Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 2, 2024 (New Series), pp 203–216
http://www.ssca.org.in/journal

A Comprehensive Study of the Power Modified Lindley-Geometric Distribution in the T-X Family: COVID-19 Applications

Meenu Jose[1] **and Lishamol Tomy**[2]
[1]*Department of Statistics, Carmel College(Autonomous) Mala,Thrissur, Kerala, India*
[2]*Department of Statistics, Deva Matha College, Kuravilangad, Kerala, 686633, India*

Received: 08 October 2023; Revised: 30 December 2023; Accepted: 05 January 2024

Abstract

In this paper we develop a superior and ideal statistical model to provide optimal modelling for the number of deaths resulting from COVID-19 infections. This paper introduces the power modified Lindley-geometric distribution, a novel versatile three-parameter discrete model built on the T-X methodology. In addition, to providing a generalized geometric distribution we offer a thorough list of its mathematical characteristics. The parameter of the new model is estimated using four different estimation techniques: maximum likelihood, Cramer-von Mises, least-square, and weighted least-square. The simulation experiment uses four distinct estimating approaches to test the accuracy of the model parameters. Additionally, we applied two datasets to the COVID-19 mortality data for the United Kingdom and Egypt. These two instances of actual data were used to highlight the significance of our distribution for modelling and fitting this particular kind of discrete data.

Key words: T-X family, Maximum likelihood; Cramer-von Mises; Least-square; Weighted least-square; Data analysis.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

In our modern period, the abundance of data coming in from all fields has submerged the interest in defining new flexible distributions. "Thoroughly changing" a baseline distribution is an easy and quick way to define these mathematical objects. The study of tail properties and improving the goodness-of-fit of the associated models have both been demonstrated to benefit from the addition of parameter(s). The most well-liked distribution among those that have been suggested is the T-X family of distributions by Alzaatreh *et al.* (2013). The following peculiar transformation is one of the most practical transformers for T-X family of distributions $W(F(x)) = -log(1 - F(x))$, where the cumulative density function *(CDF)* of random variable X is represented by the notation $F(x)$. To put it another way, $W(F(x))$ is used to modify the distribution described by $F(x)$ and define a new family

Corresponding Author: Meenu Jose
Email: meenusgc@gmail.com

of distributions based on a changed *CDF*. With the help of the T-X family, one may quickly build discrete distributions in addition to continuous distributions. The T-geometric family, which are the discrete analogues of the distribution of the random variable T, was defined and explored by Alzaatreh *et al.* (2012) as a result. The *CDF* of T-geometric family is given by

$$G(x;\phi,b)=\int_{0}^{-log(p^{(x+1)})} r(t;\phi)dt = R[-log(p^{(x+1)});\phi,p)] = R[b(x+1);\phi,b], \quad x=0,1,2,..., \tag{1}$$

where $b=-log\ p>0$ and ϕ the parameters of the *CDF* $R(x;\phi)$. Some of the families available in the modern literature are the Pareto-geometric, Weibull-geometric, Burr-geometric and exponentiated exponential -
geometric distribution by Alzaatreh *et al.* (2012), Kumaraswamy-geometric distribution by Akinsete *et al.* (2014) and exponentiated Weibull-geometric distribution by Famoye (2019). The comprehensive review of T-X family of distributions may be found in Tomy *et al.* (2019).

Chesneau *et al.* (2021b) introduce a novel two-parameter lifetime distribution that is the power version of the modified Lindley distribution and call it as power modified Lindley (PML) distribution. It offers a compelling substitute for the Weibull and power Lindley distributions as its primary goal. Let T be a random variable with the PML distribution. The probability density function (*PDF*) and the *CDF* are each defined as

$$r(t;\alpha,\theta)=\frac{\theta\alpha}{1+\theta}t^{\alpha-1}e^{-2\theta t^{\alpha}}\left[(1+\theta)e^{\theta t^{\alpha}}+2\theta t^{\alpha}-1\right], \quad t>0, \tag{2}$$

$$R(t;\alpha,\theta)=1-\left[1+\frac{\theta t^{\alpha}}{1+\theta}e^{-\theta t^{\alpha}}\right]e^{-\theta t^{\alpha}}, \quad t>0, \tag{3}$$

where α >0 and θ >0. This distribution is derived by using the power parameter α in modified Lindely distribution, has been proposed by Chesneau *et al.* (2021a).

This paper introduces a flexible three parameter discrete distribution called power modified Lindley-geometric, which is based on T-geometric family of distribution and power modified Lindley distribution. The main driving force behind the development of this new discrete distribution was the fact that, in contrast to the amount of literature on continuous cases, there was a dearth of research on the discrete families of distributions. Another fact is that there are lots of researchers work to understand the patterns of the COVID-19 epidemic and offer models that better suit the data and can be used to estimate the anticipated number of cases and deaths to assist the government in making decisions on preventative measures. And the new distribution is suitable for fitting COVID-19 data sets, which is the main goal of this study. Another motivator is the characteristics of the suggested distribution itself. In other words, the newly proposed discrete distribution features a probability mass function (*PMF*) that is right-skewed, symmetric and left-skewed. Additionally, the new distribution features hazard rate functions (*HRF*) that are increasing, decreasing and upside-down bathtub-shaped. Additionally, we provided a comparison of the various estimation techniques.

Following is an outline of the remaining content: In Section 2, we provide a brand-new discrete family of distributions. Section 3, a special case of the obtained new discrete discrete family of distribution and its probabilistic characteristics are studied in detail. Section 4 discussed least-squares, weighted least squares, and the Cramer von Mises technique in addition to maximum likelihood estimation. A thorough simulation analysis is employed in Section 5 to evaluate the performance of these estimators. Applications to the two COVID-19 data sets used to demonstrate how the new distribution performs are detailed in Section 6. A few closing thoughts are provided in Section 7.

2. Discrete power modified Lindley-X family of distribution

In this section, we introduce the discrete power modified Lindley-X (PML-X) family of distributions, a new discrete family of distributions. Utilising the Alzaatreh *et al.* (2013) T-X generalization technique, we enable the transformed random variable T to have the PML distribution and the transformer random variable X is a discrete random variable, with $W(F(x)) = -log(1 - F(x))$. Then the CDF of new family is given by

$$\begin{aligned} G(x;\alpha,\theta,\Im) &= \int_0^{-log(1-F(x;\Im))} r(t;\alpha,\theta)dt = R(-log(1-F(x;\Im))) \\ &= 1 - \left[1 + \frac{\theta[-log(1-F(x;\Im))]^\alpha}{1+\theta} e^{-\theta[-log(1-F(x;\Im))]^\alpha}\right] e^{-\theta[-log(1-F(x;\Im))]^\alpha} \end{aligned} \tag{4}$$

The corresponding PMF of the PML-X family of discrete distributions becomes.

$$\begin{aligned} g(x;\alpha,\theta,\Im) &= G(x) - G(x-1) \\ &= \left[1 + \frac{\theta[-log(1-F(x-1;\Im))]^\alpha}{1+\theta} e^{-\theta[-log(1-F(x-1;\Im))]^\alpha}\right] e^{-\theta[-log(1-F(x-1;\Im))]^\alpha} \\ &\quad - \left[1 + \frac{\theta[-log(1-F(x;\Im))]^\alpha}{1+\theta} e^{-\theta[-log(1-F(x;\Im))]^\alpha}\right] e^{-\theta[-log(1-F(x;\Im))]^\alpha} \end{aligned} \tag{5}$$

where $\alpha > 0$, $\theta > 0$ and $\Im$ the parameters of the CDF $F(x;\Im)$, and the range of variation of PML-X family of distribution depends on the random variable X with CDF $F(x;\Im)$.

In the following section, we examine one member of this family, the power modified Lindley-geometric distribution, and provide its detailed features. The geometric distribution was chosen because it has a simplified CDF form.

3. Power modified Lindley-geometric distribution

Let's make the assumption that the transformed distribution is geometric with parameter p, $0 < p < 1$, and that the survival function $S(x) = 1 - F(x) = p^{(x+1)}$.Then, the

PMF of the new model using Equation (5) is given by

$$g(x;\alpha,\theta,b) = \left[1+\frac{\theta(bx)^{\alpha}}{1+\theta}e^{-\theta(bx)^{\alpha}}\right]e^{-\theta(bx)^{\alpha}} - \left[1+\frac{\theta(b(x+1))^{\alpha}}{1+\theta}e^{-\theta(b(x+1))^{\alpha}}\right]e^{-\theta(b(x+1))^{\alpha}};\quad x=0,1,2,.... \tag{6}$$

where $b = -log\ p > 0$, $\alpha > 0$ and $\theta > 0$. We call this new distribution the power modified Lindley-geometric (PMLG) distribution with parameters b, α and θ. Note that,

$$\lim_{x\to+\infty} g(x;\alpha,\theta,b) = 0,\ \lim_{x\to 0} g(x;\alpha,\theta,b) = 0\ when\ b\to 0\ and$$

$$\lim_{x\to 0} g(x;\alpha,\theta,b) = 1\ when\ b\to\infty.$$

The corresponding *CDF* is given by

$$G(x;\alpha,\theta,b) = 1-\left[1+\frac{\theta(b(x+1))^{\alpha}}{1+\theta}e^{-\theta(b(x+1))^{\alpha}}\right]e^{-\theta(b(x+1))^{\alpha}};\quad x=0,1,2,... \tag{7}$$

and the hazard rate function (*HRF*) corresponding to the *CDF* is provided by

$$h(x;\alpha,\theta,b)) = \frac{\left[1+\frac{\theta(bx)^{\alpha}}{1+\theta}e^{-\theta(bx)^{\alpha}}\right]e^{-\theta(bx)^{\alpha}} - \left[1+\frac{\theta(b(x+1))^{\alpha}}{1+\theta}e^{-\theta(b(x+1))^{\alpha}}\right]e^{-\theta(b(x+1))^{\alpha}}}{\left[1+\frac{\theta(b(x+1))^{\alpha}}{1+\theta}e^{-\theta(b(x+1))^{\alpha}}\right]e^{-\theta(b(x+1))^{\alpha}}}$$

A graphic illustration of the *PMF* of the PMLG distribution in various forms is shown in Figure 1. These graphs demonstrate the possibility of right-skewed, symmetric, left-skewed, increasing decreasing curves for the *PMF* of the PMLG distribution. The *HRF* of the PMLG distribution in Figure 2 is depicted in some of its potential shapes for various parameter values. Figures show that the *HRF* can have a variety of shapes, including increasing, decreasing and upside-down bathtub shapes. As a result, the PMLG distribution is excellent at modelling a variety of data sets.

3.1. Probability generating function, r^{th} moment function, mean and variance

The probability generating function (*PGF*) of PMLG distribution is given by

$$p(s) = 1+(s-1)\sum_{x=1}^{\infty} s^{x-1}\left[1+\frac{\theta(bx)^{\alpha}}{1+\theta}e^{-\theta(bx)^{\alpha}}\right]e^{-\theta(bx)^{\alpha}}. \tag{8}$$

Using Equation (6), the non-central r^{th} moment of the PMLG distribution can be calculated as follows:

$$\mu'_r = \sum_{x=0}^{\infty} x^r g(x;\alpha,\theta,b)$$
$$= \sum_{x=0}^{\infty} x^r\left[1+\frac{\theta(bx)^{\alpha}}{1+\theta}e^{-\theta(bx)^{\alpha}}\right]e^{-\theta(bx)^{\alpha}} - \sum_{x=0}^{\infty} x^r\left[1+\frac{\theta(b(x+1))^{\alpha}}{1+\theta}e^{-\theta(b(x+1))^{\alpha}}\right]e^{-\theta(b(x+1))^{\alpha}}.$$

In particular, the first two moments of the PMLG distribution are given by

$$\mu_1' = E(X) = \sum_{x=1}^{\infty} \left[1 + \frac{\theta(bx)^\alpha}{1+\theta} e^{-\theta(bx)^\alpha}\right] e^{-\theta(bx)^\alpha}. \tag{9}$$

$$\mu_2' = \sum_{x=1}^{\infty} (2x-1) \left[1 + \frac{\theta(bx)^\alpha}{1+\theta} e^{-\theta(bx)^\alpha}\right] e^{-\theta(bx)^\alpha}. \tag{10}$$

The variance of PMLG distribution is given as

$$V(X) = \sum_{x=1}^{\infty} (2x-1) \left[1 + \frac{\theta(bx)^\alpha}{1+\theta} e^{-\theta(bx)^\alpha}\right] e^{-\theta(bx)^\alpha} - \left[\sum_{x=1}^{\infty} \left[1 + \frac{\theta(bx)^\alpha}{1+\theta} e^{-\theta(bx)^\alpha}\right] e^{-\theta(bx)^\alpha}\right]^2. \tag{11}$$

Table 1 shows the mean and variance of the PMLG distribution for different values of b, α and θ using statistical software. From this, we are able to understand that the variance decreases with α and θ for different values of b. Furthermore, based on the values of b, α and θ, the mean can be equal, lower or larger than its variance. As a result, many data sets can be modelled using the characteristics of the PMLG distribution.

Table 1: The mean (variance) of PMLG for various choices of parameters

	$\alpha \longrightarrow$ $\theta \downarrow$	0.5	1	2
$b = 0.25$	0.5	15.1411(337.0099)	8.8333(61.9832)	5.1041(6.0108)
	1.5	3.3854(49.6625)	2.4523(7.0107)	2.5990(2.2115)
	2.5	1.0478(7.7934)	1.2522(2.4986)	1.8552(1.3915)
$b = 1$	0.5	7.4605(143.5667)	1.8484(3.8987)	0.9007(0.4569)
	1.5	0.6757(3.7225)	0.3203(0.3896)	0.2555(0.1952)
	2.5	0.1470(0.3978)	0.0943(0.1015)	0.0869(0.0795)
$b = 1.75$	0.5	4.6069(72.9978)	0.8633(1.2649)	0.2662(0.1998)
	1.5	0.3096(1.1287)	0.0837(0.0890)	0.0103(0.0102)
	2.5	0.0532(0.1039)	0.0129(0.0131)	0.0005(0.0005)

3.2. Infinite divisibility

The Central Limit Theorem and waiting time distributions are closely related to infinite divisibility. In accordance with Steutel and Van Harn (2003), If $p(x), x \in \mathbb{N}_0$ is infinitely divisible, then $p(x) < e^{-1}$ for all $x \in \mathbb{N}$. We can observe that for the PMLG distribution with parameters $b = 0.4$, α =2 and $\theta = 5$, $r(1) = 0.4346001 > e^{-1} = 0.367$. It follows that the PMLG distribution is not infinitely divisible. Additionally, since the discrete concepts of self-decomposable and stable distributions are subclasses of infinitely divisible distributions, we are able to conclude that the PMLG distribution cannot be either of these properties.

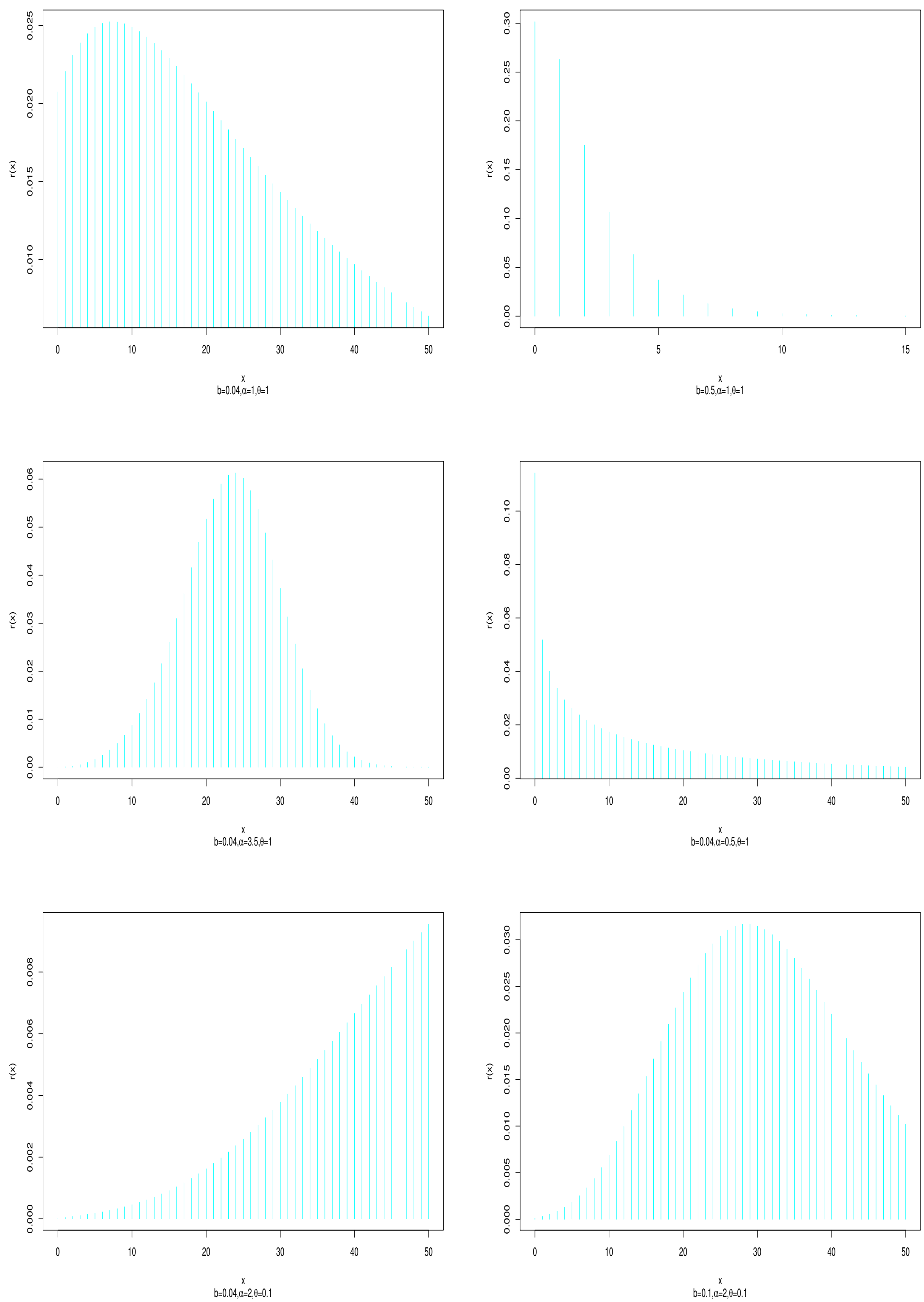


Figure 1: *PMF*s of some parameter values for the PMLG distribution

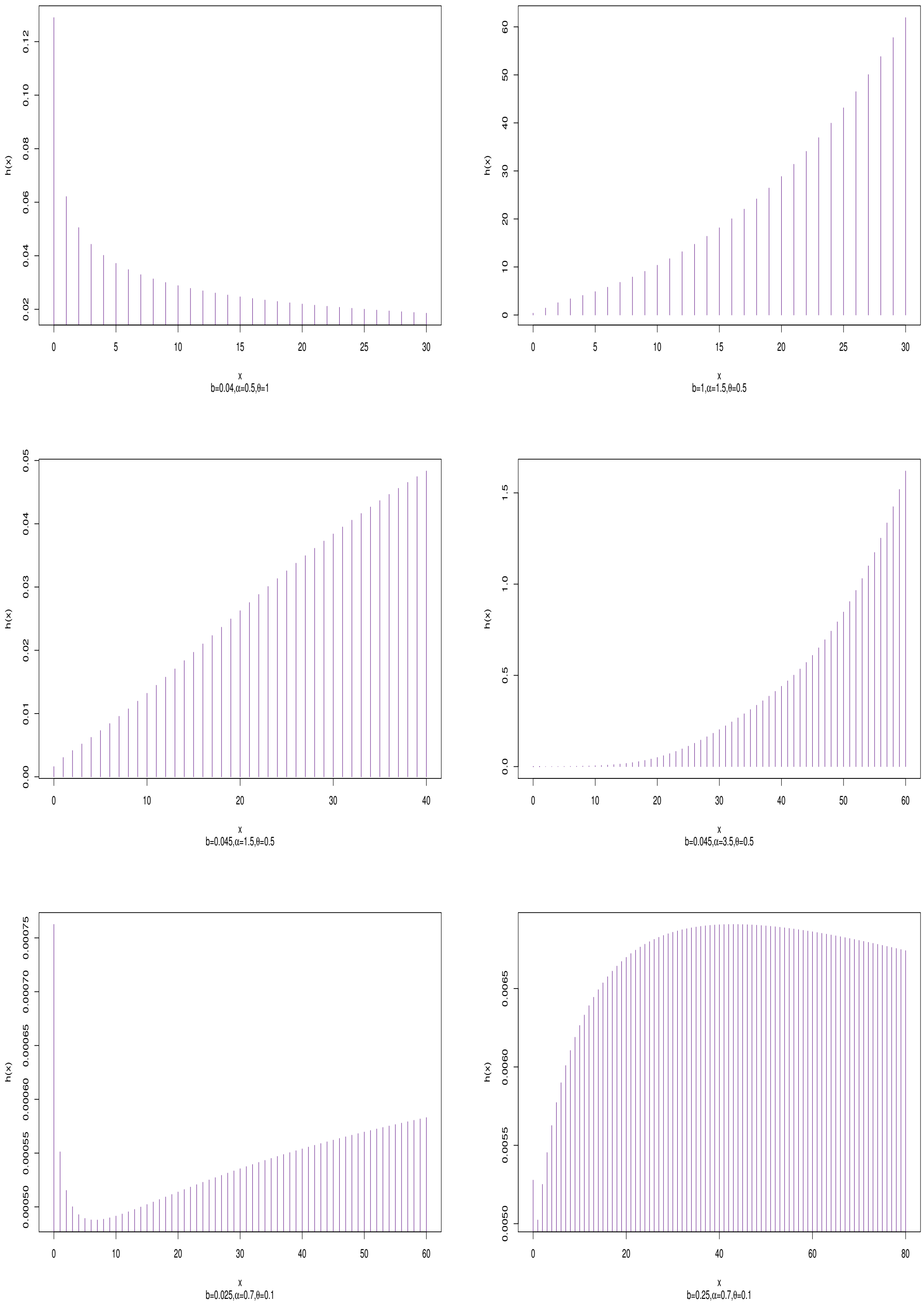


Figure 2: *HRF*s of some parameter values for the PMLG distribution

4. Parameter estimation

In this section, we focus on the many classical estimating techniques. There are numerous and different classical approaches, some of which rely on the theory of maximisation and others on the theory of minimization. This section includes, maximum likelihood, Cramer-von-Mises, least squares and weighted least squares approaches of estimation as part of four classical estimation methods.

4.1. Maximum likelihood approach of estimation

If we choose $x_1, x_2, \ldots, x_n$ to be a random sample from the PMLG distribution with unknown parameters b, α and θ and, the likelihood function is given by

$$
\begin{aligned}
L(\alpha,\theta,b) &= \prod_{i=1}^{n} g(x_i;\alpha,\theta,b)\\
&= \prod_{i=1}^{n}\left[1+\frac{\theta(bx_i)^{\alpha}}{1+\theta}e^{-\theta(bx_i)^{\alpha}}\right]e^{-\theta(bx_i)^{\alpha}}-\left[1+\frac{\theta(b(x_i+1))^{\alpha}}{1+\theta}e^{-\theta(b(x_i+1))^{\alpha}}\right]e^{-\theta(b(x_i+1))^{\alpha}}.
\end{aligned}
$$

The log-likelihood function follows immediately as

$$
\begin{aligned}
\ell(b,\alpha,\theta) &= \log\left[L(b,\alpha,\theta)\right]\\
&= \sum_{i=1}^{n} log\Bigg\{\left[1+\frac{\theta(bx_i)^{\alpha}}{1+\theta}e^{-\theta(bx_i)^{\alpha}}\right]e^{-\theta(bx_i)^{\alpha}}\\
&\quad -\left[1+\frac{\theta(b(x_i+1))^{\alpha}}{1+\theta}e^{-\theta(b(x_i+1))^{\alpha}}\right]e^{-\theta(b(x_i+1))^{\alpha}}\Bigg\}.
\end{aligned}
$$

The first derivatives of $\ell(b,\alpha,\theta)$ with respect to b, α and θ are

$$
\begin{aligned}
\frac{\partial\ell(b,\alpha,\theta)}{\partial b} &= \sum_{i=1}^{n}\frac{\frac{\alpha\theta\Delta_1}{b}\left\{\frac{e^{-\theta(bx_i)^{\alpha}}}{1+\theta}\left[1-\theta(bx_i)^{\alpha}\right]-\left[1+\frac{\theta}{1+\theta}\Delta_1\right]\right\}}{\left[1+\frac{\theta}{1+\theta}\Delta_1\right]e^{-\theta(bx_i)^{\alpha}}-\left[1+\frac{\theta}{1+\theta}\Delta_2\right]e^{-\theta(b(x_i+1))^{\alpha}}}\\
&\quad -\sum_{i=1}^{n}\frac{\frac{\alpha\theta\Delta_2}{b}\left\{\frac{e^{-\theta(b(x_i+1))^{\alpha}}}{1+\theta}\left[1-\theta(b(x_i+1))^{\alpha}\right]-\left[1+\frac{\theta}{1+\theta}\Delta_2\right]\right\}}{\left[1+\frac{\theta}{1+\theta}\Delta_1\right]e^{-\theta(bx_i)^{\alpha}}-\left[1+\frac{\theta}{1+\theta}\Delta_2\right]e^{-\theta(b(x_i+1))^{\alpha}}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\ell(b,\alpha,\theta)}{\partial\alpha} &= \sum_{i=1}^{n}\frac{\theta\Delta_1 log(bx_i)\left\{\frac{e^{-\theta(bx_i)^{\alpha}}}{1+\theta}\left[1-\theta(bx_i)^{\alpha}\right]-\left[1+\frac{\theta}{1+\theta}\Delta_1\right]\right\}}{\left[1+\frac{\theta}{1+\theta}\Delta_1\right]e^{-\theta(bx_i)^{\alpha}}-\left[1+\frac{\theta}{1+\theta}\Delta_2\right]e^{-\theta(b(x_i+1))^{\alpha}}}\\
&\quad -\sum_{i=1}^{n}\frac{\theta\Delta_2 log(b(x_i+1))\left\{\frac{e^{-\theta(b(x_i+1))^{\alpha}}}{1+\theta}\left[1-\theta(b(x_i+1))^{\alpha}\right]-\left[1+\frac{\theta}{1+\theta}\Delta_2\right]\right\}}{\left[1+\frac{\theta}{1+\theta}\Delta_1\right]e^{-\theta(bx_i)^{\alpha}}-\left[1+\frac{\theta}{1+\theta}\Delta_2\right]e^{-\theta(b(x_i+1))^{\alpha}}}
\end{aligned}
$$

$$\begin{aligned}\frac{\partial \ell(b,\alpha,\theta)}{\partial \theta} &= \sum_{i=1}^{n} \frac{\Delta_1 \left\{ \frac{e^{-\theta(bx_i)^\alpha}}{(1+\theta)^2} \left[(1+\theta)(1-\theta(bx_i)^\alpha)-\theta\right] - \left[1+\frac{\theta}{1+\theta}\Delta_1\right]\right\}}{\left[1+\frac{\theta}{1+\theta}\Delta_1\right] e^{-\theta(bx_i)^\alpha} - \left[1+\frac{\theta}{1+\theta}\Delta_2\right] e^{-\theta(b(x_i+1))^\alpha}} \\ &- \sum_{i=1}^{n} \frac{\Delta_2 \left\{ \frac{e^{-\theta(b(x_i+1))^\alpha}}{(1+\theta)^2} \left[(1+\theta)(1-\theta(b(x_i+1))^\alpha)-\theta\right] - \left[1+\frac{\theta}{1+\theta}\Delta_2\right]\right\}}{\left[1+\frac{\theta}{1+\theta}\Delta_1\right] e^{-\theta(bx_i)^\alpha} - \left[1+\frac{\theta}{1+\theta}\Delta_2\right] e^{-\theta(b(x_i+1))^\alpha}}\end{aligned}$$

Where, $\Delta_1 = e^{-\theta(bx_i)^\alpha}(bx_i)^\alpha$ and $\Delta_2 = e^{-\theta(b(x_i+1))^\alpha}(b(x_i+1))^\alpha$.

Setting $\frac{\partial \ell(b,\alpha,\theta)}{\partial b} = 0$, $\frac{\partial \ell(b,\alpha,\theta)}{\partial \alpha} = 0$ and $\frac{\partial \ell(b,\alpha,\theta)}{\partial \theta} = 0$, and then solving the equations iteratively will yield the maximum likelihood (ML) estimators of b, α and θ. These equations are complicated to solve analytically. One can use mathematical software to get numerical solutions.

4.2. Cramer-von-Mises approach of estimation

The Cramer-von-Mises (CVM) estimation approach is a significant estimation method that was discussed in Macdonald (1971). The CVM estimation technique's parameters can be calculated by minimising the function CVM in respect to the unknown parameters.

$$\begin{aligned}CVM &= \frac{1}{12} + \sum_{i=1}^{n} \left\{ G(x_i;\alpha,\theta,b) - \frac{2i-1}{2n} \right\} \\ &= \frac{1}{12} + \sum_{i=1}^{n} \left\{ 1 - \left[1 + \frac{\theta(b(x_i+1))^\alpha}{1+\theta} e^{-\theta(b(x_i+1))^\alpha}\right] e^{-\theta(b(x_i+1))^\alpha} - \frac{2i-1}{2n} \right\}\end{aligned}$$

4.3. Least square approach of estimation

Assume that $x_1, x_2, \ldots, x_n$ is a randomly selected sample of size n from the PMLG distribution and that $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ signifies a corresponding ordered sample. Consequently, the following quantity can be minimized to produce least squares (LS) estimators for PMLG parameters

$$\begin{aligned}LS &= \sum_{i=1}^{n} \left\{ G(x_{i:n};\alpha,\theta,b) - \frac{i}{n+1} \right\}^2 \\ &= \sum_{i=1}^{n} \left\{ 1 - \left[1 + \frac{\theta(b(x_{i:n}+1))^\alpha}{1+\theta} e^{-\theta(b(x_{i:n}+1))^\alpha}\right] e^{-\theta(b(x_{i:n}+1))^\alpha} - \frac{i}{n+1} \right\}^2\end{aligned}$$

with respect to b, α and θ respectively.

4.4. Weighted least square approach of estimation

The weighted least square (WLS) estimators of the unknown parameters for the PMLG distribution are derived in this subsection. Let $x_1, x_2, \ldots, x_n$ be a random sample and $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ be the corresponding ordered sample of size n from the PMLG distribution. The following sum of squares errors can be minimised to generate the PMLG estimators

$$WLS = \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left\{ G(x_{i:n}; \alpha, \theta, b) - \frac{i}{n+1} \right\}^2$$
$$= \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left\{ 1 - \left[1 + \frac{\theta(b(x_{i:n}+1))^\alpha}{1+\theta} e^{-\theta(b(x_{i:n}+1))^\alpha} \right] e^{-\theta(b(x_{i:n}+1))^\alpha} - \frac{i}{n+1} \right\}^2$$

with respect to b, α and θ respectively.

5. Simulation

Here, a simulation study is used to examine how well various estimates of the PMLG distribution work. Using the PMLG distribution, we produce random data with varying sample sizes and parameter values. The simulation research is run N= 1000 times with n = 50, 100, 150, and 200 as the sample size and the chosen parameter values. We compute the *ML*, *CVM*, *LS* and *WLS* estimates of b, α and θ. Based on the calculated results estimates, average biases (*Bias*) and mean squared errors (*MSE*s) measurements are calculated. The results of this simulation are shown in Tables 2 and 3. We can draw the following interpretations from the tables:

- With larger sample sizes, all estimates experience a decreasing trend in *MSE*s and *Bias* decays towards zero.
- The *LS* estimates *MSE*s are lower than those for the *ML*, *WLS*, and *CVM* estimates.

Table 2: The *Bias* and *MSE* of the *ML*, CVM, LS and WLS estimates for b=0.5, α=0.4 and θ=0.045

n		$Bias(\hat{b})$	$MSE(\hat{b})$	$Bias(\hat{\alpha})$	$MSE(\hat{\alpha})$	$Bias(\hat{\theta})$	$MSE\ (\hat{\theta})$
50	*ML*	0.0322	1.3889	-0.4728	0.2365	-0.3687	0.2769
	CVM	-0.0348	0.0829	-0.4892	0.2447	-0.3869	0.1582
	LS	-0.0005	0.2704e-04	-0.0001	0.4885e-06	-0.0001	0.1155e-05
	WLS	-0.0303	0.2046	-0.4945	0.2471	-0.3941	0.1595
100	*ML*	-0.0168	0.5318	-0.4434	0.2217	-0.3553	0.1421
	CVM	-0.0050	0.0335	-0.1889	0.0945	-0.1284	0.0627
	LS	-0.0002	0.2899e-05	-0.3202e-04	0.6908e-07	-0.4317e-04	0.1450e-06
	WLS	-0.0178	0.1858	-0.4163	0.2080	-0.3249	0.1388
150	*ML*	0.0267	0.4623	0.0003	0.4505e-04	0.0093	0.0537
	CVM	0.0009	0.0001	-0.0045	0.0023	0.0042	0.0020
	LS	-0.6792e-05	0.4697e-06	-0.2010e-05	0.1184e-07	-0.2032e-06	0.2626e-07
	WLS	-0.0084	0.0997	-0.2404	0.1202	-0.1848	0.0779
200	*ML*	0.0003	0.6853e-04	-0.1957e-04	0.3828e-06	-0.0002	0.5283e-04
	CVM	0.0003	0.5130e-04	-0.0010	0.0005	0.0009	0.0005
	LS	-0.3016e-06	0.9838e-07	-0.6776e-06	0.2865e-08	-0.7719e-06	0.590e-08
	WLS	0.37691e-04	0.1421e-05	-0.0005	0.0002	0.0005	0.0002

Table 3: The *Bias* and *MSE* of the *ML*, *CVM*, *LS* and *WLS* estimates for b=0.6, α=0.5 and θ=0.05

n		$Bias(\hat{b})$	$MSE(\hat{b})$	$Bias(\hat{\alpha})$	$MSE(\hat{\alpha})$	$Bias(\hat{\theta})$	$MSE\ (\hat{\theta})$
50	*ML*	-0.0143	0.3723	-0.5719	0.3436	-0.4532	0.5205
	CVM	-0.0172	0.1487	-0.4736	0.2841	-0.3554	0.19245
	LS	-0.0007	0.1823e-04	-0.0003	0.383e-04	0.0001	0.4367e-04
	WLS	-0.0459	0.0101	-0.5860	0.3519	-0.4846	0.2451
100	*ML*	-0.0043	-0.0043	-0.3523	0.2123	-0.2616	0.4132
	CVM	-0.0107	0.0010	-0.1783	0.1071	-0.1120	0.0706
	LS	-0.0002	0.2613e-05	-0.1761e-04	0.2295e-07	-0.2939e-04	0.7558 e-07
	WLS	-0.0220	0.0301	-0.3597	0.2156	-0.282	0.1592
150	*ML*	0.0028	0.0070	0.1515e-04	0.21181e-06	-0.0007	0.0003
	CVM	-0.0030	0.0002	-0.0389	0.0233	-0.0291	0.0157
	LS	-0.3115e-04	0.2720e-06	-0.3128e-05	0.2699e-08	-0.5645e-05	0.8846e-08
	WLS	0.0030	0.0095	-0.0027	0.0015	0.0026	0.0017
200	*ML*	0.0007	0.0002	-0.6898e-05	0.1166e-06	-0.0007	0.0002
	CVM	0.9024e-04	0.4348e-05	-0.0011	0.0006	0.0006	0.0002
	LS	-0.1707e-04	0.1457e-06	-0.1788e-05	0.1598e-08	-0.3128e-06	0.4894e-08
	WLS	0.0009	0.0007	-0.0012	0.0007	0.0007	0.0003

6. Application

This section uses two actual count data sets to demonstrate the significance of the PMLG distribution over the existing models, namely exponentiated exponential-geometric (EEG) distribution and Kumaraswamy-geometric (KG) distribution, in modelling count data from the field of medicine. We used the maximum likelihood method to estimate the values of the unknown parameters in order to compare these distributions. Additionally, the estimated log-likelihood function ($\hat{\ell}$), Akaike Information Criterion (AIC), correct Akaike information criterion ($AICc$), Anderson-Darling statistic (A), Cramér von Mises statistic (W) and Kolmogorov-Smirnov (K-S) statistic with p-value (p-V) are used to compare the fitted distributions. The following displays the considered data sets.

Data set I:The first data set shows the number of COVID-19-related deaths that occurred on a daily basis in the United Kingdom from August 1 through August 28, 2021. This information is obtained from the website
https://www.worldometers.info/coronavirus/country/uk/, which lists the number of deaths caused by COVID-19 in the United Kingdom on a daily basis. The data set is provided below.
{65, 24, 138, 119, 86, 92, 103, 39, 37, 140, 104, 94, 100, 91, 61, 26, 170, 111, 113, 114, 104, 49, 40, 174, 149, 140, 100, 133}

Data set II:The second data set, which has 42 observations and is available on the Worldometer website through
https://www.worldometers.info/coronavirus/country/Egypt/, shows the number of daily COVID-19 infection-related deaths that occurred in Egypt from 13 March to 30 April 2020. The data are as follows.
{1, 2, 4, 5, 1, 1, 3, 6, 6, 4, 1, 5, 6, 6, 8, 5, 7, 7, 9, 9, 15, 17, 11, 13, 5, 14, 5, 13, 9, 19, 15, 11,

14, 12, 11, 7, 13, 10, 20, 22, 21, 12}

Tables 4, 5, 6 and 7, contain the MLEs, $(-\hat{\ell})$, AIC and goodness-of-fit tests for COVID-19 data sets. The analysis yields the PMLG distribution with the lowest $-\hat{\ell}$, AIC, $AICc$, $HQIC$, A, W, K-S statistic, and highest p-Vs. The PMLG distribution is the appropriate one based on these results. We can say from the two applications that the PMLG distribution is the best model for capturing the daily deaths by COVID-19.

Figure 3 gives the total time test (TTT)-plots of PMLG distribution for the COVID-19 data sets. The TTT-plots shows increasing HRF, allowing us to fit PMLG distribution. Figure 4 display the probability-probability (PP) plots for the two data sets, respectively. The PMLG distribution offers a better fit for the COVID-19 data sets, which support the findings in Tables 4, 5, 6 and 7.

Table 4: Estimated values, $-\hat{\ell}$, AIC, and $AICc$ for the data set I

Distribution	Estimates	$-\hat{\ell}$	AIC	AICc
PMLG	$\hat{\alpha} = 2.5303$ $\hat{\theta} = 8.6018$ $\hat{b} = 0.0039$	**143.5036**	**293.0072**	**294.0072**
EEG	$\hat{\alpha} = 4.8971$ $\hat{\theta} = 0.9774$	146.008	296.0161	296.4961
KG	$p = 0.9941$ $\hat{\alpha} = 3.1638$ $\hat{\theta} = 10.4923$	144.3038	294.6075	295.6075

Table 5: A, W and K-S with p-Vs for the data set I

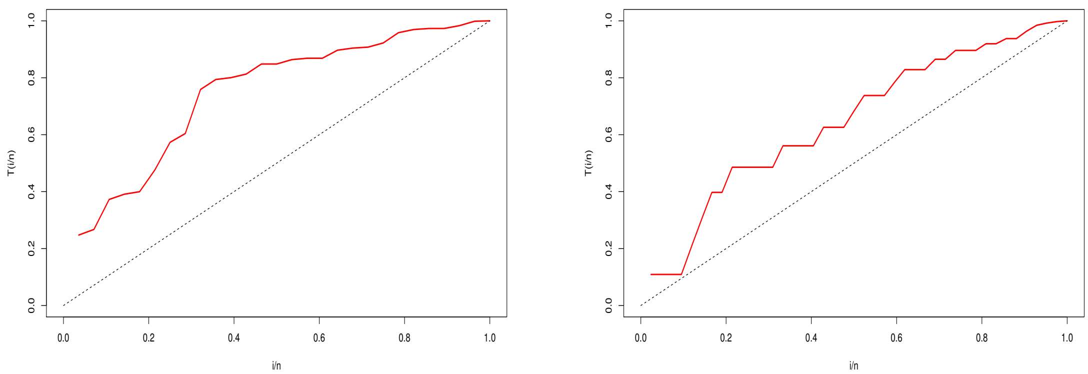
Distribution	A	W	K-S	p-Vs
PMLG	**0.5679**	**0.0941**	**0.14277**	**0.6179**
EEG	1.0461	0.2023	0.2072	0.1805
KG	0.7393	0.1351	0.1720	0.3785

Table 6: Estimated values, $-\hat{\ell}$ and AIC and $AICc$ for the data set II

Distribution	Estimates	$-\hat{\ell}$	AIC	AICc
PMLG	$\hat{\alpha} = 1.7357$ $\hat{\theta} = 11.6800$ $\hat{b} = 0.0227$	**129.0244**	**264.0489**	**264.6805**
EEG	$\hat{\alpha} = 2.6088$ $\hat{\theta} = 0.8385$	130.1956	264.3913	264.699
KG	$p = 0.9734$ $\hat{\alpha} = 1.9437$ $\hat{\theta} = 14.2437$	129.243	264.486	265.1176

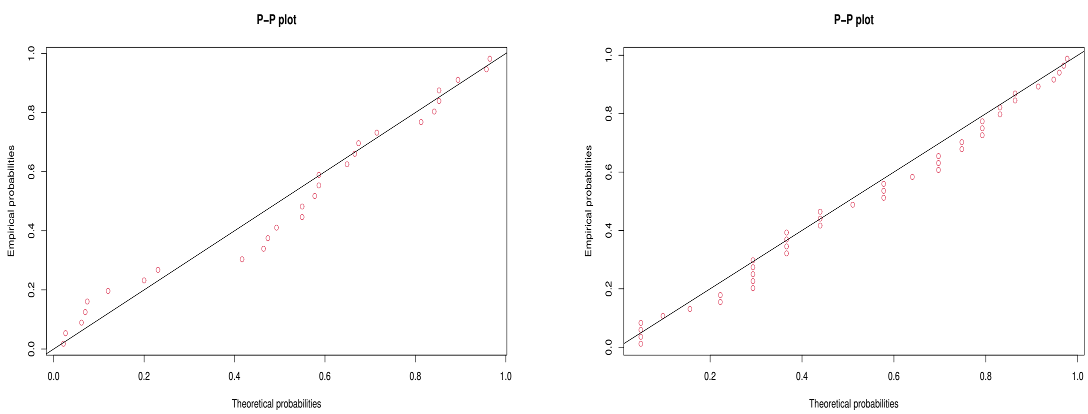
Table 7: A, W and K-S with p-Vs for the data set II

Distribution	A	W	K-S	p-Vs
PMLG	**0.4625**	**0.0725**	**0.10275**	**0.767**
EEG	0.8018	0.1447	0.13774,	0.403
KG	0.5264	0.0871	0.10979	0.692



(a) (b)

Figure 3: *TTT*-plots for the COVID-19 (a) data set I and (b) data set II



(a) (b)

Figure 4: *PP*-plots for the COVID-19 (a) data set I and (b) data set II

7. Conclusion

In this study, we suggested an entirely novel family of discrete PML-X distributions. PMLG distribution is a specific instance of this family that is thoroughly researched. *ML*, *CVM*, *OLS* and *WLS* techniques have been used to estimate the model parameters. A simulation study is conducted to evaluate the effectiveness of the various estimating techniques. In order to demonstrate the significance and adaptability of defined distribution, two real data sets are analysed at the end. We anticipate that the suggested model will replace various types of discrete distributions found in the statistical literature.

References

Akinsete, A., Famoye, F., and Lee, C. (2014). The Kumaraswamy-geometric distribution. *Journal of Statistical Distributions and Applications*, **1**, 1–21.

Alzaatreh, A., Lee, C., and Famoye, F. (2012). On the discrete analogues of continuous distributions. *Statistical Methodology*, **9**, 589–603.

Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, **71**, 63–79.

Chesneau, C., Tomy, L., and Gillariose, J. (2021a). A new modified Lindley distribution with properties and applications. *Journal of Statistics and Management Systems*, **24**, 1383–1403.

Chesneau, C., Tomy, L., and Jose, M. (2021b). Power modified Lindley distribution: Theory and applications. *Journal of Mathematical Extension*, **16**, 1–32.

Famoye, F. (2019). Exponentiated Weibull-geometric distribution and its application to count data. *Journal of Data Science*, **17**, 712–725.

Macdonald, P. (1971). Comments and queries comment on "an estimation procedure for mixtures of distributions" by choi and bulgren. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **33**, 326–329.

Steutel, F. W. and Van Harn, K. (2003). *Infinite Divisibilty of Probability Distributions on the Real Line*. CRC Press.

Tomy, L., Jose, M., and Jose, M. (2019). The T-X family of distributions: A retrospect. *Think India Journal*, **22**, 9407–9420.