STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 2024, pp 0–21.
Published online in International Academic Press (www.IAPress.org)

On a two-parameter weighted geometric distribution: properties, computation and applications

Mohammed K. Shakhatreh [1,*], Hazem Al-Mofleh [2,3]

[1] *Department of Mathematics and Statistics, Jordan University of Science and Technology, Irbid, Jordan*
[2] *Department of Mathematical Sciences, Ball State University, Muncie, IN 47306, USA*
[3] *Department of Mathematics, Tafila Technical University, Al-Tafila, Jordan*

Abstract In this article, we introduce a flexible two-parameter weighted geometric distribution characterized by its appealing properties. A key feature of this distribution is its ability to accommodate a wide range of skewness, making it particularly suitable for modeling right-skewed data. The distribution also exhibits a unimodal shape and an increasing failure rate. Several statistical measures are derived, and the distribution parameters are estimated using the method of maximum likelihood. The accuracy of these estimates is evaluated through a Monte Carlo simulation study. To demonstrate the flexibility, versatility, and practical importance of the proposed model, we analyze three real count data sets, showing its superiority over several existing models.

Keywords Geometric distribution, Infinite divisibility, Maximum likelihood estimates, Monte Carlo simulations, Weighted distribution

AMS 2010 subject classifications 60E05, 60E07, 62F10 , 62F12

DOI: 10.19139/soic-2310-5070-2139

1. Introduction

Discrete data frequently arise in various contexts, such as infant mortality rates in specific regions, car accidents within a certain time frame, disease recovery cases, and insurance claim liabilities. Accurately modeling such data requires selecting an appropriate discrete probability mass function. Among the various discrete probability distributions, the Poisson distribution is commonly used. However, a significant limitation of the Poisson distribution is that its mean equals its variance, which can be restrictive in many scenarios. In addition, it is suitable only when events occur independently at a constant rate. Often, count data exhibit an overdispersion, where the variance exceeds the mean. In such cases, the negative binomial distribution is typically employed to model overdispersed data, but it only performs well when the probability of success, denoted as p, is close to unity. Moreover, zero-inflated data, which frequently occur across various fields of real-world applications, present another challenge. The negative binomial distribution suffers from the need to simultaneously account for both overdispersion and zero inflation accurately, as demonstrated in the analysis of three real datasets reported in this paper.

To address these challenges, other discrete distributions, such as the zero-inflated Poisson [1], zero-inflated negative binomial [2], and zero-inflated Bell [3] distributions, have been considered. Although these distributions, particularly the latter two, can effectively analyze both excess zeros and overdispersion, they have been shown

*Correspondence to: Mohammed K. Shakhatreh (Email: mkshakhatreh6@just.edu.jo). Department of Mathematics and Statistics, Faculty of Science, Jordan University of Science and Technology, Irbid.22001, Jordan.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

to perform inaccurately in certain cases, such as when analyzing the number of mothers receiving antenatal care visits, as noted in [4].

Based on these issues, many authors have been motivated to define new discrete probability distributions to overcome the deficiencies that appear in some classical discrete probability models. Notable among the methods of constructing such models is the discretizing of a continuous probability distribution. In this context, we refer the reader, but not limited to, [5, 6, 7, 8, 9, 10], among others. Moreover, another method is by mixing continuous and discrete distributions [11, 12] or by mixing discrete distributions [13].

Recently, Bhati and Joshi [14] implemented Azzalini's [18] method to define a new geometric distribution. It is worth mentioning that this method was originally used to define what is called in the literature the skewed normal distribution. Later on, many authors used the idea to implement symmetric non-normal distributions, see, for example, [19], and asymmetric continuous distributions, see, for example, [20].

It is worth emphasizing that Azzalini's method depends on the form of the cumulative distribution of the baseline distribution. This implies that the new distribution will be tractable as long as the cumulative distribution is in a closed form. Notably, the geometric distribution possesses a closed-form cumulative distribution. Consequently, [14] derived a tractable weighted geometric distribution and defined as follows: a random variable X is said to have a weighted geometric distribution with shape parameters $\alpha > 0$, and $0 < \theta < 1$, denoted by WG(α, θ), if its probability mass function (PMF) is given below

$$q_X(x;\theta,\alpha) = \frac{(1-\theta)(1-\theta^{\alpha+1})}{1-\theta^{\alpha}}\theta^{x}(1-\theta^{-\alpha(x+1)}), \quad x = 0, 1, \ldots. \tag{1}$$

.

When $\alpha \to \infty$, then (1) converges to the geometric distribution and when $\alpha \to 0$, it converges to the negative binomial with $\alpha = 2$ and θ. Moreover it is a unidomal distribution. Additionally, the range of coefficient of variation assumes values in $(1/2\ , 2/\sqrt{2\theta})$ and the range of the skewness measures assume values $(1/2\sqrt{2}\ , 4/\sqrt{\theta})$ though these ranges were not mentioned in [14]. Notice that this range can be determined analogously to the one reported in proposition 8. Indeed, such measures could, in principle, provide insights into the flexibility of the model. Note that when these measures provide wider ranges, it implies greater flexibility in modeling various types of data. While the WG distribution has several good properties and can be used effectively for modeling count data, a limitation arises when the data exhibits greater skewness than the range assumed by the WG distribution. In such cases, the WG distribution may not capture the skewness in the data well. This spurs a closer look at a new distribution with a wider range of skewness to capture better the skewness that some models may lose.

A closing remark about generalizing the WG distribution using the idea of weighted distributions is found in [14] and in a parallel work by [21], which provides a three-parameter geometric distribution. Specifically, they used the weight, $\omega = (1-\theta^{\alpha x})^{\beta}$, and as a results they obtained,

$$qq_X(x;\theta,\alpha,) \propto C\ (1-\theta)\theta^{x-1}(1-\theta^{-\alpha x})^{\beta}, \quad x = 1, \ldots, \tag{2}$$

where C is the normalizing constant and $\beta > 0$. The PMF in (2) is an interesting distribution. However, it can not be considered a generalization to the one in (1) since the two probability distributions do not have the same support. The first one, i.e., (1) assumes support in $\mathbb{N}_0 = \{0,\ 1,\ 2, \ldots\}$ whereas the second one, i.e., (2) assumes values in $\mathbb{N} = \{1,\ 2, \ldots\}$. Consequently,(2) can not be used to model zero-inflated data which occurs very frequently in many different real count data. Additionally, the form of (2) is complex, particularly when β is a real positive number. Even when $\beta = m$ is an integer positive number it is difficult to study the properties and estimation smoothly.

This paper introduces a novel distribution that naturally emerges in the study of hidden truncation and selection models. The proposed distribution is notably derived from the concept of weighted distributions, making the choice of the weighting function more meaningful, in contrast to the common practice of subjectively selecting a weighting function. Additionally, these discrete weighted distributions are derived by discretizing their continuous counterparts, as demonstrated in examples like the discrete weighted Lindley distribution [15], the weighted discretized Frechet-Weibull distribution [16], the discrete weighted exponential distribution [17], and the references therein.

Our motivations for studying the proposed model are as follows: (i) to provide a flexible and interpretable extension of the geometric distribution; (ii) to derive a discrete distribution directly from a discrete parent distribution, unlike many discrete distributions that are derived from continuous counterparts; (iii) to rigorously explore its mathematical and statistical properties; (iv) to accurately analyze right-skewed data; (v) to offer significant flexibility in modeling overdispersed data and excess zeros simultaneously; and (vi) to improve upon the recently introduced weighted geometric distribution in (1).

Consequently, our main contributions in this article are as follows: Firstly, we introduce a two-parameter TWG distribution, derived through the concept of hidden and selection models. This distribution is based on a geometric series, unlike many discrete distributions that are derived from their continuous counterparts. Notably, the PMF of the TWG distribution is tractable without involving any complex quantities, and its CDF is available in closed form—an uncommon feature among discrete distributions. Secondly, we offer several interpretations of the model, emphasizing its flexibility. The proposed model can be derived hierarchically, making it both interpretable and applicable in various contexts. Additionally, we provide rigorous proofs for the majority of its statistical and mathematical properties. Thirdly, we show that the TWG distribution is particularly effective for accurately modeling right-skewed data. We also establish the existence and uniqueness of the maximum likelihood estimates (MLE) and offer an efficient simulation algorithm for the model. Finally, we validate the model's performance through the analysis of real data that exhibits right-skewness, overdispersion, and excess zeros, demonstrating that it outperforms many other models.

The rest of the paper is structured as follows. In Section 2, we introduce the new distribution, including some of its structural properties such as special sub-cases, unimodality, and infinite divisibility. The interpretation results of the proposed model are reported in Section 3. In Section 4, various statistical measures, including but not limited to the coefficient of variation and skewness along with their corresponding ranges, are derived. The asymptotic behavior of extreme values is presented in Section 5. In Section 6, the model parameters are estimated via maximum likelihood estimates. Simulation experiments are conducted to evaluate the finite sample behavior of the ML estimates is given in section 7. The assessment of the model using empirical real data is provided in Section 8. Finally, we conclude with a discussion, summary of conclusions, and suggestions for future work in Section 9.

2. Definition and Basic Properties

This section introduces the definition of the proposed distribution followed by listing some of its basic properties such as the special sub-cases of the model, unimodality, and infinite divisibility.

Definition 1
A random variable X is said to have a two-parameter weighted geometric distribution with parameters $0 < \theta < 1, \alpha > 0$, if the PMF of X is given below

$$p_X(x) := \Pr[X = x] = \frac{(1-\theta^{1+\alpha})(1-\theta^{1+2\alpha})}{(1-\theta^{\alpha})^2\,(1+\theta^{\alpha+1})}\,(1-\theta)\theta^x\left(1-\theta^{\alpha(1+x)}\right)^2, \quad x = 0, 1, 2, \dots \tag{3}$$

and 0 otherwise.

The cumulative distribution, CDF, and the hazard rate function, HRF, are given in the following

$$F_X(x;\alpha,\theta) = 1 - c(\theta,\alpha)\theta^{x+1}\left[\frac{1}{1-\theta} - 2\frac{\theta^{\alpha(x+2)}}{1-\theta^{1+\alpha}} + \frac{\theta^{2\alpha(x+2)}}{1-\theta^{1+2\alpha}}\right], \tag{4}$$

and

$$HRF_X(x;\alpha,\theta) = \frac{\left(1-\theta^{\alpha(1+x)}\right)^2}{\theta\left[\frac{1}{1-\theta} - 2\frac{\theta^{\alpha(x+2)}}{1-\theta^{1+\alpha}} + \frac{\theta^{2\alpha(x+2)}}{1-\theta^{1+2\alpha}}\right]}, \tag{5}$$

where

$$c(\theta,\alpha) = \frac{(1-\theta)(1-\theta^{1+\alpha})(1-\theta^{1+2\alpha})}{(1-\theta^{\alpha})^2\,(1+\theta^{\alpha+1})}.$$

Remark 1
The derivation of the PMF of the TWG distribution can be obtained by determining the normalizing constant in the following equation $d\sum_{x=0}^{\infty}(1-\theta)\theta^x\left(1-\theta^{\alpha(1+x)}\right)^2=1$. By expanding the term $\left(1-\theta^{\alpha(1+x)}\right)^2$ and applying the geometric series property, the normalizing constant can be identified as $d=(1-\theta^{1+\alpha})(1-\theta^{1+2\alpha})/(1-\theta^{\alpha})^2\,(1+\theta^{\alpha+1})$. Notice that the parameter α is a quality parameter that may control the effect of skewness

From now and on, we refer to a random variable X whose probability mass function (PMF) is given in Eq.(3) as the two-parameter weighted geometric random variable, and is denoted by *TWG*(α,θ). The PMF behavior of the TWG distribution for different values of parameters α and θ is shown in Figure 1. It is found that the PMF exhibits unimodal, right skewed or decreasing behavior form. Figure 2 shows the HRF of the TWG distribution for different values of parameters α and θ, it can be seen that its HRF is an increasing function behavior for all $\alpha>0$ and $0<\theta<1$ [see Proposition 3] .

2.1. Preliminary properties

In this section, we provide some basic properties of the TWG(α,θ) distribution, including its special cases, unimodality, decreasing failure rate, and infinite divisibility.

Proposition 1
Let $\mathcal{A}=\{\mathit{TWG}(\alpha,\theta):\alpha>0 \text{ and } 0<\theta<1\}$ be a family of TWG distributions then we have the following properties (1) As $\alpha\to\infty$, the TWG(α, θ) distribution became *Geo*(θ); i.e., $\lim_{\alpha\to\infty}p_X(x;\alpha,\theta)=\mathrm{Geo}(\theta)$; (2) If $\alpha\to 0$, then the TWG(α, θ) distribution reduces to the *WNB*(3, θ) with weight function $w(x,\theta)=2(x+1)/(x+2)(1+\theta)$, where *WNB*($r$,$\theta$) stands for a weighted negative distribution, i.e., $\lim_{\alpha\to 0}p_X(x;\alpha,\theta)=\frac{(1+x)^2}{1+\theta}(1-\theta)^3\theta^x$, $x=0,1,\ldots$

Proposition 2
Let X be a TWG(α,θ) random variable. Then TWG(α,θ) is strongly unimodal.

Proof
It is enough to show that $p_X^2(x)>p_X(x+1)\cdot p_X(x-1)$ for all $x\geq 1$. Consider the ratio between two consecutive probabilities,

$$R_X(x)=\frac{p_X(x)}{p_X(x-1)}=\frac{\theta\left(1-\theta^{\alpha(x+1)}\right)^2}{\left(1-\theta^{\alpha x}\right)^2},\ x=1,2,\ldots.$$

The change in the ratio between two consecutive points can be shown as

$$\Delta R_X(x)=R_X(x)-R_X(x-1)=-2\frac{\theta^{\alpha(x-1)}\left[(1-\theta^{\alpha})^2\left(1+\theta^{2\alpha x}\right)+\theta^{\alpha(x+1)}\right]}{\left(1-\theta^{\alpha(x-1)}\right)^2\left(1-\theta^{\alpha x}\right)^2}.$$

Therefore $R(y)$ is a decreasing function in y, and hence the proposed distribution is unimodal. Additionally, it implies log-concavity. □

Proposition 3
The HRF is an increasing function for all $\alpha>0$ and $0<\theta<1$.

Proof
To show that $HRF_X(x)=p_X(x)/\overline{F}_X(x)$, where $p_X(x)$ and $\overline{F}_X(x)=1-F_X(x)$ are the PMF and survival function of TWG distribution is an increasing function in x, we need to show that for considering points such that $x_1<x_2$ implies that $HRF_X(x_1)<HRF_X(x_2)$. Toward this goal, we show that for fixed $y\in\mathbb{N}$ the following ratio is an a decreasing function in x $r_X(x)=p_X(x+y)/p_X(x)=\theta^y\left(1-\theta^{\alpha(x+y+1)}\right)^2/\left(1-\theta^{\alpha(x+1)}\right)^2$. The log-ratio is

$$lr(x):=\log(r(x))=2\left[\log\left(1-\theta^{\alpha(x+y+1)}\right)-\log\left(1-\theta^{\alpha(x+1)}\right)\right]+y\log(\theta) \tag{6}$$

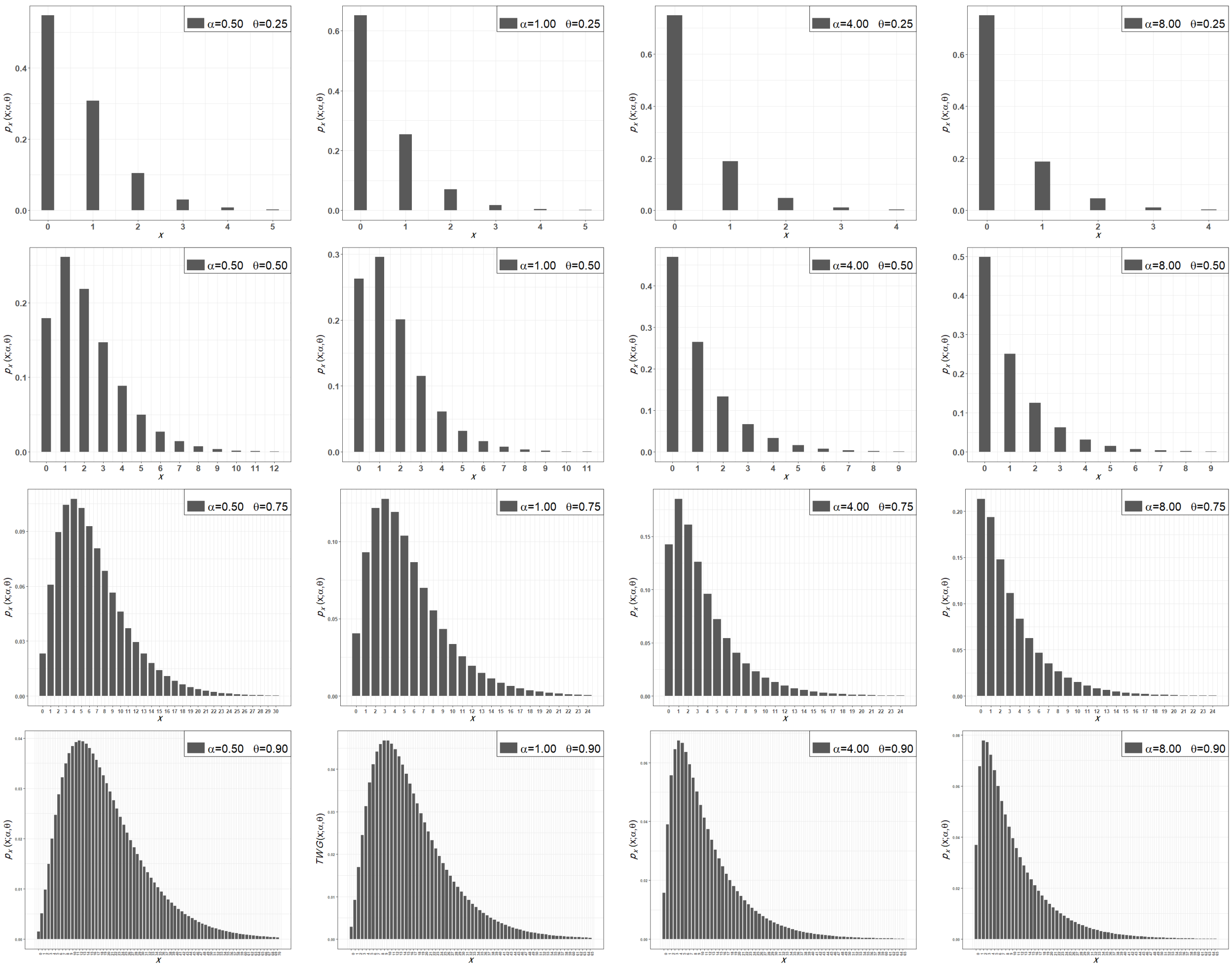

Figure 1. PMF plots of $TWG(\alpha, \theta)$ distribution for different values of parameters α and θ.

On differentiating Eq. (6) in x, we have that

$$lr^{'}(x) := \log(r_X(x)) = -2\alpha\log(\theta)\theta^{\alpha(x+1)}\left\{\frac{\theta^{\alpha y}-1}{\left(1-\theta^{\alpha(x+y+1)}\right)\left(1-\theta^{\alpha(x+1)}\right)}\right\} \tag{7}$$

Clearly, Eq. (7) is negative for all $x > 0$, i.e., $lr^{'}(x) < 0$ for $\forall x > 0$, implying that $r(x)$ is a decreasing function for $x > 0$. Hence, we have for $x_1 < x_2$, we have

$$r_X(x_1) > r_X(x_2) = \frac{p_X(x_1+y)}{p_X(x_1)} > \frac{p_X(x_2+y)}{p_X(x_2)}. \tag{8}$$

Observe that Eq. (8) can be displayed as $p_X(x_2)p_X(x_1+y) - p_X(x_1)p_X(x_2+y) > 0$, which equivalent to say that $p_X(x)$ belongs to the class of Pólya frequency of order 2 (PF2). Now, we are in a position to proof our claim, since $x_1 < x_2$, we have that

$$p_X(x_1)\overline{F}_X(x_2) - p_X(x_2)\overline{F}_X(x_1) = p_X(x_1)\sum_{y=0}^{\infty} p_X(x_2+y) - p_X(x_2)\sum_{y=0}^{\infty} p_X(x_1+y)$$
$$= \sum_{y=0}^{\infty}\left\{\underbrace{p_X(x_1)p_X(x_2+y) - p_X(x_2)p_X(x_1+y)}_{\leq 0 \text{ inview of } Eq.(8)}\right\}$$

Hence, we have that $p_X(x_1)\overline{F}_X(x_2) - p_X(x_2)\overline{F}_X(x_1) \leq 0$, implying that $p_X(x_1)/\overline{F}_X(x_1) \leq (x_2)/\overline{F}_X(x_2)$, as required. □

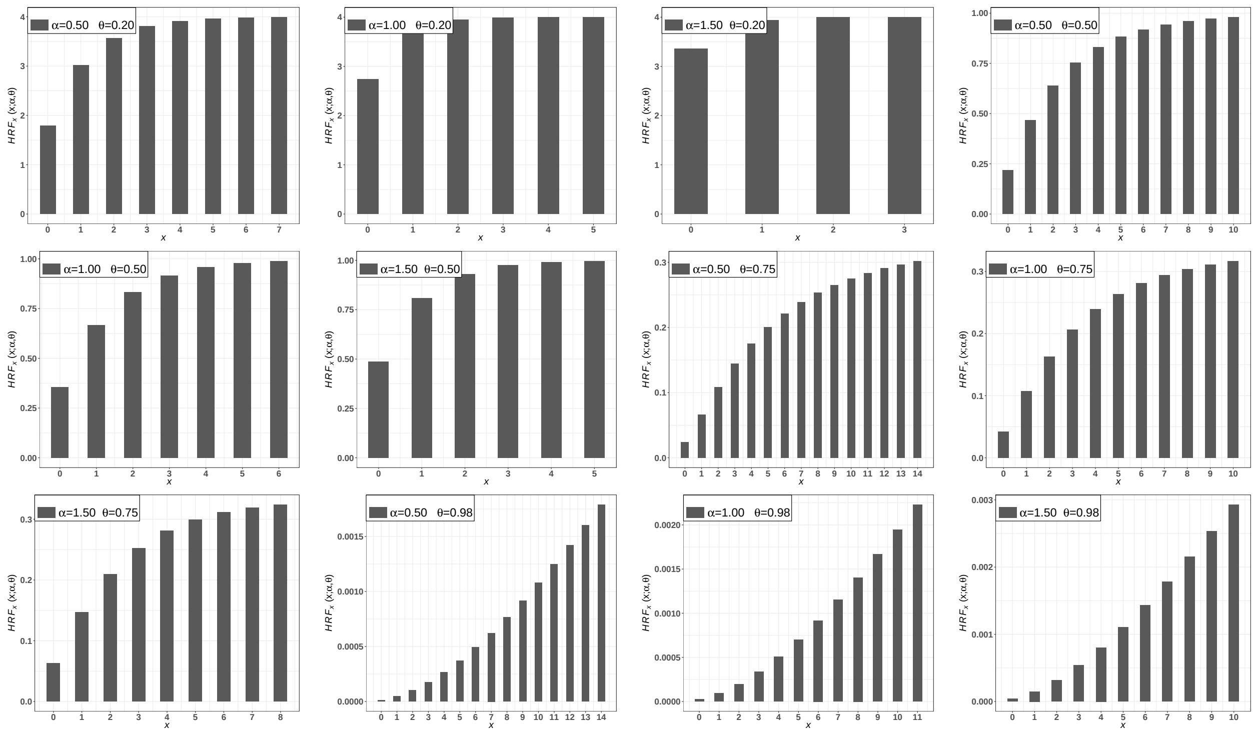

Figure 2. HRF plots of $TWG(\alpha, \theta)$ distribution for different values of parameters α and θ.

3. Important Properties

The proposed model can be explained as follows:

Proposition 4
Let X_1, X_2 and X_3 be independent and identically random variables such that each one follows the geometric distribution with parameter θ. For any positive integer $\alpha \in \mathbb{N}$, define the following random variable

$$X := X_1 \mathbb{I}\left\{X_2 \leq \alpha(X_1+1) - 1, X_3 \leq \alpha(X_1+1) - 1\right\},$$

where $\mathbb{I}(A)$ denotes the indicator function of a set A. Then the PMF of X is given via Eq.(3)

Proof
We have that $p(x_1,x_2,x_3) = (1-\theta)^3\theta^{x_1+x_2+x_3},\ x_i = 0,1,\ldots;\ i=1,2,3$. The pmf of X is

$$\begin{aligned} p_X(x) &= \Pr\left(X_1 = x \,|\, X_2 \le \alpha(X_1+1)-1, X_3 \le \alpha(X_1+1)-1\right) \\ &= \frac{\Pr\left(X_1 = x,\ X_2 \le \alpha(x+1)-1, X_3 \le \alpha(x+1)-1\right)}{\Pr\left(X_2 \le \alpha(X_1+1)-1,\ X_3 \le \alpha(X_1+1)-1\right)} \\ &= \frac{\Pr(X_1 = x)\Pr(X_2 \le \alpha(x+1)-1)\Pr(X_3 \le \alpha(x+1)-1)}{\Pr\left(X_2 \le \alpha(X_1+1)-1,\ X_3 \le \alpha(X_1+1)-1\right)} \\ &= \frac{p_{X_1}(x)F_{X_2}(\alpha(x+1)-1)F_{X_3}(\alpha(x+1)-1)}{\Pr\left(X_2 \le \alpha(X_1+1)-1,\ X_3 \le \alpha(X_1+1)-1\right)}. \end{aligned}$$

On using $p_Y(y) = (1-\theta)\theta^y$, and $F_Y(y) = 1-\theta^{(y+1)}$, it then follows that,

$$p_X(x) = \frac{(1-\theta)\theta^x(1-\theta^{\alpha(x+1))^2}}{\Pr\left(X_2 \le \alpha(X_1+1)-1,\ X_3 \le \alpha(X_1+1)-1\right)}.$$

Some algebra lead that

$$\begin{aligned} \Pr\left(X_2 \le \alpha(X_1+1)-1,\ X_3 \le \alpha(X_1+1)-1\right) &= \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\alpha(x_1+1)-1}\sum_{x_3=0}^{\alpha(x_1+1)-1}(1-\theta)^3\theta^{x_1+x_2+x_3} \\ &= \frac{(1-\theta^\alpha)^2\left(1+\theta^{\theta+1}\right)}{\left(1-\theta^{\alpha+1}\right)\left(1-\theta^{2\alpha+1}\right).} \end{aligned}$$

□

The following proposition establishes that the TWG distribution can be obtained hierarchically as follows.

Proposition 5
Suppose that V, X_1 and X_2 are random variables such that

$$X_1, X_2|v,\theta \sim^{i.i.d} \mathit{Geo}(\theta^{v+1}), \quad v \sim \mathit{Geo}(\theta).$$

If we define the random X by $X = V\mathbb{I}\left\{Y_1 \le \alpha, Y_2 \le \alpha\right\}$, it then follows that the PMF of X is given via Eq. (3).

Proof
Put $A = \left\{X_1 < \alpha, X_2 < \alpha\right\}$. Clearly, the definition of X implies that $X \overset{d}{=} (V|(X_1,X_2) \in A)$. Thus, the PMF of X is;

$$p_X(x) = \Pr(V = x \,|\, (X_1,X_2) \in A) = \frac{p_V(x)\ \Pr\left((X_1,X_2) \in A \,|\, V = x\right)}{\Pr((X_1,X_2) \in A)}.$$

Notice that $p_V(x) = (1-\theta)\theta^x$. Moreover, we have that

$$\Pr\left[(X_1,X_2) \in A \,|\, V = x\right] = \left(1-\theta^{v+1}\right)^2 \sum_{x_1=0}^{\alpha-1}\sum_{x_2=0}^{\alpha-1}\theta^{(x+1)x_1}\cdot\theta^{(x+1)x_2} = \left(1-\theta^{\alpha(x+1)}\right)^2.$$

Similarly, a computation show that

$$\Pr\left[(X_1,X_2) \in A\right] = \sum_{v=0}^{\infty}\sum_{x_1=0}^{\alpha-1}\sum_{x_2=0}^{\alpha-1}(1-\theta)\theta^v\theta^{(x+1)x_1}\cdot\theta^{(x+1)x_2} = \frac{1}{c(\alpha,\theta)},$$

and the required result follows. □

Observe that Proposition (5) has a relation to the selection models proposed in [22] but for discrete distributions. The following proposition shows that the TWG distribution can be obtained from a mixture of geometric distributions.

Proposition 6
Let $X_1,\ X_2$ and X_3 be independent geometric random variables such that $X_1 \sim \textit{Geo}(\theta)$, $X_2 \sim \textit{Geo}(\theta^{1+\alpha})$ and $X_3 \sim \textit{Geo}(\theta^{1+2\alpha})$ respectively. Let X_4 be a geometric random variable with PMF; $p_{X_4}(x) = (1-\theta)\theta^{x-1},\ x = 1, 2, \ldots,$ denoted by $Ge^{\star}(\theta)$, and assume it is independent of $X_1,\ X_2$ and X_3. If $Z = X_1 + X_2 + X_3$ and $W = X_4 + X_2 + X_3$, then Eq (3) can be obtained from

$$p_X(x;\alpha,\theta) = \delta\ p_Z(x;\alpha,\theta) + (1-\delta)\ p_W(x,\alpha,\theta) \text{ with } \delta = \frac{1}{1+\theta^{1+\alpha}},$$

and the PMF of Z

$$p_Z(x;\alpha,\theta) = c_1\ \textit{Geo}(\theta) + c_2\ \textit{Geo}(\theta^{1+\alpha}) + c_3\ \textit{Geo}(\theta^{1+2\alpha}),$$

where the constants $c_1\ c_2$ and c_3 are given below

$$c_1 = \frac{(1-\theta^{1+\alpha})(1-\theta^{1+2\alpha})}{(1-\theta^{\alpha})(1-\theta^{2\alpha})},\ c_2 = \frac{(1-\theta)(1-\theta^{1+2\alpha})}{(1-\theta^{\alpha})(1-\theta^{-\alpha})}\ \textit{and}\ c_3 = \frac{(1-\theta)(1-\theta^{1+\alpha})}{(1-\theta^{-\alpha})(1-\theta^{-2\alpha})}, \tag{9}$$

whereas the PMF of W is

$$f_W(x;\alpha,\theta) = c_1\ \textit{Geo*}(\theta) + c_2\ \textit{Geo}(\theta^{1+\alpha}) + c_3\ \textit{Geo}(\theta^{1+2\alpha}), \tag{10}$$

with $d_1 = \theta c_1,\ d_2 = c_2$ and $d_3 = c_3$.

Proof
The moment generating function of X (see Section 4) is given by

$$\begin{aligned} M_X(t) &= \frac{(1-\theta)(1-\theta^{1+\alpha})(1-\theta^{1+2\alpha})(1+\theta^{\alpha+1}\exp(t))}{(1-\theta\exp(t))(1-\theta^{1+\alpha}\exp(t))(1-\theta^{1+2\alpha}\exp(t))(1+\theta^{1+\alpha})}, \\ &= \left(\frac{1}{1+\theta^{1+\alpha}}\right) M_W(t) + \left(\frac{\theta^{1+\alpha}}{1+\theta^{1+\alpha}}\right) M_Z(t), \end{aligned}$$

where $M_Z(t)$ and $M_W(t)$ are the moment generating functions of Z and W respectively. Now the MGF of Z can be written as a partial fraction; i.e.,

$$M_Z(t) = c_1\,\frac{(1-\theta)}{1-\theta\exp(t)} + c_2\,\frac{(1-\theta^{1+\alpha})}{1-\theta^{1+\alpha}\exp(t)} + c_3\,\frac{(1-\theta^{1+2\alpha})}{(1-\theta^{2\alpha+1}\exp(t))},$$

where c_1, c_2, and c_3 are given in (9). From the uniqueness and the linearity of the inverse Laplace transform, the PMF of W is given via (10). The PMF of W, can be obtained similarly and this is complete the proof □

Interestingly, the following proposition establishes that the connection between the continuous version and discrete version. [20] introduced a two-parameter weighted exponential distributions whose its PDF is given by $f(x,\lambda,\alpha) = \frac{(1+2\alpha)(1+\alpha)}{\alpha^2}\lambda\exp(-\lambda x)\left(1-\exp(-\lambda\alpha x)\right)^2$, and referred to it as TWE(α, λ).

Proposition 7
If X:TWE(α, λ) then the discrete version of X follows X:*TWG*(α, $\exp(-\lambda)$)

Proof
It is well-known that a discrete version analogous of continuous random if its PMF is

$$p_Y(y) = \frac{f_X(y)}{\sum_{k=0}^{\infty} f_X(k)},\ y = 0, 1, \ldots.$$

Clearly,

$$p_Y(y) = \frac{\exp(-\lambda x)\,(1-\exp(-\alpha\lambda x))^2}{\sum_{k=0}^{\infty}\exp(-\lambda k)\,(1-\exp(-\alpha\lambda k))^2}, \quad y = 0, 1, \ldots$$

It is readily seen that

$$\sum_{k=0}^{\infty}\exp(-\lambda k)\,(1-\exp(-\alpha\lambda k))^2 = \frac{\exp(-\lambda)\,[1-\exp(-\alpha\lambda)]^2\,[1+\exp(-\lambda(\alpha+1))]}{[1-\exp(-\lambda)]\,[1-\exp(-\lambda(\alpha+1)]\,[1-\exp(-\lambda(2\alpha+1))]}.$$

□

4. Statistical Properties

In this section, we introduce some statistical measures including the moment generating function, moments, skewness and kurtosis. Moreover, we determine the range of skewness and kurtosis. The moment generating function (MGF) of X for $|t| < -(\alpha+1)\ln(\theta)$, is given by

$$M_X(t) = \frac{(1-\theta)(1-\theta^{\alpha+1})(1-\theta^{1+2\alpha})(1+\theta^{\alpha+1}\exp(t))}{(1-\theta\exp(t))(1-\theta^{\alpha+1}\exp(t))\,(1-\theta^{1+2\alpha}\exp(t))\,(1+\theta^{\alpha+1})}.$$

To find the mean, the variance , and the k^{th} moment we use the cummulant function . The cummulant function $(K_X(t) = \log M_X(t))$ is given by

$$\kappa_X(t) = \ln(1+\theta^{\alpha+1}\exp(t)) - \ln(1-\theta\exp(t)) - \ln(1-\theta^{\alpha+1}\exp(t)) - \ln(1-\theta^{2\alpha+1}\exp(t)) + c. \tag{11}$$

On differentiating Eq.(11) and letting $t = 0$, we have that

$$\mu = \theta\left[\frac{1}{1-\theta} + \frac{\theta^{\alpha}}{1-\theta^{\alpha+1}} + \frac{\theta^{\alpha}}{1+\theta^{\alpha+1}} + \frac{\theta^{2\alpha}}{1-\theta^{2\alpha+1}}\right], \tag{12}$$

$$Var(X) = \theta\left[\frac{1}{(1-\theta)^2} + \frac{\theta^{\alpha}}{(1-\theta^{\alpha+1})^2} + \frac{\theta^{\alpha}}{(1+\theta^{\alpha+1})^2} + \frac{\theta^{2\alpha}}{(1-\theta^{2\alpha+1})^2}\right], \tag{13}$$

$$\beta_1 = \theta\left[\frac{1+\theta}{(1-\theta)^3} + \frac{(1+\theta^{\alpha+1})\theta^{\alpha}}{(1-\theta^{\alpha+1})^3} + \frac{(1-\theta^{\alpha+1})\theta^{\alpha}}{(1+\theta^{\alpha+1})^3} + \frac{(1+\theta^{2\alpha+1})\theta^{2\alpha}}{(1-\theta^{2\alpha+1})^3}\right], \tag{14}$$

Apparently, the TWG distribution can be used to model data with overdispersion since the dispersion index defined as $I_X = Var(X)/\mu$ is greater than 1/4 and hence can model data with index of dispersion equal to 1 or greater. To see this, we have that $I_X \geq (1+\theta)/\{2(2+\theta)(1-\theta)\} > 1/4$, since $1+\theta > 1$ and that $1/(2+\theta)(1-\theta)$ is an increasing function in θ.

Proposition 8
If $X : TWG(\alpha, \theta)$, then we have the following facts: (1) The range of the coefficient of variation (c.v.) is, $(1/3,\, 3/\sqrt{3\theta})$; (2) the range of the of the skewness measure (β_1), is $(1/24\sqrt{3},\, 13/\sqrt{\theta})$.

Proof
(1)observe that the mean and the variance are decreasing functions in α. To see this, on differentiating the mean in (12) with respect to α, we have $d\mu/d\alpha = \theta^{\alpha}\ln(\theta)\{(1-\theta^{\alpha+1})^{-2} + (1+\theta^{\alpha+1})^{-2} + 2\theta^{\alpha}(1-\theta^{\alpha+1})^{-2}\}$, implying that $d\mu/d\alpha < 0$ since $\ln(\theta) < 0$. Therefore, we have that $\mu \geq \lim_{\alpha\to\infty}\theta\{(1-\theta)^{-1} + \theta^{\alpha}(1-\theta^{\alpha+1})^{-1} + \theta^{\alpha}(1+\theta^{\alpha+1})^{-1} + \theta^{2\alpha}(1-\theta^{2\alpha+1})^{-1}\} = \theta/(1-\theta)$. On the other hand, we have that $\mu \leq \lim_{\alpha\to 0}\theta\{(1-\theta)^{-1} + \theta^{\alpha}(1-\theta^{\alpha+1})^{-1} + \theta^{\alpha}(1+\theta^{\alpha+1})^{-1} + \theta^{2\alpha}(1-\theta^{2\alpha+1})^{-1}\} = 2\theta(2+\theta)/(1-\theta^2)$. Thus, we have that

$$\frac{\theta}{1-\theta} < \mu < \frac{2\theta(2+\theta)}{1-\theta^2}. \tag{15}$$

Analogously to the above analysis, the variance, $\sigma^2 := Var(X)$ in (13) is a decreasing function in α. To show this, we differentiate σ^2 with respect to α, which is given by $d\sigma^2/d\alpha = (1+\theta^{\alpha+1})\theta^{2\alpha}\ln(\theta)\{(1+\theta^{\alpha+1})(1-\theta^{\alpha+1})^{-3} + (1+\theta^{\alpha+1})(1+\theta^{\alpha+1})^{-3} + 2\theta^{\alpha}(1+\theta^{2\alpha+1})(1-\theta^{\alpha+1})^{-3}\} < 0$ due to the fact that $\theta < 1$. Therefore, we have $\sigma^2 \geq \lim_{\alpha\to\infty}\theta\{(1-\theta)^{-2} + \theta^{\alpha}(1-\theta^{\alpha+1})^{-2} + \theta^{\alpha}(1+\theta^{\alpha+1})^{-2} + \theta^{2\alpha}(1-\theta^{2\alpha+1})^{-2}\} = \theta/(1-\theta)^2$. Also, $\sigma^2 \leq \lim_{\alpha\to 0}\theta\{(1-\theta)^{-2} + \theta^{\alpha}(1-\theta^{\alpha+1})^{-2} + \theta^{\alpha}(1+\theta^{\alpha+1})^{-2} + \theta^{2\alpha}(1-\theta^{2\alpha+1})^{-2}\} = (4\theta(1+\theta+\theta^2))/((1-\theta)^2(1+\theta)^2)$. Therefore, it is clear that the variance function can be bounded as shown by the following bounds

$$\frac{\theta}{(1-\theta)^2} < \sigma^2 < \frac{4\theta(1+\theta+\theta^2)}{(1-\theta)^2(1+\theta)^2} \tag{16}$$

Now the coefficient of variation is defined via $c.v := \sigma/\mu$. To obtain its lower bound, we have from (15) that $\mu \geq \theta/1-\theta$ and from (16) that $\sigma^2 < 2\sqrt{\theta(1+\theta+\theta^2}/\{(1-\theta)(1+\theta).\}$ Hence, it follows that

$$c.v < \frac{2\sqrt{1+\theta+\theta^2}}{\sqrt{\theta}(1+\theta))} \overset{\text{since } \theta<1}{<} \frac{2\sqrt{3}}{\sqrt{\theta}(1+\theta)} < \frac{3}{\sqrt{3\theta}},$$

where the last inequality followed from the fact that $1/(1+\theta)$ is a decreasing function in θ, and hence the upper bound is obtained. Similar the lower bound can be easily obtained in veiw of eqs (12)and (13). Similar to the prove of part (1), the third centered measure μ_3 is a decreasing function in α. Therefore, after some algebra , the skewness measure; $\beta_1 = \mu 3/\sigma^{3/2}$ can be bounded as given in the following inequality

$$v_1(\theta) := \frac{(1+\theta)^4}{8\,\theta^{1/2}\,(1+\theta+\theta^2)^{3/2}} \leq \beta_1 \leq v_2(\theta) := \frac{3(1+\theta)^4 + (1-\theta)^4}{\theta^{1/2}(1+\theta)^2},$$

Notice that $1/(1+\theta+\theta^2)$ is a decreasing function and on using the fact that in $(1+\theta)^4 > 1$, it then follows that the lower bound $v_1(\theta) \geq 1/24\sqrt{3}$. On the other hand, the upper bound can shown to be less than or equal to $13/\theta^{1/2}$, and hence the proof is completed. □

Remark 2
Proposition 8 shows that the TWG distribution has a wider range of coefficient of variation and skewness compared to the WG model. Specifically, the coefficient of variation (c.v.) for the WG model from [14] is is $(1/2, 2/\sqrt{2\theta})$ whereas for the TWG it is $(1/3, 3/\sqrt{3\theta})$ indicating a wider for the TWG model. Similarly, the range of skewness is $(1/24\sqrt{3},\ 13/\sqrt{\theta})$, which is wider than the corresponding range for the WG distribution, $(1/\sqrt{2},\ 4/\sqrt{\theta})$. This demonstrates that the TWG model is more flexible in modeling count data compared to the WG distribution.

5. Large behavior of extreme values

Let $X_1, \ldots, X_n$ be an independent and identically distributed (i.i.d) random variables from TWG(α,θ). Certainly, the central limit theorem asserts that as $n \to \infty$, the sample mean $\overline{X}$ converges to a normal distribution with mean μ and variance σ^2/n, where μ and σ^2 are the mean and variance of TWG, respectively. On the other hand, understanding the asymptotic behavior of extreme statistics, particularly the maximum value, is important due to its wide-ranging applications across disciplines. If we denote $M_n := \max(X_1, \ldots, X_n)$, as the maximum value, we seek deterministic sequences of real numbers $(a_n)_{n\geq 1}$ and positive real numbers $(b_n)_{n\geq 1}$ such that $W_n := b_n\,(\max(X_1, \ldots, X_n) - a_n)$, in distribution to some probability distribution. According to Theorem 1.7.13 of [23] admits a non-degenerate probability distribution if and only if $1 - F(x)/(1 - F(x-1)) \to 1, \quad n \to \infty$. The following theorem establishes that W_n using the TWG distribution admits a non-degenerate probability distribution.

Theorem 5.1
Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d from TWG distribution. Then there is no non-degenerate limiting distribution for W_n.

Proof
$\lim_{n\to\infty}(1-F(n))/(1-F(n-1)) = \lim_{n\to\infty}\theta^{n+1}[(1-\theta)^{-1} - 2\theta^{\alpha(n+2)}(1-\theta^{1+\alpha})^{-1} + \theta^{2\alpha(n+2)}(1-\theta^{1+2\alpha})^{-1}]/\theta^{n}[(1-\theta)^{-1} - 2\theta^{\alpha(n+1)}(1-\theta^{1+\alpha})^{-1} + \theta^{2\alpha(n+1)}(1-\theta^{1+2\alpha})^{-1}] = \theta \neq 1.$ □

Remark 3
The geometric and the weighted negative binomial distributions (specifically, WNBD(3, θ)) are sub-special cases derived from the TWG distribution. Consequently, W_n corresponding to these distributions does not converge to any non-degenerate limiting distribution

6. Maximum likelihood estimates

Estimating the unknown parameters of the proposed distribution is performed via the maximum likelihood estimates (MLEs) method. We discuss the MLEs thoroughly, particularly their existence, uniqueness, and asymptotic distributions. In order to provide reliable estimates of model parameters, mainly in terms of calculations, we re-parametrize the model parameters by letting $\lambda = \theta^{\alpha}$. The log-likelihood of λ and θ using a set of observations of size n, say; $x^n := (x_1, \ldots, x_n)$ is

$$\begin{aligned}\ell(\theta,\lambda) &= n\left[\ln(1-\theta) + \ln(1-\lambda\theta) + \ln(1-\lambda^2\theta) - 2\ln(1-\lambda) - \ln(1+\lambda\theta)\right] \\ &\quad + \ln(\theta)\sum_{i=1}^{n} x_i + 2\sum_{i=1}^{n}\ln(1-\lambda^{1+x_i}). \end{aligned} \tag{17}$$

The score function is given by $\mathbf{U} = (\ell(\theta,\lambda)/\partial\theta, \ell(\theta,\lambda)/\partial\lambda)^t$, where

$$\frac{\ell(\theta,\lambda)}{\partial\theta} = -n\left[\frac{1}{1-\theta} + \frac{2\lambda}{1-\lambda^2\theta^2} + \frac{\lambda^2}{1-\lambda^2\theta}\right] + \frac{\sum_{i=1}^{n} x_i}{\theta}, \tag{18}$$

$$\frac{\ell(\theta,\lambda)}{\partial\lambda} = -n\left[\frac{2\theta}{1-\lambda^2\theta^2} + \frac{2\theta\lambda^2}{1-\lambda^2\theta}\right] + \frac{2n}{1-\lambda} - 2\sum_{i=1}^{n}\frac{(x_i+1)\lambda^{x_i}}{1-\lambda^{x_i+1}}. \tag{19}$$

The ML estimates; $\widehat{\lambda}$ and $\widehat{\theta}$ are obtained by solving $\mathbf{U} = \mathbf{0}$. The following theorem shows that the score function has unique solutions, and hence the existence and uniqueness of the ML estimates of the parameters can be proved.

Theorem 6.1
The right-hand-side of equation (18) has a unique solution in terms of θ, and the root, $\widehat{\theta}$, lies in

$$\left(\frac{\sum_{i=1}^{n} x_i}{4n + \sum_{i=1}^{n} x_i}, \frac{\sum_{i=1}^{n} x_i}{n + \sum_{i=1}^{n} x_i}\right). \tag{20}$$

Proof
Put $g(\theta;\lambda,x^n) := \omega(\theta,\lambda,x^n) - \frac{n}{1-\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta}$, where $\omega(\theta,\lambda,x^n) = -n[2\lambda(1-\lambda^2\theta^2)^{-1} + \lambda^2(1-\lambda^2\theta)^{-1}]$. Since $\omega(\theta,\lambda,x^n) \leq 0$, it then follows that

$$g(\theta;\lambda,x^n) \leq -\frac{n}{1-\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta},$$

and hence $g(\theta;\lambda,x^n) < 0$ if and only if $\theta > \sum_{i=1}^{n} x_i/(n + \sum_{i=1}^{n} x_i)$. On the other hand, we have that

$$g(\theta;\lambda,x^n) \geq \frac{4n}{1-\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta}.$$

It then follows that $g(\theta;\lambda,x^n) > 0$ if and only if $\theta < \sum_{i=1}^{n} x_i/(4n + \sum_{i=1}^{n} x_i)$. So there exists at least one root of $g(\theta;\lambda,x^n)$ in the interval

$$\left(\frac{\sum_{i=1}^{n} x_i}{4n + \sum_{i=1}^{n} x_i}, \frac{\sum_{i=1}^{n} x_i}{n + \sum_{i=1}^{n} x_i}\right).$$

To show the uniqueness, the first derivative of $g(\theta;\lambda,x^n)$ is $g^{'}(\theta;\lambda,x^n)=\omega^{'}(\theta,\lambda,x^n)-\frac{n}{(1-\theta)^2}-\frac{\sum_{i=1}^n x_i}{\theta^2}$. Clearly, $g^{'}(\theta;\lambda,x^n)<0$ due to $\omega^{'}(\theta,\lambda,x^n)<0$, and hence the the uniqueness is proven. □

Theorem 6.2
Let $\Lambda(\lambda;\theta,x^n)$ denote the function on the right-hand-side of equation (19), then there exists a root say, $\widehat{\lambda}$ such that $\Lambda(\widehat{\lambda};\theta,x^n)=0$.

Proof
We have that $\lim_{\lambda\to 0}\Lambda(\lambda;\theta,x^n)=n(1-\theta)>0$. On the other hand, we have that

$$\lim_{\lambda\to 1}\Lambda(\lambda;\theta,x^n)=-n\left[\frac{2\theta}{1-\theta^2}+\frac{2\theta}{1-\theta}\right]-2\lim_{\lambda\to 1}\left[\sum_{i=1}^n\frac{(1+x_i)\lambda^{x_i}}{1-\lambda^{1+x_i}}-\frac{1}{1-\lambda},\right]\le-\frac{2n\theta}{1-\theta}+\sum_{i=1}^n x_i.$$

The upper bound is negative if and only if $\theta>\frac{\sum_{i=1}^n x_i}{2n+\sum_{i=1}^n x_i}$. In view of (20), it follows that when $\theta\in\left(\frac{\sum_{i=1}^n x_i}{2n+\sum_{i=1}^n x_i},\frac{\sum_{i=1}^n x_i}{n+\sum_{i=1}^n x_i}\right)$, the ML of λ say $\widehat{\lambda}$, exists and satisfy $\Lambda(\widehat{\lambda};\theta,x^n)=0$. □

6.1. *Asymptotic distributions*

This subsection provides the asymptotic distributions of the ML estimates; $\widehat{\lambda}$ and $\widehat{\theta}$. Notice that the likelihood function satisfies all the regularities conditions required to establish the asymptotic normality distribution of $\widehat{\lambda}$ and $\widehat{\theta}$. Consequently, as n goes to infinity, we have $\sqrt{n}(\widehat{\theta}-\theta,\widehat{\lambda}-\lambda)\Longrightarrow N_2(0,\mathbf{I}^{-1})$. where $\mathbf{I}$ is the Fisher information per unit observation. The asymptotic normality of the ML estimates of the original parameters say, $\widehat{\theta}$ and $\widehat{\alpha}$ is $\sqrt{n}(\widehat{\theta}-\theta,\widehat{\lambda}-\lambda)\Longrightarrow N_2(0,\mathbf{G}\,\mathbf{I}^{-1}\,\mathbf{G}^T)$, $\mathbf{G}=\begin{pmatrix}1 & 0\\ -\frac{\alpha}{\theta\ln(\theta)} & \frac{1}{\theta^{\alpha}\ln(\theta)}\end{pmatrix}$.

7. Data Generation and Simulation Study

Here, we provide an efficient algorithm for simulating random samples from the CILPS distribution given via TWG(λ,θ). Additionally, Monte Carlo simulation experiments are conducted in order to evaluate the performance of the ML estimates model parameters.

7.1. *Generation algorithm*

Following propositions 5 or 6, we can generate random samples from the TWG distribution. However, on using proposition 5 we can generate the required sample as long as α is a positive integer number. On the other hand proposition 6 can be used to generate irrespective that α is integer or not. However, we provide an efficient algorithm of generating random samples from the TWG.

Algorithm 1
This algorithm can be used for generating random samples from CILPS using the CDF given (4)

1. Slicing up the interval $[0,1]$ into sub-intervals which define a partition of $[0,1]$ as: $I_0=[0,F_X(0))$, $I_1=[F_X(0),F_X(1))$, $I_2=[F_X(1),F_X(2))$,. . ., $I_k=[F_X(x_{k-1}),F_X(x_k))$,. . ., where $F_X(.)$ is the CDF in Eq.(4) with $\lambda=\theta^{\alpha}$.
2. Generate $U_i\sim Uniform(0,1), i=1,2,\dots,n$.
3. If $U_i\in I_j$, let $Y=x_i$ for $i=1,2,\dots,n$ and $j=0,1,2,\dots$.
4. Then $\Pr[U_i\in I_j]=\Pr[F_X(x_{j-1})\le U_i<F_X(x_j)]=F_X(x_j)-F_X(x_{j-1})=p_X(x_i)=\Pr[X=x_j]$ for $i=1,2,\dots,n$ and $j=0,1,2,\dots$.
5. So, Y is the required sample.

7.2. Monte Carlo Simulation study

In this section, we analyze the performance of the estimators from four different methods for estimating the parameters of the TWG distribution through simulation experiments. The simulation steps are: (i) select the sample size n; (ii) use Algorithm 1 to generate a random sample of size n from the TWG model; (iii) calculate the TWG model parameter estimates; and (iv) repeat steps (ii) and (iii) N times. Sample sizes are $n = 30, 50, 80, 120, 200,$ and 350, with parameter combinations; $\lambda = 0.50, 0.60, 0.80, 0.90$ and $\theta = 0.40, 0.50, 0.80, 0.95$. The number of replications is $N = 1000$. We evaluate the finite sample performance of the ML estimates using the following metrics for each sample size: the average of absolute value of biases ($|Bias(\widehat{\boldsymbol{\Theta}})|$), $|Bias(\widehat{\boldsymbol{\Theta}})| = \frac{1}{N}\sum_{i=1}^{N}|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|$, the mean square error of the estimates (MSEs), $MSEs = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^2$, and the mean relative estimates (MREs), $MREs = \frac{1}{N}\sum_{i=1}^{N}|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|/\boldsymbol{\Theta}$, where $\boldsymbol{\Theta} = (\lambda, \theta)^{\intercal}$ All calculations are performed using R software (version 4.2.2)[31]. Table1 shows the Monte Carlo simulation results for , MSE, and MRE of the TWG parameters. The findings provide significant insights. The parameter estimates demonstrate strong performance, with minimal bias and acceptable MSE across all scenarios. These estimates are reliable, closely matching the true values, and indicate asymptotic unbiasedness. Moreover, the MSE trends towards zero, confirming the consistency of the estimates.This is also confirmed, as demonstrated in Figures 3 through 6. Specifically, for $n \geq 50$, the convergence of the considered measures tends to zero.

Table 1. Simulation results for various values of $\boldsymbol{\Theta} = (\lambda, \theta)^{\intercal}$.

Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.50, \theta = 0.40)^{\intercal}$						$\boldsymbol{\Theta} = (\lambda = 0.50, \theta = 0.50)^{\intercal}$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS\rvert$	$\hat{\lambda}$	0.35635	0.31591	0.27530	0.24046	0.19308	0.15172	0.32314	0.26914	0.22878	0.19473	0.14652	0.10432
	$\hat{\theta}$	0.11500	0.10016	0.08845	0.07797	0.06339	0.05032	0.11357	0.09668	0.08215	0.07011	0.05378	0.03803
MSE	$\hat{\lambda}$	0.15676	0.13173	0.10821	0.08915	0.06323	0.04304	0.13710	0.10530	0.08280	0.06373	0.03988	0.02087
	$\hat{\theta}$	0.01687	0.01337	0.01105	0.00912	0.00653	0.00445	0.01815	0.01389	0.01084	0.00842	0.00540	0.00277
MRE	$\hat{\lambda}$	0.71271	0.63182	0.55061	0.48092	0.38616	0.30343	0.64629	0.53827	0.45756	0.38945	0.29303	0.20863
	$\hat{\theta}$	0.28751	0.25039	0.22113	0.19493	0.15849	0.12579	0.22713	0.19336	0.16430	0.14022	0.10757	0.07605
Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.50, \theta = 0.80)^{\intercal}$						$\boldsymbol{\Theta} = (\lambda = 0.50, \theta = 0.95)^{\intercal}$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS\rvert$	$\hat{\lambda}$	0.20498	0.15315	0.11576	0.09422	0.07082	0.05173	0.20160	0.16379	0.13186	0.11037	0.08377	0.06413
	$\hat{\theta}$	0.05453	0.03697	0.02677	0.02050	0.01464	0.01069	0.00988	0.00683	0.00502	0.00393	0.00298	0.00227
MSE	$\hat{\lambda}$	0.06681	0.03883	0.02276	0.01448	0.00797	0.00418	0.06028	0.04248	0.02791	0.02002	0.01126	0.00659
	$\hat{\theta}$	0.00666	0.00317	0.00158	0.00081	0.00037	0.00019	0.00023	0.00009	0.00004	0.00003	0.00001	0.00001
MRE	$\hat{\lambda}$	0.40995	0.30631	0.23151	0.18845	0.14163	0.10345	0.40319	0.32757	0.26372	0.22073	0.16755	0.12827
	$\hat{\theta}$	0.06816	0.04621	0.03346	0.02562	0.01830	0.01336	0.01040	0.00718	0.00529	0.00414	0.00313	0.00239
Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.60, \theta = 0.40)^{\intercal}$						$\boldsymbol{\Theta} = (\lambda = 0.60, \theta = 0.50)^{\intercal}$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS\rvert$	$\hat{\lambda}$	0.33408	0.29775	0.26609	0.24060	0.20388	0.16220	0.30375	0.26509	0.22755	0.19867	0.16090	0.12123
	$\hat{\theta}$	0.10846	0.09959	0.08808	0.07998	0.06751	0.05473	0.11136	0.09821	0.08549	0.07448	0.06077	0.04592
MSE	$\hat{\lambda}$	0.13156	0.10910	0.09190	0.07862	0.06133	0.04373	0.11448	0.09183	0.07350	0.05917	0.04275	0.02660
	$\hat{\theta}$	0.01465	0.01245	0.01016	0.00861	0.00658	0.00481	0.01603	0.01306	0.01055	0.00848	0.00618	0.00384
MRE	$\hat{\lambda}$	0.55681	0.49625	0.44348	0.40099	0.33981	0.27033	0.50626	0.44181	0.37924	0.33111	0.26816	0.20205
	$\hat{\theta}$	0.27114	0.24897	0.22019	0.19995	0.16876	0.13683	0.22271	0.19642	0.17097	0.14895	0.12154	0.09183
Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.60, \theta = 0.80)^{\intercal}$						$\boldsymbol{\Theta} = (\lambda = 0.60, \theta = 0.95)^{\intercal}$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS\rvert$	$\hat{\lambda}$	0.19029	0.15118	0.11002	0.08771	0.06603	0.04894	0.17994	0.14114	0.10975	0.08934	0.06864	0.05181
	$\hat{\theta}$	0.05935	0.04401	0.03059	0.02326	0.01650	0.01207	0.01069	0.00705	0.00518	0.00417	0.00317	0.00232
MSE	$\hat{\lambda}$	0.05684	0.03761	0.02063	0.01321	0.00715	0.00384	0.05299	0.03426	0.02071	0.01361	0.00788	0.00435
	$\hat{\theta}$	0.00718	0.00432	0.00213	0.00117	0.00051	0.00024	0.00028	0.00010	0.00005	0.00003	0.00002	0.00001
MRE	$\hat{\lambda}$	0.31714	0.25197	0.18337	0.14618	0.11004	0.08157	0.29991	0.23524	0.18291	0.14890	0.11440	0.08635
	$\hat{\theta}$	0.07419	0.05501	0.03824	0.02907	0.02062	0.01509	0.01125	0.00742	0.00546	0.00439	0.00334	0.00244

Table 1: Simulation results for various values of $\boldsymbol{\Theta} = (\lambda, \theta)^\intercal$ (Continued).

Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.80, \theta = 0.40)^\intercal$						$\boldsymbol{\Theta} = (\lambda = 0.80, \theta = 0.50)^\intercal$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS \rvert$	$\hat{\lambda}$	0.25218	0.23295	0.21518	0.19822	0.17979	0.15838	0.22875	0.20481	0.18975	0.17226	0.15323	0.13376
	$\hat{\theta}$	0.09030	0.08288	0.07727	0.07127	0.06506	0.05711	0.09080	0.08220	0.07751	0.07146	0.06405	0.05638
MSE	$\hat{\lambda}$	0.08521	0.06941	0.05743	0.04807	0.03894	0.03062	0.06858	0.05315	0.04386	0.03640	0.02933	0.02306
	$\hat{\theta}$	0.01116	0.00902	0.00763	0.00638	0.00520	0.00404	0.01055	0.00847	0.00736	0.00630	0.00515	0.00413
MRE	$\hat{\lambda}$	0.31522	0.29119	0.26897	0.24777	0.22473	0.19798	0.28594	0.25601	0.23718	0.21532	0.19154	0.16720
	$\hat{\theta}$	0.22576	0.20721	0.19317	0.17818	0.16266	0.14278	0.18160	0.16439	0.15502	0.14293	0.12811	0.11277
Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.80, \theta = 0.80)^\intercal$						$\boldsymbol{\Theta} = (\lambda = 0.80, \theta = 0.95)^\intercal$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS \rvert$	$\hat{\lambda}$	0.15123	0.12519	0.10191	0.08449	0.06714	0.04804	0.10551	0.07598	0.05731	0.04668	0.03475	0.02624
	$\hat{\theta}$	0.06450	0.05435	0.04487	0.03614	0.02865	0.02002	0.01357	0.00881	0.00638	0.00489	0.00358	0.00268
MSE	$\hat{\lambda}$	0.03264	0.02250	0.01574	0.01147	0.00759	0.00407	0.02392	0.01230	0.00590	0.00379	0.00199	0.00111
	$\hat{\theta}$	0.00621	0.00473	0.00357	0.00254	0.00168	0.00085	0.00047	0.00019	0.00008	0.00004	0.00002	0.00001
MRE	$\hat{\lambda}$	0.18903	0.15648	0.12739	0.10561	0.08393	0.06004	0.13189	0.09498	0.07164	0.05835	0.04343	0.03280
	$\hat{\theta}$	0.08063	0.06793	0.05608	0.04517	0.03582	0.02503	0.01428	0.00928	0.00672	0.00515	0.00377	0.00282
Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.90, \theta = 0.40)^\intercal$						$\boldsymbol{\Theta} = (\lambda = 0.90, \theta = 0.50)^\intercal$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS \rvert$	$\hat{\lambda}$	0.21145	0.19393	0.17947	0.16426	0.15027	0.13413	0.18024	0.16725	0.15190	0.13989	0.12690	0.11227
	$\hat{\theta}$	0.08095	0.07361	0.06815	0.06256	0.05700	0.05172	0.07721	0.07106	0.06585	0.06122	0.05620	0.05021
MSE	$\hat{\lambda}$	0.08204	0.06429	0.05138	0.04081	0.03257	0.02459	0.05709	0.04504	0.03442	0.02788	0.02176	0.01607
	$\hat{\theta}$	0.01111	0.00902	0.00743	0.00600	0.00481	0.00373	0.00909	0.00751	0.00621	0.00517	0.00419	0.00319
MRE	$\hat{\lambda}$	0.23495	0.21548	0.19941	0.18251	0.16697	0.14903	0.20026	0.18583	0.16878	0.15543	0.14100	0.12474
	$\hat{\theta}$	0.20237	0.18402	0.17038	0.15639	0.14250	0.12930	0.15442	0.14212	0.13170	0.12243	0.11241	0.10043
Est.	Est. Par.	$\boldsymbol{\Theta} = (\lambda = 0.90, \theta = 0.80)^\intercal$						$\boldsymbol{\Theta} = (\lambda = 0.90, \theta = 0.95)^\intercal$					
		n						n					
		30	50	80	120	200	350	30	50	80	120	200	350
$\lvert BIAS \rvert$	$\hat{\lambda}$	0.10672	0.09289	0.08268	0.07305	0.06237	0.05118	0.06429	0.04658	0.03569	0.02822	0.02129	0.01566
	$\hat{\theta}$	0.05352	0.04836	0.04436	0.03956	0.03422	0.02833	0.01792	0.01285	0.00906	0.00686	0.00486	0.00340
MSE	$\hat{\lambda}$	0.01658	0.01168	0.00892	0.00708	0.00530	0.00384	0.00894	0.00397	0.00216	0.00137	0.00074	0.00039
	$\hat{\theta}$	0.00364	0.00300	0.00257	0.00213	0.00170	0.00127	0.00065	0.00038	0.00020	0.00011	0.00005	0.00002
MRE	$\hat{\lambda}$	0.11857	0.10321	0.09187	0.08117	0.06929	0.05686	0.07144	0.05176	0.03966	0.03135	0.02365	0.01741
	$\hat{\theta}$	0.06690	0.06045	0.05544	0.04945	0.04278	0.03541	0.01886	0.01352	0.00953	0.00722	0.00511	0.00357

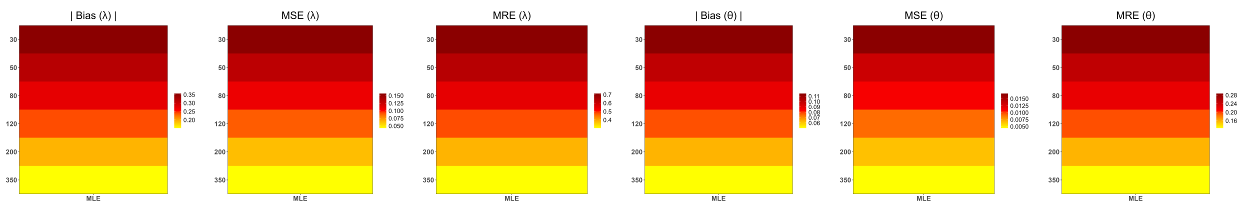

Figure 3. The heatmaps of the simulated biases, MSE and MRE of the MLE simulation method for $\boldsymbol{\Theta} = (\lambda = 0.50, \theta = 0.40)^\intercal$.

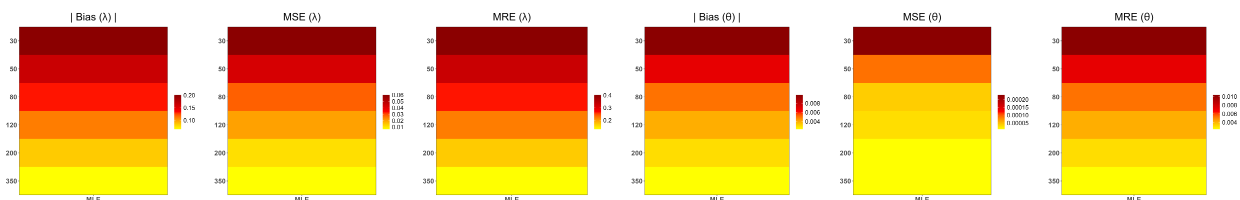

Figure 4. The heatmaps of the simulated biases, MSE and MRE of the MLE simulation method for $\boldsymbol{\Theta} = (\lambda = 0.50, \theta = 0.95)^\intercal$.

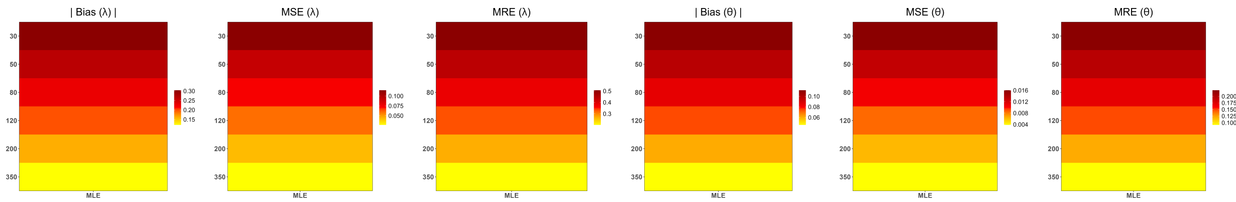

Figure 5. The heatmaps of the simulated biases, MSE and MRE of the MLE simulation method for $\boldsymbol{\Theta} = (\lambda = 0.60, \theta = 0.50)^\intercal$.

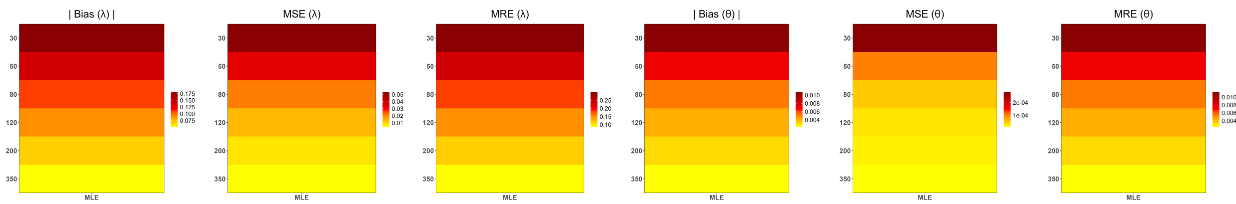

Figure 6. The heatmaps of the simulated biases, MSE and MRE of the MLE simulation method for $\boldsymbol{\Theta} = (\lambda = 0.60, \theta = 0.95)^\intercal$.

8. Data Illustration

This section demonstrates the potential of fitting the TWG model to three real data sets. Importantly, all three data sets are zero-inflated and exhibit overdispersion. Additionally, the TWG model will be compared with several competitive models, namely: the weighted geometric distribution (WG) from [14], the negative binomial (NB), the new discrete distribution (ND) from [27], the generalized geometric distribution (GGD) from [26], the general Poisson Lindley distribution (GPLD) from [30], and the new generalized Poisson Lindley distribution (NGPLD) from [25].

The competitive models are compared by using goodness-of-fit criteria and information-theoretic criteria including the maximized log-likelihood under the model ($-\widehat{\ell}$), Akaike information criterion (AIC), corrected Akaike information criterion ($CAIC$), Bayesian information criterion (BIC), Hannan-Quinn information criterion ($HQIC$) and Kolmogorov Smirnov (K-S) statistics with its corresponding p-value. Moreover, the chi-square (χ^2) test is used with its corresponding p-value where the estimated probabilities are under the null hypothesis. The model with minimum values for these statistics and the highest p-value for the chi-square (χ^2) test could be chosen as the best model to fit the data.

8.1. Data set 1: Number of ticks on 82 sheeps

As a first example, we use a dataset from [28] that represents the number of ticks counted on 82 sheep. Preliminary analysis shows that the sample mean is $\overline{x} = 6.561$, the sample variance is $s^2 = 34.768$, and the sample skewness is $b_1 = 1.53$. The data include some zero count observations. To determine whether this data can be classified as zero-inflated, we calculate the sample index dispersion as $i = s/\overline{x} = 2.99 > 1$, indicating that the data exhibit overdispersion. This suggests that the TWG model can be used to fit this data. The theoretical skewness measure of the TWG and WG distributions involves an unknown parameter, θ, which needs to be estimated for comparison purposes. The ML estimate of θ under the WG distribution is $\widehat{\theta}_{\text{WG}} = 0.83397$, and under the TWG distribution, it is $\widehat{\theta}_{\text{TWG}} = 0.8363$. Consequently, the estimated range of the WG's skewness measure is $(0.35, 4.38)$, and for the TWG's measure, it is $(0.024, 10.051)$. The sample skewness $b_1 = 1.53$ falls within the estimated ranges of skewness for both models. However, the TWG distribution provides a wider range, suggesting it may better capture the features of the current data compared to the TW distribution. The numerical values of ML estimates and their corresponding standard errors (SE), $-\widehat{\ell}$, AIC, CAIC, BIC, and K-S statistics with their corresponding p-values are given in Table 2. Additionally, the table includes the chi-square (χ^2) test along with its corresponding p-value. Considering the p-value of the K-S statistic, we may conclude that the candidate models fit this data well. However, a closer look at the value of χ^2 along with its corresponding p-value suggests that the TWG is better than all considered models. More precisely, the value of χ^2 along with its p-value (in parentheses) is 6.93157(0.73189) for TWG; 8.47572(0.58247) for WG; 9.12373(0.52040) for NB; 14.17070(0.16535) for ND; 12.66584(0.24296) for GGD; 9.84419(0.45427) for GPLD; 8.95550(0.53633) for NGPLD; and 437.24582 (< 0.00001) for POID. Clearly, the TWD outperforms all considered models in terms of having the smallest value of χ^2 and the highest p-value.The performance of WG and NGPLD is almost similar, followed by NB as a close second, whereas the performance of ND, GGD, and POID is poor, with POID being the worst candidate for this data. Figures 7 demonstrate the empirical distribution of the number of claims versus the fitted models, revealing that the fitted count is almost close to the empirical count counterpart. This can also be confirmed by examining Table 2. Figure 8 (right panel) confirms that the overall pattern follows almost a straight line, indicating that the data come from the TWG. Finally, the violin plot for the number of ticks is shown in Figure 8 (left panel), with the mean of the empirical data being 6.56.

Table 2. The goodness-of-fit for the number of ticks on sheep data.

Number of ticks	Frequency	TWG	WG	NB	ND	GGD	GPLD	NGPLD	POID
0	4	3.84899	5.36429	5.25564	8.62476	7.60945	5.46171	4.53062	0.11599
1	5	7.62239	7.72433	7.35044	8.15383	7.65695	7.11952	6.87624	0.76100
2	11	9.00703	8.41172	8.03188	7.64244	7.53404	7.74641	7.88163	2.49646
3	10	8.87624	8.20883	7.95758	7.09696	7.25145	7.75768	8.04428	5.45975
4	9	8.06696	7.56929	7.47830	6.52613	6.83284	7.39048	7.70253	8.95531
5	11	7.04584	6.75091	6.79914	5.94055	6.31081	6.80587	7.08277	11.75112
6	3	6.02967	5.89570	6.04301	5.35202	5.72181	6.11455	6.33324	12.84980
7	5	5.10505	5.07781	5.28296	4.77256	5.10141	5.39079	5.54816	12.04389
8	3								
9	2 } 7	10.92693	11.10266	11.76539	11.09105	11.69298	12.11448	12.29804	21.80242
10	2								
11	5								
12	0								
13	2 } 9	7.90013	8.15474	8.68048	8.74151	8.83863	8.93497	8.85463	5.50042
14	2								
15	1								
16	1								
17	0								
18	0								
19	1								
20	0 } 8	7.57076	7.73973	7.35518	8.05820	7.44962	7.16354	6.84786	0.26384
21	1								
22	1								
23	1								
24	0								
$\geq$25	2								
Total	82	82	82	82	82	82	82	82	82
Parameter		$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(r,\widehat{p})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$\widehat{\alpha}$
ML Estimate		(0.83632,3.45951)	(0.83397,1.75893)	(1.77748,0.21317)	(0.81583,-3.21969)	(0.80878,2.31141)	(0.31070,1.27577)	(0.30408,61.52066)	(6.56098)
SE		(0.02100,1.54830)	(0.02626,1.46955)	(0.35158,0.03669)	(0.03315,3.36365)	(0.02970,0.82881)	(0.06192,0.39822)	(0.03024,1106.74777)	(0.28286)
-2ℓ		**473.53610**	474.97729	475.92369	480.62995	479.43766	476.88396	476.27931	650.23707
AIC		**477.53610**	478.97729	479.92369	484.62995	483.43766	480.88396	480.27931	652.23707
CAIC		**477.68800**	479.12919	480.07559	484.78184	483.58955	481.03585	480.43121	652.28707
BIC		**482.34954**	483.79073	484.73713	489.44338	488.25109	485.69739	485.09275	654.64379
HQIC		**479.46862**	480.90981	481.85622	486.56247	485.37018	482.81648	482.21183	653.20334
K-S (stat)		**0.13998**	0.93458	0.93591	0.89482	0.90720	0.93339	0.94475	0.99859
K-S (p-value)		**0.08043**	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001	< 0.00001
χ^2 (stat)		**6.93157**	8.47572	9.12373	14.17070	12.66584	9.84419	8.95550	437.24582
d.f.		10	10	10	10	10	10	10	10
χ^2 (p-value)		**0.73189**	0.58247	0.52040	0.16535	0.24296	0.45427	0.53633	< 0.00001

8.2. Data set 2: Number of claims by auto insurance policyholders

The second dataset we use is from [29], representing the number of claims made by automobile insurance policyholders. The sample for this data has a mean of $\overline{x} = 0.194$, a variance of $s^2 = 0.226$, and a skewness of $b_1 = 2.901$. The sample index of dispersion is $i = 1.163$, indicating an excess of zero counts, as seen in Table 3. Consequently, the TWG model could potentially fit this data. The estimated range index of dispersion I_X under the WG and TWG distributions, considering the ML estimates of θ under these models, are $(0.353, 10.5778)$ and $(0.024, 34.378)$, respectively. Notice that the estimated range of the skewness measure provides a wider range than the corresponding range under the WG distribution. This flexibility in the range may capture the skewness present in the data much better. The numerical values of ML estimates and their corresponding standard errors

Figure 7. The fitted PMFs for the number of ticks on sheep data.

(SE), $-\widehat{\ell}$, AIC, CAIC, BIC, and K-S statistics with their corresponding p-values are given in Table 3. Additionally, the table includes the chi-square (χ^2) test along with its corresponding p-value. Based on the p-value of the K-S statistic, we conclude that the candidate models fit this data well. On the other hand, the value of the χ^2 and its corresponding p-value suggest that the TWG is better than all considered models. Specifically, the χ^2 values along with corresponding p-values (in parentheses) are 2.23057 (0.52595) for TWG, 2.31023 (0.51056) for WG, 2.50381 (0.47460) for NB, 2.32019 (0.50866) for ND, 2.30978 (0.51065) for GGD, 2.52653 (0.47052) for GPLD, 2.66315 (0.44653) for NGPLD, and 6.16087 (0.04594) for POID. Clearly, the TWG model performs better than all considered models in terms of having the smallest χ^2 value. Notice that the performance of GGD is slightly better

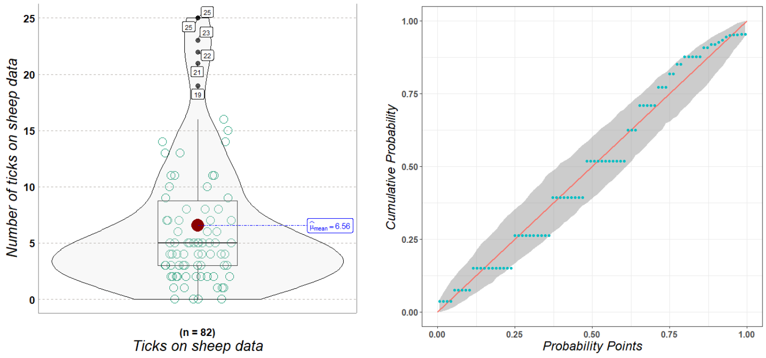

Figure 8. The Violin plot for the number of ticks on sheep data and the PP plot of TWG model.

than WG, followed closely by ND, whereas the performance of NB, GPLD, NGPLD and POID is unsatisfactory, with POID being the worst candidate for this data. Figures 9 demonstrate the empirical distribution of the number of claims versus the fitted models, showing that the TWG model is slightly better than the WG distribution and much better than the other considered models. Figure 10 (right panel) shows the P-P plot for the TWG, revealing that the data can be reasonably fitted by this model. Finally, the violin plot for the number of claims is shown in Figure 10 (left panel), with the mean of the data being 0.19.

Table 3. The goodness-of-fit for the number of claims in automobile insurance data.

Number of claims	Frequency	TWG	WG	NB	ND	GGD	GPLD	NGPLD	POID
0	1563	1563.64977	1564.27976	1564.54711	1563.70288	1563.61628	1564.47933	1564.60266	1544.15002
1	271	266.23346	265.10615	264.48258	266.14402	266.29933	264.58780	264.25293	299.77364
2	32	38.65674	39.05115	39.44336	38.75173	38.69618	39.43853	39.68639	
3	7 } 9 } 41								31.07634
4	2	6.46003	6.56293	6.52695	6.40137	6.38821	6.49434	6.45803	
Total	1875	1875	1875	1875	1875	1875	1875	1875	1875
Parameter		$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(r,\widehat{p})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$\widehat{\alpha}$
ML Estimate		(0.14301,1.23178)	(0.14324,0.87361)	(1.30821,0.87078)	(0.14122,-0.45415)	(0.14118,1.21138)	(7.28427,1.29334)	(7.87479,8.83600)	(0.19414)
SE		(0.02050,0.62647)	(0.02311,0.70663)	(0.41723,0.03644)	(0.02102,0.52387)	(0.02073,0.22442)	(2.09865,0.43049)	(2.33896,18.40592)	(0.01018)
-2ℓ		**1986.67188**	1986.77573	1987.00211	1986.76708	1986.75321	1987.02640	1987.18774	2003.50320
AIC		**1990.67188**	1990.77573	1991.00211	1990.76708	1990.75321	1991.02640	1991.18774	2005.50320
CAIC		**1990.67829**	1990.78214	1991.00852	1990.77349	1990.75962	1991.03281	1991.19415	2005.50533
BIC		**2001.74461**	2001.84845	2002.07484	2001.83981	2001.82594	2002.09913	2002.26047	2011.03956
HQIC		**1994.75084**	1994.85469	1995.08107	1994.84604	1994.83217	1995.10536	1995.26670	2007.54268
K-S (stat)		**0.83395**	0.83428	0.83443	0.83397	0.83393	0.83439	0.83445	0.82355
K-S (p-value)		$<$ **0.00001**	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
χ^2 (stat)		**2.23057**	2.31023	2.50381	2.32019	2.30978	2.52653	2.66315	6.16087
d.f.		3	3	3	3	3	3	3	2
χ^2 (p-value)		**0.52595**	0.51056	0.47460	0.50866	0.51065	0.47052	0.44653	0.04594

8.3. *Data set 3: European corn borer biological experiment data*

As a Final example, we consider the European corn borer biological experiment data [24], which is given in Table 4. For this data we have that ,$\overline{x} = 1.434$, the sample variance is $s^2 = 2.446$, and the sample skewness is $b_1 = 1.156$. The sample index of dispersion is $i = 1.163$, indicating an excess of zero counts, as seen in Table 3. So, the TWG model can be used to fit this data. Likewise, the estimated range index of dispersion I_X under the WG and TWG distributions, using the ML estimates of θ considering these models, are $(0.353, 5.318)$ and $(0.024, 17.283)$, respectively. Clearly, the sample skewness belongs to both range intervals. However, the estimated range of skewness under the TWG distribution is wider than the corresponding one under TW distribution. Prehaps, such property maybe useful in analyzing the considered data fairly eoungh. Table 4 presents the numerical values of ML estimates, their standard errors (SE), $-\widehat{\ell}$, AIC, CAIC, BIC, and K-S statistics with corresponding p-values. Additionally, the table includes the chi-square (χ^2) test and its p-value. The p-value of the K-S statistic indicates that the candidate models fit this data well.Conversely, the χ^2 value and its p-value suggest that the TWG model is superior to all other models considered. Specifically, the χ^2 values and their p-values (in parentheses) are 0.7696

Figure 9. The fitted PMFs for the number of claims in automobile insurance data.

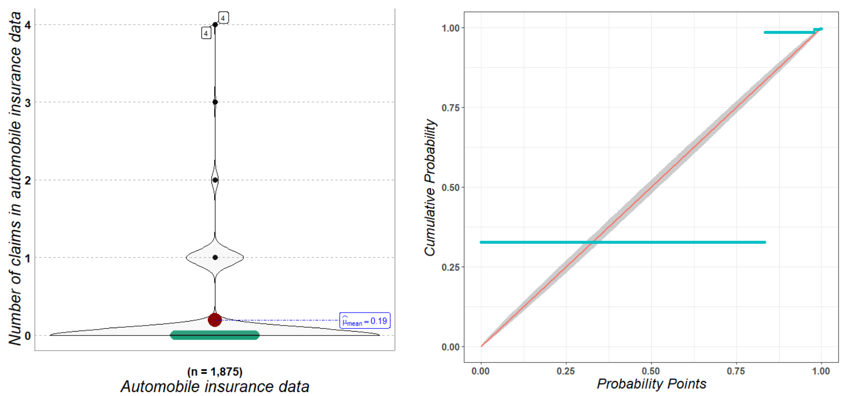

Figure 10. The Violin plot for the number of claims in automobile insurance data and the PP plot of TWG model.

(0.979) for TWG, 0.9113 (0.96940) for WG, 1.1866 (0.946) for NB, 1.4621 (0.917)for ND, 1.38956 (0.925) for GGD, 1.38133 (0.926) for GPLD, 1.4445 (0.919)for NGPLD , 38.53494 (< 0.00001) for POID. The TWG model outperforms all other models, exhibiting the smallest (χ^2) test and its p-value, with the WG model as a close second. While the performance of the NB model is satisfactory, it falls short compared to the TWG and WG models. The GGD and GPLD models show similar performance, as do the ND and NGPLD models. Lastly, the POID model proves to be the least effective candidate for fitting this count data.. Figures 11 depict the empirical distribution of the number of claims versus the fitted models, showing that the TWG model is better than the WG distribution and much better than the other models. Figure 12 (right panel) reveals that the P-P plot is approximately a straight line, implying that the current follows the TWG distributions. Finally, the violin plot for number of borers is displayed in Figure 12 (left panel) with the mean of the current data being 1.434

Table 4. The goodness-of-fit for the number of borers per hill of corn data.

Number of borers	Frequency	TWG	WG	NB	ND	GGD	GPLD	NGPLD	POID
0	43	43.54439	44.03339	44.27292	45.18934	44.86048	44.26714	44.61771	27.22561
1	35	32.43318	31.78375	31.08465	30.23314	30.48909	30.75575	30.46035	40.38467
2	17	19.05118	19.03365	19.09808	18.86369	19.00636	19.19999	19.06579	29.95196
3	11	10.83777	10.91771	11.17520	11.19163	11.21716	11.36347	11.33580	14.80958
4	5	6.13667	6.19118	6.37575	6.42545	6.40738	6.50053	6.51439	5.49189
5	4								
6	1								
7	2 } 9	7.99680	8.04032	7.99340	8.09675	8.01952	7.91312	8.00596	2.13628
8	2								
Total	120	120	120	120	120	120	120	120	120
Parameter		$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(r,\widehat{p})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$(\widehat{\theta},\widehat{\alpha})$	$\widehat{\alpha}$
ML Estimate		(0.56577,3.36179)	(0.56461,2.23691)	(1.33313,0.47334)	(0.54812,-0.87792)	(0.54747,1.38447)	(1.07982,1.12032)	(1.05837,1.40250)	(1.48333)
SE		(0.04164,1.84988)	(0.04696,1.92865)	(0.37334,0.07491)	(0.06064,1.30187)	(0.05691,0.45802)	(0.26916,0.42338)	(0.27516,2.49000)	(0.11118)
-2ℓ		**400.26108**	400.39705	400.60973	400.89437	400.82906	400.77618	400.83031	438.37586
AIC		**404.26108**	404.39705	404.60973	404.89437	404.82906	404.77618	404.83031	440.37586
CAIC		**404.36365**	404.49962	404.71229	404.99694	404.93163	404.87874	404.93287	440.40976
BIC		**409.83607**	409.97204	410.18471	410.46936	410.40404	410.35116	410.40529	443.16335
HQIC		**406.52511**	406.66108	406.87376	407.15840	407.09309	407.04021	407.09434	441.50787
K-S (stat)		**0.36287**	0.63306	0.63106	0.62342	0.62616	0.63111	0.62819	0.77312
K-S (p-value)		$<$ **0.00001**	$<$ 0.00001	$<$ 0.00001	$<$ 0.00001	$<$ 0.00001	$<$ 0.00001	$<$ 0.00001	$<$ 0.00001
χ^2 (stat)		**0.76961**	0.91134	1.18662	1.46206	1.38956	1.38133	1.44447	38.53494
d.f.		5	5	5	5	5	5	5	5
χ^2 (p-value)		**0.97893**	0.96940	0.94615	0.91741	0.92545	0.92634	0.91939	$<$ 0.00001

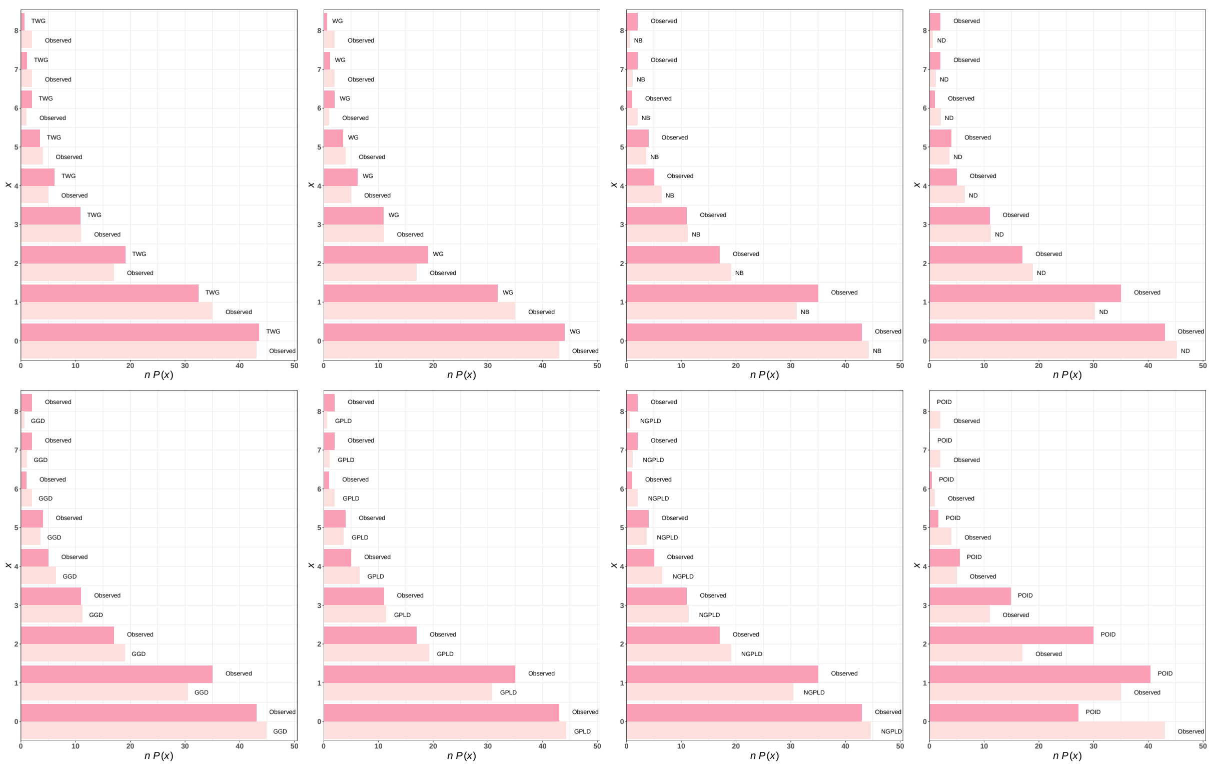

Figure 11. The fitted PMFs for the number of borers per hill of corn data.

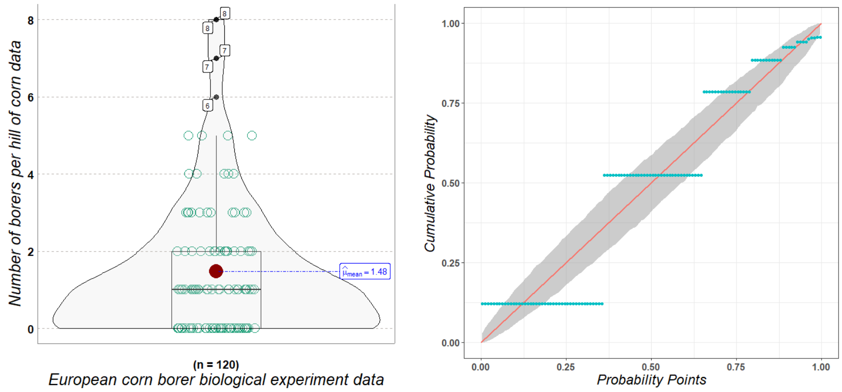

Figure 12. The Violin plot for the number of borers per hill of corn data and the PP plot of TWG model.

9. Discussion and Conclusion

This article introduces a new flexible weighted geometric (TWG) distribution. Among its many notable properties, the TWG distribution excels in capturing skewness due to its broader range of skewness measures, resulting in satisfactory performance when fitting real data. Several important statistical and mathematical features, including unimodality, infinite divisibility, and the moment-generating function, are obtained and interpreted. The model parameters are estimated using the maximum likelihood method, and their existence is discussed. The finite sample behavior of these estimates is investigated through a Monte Carlo simulation study. Finally, the proposed model is fitted using real data that exhibits overdispersion, and it is compared to some existing models, with the proposed model outperforming all others.

In this section, we provide a further discussion of the proposed distribution. The TWG distribution introduced in (3), along with its properties discussed in previous sections, can be viewed as a special case of the following family of probability mass functions (PMF):

Definition 2
A random variable X is said to follow a two-parameter weighted geometric distribution with parameters $\alpha_1,\ \alpha_2,\ \theta > 0$, if its PMF of is given below

$$gp_X(x;\theta,\alpha_1,\alpha_2) := \Pr[X=x] = k(\theta,\alpha_1,\alpha_2)(1-\theta)\,\theta^x \left(1-\theta^{\alpha_1(1+x)}\right)\left(1-\theta^{\alpha_2(1+x)}\right), \quad x = 0,1,2,\ldots \tag{21}$$

and 0 otherwise, where $k(\theta,\alpha_1,\alpha_2$ is a normalizing constant

Observe that when $\alpha_1 \to \infty$, then $gp_X(x;\theta,\alpha_1,\alpha_2)$ converges to WG(θ , α_2,) thus generalizing the WG distribution proposed in [14]. Similarly, when $\alpha_2 \to \infty$, then $gp_X(x;\theta,\alpha_1,\alpha_2)$ converges to WG(θ , α_1,) again generalizing the WG distributions. Therefore, the family of PMF distribution generated via (21) generalizes the WG distribution. Moreover, when $\alpha_1 = \alpha_2 = \alpha$, we obtain the TWG distribution, with its PMF given via (3).

While the family of PMF described via in (21) is interesting and includes the WG and TWG as special cases, it presents several challenges. First, the normalization is intractable. Second, the PMF in (21) is not identifiable, leading to difficulties in statistical inference, particularly in obtaining the ML estimates, unless a constraint is imposed on the parameters α_1 and α_2, for example, $\alpha_2 = w(\alpha_1)$. A specific case of this is when $\alpha_1 = \alpha_2$, which yields the TWG distribution discussed in this paper. A detailed statistical analysis, along with numerical computations, is warranted and would be better explored in a separate paper. We plan to address these points in future research. Additionally, statistical inferences about the unknown parameters can be made using different estimation methods, including Bayesian approaches and regression modeling, particularly median regression, since the mean function has a complex structure.

Acknowledgement

We thank the Editor and the two anonymous reviewers for their invaluable comments, which improved the manuscript.

REFERENCES

1. B. C. Truong, K.H Pho, C. C. Dinh, M. McALEER, *Zero-inflated poisson regression models: Applications in the sciences and social sciences* Annals of Financial Economics, vol. 16, pp. 1–19, 2021
2. M. N. Akram, M. Amin, N. Afzal, B.M.G. Kibria, *Kibria—Lukman estimator for the zero inflated negative binomial regression model: theory, simulation and applications* Communications in Statistics – Simulation and Computation, pp. 1–17, 2023
3. M. Amin, M.N. Akram, A. Majid, *On the estimation of Bell regression model using ridge estimator*, Communication in Statistics–Simulation and Computation, vol.52, no.3,pp. 854—867 2021.
4. G. Zuo, K. Fu, X. Dai, L. Zhang, *Generalized Poisson Hurdle Model for Count Data and Its Application in Ear Disease*, Entropy, vol. 23, pp. 1206, 2021
5. D., Roy, *Discrete Rayleigh distribution*, IEEE Trans. Reliab., vol. 53, pp. 255–260, 2004.
6. H. Krishna, and P. S. Pundir, *Discrete Burr and discrete Pareto distributions*, Statistical Methodology, vol. 6, no. 2, pp. 177–188, 2009.
7. J. D. Englehardt, and R. C. Li., *The discrete Weibull distribution: an alternative for correlated counts with confirmation for microbial counts in water*, Risk Analysis, vol. 31, no. 3, pp. 370–381, 2011.
8. V. Nekoukhou, M. H. Alamatsaz, and H. Bidram, *Discrete generalized exponential distribution*, Statistics, vol.47, no. 4, pp. 876–887, 2013.
9. A.A. Al-Babtain, A.H.N. Ahmed, and A.Z. Afify, *A new discrete analog of the continuous Lindley distribution, with reliability applications*, Entropy , vol. 22, pp. 603., 2020.
10. M. El-Morshedy, H.M. Aljohani, M.S. Eliwa, M. Nassar, M.K. Shakhatreh, A.Z. Afify, *The Exponentiated Burr–Hatke Distribution and Its Discrete Version: Reliability Properties with CSALT Model, Inference and Applications*, Mathematics, vol. 9, pp. 2277, 2021.
11. G. E. Willmot, *The Poisson-inverse Gaussian distribution as an alternative to the negative binomial*, Scandinavian Actuarial Journal, vol. 3–4 pp. 113–127, 1987.
12. K. Karakaya, *A new discrete distribution with applications to radiation, smoking, and health data* , Journal of Radiation Research and Applied Sciences, vol. 16, pp. 100735, 2023.
13. H.M. Aljohani, Y. Akdoğan, G.M. Cordeiro, A.Z. Afify, *The Uniform Poisson–Ailamujia Distribution: Actuarial Measures and Applications in Biological Science*, Symmetry, vol. 13, pp. 1258, 2021
14. D. Bhati, S. Joshi, *Weighted geometric distribution with new characterizations of geometric distribution*, Communications in Statistics – Theory and Methods, vol.47, pp. 1510–1527, 2018.
15. W. Bodhisuwan, and S. Sangpoom, *The discrete weighted Lindley distribution*, 2016 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA), Banda Aceh, Indonesia, 99-103, 2016.
16. D. Das, and B. Das, *On Weighted Discretized Fréchet–Weibull Distribution ´ with Application to real life data*, Applied Mathematics & Information Sciences, vol. 7, no. 5, pp. 791–806, 2023.
17. C. Tapan, *A Simple Discrete-Weighted Exponential Distribution and Applications*, PBRU SCIENCE JOURNAL, vol. 20, no.1 ,pp. 85–96, 2023.
18. A. Azzalini, *A class of distributions which includes the normal ones*, Scandinavian Journal of Statistics, vol. 82, no.12, pp. 171–178, 1985
19. S. Nadarajah, The skew logistic distribution. *AStA Advances in Statistical Analysis*, vol. 93, pp. 187–203,2009
20. M.K. Shakhatreh, *A two-parameter of weighted exponential distributions*, Statistics and Probability letters, vol. 82, pp. 252–261,2012
21. H. Najarzadegan, and M. H. Alamatsaz, *A New Generalization of Weighted Geometric Distribution and its Properties*, Journal of Statistical Theory and Applications, vol. 16, pp. 522—546, 2017
22. R.B. Arellano-Valle, D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections. *The Canadian Journal of Statistics*, vol. 34, pp. 581–601, 2006
23. MR, Lindgren G, Rootzen H. *Extremes and related properties of random sequences and processes.* New York: Springer; 1983 .
24. G. Beall, The fit and significance of contagious distributions when applied to observations on larval insects. *Ecology*, vol.21, no. 4, pp. 460—474, 1940
25. D. Bhati, D.V.S. Sastry, and P.Z. Maha Qadri, A new generalized Poisson Lindley distribution. *Austrian Journal of Statistics*, vol.44, no. 4, pp. 35—51, 2015
26. D.E. Gomez, *Another generalization of the geometric distribution*, *Test*, vol. 19, pp. 399—415, 2010
27. D.E. Gomez, J. M. Sarabia, and E.C. Ojeda, *A new discrete distribution with actuarial applications*, *Insurance, Mathematics and Economics*, vol.48, pp. 406—412, 2011
28. R.A. Fisher, The negative binomial distribution. *Annals of Eugenics*, vol.11, no.1, pp. 182-187, 1941
29. S.A. Klugman, H.H. Panjer, and G.E. Willmot, *Loss models: From data to decisions.* 3rd ed, John Wiley and Sons, New York, 2021
30. E. Mahmoudi and H. Zakerzadeh, Generalized Poisson-Lindley distribution. *Communications in Statistics - Theory and Methods*, vol.39, no.10,pp. 1785—1798, 2010
31. R Core Team. (2022). *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria. Available online: https://www.R-project.org/(accessed on: 31 October 2022).