boldsymbol{x_1'}, \boldsymbol{x_2'}, ..., \boldsymbol{x_n'}$ as

$$\begin{aligned} l(\boldsymbol{\delta}) = n_0 \log\{(1-e^{\boldsymbol{x'_i\beta}}+h(\theta,\alpha))(1-\theta)(1+\alpha\,\theta)\} + (1-A_i) \\ \sum_{i=1}^{n_1} \log\{(1-\alpha)\,w_1(y_i)+\alpha\,w_2(y_i)\} \end{aligned} \quad (30)$$

where, $A_i = \begin{cases} 1, & \text{if } y_i = 0 \\ 0, & \text{otherwise} \end{cases}$, $n_0 = \sum_{i=1}^{n} I_0$, $n_1 = n - n_0$ and

$$w_j(y_i) = (1-\theta^j)\,\theta^{j(y_i-1)}(\theta^j(1-e^{\boldsymbol{x'_i\beta}}+h(\theta,\alpha)) + (e^{\boldsymbol{x'_i\beta}} - h(\theta,\alpha))),\ j = 1, 2.$$

We use numerical methods directly to maximize $l(\boldsymbol{\delta})$ and obtain the maximum likelihood estimates of $\boldsymbol{\beta}, q$ and α.

9. Applications

In this section, we compare the fitted values of the BerTG distribution with those of several well-known discrete distributions commonly employed to model over-dispersed count data, such as the geometric (Geo) distribution, the generalized geometric (GGeo) distribution, the Poisson (Poi) distribution, the negative binomial (NB) distribution, the discrete Burr type II (DBX-II) distribution (Para and Jan, 2016), the discrete Weibull (DW) distribution, the discrete inverse Weibull (DIW) distribution (Jazi et al., 2010), the discrete inverted Nadarajah-Haghighi (DINH) distribution (Singh et al., 2022). The fitted models are compared using some criteria, namely, negative log-likelihood ($-ll$) and Chi-square (χ^2) test with its corresponding P-value.

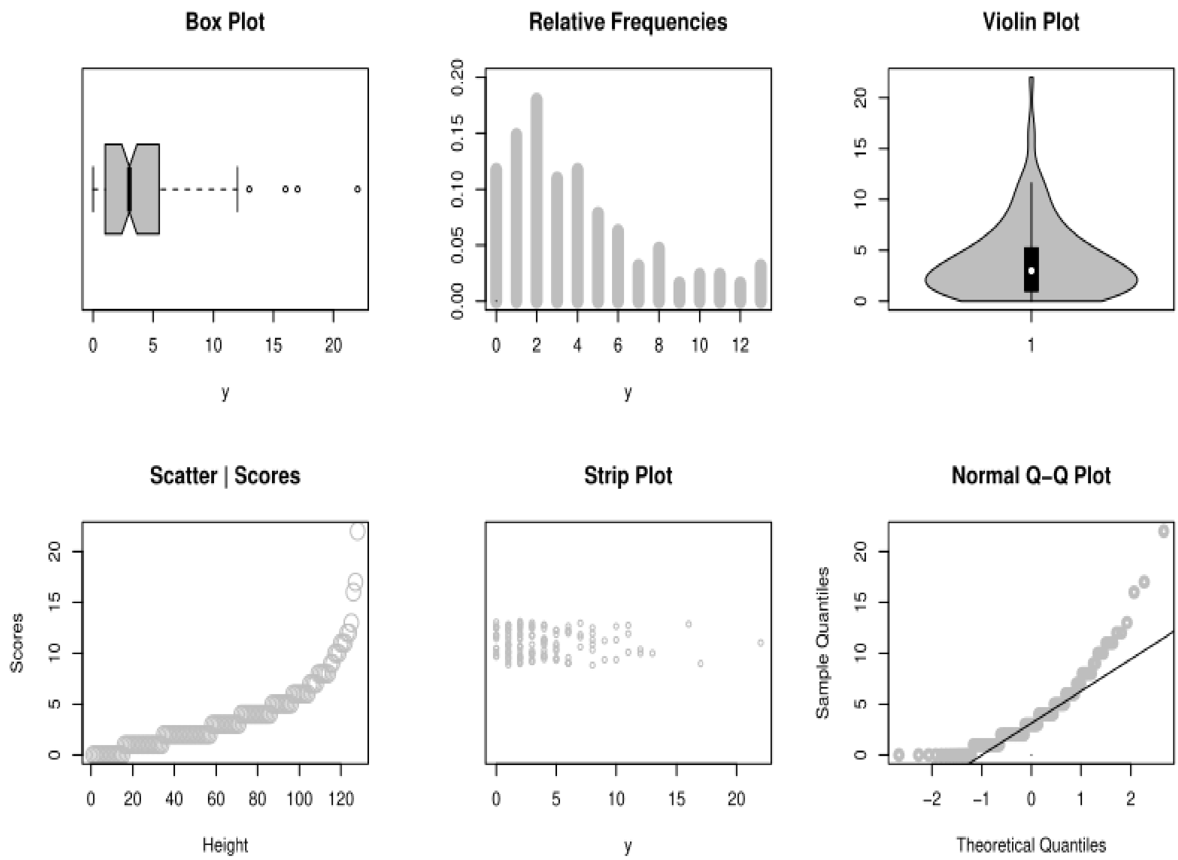

Figure 7: Non-parametric plots for data set I.

9.1. Data set I: Computer break down

Below is the provided data, which represents the number of computer breakdowns over 128 consecutive weeks of operation (Chesneau et al., 2022): 4, 0, 0, 0, 3, 2, 0, 0, 6, 7, 6, 2, 1, 11, 6, 1, 2, 1, 1, 2, 0, 2, 2, 1, 0, 12, 8, 4, 5, 0, 5, 4, 1, 0, 8, 2, 5, 2, 1, 12, 8, 9, 10, 17, 2, 3, 4, 8, 1, 2, 5, 1, 2, 2, 3, 1, 2, 0, 2, 1, 6, 3, 3, 6, 11, 10, 4, 3, 0, 2, 4, 2, 1, 5, 3, 3, 2, 5, 3, 4, 1, 3, 6, 4, 4, 5, 2, 10, 4, 1, 5, 6, 9, 7, 3, 1, 3, 0, 2, 2, 1, 4, 2, 13, 0, 2, 1, 1, 0, 3, 16, 22, 5, 1, 2, 4, 7, 8, 6, 11, 3, 0, 4, 7, 8, 4, 4, 5.

Non-parametric plots can be used for the initial visualization of this dataset, and the corresponding results can be depicted in Figure 7 . It is apparent that the data displays an asymmetric distribution, characterized by the presence of noticeable outliers. The MLEs with their corresponding standard error (SE) for the parameter(s) and goodness of fit test for data set I are listed in Table 2.

Table 2: The goodness of fit test for data set I.

X	OF	BerTG	GGeo	Geo	Poi	NB	DB-XII	DW	DIW	DINH
0	**15**	14.99	18.72	25.52	2.31	16.31	19.89	16.99	9.69	12.40
1	**19**	20.69	18.43	20.43	9.27	19.46	36.59	19.50	33.09	30.16
2	**23**	19.07	17.22	16.36	18.61	18.45	19.05	17.96	23.10	19.93
3	**14**	16.23	15.33	13.09	24.91	15.99	10.69	15.50	14.48	12.78
4	**15**	13.20	13.06	10.49	25.01	13.19	6.83	12.89	9.54	8.72
5	**10**	10.44	10.72	8.39	20.08	10.57	4.76	10.46	6.64	6.29
6	**8**	8.12	8.54	6.72	13.44	8.29	3.52	8.32	4.84	4.73
7	**4**	6.23	6.64	5.38	7.71	6.41	2.72	6.51	3.66	3.69
8	**6**	4.74	5.06	4.31	3.87	4.89	2.17	5.03	2.85	2.95
9	**2**	3.58	3.80	3.45	1.73	3.71	1.77	3.84	2.27	2.41
10	**3**	2.69	2.83	2.76	0.69	2.79	1.48	2.90	1.85	2.01
11	**3**	2.03	2.08	2.21	0.25	2.08	1.25	2.18	1.53	1.69
12	**2**	1.52	1.53	1.77	0.08	1.55	1.08	1.62	1.29	1.46
+13	**4**	4.47	4.04	7.12	0.04	4.31	16.20	4.30	13.17	18.78
Total	**128**	**128**	**128**	**128**	**128**	**128**	**128**	**128**	**128**	**128**
$-l$		316.201	317.247	320.703	384.974	316.420	342.581	316.725	330.446	331.931
MLE$_p$		0.270	0.721	0.8006	4.0156	0.703	0.784	0.867	0.076	1.244
MLE$_\theta$		0.744	2.258	-	—	1.698	3.309	1.236	1.235	1.633
MLE$_\alpha$		−0.502	—	-	—	—	—	—	—	—
χ^2		1.983	5.046	11.651	88.995	2.873	40.182	3.887	18.709	20.199
d.f		6	8	9	6	7	5	8	6	6
P.value		0.921	0.753	0.234	0	0.897	0	0.867	0.005	0.003

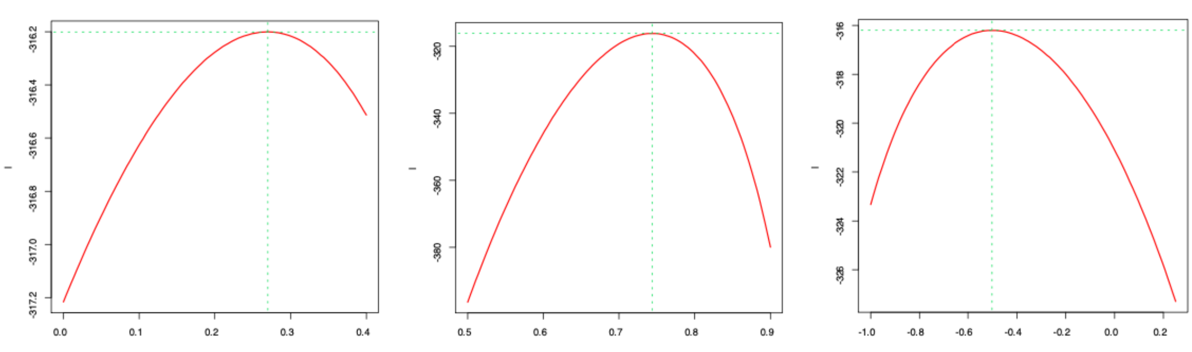

Figure 8: The profile plots of log-likelihood functions for data set I.

Among all the models considered, the BerTG model provides the best fit. Figure 8 displays the profile log-likelihood plots for Data Set I, indicating that the parameter functions are unimodal. Figure 9 presents

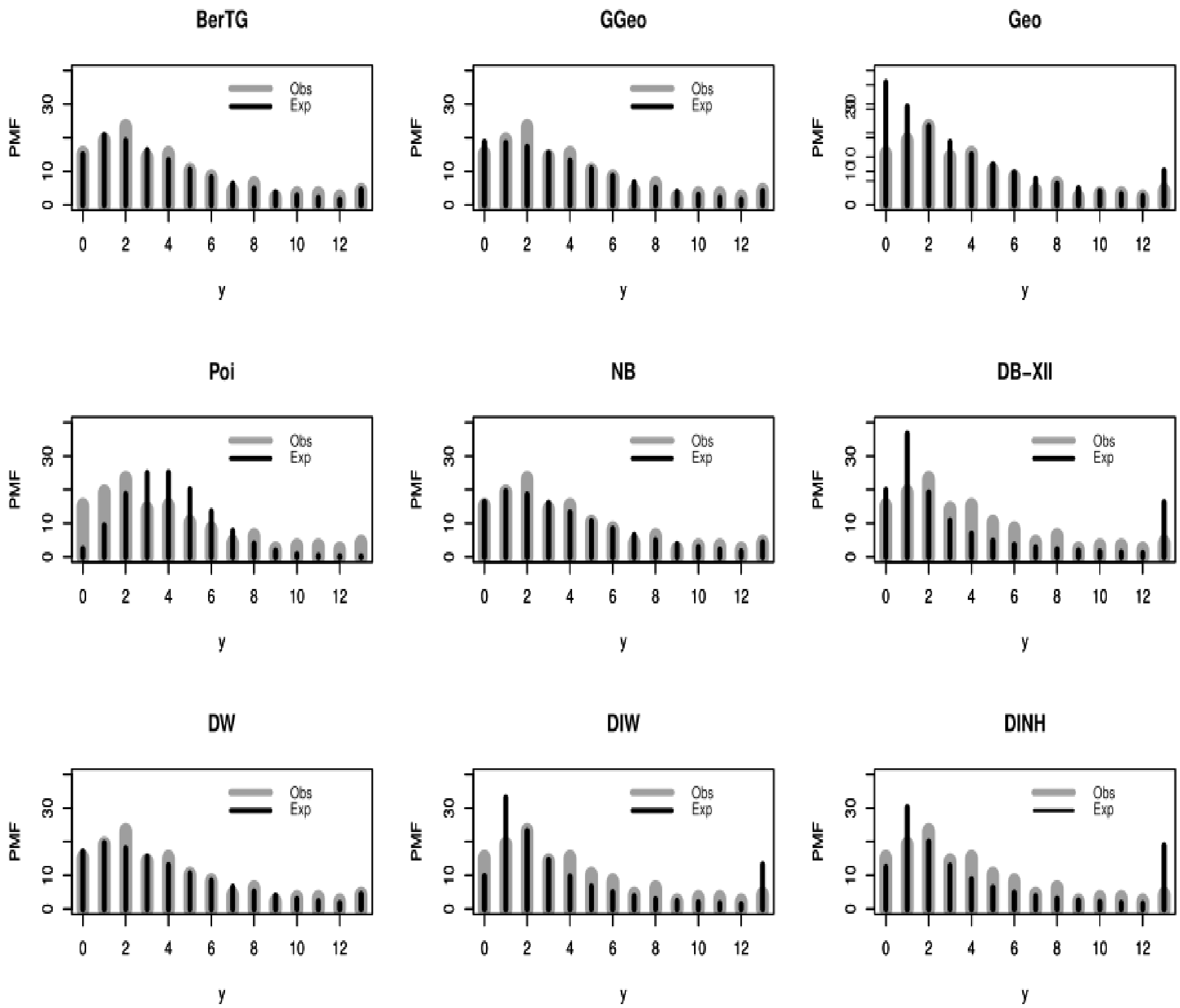

Figure 9: The estimated PMFs for data set I

the estimated PMF for Data Set I, while Table 3 summarizes key information for Data Set I based on the BerTG model.

Table 3: Some statistics for data set I.

Type ↓ Measures →	Mean	Variance	Standard Deviation	DI
Theoretical	2.33970	16.43992	4.05461	7.02651
Empirical	4.54867	13.99984	3.74164	3.07779

From Table 3, the empirical mean, variance, DI are closed to theoretical ones. Data set I suffering from overdispersion phenomena.

9.2. Data set II: Automobile insurance claims

Dataset II corresponds to the number of automobile insurance claims per policy in a car portfolio over a fixed period. In their case study, ossiaux and Lemaire (Gossiaux and Lemaire, 1981) conducted an analysis of this dataset, focusing on methods for adjusting claims distributions. The dispersion index for Dataset II is 1.16, indicating a slight over-dispersion in nature. To visualize this dataset initially, non-parametric plots can be employed, and the resulting outcomes can be presented in Figure 10 . The data exhibits an evident asymmetrical distribution with notable outliers.

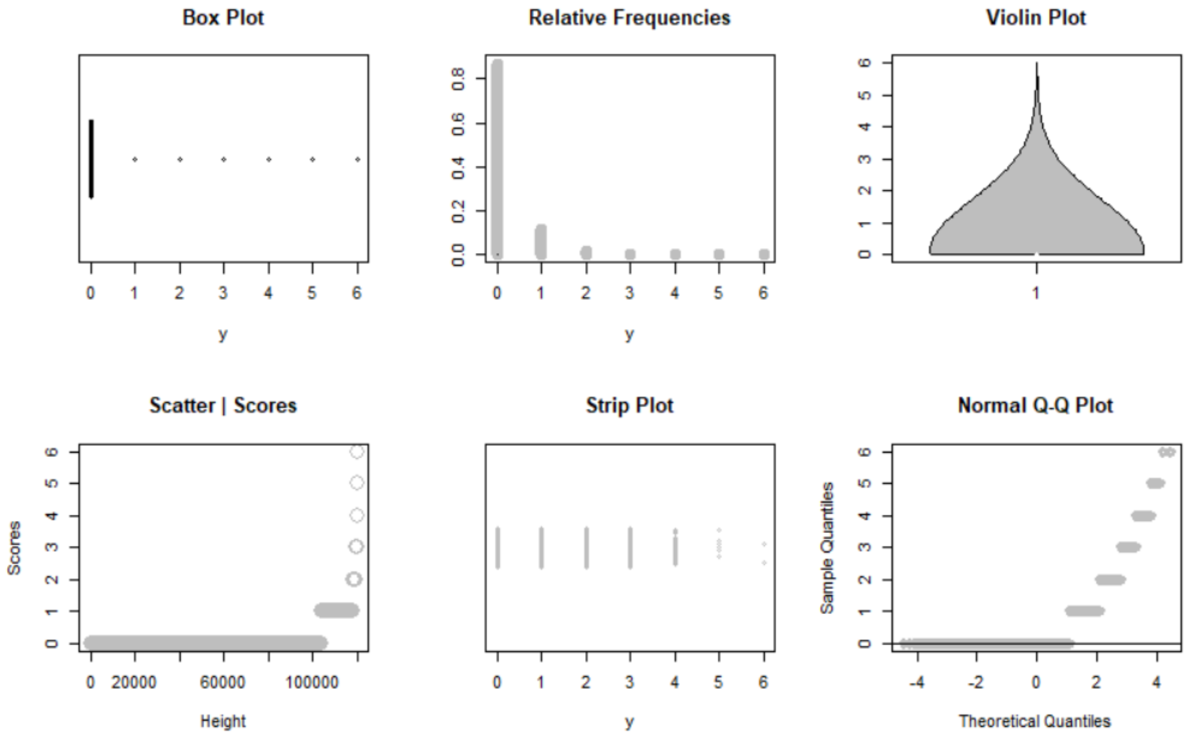

Figure 10: Non-parametric plots for data set II.

The MLEs with their corresponding SE and goodness of fit test for data set II are listed in Table 4. Among all the tested models, the BerTG model stands out as the superior distribution. The log-likelihood function plots for data set II are depicted in Figure 11, revealing that the parameters exhibit unimodal behavior. From Table 5, the empirical mean, variance, DI are closed to theoretical ones.

Table 4: The goodness of fit test results for data set II.

X	OF	BerTG	GGeo	Geo	Poi	NB	DB-XII	DW	DIW	DINH
0	103704	103700.45	104091.27	103757.11	102629.53	103724.35	103696.36	103726.01	103690.18	103806.68
1	14075	14084.53	13837.14	13934.26	15921.98	13989.19	14220.67	13985.52	14325.83	7970.05
2	1766	1748.93	1696.29	1871.33	1235.07	1857.06	1540.03	1858.53	1338.23	2682.08
3	255	268.08	201.73	251.31	63.87	245.21	279.79	245.68	301.53	1345.05
4	45	42.80	23.52	33.75	2.48	32.29	72.99	32.37	101.24	808.21
5	6	6.88	2.70	4.53	0.08	4.25	24.28	4.25	42.70	539.28
6	2	1.33	0.35	0.71	0.00	0.65	18.88	0.64	53.29	2701.65
Total	119853	119853	119853	119853	119853	119853	119853	119853	119853	119853
$-l$		54609.87	54614.96	54615.61	55108.45	54615.31	54656.4	54615.36	54768.18	61181.3
MLE$_p$		0.043	0.1315	0.134	0.155	0.1306	0.0555	0.135	0.865	7.860
MLE$_\theta$		0.161	1.026	–	–	1.032	1.6628	1.005	3.229	0.017
MLE$_\alpha$		0.483	–	–	–	–	–	–	–	–
χ^2		0.929	48.752	12.635	1332.402	12.107	76.434	12.187	260.462	9818.410
d.f		2	2	4	2	2	4	2	4	4
P.value		0.628	0	0.013	0	0.002	0	0.002	0	0

Figure 11: The profile plots of log-likelihood functions for data set II.

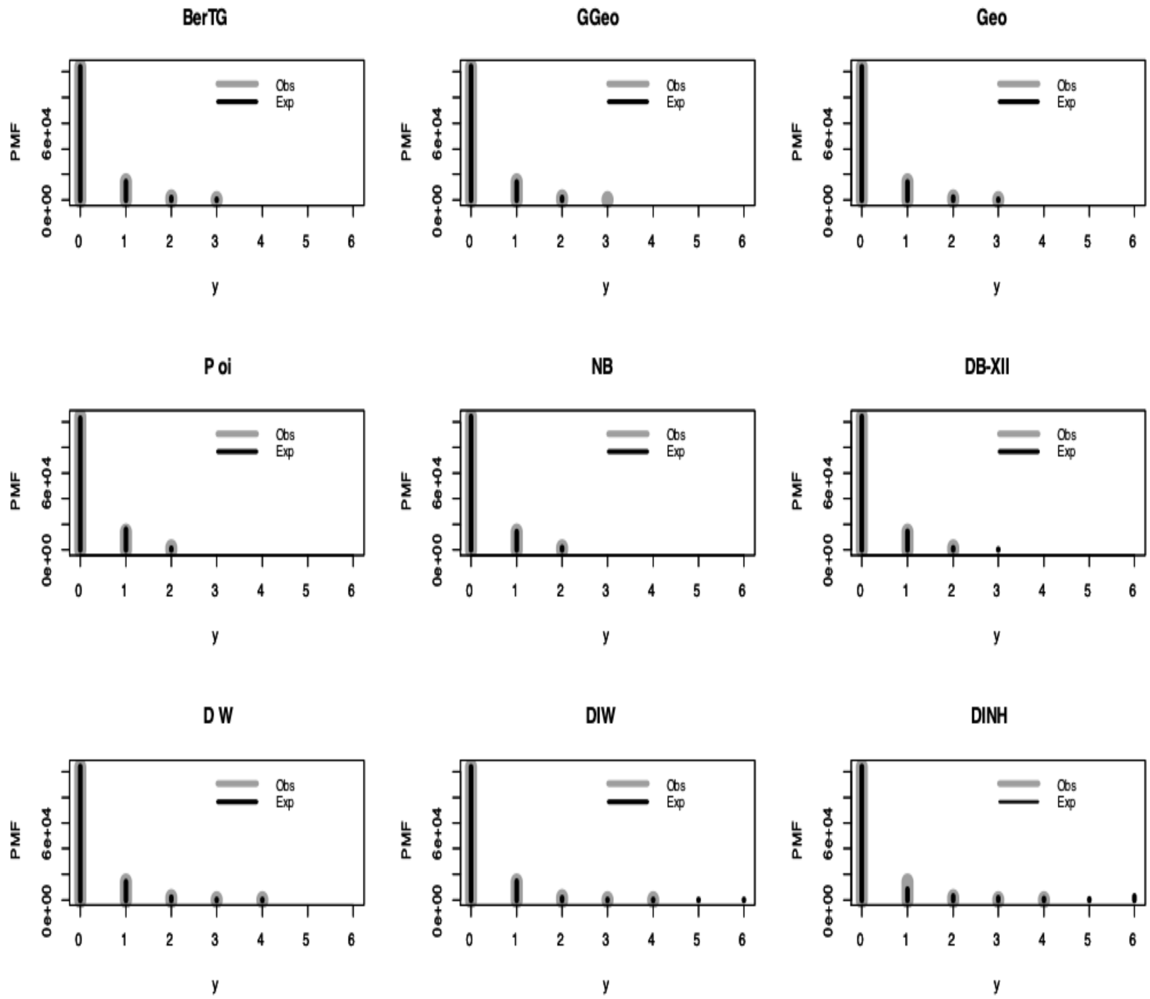

Figure 12: The estimated PMFs for data set II

Figure 12 presents the estimated PMF for Data Set II. Table 5 provides summary information for Data Set I based on the BerTG model.

Table 5: Some statistics for data set II.

Type ↓ Measures →	**Mean**	**Variance**	**DI**
Theoretical	0.15506	0.98673	6.36354
Empirical	0.15514	0.17932	1.15583

10. Discussion

In this paper, we have introduced the BerTG distribution, conducting a comprehensive examination and applying it to real-life datasets. The straightforward structural properties of the proposed distribution are expected to be valuable for practitioners. From the application point of view, the proposed model is easy to use for modeling both over-dispersed and under-dispersed count data. The maximum likelihood estimation method is employed to estimate the unknown parameters, and simulation experiments indicate favourable results for the maximum likelihood estimators. Results from two real-life datasets demonstrate that the BerTG distribution exhibits a superior fit compared to other popular count models. In two real applications, notably, the proposed model is found to be more efficient than the well-known COM-Poisson distribution for both over-dispersed and under-dispersed count responses. Furthermore, we have introduced the BerTG regression model using the generalized linear model approach. The application of such flexible discrete distributions in the analysis of time-series data is highly valuable, and further research in this aspect of the proposed model is warranted.

Declarations

The authors declare no conflict of interest.

Acknowledgement

This project was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2026/R/1447).

References

E Altun, H Alqifari, and M S Eliwa. A novel approach for zero-inflated zount regression model: Zero-inflated poisson generalized-lindley linear model with applications. *AIMS Mathematics*, 8(10):23272–23290, 2023.

G E Bardwell and E L Crow. A two parameter family of hyper-Poisson distributions. *Journal of the American Statistical Association*, 59:133–141, 1964.

M Bourguignon and C H Weiß. An INAR (1) process for modeling count time series with equidispersion, underdispersion and overdispersion. *Test*, 26(4):847–868, 2017.

M Bourguignon, D I Gallardo, and R M R Medeiros. A simple and useful regression model for underdispersed count data based on Bernoulli–Poisson convolution. *Statistical Papers*, 63(3):821–848, 2022.

S Chakraborty. On some distributional properties of the family of weighted generalized poisson distribution. *Communications in Statistics—Theory and Methods*, 39(15):2767–2788, 2010.

S Chakraborty and D Bhati. Transmuted geometric distribution with applications in modeling and regression analysis of count data. *SORT-Statistics and Operations Research Transactions*, 40(1):153–176, 2016.

S Chakraborty and R D Gupta. Exponentiated geometric distribution: another generalization of geometric distribution. *Communications in Statistics-Theory and Methods*, 44(6):1143–1157, 2015.

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

J Del Castillo and M Pérez-Casany. Weighted poisson distributions for overdispersion and underdispersion situations. *Annals of the Institute of Statistical Mathematics*, 50(3):567–585, 1998.

B Efron. Double exponential-families and their use in generalized linear-regression. *Journal of the American Statistical Association*, 81:709–721, 1986.

M El-Dawoody, M S Eliwa, and M El-Morshedy. An extension of the poisson distribution: Features and application for medical data modeling. *Processes*, 11(4):1195, 2023.

R A Fisher, A S Corbet, and C B Williams. The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, 12(1):42–58, 1943.

A M Gossiaux and J Lemaire. Méthodes d'ajustement de distributions de sinistres. *Bulletin of the Association of Swiss Actuaries*, 81:87–95, 1981.

E. Gómez-Déniz. Another generalization of the geometric distribution. *Test*, 19(2):399–415, 2010.

P G Hoel. On indices of dispersion. *The Annals of Mathematical Statistics*, 14(2):155–162, 1943.

G C. Jain and P C. Consul. A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics*, 21(4):501–513, 1971. doi: 10.1137/0121056.

M A Jazi, C D Lai, and M H Alamatsaz. A discrete inverse weibull distribution and estimation of its parameters. *Statistical Methodology*, 7(2):121–132, 2010.

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

A Nandi, A Biswas, P J Hazarika, and J Das. A new regression model for over-dispersed count responses based on poisson and geometric convolution. *Journal of the Indian Society for Probability and Statistics*, 26:1–18, 2024a.

A Nandi, P J Hazarika, A Biswas, and G G Hamedani. A new three-parameter discrete distribution to model over-dispersed count data. *Pakistan Journal of Statistics and Operation Research*, 20(2):197–215, 2024b.

A Nandi, A Biswas, P J Hazarika, and G G Hamedani. A flexible count data model based on bernoulli-poisson-geometric convolution. *Communications in Statistics - Theory and Methods*, pages 1–26, 2025. doi: 10.1080/03610926.2025.2464081.

B A Para and T R Jan. On discrete three-parameter burr type xii and discrete lomax distributions and their applications to model count data from medical science. *Biometrics and Biostatistics International Journal*, 4(2):1–15, 2016.

W T Shaw and I R Buckley. The alchemy of probability distributions: Beyond gram-charlier & cornish-fisher expansions, and skew-normal or kurtotic-normal distributions. *Submitted, Feb*, 7:64, 2007.

G Shmueli, T P Minka, J B Kadane, S Borle, and P Boatwright. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142, 2005.

B Singh, R P Singh, A S Nayal, and A Tyagi. Discrete inverted nadarajah-haghighi distribution: Properties and classical estimation with application to complete and censored data. *Statistics, Optimization & Information Computing*, 10(4):1293–1313, 2022.

R C Tripathi, R C Gupta, and T J White. Some generalizations of the geometric distribution. *Sankhya Ser. B*, 49(3):218–223, 1987.