ISSN: 2832-4293

Pan-American Journal of Mathematics 5 (2026), 12
https://doi.org/10.28919/cpr-pajm/5-12


# A GENERALIZED COMPOSITE POWER SERIES FAMILY OF DISTRIBUTIONS WITH APPLICATION TO REAL DATA

RAHMAT AL KAFI[1,2] AND PAWAT PAKSARANUWAT[2,*]

ABSTRACT. In this article, we introduce a new flexible family of discrete probability distributions called the generalized composite power series (GCPS). This family was constructed by combining two classes of discrete distributions to generate a wide range of flexible discrete probability models. We derived the probability mass functions and cumulative distribution function used to characterize members of the GCPS family. To demonstrate its flexibility, we herein present two specific members: the generalized geometric zero-truncated Poisson (GGZTP) and the generalized geometric logarithmic (GGL) distributions. Various statistical properties of the GGZTP and GGL distributions are explored, and parameter estimation methods are discussed. To showcase their potential as alternative discrete distributions for count data, we demonstrate their applicability to real datasets, along with a comparative analysis involving two well-known flexible discrete distributions.

## 1. INTRODUCTION

Probability distributions provide a mathematical framework for studying real-world phenomena, particularly for describing the likelihood of different outcomes in uncertain events. For example, global disease incidence can be analyzed using the number of reported cases, while insurance problems often involve modeling the frequency of policyholder claims. Events such as the number of reported cases or claims can be effectively examined using non-negative discrete probability distributions. However, when developing a new probability distribution as an alternative to existing models, sufficient consideration of several factors, particularly flexibility, is required. The concept of flexibility varies depending on the type of distribution. In the context of discrete distributions, flexibility refers to the ability of a model to represent various dispersion behaviors in count data. Specifically, three types of dispersion are commonly considered: underdispersion, equidispersion, and overdispersion. These are quantified using the index of dispersion (IOD), which is defined as the ratio of the variance to the mean. A distribution is said to be underdispersed when its IOD is less than one, equidispersed when the IOD equals one, and overdispersed when the IOD exceeds one.

Classical discrete distributions, such as Poisson, geometric, and negative binomial, exhibit limited flexibility because each can model specific dispersion patterns only. For instance, the well-known Poisson distribution can model only equidispersed data, where the variance equals the mean. In response, researchers

[1]DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCES, UNIVERSITAS INDONESIA, DEPOK 16424, INDONESIA

[2]DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE, CHIANG MAI UNIVERSITY, CHIANG MAI 50200, THAILAND

*Email address*: rahmatal_kafi@cmu.ac.th, pawat.pak@cmu.ac.th.

Submitted on May 29, 2026.

2020 *Mathematics Subject Classification.* Primary 60E05, 62F10.

*Key words and phrases.* geometric distribution; index of dispersion; logarithmic distribution; maximum likelihood estimation; Poisson distribution.

*Corresponding author.

have increasingly focused on developing more flexible discrete distributions that can accommodate a wider range of count data behaviors. Many of these efforts involve modifying or compounding existing distributions to enhance their flexibility [1–6]. However, many of these models still lack full flexibility. For example, the convoluted Poisson–negative binomial distribution [6], despite having two parameters, has an IOD greater than one, thus making it suitable for overdispersed data only.

Nevertheless, there have been successful efforts to introduce more flexible discrete distributions. For example, the Conway–Maxwell–Poisson (CMP) distribution generalizes the Poisson distribution and includes an additional parameter to control dispersion [7]. Similarly, the distributions introduced in [8] and the partial-geometric (PG) distribution proposed in [9] also offer greater flexibility. However, some of these models, such as those in [7] and [8], do not have closed-form expressions for their probability mass functions (PMFs), which makes computations involving them challenging. Moreover, the modified partial-geometric distribution proposed in [10] is flexible but contains three parameters, one of which is subject to a nonlinear constraint, leading to computationally expensive parameter estimation.

Rather than exploiting a single flexible distribution, it is more beneficial to introduce a family of discrete distributions that can generate a variety of flexible models. For years, the power series family, comprising the Poisson, geometric, binomial, negative binomial, and logarithmic distributions [11], has been the only one suitable for discrete distributions. However, as mentioned earlier, each of these distributions is still limited to a specific type of dispersal behavior. In the present work, we introduce a new family of discrete distributions called generalized composite power series (GCPS). This family, which can be regarded as an extension of the classic power series family, is primarily defined via its cumulative distribution function (CDF). The GCPS family is constructed using two discrete distributions drawn from the power series family. To illustrate the flexibility of this new family, we explore several of its specific members. We also demonstrate that this novel family of discrete distributions encompassing flexible models is capable of representing a wide range of dispersion behaviors in count data.

## 2. The GCPS Family of Distributions

The basic idea behind GCPS family involves incorporating two distinct power series distributions. To this end, let $N$ be a positive integer-valued random variable; and $Y_1, Y_2, \dots, Y_N$ be independent and identically distributed (i.i.d.) random variables from the power series family of distributions with a PMF formulated as

$$p_Y(y) = \Pr(Y = k) = \frac{a_k \phi^k}{g(\phi)}, \quad \phi > 0;\ k = 0, 1, 2, \dots, \tag{2.1}$$

where $\phi$ denotes the parameter for the baseline distribution, $a_k$ is a non-negative real number for each $k$, and

$$g(\phi) = \sum_{k=0}^{\infty} a_k \phi^k$$

is finite for some values of $\phi$ [11]. The corresponding CDF is

$$F_Y(y) = \Pr(Y \leq y) = \sum_{j=0}^{\lfloor y \rfloor} \frac{a_j \phi^j}{g(\phi)} = \frac{g_y(\phi)}{g(\phi)}, \quad y \geq 0, \tag{2.2}$$

where $g_y(\phi) = \sum_{j=0}^{\lfloor y \rfloor} a_j \phi^j$, and $\lfloor \cdot \rfloor$ denotes the floor function [11].

Assume that random variable $N$ follows a zero-truncated power series distribution with a PMF defined as

$$p_N(n) = \Pr(N = n) = \frac{b_n \vartheta^n}{c(\vartheta)}, \quad \vartheta > 0;\ n = 1, 2, 3, \dots, \tag{2.3}$$

where $\vartheta$ denotes the parameter of the baseline distribution, $b_n$ is a non-negative real number for each $n$, and $c(\vartheta) = \sum_{n=1}^{\infty} b_n \vartheta^n$ is finite for some values of $\vartheta$ [11].

Let

$$X = \max\{Y_1, Y_2, \ldots, Y_N\},$$

where the $Y_i$'s and $N$ are independent. The conditional CDF for $X$ given $N = n$ is obtained as

$$\begin{aligned} \Pr(X \le x | N = n) &= \Pr\left(\max\{Y_1, Y_2, \ldots, Y_N\} \le x \mid N = n\right) \\ &= \Pr(Y_1 \le x, Y_2 \le x, \ldots, Y_n \le x) \\ &= \Pr(Y_1 \le x)\Pr(Y_2 \le x)\cdots\Pr(Y_n \le x) \\ &= \left(\sum_{j=0}^{\lfloor x \rfloor} \frac{a_j \phi^j}{g(\phi)}\right)^n, \quad x \ge 0. \end{aligned} \tag{2.4}$$

Therefore, the joint probability of $X \le x$ and $N = n$ becomes

$$\Pr(X \le x, N = n) = \left(\sum_{j=0}^{\lfloor x \rfloor} \frac{a_j \phi^j}{g(\phi)}\right)^n \frac{b_n \vartheta^n}{c(\vartheta)}. \tag{2.5}$$

The GCPS family is defined as the marginal distribution of $X$. Hence, the CDF of $X$ can be written as

$$\begin{aligned} F_X(x) = \sum_{n=1}^{\infty} \Pr(X \le x, N = n) &= \sum_{n=1}^{\infty} \left(\sum_{j=0}^{\lfloor x \rfloor} \frac{a_j \phi^j}{g(\phi)}\right)^n \frac{b_n \vartheta^n}{c(\vartheta)} \\ &= \frac{1}{c(\vartheta)} \sum_{n=1}^{\infty} b_n \left(\vartheta \sum_{j=0}^{\lfloor x \rfloor} \frac{a_j \phi^j}{g(\phi)}\right)^n, \quad x \ge 0. \end{aligned}$$

Therefore, we have the following definition about our proposed family of distributions.

**Definition 2.1.** A discrete distribution with non-negative integer support is said to belong to the GCPS family of distributions if its CDF can be expressed in the following form:

$$F_X(x) = \Pr(X \le x) = \begin{cases} 0, & x < 0 \\ \frac{1}{c(\vartheta)} \sum_{n=1}^{\infty} b_n \left(\vartheta \frac{\sum_{j=0}^{\lfloor x \rfloor} a_j \phi^j}{g(\phi)}\right)^n, & x \ge 0 \end{cases}, \tag{2.6}$$

where $\lfloor \cdot \rfloor$ denotes the floor function, $\phi, \vartheta, a_j, b_n > 0$,

$$g(\phi) = \sum_{k=0}^{\infty} a_k \phi^k,$$

which is finite for some values of $\phi$, and

$$c(\vartheta) = \sum_{n=1}^{\infty} b_n \vartheta^n,$$

which is finite for some values of $\vartheta$.

The corresponding PMF for the GCPS family is obtained as

$$p_X(k) = \begin{cases} \frac{1}{c(\vartheta)} \sum_{n=1}^{\infty} b_n \left(\vartheta G(0)\right)^n, & k = 0 \\ \frac{\sum_{n=1}^{\infty} b_n \left(\vartheta G(k)\right)^n}{c(\vartheta)} - \frac{\sum_{n=1}^{\infty} b_n \left(\vartheta G(k-1)\right)^n}{c(\vartheta)}, & k = 1, 2, \ldots \end{cases}, \tag{2.7}$$

where

$$G(x)=\sum_{j=0}^{x}\frac{a_j\phi^j}{g(\phi)}.$$

## 3. Some Members of the GCPS Family

To show the flexibility of the GCPS family of distributions, we studied some of its members. In this section, selected members of the GCPS family, namely the generalized geometric zero-truncated Poisson (GGZTP) and the generalized geometric logarithmic (GGL) distributions, are discussed. Kafi *et al.* [12] introduced the GGZTP distribution, derived several of its statistical functions, and explored its properties. However, they did not discuss its sub-models and applications, which we address here.

3.1. **The GGZTP distribution.** First, we show that the GGZTP distribution belongs to the GCPS family. To this end, let $Y_1, Y_2, \ldots, Y_N$ be i.i.d. random variables following a geometric distribution with parameter $p \in (0,1)$, and let $N$ be a zero-truncated Poisson random variable with parameter $\theta > 0$.

The geometric distribution is a member of the power series family since its PMF can be written as

$$p_Y(k)=(1-p)^k p=\frac{(1-p)^k}{1/p}=\frac{a_k\phi^k}{g(\phi)},\quad k=0,1,2,\ldots, \tag{3.1}$$

where $\phi=1-p$, $a_k=1$ for all $k=0,1,2,\ldots$, and

$$g(\phi)=\sum_{k=0}^{\infty}(1-p)^k=\frac{1}{1-(1-p)}=\frac{1}{1-\phi}.$$

Meanwhile, the zero-truncated Poisson distribution with parameter $\theta > 0$ is a member of the zero-truncated power series family since its PMF can be expressed as

$$p_N(n)=\frac{\theta^n}{n!(e^\theta-1)}=\frac{b_n\vartheta^n}{c(\vartheta)},\quad n=1,2,3,\ldots, \tag{3.2}$$

where $\vartheta=\theta$, $b_n=1/n!$, and

$$c(\vartheta)=\sum_{n=1}^{\infty}\frac{\theta^n}{n!}=e^\theta-1=e^\vartheta-1.$$

Following the procedure in Section 2, we can obtain a GCPS family member, denoted as $\mathrm{GGZTP}(\theta,p)$.

The CDF of the GGZTP distribution is obtained from (2.6) as

$$F_X(x)=\begin{cases}0, & x<0\\ \frac{1}{e^\vartheta-1}\sum_{n=1}^{\infty}\left[\left(\sum_{j=0}^{\lfloor x\rfloor}\frac{\phi^j}{1/(1-\phi)}\vartheta\right)^n\frac{1}{n!}\right], & x\geq 0\end{cases}, \tag{3.3}$$

Hence, in accordance with Definition 2.1, the GGZTP distribution belongs to the GCPS family. Moreover, since $\phi=1-p$ and $\vartheta=\theta$, the CDF in Equation (3.3) can be simplified. According to [12], the CDF of the GGZTP distribution can be written as

$$F_X(x)=\Pr(X\leq x)=\begin{cases}0, & x<0\\ \dfrac{\exp\left\{\theta\left[1-(1-p)^{\lfloor x\rfloor+1}\right]\right\}-1}{e^\theta-1}, & x\geq 0\end{cases}. \tag{3.4}$$

Meanwhile, the corresponding PMF is given by

$$p_X(k)=\begin{cases}\dfrac{e^{\theta p}-1}{e^\theta-1}, & k=0\\ \dfrac{e^\theta\left[e^{-\theta(1-p)^{k+1}}-e^{-\theta(1-p)^k}\right]}{e^\theta-1}, & k=1,2,\ldots\end{cases}, \tag{3.5}$$

where $\theta > 0$ and $0 < p < 1$.

Since the asymptotic behavior of the GGZTP PMF was not established in [12], we did so via the following proposition.

**Theorem 3.1.** *The asymptotic behavior of the* $\mathrm{GGZTP}(\theta, p)$ *PMF as* $\theta \to 0^+$ *is*

$$\lim_{\theta\to 0^+} p_X(k) = (1-p)^k p, \quad k = 0, 1, 2, \dots,$$

*which is the PMF of the geometric distribution with parameter* $p$.

*Proof.* Consider Equation (3.5) under the condition that $\theta \to 0^+$. For $k = 0$, we obtain

$$\lim_{\theta\to 0^+} \frac{e^{\theta p} - 1}{e^{\theta} - 1} = p.$$

Similarly, $k = 1, 2, 3, \dots$:

$$\lim_{\theta\to 0^+} p_X(k) = \lim_{\theta\to 0^+} \frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^k}\right]}{e^{\theta} - 1} = (1-p)^k p.$$

Hence, as $\theta \to 0^+$, the PMF of the GGZTP distribution reduces to

$$p_X(k) = (1-p)^k p, \quad k = 0, 1, 2, \dots, \tag{3.6}$$

which is the PMF for geometric$(p)$. □

For a fixed value of $p$, plots showing the PMFs for the geometric and GGZTP distributions are presented in Fig. 1. As can be seen in Fig. 1, the PMF plots for both distributions are almost identical, especially for small values of $\theta$. Therefore, the geometric distribution can be regarded as a sub-model of the GGZTP distribution as $\theta \to 0^+$. However, for arbitrary values of $p$, the PMF given in (3.6) is a decreasing function of $x$ only. In contrast, the PMF defined in Equation (3.5) can be either decreasing or increasing/decreasing, depending on the choice of parameters. Plots for the PMF and CDF of the GGZTP distribution for selected values of $(\theta, p)$ are presented in Figs. 2 and 3, respectively.

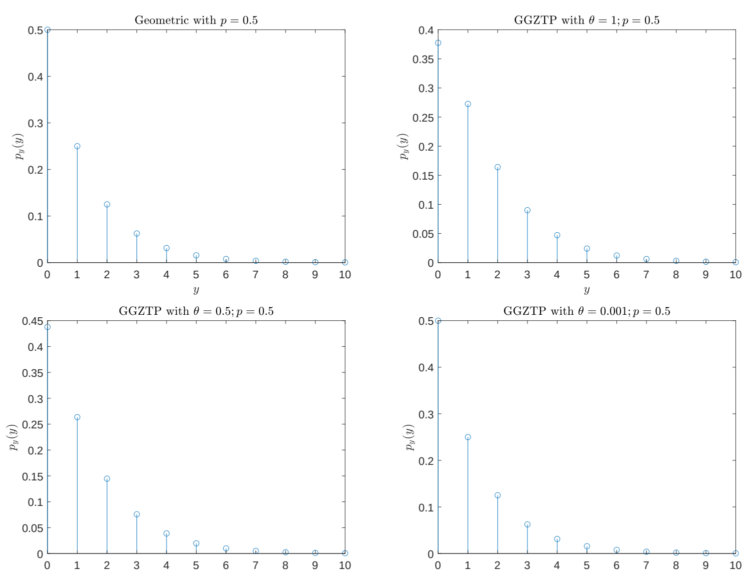


FIGURE 1. Plots for the PMFs for the geometric and GGZTP distributions for $p = 0.5$.

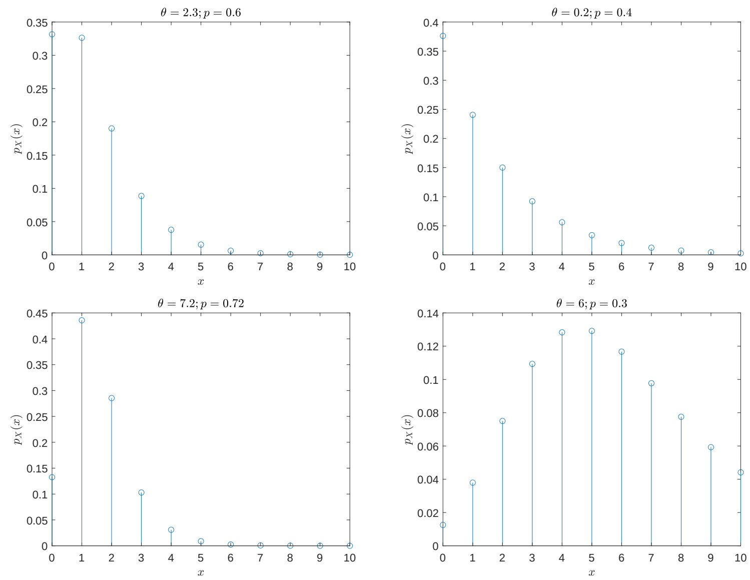


FIGURE 2. Plots for the PMF for the GGZTP distribution for various values of $\theta$ and $p$.

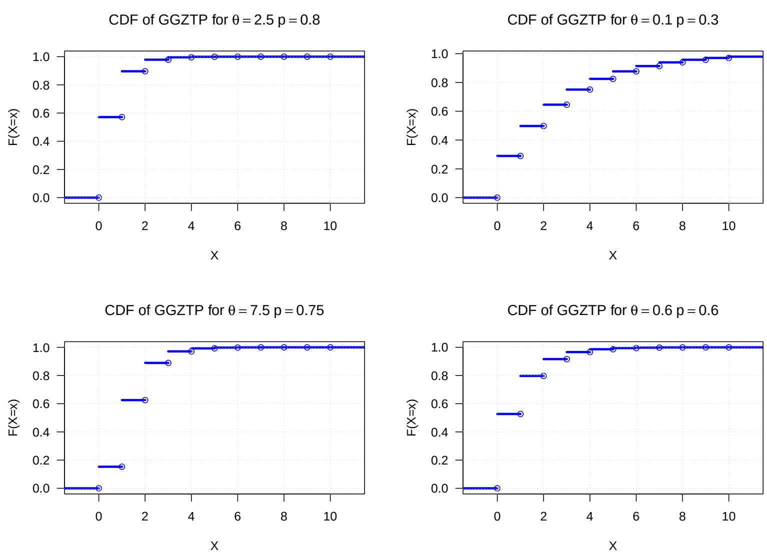


FIGURE 3. Plots for the CDF for the GGZTP distribution for various values of $\theta$ and $p$.

The plots in Fig. 2 confirm that the PMF for the GGZTP distribution is either decreasing or unimodal. Fig. 3 illustrates that the CDF for the GGZTP distribution increases gradually when $\theta < p$; but steeply otherwise, the latter indicating a higher probability mass concentrated in the lower range of values. In [12], the mean and variance of the GGZTP distribution are provided, but not the $k$-th moment. For $X \sim \mathrm{GGZTP}(\theta, p)$, the $m$-th moment for the GGZTP distribution is derived as

$$E(X^m) = \sum_{k=0}^{\infty} k^m p_X(k) \quad = \quad \sum_{k=1}^{\infty} k^m \left( \frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right]}{e^{\theta}-1} \right)$$

$$= \frac{e^{\theta}}{e^{\theta}-1}\sum_{k=1}^{\infty} k^m \left(e^{-\theta(1-p)^k}\right)\left[e^{\theta p(1-p)^k}-1\right]$$

$$= \frac{e^{\theta}}{e^{\theta}-1}\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} k^m(-1)^i g(k,p,i,j), \tag{3.7}$$

where $g(k,p,i,j) = \frac{(\theta(1-p)^k)^{i+j}p^j}{i!j!}$. Although the infinite series in Equation (3.7) does not have a closed-form solution, it can be evaluated numerically using mathematical software [12].

**Theorem 3.2.** *For $0 \le r < 1$, the $r$ quantiles (or $100r$-th percentiles) for GGZTP random variable $X$ vary depending on the condition; i.e.,*

- *If $r \le \frac{e^{\theta p}-1}{e^{\theta}-1}$, then the $r$ quantile is $x_r = 0$.*
- *If $r > \frac{e^{\theta p}-1}{e^{\theta}-1}$, then the $r$ quantile is positive integer $x_r$ satisfying*

$$c-1 \;\le\; x_r \;\le\; c,$$

*where $c = \frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)}$. In particular, if $\frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)}$ is not an integer, then the $r$ quantiles for the GGZTP distribution become*

$$x_r = \left\lfloor \frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)} \right\rfloor,$$

*where $\lfloor\cdot\rfloor$ denotes the floor function.*

*Proof.* The $r$ quantiles for an integer-valued discrete distribution are non-negative integers $x_r$ satisfying

$$\Pr(X < x_r) \le r \le \Pr(X \le x_r).$$

For case $0 \le r \le \frac{e^{\theta p}-1}{e^{\theta}-1}$, it is clear that $x_r = 0$ is the only $r$ quantile satisfying

$$\Pr(X < 0) = 0 \le r \le \frac{e^{\theta p}-1}{e^{\theta}-1} = \Pr(X \le 0).$$

For $\frac{e^{\theta p}-1}{e^{\theta}-1} < r < 1$, we note that inequality $\Pr(X \le x_r) \ge r$ can be written as

$$\frac{\exp\left\{\theta\left(1-(1-p)^{x_r+1}\right)\right\}-1}{e^{\theta}-1} \ge r,$$

which implies

$$x_r \ge \frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)} - 1.$$

Moreover, inequality $\Pr(X < x_r) \le r$ can be written as

$$\frac{e^{\theta p}-1}{e^{\theta}-1} + \sum_{k=1}^{x_r-1}\frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^k}\right]}{e^{\theta}-1} \le r,$$

which implies

$$x_r \le \frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)}.$$

Therefore, $x_r$ is any non-negative integer satisfying

$$c-1 \;\le\; x_r \;\le\; c.$$

where $c = \frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)}$. Furthermore, if $c$ is not an integer, the last inequality simplifies to

$$x_r = \lfloor c \rfloor = \left\lfloor \frac{\ln\left\{1-\frac{\ln\left[r(e^{\theta}-1)+1\right]}{\theta}\right\}}{\ln(1-p)} \right\rfloor.$$

This ends the proof. □

**Theorem 3.3.** *For $X \sim \mathrm{GGZTP}(\theta, p)$, the mode of the GGZTP distribution is given by*

- *if $\frac{e^{\theta p}-1}{e^{\theta}-1} > \frac{e^{\theta}}{e^{\theta}-1} \max_{k\geq 1}\left\{e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right\}$ and $\theta \leq \frac{1}{1-p}$, then mode $= 0$.*
- *if $\frac{e^{\theta p}-1}{e^{\theta}-1} < \frac{e^{\theta}\left[e^{-\theta(1-p)^2}-e^{-\theta(1-p)}\right]}{e^{\theta}-1}$, then mode $= \left\lfloor \frac{\ln(1/\theta)}{\ln(1-p)} \right\rfloor$.*

*where $\lfloor \cdot \rfloor$ denotes the floor function.*

*Proof.* Recall that the mode of a discrete probability distribution can be any local maximum of its PMF. The proof here is given case by case.

**Case 1:** Suppose

$$\frac{e^{\theta p}-1}{e^{\theta}-1} > \frac{e^{\theta}}{e^{\theta}-1} \max_{k\geq 1}\left\{e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right\},$$

then $p_X(0) > p_X(k)$ for all $k \geq 1$, in which case 0 is the mode.

For $x \geq 1$, recall the PMF for the GGZTP distribution:

$$p_X(x) = \frac{e^{\theta}}{e^{\theta}-1}\left(e^{-\theta(1-p)^{x+1}} - e^{-\theta(1-p)^{x}}\right).$$

Let $g(x) = e^{-\theta(1-p)^x}$ such that $p_X(x) \propto g(x+1) - g(x)$. Hence, the mode occurs where the forward difference of $g(x)$ is maximized.

The derivatives are computed as

$$g'(x) = e^{-\theta(1-p)^x}\left(-\theta(1-p)^x \ln(1-p)\right),$$
$$g''(x) = e^{-\theta(1-p)^x}(\ln(1-p))^2\left(\theta^2(1-p)^{2x} - \theta(1-p)^x\right).$$

Thus,

$$g''(x) > 0 \; if \; \theta(1-p)^x > 1 \quad \text{(convex)},$$
$$g''(x) < 0 \; if \; \theta(1-p)^x < 1 \quad \text{(concave)}.$$

Hence, the difference $(g(k+1) - g(k))$ increases while $g$ is convex and decreases after it becomes concave. If $\theta \leq \frac{1}{1-p}$, then $g''(k) < 0$ for all $k \geq 1$, so $p_X(k)$ is decreasing when $k \geq 1$. Therefore, the GGZTP distribution is unimodal with the mode at $0$.

**Case 2:** Suppose

$$\frac{e^{\theta p}-1}{e^{\theta}-1} < \frac{e^{\theta}\left[e^{-\theta(1-p)^2} - e^{-\theta(1-p)}\right]}{e^{\theta}-1},$$

then $p_X(0) < p_X(k)$ for all $k \geq 1$, and hence zero is not a mode.

Let $a = 1-p$ with $0 < a < 1$. For $x \geq 1$, the PMF becomes

$$p_X(x) = \frac{e^{\theta}}{e^{\theta}-1}\left(e^{-\theta a^{x+1}} - e^{-\theta a^{x}}\right).$$

Let $g(x) = e^{-\theta a^x}$ so that $p(x) \propto g(x+1) - g(x)$, then

$$g'(x) = e^{-\theta a^x}\left(-\theta a^x \ln a\right),$$
$$g''(x) = e^{-\theta a^x}(\ln a)^2\left(\theta^2 a^{2x} - \theta a^x\right).$$

Thus, the sign for $g''(x)$ is the same as the sign for $\theta a^x(\theta a^x - 1)$, and hence

$$g''(x) > 0 \; if \; \theta(1-p)^x > 1,$$

$$g''(x) < 0 \; if \; \theta(1-p)^x < 1.$$

Therefore, $p_X(k) = g(k+1) - g(k)$ increases while $\theta(1-p)^k > 1$ and decreases afterward. The maximum occurs near the point where

$$\theta a^k \approx 1.$$

Solving

$$\theta(1-p)^k = 1$$

results in

$$k = \frac{\ln(1/\theta)}{\ln(1-p)}.$$

Since $k$ must be an integer, the mode is

$$\text{mode} = \left\lfloor \frac{\ln(1/\theta)}{\ln(1-p)} \right\rfloor.$$

□

*Remark* 3.4. The mode for the geometric distribution is zero. Combining it with the zero-truncated Poisson distribution enables the established model (i.e., the GGZTP distribution) to accommodate another possible mode, which is influenced by the selected values for parameters $\theta$ and $p$.

3.2. **The GGL Distribution.** First, we show that the GGL distribution belongs to the GCPS family. Let $Y_1, Y_2, \ldots, Y_N$ be i.i.d. random variables following a geometric distribution with parameter $p \in (0,1)$ and let $N$ be a logarithmically distributed random variable with parameter $q \in (0,1)$. As explained in Subsection 3.1, the geometric distribution is a member of the power series family. The logarithmic distribution with parameter $q \in (0,1)$ is a member of the zero-truncated power series family because its PMF can be expressed as in Equation (2.3); i.e.,

$$p_N(n) = \frac{-1}{\ln(1-q)} \frac{q^n}{n} = \frac{b_n \vartheta^n}{c(\vartheta)}, \qquad n = 1, 2, 3, \ldots, \tag{3.8}$$

where, in this case, $\vartheta = q$, $b_n = 1/n$ for all $n \in \{1, 2, 3, \ldots\}$ and

$$c(\vartheta) = \sum_{n=1}^{\infty} \frac{q^n}{n} = -\ln(1-q) = -\ln(1-\vartheta).$$

Therefore, since we have one power series (geometric) distribution and one zero-truncated power series (logarithmic) distribution, then by following the procedure described in Section 2, we can show that the $\mathrm{GGL}(q,p)$ distribution belongs to the GCPS family. The CDF for the GGL distribution can be expressed as in Equation (2.6) as follows:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ -\frac{1}{\ln(1-\vartheta)} \sum_{n=1}^{\infty} \left[ \left( \sum_{j=0}^{\lfloor x \rfloor} \frac{\phi^j}{1/(1-\phi)} \vartheta \right)^n \frac{1}{n} \right], & x \geq 0 \end{cases}, \tag{3.9}$$

where $c(\vartheta) = -\ln(1-\vartheta)$, $g(\phi) = (1-\phi)^{-1}$ and $b_n = 1/n$. Letting $\phi = 1-p$ and $\vartheta = q$ allows us to simplify Equation (3.9). Thus, the simplest expression for the GGL CDF becomes

$$F_X(x) = -\frac{1}{\ln(1-q)} \sum_{n=1}^{\infty} \frac{\left[q\left(1-(1-p)^{\lfloor x \rfloor + 1}\right)\right]^n}{n}. \tag{3.10}$$

Since $q \in (0,1)$ and $1-(1-p)^{\lfloor x \rfloor+1} \in (0,1)$ for all $x \geq 0$, it follows that $q\left(1-(1-p)^{\lfloor x \rfloor+1}\right) \in (0,1)$. Hence, using the Maclaurin series expansion of $-\ln(1-u)$ for $u \in (0,1)$, Equation (3.10) can be rewritten as

$$F_X(x) = \Pr(X \leq x) = \begin{cases} 0, & x < 0 \\ \dfrac{\ln\left(1-q\left[1-(1-p)^{\lfloor x \rfloor+1}\right]\right)}{\ln(1-q)}, & x \geq 0 \end{cases}. \tag{3.11}$$

The corresponding PMF is given by

$$p_X(k) = \begin{cases} \dfrac{\ln(1-qp)}{\ln(1-q)}, & k = 0 \\ \dfrac{1}{\ln(1-q)} \ln\left(\dfrac{1-q\left[1-(1-p)^{k+1}\right]}{1-q\left[1-(1-p)^{k}\right]}\right), & k = 1, 2, \ldots \end{cases}, \tag{3.12}$$

where $p, q \in (0,1)$.

**Theorem 3.5.** *The limiting distribution for* $\mathrm{GGL}(q,p)$ *as* $q \to 0^+$ *is*

$$\lim_{q \to 0^+} F_X(x) = 1-(1-p)^{\lfloor x \rfloor+1},$$

*which is the CDF for the geometric distribution with parameter* $p$.

*Proof.* Consider Equation (3.11) under the condition that $q \to 0^+$. By letting $a(x) = 1-(1-p)^{\lfloor x \rfloor+1}$, we obtain

$$\begin{aligned} \lim_{q \to 0^+} F_X(x) &= \lim_{q \to 0^+} \frac{\ln(1-a(x)q)}{\ln(1-q)} = \lim_{q \to 0^+} \frac{-a(x)/(1-a(x)q)}{-1/(1-q)} \\ &= \lim_{q \to 0^+} \frac{a(x)(1-q)}{1-a(x)q} = a(x) = 1-(1-p)^{\lfloor x \rfloor+1}. \end{aligned}$$

Thus, as $q \to 0^+$, the CDF for the GGL distribution converges to the CDF for the geometric distribution. □

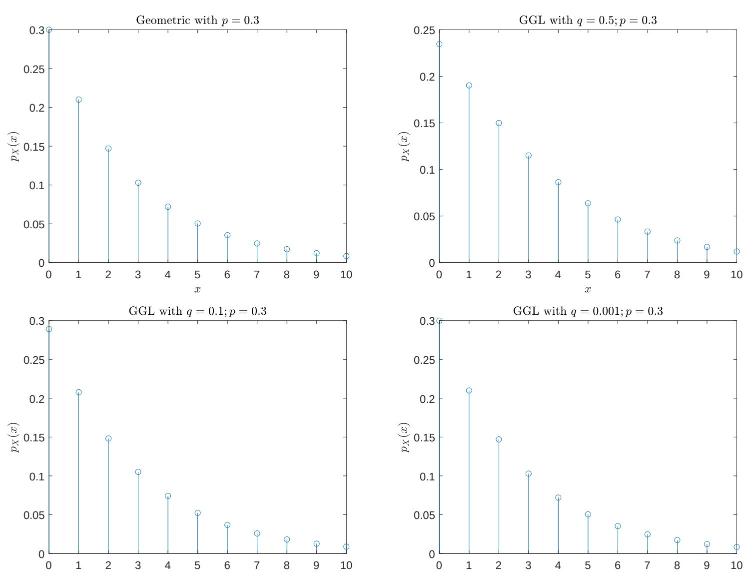


FIGURE 4. Plots for the PMF for the geometric and GGL distributions for $p = 0.3$.

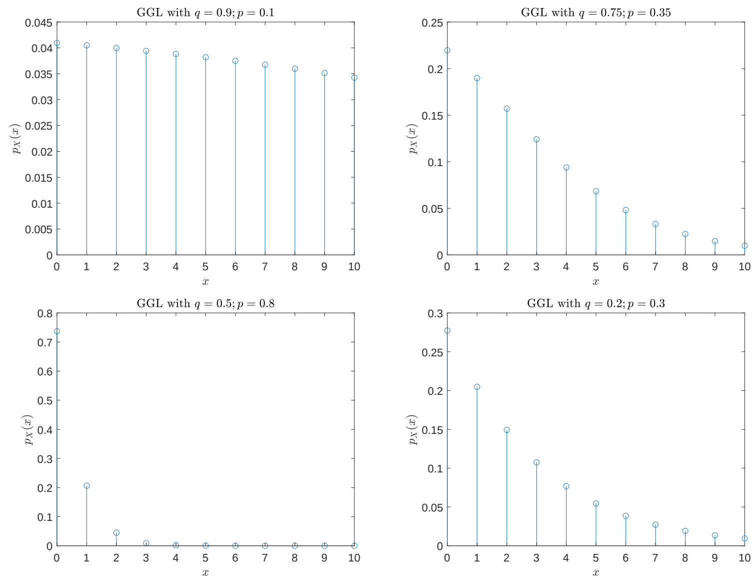


FIGURE 5. Plots for the PMF for the GGL distribution for various values of $q$ and $p$.

Plots for the PMF for the geometric($p$) and GGL($q, p$) distributions for a fixed value of $p$ are presented in Fig. 4. As can be seen in Fig. 4, the PMF plots of both distributions are almost identical, especially for smaller values of $q$. This observation further supports the conclusion that the geometric distribution is a sub-model of the GGL distribution as $q \to 0^+$. Moreover, for any values of $q$ and $p$, the PMF given in Equation (3.12) is a decreasing function of $x$. The plots for the PMF and CDF for the GGL distribution for selected values of $(q, p)$ are presented in Figs. 5 and 6, respectively.

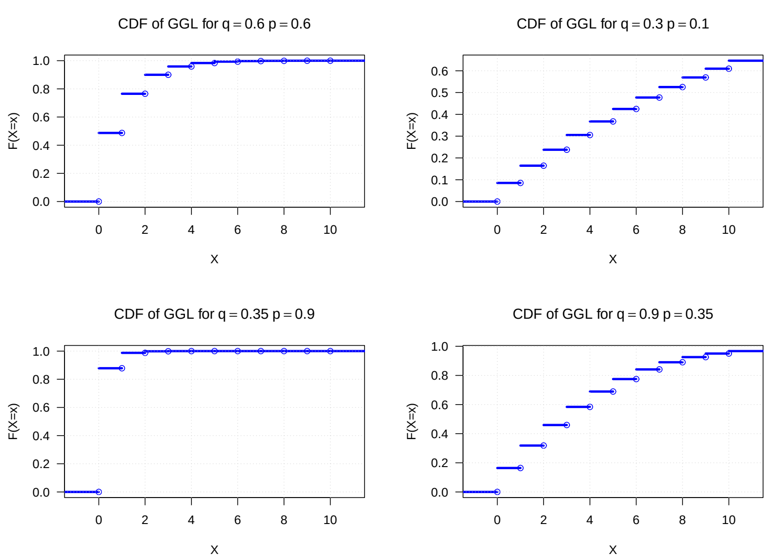


FIGURE 6. Plots for the CDF for the GGL distribution for various values of $q$ and $p$.

The evidence in Fig. 5 confirms that the PMF for the GGL distribution is a decreasing function. Meanwhile, Fig. 6 illustrates that, when $q > p$, the CDF for the GGL distribution increases gradually. Otherwise, the CDF exhibits a steeper increase, which indicates a higher probability mass centered around the lower values of random variable $X$.

The $m$-th moment of the GGL distribution is derived as

$$E(X^m) = \sum_{k=0}^{\infty} k^m p_X(k) = \frac{\sum_{k=1}^{\infty} k^m \ln\left(\frac{1-q\left(1-(1-p)^{k+1}\right)}{1-q(1-(1-p)^k)}\right)}{\ln(1-q)} \tag{3.13}$$

Although the infinite series in Equation (3.13) does not have a closed-form expression, it can be evaluated numerically using mathematical software. From Equation (3.13), the first moment (mean) and second moment of the GGL distribution are respectively given by

$$E(X) = \frac{\sum_{k=1}^{\infty} k \ln\left(\frac{1-q\left(1-(1-p)^{k+1}\right)}{1-q(1-(1-p)^k)}\right)}{\ln(1-q)}, \tag{3.14}$$

and

$$E(X^2) = \frac{1}{\ln(1-q)} \sum_{k=1}^{\infty} k^2 \ln\left(\frac{1-q\left(1-(1-p)^{k+1}\right)}{1-q\left(1-(1-p)^k\right)}\right). \tag{3.15}$$

In addition, the variance for the GGL distribution is defined as

$$\begin{aligned}\mathrm{Var}(X) &= \frac{\sum_{k=1}^{\infty} k^2 \ln\left(\frac{1-q\left(1-(1-p)^{k+1}\right)}{1-q(1-(1-p)^k)}\right)}{\ln(1-q)} \\ &- \left[\frac{1}{\ln(1-q)}\right]^2 \left[\sum_{k=1}^{\infty} k \ln\left(\frac{1-q\left(1-(1-p)^{k+1}\right)}{1-q\left(1-(1-p)^k\right)}\right)\right]^2.\end{aligned} \tag{3.16}$$

Equations (3.14), (3.15), and (3.16) should also be evaluated numerically using mathematical software.

**Theorem 3.6.** *For $0 \le r < 1$, the $r$ quantile (or $100r$-th percentile) for GGL random variable $X$ varies depending on the condition; i.e.,*

- *If $r \le \frac{\ln(1-qp)}{\ln(1-q)}$, then the $r$ quantile is $x_r = 0$.*
- *If $r > \frac{\ln(1-qp)}{\ln(1-q)}$, then the $r$ quantile is a positive integer $x_r$ satisfying*

$$c - 1 \ \le\ x_r \ \le\ c,$$

*where $c = \frac{\ln\left(1-\frac{1-(1-q)^r}{q}\right)}{\ln(1-p)}$. In particular, if $\frac{\ln\left(1-\frac{1-(1-q)^r}{q}\right)}{\ln(1-p)}$ is not an integer, then the $r$ quantile for the GGL distribution becomes*

$$x_r = \left\lfloor \frac{\ln\left(1-\frac{1-(1-q)^r}{q}\right)}{\ln(1-p)} \right\rfloor,$$

*where $\lfloor\cdot\rfloor$ denotes the floor function.*

*Proof.* The $r$ quantiles of an integer-valued discrete distribution are non-negative integers $x_r$ satisfying

$$\Pr(X < x_r) \le r \le \Pr(X \le x_r).$$

For case $0 \le r \le \frac{\ln(1-qp)}{\ln(1-q)}$, it is clear that $x_r = 0$ is the only $r$ quantile satisfying

$$\Pr(X < 0) = 0 \le r \le \frac{\ln(1-qp)}{\ln(1-q)} = \Pr(X \le 0).$$

For $\frac{\ln(1-qp)}{\ln(1-q)} < r < 1$, note that inequality $\Pr(X \le x_r) \ge r$ can be written as

$$\frac{\ln\left(1-q\left(1-(1-p)^{x_r+1}\right)\right)}{\ln(1-q)} \ge r,$$

which implies

$$x_r \geq \frac{\ln\left\{1 - \frac{1-(1-q)^r}{q}\right\}}{\ln(1-p)} - 1.$$

Furthermore, inequality $\Pr(X < x_r) \leq r$ can be defined as

$$\frac{\ln(1-qp)}{\ln(1-q)} + \sum_{k=1}^{x_r-1} \frac{\left[\ln\left(\frac{1-q\left(1-(1-p)^{k+1}\right)}{1-q(1-(1-p)^k)}\right)\right]}{\ln(1-q)} \leq r,$$

which implies

$$x_r \leq \frac{\ln\left\{1 - \frac{1-(1-q)^r}{q}\right\}}{\ln(1-p)}.$$

Therefore, $x_r$ is any positive integer satisfying

$$\frac{\ln\left(1 - \frac{1-(1-q)^r}{q}\right)}{\ln(1-p)} - 1 \leq x_r \leq \frac{\ln\left(1 - \frac{1-(1-q)^r}{q}\right)}{\ln(1-p)},$$

Furthermore, if $\frac{\ln\left(1-\frac{1-(1-q)^r}{q}\right)}{\ln(1-p)}$ is not an integer, the last inequality can be simplified to

$$x_r = \left\lfloor \frac{\ln\left(1 - \frac{1-(1-q)^r}{q}\right)}{\ln(1-p)} \right\rfloor.$$

This ends the proof. □

**Theorem 3.7.** *For $X \sim \mathrm{GGL}(q,p)$, the mode of the GGL distribution is zero.*

*Proof.* Recall that the mode of a discrete probability distribution is any local maximum of its PMF. For $r = 1-p$, we obtain

$$p_X(0) = \frac{\ln(1-qp)}{\ln(1-q)}$$

and

$$p_X(1) = \frac{1}{\ln(1-q)} \ln\left(\frac{1-q(1-r^2)}{1-q(1-r)}\right).$$

Suppose $p_X(1) \geq p_X(0)$. Since $\ln(1-q) < 0$, this inequality implies

$$\ln\left(\frac{1-q(1-r^2)}{1-q(1-r)}\right) \leq \ln(1-qp),$$

which gives

$$\frac{1-q(1-(1-p)^2)}{1-q(1-p)} \leq 1-qp,$$
$$\Leftrightarrow 1-q(1-(1-p)^2) \leq (1-qp)^2,$$
$$\Leftrightarrow -q(-p^2+2p) \leq -2qp + q^2p^2,$$
$$\Leftrightarrow qp^2 \leq q^2p^2,$$

which implies $1 \leq q$.

However, this never holds because $q \in (0,1)$. Thus,

$$p(1) < p(0),$$

which makes $k = 0$ a local maximum.

For $k = 1, 2, 3, \ldots$, we can define

$$A_k = \frac{1 - q\left[1 - (1-p)^{k+1}\right]}{1 - q\left[1 - (1-p)^{k}\right]}.$$

Subsequently, the PMF for the GGL distribution can be simplified to

$$p_X(k) = \frac{1}{\ln(1-q)} \ln(A_k).$$

Note that $A_{k+1} \geq A_k$ because $(1-p)^k$ is decreasing in $k$. Therefore, $\ln(A_x)$ increases with $x$. However, since $\ln(1-q) < 0$, PMF $p_X(k)$ is strictly decreasing for $k \geq 1$.

Consequently,

$$p_X(0) > p_X(1) > p_X(2) > \cdots,$$

and hence the mode is zero, which is unique mode for the GGL distribution. □

*Remark* 3.8. The mode of the GGL distribution is similar to the mode of the geometric distribution (i.e., zero). Consequently, combining the geometric distribution with the logarithmic distribution preserves the mode of the original geometric distribution.

3.3. **The IOD for the GGZTP and GGL Distributions.** As dispersion behavior reflects the flexibility of a discrete distribution, studying it is essential and is therefore presented in this subsection. Dispersion behavior describes how the observed counts vary or scatter around the mean. This behavior is quantified using the index of dispersion (IOD), defined as the ratio of the variance to the mean. As mentioned earlier, the three types of dispersion behavior are underdispersion, equidispersion, and overdispersion [13].

A discrete distribution is said to be flexible if it can exhibit all three behaviors [13]. To demonstrate the flexibility of the GGZTP distribution, Kafi et al. [12] derived its IOD, along with its mean and variance. The IOD for the GGZTP distribution is given by

$$\text{IOD} = \frac{\frac{(E(X))^2(e^{\theta p}-1)}{e^{\theta}-1} + \sum_{k=1}^{\infty}(k - E(X))^2 \left(\frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right]}{e^{\theta}-1}\right)}{\sum_{k=1}^{\infty} k \left(\frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right]}{e^{\theta}-1}\right)}. \tag{3.17}$$

Here, $E(X)$ denotes the expected value (mean) of GGZTP distributed random variable $X$. Since Equation (3.17) does not allow a closed-form expression, numerical simulations must be performed to obtain the IOD values for selected values of $\theta$ and $p$; the results reported in Table 1 indicate that the IOD values can be less than, equal to, or greater than one, so the GGZTP distribution is flexible.

For the GGL distribution, the ratio of the variance to the mean leads to the IOD, which is given by

$$\text{IOD} = \frac{\frac{1}{\ln(1-q)}\sum_{k=1}^{\infty} k^2 \left[\ln\left(\frac{1 - q\left(1-(1-p)^{k+1}\right)}{1 - q\left(1-(1-p)^{k}\right)}\right)\right] - \left(\frac{1}{\ln(1-q)}\right)^2 \left(\sum_{k=1}^{\infty} k \left[\ln\left(\frac{1 - q\left(1-(1-p)^{k+1}\right)}{1 - q\left(1-(1-p)^{k}\right)}\right)\right]\right)^2}{\frac{1}{\ln(1-q)}\sum_{k=1}^{\infty} k \left[\ln\left(\frac{1 - q\left(1-(1-p)^{k+1}\right)}{1 - q\left(1-(1-p)^{k}\right)}\right)\right]}. \tag{3.18}$$

Equation (3.18) must be evaluated using mathematical software. Similar to the GGZTP distribution, numerical simulations are required to obtain the IOD values for the GGL distribution for selected values of $q$ and $p$, the results for which are reported in Table 2; it can be seen that the IOD values can be less than, equal to, or greater than one, so the GGL distribution is capable of exhibiting underdispersion, equidispersion, and overdispersion, and is thus flexible.

*Remark* 3.9. Members of the power series family, such as the Poisson, negative binomial, and geometric distributions, can only capture a single type of dispersion behavior, whereas members of the GCPS family are capable of accommodating underdispersed, overdispersed, and equidispersed data within a unified framework and are thus flexible.

TABLE 1. The mean, variance, and IOD of the GGZTP distribution for various values of $(\theta, p)$

| $\theta$ | $p$ | Mean | Variance | IOD |
|---|---|---|---|---|
| 9 | 0.83 | 1.0649 | 0.619918 | 0.582137 |
| 8.7 | 0.87 | 0.83595 | 0.504364 | 0.603346 |
| 6.5 | 0.85 | 0.78561 | 0.552475 | 0.703243 |
| 0.000001 | 0.9999 | 0.0001 | 0.00010002 | 1.0001 |
| 2 | 0.9 | 0.23512 | 0.236476 | 1.005813 |
| 0.1 | 0.5 | 1.03357 | 2.055363 | 1.988606 |
| 0.05 | 0.25 | 3.04298 | 12.15267 | 3.993675 |
| 4 | 0.12 | 15.1774 | 97.08553 | 6.396717 |

TABLE 2. The mean, variance, and IOD of the GGL distribution for various values of $(q, p)$

| $q$ | $p$ | Mean | Variance | IOD |
|---|---|---|---|---|
| 0.98 | 0.98 | 0.179679 | 0.157731 | 0.877846 |
| 0.95 | 0.9 | 0.420463 | 0.389321 | 0.925934 |
| 0.75 | 0.98 | 0.042915 | 0.044717 | 0.999066 |
| 0.75 | 0.95 | 0.106492 | 0.107099 | 1.005704 |
| 0.75 | 0.8 | 0.442213 | 0.506671 | 1.145762 |
| 0.2 | 0.5 | 1.07674 | 2.127881 | 1.976225 |
| 0.5 | 0.2 | 4.83093 | 23.90692 | 4.948719 |
| 0.95 | 0.15 | 11.6799 | 83.02194 | 7.108103 |

## 4. PARAMETER ESTIMATIONS

The maximum likelihood estimation (MLE) method was employed to estimate the values of the GGZTP and GGL distribution parameters. This is because the percentile matching method may lead to non-unique solutions, while the moment-matching method is complicated by the $k$-th moments of both the GGZTP and GGL distributions not having closed-form expressions. Moreover, the Bayesian approach is computationally demanding due to the difficulty in specifying appropriate prior distributions.

4.1. **The MLE for the GGZTP distribution.** Kafi et al. [12] employed the MLE method to estimate the GGZTP distribution parameter values. Let $X_1, X_2, \ldots, X_z$ be an i.i.d. random sample of size $z$ from a GGZTP-distributed population. The log-likelihood function of the GGZTP distribution is given by

$$\ell(\theta, p; X_1, X_2, \ldots, X_z) = z_0 \ln\left(\frac{e^{\theta p} - 1}{e^{\theta} - 1}\right) + \sum_{k=1}^{\infty} z_k h_1(k; p, \theta), \tag{4.1}$$

where $h_1(k; p, \theta) = \ln\left(\frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right]}{e^{\theta}-1}\right)$, $z = \sum_{k=0}^{\infty} z_k$ and $z_k$ denotes the number of observations for which $X = k$. Kafi et al. [12] stated that the values of $\theta$ and $p$ that maximize Equation (4.1) must be obtained numerically; e.g., by using the Newton–Raphson method.

4.2. **The MLE for the GGL distribution.** Let $X_1, X_2, \ldots, X_z$ be an i.i.d. random sample of size $z$ from a GGL distributed population. The values of parameters $q$ and $p$ can be estimated by maximizing the likelihood or log-likelihood functions of the GGL distribution, which are respectively given by

$$L(q, p; X_1, X_2, \ldots, X_z) = \left[\frac{\ln(1 - qp)}{\ln(1 - q)}\right]^{z_0} \prod_{k=1}^{\infty} \left[h_2(k; q, p)\right]^{z_k}, \tag{4.2}$$

and

$$\ell(q, p; X_1, X_2, \ldots, X_z) = z_0 \ln\left(\frac{\ln(1 - qp)}{\ln(1 - q)}\right) + \sum_{k=1}^{\infty} z_k \ln[h_2(k; q, p)], \tag{4.3}$$

where

$$h_2(k; q, p) = \ln\left[\frac{1}{\ln(1-q)} \ln\left(\frac{1 - q\left(1 - (1-p)^{k+1}\right)}{1 - q\left(1 - (1-p)^{k}\right)}\right)\right].$$

Here, $z = \sum_{k=0}^{\infty} z_k$ and $z_k$ denotes the number of observations for which $X = k$. The values of $q$ and $p$ that maximize Equation (4.3) are obtained by simultaneously solving the following score equations:

$$\frac{\partial \ell}{\partial q} = z_0 \left[\frac{1}{(1-q)\ln(1-q)} - \frac{p}{(1-pq)\ln(1-pq)}\right] + \sum_{k=1}^{\infty} z_k \frac{\partial}{\partial q}\left[h_2(k; q, p)\right] = 0, \tag{4.4}$$

and

$$\frac{\partial \ell}{\partial p} = -\frac{z_0 q}{(1-pq)\ln(1-pq)} + \sum_{k=1}^{\infty} z_k \frac{\partial}{\partial p}\left[h_2(k; q, p)\right] = 0. \tag{4.5}$$

The closed-form solutions for Equations (4.4) and (4.5), denoted as $\hat{q}$ and $\hat{p}$, respectively, cannot be obtained. Therefore, the Newton–Raphson method was employed in the present study to approximate the MLEs $\hat{q}$ and $\hat{p}$.

## 5. Numerical Experiments

As discussed in the preceding section, the MLEs of the parameters of the GGZTP and GGL distributions must be obtained using numerical methods. Therefore, numerical simulations are required to evaluate the performance of the MLEs for selected values of the parameters in both distributions and for several sample sizes. To this end, Monte Carlo simulations were repeated $10{,}000$ times for each selected parameter setting. The performances of the MLEs were evaluated using two measures: the average estimate (AE) and the mean squared error (MSE), respectively, defined as

$$\mathrm{AE} = \frac{1}{10{,}000} \sum_{j=1}^{10{,}000} \hat{\tau}_j; \quad \mathrm{MSE} = \frac{1}{10{,}000} \sum_{j=1}^{10{,}000} (\hat{\tau}_j - \tau)^2,$$

where $\hat{\tau}_j$ denotes the estimated value of the parameter in the $j$th Monte Carlo replication, and $\tau$ represents the true value of the parameter.

For the GGZTP distribution, the robustness of the MLEs was investigated via additional simulations in the present study using different selected values of $\theta$ and $p$ over those performed by Kafi et al. [12]. Following procedures similar to theirs, numerical simulations were also carried out for the GGL distribution. To this end, sample sizes of $100$, $500$, and $1000$ were generated from the GGL distribution for cases $q < p$, $q = p$, and $q > p$, representing underdispersed, equidispersed, and overdispersed count data, respectively. The numerical results for GGZTP and GGL are provided in Tables 3 and 4, respectively.

TABLE 3. Numerical results for estimating the GGZTP parameter values for simulated data of size $z$

| Parameter | Average Estimate | | | MSE | | |
|---|---|---|---|---|---|---|
| | $z = 100$ | $z = 500$ | $z = 1000$ | $z = 100$ | $z = 500$ | $z = 1000$ |
| $\theta = 1.45$ | 1.5465 | 1.4665 | 1.4588 | 0.6084 | 0.1253 | 0.0620 |
| $p = 0.68$ | 0.6823 | 0.6804 | 0.6801 | 0.0031 | 0.0007 | 0.0003 |
| $\theta = 4$ | 4.1109 | 4.0257 | 4.0089 | 0.5941 | 0.1046 | 0.0536 |
| $p = 0.12$ | 0.1211 | 0.1203 | 0.1201 | 0.0001 | 0.0000 | 0.0000 |
| $\theta = 0.32$ | 0.4076 | 0.3647 | 0.3474 | 0.0556 | 0.0209 | 0.0124 |
| $p = 0.65$ | 0.6585 | 0.6540 | 0.6523 | 0.0013 | 0.0002 | 0.0001 |
| $\theta = 0.87$ | 0.9523 | 0.8776 | 0.8741 | 0.2780 | 0.0536 | 0.0260 |
| $p = 0.10$ | 0.1023 | 0.1004 | 0.1001 | 0.0002 | 0.0000 | 0.0000 |

TABLE 4. Numerical results for estimating the GGL parameter values for simulated data of size $z$

| Parameter | Average Estimate | | | MSE | | |
|---|---|---|---|---|---|---|
| | $z = 100$ | $z = 500$ | $z = 1000$ | $z = 100$ | $z = 500$ | $z = 1000$ |
| $q = 0.98$ | 0.9378 | 0.9727 | 0.9766 | 0.0252 | 0.0011 | 0.0004 |
| $p = 0.98$ | 0.9782 | 0.9793 | 0.9797 | 0.0010 | 0.0002 | 0.0001 |
| $q = 0.75$ | 0.7987 | 0.6895 | 0.7119 | 0.0559 | 0.0910 | 0.0536 |
| $p = 0.98$ | 0.9846 | 0.9803 | 0.9803 | 0.0002 | 0.0001 | 0.0001 |
| $q = 0.50$ | 0.5435 | 0.5044 | 0.4999 | 0.0381 | 0.0073 | 0.0035 |
| $p = 0.20$ | 0.2112 | 0.2016 | 0.2005 | 0.0010 | 0.0001 | 0.0001 |
| $q = 0.95$ | 0.9545 | 0.9509 | 0.9505 | 0.0005 | 0.0001 | 0.0000 |
| $p = 0.15$ | 0.1581 | 0.1512 | 0.1506 | 0.0005 | 0.0001 | 0.0000 |

Based on the results, the average estimates of the parameters were generally close to their true values, with the distance never exceeding 0.1. Moreover, the average estimates were close to the true parameter values as the sample size was increased. This was also the case for the MSE: as the sample size was increased, the MSE decreased. Therefore, we can conclude that the proposed estimators were suitable for the point estimation of each parameter.

## 6. Application to Real Data

This section presents the results of applying the GGZTP and GGL distributions to four real datasets to illustrate the usefulness of the distributions and to provide a comparison of their efficacies with those of the Conway-Maxwell Poisson (CMP) and PG distributions. The respective PMFs for these competing models are as follows:

$$p_X(k) = \frac{\lambda^k}{(k!)^\nu Z(\lambda,\nu)}, \quad k = 0, 1, 2, \ldots,$$

where

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}; \quad \lambda > 0 \text{ and } \nu > 0.$$

and

$$p_X(k) = \begin{cases} 1-\alpha(1-p)/p, & k = 0, \\ \alpha(1-p)^k, & k = 1, 2, \ldots, \end{cases}$$

where $0 < p < 1$ and $0 \leq \alpha \leq p/(1-p)$.

The first dataset comprises the number of strikes in the UK coal mining industry over 4-week periods during 1948–1959 [14], the second is the number of sperm cells in sea urchins' eggs with respect to fertilization [15], the third is the claim frequency for 10,814 automobile insurance policyholders in a Turkish insurance company recorded between 2012 and 2014 [16], and the fourth is the number of dicentric chromosomes observed after exposure to radiation after being exposed to 0.6 Sv [4]. Datasets 1 and 2 are listed in Table 5, while datasets 3 and 4 are listed in Table 6.

Table 5. Details of datasets 1 and 2

| Dataset 1 | | Dataset 2 | |
|---|---|---|---|
| Number of Strikes | Observed Freq. | Number of Sperm Cells | Observed Freq. |
| 0 | 46 | 0 | 28 |
| 1 | 76 | 1 | 44 |
| 2 | 24 | 2 | 7 |
| 3 | 9 | 3 | 1 |
| 4 | 1 | | |
| Total | 156 | Total | 80 |
| Mean | 0.9935897 | Mean | 0.7625 |
| Variance | 0.7418941 | Variance | 0.4365506 |
| IOD | 0.7466805 | IOD | 0.5725254 |

TABLE 6. Details of datasets 3 and 4

| Dataset 3 | | Dataset 4 | |
|---|---|---|---|
| Number of Claims | Observed Freq. | Number of Chromsms | Observed Freq. |
| 0 | 8,544 | 0 | 473 |
| 1 | 1,796 | 1 | 119 |
| 2 | 370 | 2 | 34 |
| 3 | 81 | 3 | 3 |
| 4 | 22 | 4 | 2 |
| 5 | 1 | | |
| Total | 10,814 | Total | 631 |
| Mean | 0.2655817 | Mean | 0.3232964 |
| Variance | 0.3347128 | Variance | 0.3937263 |
| IOD | 1.260300691 | IOD | 1.217849 |

The results in Table 5 indicate that datasets 1 and 2 exhibited underdispersion, while Table 6 indicates that datasets 3 and 4 exhibited overdispersion. The results for fitting datasets 1, 2, 3, and 4 to all four distributions evaluated using adequate measures such as the maximum log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and chi-squared ($\chi^2$) goodness-of-fit with corresponding $p$-values are provided in Tables 7, 8, 9, and 10, respectively.

TABLE 7. Fitting results for dataset 1

| Number of Strikes | Observed Freq. | GGZTP | CMP | PG | GGL |
|---|---|---|---|---|---|
| 0 | 46 | 46.1865 | 47.4922 | 45.9999 | 60.6275 |
| 1 | 76 | 74.8944 | 70.4309 | 78.0645 | 55.3159 |
| 2 | 24 | 26.8862 | 30.6551 | 22.6639 | 32.0205 |
| 3 | 9 | 6.3213 | 6.5134 | 6.5798 | 7.1062 |
| 4 | 1 | 1.3528 | 0.8320 | 1.9103 | 0.8346 |
| Fitted parameters | | $\hat{p} = 0.7913$<br>$\hat{\theta} = 5.7977$ | $\hat{\lambda} = 1.4830$<br>$\hat{\nu} = 1.7686$ | $\hat{p} = 0.7097$<br>$\hat{\alpha} = 1.7237$ | $\hat{q} = 0.997$<br>$\hat{p} = 0.899$ |
| $\chi^2$ | | 1.5540 | 2.9153 | 1.4573 | 13.809 |
| $p$-value | | 0.4598 | 0.2328 | 0.4826 | 0.0010 |
| Log-likelihood | | -187.556 | -188.011 | -187.985 | -193.306 |
| AIC | | 379.111 | 380.023 | 379.971 | 390.611 |
| BIC | | 385.211 | 386.123 | 386.071 | 396.711 |

TABLE 8. Fitting results for dataset 2

| Number of Sperm | Observed Freq. | GGZTP | CMP | PG | GGL |
|---|---|---|---|---|---|
| 0 | 28 | 28.0111 | 28.2859 | 28.0000 | 37.8941 |
| 1 | 44 | 43.9346 | 42.9323 | 44.3279 | 33.3839 |
| 2 | 7 | 7.2005 | 8.2906 | 6.5402 | 8.4737 |
| 3 | 1 | 0.7671 | 0.4793 | 0.9649 | 0.2439 |
| Fitted parameters | | $\hat{p} = 0.8989$ $\hat{\theta} = 10.3775$ | $\hat{\lambda} = 1.5178$ $\hat{\nu} = 2.9745$ | $\hat{p} = 0.8525$ $\hat{\alpha} = 3.7556$ | $\hat{q} = 0.9998$ $\hat{p} = 0.9825$ |
| $\chi^2$ | | 0.0764 | 0.7960 | 0.0360 | 8.559 |
| $p$-value | | 0.7822 | 0.3723 | 0.8495 | 0.0034 |
| Log-likelihood | | -77.2566 | -77.4821 | -77.3193 | -80.8847 |
| AIC | | 158.5131 | 158.9642 | 158.6387 | 165.7695 |
| BIC | | 163.2772 | 163.7282 | 163.4027 | 170.5335 |

TABLE 9. Fitting results for dataset 3

| Number of Claims | Observed Freq. | GGZTP | CMP | PG | GGL |
|---|---|---|---|---|---|
| 0 | 8544 | 8544.2100 | 8542.8500 | 8544.0004 | 8544.2170 |
| 1 | 1796 | 1793.8290 | 1795.7330 | 1794.1850 | 1793.8260 |
| 2 | 370 | 376.1750 | 376.0100 | 376.0791 | 376.1728 |
| 3 | 81 | 78.8668 | 78.6035 | 78.8299 | 78.8659 |
| 4 | 22 | 16.5339 | 16.4031 | 16.5236 | 16.5337 |
| 5 | 1 | 3.4662 | 3.4190 | 3.4635 | 3.4661 |
| Fitted parameters | | $\hat{p} = 0.7904$ $\hat{\theta} = 0.0031$ | $\hat{\lambda} = 0.2102$ $\hat{\nu} = 0.0052$ | $\hat{p} = 0.7904$ $\hat{\alpha} = 0.7915$ | $\hat{q} = 0.0031$ $\hat{p} = 0.7904$ |
| $\chi^2$ | | 3.72347 | 3.79341 | 3.72715 | 3.72358 |
| $p$-value | | 0.2929 | 0.2847 | 0.2925 | 0.2929 |
| Log-likelihood | | -7031.28 | -7031.27 | -7031.28 | -7031.28 |
| AIC | | 14066.56 | 14066.54 | 14066.56 | 14066.56 |
| BIC | | 14081.14 | 14081.11 | 14081.14 | 14081.14 |

TABLE 10. Fitting results for dataset 4

| Number of Chromsm | Observed Freq. | GGZTP | CMP | PG | GGL |
|---|---|---|---|---|---|
| 0 | 473 | 472.3589 | 471.7168 | 472.9999 | 472.1554 |
| 1 | 119 | 123.1001 | 123.7125 | 122.3726 | 123.3557 |
| 2 | 34 | 27.7829 | 28.2024 | 27.5938 | 27.8052 |
| 3 | 3 | 6.0742 | 5.9232 | 6.2221 | 6.0318 |
| 4 | 2 | 1.3189 | 1.1737 | 1.4030 | 1.2973 |
| Fitted parameters | | $\hat{p} = 0.7833$ $\hat{\theta} = 0.3951$ | $\hat{\lambda} = 0.2623$ $\hat{\nu} = 0.2022$ | $\hat{p} = 0.7745$ $\hat{\alpha} = 0.8601$ | $\hat{q} = 0.3361$ $\hat{p} = 0.7854$ |
| $\chi^2$ | | 3.4363 | 3.3991 | 3.5028 | 3.4401 |
| $p$-value | | 0.1794 | 0.1828 | 0.1735 | 0.1791 |
| Log-likelihood | | -463.8967 | -463.7487 | -463.9968 | -463.8804 |
| AIC | | 931.7934 | 931.4973 | 931.9936 | 931.7608 |
| BIC | | 940.6880 | 940.3920 | 940.8883 | 940.6554 |

As can be seen in Tables 7 and 8, the $\chi^2$ goodness-of-fit $p$-values for the GGZTP distribution exceeded the significance level of $0.05$ for datasets 1 and 2, respectively, thereby indicating adequate fitting of the data therein. In contrast, the GGL distribution did not satisfactorily fit these two datasets, as the corresponding $p$-values fall below the $0.05$ threshold. Furthermore, the results in Tables 9 and 10 for datasets 3 and 4, respectively, indicate that the $\chi^2$ goodness-of-fit $p$-values for both the GGZTP and GGL distributions exceeded $0.05$, thereby suggesting that both distributions are suitable for fitting these datasets.

For datasets 1 and 2, the GGZTP distribution is the most suitable model, as it yielded the lowest AIC and BIC values. For dataset 3, although the fitting results are very similar, the CMP distribution performed slightly better than the other three distributions, as indicated by its achieving the lowest AIC and BIC values, although the differences are very small. For dataset 4, the AIC and BIC values for the GGL and GGZTP distributions indicate that they are the second and third best options, respectively. Nevertheless, the results suggest that GGZTP and GGL distributions could be viable alternatives to the CMP distribution for modeling dataset 4.

Furthermore, the insurer may utilize historical data, namely the claim frequency from $10{,}814$ automobile insurance policyholders between 2012 and 2014 [16], to estimate the expected loss for the year 2015 and price accordingly. The gross premium charged to applicants intending to purchase a motor insurance policy is calculated as

$$\text{Gross Premium} = \frac{\text{Pure Premium}}{1 - \text{Total Loadings}}.$$

The pure premium is based on the past claim amount data and is calculated using the expected value; i.e.,

$$\text{Pure Premium} = E(C)\,E(X),$$

where $E(C)$ is the average cost per claim, $E(X)$ is the average number of claims per exposure unit (policy), and total loadings cover administrative costs, commissions, and profit margins [17].

According to the results in Table 9, the insurance data follow either a GGZTP or GGL distribution, as indicated by the $p$-values for both exceeding the significance level of $0.05$. In this case, the total loadings are $30\%$ of the gross premium, and the average cost per claim is the same as in 2024 (31,520 Turkish Lira per claim) [18].

Assuming the number of claims follows a GGZTP distribution with parameter estimates $\hat{p} = 0.7904$ and $\hat{\theta} = 0.00305$, the expected number of claims per policy is

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \frac{e^{\theta}\left[e^{-\theta(1-p)^{k+1}} - e^{-\theta(1-p)^{k}}\right]}{e^{\theta}-1} \\ &= \frac{e^{0.00305}}{e^{0.00305}-1} \sum_{k=1}^{\infty} k \left(e^{-0.00305(1-0.0.7904)^{k+1}} - e^{-0.00305(1-0.0.7904)^{k}}\right) \\ &= 0.265581. \end{aligned}$$

Thus,

$$\text{Pure Premium} = 31{,}520 \times 0.265581 = 8{,}371.11312,$$

and

$$\text{Gross Premium} = \frac{8{,}371.11312}{1-0.3} = 11{,}958.73303.$$

Assuming instead that the insurance data follow a GGL distribution with parameter estimates $\hat{p} = 0.0.7904$ and $\hat{q} = 0.003048$, the expected number of claims is

$$E(X) = \frac{1}{\ln(1-q)} \sum_{k=0}^{\infty} k \ln\left(\frac{1 - q\left(1-(1-p)^{k+1}\right)}{1 - q\left(1-(1-p)^{k}\right)}\right)$$

$$
\begin{aligned}
&= \frac{1}{\ln(1-0.003048)}\sum_{k=1}^{\infty} k\ln\left(\frac{1-0.003048\left(1-(1-0.0.7904)^{k+1}\right)}{1-0.003048\left(1-(1-0.7904)^{k}\right)}\right)\\
&= 0.265581 \quad \text{claims per policy.}
\end{aligned}
$$

Hence,

$$\text{Pure Premium} = 31{,}520 \times 0.265581 = 8{,}371.11312,$$

and

$$\text{Gross Premium} = \frac{8{,}371.11312}{1-0.3} = 11{,}958.73303.$$

The identical gross premium values arose because both models are similar in their limiting distributions: specifically, $\hat{\theta} = 0.00305 \approx 0$ for the GGZTP distribution and $\hat{q} = 0.003048 \approx 0$ for the GGL distribution. This came about because according to Propositions 1 and 2, both distributions reduce to the geometric distribution with parameter $\hat{p} = 0.79036$.

The gross premium obtained covers the expected losses and loadings for the year 2015. It is also important to note that this premium calculation does not vary across individual policyholders unless rating factors such as age or location are incorporated. Under this framework, all policyholders are treated equally while assuming a homogeneous risk pool.

The choice of an appropriate distribution model is of critical importance, as demonstrated in the analysis of claim frequency presented in this section. An inadequate model may fail to capture the underlying characteristics of the data, leading to biased or inefficient estimates. Consequently, the selection of the claim frequency model plays a decisive role in determining the accuracy of gross premium calculations.

## 7. Conclusion

This study introduces a novel family of distributions denoted as GCPS. In particular, the characteristics of the GGZTP and GGL distributions were investigated in detail. An investigation of their IODs confirmed that they can exhibit overdispersion (IOD $> 1$), equidispersion (IOD $= 1$), or underdispersion (IOD $< 1$), thereby demonstrating their flexibility for data count applications. Furthermore, an in-depth exploration of their behavior revealed that the geometric distribution is a sub-model of both.

A comprehensive Monte Carlo simulation study indicated that their MLEs are consistent and efficient, with estimated parameters closely approximating the true values, particularly as the sample size was increased. When applied to real data, the efficacies of the GGZTP and GGL distributions were compared with those of the CMP and PG distributions. While the GGZTP distribution fitted all datasets well, the GGL distribution was only suitable in some cases. The GCPS family offers other discrete distributions that could be investigated in future studies.

**Acknowledgments.** Rahmat Al Kafi, a PhD student in the Applied Statistics Program, Faculty of Science, Chiang Mai University, and a recipient of the CMU Presidential Scholarship, sincerely acknowledges and appreciates the CMU Presidential Scholarship for providing him with full financial support for his studies and research at Chiang Mai University.

**Competing interests.** The authors declare no competing interests.

## References

[1] C. Chesneau, H.S. Bakouch, Y. Akdoğan, K. Karakaya, The Binomial-Discrete Poisson-Lindley Model: Modeling and Applications to Count Regression, Commun. Math. Res. 38 (2021), 28–51. https://doi.org/10.4208/cmr.2021-0045.

[2] S. Shafiq, S. Khan, W. Marzouk, J. Gillariose, F. Jamal, The Binomial–Natural Discrete Lindley Distribution: Properties and Application to Count Data, Math. Comput. Appl. 27 (2022), 62. https://doi.org/10.3390/mca27040062.

[3] P.B. Amjish, K. Jayakumar, On Generalized Gauss Hypergeometric Distribution, Commun. Stat. Theory Methods 55 (2026), 2144–2164. https://doi.org/10.1080/03610926.2025.2543196.

[4] K. Karakaya, A New Discrete Distribution with Applications to Radiation, Smoking, and Health Data, J. Radiat. Res. Appl. Sci. 16 (2023), 100735. https://doi.org/10.1016/j.jrras.2023.100735.

[5] A. Alrumayh, H.A. Khogeer, A New Two-Parameter Discrete Distribution for Overdispersed and Asymmetric Data: Its Properties, Estimation, Regression Model, and Applications, Symmetry 15 (2023), 1289. https://doi.org/10.3390/sym15061289.

[6] A. Nandi, P.J. Hazarika, A. Biswas, G.G. Hamedani, A New Three-Parameter Discrete Distribution to Model Over-Dispersed Count Data, Pak. J. Stat. Oper. Res. 20 (2024), 197–215. https://doi.org/10.18187/pjsor.v20i2.4554.

[7] G. Shmueli, T.P. Minka, J.B. Kadane, S. Borle, P. Boatwright, A Useful Distribution for Fitting Discrete Data: Revival of the Conway–Maxwell–Poisson Distribution, J. R. Stat. Soc. Ser. C Appl. Stat. 54 (2005), 127–142. https://doi.org/10.1111/j.1467-9876.2005.00474.x.

[8] G. Pollio, G. De Luca, A New Flexible Discrete Distribution: Theory and Empirical Evidence, Int. J. Bus. Soc. Sci. 9 (2018), 57–63. https://doi.org/10.30845/ijbss.v9n12p7.

[9] K. Khruachalee, W. Bodhisuwan, A. Volodin, On the Partial-Geometric Distribution: Properties and Applications, Lobachevskii J. Math. 42 (2021), 3141–3149. https://doi.org/10.1134/S1995080222010103.

[10] R.A. Kafi, P. Paksaranuwat, Modified Partial-Geometric Distribution: Properties, Method of Estimations and Numerical Simulation, Pan-Am. J. Math. 4 (2025), 21. https://doi.org/10.28919/cpr-pajm/4-21.

[11] P.C. Consul, New Class of Location-Parameter Discrete Probability Distributions and Their Characterizations, Commun. Stat. Theory Methods 19 (1990), 4653–4666. https://doi.org/10.1080/03610929008830465.

[12] R.A. Kafi, P. Paksaranuwat, Generalized Geometric Zero-Truncated Poisson Distribution: Properties and Numerical Analysis, J. Phys.: Conf. Ser. 3042 (2025), 012005. https://doi.org/10.1088/1742-6596/3042/1/012005.

[13] X. Bektashi, S. Rexhepi, N. Limani–Bektashi, Dispersion of Count Data: A Case Study of Poisson Distribution and Its Limitations, Asian J. Probab. Stat. 19 (2022), 18–28. https://doi.org/10.9734/ajpas/2022/v19i230464.

[14] M.S. Ridout, P. Besbeas, An Empirical Model for Underdispersed Count Data, Stat. Model. 4 (2004), 77–89. https://doi.org/10.1191/1471082X04st064oa.

[15] B.P. Singh, G. Singh, U.D. Das, D.K. Maurya, An Under-Dispersed Discrete Distribution and Its Application, J. Stat. Appl. Probab. Lett. 8 (2021), 205–213. https://doi.org/10.18576/jsapl/080306.

[16] L.S. Sarul, S. Sahin, An Application of Claim Frequency Data Using Zero Inflated and Hurdle Models in General Insurance, J. Bus. Econ. Finance 4 (2015), 732–743. https://doi.org/10.17261/Pressacademia.2015414539.

[17] P. Parodi, *Pricing in General Insurance*, 2nd ed., Chapman & Hall/CRC, Boca Raton, 2023. https://doi.org/10.1201/9781003168881.

[18] Cumhuriyet Gazetesi, 2024'te Ortalama Tazminat Kaskoda Yüzde 61, Zorunlu Trafikte Yüzde 50 Arttı: 4.3 Milyon Dosya Ödendi, 25 February 2025, citing Türkiye Sigorta Birliği, Motorlu Taşıtlar Sigorta İstatistikleri, 2024. https://www.cumhuriyet.com.tr/ekonomi/2024te-ortalama-tazminat-kaskoda-yuzde-61-zorunlu-trafikte-yuzde-50-2303388.