ISSN: 2832-4293

Pan-American Journal of Mathematics 4 (2025), 21
https://doi.org/10.28919/cpr-pajm/4-21

MODIFIED PARTIAL-GEOMETRIC DISTRIBUTION: PROPERTIES, METHOD OF ESTIMATIONS AND NUMERICAL SIMULATION

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

ABSTRACT. This article introduces a modified version of the partial-geometric distribution, derived by raising the partial-geometric cumulative distribution function to the power of a positive real number. Several distributional functions and quantities of the proposed distributions are derived. A significant property of the proposed distribution, other than exhibiting various dispersal behaviors, is its non-constant hazard rate function, a characteristic not present in the original partial-geometric distribution. Several estimation methods for the model parameters are discussed. However, maximum likelihood estimation (MLE) is ultimately employed due to its simplicity and unbiasedness. Numerical studies are conducted to examine the quality of MLE estimators, demonstrating that the average estimate for each parameter tends its true value as the sample size gets larger.

1. INTRODUCTION

The development of probability distributions has long been a central focus in statistical research, particularly in the context of modeling count data. Classic discrete distributions, such as the Poisson, binomial, negative binomial, and geometric, often exhibit limited flexibility in capturing the dispersion characteristics commonly observed in empirical count data. This limitation presents a substantial concern for researchers, as the Poisson distribution is appropriate only for equi-dispersed data, the binomial distribution is suited to under-dispersed data, and both the negative binomial and geometric distributions are typically employed for over-dispersed data.

In recent years, there has been increasing demand for novel discrete distributions capable of capturing the diverse nature of randomness observed in real-world phenomena. To enhance flexibility in modeling such data, many researchers have proposed modified or transformed versions of well-established distributions. Additionally, various discretization techniques, such as those based on survival functions or cumulative distribution functions, have been employed to derive discrete analogues from continuous distributions [1–7]. Others have combined two existing distributions, as seen in the work of [8–14]. However, these distributions are generally suited to specific dispersion behaviors. In contrast, Krisada et al. introduced the partial-geometric (PG) distribution, an extension of the geometric distribution that can model

[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

E-mail addresses: rahmat.alkafi@sci.ui.ac.id, pawat.pak@cmu.ac.th.

Submitted on October 27, 2025.

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

Key words and phrases. maximum likelihood estimation; moment generating function; probability generating function; quantile function.

*Corresponding author.

under-dispersed, equi-dispersed and over-dispersed data [15]. As the PG distribution has closed-form expressions for its cumulative distribution functions (CDF), developing its CDF would be interesting to create a new discrete distribution with the PG as the baseline distribution.

In this article, we introduce a modified partial-geometric (MPG) distribution. The MPG is constructed by exponentiating the CDF of the PG distribution through the inclusion of an additional parameter. The technique of exponentiating a CDF has never been attempted to discrete type distributions, except the geometric distribution by [16]. However, the proposed distibution by [16] does not offer flexibility since it is only suitable for over-dispersed data. Nevertheless, the exponentiating method has been widely applied to continuous distributions, such as the work of [17–22], and has been proven to enhance their flexibility. Herein, the modified partial-geometric distribution provides wider flexibility than the partial-geometric, which is able to model and generate count data that exhibit a variety of dispersal behavior, as well as diverse hazard rate and probability mass function shapes. By an extra parameter: (1) the hazard rate function of MPG can be in the form of increasing and decreasing in addition to constant. (2) the mass function of MPG can be in the form of decreasing-upside-down in addition to decreasing and upside-down.

2. Basic Theory: A Brief Review on Partial-Geometric Distribution

The partial-geometric (PG) distribution, introduced by [15] as an extension of geometric distribution, provides more flexibility than the geometric distribution as it is potential to capture not only over-dispersed data but also under-dispersed data. A non-negative random variable X has the PG distribution with parameters $p \in (0,1)$ and $\alpha \in \left[0, \frac{p}{1-p}\right]$ if its probability mass function (PMF) takes the form

$$p_X(k) = \begin{cases} 1 - \frac{\alpha(1-p)}{p} & \text{if } k = 0 \\ \alpha(1-p)^k & \text{if } k = 1, 2, \ldots \end{cases} \tag{2.1}$$

3. The modified Partial-Geometric Distribution

3.1. **Cumulative distribution and probability mass functions.** The basic idea of constructing the modified partial-geometric (MPG) distribution came by investigating the cumulative distribution function (CDF) of the partial-geometric (PG) distribution with the corresponding probability mass function (PMF) given in Eq. (2.1). However, the study in [15] does not provide CDF of the PG distribution. For this purpose, we derive CDF of the PG distribution in advance. Let X be a non-negative integer random variable following PG(p, α) distribution with its PMF given in Eq. (2.1). Then the corresponding CDF is obtained as follows:

$$\begin{aligned} F_X^*(k) = \sum_{i=0}^{k} f_X(i) &= \left(1 - \frac{\alpha(1-p)}{p}\right) + \alpha(1-p) + \alpha(1-p)^2 + \cdots + \alpha(1-p)^k \\ &= \left(1 - \frac{\alpha(1-p)}{p}\right) + \alpha(1-p)\left(\frac{1-(1-p)^k}{1-(1-p)}\right) \\ &= 1 - \frac{\alpha(1-p)^{k+1}}{p} \end{aligned} \tag{3.1}$$

By powering a positive real number λ to the two-hand side of Eq. (3.1), it transforms into Eq. (3.2).

$$[F_X^*(k)]^\lambda = \left(1 - \frac{\alpha(1-p)^{k+1}}{p}\right)^\lambda. \tag{3.2}$$

The value of λ must be positive in order to preserve the properties of the CDF, and leads to the following definition of MPG distribution.

Definition 3.1. A non-negative integer random variable Y has the MPG distribution, denoted by MPG(p, α, λ), if its CDF takes the form

$$F_Y(y) = \left[1 - \frac{\alpha(1-p)^{\lfloor y \rfloor + 1}}{p}\right]^\lambda, \;\; y \geq 0, \tag{3.3}$$

where $0 < p < 1$, $0 \leq \alpha \leq \frac{p}{1-p}$, and $\lambda > 0$ is a shape parameter.

In particular, if $\lambda = 1$, then the MPG(p, α, λ) simplifies to PG(p, α). Therefore, the partial-geometric distribution is a submodel of modified partial-geometric distribution.

By the CDF in Eq. (3.3), other distributional functions alongside basic distributional quantities can be derived. For more convenient, the CDF in Eq. (3.3) can be rewritten into series expansion. By using the binomial expansion

$$(1+x)^\beta = \sum_{i=0}^{\infty} \binom{\beta}{i} x^i, \tag{3.4}$$

Eq. (3.3) becomes

$$F_Y(y) = \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j (1-p)^{\lfloor y \rfloor j}, \;\; y \geq 0. \tag{3.5}$$

The corresponding PMF for the CDF in Eq. (3.3) is

$$p_Y(y) = F_Y(y) - F_Y(y-1) = \left[1 - \frac{\alpha(1-p)^{y+1}}{p}\right]^\lambda - \left[1 - \frac{\alpha(1-p)^y}{p}\right]^\lambda, \tag{3.6}$$

for $y = 1, 2, \ldots$ and $p_Y(0) = F_Y(0) = \left[1 - \frac{\alpha(1-p)}{p}\right]^\lambda$. However, by using expansion in Eq. (3.4), Eq. (3.6) becomes

$$\begin{aligned} p_Y(y) &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j (1-p)^{yj} - \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j (1-p)^{(y-1)j} \\ &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j (1-p)^{(y-1)j}[(1-p)^j - 1], \text{ for } y = 1, 2, \ldots \\ p_Y(0) &= \left[1 - \frac{\alpha(1-p)}{p}\right]^\lambda. \end{aligned} \tag{3.7}$$

The PMF plots of the MPG distribution are presented in Fig. 1. While the PMF of the original PG distribution exhibits only decreasing or unimodal (upside-down) shapes [15], the PMF of the proposed MPG distribution demonstrates greater flexibility, allowing for decreasing, unimodal, and more complex forms such as a decreasing-unimodal pattern.

3.2. **Shape parameter λ analysis.** This section provides the analysis of shape parameter λ in the MPG distribution. A shape parameter has been able to add more shapes to the PMF and hazard rate graphs of a probability distribution. This parameter is important because it allows the users to see if a distribution can be fitted to a given dataset by comparing the shape of the histogram (empirical distribution) with possible curves of PMF. Moreover, shape parameter is also useful to see if a distribution is suitable for data with monotonic or non-monotonic hazard rates.

Let $Y \sim$ MPG(p, α, λ), the hazard rate function of Y is

$$h_Y(y) = \frac{p_Y(y)}{S_Y(y)} = \frac{\left[1 - \frac{\alpha(1-p)^{y+1}}{p}\right]^\lambda - \left[1 - \frac{\alpha(1-p)^y}{p}\right]^\lambda}{1 - \left[1 - \frac{\alpha(1-p)^{y+1}}{p}\right]^\lambda}, \;\; y = 1, 2, \ldots. \tag{3.8}$$

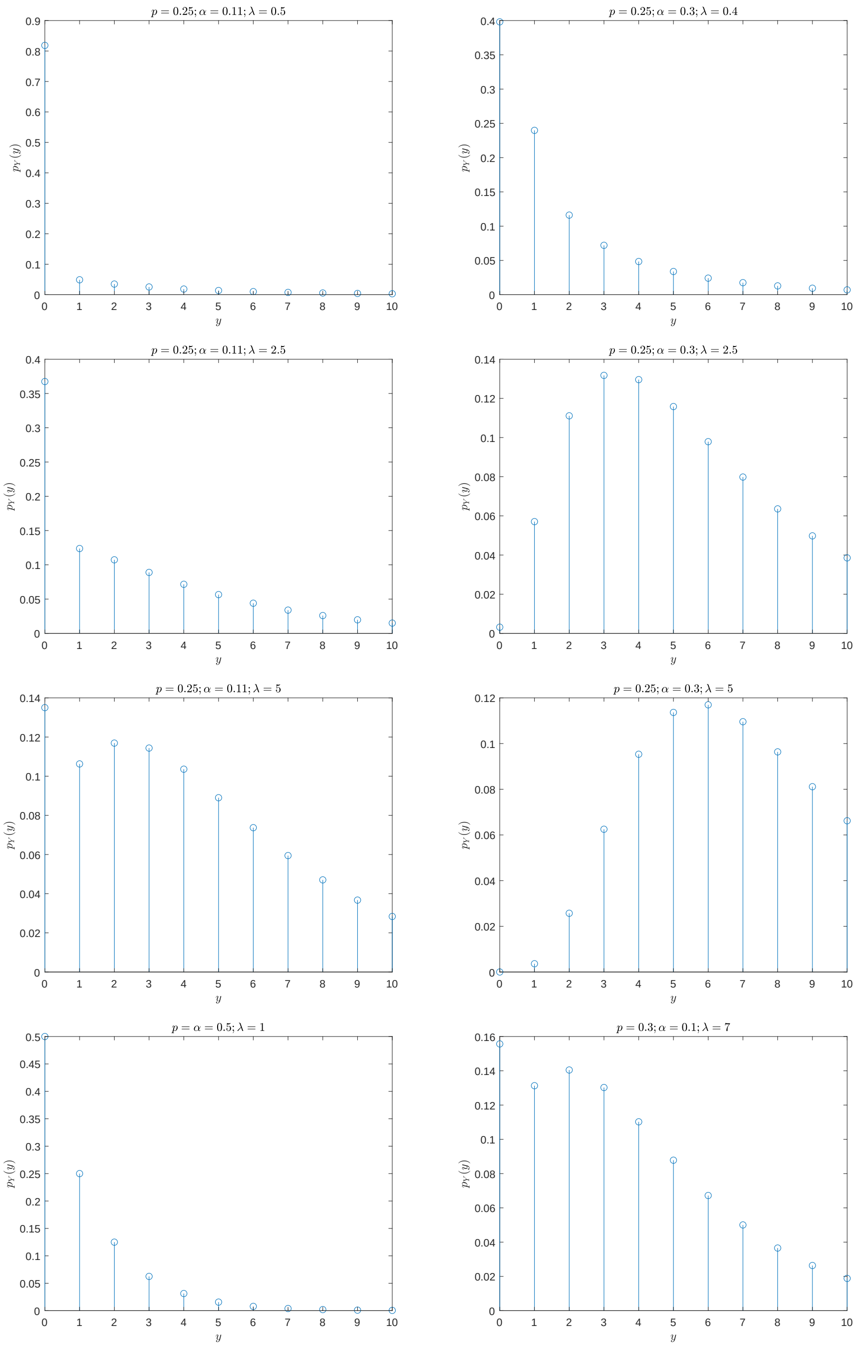

FIGURE 1. The PMF plots of the modified partial-geometric distribution with $p > \alpha$, $p < \alpha$ and different values of λ.

Since Eq. (3.8) is too complex, a graphical approach becomes relevant to be considered to show the shapes (patterns) of hazard rate. Fig. 2 and Fig. 3 present the hazard rate plots of the MPG distribution with selected values of shape parameter λ and it shows that for any combination of p and α, the hazard rate is monotonically decreasing when $\lambda < 1$ or monotonically increasing when $\lambda > 1$. On the other hand, when $\lambda = 1$ (PG distribution), the hazard rate is only a constant function. Therefore, the shape parameter λ enhances the flexibility to the hazard rate function of the MPG.

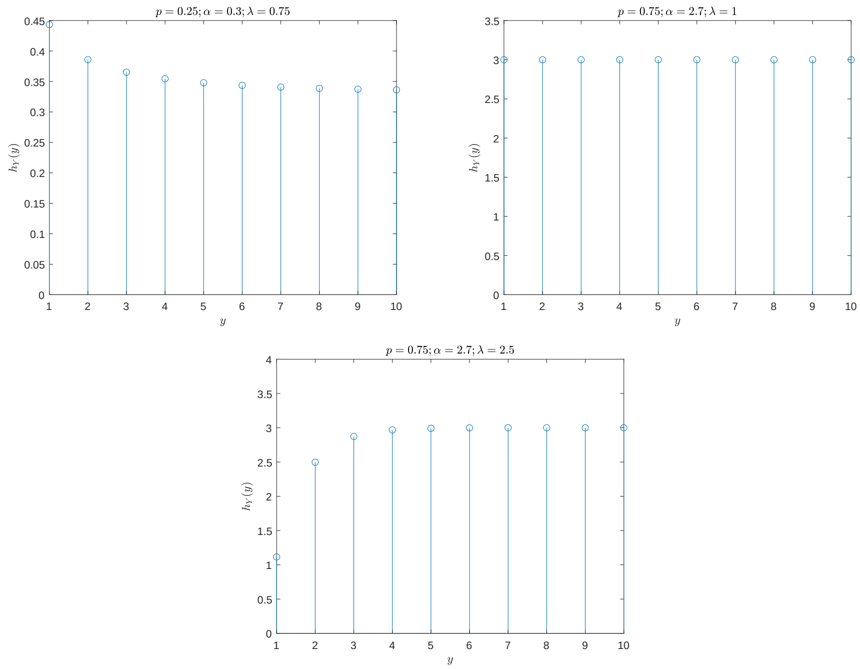

FIGURE 2. The hazard plots of the modified partial-geometric distribution with $p < \alpha$ and different values of λ.

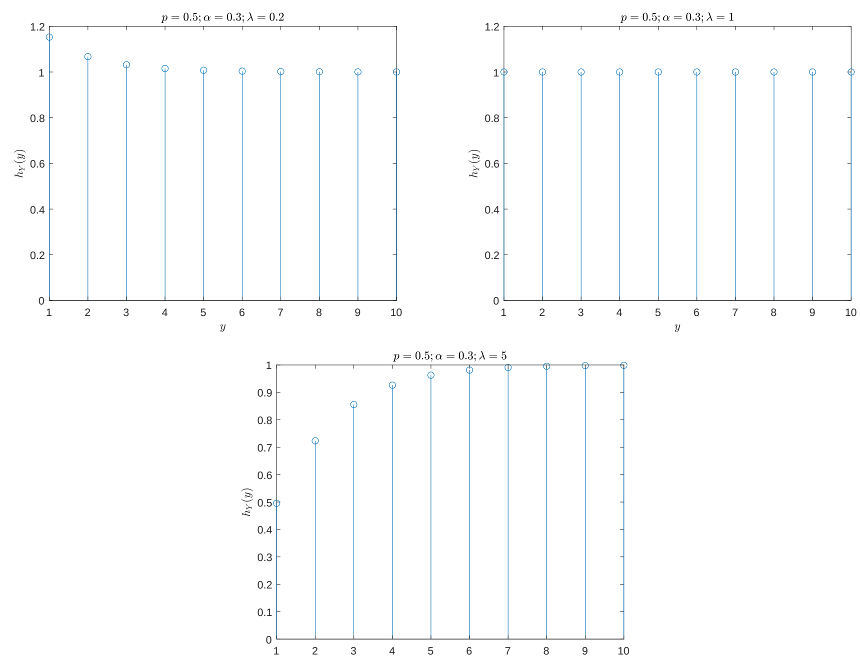

FIGURE 3. The hazard plots of the modified partial-geometric distribution with $p > \alpha$ and different values of λ.

3.3. **Moment and probability generating functions.** Moment generating function (MGF) and probability generating function (PGF) are useful to calculate some measures, such as moments of distribution.

Theorem 3.2. *Let Y denotes the MPG random variable in which PMF given in Eq. (3.7). The MGF of Y is*

$$M_Y(t)=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[1-(1-p)^{-j}]\frac{e^t(1-p)^j}{1-e^t(1-p)^j},\ \ t<-j\log(1-p).$$

Proof.

$$\begin{aligned}
M_Y(t)&=E[e^{tY}]\\
&=\sum_{y=0}^{\infty}e^{ty}\left[\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j(1-p)^{(y-1)j}[(1-p)^j-1]\right]\\
&=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[(1-p)^j-1]\left[\sum_{y=0}^{\infty}e^{ty}(1-p)^{(y-1)j}\right]\\
&=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[(1-p)^j-1]\left[(1-p)^{-j}\sum_{y=0}^{\infty}(e^t(1-p)^j)^y\right]\\
&=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[1-(1-p)^{-j}]\frac{e^t(1-p)^j}{1-e^t(1-p)^j}.
\end{aligned}\tag{3.9}$$

Eq. (3.9) holds if and only if

$$|e^t(1-p)^j|<1,$$

or equivalent to $t<-j\log(1-p)$. □

Theorem 3.3. *Let Y denotes the MPG random variable in which PMF given in Eq. (3.7). The PGF of Y is*

$$G_Y(t)=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[1-(1-p)^{-j}]\frac{t(1-p)^j}{1-t(1-p)^j},$$

where $-(1-p)^{-j}<t<(1-p)^{-j}$.

Proof.

$$\begin{aligned}
G_Y(t)=E[t^Y]&=\sum_{y=0}^{\infty}t^y\left[\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j(1-p)^{(y-1)j}[(1-p)^j-1]\right]\\
&=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[(1-p)^j-1]\left[\sum_{y=0}^{\infty}t^y(1-p)^{(y-1)j}\right]\\
&=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[(1-p)^j-1]\left[(1-p)^{-j}\sum_{y=0}^{\infty}(t(1-p)^j)^y\right]\\
&=\sum_{j=0}^{\infty}\binom{\lambda}{j}\left(-\frac{\alpha(1-p)}{p}\right)^j[1-(1-p)^{-j}]\frac{t(1-p)^j}{1-t(1-p)^j}.
\end{aligned}\tag{3.10}$$

Eq. (3.10) holds if and only if

$$|t(1-p)^j|<1,$$

or equivalent to $-(1-p)^{-j}<t<(1-p)^{-j}$. □

3.4. **Mean, variance, and index of dispersion.** This subsection provides some basic distributional quantities of the MPG, such as mean, variance and index of dispersion (IOD).

Theorem 3.4. *Let Y denotes the MPG random variable in which PMF given in Eq. (3.7). The mean (first moment) of Y is*

$$E(Y) = \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j}{[1-(1-p)^j]^2}. \tag{3.11}$$

Proof.

$$\begin{aligned} E(Y) &= \frac{d}{dt} M_Y(t)|_{t=0} \\ &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{e^t(1-p)^j}{[1-e^t(1-p)^j]^2}|_{t=0} \\ &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j}{[1-(1-p)^j]^2}. \end{aligned}$$

□

Theorem 3.5. *Let Y denotes the MPG random variable in which PMF given in Eq. (3.7). The second moment of Y is*

$$E(Y^2) = \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j[1+(1-p)^j]}{[1-(1-p)^j]^3}. \tag{3.12}$$

Proof.

$$\begin{aligned} E(Y^2) &= \frac{d^2}{dt^2} M_Y(t)|_{t=0} \\ &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{e^t(1-p)^j[1+e^t(1-p)^j]}{[1-e^t(1-p)^j]^3}|_{t=0} \\ &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j[1+(1-p)^j]}{[1-(1-p)^j]^3}. \end{aligned}$$

□

Theorem 3.6. *Let Y denotes the MPG random variable in which PMF given in Eq. (3.7). The variance of Y is*

$$\begin{aligned} \mathit{Var}(Y) = &\sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j[1+(1-p)^j]}{[1-(1-p)^j]^3} \\ &- \left[\sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j}{[1-(1-p)^j]^2}\right]^2. \end{aligned} \tag{3.13}$$

Proof.

$$\begin{aligned} \mathrm{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\ &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j[1+(1-p)^j]}{[1-(1-p)^j]^3} \\ &\quad - \left[\sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j}{[1-(1-p)^j]^2}\right]^2. \end{aligned}$$

□

Furthermore, we can obtain the index of dispersion (IOD) which is the ratio of the variance to the mean. There are three types of dispersal behavior, that is underdispersion, equidispersion, and overdispersion, respectively, when its IOD is less than one, equals to one, and more than one. However, since mathematical expressions of the variance and mean of the MPG distribution are too complex, the form of its IOD is complex as well. Hence, numerical simulation of IOD is conducted by selecting several parameter values to produce various values of the IOD. Table 1 presents IOD values of the MPG distribution, and it indicates that the MPG distribution can be under-dispersed, equi-dispersed and over-dispersed, which are resulted from various combinations of p, α and λ. Moreover, in any fixed combinations of p and α, the IOD is larger as the value of λ decreases. Therefore, it is sufficient to conclude that the shape parameter λ can also influence the way data in the MPG distribution likely dispersed relative to its mean.

TABLE 1. Index of dispersion of the MPG distribution for different value of p, α and λ.

p	α	λ	Mean	Second Moment	Variance	IOD
0.5	0.3	2.5	1.285	4.1378	2.4866	1.9351
0.5	0.3	1	0.6	1.8	1.44	2.4
0.5	0.3	0.5	0.3174	0.928	0.8273	2.6065
0.75	1.5	2.5	1.2093	2.255	0.7926	0.6554
0.75	1.5	1	0.6667	1.1111	0.6667	1
0.75	1.5	0.5	0.3784	0.6054	0.4622	1.2215
0.75	2.7	2.5	1.6493	3.4409	0.7207	0.437
0.75	2.7	1	1.2	2	0.56	0.4667
0.75	2.7	0.5	0.8414	1.2575	0.5496	0.6532
0.5	0.5	2.5	1.9227	6.5192	2.8224	1.4679
0.5	0.5	1	1	3	2	2
0.5	0.5	0.5	0.5546	1.5854	1.2778	2.304

3.5. **Quantiles of the MPG distribution.**

Theorem 3.7. *Let Y denotes the MPG random variable in which CDF given in Eq. (3.3). The r quantile (or $100r$-th percentile) of Y is any non-negative integer y_r belong to the following closed interval.*

$$\frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} - 1 \le y_r \le \frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)}, \tag{3.14}$$

where $0 < r < 1$.

Proof. The r quantile (or $100r$-th percentile) of Y is any non-negative integer y_r such that the following inequality holds.

$$F_Y(y_r^-) \le r \le F_Y(y_r), \tag{3.15}$$

where $0 < r < 1$. The inequality in Eq. (3.15) leads to the following inequalities

$$y_r \ge \frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} - 1 \text{ and } y_r \le \frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)},$$

or it can be rewritten as follows

$$\frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} - 1 \le y_r \le \frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)}.$$

Hence, the quantile y_r is the non-negative integers between these values. $\square$

Moreover, if $\frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)}$ is not an integer, the r quantile of the MPG distribution is unique and given as follows:

$$y_r = \left\lfloor \frac{\log\left\{\frac{p(1-r^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} \right\rfloor, \tag{3.16}$$

where $\lfloor \cdot \rfloor$ is the floor function.

4. Parameter Estimations

4.1. **Moment matching.** Moment matching is a statistical technique used to estimate parameters of a probability distribution by equating the theoretical moments of the distribution (like the mean, variance, etc.) to the corresponding empirical (sample) moments computed from data. Given a random sample of size n, the k-th empirical moment is defined as follows:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} y_i^k \tag{4.1}$$

Since there are three parameters within the modified partial-geometric (MPG) distribution, three moment matchings are then necessary. For example, we can match the first and second moments of the MPG with the first and second empirical moments, respectively, and also match the variance of the MPG with the empirical variance. Therefore, we obtain

$$E(Y) = \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j}{[1-(1-p)^j]^2} = m_1 \tag{4.2}$$

$$E(Y^2) = \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j + (1-p)^{2j}}{[1-(1-p)^j]^3} = m_2 \tag{4.3}$$

$$\begin{aligned} \mathrm{Var}(Y) &= \sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j + (1-p)^{2j}}{[1-(1-p)^j]^3} \\ &- \left[\sum_{j=0}^{\infty} \binom{\lambda}{j} \left(-\frac{\alpha(1-p)}{p}\right)^j [1-(1-p)^{-j}] \frac{(1-p)^j}{[1-(1-p)^j]^2}\right]^2 \\ &= m_2 - (m_1)^2, \end{aligned} \tag{4.4}$$

where m_1 and m_2 are obtained from Eq. (4.1). However, the above equations are too difficult to be solved even using numerical approaches.

4.2. **Percentile matching.** Percentile matching method is a technique used to estimate parameters of a probability distribution by matching empirical (sample) percentiles to the theoretical percentiles of the distribution. Given a random sample of size n, the $100r$-th empirical estimate of a percentile is calculated as follows:

$$\pi_r = (1-h)y_{(j)} + hy_{(j+1)}, \tag{4.5}$$

where

$$0 < r < 1; \;\; j = \lfloor (n+1)r \rfloor; \;\; h = (n+1)r - j,$$

and $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ are the order observed values from the sample.

As there are three parameters within the modified partial-geometric (MPG) distribution, three percentile matchings are then necessary. For example, we can match the 25-th, 50-th, and 75-th percentiles of the MPG with the 25-th, 50-th, and 75-th empirical percentiles, respectively. For more convenient, assume that the condition for Eq. (3.16) is fulfilled. Then, we obtain

$$y_{0.25} = \left\lfloor \frac{\log\left\{\frac{p(1-(0.25)^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} \right\rfloor = \pi_{0.25}, \tag{4.6}$$

$$y_{0.5} = \left\lfloor \frac{\log\left\{\frac{p(1-(0.5)^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} \right\rfloor = \pi_{0.5}, \tag{4.7}$$

$$y_{0.75} = \left\lfloor \frac{\log\left\{\frac{p(1-(0.75)^{\frac{1}{\lambda}})}{\alpha}\right\}}{\log(1-p)} \right\rfloor = \pi_{0.75}, \tag{4.8}$$

where $\pi_{0.25}$, $\pi_{0.5}$ and $\pi_{0.75}$ are obtained from Eq. (4.5). However, the above equations are too difficult to be solved even using numerical approaches. Moreover, percentile method produce different solutions for different selected percentiles. It is not efficient for users to determine the right percentiles to obtain good estimators.

4.3. **Maximum likelihood estimation.** In this research, the maximum likelihood estimation (MLE) is applied to estimate the parameters of MPG distribution as the MLE will result in minimum-variance unbiased estimator for large sample size. Let $Y_1, Y_2, \ldots, Y_n$ be an independent and identically distributed (i.i.d) random sample of size n from the MPG distributed population, MPG(p, α, λ), and $y_1, y_2, \ldots, y_n$ be the realization values of random variables $Y_1, Y_2, \ldots, Y_n$ respectively. For every $k = 0, 1, 2, \ldots$, n_k denotes the number of observations where $Y = k$. Let n denotes total number of observations (sample size). Hence

$$n = \sum_{k=0}^{\infty} n_k$$

The likelihood function can be written as

$$\begin{aligned} L(p, \alpha, \lambda \mid y_1, y_2, \ldots, y_n) &= [\Pr(Y=0)]^{n_0} \prod_{k=1}^{\infty} [\Pr(Y=k)]^{n_k} \\ &= \left(\left[1 - \frac{\alpha(1-p)}{p}\right]^{\lambda}\right)^{n_0} \\ &\quad \prod_{k=1}^{\infty} \left(\left[1 - \frac{\alpha(1-p)^{k+1}}{p}\right]^{\lambda} - \left[1 - \frac{\alpha(1-p)^{k}}{p}\right]^{\lambda}\right)^{n_k}. \end{aligned}$$

Taking the logarithm of the likelihood function, we obtain the log-likelihood function:

$$\begin{aligned} l(p, \alpha, \lambda \mid y_1, y_2, \ldots, y_n) &= \lambda n_0 \log\left[1 - \frac{\alpha(1-p)}{p}\right] \\ &\quad + \sum_{k=1}^{\infty} n_k \log\left(\left[1 - \frac{\alpha(1-p)^{k+1}}{p}\right]^{\lambda} - \left[1 - \frac{\alpha(1-p)^{k}}{p}\right]^{\lambda}\right), \end{aligned} \tag{4.9}$$

where $0 < p < 1$, $0 \leq \alpha \leq \frac{p}{1-p}$, and $\lambda > 0$.

The value of p, α and λ that maximizes Eq. (4.9) can be obtained by solving the following system of partial differential equations:

$$\frac{\partial l}{\partial p} = \frac{\alpha\lambda n_0}{p^2\left[1-\frac{\alpha(1-p)}{p}\right]} + \sum_{k=1}^{\infty} n_k \frac{\partial}{\partial p}\left[\log(g(k;p,\alpha,\lambda))\right] = 0 \tag{4.10}$$

$$\frac{\partial l}{\partial \alpha} = -\frac{\lambda n_0}{\frac{p}{1-p}-\alpha} + \sum_{k=1}^{\infty} n_k \frac{\partial}{\partial \alpha}\left[\log(g(k;p,\alpha,\lambda))\right] = 0 \tag{4.11}$$

$$\frac{\partial l}{\partial \lambda} = n_0 \log\left[1-\frac{\alpha(1-p)}{p}\right] + \sum_{k=1}^{\infty} n_k \frac{\partial}{\partial \lambda}\left[\log(g(k;p,\alpha,\lambda))\right] = 0 \tag{4.12}$$

where $g(k;p,\alpha,\lambda) = \left[1-\frac{\alpha(1-p)^{k+1}}{p}\right]^{\lambda} - \left[1-\frac{\alpha(1-p)^{k}}{p}\right]^{\lambda}$.

It requires a numerical approach to find $\hat{p}$, $\hat{\alpha}$ and $\hat{\lambda}$. However, instead of maximizing the log-likelihood function in Eq. (4.9), minimizing the negative log-likelihood function is more preferable. This study employs the augmented Lagrangian minimization method (see [23]) to find $\hat{p}$, $\hat{\alpha}$ and $\hat{\lambda}$ that minimize the negative log-likelihood function formed by the MPG random sample. This method is chosen due to its suitability in optimizing nonlinear objective functions with linear or nonlinear equality and inequality constraints [23]. We run this algorithm using "auglag" function in the R programming.

5. Numerical Simulation

This section presents numerical simulations to assess the performance of proposed estimators for some specified values of the parameters of the MPG distribution, using simulated data with several sizes. The simulation is begun by performing Monte Carlo experiment to generate data of size n from the MPG distribution and it is repeated 10,000 times. The estimated parameters are obtained using the MLE and augmented Lagrangian minimization method as discussed in Subsection 4.3. The performance of the MLE estimator is evaluated using two measures, the average estimate and the mean squared error (MSE), which are given as follows:

$$\text{Average estimate} = \frac{\sum_{j=1}^{10,000}\hat{\tau}_j}{10,000};\ \ \text{MSE} = \frac{\sum_{j=1}^{10,000}(\hat{\tau}_j-\tau)^2}{10,000}$$

where $\hat{\tau}_j$ is the fitted value for the considered parameter on the j-th experiment and τ is the true value of the considered parameter. The numerical results are presented in Table 2.

According to Table 2, the average estimates of the parameters are generally close to their true values, with deviations not exceeding 0.5. Furthermore, the accuracy of the estimates improves as the sample size increases, as evidenced by the corresponding decrease in mean squared error (MSE). These findings support the conclusion that the MLE estimators, as presenting in Subsection 4.3, are suitable as point estimators of the respective parameters.

6. Conclusion

This research introduces the novel modified partial-geometric (MPG) distribution which is constructed by taking the CDF of partial-geometric (PG) distribution to the power of positive real number. As confirmed by its IOD, the MPG distribution offers greater flexibility than the majority of existing distributions, as it can generate under-dispersed, equi-dispersed, and over-dispersed data. Moreover, the hazard rate shapes of the MPG distribution can also be increasing and decreasing which cannot be achieved by partial-geometric (PG) distribution. This study also derived essential probability properties of the MPG distribution, including its MGF and PGF.

TABLE 2. Numerical results of MPG distribution parameter estimation for simulated data of size n.

Para meter	Average Estimate			MSE		
	$n = 100$	$n = 500$	$n = 1000$	$n = 100$	$n = 500$	$n = 1000$
$p = 0.5$	0.4969	0.4832	0.4872	0.0203	0.0050	0.0027
$\alpha = 0.2$	0.2398	0.2483	0.2449	0.1054	0.0425	0.0333
$\lambda = 0.5$	1.7840	0.9533	0.7821	4.4894	1.1933	0.5985
$p = 0.3$	0.3031	0.2994	0.2993	0.0009	0.0002	0.0001
$\alpha = 0.4$	0.4032	0.4033	0.4014	0.0044	0.0007	0.0004
$\lambda = 2.5$	2.8067	2.4805	2.4807	3.2493	0.1317	0.0660
$p = 0.7$	0.7117	0.6986	0.6992	0.0179	0.0076	0.0048
$\alpha = 1.65$	1.1198	1.2043	1.3392	1.7680	0.5545	0.3434
$\lambda = 0.15$	1.5565	0.6553	0.4342	5.5447	1.2450	0.5977
$p = 0.2$	0.1973	0.1988	0.1996	0.0006	0.0001	0.0001
$\alpha = 0.05$	0.0592	0.0535	0.0499	0.0018	0.0007	0.0004
$\lambda = 5$	6.3993	5.6344	5.8804	21.0097	7.0537	7.1241

Parameter estimation for the MPG distribution is performed using MLE with the support of augmented Lagrangian minimization method because the solution of MLE cannot be obtained analytically. The performance of the MLE estimator is examined by conducting Monte Carlo simulations with various parameter values and sample sizes. The simulation results show that the estimated parameters are very close to the true parameter values, particularly for large sample sizes.

Competing interests. The authors declare no competing interests.

REFERENCES

[1] E. Gómez-Déniz, E. Calderín-Ojeda, The Discrete Lindley Distribution: Properties and Applications, J. Stat. Comput. Simul. 81 (2011), 1405–1416. https://doi.org/10.1080/00949655.2010.487825.

[2] S. Chakraborty, D. Chakravarty, Discrete Gamma Distributions: Properties and Parameter Estimations, Commun. Stat. - Theory Methods 41 (2012), 3301–3324. https://doi.org/10.1080/03610926.2011.563014.

[3] M.S. Noughabi, A.H. Rezaei Roknabadi, G.R. Mohtashami Borzadaran, Some Discrete Lifetime Distributions with Bathtub-Shaped Hazard Rate Functions, Qual. Eng. 25 (2013), 225–236. https://doi.org/10.1080/08982112.2013.769055.

[4] M. El-Morshedy, M.S. Eliwa, E. Altun, Discrete Burr-Hatke Distribution with Properties, Estimation Methods and Regression Model, IEEE Access 8 (2020), 74359–74370. https://doi.org/10.1109/access.2020.2988431.

[5] V. Nekoukhou, M.H. Alamatsaz, H. Bidram, Discrete Generalized Exponential Distribution of a Second Type, Statistics 47 (2013), 876–887. https://doi.org/10.1080/02331888.2011.633707.

[6] J. Gillariose, O.S. Balogun, E.M. Almetwally, R.A.K. Sherwani, F. Jamal, et al., On the Discrete Weibull Marshall–Olkin Family of Distributions: Properties, Characterizations, and Applications, Axioms 10 (2021), 287. https://doi.org/10.3390/axioms10040287.

[7] E.M. Almetwally, H.M. Almongy, H.A. Saleh, Managing Risk of Spreading "COVID-19" in Egypt: Modelling Using a Discrete Marshall-Olkin Generalized Exponential Distribution, Int. J. Probab. Stat. 9 (2020), 33–41.

[8] E.G. Déniz, A New Discrete Distribution: Properties and Applications in Medical Care, J. Appl. Stat. 40 (2013), 2760–2770. https://doi.org/10.1080/02664763.2013.827161.

[9] C. Kuş, Y. Akdoğan, A. Asgharzadeh, İ. Kınacı, K. Karakaya, Binomial-Discrete Lindley Distribution, Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat. 68 (2018), 401–411. https://doi.org/10.31801/cfsuasmas.424228.

[10] D. Bhati, I.S. Ahmed, On Uniform-Negative Binomial Distribution Including Gauss Hypergeometric Function and Its Application in Count Regression Modeling, Commun. Stat. - Theory Methods 50 (2019), 3106–3122. https://doi.org/10.1080/03610926.2019.1682163.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] S. Nadarajah, S. Bakar, An Exponentiated Geometric Distribution, Appl. Math. Model. 40 (2016), 6775–6784. https://doi.org/10.1016/j.apm.2015.11.010.

[17] R.D. Gupta, D. Kundu, Theory & Methods: Generalized Exponential Distributions, Aust. New Zealand J. Stat. 41 (1999), 173–188. https://doi.org/10.1111/1467-842x.00072.

[18] M. Pal, M.M. Ali, J. Woo, Modified Weibull Distribution, Statistica. 66 (2006), 139–147. https://doi.org/10.6092/issn.1973-2201/493.

[19] A.M. Abd-Elrahman, A New Two-Parameter Lifetime Distribution with Decreasing, Increasing or Upside-Down Bathtub-Shaped Failure Rate, Commun. Stat. - Theory Methods 46 (2017), 8865–8880. https://doi.org/10.1080/03610926.2016.1193198.

[20] M. Mahmoud, M. Ghazal, Estimations from the Exponentiated Rayleigh Distribution Based on Generalized Type-II Hybrid Censored Data, J. Egypt. Math. Soc. 25 (2017), 71–78. https://doi.org/10.1016/j.joems.2016.06.008.

[21] S. Khazaei, A.A. Nanvapisheh, The Comparison Between Gumbel and Modified Gumbel Distributions and Their Applications in Hydrological Process, Adv. Mach. Learn. Artif. Intell. 2 (2021), 49–54. https://doi.org/10.33140/AMLAI.02.01.08.

[22] H.S. Bakouch, B.M. Al-Zahrani, A.A. Al-Shomrani, V.A. Marchi, F. Louzada, An Extended Lindley Distribution, J. Korean Stat. Soc. 41 (2012), 75–85. https://doi.org/10.1016/j.jkss.2011.06.002.

[23] E.G. Birgin, J.M. Martínez, Practical Augmented Lagrangian Methods for Constrained Optimization, SIAM, Philadelphia, 2014. https://doi.org/10.1137/1.9781611973365.