doi:10.17713/ajs.v50i4.1091



New Distribution for Fitting Discrete Data: The Poisson-Gold Distribution and Its Statistical Properties

Ahmad A. Hanandeh

Amjad D. Al-Nasser

Yarmouk University

Yarmouk University & Al Falah University

Abstract

Motivated mainly by lifetime issues, a new lifetime distribution coined "Discrete Poisson-Gold distribution" is introduced in this paper. Different structural properties of the new distribution are derived including moment generating function and the r^{th} moment and others are presented. In addition, we discussed various important mathematical properties of the new distribution including estimation procedures for estimating the distribution parameters using the maximum likelihood and method of moments. The usefulness and credibility of the distribution are illustrated by means of two real-data applications to show its superior performance over some other well-known lifetime distributions and to prove its versatility in practical applications.

Keywords: gold distribution, lifetime distribution, mixture distributions, parameter estimation, statistical properties.

1. Introduction

Studying and analyzing lifetime data play a crucial role in a wide variety of fields in applied sciences such as engineering, medicine, insurance, economics and marketing. The exponential distribution is widely used for modeling such dataset. Nonetheless, using exponential distribution for some practical applications does not provide a reasonable parametric fit when the underlying failure rates present non-constant (monotone) shapes. To overcome such problem, several recent probability distributions have been developed to model such data by compounding some useful life distributions (Gómez-Déniz, Sordo, and Calderín-Ojeda (2014); Shanker (2015); Al-Omari, Al-Nasser, and Ciavolino (2019); Al-Omari and Shraa (2019); Altun (2019)).

Compounding Poisson distribution with many other continuous distributions plays a prominent role in deriving new families of distributions. Recall that a random variable X is said to have a Poisson distribution (PD) with parameter $\theta > 0$ if its probability mass function is:

$$\frac{e^{-\lambda}\lambda^x}{x!}, \lambda > 0, x = 0, 1, \dots$$

This distribution, in particular, have been widely used in stochastic process as claim frequency distributions where it counts the number of events in a specific period of time.

The resulting distributions are more flexible for modeling lifetime data and, sometimes, provide reasonable parametric fits to practical applications as in lifetime and reliability studies. The flexibility of the new distributions comes from the fact that one or more failure rate shapes can be decreasing, increasing, bathtub shaped or unimodal. Indeed, it is possible to construct the mathematical features of a lifetime distribution for a specific life phenomenon based on its failure rate pattern.

Lindley (1958) introduced the Lindley distribution (LD) for analyzing lifetime data that belongs to an exponential family. It has been used as an alternative to the exponential lifetime distributions when the failure rate is unimodal (Bakouch, Al-Zahrani, Al-Shomrani, Marchi, and Louzada 2012). It has also been widely used in the field of medical sciences, biological sciences, even in engineering for the data having increasing hazard rate function. The probability density function (pdf) of the LD with scale parameter $\theta > 0$ is defined by

$$f_{LD}(x,\theta) = \frac{\theta^2}{\theta + 1}(1+x)e^{-\theta x}, x > 0, \theta > 0.$$

The corresponding cumulative distribution function (cdf) is given by

$$F_{LD}(x,\theta) = 1 - (1 + \frac{\theta x}{\theta + 1})e^{-\theta x}, x > 0, \theta > 0$$

The pdf of the LD is a two-component mixture of exponential (θ) and $gamma(2, \theta)$. Figure 1 shows the pdf and the cdf of LD for some parameter values.

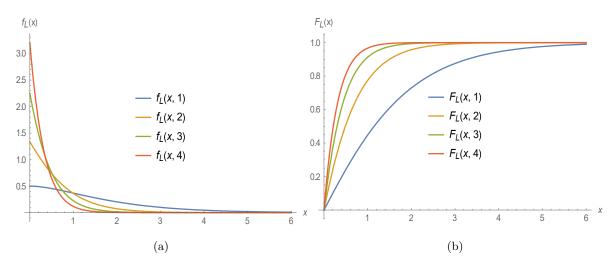


Figure 1: The pdf and cdf of LD for some parameter values

Statistical properties and parameter estimation of the LD were studied and applied to a waiting time data by Ghitany, Atieh, and Nadarajah (2008) and Ghitany, Alqallaf, Al-Mutairi, and Husain (2011). It was shown that LD is better than the exponential distribution in modeling lifetime data using some comparison criteria. Due to the failure rate property of LD, there are some situations where the distribution fails to provide a good fit in modeling real lifetime data. To address this situation, many researchers have proposed new classes of distributions based on generalizations and modifications of the LD. Recently, Al-Talib, Al-Nasser, and Ciavolino (2019) proposed a new continuous distribution named as Gold distribution (GD); which offers a flexible distribution for modeling lifetime data and generalization of many statistical distributions such as exponential, gamma and Lindley distributions.

On the other hand, Sankaran (1970) proposed the discrete Poisson-Lindley distribution by compounding the Poisson and Lindley distributions. Adamidis and Loukas (1998) introduced a two-parameter exponential-geometric distribution with decreasing failure rate by compounding the exponential and geometric distributions. Gupta and Kundu (1999) and Gupta and

Kundu (2001) studied and introduced the generalized exponential distribution as an extension of the exponential distribution. Kuş (2007), proposed an exponential-Poisson distribution by compounding an exponential and zero truncated Poisson distributions. Chahkandi and Ganjali (2009) introduced a new two-parameter distribution family with decreasing failure rate by compounding power-series distribution and exponential distribution.

In this paper, a new discrete distribution is introduced by compounding the Poisson and Gold distributions. The main motivations for introducing this lifetime model are to: (i) propose a new, more flexible lifetime distribution that can be used for modeling lifetime data in a wider class of reliability problems; (ii) extend both Poisson and Gold distributions; (iii) accommodate a broad class of failure rate functions.

The rest of the paper is organized as follows: in Section 2, we review the GD. Section 3 describes the new Poisson-Gold distribution (PGD) and study its basic properties, including defining its pmf and cdf. Section 4 introduced some related discrete distribution to the PGD. Section 5 investigates the factorial moments of PGD including deriving its mean and variance. Moment and probability generating functions are also explored in this Section. The hazard rate function is discussed in Section 6. Estimation of the distribution parameters by maximum likelihood and method of moments are investigated in Section 7. In Section 8, two real data applications illustrate the performance of the PGD over other competing reliability distributions and reports the results. The article ends with concluding remarks in Section 9.

2. The Gold distribution

The GD is a mixture of k independent gamma distributions with a scale parameter (θ) and different shape parameters (j); where j = 1, 2, ..., k.

Definition 1. A random variable X is said to have a GD if the pdf and cdf of X are given by

$$f_{GD}(x;\theta,k) = \frac{1}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}} \cdot \frac{x^{k} - 1}{x - 1} \frac{1}{\theta} e^{\frac{-x}{\theta}} , x > 0, \theta > 0 ; k = 1, 2, \dots \text{ and } x \neq 1, (1)$$

$$F_{GD}(x;\theta,k) = 1 - \frac{1}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}} \sum_{j=1}^{k} \theta^{j-1} \Gamma\left(j, \frac{x}{\theta}\right).$$

Figures 2 and 3 illustrate the pdf and cdf shapes with different parameter values, respectively.

The mean and variance are found to be:

$$E(X) = \frac{\sum_{j=1}^{k} \Gamma(j+1) \theta^{j}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}}.$$

and

$$Var(X) = \sum_{j=1}^{k} \frac{\Gamma(j) \Gamma(j+1) \theta^{2j}}{\left(\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}\right)^{2}}.$$

Al-Talib et al. (2019) obtained many interesting properties of the GD.

3. Poisson-Gold distribution

The GD is a mixture of several Gamma distributions, therefore, it could be considered as a conjugate distribution of Poisson distribution (PD), which is one of the well-utilized discrete distributions. This fact leads to an analytically tractable compound distribution where one think of the λ parameter in PD as being random variable drawn from GD.

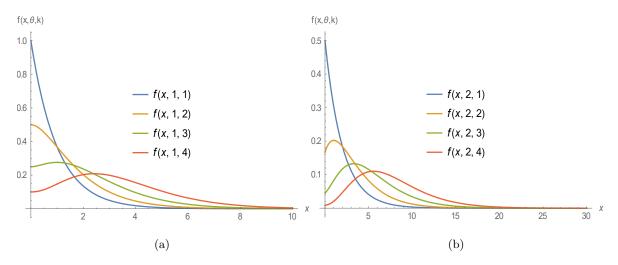


Figure 2: The pdf of GD for some parameter values

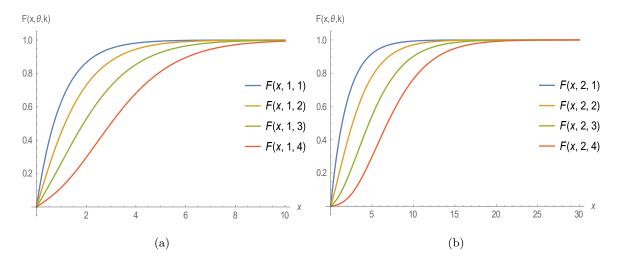


Figure 3: The cdf of GD for some parameter values

Definition 2. A random variable X follows a mixed PD with mixing distribution having probability density function g if its probability function is given by

$$P(X=x) = \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} g(\lambda) d\lambda; x = 0, 1, \dots$$
 (2)

This can be rewritten in terms of the probability generating function, $M_X(t)$ as

$$M_X(t) = \int_0^\infty e^{\lambda(t-1)} g(\lambda) d\lambda; x = 0, 1, \dots$$

Note that the right-hand side of this equation is the moment generating function of the mixing distribution evaluated at t-1. This, immediately, implies that the probability generating function of the mixed PD uniquely determines the mixing distribution through its moment generating function. In the sequel, the mixed PD with mixing distribution with density function g is denoted by the MP(g), while its probability function is denoted by P(x). Note that λ can be a continuous/discrete random variable or it may be a finite number of values.

Corollary 1. Suppose that the parameter λ of the PD follows GD (1). Then, the Poisson mixture of GD "PGD (k, θ) " is given by

$$PGD(k,\theta) = \frac{\sum_{j=1}^{k} \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^{x+j}}{\Gamma(x+1) \sum_{j=1}^{k} \Gamma(j) \theta^{j}}; \quad x = 0, 1, 2, \dots$$

Proof. Using the fact given in (2), then

$$\begin{split} P\left(X=x\right) &= \int_{0}^{\infty} P\left(X=x|\lambda\right) f\left(\lambda;\theta\right) d\lambda \\ &= \int_{0}^{\infty} \frac{e^{-\lambda} \ \lambda^{x}}{x!} \ \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \left(\sum_{j=1}^{k} \lambda^{j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} \ d\lambda \\ &= \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \frac{1}{x!} \int_{0}^{\infty} \left(\sum_{j=1}^{k} \lambda^{x+j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta} - \lambda} \ d\lambda \\ &= \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \frac{1}{x!} \sum_{j=1}^{k} \int_{0}^{\infty} \lambda^{x+j-1} \frac{1}{\theta} \ e^{\frac{-\lambda(1+\theta)}{\theta}} \ d\lambda \\ &= \frac{1}{1+\theta} \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \frac{1}{x!} \sum_{j=1}^{k} \int_{0}^{\infty} \lambda^{x+j-1} \frac{1+\theta}{\theta} \ e^{\frac{-\lambda(1+\theta)}{\theta}} \ d\lambda \end{split}$$

which is equivalent to

$$P\left(X=x\right) = \ \frac{1}{1+\theta} \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \frac{1}{x!} \sum_{j=1}^{k} E(\lambda^{x+j-1}) \ ; \ \text{given that} \ \lambda \sim Exp\left(\frac{\theta}{1+\theta}\right)$$

Therefore,

$$P(X = x) = \frac{1}{1+\theta} \frac{\sum_{j=1}^{k} \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^{x+j-1}}{x! \sum_{j=1}^{k} \Gamma(j) \theta^{j-1}}; \quad x = 0, 1, 2, \dots$$

Or simply the probability mass function (pmf) of PGD can be written as

$$P(X=x) = \frac{\sum_{j=1}^{k} \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^{x+j}}{\Gamma(x+1) \sum_{j=1}^{k} \Gamma(j) \theta^{j}}; \quad x = 0, 1, 2, \dots$$
(3)

which completes the proof of Corollary 1.

Figures 4 shows different bar diagrams expressing different shapes of the pmf of PGD for different values of k and θ : In particular, the null probability can be expressed as:

$$P(X = 0) = \frac{\sum_{j=1}^{k} \Gamma(j) \left(\frac{\theta}{1+\theta}\right)^{j}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j}}$$

Moreover, the cumulative distribution function (cdf) of PG distribution will be

$$P(X \le x) = \sum_{i=0}^{x} \frac{\sum_{j=1}^{k} \Gamma(i+j) \left(\frac{\theta}{1+\theta}\right)^{i+j}}{\Gamma(i+1) \sum_{j=1}^{k} \Gamma(j) \theta^{j}}$$

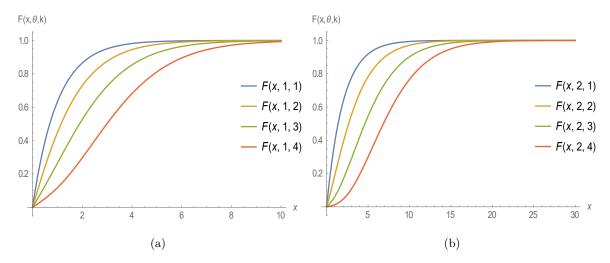


Figure 4: The pmf of the PGD for some parameter values

4. PGD related distributions

The PGD is a generalization of some well-known discrete distributions. When k = 1, the pmf in (3) becomes:

$$P(X = x) = \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta}\right)^x \; ; \; x = 0, 1, 2, \dots$$
 (4)

which is the geometric distribution with probability of fail $q = \frac{\theta}{1+\theta}$. Note that this is exactly the Poisson-Exponential distribution. In fact, based on the GD properties, the PGD can be considered as a family of discrete distributions. The value of the parameter k plays the main role in redefining a new discrete distribution. Table 1 represents several special cases of $PGD(\theta, k)$:

$PGD(\theta,k)$	Special cases of $PGD(\theta, k)$	Reference
$PGD(\theta, 1)$	Poisson-Exponential Distribution	Equation (4)
$PGD(\frac{1}{\theta},2)$	Poisson-Lindley Distribution	Sankaran (1970)
$PGD(\frac{1}{\theta},3)$	Poisson-Sujatha distribution	Shanker and Fesshaye (2016)
$PGD(\frac{1}{\theta},4)$	Poisson-Amarendra distribution	Shanker (2016)
$PGD(\frac{1}{\theta},5)$	Poisson-Devya distribution	New discrete distribution
$PGD(\frac{9}{\theta},6)$	Poisson-Shambhu distribution	New discrete distribution

Table 1: PGD related distributions

5. Moments and generating functions

Many of the most important measures and characteristics of a distribution can be derived and investigated through its moments, therefore, in this section we consider different moments of $PGD(k, \theta)$ distribution such as moment generating function, probability generation function and factorial moments.

5.1. Moment generating function

Corollary 2. Suppose that $X \sim PGD(k, \theta)$; then the moment generating function (mgf) is

given as:

$$M_X(t) = \frac{\theta}{\left(2 - e^t\right)^2 \sum_{j=1}^k \Gamma(j) \theta^{j-1}} \left(\sum_{j=1}^k \Gamma(j) \left(\frac{\left(2 - e^t\right)}{\theta} \right)^j \right)$$

Proof.

$$M_X(t) = E(e^{tx}) = \int_0^\infty \sum_{x=0}^\infty \left(e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} \right) g(\lambda) d\lambda$$

It is well-known that the mgf of PD is given by

$$\sum_{x=0}^{\infty} \left(e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} \right) = e^{\lambda (e^t - 1)}$$

As a result,

$$\begin{split} M_X(t) &= E\left(e^{tx}\right) \\ &= \int_0^\infty e^{\lambda \left(e^t - 1\right)} g\left(\lambda\right) d\lambda \\ &= \int_0^\infty e^{\lambda \left(e^t - 1\right)} \frac{1}{\sum_{j=1}^k \Gamma(j) \, \theta^{j-1}} \left(\sum_{j=1}^k \lambda^{j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} d\lambda \\ &= \frac{1}{\sum_{j=1}^k \Gamma(j) \, \theta^{j-1}} \left(\sum_{j=1}^k \int_0^\infty e^{\lambda \left(e^t - 1\right)} \, \frac{\lambda^{j-1}}{\theta} e^{\frac{-\lambda}{\theta}} d\lambda\right) \\ &= \frac{1}{\sum_{j=1}^k \Gamma(j) \, \theta^{j-1}} \left(\sum_{j=1}^k \int_0^\infty \, \frac{\lambda^{j-1}}{\theta} e^{\frac{-2\lambda + \lambda \cdot e^t}{\theta}} \, d\lambda\right) \\ &= \frac{1}{(2 - e^t) \, \sum_{j=1}^k \Gamma(j) \, \theta^{j-1}} \left(\sum_{j=1}^k \int_0^\infty \lambda^{j-1} \, \frac{\left(2 - e^t\right)}{\theta} e^{\frac{-\lambda(2 - e^t)}{\theta}} \, d\lambda\right) \\ &= \frac{1}{(2 - e^t) \, \sum_{j=1}^k \Gamma(j) \, \theta^{j-1}} \left(\sum_{j=1}^k \Gamma(j) \left(\frac{\left(2 - e^t\right)}{\theta}\right)^{j-1}\right) \end{split}$$

Thus,

$$M_{X}\left(t\right) = \frac{\theta}{\left(2 - e^{t}\right)^{2} \sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \left(\sum_{j=1}^{k} \Gamma\left(j\right) \left(\frac{\left(2 - e^{t}\right)}{\theta}\right)^{j}\right)$$

which completes the proof of Corollary 2.

5.2. Probability generating function

Corollary 3. Suppose that $X \sim PGD(k, \theta)$; then the probability generating function (PGF) of X is defined as:

$$G_X(\eta) = \sum_{x=0}^{\infty} \frac{\left(\frac{\eta\theta}{1+\theta}\right)^x}{\Gamma(x+1)} \frac{\sum_{j=1}^k \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^j}{\sum_{j=1}^k \Gamma(j) (\theta)^j}$$

Proof. The PGF of X is defined as

$$G_X(\eta) = E(\eta^X)$$

$$= \sum_{x=0}^{\infty} \eta^x P(X = x)$$

$$= \int_0^{\infty} \sum_{x=0}^{\infty} \left(\eta^x \frac{e^{-\lambda} \lambda^x}{x!} \right) \frac{1}{\sum_{j=1}^k \Gamma(j) \theta^{j-1}} \left(\sum_{j=1}^k \lambda^{j-1} \right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} d\lambda$$

By using Taylor expansion, we have

$$\sum_{x=0}^{\infty} \left(\eta^x \frac{e^{-\lambda} \lambda^x}{x!} \right) = e^{\lambda(\eta - 1)}$$

Therefore,

$$\begin{split} E\left(\eta^{X}\right) &= \int_{0}^{\infty} \frac{e^{\lambda(\eta-1)}}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \left(\sum_{j=1}^{k} \lambda^{j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} \, d\lambda \\ &= \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \int_{0}^{\infty} \left(\sum_{j=1}^{k} \lambda^{j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda + \lambda(\eta-1)}{\theta}} \, d\lambda \\ &= \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \sum_{j=1}^{k} \int_{0}^{\infty} \frac{\lambda^{j-1}}{\theta} e^{\frac{-2\lambda + \lambda\eta}{\theta}} \, d\lambda \\ &= \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \sum_{j=1}^{k} \int_{0}^{\infty} \frac{\lambda^{j-1}}{\theta} e^{\frac{-\lambda(2-\eta)}{\theta}} \, d\lambda \\ &= \frac{1}{2-\eta} \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \sum_{j=1}^{k} \int_{0}^{\infty} \lambda^{j-1} \, \frac{2-\eta}{\theta} e^{\frac{-\lambda(2-\eta)}{\theta}} \, d\lambda \end{split}$$

which is equivalent to

$$E\left(\eta^{X}\right) = \frac{1}{2-\eta} \frac{1}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} \sum_{j=1}^{k} E(\lambda^{j-1}) \; ; \; \text{given that} \; \; \lambda \sim Exp\left(\frac{\theta}{2-\eta}\right)$$

Thus,

$$E(\eta^{X}) = \frac{1}{2 - \eta} \frac{\sum_{j=1}^{k} \Gamma(j) \left(\frac{\theta}{2 - \eta}\right)^{j-1}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}}; \quad x = 0, 1, 2, \dots$$
$$= \frac{\sum_{j=1}^{k} \Gamma(j) \left(\frac{\theta}{2 - \eta}\right)^{j}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j}}; \quad x = 0, 1, 2, \dots$$

Hence,

$$G_X(\eta) = \sum_{x=0}^{\infty} \eta^x \frac{\sum_{j=1}^k \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^{x+j}}{\Gamma(x+1) \sum_{j=1}^k \Gamma(j) \theta^j}$$
$$= \sum_{x=0}^{\infty} \frac{\left(\frac{\eta\theta}{1+\theta}\right)^x}{\Gamma(x+1)} \frac{\sum_{j=1}^k \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^j}{\sum_{j=1}^k \Gamma(j) (\theta)^j}$$

which completes the proof of Corollary 3.

5.3. Factorial moments of PGD

Corollary 4. The r^{th} factorial moment about the origin of a random variable X, where $X \sim PGD(k, \theta)$, and $\theta, k > 0$; is given by the following formula

$$\mu_{(r)} = \frac{\sum_{j=1}^{k} \Gamma(j+r) (\theta)^{j+r-1}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}}; r = 1, 2, \dots$$

Proof. Let $X^{(r)} = X (X-1)(X-2)...(X-r+1) = \frac{X!}{(X-r)!}$ be the r^{th} factorial of X, then the factorial mean can be obtain as

$$\begin{split} \mu_{(r)} &= E\left(E[X^{(r)} \mid \lambda]\right) \\ &= \int_0^\infty \sum_{x=0}^\infty \left(x^{(r)} \frac{e^{-\lambda} \lambda^x}{x!}\right) \, \frac{1}{\sum_{j=1}^k \Gamma\left(j\right) \theta^{j-1}} \, \left(\sum_{j=1}^k \lambda^{j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} \, d\lambda \\ &= \int_0^\infty \lambda^r \sum_{x=r}^\infty \left(\frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!}\right) \, \frac{1}{\sum_{j=1}^k \Gamma\left(j\right) \theta^{j-1}} \, \left(\sum_{j=1}^k \lambda^{j-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} \, d\lambda \end{split}$$

However, it can be shown with simple transformation that $\sum_{x=r}^{\infty} \left(\frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} \right) = 1$. It follows that

$$\begin{split} \mu_{(r)} &= E\left(E[X^{(r)} \mid \lambda]\right) \\ &= \int_0^\infty \frac{1}{\sum_{j=1}^k \Gamma\left(j\right) \theta^{j-1}} \left(\sum_{j=1}^k \lambda^{j+r-1}\right) \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} \ d\lambda \\ &= \frac{1}{\sum_{j=1}^k \Gamma\left(j\right) \theta^{j-1}} \sum_{j=1}^k \int_0^\infty \lambda^{j+r-1} \frac{1}{\theta} e^{\frac{-\lambda}{\theta}} \ d\lambda \\ &= \frac{1}{\sum_{j=1}^k \Gamma\left(j\right) \theta^{j-1}} \sum_{j=1}^k E(\lambda^{j+r-1}) \ ; given \ that \ \lambda \sim Exp\left(\theta\right) \end{split}$$

Thus,

$$\mu_{(r)} = \frac{\sum_{j=1}^{k} \Gamma(j+r) (\theta)^{j+r-1}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}}; r = 1, 2, \dots$$

which completes the proof of Corollary 4.

Accordingly, the mean of the PGD is given directly as

$$\mu_{(1)} = \frac{\sum_{j=1}^{k} \Gamma(j+1) \theta^{j}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}}$$

However, the variance of the PGD can be found based on the second factorial moment:

$$\mu_{(2)} = \frac{\sum_{j=1}^{k} \Gamma(j+2) (\theta)^{j+2-1}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}} = E(X(X-1)) = E(X^{2}) - E(X)$$

Therefore,

$$Variance = \frac{\sum_{j=1}^{k} \Gamma\left(j+2\right) \left(\theta\right)^{j+2-1} + \sum_{j=1}^{k} \Gamma\left(j+1\right) \theta^{j}}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}} - \left(\frac{\sum_{j=1}^{k} \Gamma\left(j+1\right) \theta^{j}}{\sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}}\right)^{2}$$

6. Hazard rate function

The hazard rate is used to monitor the lifetime of a unit since it is more informative about the underlying mechanism of failure than the lifetime distribution. The hazard rate function of the PGD is given by

$$\frac{P(X = x + 1)}{P(X = x)} = \frac{\frac{\sum_{j=1}^{k} \Gamma(x+1+j) \left(\frac{\theta}{1+\theta}\right)^{x+1+j}}{\Gamma(x+2) \sum_{j=1}^{k} \Gamma(j)\theta^{j}}}{\frac{\sum_{j=1}^{k} \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^{x+j}}{\Gamma(x+1) \sum_{j=1}^{k} \Gamma(j)\theta^{j}}};$$

$$= \frac{1}{x+1} \frac{\sum_{j=1}^{k} \Gamma(x+1+j) \left(\frac{\theta}{1+\theta}\right)^{x+1+j}}{\sum_{j=1}^{k} \Gamma(x+j) \left(\frac{\theta}{1+\theta}\right)^{x+j}};$$

Thus,

$$P\left(X=x+1\right) = \frac{1}{x+1} \frac{\sum_{j=1}^{k} \Gamma\left(x+1+j\right) \left(\frac{\theta}{1+\theta}\right)^{x+1+j}}{\sum_{j=1}^{k} \Gamma\left(x+j\right) \left(\frac{\theta}{1+\theta}\right)^{x+j}} \ P\left(X=x\right);$$

The above expression is a decreasing function in x implying unimodality. Furthermore,

$$\frac{P\left(X=x+2\right)\ P\left(X=x\right)}{P\left(X=x+1\right)^{2}} = \frac{\left(\sum_{j=1}^{k} \Gamma\left(x+j\right) \left(\frac{\theta}{1+\theta}\right)^{x+j}\right) \left(\sum_{j=1}^{k} \Gamma\left(x+2+j\right) \left(\frac{\theta}{1+\theta}\right)^{x+2+j}\right)}{\left(x+2\right) \Gamma\left(x+1\right) \left(\sum_{j=1}^{k} \Gamma\left(x+1+j\right) \left(\frac{\theta}{1+\theta}\right)^{x+1+j}\right)^{2}} < 1$$

Now, since the density function P(X = x) is log-concave, the distribution in (3) has an increasing failure rate, see Johnson, Kemp, and Kotz (2005) (p. 209). For instance, for k = 1 we have

$$\frac{P(X = x + 1)}{P(X = x)} = \frac{\theta}{1 + \theta};$$

Therefore,

$$P(X = x + 1) = \frac{\theta}{1 + \theta} P(X = x);$$

7. Parameter estimation

In this section, we investigated the estimation of the parameters of the PGD by the maximum likelihood and the method of moment.

7.1. Maximum likelihood estimation

A maximum likelihood method is one of the most widely used technique of estimation in statistics. This method finds the parameter value that maximizes the likelihood function. More precisely, if the data are independent and identically distributed, then we have

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Consequently,

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{1+\theta} \frac{\sum_{j=1}^{k} \Gamma\left(x_i+j\right) \left(\frac{\theta}{1+\theta}\right)^{x_i+j-1}}{x_i! \sum_{j=1}^{k} \Gamma\left(j\right) \theta^{j-1}};$$

Equivalently, to find the value of θ that maximizes the probability function of the PGD, we can use the logarithm of the likelihood function

$$\log L(\theta) = \ln \left(\prod_{i=1}^{n} f(x_i; \theta) \right)$$

$$= \sum_{i=1}^{n} \ln(f(x_i; \theta))$$

$$= -n \ln(1+\theta) - \sum_{i=1}^{n} \ln(x_i!) - n \ln \left(\sum_{j=1}^{k} \Gamma(j) \theta^{j-1} \right) + \sum_{i=1}^{n} \ln \left(\sum_{j=1}^{k} \Gamma(x_i+j) \left(\frac{\theta}{1+\theta} \right)^{x_i+j-1} \right)$$

which can be solved by finding the first derivative with respect to θ .

As a special case, when k=1

$$log L(\theta) = -n \ln (1 + \theta) - \sum_{i=1}^{n} x_i (\ln (\theta) - \ln (1 + \theta))$$

It follows that,

$$\frac{dlog~L\left(\theta\right)}{d\theta} = \left(\frac{-n}{1+\theta}\right) - \sum_{i=1}^{n} x_i~\left(\frac{1}{\theta} - \frac{1}{1+\theta}\right) ~= 0$$

by solving the above equation, we find that

$$\hat{\theta} = \bar{X}$$

Another special case, when k=2

$$\begin{split} \frac{dlog\ L\left(\theta\right)}{d\theta} &= \frac{-n}{1+\theta} - n \frac{\Gamma\left(2\right)}{\Gamma\left(1\right) + \Gamma\left(2\right)\theta} \\ &+ \left(\frac{1}{1+\theta}\right)^2 \sum_{i=1}^n \frac{x_i \Gamma\left(x_i+1\right) \left(\frac{\theta}{1+\theta}\right)^{x_i-1} + \left(x_i+1\right) \Gamma\left(x_i+2\right) \left(\frac{\theta}{1+\theta}\right)^{x_i}}{\Gamma\left(x_i+1\right) \left(\frac{\theta}{1+\theta}\right)^{x_i} + \Gamma\left(x_i+2\right) \left(\frac{\theta}{1+\theta}\right)^{x_i+1}} = 0 \\ &= -2n + \frac{1}{\theta} \sum_{i=1}^n \frac{x_i + \left(x_i+1\right)^2 \left(\frac{\theta}{1+\theta}\right)}{1 + \left(x_i+1\right) \left(\frac{\theta}{1+\theta}\right)} = 0 \end{split}$$

Thus,

$$\frac{d \log L(\theta)}{d \theta} = -2n + \frac{1}{\theta} \sum_{i=1}^{n} \frac{\theta(x_i + (x_i + 1)^2) + x_i}{\theta(x_i + 2) + 1} = 0$$

Solving such non-linear equation can be done using any numerical iteration methods such as Newton Raphson method.

In general, the maximum likelihood estimate of θ can be found numerically by solving the following non-linear equation:

$$\frac{d\log L(\theta)}{d\theta} = \frac{-n}{1+\theta} - n \frac{\sum_{j=1}^{k} (j-1) \Gamma(j) \theta^{j-2}}{\sum_{j=1}^{k} \Gamma(j) \theta^{j-1}} + \left(\frac{1}{1+\theta}\right)^{2} \sum_{i=1}^{n} \frac{\sum_{j=1}^{k} (x_{i}+j-2) \Gamma(x_{i}+j) \left(\frac{\theta}{1+\theta}\right)^{x_{i}+j-2}}{\sum_{j=1}^{k} \Gamma(x_{i}+j) \left(\frac{\theta}{1+\theta}\right)^{x_{i}+j-1}} = 0$$
(5)

7.2. Method of moments (MoM)

The MoM is one of the most popular techniques and probably the oldest technique in statistics for constructing an estimator. Let $\mu_j(\theta) = E_{\theta}X^j$ be the theoretical moment and $\hat{\mu}_j(\theta) =$

 $\int x^j d\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n X_i^j$ be the empirical moment. From the law of large numbers, it is well-known that the empirical moments are close to theoretical ones. MoM estimators of $\theta_1, \ldots, \theta_k$, are $\hat{\theta}_1, \ldots, \hat{\theta}_k$, which defined as solutions to the following system of equations

$$\mu_j(\theta) = \hat{\mu}_j, j = 1, \dots, k,$$

Table 2 shows some parameter estimation results for different values of k.

Table 2: MoM estimators for several values of k

Value of k	$\hat{\theta}$ using MoM
1	\bar{X}
2	$\frac{-(1+\bar{X})\pm\sqrt{\bar{X}^2-6\bar{X}+1}}{}$
3	$\overset{4}{A}$
4	B

Where

$$A = \frac{\sqrt[3]{-4\bar{X}^3 - 15\bar{X}^2 + 9\sqrt{3}\left(\sqrt{-9\bar{X}^4 - 32\bar{X}^3 + 339\bar{X}^2 - 60\bar{X} + 5}\right) + 285\bar{X} - 23}}{9\times 2^{(2/3)}} - \frac{(-4\bar{X}^2 - 10\bar{X} + 14)}{182^{(1/3)}\sqrt[3]{-4\bar{X}^3 - 15\bar{X}^2 + 9\sqrt{3}\sqrt{-9\bar{X}^4 - 32\bar{X}^3 + 339\bar{X}^2 - 60\bar{X} + 5} + 285\bar{X} - 23}}{+(1 - \bar{X})/9}$$

and B is the solution of the following equation

$$\bar{X} \frac{1}{\hat{\theta}^4} + (\bar{X} - 1) \frac{1}{\hat{\theta}^3} + 2(\bar{X} - 1) \frac{1}{\hat{\theta}^2} + 6(\bar{X} - 1) \frac{1}{\hat{\theta}} - 24 = 0$$

where \bar{X} is the sample mean.

8. Real data application

Despite their small size, thunderstorms are among the most common and dangerous of all weather phenomenon. One of the main concerns of the meteorologists is to forecast afternoon convective thunderstorm activity at the Cape Kennedy Florida. This is because of the importance of thunderstorms in designing launch vehicles and planning of space missions.

PD is a statistical model suitable to be used for the case of independent events and mean equals variance, which is impractical condition in most of data sets such as thunderstorms. One may think to use the negative binomial distribution (NBD) as an alternative of the PD. Indeed, fitting NBD to the count data require the mean to be less than the variance which is not valid in the thunderstorms data.

Therefore, to clarify the application of the PGD, we consider the fit of the distribution to the count data on thunderstorm events at Cape Kennedy, Florida in two cases; for the month of June and for summer seasons (Jun, July and August) within eleven years period; from 1957 - 1967; in Tables 3 and 4 (Falls, Williford, and Carter 1971). The ML estimates have been evaluated by numerical optimization of the log-likelihood function instead of solving (5). The Poisson-Lindley distribution (PLD) and Poisson-Shanker distribution (PSD) introduced by (Shanker, Hagos, and Sujatha (2015), Shanker and Fesshaye (2017)) have also been fitted for comparison.

It is customary to use the chi-square test only when none of the expected frequencies (E_i) is less than 5; sometimes this requires that we combine some of the cells. As a result, we have combined some cells. The calculated value of chi-square is shown in the below Table 3. Note

that the number of degrees of freedom is given by df = c - p - 1, where c is the number of terms in the chi-square statistic and p is the number of estimated parameters in the fitted distribution.

Table 3: Goodness of fit test for the number of days that experienced X thunderstorms events at Cape Kennedy, Florida for the month of June within 11-year period (June 1957 through June 1967), Falls et al. (1971)

		Expected Frequency						
Number of	Observed				PGD	PGD	PGD	PGD
Thunderstorms	Frequency	PD	PLD	PSD	(k=1)	(k=2)	(k=3)	(k=4)
0	187	155.6	185.3	186.4	188.4	185.3	184.8	185.4
1	77	117.0	83.5	82.3	80.8	83.5	83.6	82.7
2	40	43.6	35.9	35.3	34.7	35.9	36.3	36.3
3	17	11.0	15.0	15.0	14.9	15.0	15.2	15.4
4	6	2.1	6.1	6.3	6.4	6.1	6.1	6.3
5	2	0.3	2.5	2.6	2.7	2.5	2.4	2.4
6	1	0.1	1.7	1.9	2.1	1.7	1.7	1.4
Total	330	330	330	330	330	330	330	330
ML Estimate $\hat{\theta}$		0.7515	0.5542	0.5956	0.7515	0.5542	0.4485	0.3809
Chi-Square		31.93	1.43	1.48	1.73	1.37	1.25	1.09
d.f.		2	3	3	3	3	3	3
p-value		0.0000	0.6985	0.6869	0.6303	0.7126	0.7410	0.7795

The expected Poisson frequencies along with the test statistics of this data set are provided in Table 3 and indicate a poor fit to this model ($\chi^2 = 31.93, df = 2, p - value = 1.16 \times 10^{-7}$). The other models have a much better fit than that given by the Poisson model and the best one is PGD with k = 4 ($\chi^2 = 1.09, df = 3, p - value = 0.7795$) at the price of 3 df. The empirical and fitted distribution using the different methods are compared in Figure 5.

Table 4: Goodness of fit test for the number of days that experienced X thunderstorms events at Cape Kennedy, Florida for the 11-year period of record for the summer season (June, July, August), (Summer 1957 through **Summer** 1967); Falls et al. (1971)

		Expected Frequency						
Number of Thunderstorms	Observed Frequency	PD	PLD	PSD	$ PGD \\ (k=1) $	$ PGD \\ (k=2) $	$ PGD \\ (k=3) $	$ \begin{array}{c} \text{PGD} \\ (k=4) \end{array} $
0	549	547.5	547.5	550.8	558.4	547.5	545.1	546.5
1	246	364.8	259.0	255.7	250.3	259	259.9	257.2
2	117	148.2	116.9	115.5	112.2	116.9	118.5	118.6
3	67	40.1	51.2	51.1	50.3	51.2	51.8	52.5
4	25	8.1	21.9	22.3	22.5	21.9	21.9	22.3
5	7	1.3	9.2	9.6	10.1	9.2	9.0	9.1
6	1	0.3	6.3	7.0	8.2	6.3	5.8	5.8
Total	1012	1012	1012	1012	1012	1012	1012	1012
ML Estimate $\hat{\theta}$		0.8123	0.5926	0.6319	0.8123	0.5926	0.4746	0.3995
Chi-Square		141.42	9.60	10.09	12.06	9.60	8.81	8.05
d.f.		3	4	4	4	4	4	4
p-value		0.0000	0.0477	0.0389	0.0169	0.0477	0.0660	0.0898

The expected frequencies of PD, PLD and PSD along with the test statistics of this data set are provided in Table 4 and indicate a poor fit to this model. The PGD with k = 3 and k = 4 have a slightly better fit than the other models. The empirical and fitted distribution using the different methods are compared in Figure 6.

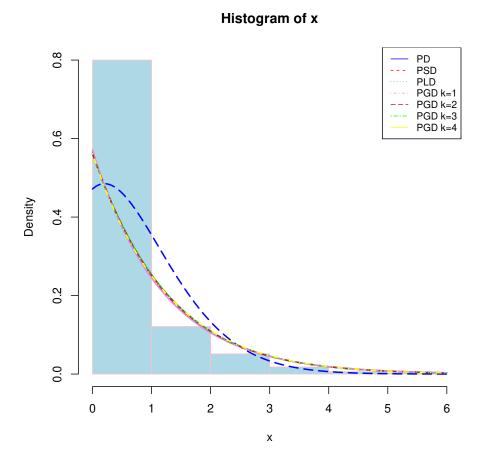


Figure 5: The empirical and fitted distribution of thunderstorms for the month of June

9. Concluding remarks

In this paper, a new discrete distribution has been introduced by compounding the Poisson and Gold distributions (PGD). The properties of the PGD were derived and discussed, the properties include the factorial moments, moment generating function, probability generating functions and hazard rate. Moreover, the unknown parameters are also estimated by using the maximum likelihood estimation method and the method of moments. In addition, the goodness of fit of PGD has been discussed with two applications from thunderstorms and the fit has been compared with PD, Poisson-Lindley distribution (PLD) and Poisson-Shanker distribution (PSD). The data analysis shows that the PGD gives satisfactory fit in the two datasets and slightly better than the other competitive distributions. It appears that the PD is not an appropriate choice to fit thunderstorm data. In fact, the PGD is recommended for fitting such data and may help the meteorologists for proper designing launch vehicles, planning of space missions and in launch operations at different stations in the case of having thunderstorms activities.

Acknowledgment

The authors are very thankful to the Editor and the anonymous referee for their useful comments, which led to a considerable improvement of the previous draft of this paper.

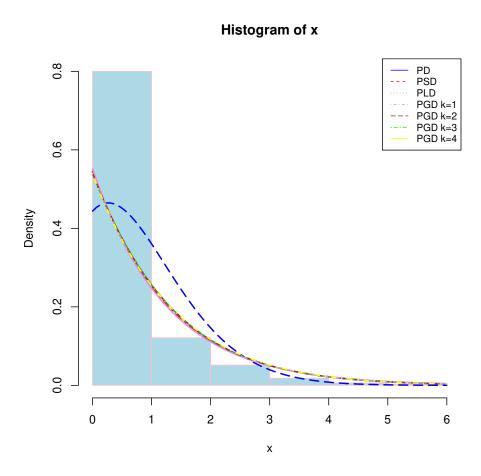


Figure 6: The empirical and fitted distribution of thunderstorms for the summer season

References

- Adamidis K, Loukas S (1998). "A Lifetime Distribution with Decreasing Failure Rate." Statistics & Probability Letters, **39**(1), 35–42.
- Al-Omari AI, Al-Nasser AD, Ciavolino E (2019). "A Size-Biased Ishita Distribution and Application to Real Data." Quality & Quantity, 53(1), 493–512.
- Al-Omari AI, Shraa D (2019). "Darna Distribution: Properties and Application." *Electronic Journal of Applied Statistical Analysis*, **12**(2), 520–541.
- Al-Talib MM, Al-Nasser AD, Ciavolino E (2019). "Gold Distribution: Another Look on the Generalization of Lindley Distribution." Submitted to: Sankhya (under review).
- Altun E (2019). "A New Generalization of Geometric Distribution with Properties and Applications." Communications in Statistics-Simulation and Computation, pp. 1–15.
- Bakouch HS, Al-Zahrani BM, Al-Shomrani AA, Marchi VAA, Louzada F (2012). "An Extended Lindley Distribution." *Journal of the Korean Statistical Society*, **41**(1), 75–85.
- Chahkandi M, Ganjali M (2009). "On Some Lifetime Distributions with Decreasing Failure Rate." Computational Statistics & Data Analysis, 53(12), 4433–4440.
- Falls LW, Williford WO, Carter MC (1971). "Probability Distributions for Thunderstorm Activity at Cape Kennedy, Florida." Journal of Applied meteorology, 10(1), 97–104.
- Ghitany ME, Alqallaf F, Al-Mutairi DK, Husain HA (2011). "A Two-Parameter Weighted Lindley Distribution and its applications to Survival Data." *Mathematics and Computers in simulation*, **81**(6), 1190–1201.
- Ghitany ME, Atieh B, Nadarajah S (2008). "Lindley Distribution and Its Application." *Mathematics and Computers in Simulation*, **78**(4), 493–506.
- Gómez-Déniz E, Sordo MA, Calderín-Ojeda E (2014). "The Log-Lindley Distribution as an Alternative to the Beta Regression Model with Applications in Insurance." *Insurance: Mathematics and Economics*, **54**, 49–57.
- Gupta RD, Kundu D (1999). "Theory & Methods: Generalized Exponential Distributions." Australian & New Zealand Journal of Statistics, 41(2), 173–188.
- Gupta RD, Kundu D (2001). "Generalized Exponential Distribution: Different Method of Estimations." *Journal of Statistical Computation and Simulation*, **69**(4), 315–337.
- Johnson NL, Kemp AW, Kotz S (2005). *Univariate Discrete Distributions*, volume 444. John Wiley & Sons.
- Kuş C (2007). "A New Lifetime Distribution." Computational Statistics & Data Analysis, 51(9), 4497–4509.
- Lindley DV (1958). "Fiducial Distributions and Bayes' Theorem." Journal of the Royal Statistical Society. Series B (Methodological), pp. 102–107.
- Sankaran M (1970). "The Discrete Poisson-Lindley Distribution." Biometrics, pp. 145–149.
- Shanker R (2015). "The Discrete Poisson-Shanker Distribution." *Jacobs Journal of Biostatistics*.
- Shanker R (2016). "The Discrete Poisson-Amarendra Distribution." International Journal of Statistical Distributions and Applications, 2(2), 14–21.

- Shanker R, Fesshaye H (2016). "On Poisson-Sujatha Distribution and Its Applications to Model Count Data from Biological Sciences." *Biometrics & Biostatistics International Journal*, 4(3), 1–7.
- Shanker R, Fesshaye H (2017). "On Discrete Poisson-Shanker Distribution and its Applications." Biometrics & Biostatistics International Journal, 5(1), 00121.

Shanker R, Hagos F, Sujatha S (2015). "On Modeling of Lifetimes Data using Exponential and Lindley Distributions." *Biometrics & Biostatistics International Journal*, **2**(5), 1–9.

Affiliation:

Ahmad A. Hanandeh Department of Statistics Science Faculty Yarmouk University Irbid 21163, Jordan

E-mail: ahmad.hanandeh@yu.edu.jo

URL:http://faculty.yu.edu.jo/ahmad.hanandeh/SitePages/Home.aspx

Amjad D. Al-Nasser Department of Statistics Science Faculty Yarmouk University Irbid 21163, Jordan

URL:https://faculty.yu.edu.jo/alnasser/SitePages/Home.aspx

&

College of Business Administration

Al Falah University

Dubai, UAE

E-mail: amjadyu@yahoo.com

Volume 50 July 2021 http://www.ajs.or.at/ http://www.osg.or.at/

> Submitted: 2019-12-26 Accepted: 2020-07-09