JIRSS (2024)
Vol. 23, No. 01, pp 99-115
DOI: 10.22034/jirss.2024.2031151.1065

Another Generalization of the Geometric Distribution

Vahid Nekoukhou [1] **and Ashkan Khalifeh** [2]

[1] Department of Statistics, Khansar Campus, University of Isfahan, Iran.
[2] Department of Statistics, Yazd University, Yazd, Iran.

Received: 04/06/2024, Accepted: 25/08/2024, Published online: 04/11/2024

Abstract. This paper examines a novel extension of the geometric distribution characterized by two parameters, that is not created based on discretizing existing continuous models. This model, due to its analytical form of the cumulative distribution function and simple structure, can be of interest from mathematical perspectives, particularly in cases where the analysis of stochastic orders is desired. In addition, it is a suitable candidate for analyzing monotone hazard rate discrete data, in view of the fact that its hazard rate function exhibits monotonicity in both increasing and decreasing directions. Additionally, the behavior of the survival function of residual lifetime is briefly addressed. The parameters of the distribution are estimated using the maximum likelihood method, and a real-world data set is scrutinized to assess the distribution's adequacy in providing satisfactory fits.

Keywords. Geometric Distribution, Hazard Rate Function, Infinite Divisibility, Maximum Likelihood Estimation, Residual Lifetime, Stochastic Orders.

MSC: 62F10, 62H10.

1 Introduction

In recent years, there has been a surge of interest in the development and examination of various discrete probability distributions. Many of these efforts have involved discretizing existing continuous distributions. For instance, Gómez-Déniz (2010) investigated a discrete version of a member belonging to the Marshall-Olkin (1997) family of continuous distributions, resulting in a generalization of the geometric distribution. Similarly,

---

Corresponding Author: Vahid Nekoukhou (v.nekoukhou@khc.ui.ac.ir)
Ashkan Khalifeh (khalifeh68@yahoo.com).

Nekoukhou et al. (2013) introduced the discrete generalized exponential (DGE) distribution, which serves as a discrete counterpart to the widely recognized continuous generalized exponential distribution proposed by Gupta and Kundu (1999). The DGE distribution can be considered as an exponentiated geometric distribution which contains the geometric distribution as a special case. Furthermore, the discrete Weibull distribution introduced by Nakagawa and Osaki (1975) can be regarded as a discrete analogue of the continuous Weibull distribution. The geometric distribution is a special case of the discrete Weibull distribution. Moreover, Gillariose et al. (2022) discusses a discrete Kumaraswamy Marshall-Olin exponential distribution as a generalization of the geometric distribution. Jayakumar and Sankaran (2017) explores a discrete generalized Marshall-Olin exponential distribution, which includes sub-models of the generalized and geometric distributions. Altun (2020), and Rattihalli and Rattihalli (2023) introduced some new generalizations of the geometric distribution and studied their properties and applications. Krishna and Pundir (2007) introduced the discrete Maxwell distribution, and Krishna and Pundir (2009) introduced the discrete Burr distribution, investigating a special case leading to the discrete Pareto distribution. Jazi et al. (2010) examined the discrete inverse Weibull distribution and proposed significant properties of their model. Gómez-Déniz and Calderin-Ojeda (2011) explored the discrete Lindley distribution, investigating its properties and applications. Additionally, Hussain and Ahmad (2014) introduced the discrete inverse Rayleigh distribution.

Within this context, arises a crucial question: Can we develop a discrete distribution that possesses unique characteristics, independent of existing continuous distributions? Nekoukhou and Bidram (2020) gave a positive response to this question while they studied a new discrete distribution based on geometric odds ratio. In addition, Akdogan et al. (2019), using the compounding method, studied an extension of the geometric distribution. Here, we propose a discrete distribution as a generalization of the geometric distribution, termed the new generalized geometric (NGG) distribution. This distribution, introduced in this paper, features an analytical cumulative distribution function (CDF) whose probability mass function (PMF) exhibits either log-concave or log-convex behavior, depending on parameter values. Consequently, the hazard rate function of an NGG distribution shows monotonic behavior, either increasing or decreasing.

The remaining sections of this paper are structured as follows. Section 2 introduces the NGG distribution, discussing its key features and properties such as the CDF, survival function, mean, and variance. Furthermore, the behavior of the hazard rate function is analyzed, demonstrating that the NGG distribution exhibits monotonically increasing and decreasing hazard rates. The section also explores the infinite divisibility of the distribution and investigates various stochastic orders. In Section 3, the estimation process of the distribution parameters is presented. Moreover, the application of the NGG distribution to real-world data, specifically concerning a renowned clothing brand, is examined in this section. Finally, Section 4 provides concluding remarks summarizing the findings and contributions of the paper.

2 The Proposed Distribution

2.1 Definition and Interpretations

Let X represent the number of failures before the first success in a Bernoulli experiment with success probability p. In this scenario, the survival function (SF) of X is given by:

$$S(x) = P(X \geq x) = q^x, \quad x \in \mathbb{N}_0 = 0, 1, 2, \ldots, \tag{2.1}$$

where $q = 1 - p$. This is known as the geometric survival function, and X follows a geometric distribution denoted by $X \sim \mathrm{G}(q)$.

The CDF of the geometric random variable X, given by $F(x) = P(X \leq x)$, is:

$$F(x) = 1 - S(x+1) = 1 - q^{x+1}, \quad x \in \mathbb{N}_0. \tag{2.2}$$

These equations define the geometric distribution of X, where q is the probability of failure and p is the probability of success.

Definition 2.1. A discrete random variable X follows the new generalized geometric (NGG) distribution if its survival function (SF) is defined as:

$$S(x; \theta, q) = \frac{q^x}{\theta^{q^x - 1}}, \quad x \in \mathbb{N}_0, \tag{2.3}$$

where $0 < q < 1$ and $0 < \theta \leq \exp\left(\frac{\log q}{q-1}\right)$ are the model parameters. If q tends to 0, the upper limit on θ relaxes, and as q approaches 1, θ is constrained to be less than or equal to e. We denote an NGG distribution with parameters θ and q as $\mathrm{NGG}(\theta, q)$.

It can be shown that the survival function S in Eq. (2.3) is a bona fide survival function, and its corresponding CDF for non-negative integer values of x is given by:

$$F(x; \theta, q) = 1 - \frac{q^{x+1}}{\theta^{q^{x+1} - 1}}. \tag{2.4}$$

Furthermore, the PMF of an $\mathrm{NGG}(\theta, q)$ distribution, for $x \in \mathbb{N}_0$, is expressed as:

$$f(x; \theta, q) = p_x = \frac{q^x}{\theta^{q^x - 1}} - \frac{q^{x+1}}{\theta^{q^{x+1} - 1}}. \tag{2.5}$$

When $\theta = 1$, the distribution simplifies to:

$$p_x = (1 - q) q^x, \quad x \in \mathbb{N}_0,$$

which is the well-known geometric distribution.

Figure 1 depicts examples of probability mass functions of NGG distributions for various values of θ and q.

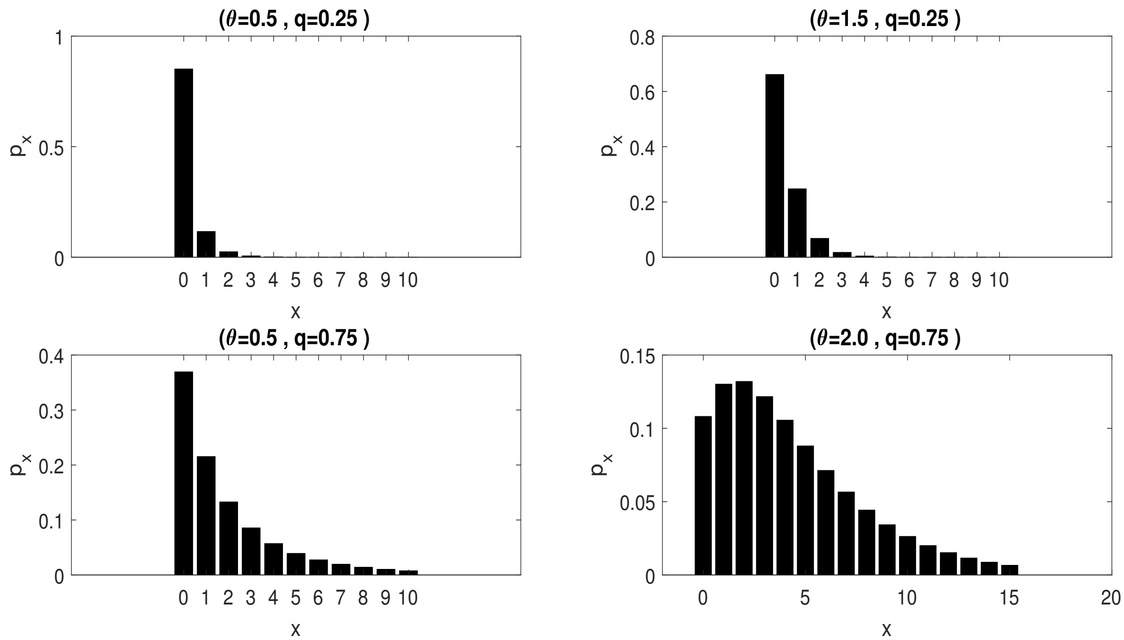

Figure 1: PMF plots of an NGG(θ, q) distribution for some parameters values.

The probability generating function of the random variable X following an NGG(θ, q) distribution, for $|z| < 1$, is given by

$$G_X(z) = \sum_{x=1}^{\infty} \frac{(zq)^x}{\theta^{q^x - 1}} - z^{-1} \sum_{x=1}^{\infty} \frac{(zq)^{x+1}}{\theta^{q^{x+1}-1}}.$$

The r-th moment of the NGG(θ, q) distribution can be expressed as follows:

$$\begin{aligned} E(X^r) &= \sum_{x=1}^{\infty} \{x^r - (x-1)^r\} P(X \geq x) \\ &= \sum_{x=1}^{\infty} \{x^r - (x-1)^r\} \frac{q^x}{\theta^{q^x - 1}}. \end{aligned} \tag{2.6}$$

The existence of the r-th moment is justified by the ratio test. More precisely, let us consider

$$a_x = \{x^r - (x-1)^r\} \frac{q^x}{\theta^{q^x - 1}}.$$

It is seen that

$$\begin{aligned}\lim_{x\to\infty}\frac{a_{x+1}}{a_x} &= \lim_{x\to\infty}\frac{\{(x+1)^r - x^r\}\frac{q^{x+1}}{\theta^{q^{x+1}-1}}}{\{x^r-(x-1)^r\}\frac{q^x}{\theta^{q^x-1}}}\\ &= \lim_{x\to\infty}\frac{rx^{r-1}}{rx^{r-1}}\times\frac{\theta^{q^x-1}q^{x+1}}{\theta^{q^{x+1}-1}q^x}\\ &= q\lim_{x\to\infty}\theta^{q^x(1-q)}\\ &= q<1.\end{aligned}$$

Therefore, the series in (2.6) is convergent, and one can conclude that different moments of the random variable X exist. Specifically, the mean and the second moment of X are also exist and given by:

$$E(X)=\theta\sum_{x=1}^{\infty}\frac{q^x}{\theta^{q^x}}, \tag{2.7}$$

and

$$E(X^2)=\theta\sum_{x=1}^{\infty}(2x-1)\frac{q^x}{\theta^{q^x}}, \tag{2.8}$$

respectively.

The mean and variance of the NGG distribution, computed for various parameter values, are presented in Table 1.

Table 1: Mean (above) and variance (below) of NGG(θ, q) distributions for different values of θ and q.

θ/q	0.25	0.5	0.75
0.25	0.1107	0.4087	1.4370
	0.1572	0.8296	5.3298
0.75	0.2649	0.8266	2.5515
	0.3633	1.6926	10.2890
1.00	0.3333	1.0000	3.0000
	0.4444	2.0000	12.0000
1.50	0.4613	1.3138	3.8008
	0.5765	2.4574	14.4685
2.00	0.5815	1.5997	4.5223
	0.6757	2.7512	15.9555

The observed increase in mean with higher values of θ or q can be explained by the following derivatives:

$$\frac{\partial}{\partial\theta}E(X)=\sum_{x=1}^{\infty}\frac{q^x(1-q^x)}{\theta^{q^x}}>0,$$

and

$$\frac{\partial}{\partial q}E(X) = \sum_{x=1}^{\infty} q^{x-1}\theta^{1-q^x} x\,(1 - q^x \log\theta) > 0.$$

Specifically, when $0 < \theta < 1$, the term $1 - q^x \log\theta$ is positive. In the case of $\theta > 1$, it's evident that $x \log q < \log\log^{-1}\theta$, or equivalently, $1 - q^x \log\theta > 0$.

The variance follows a similar pattern. When $\theta = 1$, the mean and variance correspond to those of the geometric distribution.

2.2 Stochastic Orders

Various stochastic orders between random variables X and Y have been introduced in the literature. For a comprehensive overview, readers are referred to Shaked and Shanthikumar (2007). In the context of this paper, we will review some relevant notions of stochastic orders in the discrete setup

Simple stochastic order: X is said to be stochastically smaller than Y (written as $X \le_{st} Y$) if for all integer values of x, $S_X(x+1) \le S_Y(x+1)$, where S_X and S_Y are the SFs of X and Y, respectively.

Hazard rate order: X is smaller than Y in hazard rate order (written as $X \le_{hr} Y$), if $h_X(x) \ge h_Y(x)$, such that h_X and h_Y are the hazard rate functions of X and Y, respectively.

It is known that, $X \le_{hr} Y \Rightarrow X \le_{st} Y$; see, Shaked and Shanthikumar (2007).

The family of univariate NGG distributions exhibits the following stochastic ordering properties.

Theorem 2.1. *Let $X \sim NGG(\theta, q_1)$ and $Y \sim NGG(\theta, q_2)$,*
i) If $q_1 < q_2$ and $0 < \theta < 1$, then $X \ge_{hr} Y$.
ii) If $q_1 < q_2$ and $\theta > 1$, then $X \le_{hr} Y$.

Proof. i) In this case, $\theta^{q_1^x} \ge \theta^{q_2^x}$ and $-q_1\theta^{q_1^x} \le -q_2\theta^{q_2^x}$ which implies that

$$1 - \frac{q_1\theta^{q_1^x}}{\theta^{q_1^{x+1}}} \le 1 - \frac{q_2\theta^{q_2^x}}{\theta^{q_2^{x+1}}}.$$

The proof of part (ii) is similar to that of (i) and hence the details are avoided. □

Corollary 2.1. *Under the conditions of Theorem 2.1, we have the following results:*
i) If $q_1 < q_2$ and $0 < \theta < 1$, then $X \ge_{st} Y$.
ii) If $q_1 < q_2$ and $\theta > 1$, then $X \le_{st} Y$.

Theorem 2.2. *Let $X \sim NGG(\theta_1, q)$ and $Y \sim NGG(\theta_2, q)$. If $\theta_1 < \theta_2$, then $X \le_{hr} Y$.*

Proof. The proof of this theorem is similar to that of Theorem 2.1 and hence the details are relaxed. □

Corollary 2.2. *Under the conditions of Theorem 2.2, we can conclude that* $X \leq_{st} Y$.

2.3 Survival Analysis

In this section, our focus is on examining the behavior of the hazard rate function. We begin by establishing the unimodal property of an NGG(θ, q) distribution with the following result:

Theorem 2.3. *The PMF of an NGG*(θ, q) *distribution is log-concave for* $\theta \geq 1$*, and log-convex for* $\theta \leq 1$.

Proof. By straightforward calculations, one can show that for all values of x and for $\theta \geq 1$, $p_x^2 \geq p_{x-1}p_{x+1}$ and for $\theta \leq 1$, $p_x^2 \leq p_{x-1}p_{x+1}$. Therefore, according to Keilson and Gerber (1971), the result is immediately obtained. □

Since log-concave probability mass functions are strongly unimodal (Keilson and Gerber, 1971) and thus unimodal, and also in view of the fact that log-convex probability mass functions are decreasing, we have the following result.

Corollary 2.3. *NGG*(θ, q) *distributions are unimodal for all values of* θ *and* q.

A discrete random variable X with a log-concave PMF, has a non-decreasing hazard rate function. This is so, because by definition it follows that for all $x < z$ and any positive integer m, we have

$$\frac{p_{x+m}}{p_x} \geq \frac{p_{z+m}}{p_z}. \tag{2.9}$$

Thus, in view of Eq. (2.9), we see that

$$p_z(\sum_{m \geq x} p_m) - p_x(\sum_{m \geq z} p_m) = \sum_{m \geq 1}(p_z p_{x+m} - p_x p_{z+m}) \geq 0,$$

which implies that the hazard rate function is non-decreasing. Similarly, when a discrete random variable X has a log-convex PMF, its hazard rate function is proved to be non-increasing. Therefore, we have the following result.

Corollary 2.4. *NGG*(θ, q) *distributions for* $\theta \geq 1$ *have non-decreasing and for* $\theta \leq 1$*, non-increasing hazard rate functions.*

The hazard rate function of the NGG(θ, q) distribution, for $x \in \mathbb{N}_0$, is given by

$$h(x) = \frac{p_x}{S(x)} = 1 - q\frac{\theta^{q^x}}{\theta^{q^{x+1}}}. \tag{2.10}$$

When $\theta = 1$, the hazard rate function of the NGG distribution simplifies to $1 - q$, which coincides with the hazard rate function of the geometric distribution. Figure 2 illustrates the behavior of the hazard rate function of the NGG distribution, confirming that it is both monotonically increasing and decreasing, as stated in Corollary 2.4.

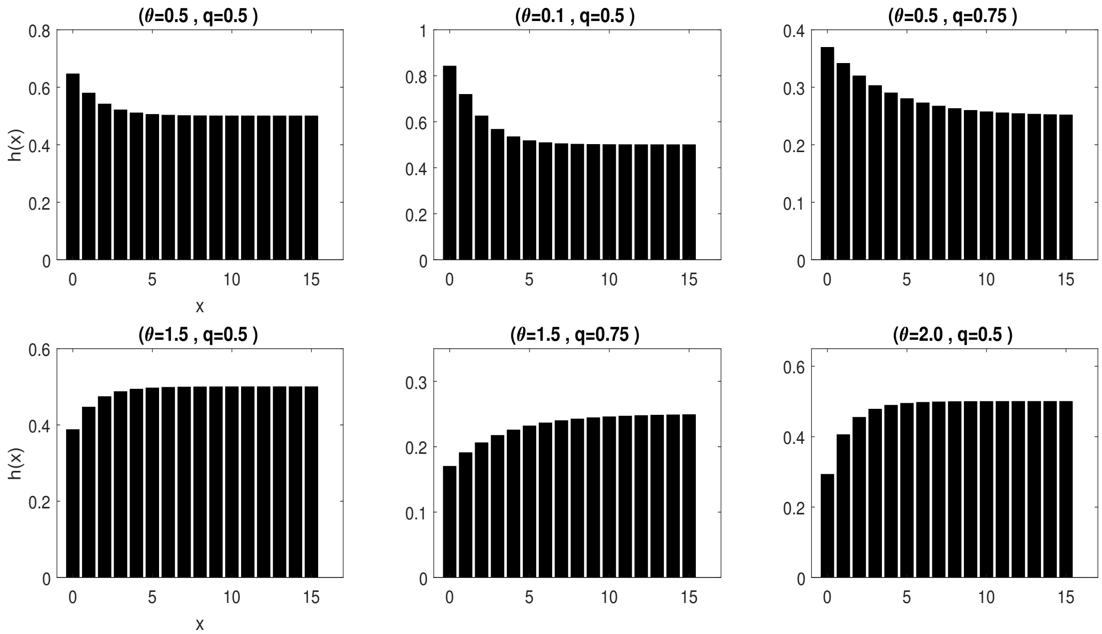

Figure 2: Hazard rate plots of an NGG(θ, q) distribution for some parameters values.

The SF of the residual lifetime of the NGG(θ, q) distribution, for constant values of $t > 0$ and $s > 0$, is also given by

$$P(X \geq t + s | X \geq t) = q^s \theta^{q^t(1-q^s)}.$$

Theorem 2.4. *Suppose that $X \sim NGG(\theta, q)$. If $0 \leq \theta \leq 1$, then for $t > 0$ and $s \in (0, t]$, we have the following result:*

$$P(X \geq t + s | X \geq t) \geq P(X \geq t).$$

Proof. $(q^t - 1)^2 \geq 0$ yields that $2q^t \leq 1 + q^{2t}$. In view of the fact that $s \in (0, t]$, one can conclude that $2q^t \leq 1 + q^{t+s}$. Thus, for $0 \leq \theta \leq 1$, we see that

$$\log \theta \left(2q^t - 1 - q^{t+s}\right) + (s - t) \log q \geq 0,$$

which is equivalent to that of

$$\frac{\theta^{2(q^t-1)} q^s}{\theta^{q^{t+s}-1} q^t} \geq 1,$$

in which implies that

$$P(X \geq t + s | X \geq t) \geq P(X \geq t).$$

□

2.4 Infinite Divisibility

We note the significant structural property of infinite divisibility of the NGG distribution, which bears a close relation to the central limit theorem (CLT) and waiting time distributions. Determining whether a given distribution is infinitely divisible is an essential question in modeling. According to Steutel and van Harn (2004, pp. 56), if p_x, $x \in \mathbb{N}_0$, is infinitely divisible, then $p_x \leq e^{-1}$ for all $x \in \mathbb{N}$. However, for example, in the NGG(3, 0.35) distribution, we find that $p_1 = 0.394 > e^{-1}$. Therefore, we arrive at the following conclusion.

Corollary 2.5. *NGG distributions are not infinitely divisible.*

As we know, the classes of self-decomposable and stable distributions, in their discrete concepts, are subclasses of infinitely divisible distributions. Hence, we conclude that the distributions in question cannot be self-decomposable or stable either.

2.5 Mixing Process

In certain scenarios, it becomes crucial to account for the variation of one or more parameters of a distribution according to a specific probability distribution, known as the mixing distribution. In such scenarios, some particular results which can be useful in different fields of research can be obtained. For example, in problems associated with accident proneness and entomological field data, among others, it is generally accepted that one or more parameters of the distribution involved vary according to certain given probability distribution, called the mixing distribution. A well-known example is based on Poisson distribution mixture with gamma mixing distribution leading to negative binomial distribution. Mixed distributions are appropriate for modeling nonhomogeneous population as occurs, for instance, in actuarial problems (cf. Gómez-Déniz (2010)).

Here, In an NGG(θ, q) distribution, in the case that $0 < \theta < 1$, it is supposed that θ is itself a continuous random variable specified by the generalized beta (GB) distribution, introduced by McDonald (1984), with the following probability density function:

$$\pi(\theta) = \frac{\zeta \theta^{a\zeta - 1}(1 - \theta^{\zeta})^{b-1}}{B(a, b)}, \quad 0 < \theta < 1,$$

where $a > 0$, $b > 0$, $\zeta > 0$ are the model parameters, and $B(a, b)$ is the known beta function.

Theorem 2.5. *If $X|\{\Theta = \theta\} \sim NGG(\theta, q)$, and $\Theta \sim \pi(\theta)$, then the marginal probability mass function of X is given by*

$$m(x) = \frac{q^x B(a + \frac{1-q^x}{\zeta}, b) - q^{x+1} B(a + \frac{1-q^{x+1}}{\zeta}, b)}{B(a,b)}, \quad x \in \mathbb{N}_0. \tag{2.11}$$

Proof. The proof is straightforward and the details are avoided. □

3 Statistical Inference

3.1 Maximum Likelihood Estimation

In this section, we discuss the process of finding the maximum likelihood estimators (MLEs) for the parameters θ and q of the NGG distribution. Due to the complexity of the score equations involved, analytical solutions for the MLEs are not readily available. Therefore, we employ numerical optimization techniques, specifically using the MATLAB *fmincon* function, to maximize the log-likelihood function and estimate the parameters.

To apply the method of maximum likelihood for estimating the unknown parameters of an NGG distribution, we start with the likelihood function for a single observation x_i:

$$L_{[i]}(\theta, q) = \frac{q^{x_i}}{\theta^{q^{x_i}-1}} - \frac{q^{x_i+1}}{\theta^{q^{x_i+1}-1}}. \tag{3.1}$$

The likelihood equations are then given by:

$$\frac{\partial L_{[i]}(\theta, q)}{\partial \theta} = g(x_i + 1; \theta, q) - g(x_i; \theta, q) = 0, \tag{3.2}$$

and

$$\frac{\partial L_{[i]}(\theta, q)}{\partial q} = u(x_i; \theta, q) - u(x_i + 1; \theta, q) - v(x_i; \theta, q) + v(x_i + 1; \theta, q) = 0, \tag{3.3}$$

where

$$g(x; \theta, q) = \frac{q^x(q^x - 1)}{\theta^{q^x}}, \; u(x; \theta, q) = xq^{x-1}\theta^{1-q^x}, \; v(x; \theta, q) = x \log \theta q^{x-1} q^x \theta^{1-q^x}. \tag{3.4}$$

Now, let $x_1, x_2, \ldots, x_n$ be a random sample of size n from the NGG(θ, q) distribution. The total likelihood function is then given by:

$$L_n(\theta, q) = \prod_{i=1}^{n} L_{[i]}(\theta, q), \tag{3.5}$$

where $L_{[i]}(\theta, q)$, $i = 1, 2, \ldots, n$, represents the likelihood function based on the i-th observation. Let $\mathbf{\Omega} = (\theta, q)'$. The MLE of $\mathbf{\Omega}$, denoted by $\hat{\mathbf{\Omega}}$, is obtained by solving the following nonlinear equation

$$\boldsymbol{M}_n = (\partial L_n/\partial\theta, \partial L_n/\partial q) = \mathbf{0}. \tag{3.6}$$

Analytically, solving the system of Equations (3.6) is hindered by the intricate relationships among the parameters, rendering it unfeasible. Consequently, numerical methods are essential for obtaining MLEs. In this context, we employ the likelihood function as the optimization objective. The MATLAB function *fmincon* is particularly advantageous due to its capability in constrained optimization, thereby maintaining parameter estimates within defined bounds. Furthermore, its robust optimization algorithms and convergence criteria are pivotal for obtaining reliable MLEs in the face of potential numerical challenges.

An inherent strength of $fmincon$ lies in its capacity to handle both equality and inequality constraints on the parameters. In our context, we incorporate the constraint $\theta \leq \exp\frac{\log q}{q-1}$ into the optimization process to guarantee adherence to this condition in the estimated parameters. By embedding the constraint within the optimization function, we achieve MLEs that not only maximize likelihood but also conform to the distribution's imposed constraints.

Careful attention is needed when establishing initial parameter values and tackling potential numerical challenges in optimization, including convergence issues and local optima. We propose initializing $\theta_{initial} = 1$ and $q_{initial} = \overline{x}/(\overline{x} + 1)$ as starting points. These initial values offer a proximity to the optimum while ensuring feasibility with respect to the constraint. Notably, when $\theta = 1$, the NGG distribution simplifies to the well-known geometric distribution.

3.2 Simulation Study

In this section, we present the results of a simulation study designed to evaluate the performance of the ML method for estimating the parameters of the NGG distribution.

To investigate how the MLEs behave across various parameter configurations and sample sizes, we opted for the subsequent parameter pairings: sample sizes of $n = 30$, 50, 100, 500, 1000, and the following parameter values: (θ, q) = (5, 0.25), (2, 0.5),(0.5, 0.75), and (0.3, 0.95). These values were arbitrarily selected within the constraints of the parameters.

We performed $N = 1000$ simulation runs to obtain reliable estimates of the mean squared error (MSE). The summary of the simulation results is presented in Table 2, which provides the estimated parameters along with the corresponding MSE for each sample size. For the optimization process using the *fmincon* function, we used $\theta_{initial} = 1$ and $q_{initial} = \overline{x}/(\overline{x} + 1)$ as initial points for all runs.

The point estimates, denoted as $\hat{\theta}$ and $\hat{q}$, are calculated as the mean of $N = 1000$

individual estimates ($\hat{\theta}_i$ and $\hat{q}_i$) derived from the ML procedure as:

$$\hat{\theta} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\theta}_i,$$

and

$$\hat{q} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{q}_i,$$

respectively. Moreover, to assess the precision of our estimates, we calculate the MSEs for each parameter. The MSE values, denoted as MSE(θ) and MSE(q) , are obtained by measuring the average squared discrepancy between the estimated values and the true parameter values. More precisely, we have

$$\mathrm{MSE}(\theta) = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_i - \theta)^2,$$

and

$$\mathrm{MSE}(q) = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{q}_i - q)^2.$$

These measures provide valuable insights into the accuracy and reliability of our estimation methodology based on ML method using MATLAB's *fmincon* function for parameters θ and q. Table 2 indicates the summary of simulated results.

The simulation study confirms the effectiveness of the ML approach for estimating the parameters of the NGG distribution while satisfying the constraint between parameters θ and q. Our findings highlight the importance of considering sample size in the estimation process, as larger samples lead to more precise parameter estimates with lower MSE. Importantly, our results also demonstrate the reliability of the ML method even for small sample sizes. This robust performance of the ML method for small sample sizes is particularly noteworthy, as it underscores the practical utility of the approach in real-world scenarios where limited data may be available. The ability to obtain accurate parameter estimates with small sample sizes is crucial in various applications, specially in fields where data collection may be challenging or expensive.

Table 2: Summary of simulated results.

n	θ	q	$\hat{\theta}$	MSE(θ)	$\hat{q}$	MSE(q)
30	5.00	0.25	5.3674	1.9349	0.2441	0.00318
50	5.00	0.25	5.1842	0.9875	0.2473	0.00211
100	5.00	0.25	5.0874	0.3950	0.2494	0.00100
500	5.00	0.25	5.0133	0.0762	0.2496	0.00020
1000	5.00	0.25	5.0092	0.0363	0.2497	0.00010
30	2.00	0.50	2.1113	0.4981	0.4926	0.00612
50	2.00	0.50	2.0832	0.3060	0.4949	0.00405
100	2.00	0.50	2.0236	0.1424	0.5011	0.00198
500	2.00	0.50	2.0062	0.0276	0.4996	0.00034
1000	2.00	0.50	2.0099	0.0153	0.4990	0.00020
30	0.50	0.75	0.8215	0.4545	0.7029	0.01422
50	0.50	0.75	0.6925	0.2550	0.7260	0.00896
100	0.50	0.75	0.5965	0.1307	0.7398	0.00579
500	0.50	0.75	0.5070	0.0292	0.7516	0.00176
1000	0.50	0.75	0.5032	0.0159	0.7506	0.00079
30	0.30	0.95	0.6795	0.4347	0.9289	0.00153
50	0.30	0.95	0.5755	0.2472	0.9344	0.00095
100	0.30	0.95	0.4362	0.1151	0.9435	0.00046
500	0.30	0.95	0.3289	0.0292	0.9492	0.00016
1000	0.30	0.95	0.3072	0.0181	0.9504	0.00011

3.3 Data Analysis

This section demonstrates the applicability of the NGG distribution through analysis of a real data set. The data set outlines quarterly sales of a well-known brand of clothing at stores of a large national retailer, discussed by Shmueli et al. (2005). Habibi and Asgharzadeh (2018) introduce a mixed Poisson distribution, by blending the Poisson distribution with the binomial exponential 2 distribution of Bakouch et al. (2014), denoted by PBE2. Initially, Habibi and Asgharzadeh (2018) evaluated this data set using the Poisson, geometric, negative binomial (NB) and generalized Poisson-Lindley (GPL) distributions; see also Mahmoudi and Zakerzadeh (2010). Here, the NGG distribution is fitted to this data set as another rival model. Table 3 indicates the observed and expected values of the fitted models. The fit in Table 3 improves the results obtained by other rival models. Hence, the NGG distribution is a suitable candidate to fit such discrete data.

The fitting computations are summarized in Table 4. Comparing the NGG model with its rival models is performed by using the χ^2 statistic, Akaike information criterion (AIC), Bayesian information criterion (BIC) and the values of log-likelihood function. Table 4 indicates the MLEs, values of χ^2 statistic with the corresponding p-values, AICs, BICs and the values of the log-likelihood function, determined by the fitted models.

These results indicate that the NGG distribution provides better fits than other competing models for this data set. In computing the test statistic, we adhered to Cochran's (1954) recommendations for grouping small expected frequencies (less than or equal to 10); see also Greenwood and Nikulin, (1996).

The likelihood ratio (LR) test statistic is used to test the null hypothesis $H_0 : \theta = 1$ (Geometric) versus the alternative hypothesis $H_1 : \theta \neq 1$ (NGG). The value of LR test statistic is 20674 and its corresponding p-value is 0, which means that the null hypothesis is rejected in favor of the alternative hypothesis. Therefore, we have enough evidence to conclude that the NGG distribution is a better choice for analyzing such data set than the geometric distribution.

Table 3: Quarterly sales of a well-known brand of clothing at stores of a large national retailer; $\bar{x} = 3.5597, s^2 = 11.3138$.

Count	Observed	NGG	PBE2	GPL	NB	Geometric	Poisson
0	514	502.61	506.05	505.62	490.92	694.79	90.12
1	503	521.07	521.00	522.66	537.62	542.41	320.81
2	457	473.77	469.65	469.28	480.49	423.45	570.99
3	423	398.78	394.57	393.68	397.18	330.58	677.51
4	326	319.49	317.32	316.67	314.98	258.08	602.93
5	233	247.66	247.69	247.42	243.45	201.48	429.25
6	195	187.68	189.21	189.22	184.90	157.29	254.66
7	139	140.01	142.18	142.35	138.66	122.80	129.50
8	101	103.29	105.47	105.70	102.98	95.87	57.62
9	77	75.61	77.43	77.65	75.91	74.84	22.79
10	56	55.04	56.36	56.54	55.61	58.43	8.11
11	40	39.91	40.73	40.86	40.54	45.61	2.63
12	37	28.86	29.26	29.34	29.43	35.61	0.78
13	22	20.82	20.90	20.95	21.29	27.80	0.21
14	9	15.00	14.87	14.89	15.35	21.70	0.05
15	7	10.80	10.53	10.54	11.04	16.94	0.01
16	10	7.77	7.44	7.43	7.92	13.23	0.00
17	9	5.58	5.23	5.22	5.67	10.23	0.00
18	3	4.01	3.67	3.66	4.05	8.06	0.00
19	2	2.88	2.57	2.55	2.89	6.29	0.00
20	2	2.07	1.80	1.78	2.06	4.91	0.00
21	2	1.49	1.25	1.24	1.47	3.84	0.00
22	0	1.07	0.87	0.86	1.04	2.99	0.00
23	0	0.77	0.60	0.59	0.74	2.34	0.00
24	0	0.55	0.42	0.41	0.52	1.83	0.00
25	0	0.40	0.29	0.28	0.37	1.42	0.00
26	0	0.28	0.20	0.20	0.26	1.11	0.00
27	0	0.20	0.14	0.13	0.19	0.87	0.00
28	0	0.15	0.10	0.09	0.13	0.68	0.00
29	0	0.10	0.07	0.06	0.09	0.53	0.00
30	1	0.08	0.05	0.04	0.07	0.41	0.00
> 30	0	0.00	0.08	0.09	0.18	1.48	0.03

Table 4: Summary.

Models	MLEs	$(\chi^2, \text{d.f.}, p-\text{value})$	log-likelihood	AIC	BIC
NGG	$(\hat{\theta}, \hat{q}) = (1.7552, 0.7179)$	(10.532, 14, 0.73)	-7524.07	15052.14	15064.26
PBE2	$(\hat{\alpha}, \hat{\theta}) = (0.50, 0.88)$	(11.00, 14, 0.68)	-7524.80	15053.70	15065.72
GPL	$(\hat{\alpha}, \hat{\theta}) = (1.17, 0.51)$	(11.30, 14, 0.66)	-7525.00	15053.90	15066.12
NB	$(\hat{r}, \hat{p}) = (1.58, 0.31)$	(13.70, 14, 0.47)	-7526.60	15057.20	15069.32
Geometric	$\hat{q} = 0.22$	(144.30, 15, 0.00)	-7598.60	15199.30	15205.26
Poisson	$\hat{\lambda} = 3.60$	(41828.80, 15, 0.00)	-8959.80	17921.60	17927.66

4 Conclusions

In summary, this paper introduced the new generalized geometric (NGG) distribution as a novel extension of the geometric distribution, aimed at providing enhanced flexibility for analyzing discrete data, particularly in cases of over-dispersion. The NGG distribution features an analytical cumulative distribution function with a probability mass function that exhibits either log-concave or log-convex behavior, contingent upon the values of its parameters. Consequently, the hazard rate function of the NGG distribution demonstrates monotonicity, either increasing or decreasing. Additionally, the paper briefly explored the behavior of the survival function of residual lifetime. Parameter estimation for the NGG distribution was conducted using the maximum likelihood method, with constraints incorporated into the optimization process. An empirical analysis of a real over-dispersed discrete data set revealed that the NGG distribution yielded satisfactory fits for the data under consideration.

Acknowledgments

The authors sincerely thank two anonymous referees and also the Editor-in-Chief for their valuable comments that led to improvement of the paper.

Ethical Conduct

This study was conducted in accordance with the ethical principles, and all participants provided informed consent prior to participation.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability

The data set, used in this study, were obtained from the previously published article: https://doi.org/10.1111/j.1467-9876.2005.00474.x.

References

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, **19**(3), 399-415.

Marshall, A. W., and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, **84**(3), 641-652.

Nekoukhou, V., Alamatsaz, M. H., and Bidram, H. (2013). Discrete generalized exponential distribution of a second type. *Statistics*, **47**(6), 876-887.

Gupta, R. D., and Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, **41**(2), 173-188.

Nakagawa, T., and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, **24**(4), 300-301.

Gillariose, J., Tomy, L., Jamal, F., and Chesneau, C. (2022). A discrete Kumaraswamy Marshall-Olkin exponential distribution. *Journal of the Iranian Statistical Society*, **20**(1), 129-152.

Jayakumar, K., and Sankaran, K. K. (2017). A discrete generalization of Marshall-Olkin scheme and its application to geometric distribution. *Journal of the Kerala Statistical Association*, **28**, 1-21.

Altun, E. (2020). A new generalization of geometric distribution with properties and applications. *Communications in Statistics - Simulation and Computation*, **49**(4), 793-807.

Rattihalli, R. N., and Rattihalli, S. R. (2023). A generalisation of geometric distribution. *Communications in Statistics - Theory and Methods*, **52**(20), 4400-4413.

Krishna, H., and Pundir, P.S. (2007). Discrete Maxwell distribution. *InterStat*.

Krishna, H., and Pundir, P.S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology*, **6**(2), 177-188.

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

Gómez-Déniz, E., and Calderin-Ojeda, E. (2011). The discrete Lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation*, **81**(10), 1405-1416.

Hussain, T., and Ahmad, M. (2014). Discrete Inverse Rayleigh Distribution. *Pakistan Journal of Statistics*, **30**(2), 203-222.

Nekoukhou, V., and Bidram, H. (2020). A new discrete distribution based on geometric odds ratio. *Journal of Statistical Modelling: Theory and Applications*, **2**(1), 153-166.

Akdogan, Y., Kus, C., Bidram, H., and Kinaci, I. (2019). Geometric-zero truncated Poisson distribution: Properties and applications. *Journal of Science*, **32**(6), 1339-1354.

Shaked, M., and Shanthikumar, J.G. (2007). Stochastic orders (Vol 3–44). Springer.

Keilson, J., and Gerber, H. (1971). Some results for discrete unimodality. *Journal of the American Statistical Association*, **66**(335), 386-389.

Steutel, F.W., and van Harn, K. (2004). Infinite Divisibility of Probability Distributions on the Real Line. New York: Marcel Dekker.

McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, **52**(3), 647-664.

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

Habibi, M., and Asgharzadeh, A. (2018). A new mixed Poisson distribution: Modeling and applications. *Journal of Testing and Evaluation*, **46**(6), 1728-1740.

Bakouch, H.S., Jazi, M.A., Nadarajah, S., Dolati, A., and Roozegar, R. (2014). A lifetime model with increasing failure rate. *Journal of Applied Mathematics and Modeling*, **38**(20), 5392-5406.

Mahmoudi, E., and Zakerzadeh, H. (2010). Generalized Poisson–Lindley Distribution. *Communications in Statistics - Theory and Methods*, **39**(10), 1785-1798.

Cochran, W. G. (1954). Some methods for strengthening the Common χ^2 tests. *Biometrics*, **10**(4), 417-451.

Greenwood,P.E., and Nikulin,M.S.(1996)A guide to chi-squared testing.Wiley Hoboken,NJ.