A NEW COUNT DATA PROBABILITY MODEL: PROPERTIES AND APPLICATIONS

Morifat[1], Bilal Ahmad Para[2,*]
•
[1,2]Department of Mathematical Sciences, Islamic University of Science and Technology, Kashmir, India
[1]morifat13@gmail.com, [2,*]Corresponding author: parabilal@gmail.com

Abstract

In the field of count data analysis, over-dispersion poses significant challenges, often limiting the effectiveness of traditional Poisson model. To address this limitation, we propose a novel two parameter distribution as an extension of Poisson distribution namely two-Parameter Poisson Garima Distribution (TPPGD). This distribution enhances modeling flexibility for over-dispersed data, offering a superior fit for real-world datasets. In this paper we derive the theoretical properties of TPPGD, including its probability mass function, cumulative distribution function and different statistical properties. Parameter estimation has been done using the maximum likelihood method and the moment method. Finally, the validity of the proposed model is checked using different real world data sets.

Keywords: Poisson distribution, Two Parameter Garima distribution, Compounding, Count data.

1. Introduction

Numerous probability models are acquired by researchers in order to analyze a wide range of data from different industries, including engineering, agriculture, transportation, medical, and other sectors. Lots of well-known techniques are employed to serve the purpose of constructing new probability distributions. Discretization, the T-X family, and compounding are a few well-known methods that offer a very effective means of extending popular parametric families of distributions to match data sets that are not well-fitted by classical distributions. In statistics, compounding technique plays an important role in discussing complex phenomena by integrating multiple distributions to model real world data effectively. The compounding technique originated in the early 20th century and the main figures in advancing use of compounding distributions were Greenwood & Udny Yule [5] who introduced concepts like negative binomial distribution in the 20th century. The negative binomial arises as a compounding distribution when a Poisson variable is mixed with a gamma-distributed parameter to model over-dispersion in count data. Compounding involves combining two or more probability distributions, where one distribution is considered conditional on the parameters of another. If the mean of Poisson follows inverse Gaussian, the resulting distribution is Poisson inverse Gaussian (Holla [6]). Despite being extensively utilized across all domains, these models are not always able to effectively characterize the relationships between variables and cannot handle all forms of count data. In this context, new models are created to provide better outcomes and serve as substitutes for traditional count models.

Sankaran [7] obtained the compound Poisson distribution by combining the Poisson distribution with the distribution attributed to Lindley. A limiting version of the generalized negative binomial distribution is derived by Cousul & Jain [2] which is a new generalization of the Poisson distribution. Poisson inverse Gaussian regression proposed by Dean, Lawless & Willmot [3] is another Poisson mixing model that has been presented to explain discrete data that is overly dispersed.

Gerstenkorn [4] proposed a compound of the generalized negative binomial distribution with the generalized beta distribution. Sujatha distribution was studied by Shanker [11] while a Poisson mixture of Sujatha distribution was introduced by Shanker et al. [10]. Sen et al. [8] introduced the two parameter X-gamma distribution while a two-parameter Lindley Distribution with application was proposed by Altun et al. [1]. Shanker et al. [13] studied a Poisson mixture of a new two-parameter Sujatha distribution. Shanker [9] introduced Garima distribution. Shanker et al. [12] proposed a Poisson mixture of Garima distribution named, Poisson Garima distribution (PGD). We now present a novel discrete count data model in this paper by combining Poisson distribution with two parameter Garima distribution, since a more adaptable approach for statistical data analysis is required.

The rest of this paper is structured as follows. In section 1, the two-parameter Poisson Garima Distribution is demonstrated. The various statistical properties of distribution such as the moments, skewness, kurtosis, moment generating function, probability generating function, order statistics, etc are summarized in Section 2. Parameter estimation has been done using maximum likelihood (ML) method and moment method in Section 3. Section 4 deals with the applicability of two parameter Poisson Garima distribution in real life data which is illustrated by three real-life data sets.

1.1 Definition of Proposed Model Two Parameter Poisson Garima Distribution

Consider a random variable $X/\lambda \sim P(\lambda)$, where λ is itself a random variable following two parameter Garima distribution with parameters α and θ. Here two parameter Garima distribution is itself a mixture of $Exp\ (\theta)$ and $Gamma\ (2,\theta)$ with mixing proportion $\frac{\alpha\theta+1}{\alpha\theta+2}$ and its probability density function is given by

$$h(x;\theta,\alpha)=\frac{\theta e^{-\theta x}}{\alpha\theta+2}(\alpha\theta+\theta x+1); \qquad x>0\ ;\ \theta>0,\ \alpha>0\ . \tag{1}$$

Then determining the distribution that results from marginalizing over λ will be known as a compound of Poisson distribution with that of two parameter Garima distribution, which has been denoted by TPPGD $(X;\theta,\alpha)$. If $\alpha=1$, the distribution reduces to Garima distribution [9].

Theorem 1: The probability mass function of a two parameter Poisson Garima distribution is given as:

$$P(X=x)=\frac{\theta}{(1+\theta)^{x+2}(\alpha\theta+2)}\left(\theta^2\alpha+\theta\alpha+\theta\ x+2\theta+1\right); \quad x=1,2,3,\dots\ ;\ \alpha,\theta>0 \tag{2}$$

Proof: The PMF of two parameter Poisson Garima distribution, i.e., TPPGD $(X;\theta,\alpha)$ can be obtained by

$g(x\mid\lambda)=\frac{e^{-\lambda}\lambda^x}{x!};\ \ x=0,1,2,3,\dots;\ \ \lambda>0$, when its parameter λ follows two parameter Garima distribution with pdf,

$$h(\lambda;\theta,\alpha)=\frac{\theta}{\alpha\theta+2}(\alpha\theta+\theta\lambda+1)e^{-\theta\lambda}\quad ;\lambda>0,\ \theta>0,\ \alpha>0. \tag{3}$$

We have, $P(X=x)=\int_0^\infty g(x\mid\lambda)\,h(\lambda;\theta,\alpha)\,d\lambda$

$$P(X=x)=\frac{\theta}{x!(\alpha\theta+2)}\left(\alpha\theta\int_0^\infty e^{-\lambda(1+\theta)}\lambda^x d\lambda+\theta\int_0^\infty e^{-\lambda(1+\theta)}\lambda^{x+1}d\lambda+\int_0^\infty e^{-\lambda(1+\theta)}\lambda^x d\lambda\right) \tag{4}$$

$$P(X=x)=\frac{\theta}{(1+\theta)^{x+2}(\alpha\theta+2)}\left(\theta^2\alpha+\theta\alpha+\theta x+2\theta+1\right);\qquad x=0,1,2,3,\ldots;\theta>0,\alpha>0. \tag{5}$$

Which is the required PMF of TPPGD $(X;\theta,\alpha)$

If $\alpha=1$, the distribution reduces to Poisson Garima distribution [12].

Figure 1 exhibits the PMF plots of the proposed model for different values of the parameters θ and α.

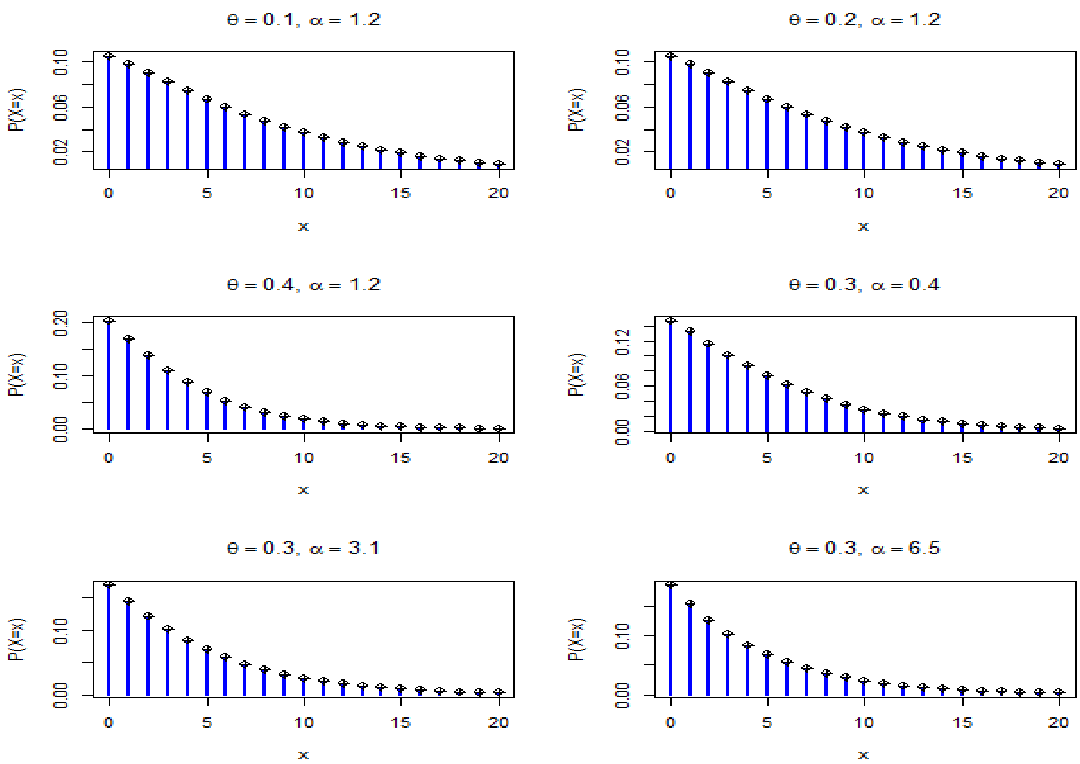

Figure 1: *PMF plots of two parameter Poisson Garima distribution.*

The CDF of two parameter Poisson Garima distribution is obtained as:

$$F(x)=\sum_{x=0}^{\infty}\frac{\theta}{(1+\theta)^{x+2}(\alpha\theta+2)}\left(\theta^2\alpha+\theta\alpha+\theta x+2\theta+1\right) \tag{6}$$

$$F(x)=\frac{1}{(1+\theta)^{x+2}(\alpha\theta+2)}\left[-2-\theta x-3\theta-\theta\alpha-\theta^2\alpha+(1+\theta)^{2+x}(\theta\alpha+2)\right]. \tag{7}$$

Figure 2 exhibits the CDF plots of the proposed model for different values of the parameters θ and α.

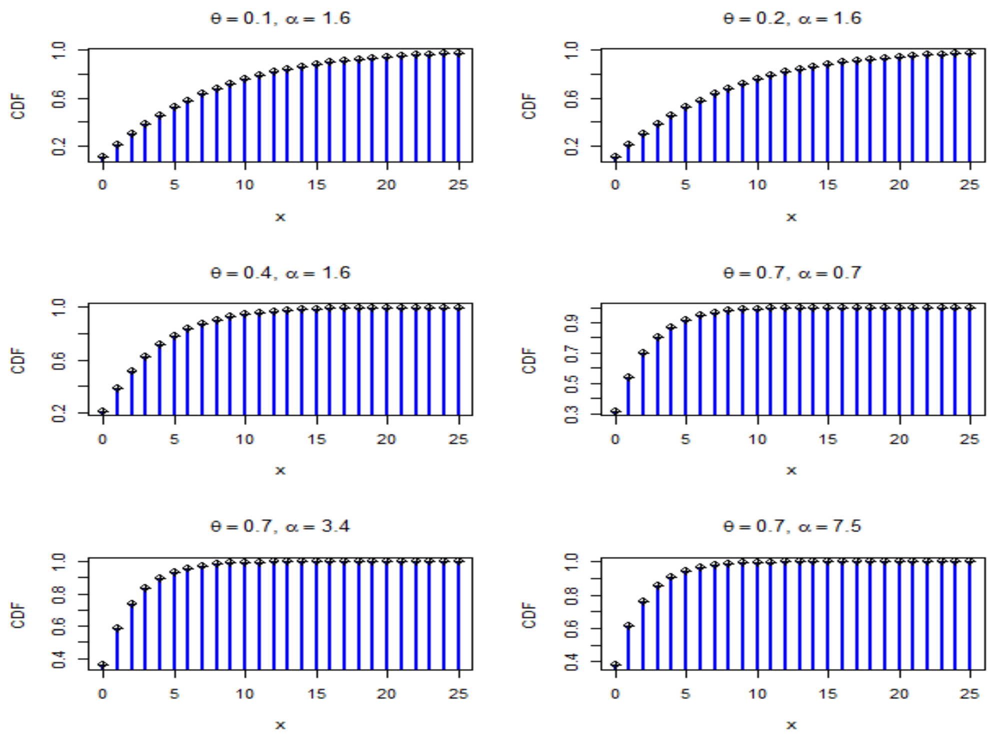

Figure 2: *CDF plots of two parameter Poisson Garima distribution.*

2. Statistical Properties of TPPGD

In this section, we will evaluate various properties of the two parameter Poisson Garima distribution. These include moments, index of dispersion, moment generating function and probability generating function, order statistics, and survival measures.

2.1 Moments

Using Eq. (2), the rth factorial moment about origin of the TPPGD $(X;\theta,\alpha)$ can be obtained as

$$\mu_{(r)}' = E[E(x^{(r)} \mid \lambda)] \quad , \text{ where } x^{(r)} = x(x-1)(x-2)...(x-r+1)$$

$$\mu_{(r)}' = \int_0^\infty \left[\sum_0^\infty x^r \frac{e^{-\lambda}\lambda^x}{x!}\right] \frac{\theta}{\alpha\theta+2}(\alpha\theta+\theta\lambda+1)e^{-\theta\lambda}\, d\lambda \tag{8}$$

$$\mu_{(r)}' = \frac{\theta}{\alpha\theta+2}\int_0^\infty \left[\lambda^r\left(\sum_{x=r}^\infty \frac{e^{-\lambda}\lambda^{x-r}}{(x-r)!}\right)\right](\alpha\theta+\theta\lambda+1)e^{-\theta\lambda}\, d\lambda \tag{9}$$

Taking $u = x - r$, we get

$$\mu_{(r)}' = \frac{\theta}{\alpha\theta+2}\int_0^\infty \left[\lambda^r\left(\sum_{u=0}^\infty \frac{e^{-\lambda}\lambda^{u}}{u!}\right)\right](\alpha\theta+\theta\lambda+1)e^{-\theta\lambda}\, d\lambda \tag{10}$$

$$\mu_{(r)}' = \frac{\theta}{\alpha\theta+2}\left(\theta\alpha\int_0^\infty \lambda^r e^{\theta\lambda} d\lambda + \theta\int_0^\infty \lambda^{r+1} e^{-\theta\lambda} d\lambda + \int_0^\infty \lambda^r e^{\theta\lambda} d\lambda\right) \tag{11}$$

$$\mu_{(r)}' = \frac{r!}{\alpha\theta+2}\left[\frac{\alpha\theta+(r+2)}{\theta^r}\right] \tag{12}$$

Taking r=1, 2, 3, 4 in Eq. (12), the first four factorial moments about origin of TPPGD can be obtained as

$$\mu'_{(1)} = \frac{\alpha\theta + 3}{\theta(\alpha\theta + 2)} \tag{13}$$

$$\mu'_{(2)} = \frac{2(\alpha\theta + 4)}{\theta^2(\alpha\theta + 2)} \tag{14}$$

$$\mu'_{(3)} = \frac{6(\alpha\theta + 5)}{\theta^3(\alpha\theta + 2)} \tag{15}$$

$$\mu'_{(4)} = \frac{24(\alpha\theta + 6)}{\theta^4(\alpha\theta + 2)} \tag{16}$$

The first four moments about origin, using the relationship between factorial moments about origin and the moments about origin of TPPGD $(X;\theta,\alpha)$.

$$\mu'_1 = \frac{\alpha\theta + 3}{\theta(\alpha\theta + 2)} \tag{17}$$

$$\mu'_2 = \frac{\alpha\theta^2 + 2\alpha\theta + 3\theta + 8}{\theta^2(\alpha\theta + 2)} \tag{18}$$

$$\mu'_3 = \frac{\alpha\theta^3 + 6\alpha\theta^2 + 3\theta^2 + 6\alpha\theta + 24\theta + 30}{\theta^3(\alpha\theta + 2)} \tag{19}$$

$$\mu'_3 = \frac{\alpha\theta^4 + 14\alpha\theta^3 + 36\alpha\theta^2 + 56\theta^2 + 3\theta^3 + 24\alpha\theta + 180\theta + 144}{\theta^3(\alpha\theta + 2)} \tag{20}$$

Mean, variance and index of dispersion of the proposed model are given by

$$\text{Mean} = \mu'_1 = \frac{\alpha\theta + 3}{\theta(\alpha\theta + 2)} \tag{21}$$

$$\text{Variance} = \mu_2 = \frac{\alpha^2\theta^3 + \alpha^2\theta^2 + 5\alpha\theta^2 + 6\alpha\theta + 6\theta + 7}{\theta^2(\alpha\theta + 2)^2} \tag{22}$$

$$\text{Index of dispersion} = 1 + \frac{\alpha^2\theta^2 + 6\alpha\theta + 7}{\theta(\alpha\theta + 2)(\alpha\theta + 3)} \tag{23}$$

The corresponding kurtosis and skewness were obtained as follows:

$$K_s = \frac{\mu_4}{(\mu_2)^2} = \frac{\begin{pmatrix} \alpha^4\theta^7 + 9\alpha^4\theta^4 + 402\alpha^2\theta^2 + 108\alpha^3\theta^3 + 10\alpha^4\theta^6 + 18\alpha^4\theta^5 + 198\alpha^3\theta^4 + 100\alpha^3\theta^5 + 9\alpha^3\theta^6 + \\ 356\alpha^2\theta^4 + 30\alpha^2\theta^5 + 726\alpha^2\theta^3 + 544\alpha\theta^3 + 44\alpha\theta^4 + 314\alpha^4\theta^7 + 612\alpha\theta + 304\theta^2 + 24\theta^3 + 612\theta + 333 \end{pmatrix}}{\theta^4(\alpha\theta + 2)^4}$$

$$S_k = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\begin{pmatrix} \alpha^3\theta^5 + 3\alpha^3\theta^4 + 24\alpha^2\theta^3 + 7\alpha^2\theta^4 + 2\alpha^3\theta^3 + 18\alpha^2\theta^2 + \\ 57\alpha\theta^2 + 16\alpha\theta^3 + 12\theta^2 + 42\alpha\theta + 42\theta + 30 \end{pmatrix}}{\left(\alpha^2\theta^3 + \alpha^2\theta^2 + 5\alpha\theta^2 + 6\alpha\theta + 6\theta + 7\right)^{\frac{3}{2}}} \tag{24}$$

2.2 Probability Generating Function (pgf) and Moment Generating Function (mgf)

We derive moment generating function and probability generating function of TPPGD in this sub section.

Theorem 2: If X follows TPPGD $(X;\theta,\alpha)$, then the probability generating function $P_X(t)$ has the following form

$$P_X(t) = \frac{\theta}{(\alpha\theta + 2)(\theta + 1)}\left[\frac{\theta t}{(1 + \theta - t)^2} + \frac{(\theta^2\alpha + \theta\alpha + 2\theta + 1)}{(1 + \theta - t)}\right] \tag{25}$$

Proof: The definition of probability generating function is given by

$$P_X(t)=\sum_{x=o}^{\infty}t^x\left[\frac{\theta}{(\alpha\theta+2)(1+\theta)^{x+2}}\left(\theta^2\alpha+\theta\alpha+\theta x+2\theta+1\right)\right] \tag{26}$$

$$P_X(t)=\frac{\theta}{(\alpha\theta+2)(1+\theta)^2}\left[\theta\sum_{x=0}^{\infty}x\left(\frac{t}{1+\theta}\right)^x+\theta^2\alpha\sum_{x=0}^{\infty}\left(\frac{t}{1+\theta}\right)^x+\theta\alpha\sum_{x=0}^{\infty}\left(\frac{t}{1+\theta}\right)^x+2\theta\sum_{x=0}^{\infty}\left(\frac{t}{1+\theta}\right)^x\sum_{x=0}^{\infty}\left(\frac{t}{1+\theta}\right)^x+...\right]$$

$$P_X(t)=\frac{\theta}{(\alpha\theta+2)(1+\theta)^2}\left[\frac{\theta(\theta+1)t}{(\theta+1-t)^2}+\frac{\left(\theta^2\alpha+\theta\alpha+2\theta+1\right)(\theta+1)}{(\theta+1-t)}\right] \tag{27}$$

$$P_X(t)=\frac{\theta}{(\alpha\theta+2)(1+\theta)}\left[\frac{\theta t}{(\theta+1-t)^2}+\frac{\theta^2\alpha+\theta\alpha+2\theta+1}{(\theta+1-t)}\right] \tag{28}$$

Which is the required pgf of TPPGD $(X;\theta,\alpha)$.

Theorem 3: If X follows TPPGD $(X;\theta,\alpha)$, then the moment generating function has the following form

$$M_x(t)=\frac{\theta}{(1+\theta)(\alpha\theta+2)}\left[\frac{\theta e^t}{\left(\theta+1-e^t\right)^2}+\frac{\theta^2\alpha+\theta\alpha+2\theta+1}{\left(\theta+1-e^t\right)}\right] \tag{29}$$

Proof: The definition of the moment generating function given by

$$M_X(t)=\sum_{x=0}^{\infty}e^{tx}\left[\frac{\theta}{(1+\theta)^{x+2}(\alpha\theta+2)}\left(\theta^2\alpha+\theta\alpha+\theta x+2\theta+1\right)\right] \tag{30}$$

$$M_x(t)=\frac{\theta}{(1+\theta)^2(\alpha\theta+2)}\left[\frac{\theta(\theta+1)e^t}{\left(\theta+1-e^t\right)^2}+\frac{\left(\theta^2\alpha+\theta\alpha+2\theta+1\right)(\theta+1)}{\left(\theta+1-e^t\right)}\right] \tag{31}$$

$$M_x(t)=\frac{\theta}{(1+\theta)(\alpha\theta+2)}\left[\frac{\theta e^t}{\left(\theta+1-e^t\right)^2}+\frac{\theta^2\alpha+\theta\alpha+2\theta+1}{\left(\theta+1-e^t\right)}\right] \tag{32}$$

Which is the required mgf of TPPGD $(X;\theta,\alpha)$.

2.3 Order Statistics

Let us suppose $x_1,x_2,x_3,...,x_n$ be a random sample of size n from two parameter Poisson Garima distribution with PMF $P(X=x)$ and CDF $F(x)$. Also, let $x_{(1)}\le x_{(2)}\le....\le x_{(n)}$, denote the corresponding order statistics.

Now let $F(x)$ be the CDF of the i^{th} order statistics for a random sample $x_1,x_2,x_3,...,x_n$ from X~TPPGD $(X;\theta,\alpha)$ distribution, then the CDF is defined by

$$F_i(x)=\sum_{k=1}^{n}\binom{n}{k}[F(x)]^k[1-F(x)]^{n-k} \tag{33}$$

By using the binomial expression for $[1-F(x)]^{n-k}$ and substituting the value in above given equation, we obtain the following result

$$F_i(x)=\sum_{k=1}^{n}\sum_{r=0}^{n-k}\binom{n}{k}\binom{n-k}{r}(-1)^r[1-F(x;\theta,\alpha)]^{k+r} \tag{34}$$

$$F_i(x)=\sum_{k=1}^{n}\sum_{r=0}^{n-k}\binom{n}{k}\binom{n-k}{r}(-1)^r\left[\frac{1}{(1+\theta)^{x+2}(\alpha\theta+2)}\left[-2-\theta x-3\theta-\theta\alpha-\theta^2\alpha+(1+\theta)^{2+x}(\theta\alpha+2)\right]\right]^{k+r}$$

The corresponding PMF of the i^{th} order statistic is $x_{(i)}$

$$P_i(X = x) = F_i(x) - F_i(x-1) \tag{35}$$

$$P_i(X = x) = \sum_{k=1}^{n}\sum_{r=0}^{n-k}(-1)^r\binom{n}{k}\binom{n-k}{r}\left[\frac{1}{(1+\theta)^{x+2}(\alpha\theta+2)}\left[-2-\theta x-3\theta-\theta\alpha-\theta^2\alpha+(1+\theta)^{2+x}(\theta\alpha+2)\right]\right]^{(k+r)}$$
$$-\left[\frac{1}{(1+\theta)^{x+1}(\alpha\theta+2)}\left[-2-\theta x-3\theta-\theta\alpha-\theta^2\alpha+(1+\theta)^{2+x}(\theta\alpha+2)\right]\right]^{(k+r)}$$

Particularly, by putting i=1 and i=n in above equation, we can obtain the PMF of first order statistics and nth order statistics.

2.4. Random Data Generation from Two Parameter Poisson Garima Distribution

In order to simulate the data from two parameter Poisson Garima distribution, we employ the inverse CDF method. Simulating a sequence of random numbers $x_1, x_2, x_3, ..., x_n$ of the two parameter Poisson Garima random variable X with PMF $P(X = x_i) = p_i, \sum_{i=0}^{k} p_i = 1$ and CDF $F(X;\theta,\alpha)$.where k may be finite or infinite can be described as in the following steps.

Step 1: Generate a random number u from uniform distribution $U(0,1)$.

Step 2: Generate random number x_i based on

$$\text{if } u \le p_o = F(x_o;\theta,\alpha) \text{ then } X = x_o$$
$$\text{if } p < u \le p_o + p_1 = F(x_1;\theta,\alpha) \text{ then } X = x_1$$
$$\vdots$$
$$\text{if } \sum_{j=0}^{k-1} p_j < u \le \sum_{j=0}^{k} p_j = F(x_k;\theta,\alpha) \text{ then } X = x_k$$

In order to generate n random numbers $x_1, x_2, x_3 ..., x_n$ from TPPGD $(X;\theta,\alpha)$ distribution, repeat the step 1 to step 2 n times. We have employed R software for running the simulation study of the proposed model.

2.5. Survival Measures of Two Parameter Poisson Garima Distribution

This sub-section deals with the survival measures of two parameter Poisson Garima distribution such as the survival function and the hazard function. The survival function, also known as the survivorship function, refers to the probability that a life, system or a component will survive beyond a specified time. In mathematical terms, it happens to be the complement of the CDF and is given by:

$$S(x;\theta,\alpha) = p(X > x) = 1 - F(x) \tag{36}$$

Using (7) in (36), we obtain the survival function of as follows:

$$S(x;\theta,\alpha) = 1 - \frac{1}{(1+\theta)^{x+2}(\alpha\theta+2)}\left[-2-\theta x-3\theta-\theta\alpha-\theta^2\alpha+(1+\theta)^{2+x}(\theta\alpha+2)\right] \tag{37}$$

Figure 3 exhibits the reliability function plot of the proposed model for different values of the parameters θ and α. The hazard function, also known as the hazard rate or failure rate or force of mortality, happens to be an important quantity used for the characterization of life phenomenon. Hazard function is defined as the conditional probability that a life, system or a component that survives up to a specified time, will undergo failure or succumb in the immediate, infinitesimally small interval of time that follows. In mathematical terms, the hazard rate or the hazard function is given by:

$$h(x) = \lim_{\Delta t \to 0} \frac{\Pr[t \le X < t + \Delta t \mid X \ge t]}{\Delta t} \tag{38}$$

Which upon simplification yields

$$h(x;\theta,\alpha) = \frac{P(X = x)}{S(x;\theta,\alpha)} \tag{39}$$

Using (2) and (10) in (11), we obtain the hazard function of TPPGD as follows:

$$h(x;\theta,\alpha) = \frac{\theta\left(\theta^2\alpha + \theta\alpha + \theta x + 2\theta + 1\right)}{1 - \left[-2 - \theta x - 3\theta - \theta\alpha - \theta^2\alpha + (1+\theta)^{2+x}(\theta\alpha + 2)\right]}. \tag{40}$$

The reverse hazard rate function of two parameter Poisson Garima distribution is given as:

$$h_r(x;\theta,\alpha) = \frac{P(X = x)}{F(x)} = \frac{\theta\left(\theta^2\alpha + \theta\alpha + \theta x + 2\theta + 1\right)}{\left[-2 - \theta x - 3\theta - \theta\alpha - \theta^2\alpha + (1+\theta)^{2+x}(\theta\alpha + 2)\right]}. \tag{41}$$

3. Estimation of parameters

In this section, we discuss the parameter estimation of the TPPGD $(X;\theta,\alpha)$ using method of maximum likelihood estimation and the frequent approach such as the method of moments.

3.1 Method of Maximum likelihood Estimate (MLE)

This is one of the most useful methods for estimating the different parameters of the distribution. Let $x_1, x_2, x_3, ..., x_n$ be the random sample of size n, drawn from two parameter Poisson Garima distribution, then the likelihood function of TPPGD $(X;\theta,\alpha)$ is given as

$$L(x_i \mid \theta) = \left(\frac{\theta}{\alpha\theta + 2}\right)^n \prod_{i=1}^{n} \left(\frac{\left(\theta^2\alpha + \theta\alpha + \theta x_i + 2\theta + 1\right)}{(1+\theta)^{x_i+2}}\right) \tag{42}$$

The log likelihood function is obtained as

$$\log L = n\log\theta - n\log(\alpha\theta + 2) + \sum_{i=1}^{n} \log\left(\theta^2\alpha + \theta\alpha + \theta x_i + 2\theta + 1\right) - \left(\sum_{i=1}^{n} x_i + 2n\right)\log(1+\theta) \tag{43}$$

The two likelihood equations are thus obtained as

$$\frac{d\log l}{d\theta} = \frac{2n}{\theta(\alpha\theta + 2)} + \sum_{i=1}^{n} \frac{(x_i + 2\alpha\theta + \alpha + 2)}{\theta x_i + \theta^2\alpha + \theta\alpha + 2\theta + 1} - \frac{n(\bar{x} + 2)}{\theta + 1} = 0 \tag{44}$$

$$\frac{d\log l}{d\alpha} = -\frac{n\theta}{\alpha\theta + 2} + \sum_{i=1}^{n} \frac{\theta^2 + \theta}{\theta x_i + \theta^2\alpha + \theta\alpha + 2\theta + 1} = 0 \tag{45}$$

The two equations (44) and (45) do not seem to be solved directly. However, the Fisher's scoring method can be applied to solve the equations.

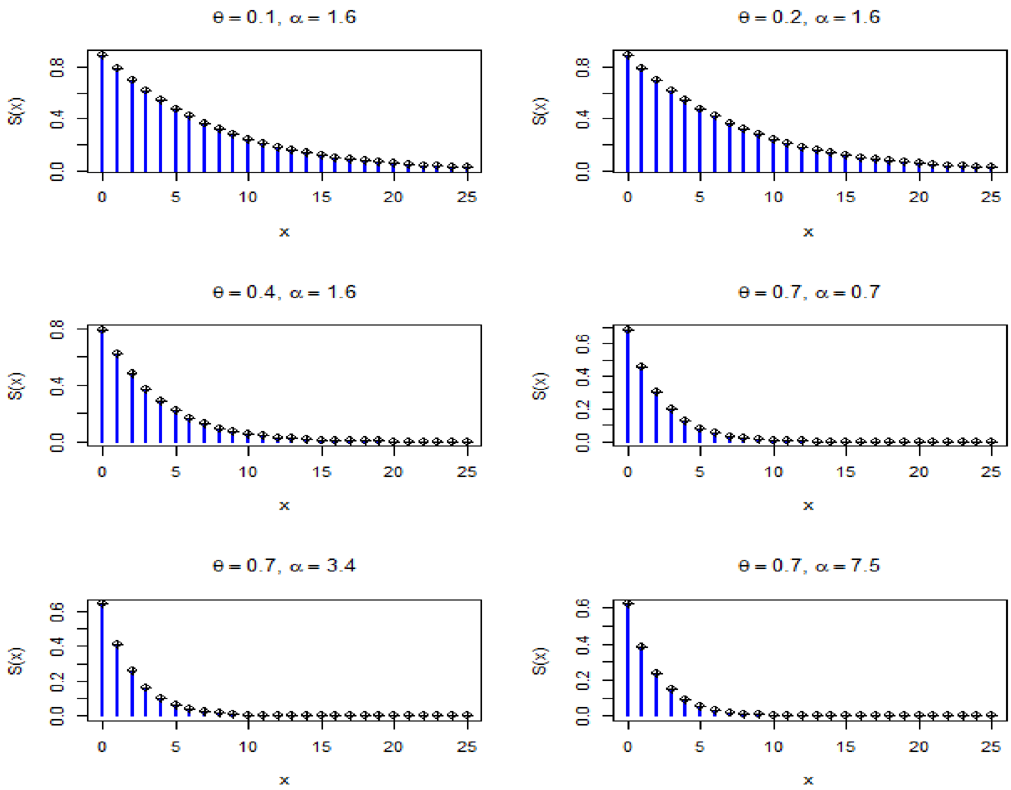

Figure 3: *Survival function plots of two parameter Poisson Garima distribution*

We have,

$$\frac{d^2\log l}{d\theta^2}=4n\left[\frac{\alpha\theta+1}{\theta^2(\alpha\theta+2)^2}\right]+\sum_{i=1}^{n}\frac{2\alpha\left(\theta^2\alpha+\theta\alpha+\theta x_i+2\theta+1\right)-\left(2\alpha\theta+\alpha+x_i+2\right)^2}{\left(\theta^2\alpha+\theta\alpha+\theta x_i+2\theta+1\right)^2}+\frac{n(\bar{x}+2)}{(1+\theta)^2}$$

$$\frac{d^2\log l}{d\alpha^2}=\frac{n\theta^2}{(\alpha\theta+2)^2}-\sum_{i=1}^{n}\frac{\left(\theta+\theta^2\right)^2}{\left(\theta^2\alpha+\theta\alpha+\theta x_i+2\theta+1\right)^2} \tag{46}$$

$$\frac{d^2\log l}{d\theta\, d\alpha}=\frac{-2n\theta^2}{(\theta(\alpha\theta+2))^2}+\sum_{i=1}^{n}\frac{\theta^2x_i+2\theta^2+2\theta+1}{\left(\theta^2\alpha+\theta\alpha+\theta x_i+2\theta+1\right)^2} \tag{47}$$

The following equation for $\hat{\theta}$ and $\hat{\alpha}$ can be solved

$$\begin{bmatrix}\dfrac{d^2\log l}{d\theta^2} & \dfrac{d^2\log l}{d\theta\, d\alpha}\\ \dfrac{d^2\log l}{d\theta\, d\alpha} & \dfrac{d^2\log l}{d\alpha^2}\end{bmatrix}_{\substack{\hat{\theta}=\theta_0\\ \hat{\alpha}=\alpha_0}}\begin{bmatrix}\hat{\theta}-\theta_0\\ \hat{\alpha}-\alpha_0\end{bmatrix}=\begin{bmatrix}\dfrac{d\log l}{d\theta}\\ \dfrac{d\log l}{d\alpha}\end{bmatrix}_{\substack{\hat{\theta}=\theta_0\\ \hat{\alpha}=\alpha_0}} \tag{48}$$

Where θ_0 and α_0 are the initial values of θ and α respectively. The equations are solved iteratively till sufficiently close estimates of $\hat{\theta}$ and $\hat{\alpha}$ are obtained.

3.2. Estimates from Moments

Using the first two moments about origin of two parameter Poisson Garima Distribution, we have

$$\frac{\mu_2'}{\mu_1'^2}=\frac{8+2\alpha\theta(\alpha\theta+2)}{(\alpha\theta+3)^2}=k \tag{49}$$

Taking $\alpha\theta = b$, we get

$$\frac{\mu_2'}{\mu_1'^2} = \frac{(8+2b)(b+2)}{(b+3)^2} = \frac{(2b^2+12b+16)}{(b^2+6b+9)} = k \tag{50}$$

This gives

$$b^2(2-k) + 6b(2-k) - 1(16-9k) = 0 \tag{51}$$

Which is a quadratic equation in b. Replacing the first and the second moments μ_1' and μ_2' by the respective sample moments $\bar{x}$ and m_2' an estimate of k can be obtained, using which, the equation (51) can be solved and an estimate of b obtained. Again substituting $\alpha\theta = b$ in this expression for the mean of the TPPGD$(X;\theta,\alpha)$, we get

$$\bar{x} = \frac{\alpha(b+2)}{b(b+1)}. \tag{52}$$

And thus, an estimate of α is given by

$$\hat{\alpha} = \left(\frac{b(b+1)}{b+2}\right)\bar{x}. \tag{53}$$

Finally, an estimate of θ is given by

$$\hat{\theta} = \left(\frac{b+2}{b+1}\right)\frac{1}{\bar{x}}. \tag{54}$$

4. Applications of Two Parameter Poisson Garima Distribution

In this section, we fit our proposed distribution to three practical datasets.

Data set 1: The first data set is regarding the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967. Shanker et Al. [14]

Table 1: *Thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of July, January 1957 to December 1967.*

No. of Thunderstorms	0	1	2	3	4	5	6	Total
Frequency	187	77	40	17	6	2	1	330

In order to give an impression of the typical form of our observed distributions and their various parameter estimates, we have provided Table 2 below. The ML estimates have been calculated using the ***fitdistr*** procedure in R studio. Furthermore, we have computed the expected frequencies of counts for Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh and two parameter Poisson Garima distributions with the help of R Studio statistical software [17] and Pearson's chi-square test is applied to check the goodness of fit of the models mentioned above. The calculated figures are given in Table 3.

The p-values of Pearson's Chi-square statistic are <0.0001, 0.46, 0.47, <0.0001 and 0.51 for Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh, and two parameter Poisson Garima distributions, respectively (see Table 3). This reveals that Poisson and discrete Rayleigh are not a good fit at all for the data in question, whereas two parameter Poisson Garima, Negative binomial and discrete Weibull are good fit distributions with two parameter Poisson Garima model being the best one. The null hypothesis that data comes from two parameter Poisson Garima distribution is strongly accepted.

Finally, we have compared Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh and two parameter Poisson Garima distributions using the Akaike Information Criterion (AIC), given by Akaike [16] and the Bayesian information criterion (BIC), given by Schwarz [17]. AIC and BIC are evaluated according to the formula −2 ∗ log likelihood + k ∗ npar, where npar represents the number of parameters in the fitted model, and k = 2 for the usual AIC, or k = log(n) (n being the number of observations) for the so-called BIC or SBC (Schwarz's Bayesian criterion).

The AIC and BIC values for the fitted distributions to dataset given in Table 1 are given in Table 4. By comparing the AIC and BIC values for the fitted models, we conclude that in case of the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967, (Falls et al. [17]), two parameter Poisson Garima distribution fits better in contrast to Poisson, NBD, discrete Weibull and, discrete Rayleigh models.

Table 2. *Estimated parameters by ML method for various fitted distributions for the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.*

Distribution	Parameter Estimates	Model function
Two parameters Poisson Garima	$\hat{\theta} = 1.89$ $\hat{\alpha} = 0.18$	$P(X = x) = \frac{\theta}{(1+\theta)^{x+2}(\alpha\theta + 2)}(\theta^2\alpha + \theta\alpha + \theta\ x + 2\theta + 1);$ $x = 1, 2, 3, ...;\ \theta > 0, \alpha > 0$
Poisson	$\hat{\lambda} = 0.75$	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, 2, ...; \lambda > 0$
Discrete Rayleigh	$\hat{q} = 0.68$	$p(x) = q^{x^2} - q^{(x+1)^2}$ $x = 0, 1, 2, ...;\quad 0 < q < 1$
NBD	$\hat{p} = 0.60,$ $\hat{r} = 1.17$	$p(x) = \begin{bmatrix} x + r - 1 \\ x \end{bmatrix} p^r q^x\ ;$ $x = 0, 1, 2, ...;\ 0 < p < 1\ ;\ r > 0$
Discrete Weibull	$\hat{q} = 0.44$, $\hat{\beta} = 1.04$	$p(x) = q^{x^\beta} - q^{(x+1)^\beta}$ $x = 0, 1, 2, ...;\ 0 < q < 1;\ \beta > 0$

Table 3: *Expected frequency of counts & goodness of fit for the thunderstorms events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.*

X	Observed frequency	Expected frequencies				
		Poisson	NB	DW	DR	TPPGD
0	187	155.64	184.66	184.45	103.02	184.33
1	77	116.97	84.54	84.71	153.12	84.43
2	40	43.95	35.87	35.95	62.49	36.32
3	17	11.01	14.82	14.85	10.54	15.01
4	6	2.07	6.04	6.04	0.80	6.04
5	2	0.31	2.44	2.42	0.03	2.38
6	1	0.04	0.98	0.96	0.00	0.92
χ^2 p-value		<0.0001	0.46	0.47	<0.0001	0.51

X: Number of thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.
Observed: Observed frequency of the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.

Table 4: *AIC and BIC values for fitted distributions for counts of thunderstorms events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967, Shanker et al. [14]*

Criterion	Poisson	NBD	DWD	DR	TPPGD
-loglik	412.253	394.5933	394.5637	470.2933	394.4194
AIC	826.506	793.1866	793.1275	942.5866	792.8389
BIC	830.305	800.7848	800.7256	946.3857	800.4371

Data set 2: Secondly, we analyze the data set regarding distribution of the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of July, January 1957 to December 1967, Shanker et al. [14].

Table 5: *Distribution of the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of July, January 1957 to December 1967.*

Number of Thunderstorms	0	1	2	3	4	5	Total
Frequency	177	80	47	26	9	2	341

In order to give an impression of the typical form of our observed distributions and their various parameter estimates, we have provided Table 6 below. The ML estimates have again been calculated using the ***fitdistr*** procedure in R studio.

Table 6. *Estimated parameters by ML method for various fitted distributions for thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.*

Distribution	Parameter Estimates	Model function
Two parameter Poisson Garima	$\hat{\theta}=1.7182$ $\hat{\alpha}=0.0001$	$P(X=x)=\frac{\theta}{(1+\theta)^{x+2}(\alpha\theta+2)}\left(\theta^2\alpha+\theta\alpha+\theta\ x+2\theta+1\right);$ $x=1,2,3,\ldots;\quad \theta>0,\alpha>0$
Poisson	$\hat{\lambda}=0.8739$	$p(x)=\frac{e^{-\lambda}\lambda^x}{x!}$ $x=0,1,2,\ldots;\lambda>0$
Discrete Rayleigh	$\hat{q}=0.724$	$p(x)=q^{x^2}-q^{(x+1)^2}$ $x=0,1,2,\ldots;\quad 0<q<1$
NBD	$\hat{p}=0.6263$ $\hat{r}=1.4652$	$p(x)=\begin{bmatrix}x+r-1\\ x\end{bmatrix}p^r q^x\ ;$ $x=0,1,2,\ldots;\ \ 0<p<1\ ;\ r>0$
Discrete Weibull	$\hat{q}=0.496$ $\hat{\beta}=1.112$	$p(x)=q^{x^\beta}-q^{(x+1)^\beta}$ $x=0,1,2,\ldots;\ \ 0<q<1;\ \beta>0$

Furthermore, we have computed the expected frequencies of counts for Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh and two parameter Poisson Garima distributions with the help of R studio statistical software [17] and Pearson's chi-square test is applied to check the goodness of fit of the models mentioned above. The calculated figures are given in Table 7. The p-values of Pearson's Chi-square statistic are<0.0001, 0.066, 0.067, <0.0001 and 0.117 for Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh and two parameter Poisson Garima distributions, respectively (see Table 7). This reveals that Poisson, discrete Rayleigh, are not a good fit at all for the data in question, whereas two parameter Poisson Garima, Negative binomial and discrete Weibull are good fit distributions with two parameter Poisson Garima model being the best one. The null hypothesis that data comes from two parameter Poisson Garima distribution is strongly accepted.

Table 7: *Expected frequency counts & goodness of fit for Number of thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.*

X	Observed Frequency	Expected Frequency				
		Poisson	NBD	DWD	DR	TPPGD
0	177	142.306	171.829	171.531	94.095	175.904
1	80	124.362	94.061	94.276	153.180	89.774
2	47	54.340	43.315	43.429	75.073	42.246
3	26	15.829	18.692	18.788	16.706	18.934
4	9	3.458	7.796	7.810	1.840	8.213
5	2	0.604	3.183	3.152	0.103	3.481
χ^2 p-value		<0.0001	0.066	0.067	<0.0001	0.117

X: Number of thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.
Observed: Observed frequency of thunderstorms events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967.

The AIC and BIC values for the fitted distributions to dataset given in Table 5 are given in Table 8. By comparing the AIC and BIC values for the fitted models, we conclude that in case of thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967, two parameter Poisson Garima distribution fits better in contrast to Poisson, NBD, discrete Weibull and discrete Rayleigh model.

Table 8: *AIC and BIC values for fitted distributions for the thunderstorm events at Cape Kennedy, Florida for the 11-year period of record for the month of June, January 1957 to December 1967*

Criterion	**Poisson**	**NBD**	**DWD**	**DR**	**TPPGD**
-loglik	455.51	440.178	440.0484	504.68	439.9534
AIC	913.02	884.356	884.0968	1011.35	883.9069
BIC	916.85	892.0198	891.7606	1015.18	891.5707

Data set 3: This data represents the daily new deaths of 111 days from 12 March to 30 June 2020 belonging to Greece country (see World Health Organization), Almetwally et.al [15].

Table 9: *Daily new deaths of 111 days from 12 March to 30 June 2020 in Greece.*

Number of deaths	0	1	2	3	4	5	6	7	8	9	Total
Frequency	39	26	17	9	6	7	6	0	0	1	111

In order to give an impression of the typical form of our observed distributions and their various parameter estimates, we have provided Table 10 below. The ML estimates have been calculated using the ***fitdistr*** procedure in R studio. Furthermore, we have computed the expected frequencies of counts for Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh and two parameter Poisson Garima distributions with the help of R studio statistical software [17] and Pearson's chi-square test is applied to check the goodness of fit of the models mentioned above. The calculated figures are given in Table 11.

Table 10 *Estimated parameters by ML method for various fitted distributions for the daily new deaths of 111 days from 12 March to 30 June 2020 belonging to Greece country (see World Health Organization).*

Distribution	Parameter Estimates	Model Function
Two Parameter Poisson Garima	$\hat{\theta}=0.86$ $\hat{\alpha}=0.08$	$P(X=x)=\frac{\theta}{(1+\theta)^{x+2}(\alpha\theta+2)}(\theta^2\alpha+\theta\alpha+\theta\ x+2\theta+1);$ $x=1,2,3,...;\ \ \theta>0,\alpha>0$
Poisson	$\hat{\lambda}=1.72$	$p(x)=\frac{e^{-\lambda}\lambda^x}{x!}$ $x=0,1,2,...;\lambda>0$
Discrete Rayleigh	$\hat{q}=0.89$	$p(x)=q^{x^2}-q^{(x+1)^2}$ $x=0,1,2,...;\quad 0<q<1$
NBD	$\hat{p}=0.415,$ $\hat{r}=1.220$	$p(x)=\begin{bmatrix}x+r-1\\x\end{bmatrix}p^r q^x\ ;$ $x=0,1,2,...;\ \ 0<p<1\ ;\ r>0$
Discrete Weibull	$\hat{q}=0.66$, $\hat{\beta}=1.08$	$p(x)=q^{x^\beta}-q^{(x+1)^\beta}$ $x=0,1,2,...;\ \ 0<q<1;\ \beta>0$

The p-values of Pearson's Chi-square statistic are <0.001, 0.471, 0.462, <0.001 and 0.483 for Poisson, Negative Binomial, discrete Weibull, discrete Rayleigh, and two parameter Poisson Garima distributions, respectively (see Table 11). This reveals that Poisson and discrete Rayleigh are not a good fit at all for the data in question, whereas two parameter Poisson Garima, Negative binomial and discrete Weibull are good fit distributions with two parameter Poisson Garima model being the best one. The null hypothesis that data comes from two parameter Poisson Garima distribution is strongly accepted.

Table 11: *Expected frequency of counts & goodness of fit for the daily new deaths of 111 days from 12 March to 30 June 2020 belonging to Greece country.*

X	Observed Frequency	Expected Frequency				
		Poisson	NBD	DWD	D-Rayleigh	TPPGD
0	39	19.862	37.946	37.665	12.210	38.051
1	26	34.177	27.089	27.172	29.146	26.609
2	17	29.405	17.594	17.686	30.754	17.606
3	9	16.866	11.050	11.147	21.688	11.236
4	6	7.255	6.821	6.892	11.175	6.991
5	7	2.497	4.167	4.204	4.354	4.268
6	6	0.716	2.528	2.537	1.305	2.568
7	0	0.176	1.526	1.518	0.304	1.527
8	0	0.038	0.917	0.902	0.055	0.900
9	1	0.007	0.550	0.532	0.008	0.526
χ^2 p-value		<0.001	0.471	0.462	<0.001	0.483

X: Number of daily new deaths of 111 days from 12 March to 30 June 2020 belong to Greece country.
Observed: Observed frequency of the daily new deaths of 111 days from 12 March to 30 June 2020

belonging to Greece country.

The AIC and BIC values for the fitted distributions to dataset given in Table 9 are given in Table 12. By comparing the AIC and BIC values for the fitted models, we conclude that in case of for the daily new deaths of 111 days from 12 March to 30 June 2020 belonging to Greece country, two parameter Poisson Garima distribution fits better in contrast to Poisson, NBD, discrete Weibull and, discrete Rayleigh models.

Table 12: *AIC and BIC values for fitted distributions the data represents the daily new deaths of 111 days from 12 March to 30 June 2020 belonging to Greece country*

Criterion	Poisson	NBD	DWD	D-Rayleigh	TPPGD
-loglik	220.10	198.3241	198.2529108	230.03	198.1523
AIC	442.21	400.6481	400.5058215	462.05	400.3046
BIC	444.92	406.0672	405.9248819	464.76	405.7237

6. Conclusion

In this paper, we have introduced a new count data model namely two parameter Poisson Garima distribution. General description of the statistical properties such as the moments, skewness, kurtosis, moment generating function, probability generating function, order statistics, etc has been provided after an extensive study. The estimation of the model parameters is performed by maximum likelihood method and method of moment. The applicability of this distribution to real life data is illustrated by three real life examples by positioning it alongside other four models (NB, Poisson, DW, DR) for comparison. The comparison was conducted by using AIC, BIC and p value. By competing the models, we find that TPPGD is superior as compared to alternative models and provides the best fit, because it has the smallest values of AIC and BIC.

References

[1] Altun, E., Alqifari, H., & Eliwa, M. S. (2023). A novel approach for zero-inflated count regression model: Zero-inflated Poisson generalized-Lindley linear model with applications. *AIMS Mathematics*, 8(10), 23272–23290. https://doi.org/10.3934/math.20231183

[2] Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4), 791–799.

[3] Dean, C., Lawless, J. F., & Willmot, G. E. (1989). A mixed Poisson–inverse-gaussian regression model. *Canadian Journal of Statistics*, 17(2), 171–181.

[4] Gerstenkorn, T. (2004). A compound of the generalized negative binomial distribution with the generalized beta distribution. *Open Mathematics*, 2(4), 527–537. https://doi.org/10.2478/bf02475961

[5] Greenwood, M., & Udny Yule, G. (1920). An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society*, 83.

[6] Holla, M. S. (1967). On a Poisson-inverse Gaussian distribution. *Metrika*, 11(1), 115–121.

[7] Sankaran, M. (1970). 275. note: The discrete Poisson-Lindley distribution. *Biometrics*, 145–149.

[8] Sen, S., Chandra, N., & Maiti, S. S. (2018). On Properties and Applications of a Two-Parameter Xgamma Distribution. *Journal of Statistical Theory and Applications*, 17(4), 674. https://doi.org/10.2991/jsta.2018.17.4.9

[9] Shanker, R. (2016a). Garima Distribution and its Application to Model Behavioral Science Data. *Biometrics & Biostatistics International Journal*, 4(7). https://doi.org/10.15406/bbij.2016.04.00116

[10] Shanker, R. (2016b). On Poisson-Sujatha Distribution and its Applications to Model Count Data from Biological Sciences. *Biometrics & Biostatistics International Journal*, 3(4), 1–7. https://doi.org/10.15406/bbij.2016.03.00069 [

[11] Shanker, R. (2016c). Sujatha distribution and its applications. *Statistics in Transition New Series*, 17(3), 391–410. https://doi.org/10.21307/stattrans-2016-029[

[12] Shanker, R. (2017). The Discrete Poisson-Garima Distribution. *Biometrics & Biostatistics International Journal*, 5(2). https://doi.org/10.15406/bbij.2017.05.00127

[13] Shanker, R., Shukla, K. K., & Leonida, T. A. (2020). A New Two-Parameter Poisson-Sujatha Distribution. *International Journal of Probability and Statistics,* 9(2), 21-32. https://doi.org/10.5923/j.ijps.20200902.01

[14] Shanker, R., Fesshaye, H., Shanker, R., Leonida, T. A., & Sium, S. (2017). On discrete Poisson-Shanker distribution and its applications. *Biometrics & biostatistics International journal*, *5*(1), 00121.

[15] Almetwally, E. M., & Dey, S. (2021). Modelling of COVID-19 Data Using Discrete Distribution. *Delta University Scientific Journal,* 4(1), 13-24.

[16] R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.