Pakistan Journal of Statistics and Operation Research

GOLD DISTRIBUTION
Another Look on the Generalization of Lindely Distribution

Mohammad Al-Talib[1*], Amjad Al-Nasser[2], Enrico Ciavolino[3]

* Corresponding Author

1. Department of Statistics, Yarmouk University, 21163 Irbid, Jordan, m.altalib@yu.edu.jo
2. Department of Statistics, Yarmouk University, 21163 Irbid, Jordan, amjadn@yu.edu.jo
3. Dipartimento di Filosofia e Scienze Sociali, Università del Salento, Lecce, Italy, enrico.ciavolino@unisalento.it

Abstract

In this paper, a new generalization of one parameter Lindely distribution is proposed. The new distribution is a mixture distribution of Gamma distributions with fixed scale parameter and variable shape parameter. The distribution is called 'GOLD Distribution' as it is a generalization for several distributions such as exponential, Lindely, Sujatha, Amarendra, Devya and Shambhu distributions. The probability density and cumulative density functions are derived. Also, the statistical properties of the GOLD distribution are discussed. Parameter estimation using the maximum likelihood and the method of moments are given. Moreover, an illustration of the usefulness of the GOLD distribution in survival data analysis is discussed based on a real lifetime data.

Key Words: Gamma Distribution, Mixture Distributions, Lindely Distribution, Survival Analysis, Statistical Measures.

Mathematical Subject Classification: 46T30, 62N02

1. Introduction

Lindely distribution (LD) is one of the most important distribution that is widely used in reliability and survivals data analyses. LD is derived from a two-parameter gamma distribution which is the most popular distribution that is used in analyzing lifetime data. A random variable X is said to have a two-parameter gamma distribution with parameters α (shape parameter) and θ (scale parameter); and denoted by Gamma(α,θ), if its probability density function (pdf) is given by:

$$f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\theta^{\alpha}}\, e^{-x/\theta}, \qquad x > 0;\ \alpha, \theta > 0.$$

It worth to say that, there is no close form for the gamma cumulative distribution function (*cdf*), and can be written based on the incomplete gamma function as:

$$F(x) = \frac{1}{\Gamma(\alpha)}\, \Gamma\left(\alpha, \frac{x}{\theta}\right),$$

where, $\Gamma(\alpha, x) = \int_x^{\infty} t^{\alpha-1}\, e^{-t}\, dt$. However, when α has only integer values, then the distribution is known as Erlang distribution and the cdf can be obtained as:

$$F(x) = 1 - \sum_{j=1}^{\alpha-1} \frac{x^j}{\Gamma(j)\theta^j}\, e^{-\frac{x}{\theta}},$$

Based on this fact, Lindely (1958) proposed a compound distribution of two independent gamma distributions of the same scale parameters but with different shape parameters with *pdf*:

$$f_{LD}(x) = \sum_{i=1}^{2} p_i \, g_i(x),$$

where,

$$p_1 = \frac{\theta}{1+\theta};\ p_2 = 1 - p_1;\ g_1(x) is\ Gamma\left(1, \frac{1}{\theta}\right) and\ g_2(x) is\ Gamma\left(2, \frac{1}{\theta}\right).$$

Hence, the closed formula of LD *pdf* is:

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

Ghitany et al. (2008) studied the statistical properties of Lindely distribution and considered it as a competitor and a replacement of the exponential distribution as it has more flexible mathematical properties. Since then, vast researches dealt with LD and it's properties (for example, see; Sankaran (1970), Zamani and Ismail (2010), Zakerzadeh and Dolati (2009), Shanker *et al*. (2013). Shanker and Mishra (2013a, 2013b)). Modification, extensions and generalizations of Lindely distribution were studied by several researchers. In fact, the studies on LD can be classified into three categories

- New formulation of LD based on other distribution families Ghitany *et al*., (2008); Ghitany *et al*., (2013); Jodra (2010); Gomez and Ojeda (2011); Kadilar and Cakmakyapan (2017); Hassan (2014); Sharma *et al*., (2015); Akbar *et al*. (2016); Deniz and Ojeda (2011); Shanker and Mishra (2013b); Singh *et al*. (2014); Merovci (2013).
- Modified LD by adding more parameters (Abd El-Monsef (2015); Shanker *et al*. (2013); Ghitany *et al*. (2011); Shanker and Mishra (2013a); Pararai *et al*. (2015); Alkarni (2015)).
- Generalized LD by extending the pdf to a more general form (Abouammoh (2015); Nadarajah *et al*. (2011); Zakerzadeh and Dolati (2009); Maya and Irshad (2017); Shibu and Irshad (2016); Bakouch *et al*. (2012), Elbatal *et al*. (2013); Oluyede and Yang (2014)).

The aim of this paper is to introduce a new generalization of the Lindley distribution which offers a more flexible distribution for modeling lifetime data, which could be also considered as a family distribution in which several distributions are a special case of the new generalization.

The remainder of this paper is organized as follows. Review of the previous generalized Lindely distribution is given in Section 2. The proposed generalization of LD (GOLD) with its properties are explored in Section 3. Various statistical measures of the GOLD are given in Section 4. Estimation of the distribution parameters by method of moment and inference of a random sample from GOLD are investigated in Section 5. In Section 6 a real data application illustrates the performance of the GOLD distribution over other competing reliability distributions. Section 7, with some concluding remarks, ends the paper.

2. Some Generalized LD

The first discussion of the statistical properties of LD was introduced by Ghitany *et al*. (2008). Since then, many generalized LD have been introduced. Ekhosuehi *et al*. (2018), proposed a new generalized two-parameter LD in order to offer more flexibility in modeling lifetime data, the *pdf*

$$g_1(x) = \frac{\theta}{\theta+1}\left(1 + \frac{\theta^{\alpha-2} x^{\alpha-1}}{\Gamma(\alpha)}\right)\theta e^{-\theta x}, x > 0;\ \alpha, \theta > 0,$$

which could be considered as a new modification of Shanker *et al*., (2013) who proposed another two-parameter LD with *pdf*

$$g_2(x) = \frac{\theta}{1+\theta}(1+\alpha x)\ \theta\ e^{-\theta x}; x > 0, \theta > 0, \alpha > -\theta.$$

Zakerzadeh and Dolati (2009) proposed a three-parameter LD having the following *pdf*

$$g_3(x) = \frac{\theta\ (\theta x)^{\alpha-1}(\alpha+\gamma x)}{(\gamma+\theta)\ \Gamma(\alpha+1)}\ \theta\ e^{-\theta x}, x > 0, \alpha, \gamma, \theta > 0.$$

Elbatal *et al*. (2013) introduced a three-parameter Lindley distribution with *pdf*

$$g_4(x) = \frac{1}{1+\theta}\left(\frac{\theta^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)} + \frac{\theta^{\beta-1}x^{\beta-1}}{\Gamma(\beta)}\right)\theta\, e^{-\theta x}, x > 0, \alpha, \beta, \theta > 0$$

A similar modification of a three-parameter Lindely distribution was proposed by Abd El-Monsef (2015) with the following *pdf*:

$$g_5(x) = \frac{\theta}{\theta+\alpha}\left(1+\alpha(x+\beta)\right)\theta e^{-\theta x}, x > 0, \alpha, \beta, \theta > 0.$$

Shibu and Irshad (2016), obtained a more generalized Lindley distribution with (n + 3) parameters having the following *pdf*:

$$g_6(x) = \sum_{i=1}^{n} p_i \frac{\theta^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i-1}e^{-\theta x}, x > 0, \alpha_i, \theta > 0,$$

where,

$p_i = \frac{1}{n-1}\frac{\beta\gamma}{\theta+\beta\gamma}$, $i = 2,3, \dots, n$ and $p_1 = 1 - \sum_{i=2}^{n} p_i$, $\beta, \gamma > 0$.

This article discusses another look to the generalization of Lindely distribution. Starting with one-parameter pdf, then based on the idea of the above modifications; a new family of distributions can be suggested.

3. The Proposed Generalization: GOLD Distribution

The suggested new formula of the LD is to be a mixture distribution of k independent gamma distributions of the same scale parameter (θ) and with known but different known shape parameter (α); then the new GOLD is introduced in the following theorem.

Theorem 1. *A random variable X is said to have a GOLD density with parameter $\theta > 0$; and a known positive integer k, if the pdf (denoted by $f_{gold}(x;\theta,k)$) of X is given by*

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

with a corresponding cdf (denoted by $F_{gold}(x;\theta,k)$);

$$F_{gold}(x;\theta,k) = 1 - \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\sum_{j=1}^{k}\theta^{j-1}\,\Gamma\left(j,\frac{x}{\theta}\right). \quad (2)$$

Proof.

Following the same concept of obtaining Lindely distribution, we define $f_{gold}(x;\theta,k)$ as

$$f_{gold}(x;\theta,k) = \sum_{j=1}^{k} p_j\, g_j(x,\theta)\ ; \sum_{j=1}^{k} p_j = 1.$$

Now, by assuming $g_j(x)$ to be the pdf of Gamma(j, θ); and

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

as the weighted multiplier to satisfy the *pdf* axioms. Hence, the probability density function is introduced as follows

$$f_{gold}(x;\theta,k) = \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\left(\sum_{j=1}^{k}x^{j-1}\right)\frac{1}{\theta}e^{-x/\theta}\ ;\ x > 0, \theta > 0, \quad (3)$$

and k is a known positive integer. It is easy to show that

$$\sum_{j=1}^{k}x^{j-1} = \frac{x^k-1}{x-1}\ ; k = 1,2, \dots\ and\ x \neq 1$$

Then, the *pdf* of the GOLD will be (as given in (1)):

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

The corresponding *cdf* can be obtain by computing:

$$F(x) = \int_0^x f_{gold}(t;\theta,k)\,dt,$$

which simply can be written as

$$F(x) = \sum_{j=1}^{k} p_j \int_0^x g_j(t,\theta)dt,$$

which leads to

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

$$= 1 - \frac{1}{\sum_{j=1}^{k} \Gamma(j)\theta^{j-1}} \sum_{j=1}^{k} \theta^{j-1}\,\Gamma\left(j,\frac{x}{\theta}\right), \quad (2)$$

Where $\Gamma\left(j,\frac{x}{\theta}\right)$ is incomplete gamma function and defined as

$$\Gamma(j,x) = \int_x^\infty t^{j-1}\, e^{-t}\, dt.$$

If *j* has an integer value, then

$$\Gamma(j,x) = \Gamma(j)e^{-x} \sum_{i=0}^{j-1} \frac{x^i}{i!}. \quad \square$$

Remark. *A random variable X with the pdf and the cdf given in theorem 1, is said to follow GOLD distribution and it will be denoted later on by* $X \sim GOLD(\theta,k)$.

The graph of the *pdf* and *cdf* of the GOLD for varying values of θ and k are presented in Figure 1 and Figure 2, respectively.

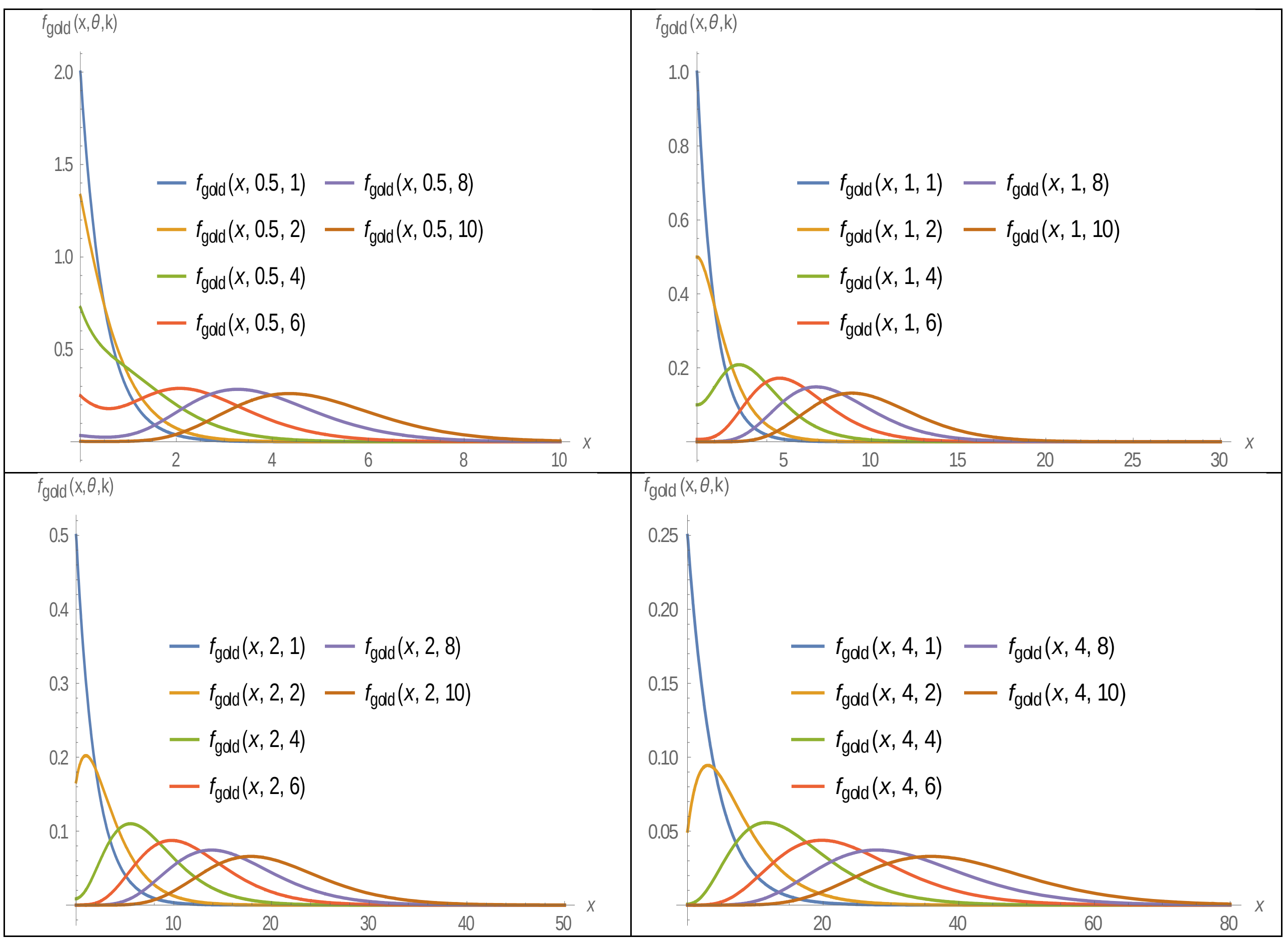


Figure 1: pdf of GOLD

 DOI: http://dx.doi.org/10.18187/pjsor.v19i2.3255

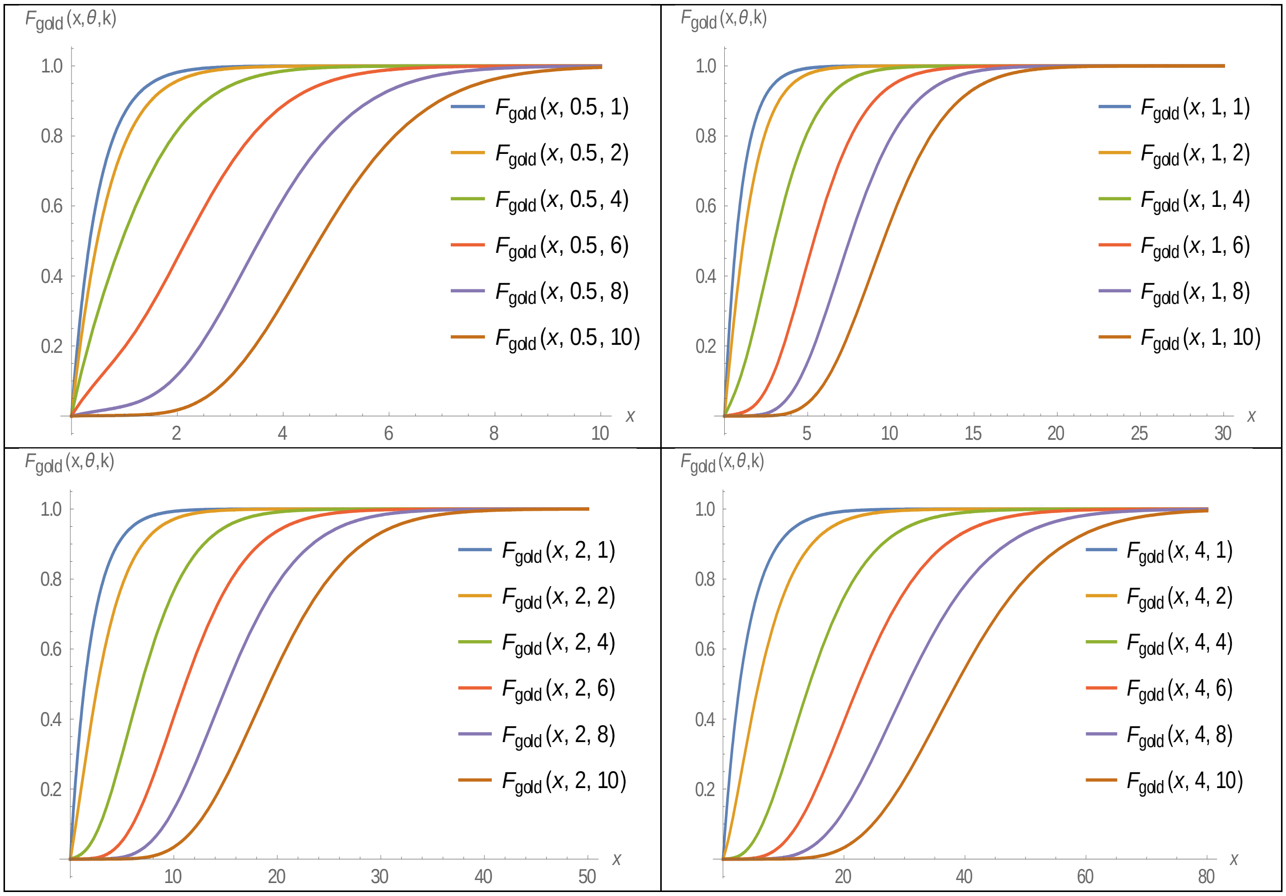


Figure 2: cdf of GOLD

The distribution in (1) is clearly a member of a one parameter exponential family, as

$$f_{gold}(x;\theta,k) = g(\theta).h(x).\ \exp[t(x).w(\theta)]$$

Where, $g(\theta) = \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j}}$, $h(x) = \sum_{j=1}^{k} x^{j-1}$, $t(x) = x$ and $w(\theta) = -\frac{1}{\theta}$.

4. Statistical Measures

4.1 Reliability Measures: Survival, Hazard and Cumulative Hazard Functions

Reliability measures are widely used to analyze time-to-event data in survival analysis. The most well-known measures are the survival, hazard or faultier rate and cumulative hazard functions; the following theorem presents these measures of GOLD family of distributions.

Theorem 2. *Let X be a random variable that follows the GOLD family of distributions, with a pdf and cdf defined in Theorem 1, then:*

1. *The survival function is given by*

$$S(x;\theta,k) = \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\sum_{j=1}^{k}\theta^{j-1}\,\Gamma\left(j,\frac{x}{\theta}\right). \tag{3}$$

2. *The hazard function is given by*

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

3. *The cumulative hazard function is given by*

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

Proof.

1. For any continuous random variable, the survival function is defined by,

$$S(x) = 1 - F(x),$$

therefore, if a random variable follows a GOLD, the survival function is given as:

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

2. The hazard function $h(x)$ which is also known as failure rate function, can be computed for any random variable using the equation

$$h(x) = \frac{f(x)}{S(x)}.$$

If a random variable follows a GOLD distribution, the hazard function is given as,

$$\begin{aligned} h(x;\theta,k) &= \frac{f_{gold}(x;\theta,k)}{S(x;\theta,k)} \\ &= \frac{\frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\cdot\frac{x^k-1}{x-1}\,\frac{1}{\theta}\,e^{-x/\theta}}{\frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\sum_{j=1}^{k}\theta^{j-1}\,\Gamma(j,\frac{x}{\theta})} \\ &= \frac{1}{\sum_{j=1}^{k}\theta^{j}\,\Gamma\left(j,\frac{x}{\theta}\right)}\frac{x^k-1}{x-1}\,e^{-x/\theta}. \end{aligned}$$

Correspondingly, the Mills ratio, $M(x)$ is defined by:

$$M(x) = \frac{1}{h(x;\theta,k)} = \left(\sum_{j=1}^{k}\theta^{j}\,\Gamma(j,\frac{x}{\theta})\right)\frac{x-1}{x^k-1}\,e^{x/\theta}.$$

3. Generally, the cumulative hazard function of a random variable is defined to be

$$ch(x) = -\log\big(S(x)\big).$$

Hence, for the GOLD family of distributions, the cumulative hazard function is

$$\begin{aligned} ch(x;\theta,k) = -\log\big(S(x;\theta,k)\big) &= -\log\left(\frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\sum_{j=1}^{k}\theta^{j-1}\,\Gamma\left(j,\frac{x}{\theta}\right)\right) \\ &= \log\left(\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}\right) - \log\left(\sum_{j=1}^{k}\theta^{j-1}\,\Gamma\left(j,\frac{x}{\theta}\right)\right). \end{aligned}$$ □

As noted earlier, theses reliability measures are important and they are of a primary interest in survival analysis. Some of their properties are worth noting and to be checked for the GOLD family of distributions.

So, it is known that the $\lim_{x\to\infty} S(x) = 0$, this fact can be proved easily using equation (3) and visualized in Figure 3 which shows that $S(x;\theta,k)$ is a decreasing function of *x*.

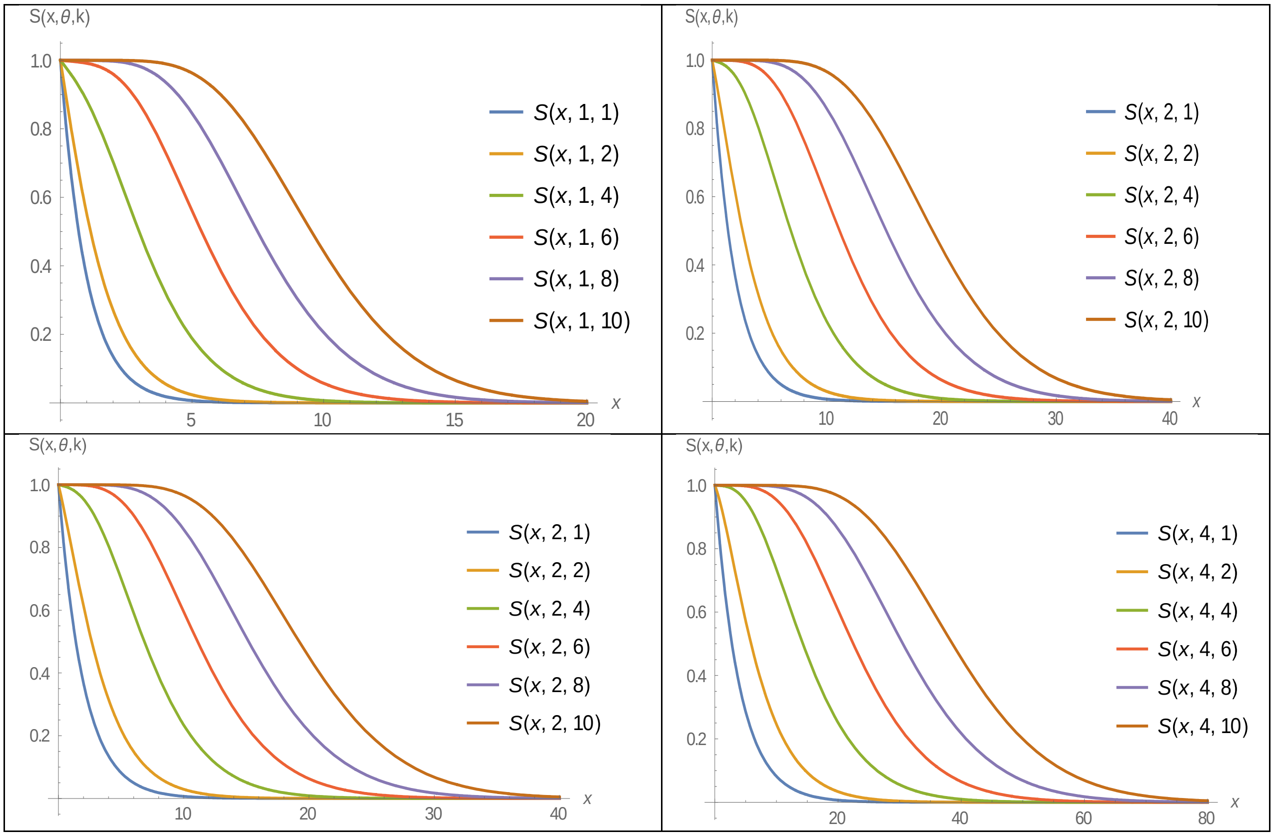
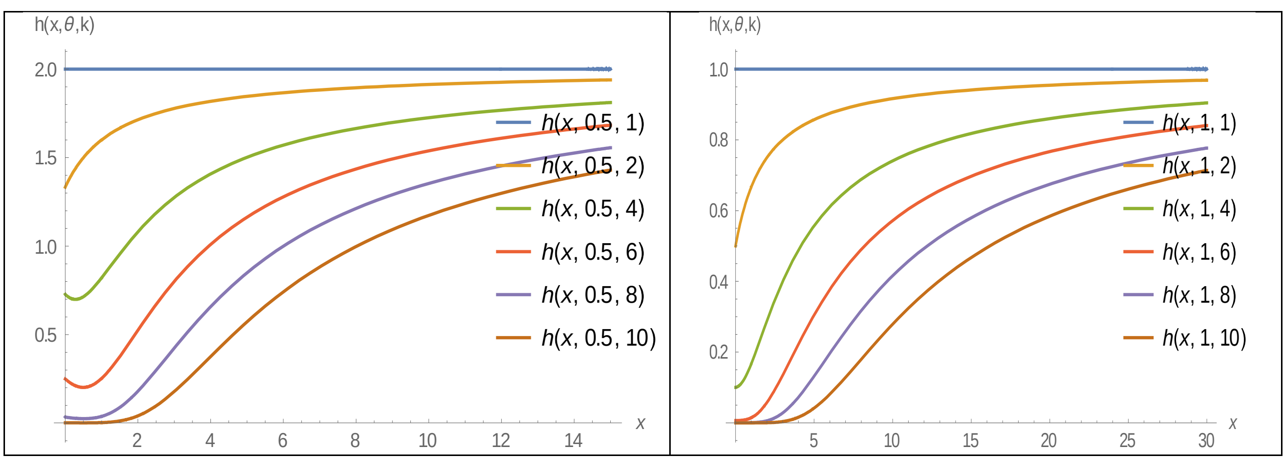


Figure 3: The graph of different forms of $S(x;\theta,k)$

On the other hand, the hazard function $h(x)$ is an estimate of the incidence rate (as a function of time) and is measured in events per unit time. It is not a probability, and it takes nonnegative values. A lower hazard rate implies a higher survival function. Figure 4 shows that $h(x;\theta,k)$ in an increasing function of x, which is an advantage compared to the exponential distribution which has constant hazard rate.



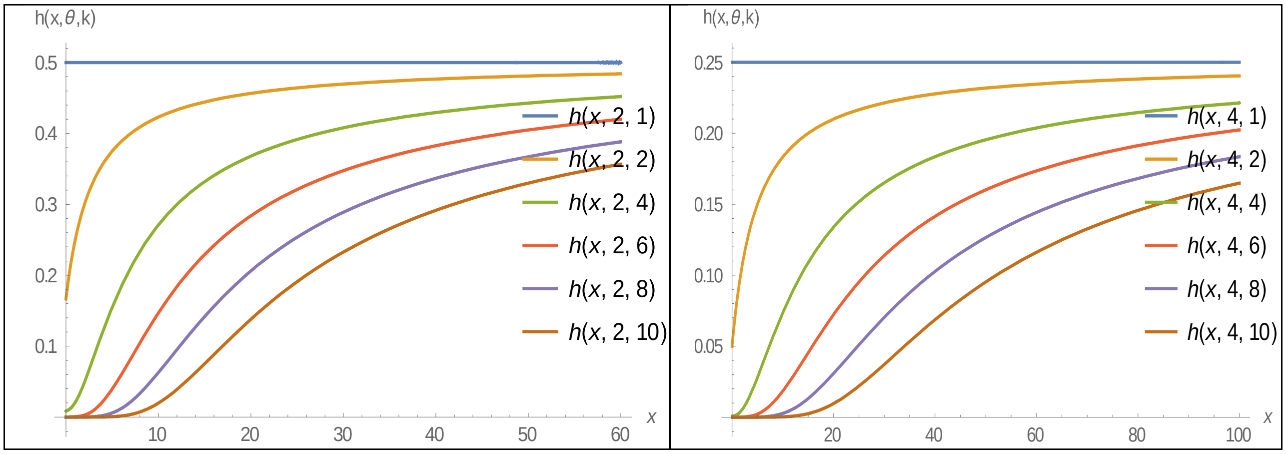


Figure 4: the graph of different forms of $h(x; \theta, k)$

Also, Cumulative Hazard Function is another measure of risk; the greater the value of $ch(x; \theta, k)$, the greater the risk of failure by time *x*. Moreover, a straight line in the cumulative hazard function plot means that the hazard rate is constant. Figure 5, presents the cumulative hazard function at varying values of θ and *k*.

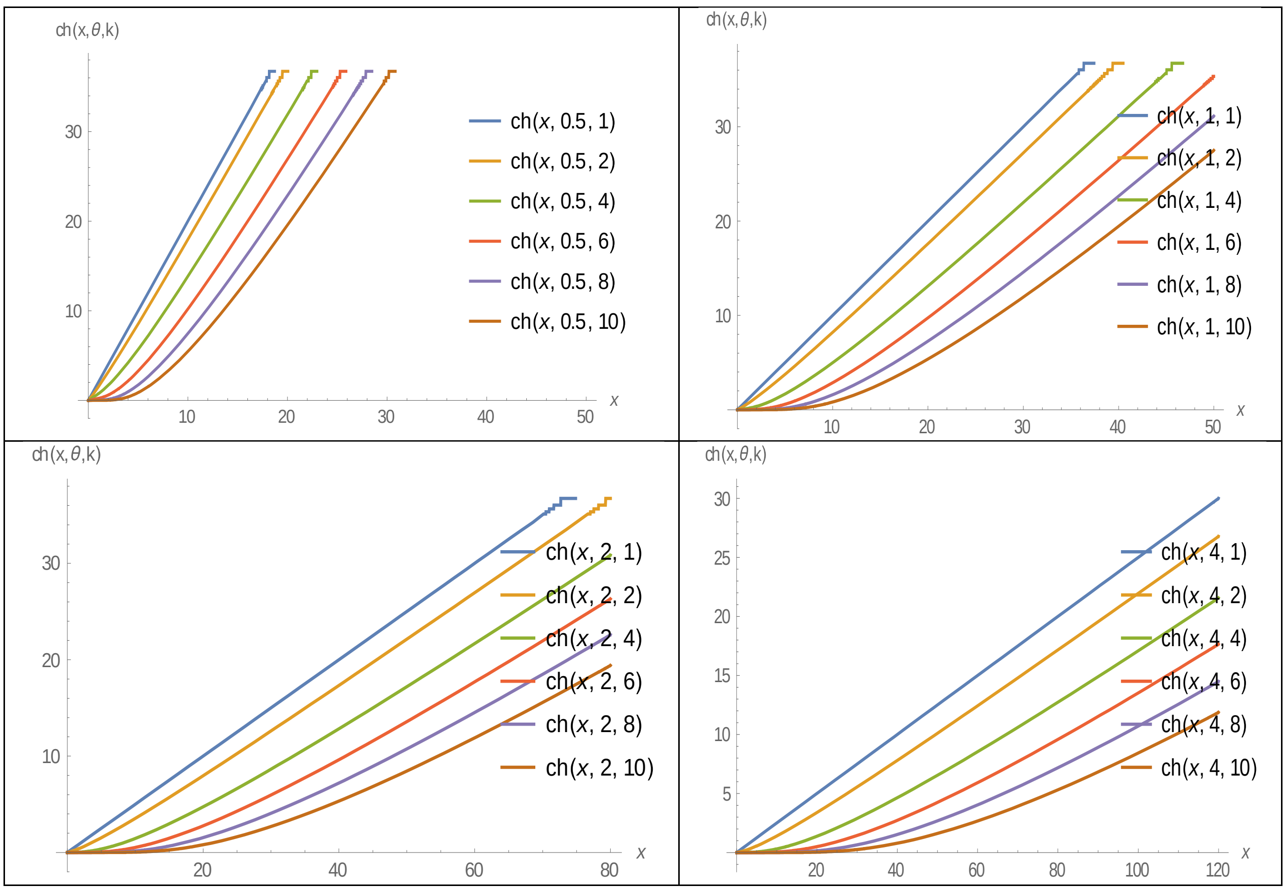


Figure 5: the graph of different forms of $ch(x; \theta, k)$

4.2 Moments and related measures

Many of the interesting characteristics and features of a distribution can be studied through its moments. Let X be a random variable following the GOLD, then expressions for the r^{th} moment about the origin,

$$E(X^r) = \sum_{j=1}^{k} p_j \, E_{g_j(x)}(X^r)$$

where $E_{g_j(x)}(X^r)$ is the r^{th} moment from a gamma distribution with parameters j and θ. Therefore,

$$E(X^r) = \frac{\sum_{j=1}^{k}\Gamma(j+r)\theta^{j+r-1}}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}, \quad r = 1,2,\ldots. \tag{4}$$

Accordingly, the following proposition expresses the mean, variance and moment generating functions of GOLD.
Proposition 1. *Let X be a random variable following the GOLD family of distributions, then*

1. *The mean of X is*

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

2. *The variance of X is*

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

3. *The moment generating function; $M_x(t)$ is given by*

$$M_x(t) = \frac{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}\left(\frac{1}{1-\theta t}\right)^j}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}, \quad \frac{1}{\theta} > t.$$

Proof.

1. Substituting $r = 1$ in equation (4), the mean of GOLD is straight forward and given by

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

2. In similar fashion, the variance can be obtained as:

$$\begin{aligned} Var(x) &= \sum_{j=1}^{k} p_j^{\,2}\, Var\,(\,g_j(x)) = \sum_{j=1}^{k} p_j^{\,2}\, j.\theta^2 \\ &= \sum_{j=1}^{k}\left(\frac{\Gamma(j)\theta^{j-1}}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\right)^2 j.\theta^2, \end{aligned}$$

and hence,

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

3. The moment generating function, $M_x(t)$ is found as follows

$$\begin{aligned} M_x(t) &= \sum_{j=1}^{k} p_j\, M_{g(x)}(t) \\ &= \sum_{j=1}^{k} p_j\,\left(\frac{1}{1-\theta t}\right)^j, \end{aligned}$$

which is equivalent to

$$M_x(t) = \frac{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}\left(\frac{1}{1-\theta t}\right)^j}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}, \quad \frac{1}{\theta} > t.$$

On the other hand, for empirical purposes, the shape of many distributions can be usefully described by the incomplete moments.

$$m_r(y) = \int_{-\infty}^{y} x^r f(x)dx.$$

For the GOLD distribution we have

$$m_r(y) = \frac{\sum_{j=1}^{k}\left(\Gamma(j+r) - \Gamma\left(j+r,\frac{y}{\theta}\right)\right)\theta^{j+r-1}}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}.$$

4.3 Mean residual life function

Mean residual life function for a continuous random variable X is known as the Borel measurable function defined as the expected remaining life given survival to age x. Mathematically in terms of expected value, the mean residual life is defined as:

$$m(x) = E(X - x \mid X > x),$$

which is equivalent to

$$m(x) = \frac{1}{S(x)}\int_x^\infty S(t)dt,$$

where, $S(x)$ is the survival function.

Corollary 1. *For a random variable X that follows GOLD family of distributions, the mean residual life function is*

$$m(x;\theta,k) = \frac{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}{\sum_{j=1}^{k}\theta^{j-1}\,\Gamma(j,\frac{x}{\theta})}\sum_{i=1}^{k}\theta^{i-1}\int_x^\infty \Gamma(i,\frac{t}{\theta})dt,$$

which is simplified to

$$\frac{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}{\sum_{j=1}^{k}\theta^{j-1}\,\Gamma(j,\frac{x}{\theta})}\sum_{i=1}^{k}\frac{x^{i+1}e^{-\frac{x}{\theta}} + \theta^{-i}(\theta\cdot i - x)\,\Gamma(i+1,\frac{x}{\theta})}{\theta\cdot i}.$$

4.4 Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample from GOLD family of distributions, then the *pdf* of the r^{th} order statistics ; $Y = X_{(r)}$ is:

$$\begin{aligned} g(y;\theta,k) &= \frac{f(y)}{\beta(r, n-r+1)}\,F(y)^r\,(1-F(y))^{n-r} \\ &= r\binom{n}{r}\,f(y)\,F(y)^r\,S(y)^{n-r} \end{aligned}$$

Therefore,

$$g(y;\theta,k) = r\binom{n}{r}\frac{\sum_{j=1}^{k}y^{j-1}}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}.\,\frac{1}{\theta}e^{-y/\theta}\left(1 - \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\sum_{j=1}^{k}\theta^{j-1}\,\Gamma\left(j,\frac{y}{\theta}\right)\right)^r\left(\frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\sum_{j=1}^{k}\theta^{j-1}\,\Gamma\left(j,\frac{y}{\theta}\right)\right)^{n-r}.$$

4.5 Inequality Measures

Many inequality measures are studied and have their applications in economics to study poverty and income; also, they have importance in reliability, demography and medical sciences. The Gini index is a well-known measure for summarizing income inequality. Its range in [0,1]. A Gini index of 0 expresses perfect equality, which means that every person in the population has an exactly equal income. A Gini index of 1 expresses a maximal inequality, that is only one person has all the income. Also, the Lorenz and Bonferroni curves are major tools for analyzing data in economics and reliability.

The next theorem presents the Gini Index, Lorenz curve and Bonferroni curve of the GOLD family of distributions.

Theorem 3. *For a random variable X which follows the GOLD family of distributions, the Gini index, Lorenz curve and the Bonferroni curve are given by*

1. *Gini Index*

$$G = 1 - \frac{\int_0^\infty \left\{\left(\sum_{j=1}^k \theta^{2j-2} \left(\Gamma\left(j,\frac{x}{\theta}\right)\right)^2\right) + \left(\sum_{i=1}^k \sum_{r=1}^k \theta^{i+r-2}\, \Gamma\left(i,\frac{x}{\theta}\right) \Gamma\left(r,\frac{x}{\theta}\right)\right)\right\} dx}{\left(\sum_{j=1}^k \Gamma(j)\theta^{j-1}\right)\left(\sum_{j=1}^k \Gamma(j+1)\theta^j\right)}.$$

2. *Lorenz Curve*

$$L(P) = \frac{1}{\mu} \frac{\sum_{j=1}^k \left(\Gamma(j+1) - \Gamma\left(j+1,\frac{q}{\theta}\right)\right)\theta^j}{\sum_{j=1}^k \Gamma(j)\theta^{j-1}}; \quad q = F^{-1}(p).$$

3. *Bonferroni curve*

$$B(P) = \frac{1}{P\mu} \frac{\sum_{j=1}^k \left(\Gamma(j+1) - \Gamma\left(j+1,\frac{q}{\theta}\right)\right)\theta^j}{\sum_{j=1}^k \Gamma(j)\theta^{j-1}}.$$

Proof.

1. For a cumulative distribution function $F(x)$ that is piecewise differentiable, has a mean μ, and is zero for all negative values of, the Gini index G is calculated using the following equation,

$$G = 1 - \frac{1}{\mu}\int_0^\infty (1 - F(x))^2 dx$$
$$= 1 - \frac{1}{\mu}\int_0^\infty S(x)^2 dx\,.$$

Therefore, the Gini index for GOLD is

$$G = 1 - \frac{1}{\mu}\int_0^\infty \left(\frac{1}{\sum_{j=1}^k \Gamma(j)\theta^{j-1}} \sum_{j=1}^k \theta^{j-1}\, \Gamma\left(j,\frac{x}{\theta}\right)\right)^2 dx$$

$$= 1 - \frac{1}{\frac{\sum_{j=1}^k \Gamma(j+1)\theta^j}{\sum_{j=1}^k \Gamma(j)\theta^{j-1}}} \int_0^\infty \left(\frac{1}{\sum_{j=1}^k \Gamma(j)\theta^{j-1}} \sum_{j=1}^k \theta^{j-1}\, \Gamma\left(j,\frac{x}{\theta}\right)\right)^2 dx$$

$$= 1 - \frac{1}{\left(\sum_{j=1}^k \Gamma(j)\theta^{j-1}\right)\left(\sum_{j=1}^k \Gamma(j+1)\theta^j\right)} \int_0^\infty \left(\sum_{j=1}^k \theta^{j-1}\, \Gamma\left(j,\frac{x}{\theta}\right)\right)^2 dx$$

$$= 1 - \frac{\int_0^\infty \left\{\left(\sum_{j=1}^k \theta^{2j-2} \left(\Gamma\left(j,\frac{x}{\theta}\right)\right)^2\right) + \left(\sum_{i=1}^k \sum_{r=1}^k \theta^{i+r-2}\, \Gamma\left(i,\frac{x}{\theta}\right) \Gamma\left(r,\frac{x}{\theta}\right)\right)\right\} dx}{\left(\sum_{j=1}^k \Gamma(j)\theta^{j-1}\right)\left(\sum_{j=1}^k \Gamma(j+1)\theta^j\right)}.$$

Which can computed for given values of k.

2. Lorenz curve for a nonnegative random variable, is the graph of $(L(F(x)), F(x))$, where

$$L(P) = \frac{1}{\mu}\int_0^q t\, f(t)dt\,, \quad q = F^{-1}(p).$$

using the incomplete moment results, then Lorenz Curve of GOLD will be

$$L(P) = \frac{1}{\mu} \frac{\sum_{j=1}^k \left(\Gamma(j+1) - \Gamma\left(j+1,\frac{q}{\theta}\right)\right)\theta^j}{\sum_{j=1}^k \Gamma(j)\theta^{j-1}}; \quad q = F^{-1}(p).$$

3. The Bonferroni curve is defined to be

$$B(P) = \frac{1}{P\mu}\int_0^q t\, f(t)dt,$$

and hence, the Bonferroni curve for GOLD is

$$B(P) = \frac{1}{P\mu}\frac{\sum_{j=1}^{k}\left(\Gamma(j+1) - \Gamma\left(j+1,\frac{q}{\theta}\right)\right)\theta^{j}}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}, \; q = F^{-1}(p).$$ □

4.6 Entropy Measure

As measures of uncertainty, entropies study the amount information gain. A popular entropy is Shannon entropy and the next proposition presents it for GOLD family of distributions.

Proposition 2. *The Shannon entropy of the GOLD family of distributions is given by*

$$H\left(f_{gold}(x;\theta,k)\right) = Log\left(\sum_{j=1}^{k}\Gamma(j)\theta^{j}\right) + \frac{\mu}{\theta} - -\int_0^\infty Log\left(\sum_{j=1}^{k}x^{j-1}\right)f(x)dx.$$

Proof.

Generally, for any *pdf* $f(x)$, the Shannon entropy is

$$H\big(f(x)\big) = -E\log\big(f(x)\big),$$

Recall the *pdf* of GOLD family of distributions defined in Theorem 1;

$$f_{gold}(x;\theta,k) = \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\left(\sum_{j=1}^{k}x^{j-1}\right)\frac{1}{\theta}e^{-x/_{\theta}} \;, x>0, \theta>0.$$

Then we get,

$$H(f(x)) = \int_0^\infty\left[-Log\left(\frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j-1}}\right) - Log\left(\sum_{j=1}^{k}x^{j-1}\right) + \frac{x}{\theta}\right]\cdot f_{gold}(x;\theta,k)\,dx\,,$$

And hence the result is obtained.

5 Parameter estimation

5.1 MLE

Assume that $X_1, X_2, \ldots, X_n$ be a random sample from GOLD. Then, the maximum likelihood estimate of θ is given in the next proposition.

Proposition 3. *The MLE of θ for the GOLD family of distributions defined in Theorem 1, is the solution of the equation*

$$\bar{x}\cdot\sum_{j=1}^{k}\Gamma(j)\theta^{j} = \sum_{j=1}^{k}\Gamma(j+1)\theta^{j+1}$$

Proof.

The likelihood function of GOLD is given by,

$$L(\theta;x) = \left(\frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta^{j}}\right)^{n}\cdot\prod_{i=1}^{n}\frac{x_i^k-1}{x_i-1}.e^{\frac{-n\bar{x}}{\theta}}; \quad \bar{x}>0\,,\theta>0$$

thus, the log-likelihood is

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

And hence,

$$\frac{\partial log\,L(\theta;x)}{\partial\theta} = \frac{\bar{x}}{\theta^2} - \frac{\sum_{j=1}^{k}\Gamma(j+1)\theta^{j-1}}{\sum_{j=1}^{k}\Gamma(j)\theta^{j}}\,,$$

the result is obtained by equating the above equation to zero.

Corollary 2. *A special case of Proposition 3, is when k=1* (*exponential distribution*) *and k=2* (*Lindley distribution*). *The MLE of θ is*

1. *When k=1, $\hat{\theta} = \bar{x}$.*

2. *When k=2,*

$$\hat{\theta} = \frac{(\bar{x}-1)+\sqrt{\bar{x}^2+6\bar{x}+1}}{4}.$$

Remark. *For values of $k \geq 3$, the MLE presented in Proposition2 can be found numerically but it depends on the value of $\bar{x}$.*

5.2 MOM

The Method of Moments can also be applied, to find the estimate of θ

Corollary 3. *The method of moments estimates of θ for the GOLD family of distribution are*

1. *When k=1, $\hat{\theta} = \bar{x}$.*
2. *When $k = 2$,*

$$\hat{\theta} = \frac{(\bar{x}-1)+\sqrt{\bar{x}^2+6\bar{x}+1}}{4},$$

3. *When $k \geq 3$, the estimate will be the solution of the equation*

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

Remark. *The maximum likelihood estimation method and the method of moments yield the same estimates of θ.*

6. GOLD Related Distribution and possible extensions

The proposed distribution $GOLD(\theta, k)$ could be considered as a family of several distributions based on a given value of *k*. The following table represents several special cases of $GOLD(\theta, k)$:

Table 1: Special cases of $GOLD(\theta, k)$:

$GOLD(\theta, k)$	Special cases of GOLD distribution	Reference
$GOLD\left(\frac{1}{\theta}, 1\right)$	Exponential Distribution	Balakrishnan and Basu (1996)
$GOLD\left(\frac{1}{\theta}, 2\right)$	Lindely Distribution	Lindely (1958)
$GOLD\left(\frac{1}{\theta}, 3\right)$	Sujatha distribution	Shanker (2016a)
$GOLD\left(\frac{1}{\theta}, 4\right)$	Amarendra distribution	Shanker (2016b)
$GOLD\left(\frac{1}{\theta}, 5\right)$	Devya distribution	Shanker (2016c)
$GOLD\left(\frac{1}{\theta}, 6\right)$	Shambhu distribution	Shanker (2016d)

Moreover, there are several possible extension of $GOLD(\theta, k)$ depending on the scale and shape parameters of the distribution:

(I) Assume the shape parameter k is a given integer number and the scale parameter vary within the combinations of the gamma distributions; that is to say

$$f_{gold}(x;\theta,k) = \sum_{j=1}^{k} p_j \; Gamma(j,\theta_j)\,;\; \sum_{j=1}^{k} p_j = 1,$$

then the new extension will have the following pdf:

$$f_{gold}(x;\theta,k) = \frac{1}{\sum_{j=1}^{k}\Gamma(j)\theta_j{}^{j}}\left(\sum_{j=1}^{k} x^{j-1}\, e^{-x/\theta_j}\right) \;, x > 0, \theta_j > 0$$

(II) Assume the shape parameter is fixed and equal to , then the possible formulation of GOLD distribution will be $f_{gold}(x;\theta,k) = \sum_{j=1}^{k} p_j \; Gamma(\alpha,\theta_j)\,;\; \sum_{j=1}^{k} p_j = 1$, this will give us the following pdf:

$$f_{gold}(x;\theta,k) = x^{\alpha-1}\left(\sum_{j=1}^{k}\frac{\theta_j{}^{\alpha}}{\sum_{j=1}^{k}\theta_j{}^{\alpha}}\; e^{-x/\theta_j}\right) \;, x > 0, \alpha, \theta_j > 0$$

(III) The third possible extension of the proposed distribution to assume random parameters (shape and scale) $f_{gold}(x;\theta,k)=\sum_{j=1}^{k} p_j \ Gamma(\alpha_j,\theta_j)$; $\sum_{j=1}^{k} p_j = 1$, then the new GOLD distribution will have the following form

$$f_{gold}(x;\theta,k)=\left(\sum_{j=1}^{k}\frac{1}{\sum_{j=1}^{k}\Gamma(\alpha_j)\theta_j^{\alpha_j}}\ x^{\alpha_j-1}\,e^{-x/\theta_j}\right),x>0,\alpha_j,\theta_j>0$$

Also, all previous suggested extensions can be generalized to a more complicated form by adding different parameters "such as power, location, rate … etc" to the distribution function or by assuming inverted type of the distribution.

7. Application to real data

We apply the data set reported by Fuller *et al*. (1994) to the GOLD family of distributions; it is about the strength of glass of the aircraft window. The data are as follows

18.83, 20.80, 21.657, 23.03, 23.23, 24.05, 24.321, 25.50, 25.52, 25.80, 26.69, 26.77, 26.78, 27.05, 27.67, 29.90, 31.11, 33.20, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381

We present the fitting of GOLD in next Table alongside other distributions. In order to compare distributions, - 2ln$L$, AIC (Akaike Information Criterion), AICC (Akaike Information Criterion Corrected), BIC (Bayesian Information Criterion), K-S Statistics (Kolmogorov-Smirnov Statistics) and they are presented with the parameter estimate at each value of *k*.

Table 2: Fitting aircraft window data using GOLD and other distributions

Special cases of GOLD distribution	k	Parameter Estimate	-2ln$L$	AIC	AICC	BIC	K-S Statistics
Exponential Distribution	1	30.8118	274.53	276.53	276.67	277.96	0.426
$GOLD(\theta,k)$	2	15.8760	253.99	255.99	256.13	257.42	0.333
Lindely Distribution		15.8760	253.99	255.99	256.13	257.42	0.333
$GOLD(\theta,k)$	3	10.4592	241.50	243.50	243.64	244.9371	0.270
Sujatha distribution		10.4592	241.50	243.50	243.64	244.94	0.270
$GOLD(\theta,k)$	4	7.79472	233.41	235.41	235.55	236.84	0.225
Amarendra distribution		7.79472	233.41	235.41	235.55	236.84	0.225
$GOLD(\theta,k)$	5	6.21612	227.68	229.68	229.82	231.12	0.193
Devya distribution		6.21612	227.68	229.68	229.82	231.82	0.193
$GOLD(\theta,k)$	6	5.17071	223.40	225.40	225.53	226.83	0.167
Shambhu distribution		5.17071	223.40	225.40	225.53	226.83	0.167
$GOLD(\theta,k)$	7	4.42679	220.07	222.07	222.21	223.51	0.145
$GOLD(\theta,k)$	8	3.87022	217.44	219.44	219.57	220.88	0.127
$GOLD(\theta,k)$	9	3.43807	215.33	217.33	217.46	218.76	0.119
$GOLD(\theta,k)$	10	3.09279	213.63	215.63	215.77	217.06	0.110

The above table shows that all defined distributions are a special case of GOLD distribution. Also, it is clear that as *k* gets larger, AIC and K-S statistic (and other measures) gets smaller. In future researches we will consider the optimum choice of *k* and weather if it may be considered as a parameter or not.

Conclusion

A new generalization of one parameter Lindely distribution is discussed. The new distribution is called GOLD distribution which could be considered as a new formulation of a lifetime distribution. Several statistical measures has been discussed and the MLE and MOM parameter estimators are given. It is shown that GOLD distribution is a generalizations of many distributions included exponential, Lindely, Sujatha, Amarendra, Devya and Shambhu distributions. Also, possible extensions of the GOLD distribution are suggested. The importance of the GOLD distribution in survival data analysis is discussed by a real lifetime data and the goodness of fit of the distribution has been compared to other distributions. The future works to be done on GOLD distribution are to study its performance assuming the shape parameter is unknown. Other possible extension could be to study the Morgenstern type bivariate gold distribution.

References

1. Abd El-Monsef M. (2015): A New Lindley Distribution with Location Parameter, Communications in Statistics - Theory and Methods, DOI:10.1080/03610926.2014.941496
2. Abouammoh A, Alshangiti A, Ragab I. (2015). A new generalized Lindley distribution. Journal of Statistical Computation and Simulation, 85(18): 3662–3678.
3. Akbar A. Hassan S , Nadarajahc S, Sharafia F. 2016. A new weighted Lindley distribution with application. Brazilian Journal of Probability and Statistics 30(1):1-27.
4. Alkarni S (2015) Extended Power Lindley distribution-A new Statistical model for non-monotone survival data. European journal of statistics and probability 3(3): 19-34.
5. Bakouch H. S., Al-Zahrani B. M., Al-Shomrani A. A., Marchi V. A., and Louzada F.,(2012). An extended Lindley distribution, Journal of the Korean Statistical Society, Vol(41), 75-85.
6. Deniz E, Ojeda E (2011) The discrete Lindley distribution-Properties and Applications. Journal of Statistical Computation and Simulation 81: 1405-1416.
7. Elbatal I, Merovi F, Elgarhy M (2013) A new generalized Lindley distribution. Mathematical theory and Modeling 3(13): 30-47.
8. Fuller, E.J., Frieman, S., Quinn, J., Quinn, G., and Carter, W. (1994): Fracture mechanics approach to the design of glass aircraft windows: A case study, SPIE Proc 2286, 419-430.
9. Ghitany M. E., Al-Mutairi D. K., Balakrishnan N., Al-Enezi L. J. (2013), Power Lindley distribution and associated inference. Computational Statistics and Data Analysis, 64: 20–33.
10. Ghitany M. E., Alqallaf F., Al-Mutairi D. K., and Husain H. A., (2011) A two-parameter weighted Lindley distribution and its applications to survival data, Mathematics and Computers in Simulation, Vol. (81), no. 6,1190-1201.
11. Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and computers in simulation*, *78*(4), 493-506.
12. Gomez D. E., Ojeda E. C. (2011). The discrete Lindley distribution: Properties and applications. Journal of Statistical Computation and Simulation, 81(11): 1405–1416.
13. Hassan M.K.,(2014), On the Convolution of Lindley Distribution, Columbia International
14. Jodra, P. (2010). “Computer Generation of Random Variables with Lindley or Poisson–Lindley Distribution via the Lambert W Function”. Mathematics and Computers in Simulation. 81, 851–859.
15. Kadilar G. O., Cakmakyapan S. (2017). The Lindley family of distributions: Properties and applications. Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics.
16. Lindely, D, V., (1958). Fiducial distributions and Bayes’ theorem, Journal of the Royal Society, series B, 20, pp. 102–107.
17. Maya R., Irshad M. R. 2017. NEW EXTENDED GENERALIZED LINDLEY DISTRIBUTION: PROPERTIES AND APPLICATIONS. STATISTICA, anno LXXVII, n. 1, 33- 52.
18. Merovci F (2013) Transmuted Lindley distribution. Int J Open Problems Compt Math 6(2): 63-72.
19. Nadarajah S., Bakouch H., Tahmasbi R. (2011). A generalized lindley distribution. Sankhya B - Applied and Interdisciplinary Statistics, 73: 331–359.
20. Oluyede BO, Yang T (2014) A new class of generalized Lindley distribution with applications. Journal of Statistical Computation and Simulation 85(10): 2072-2100.
21. Pararai M, Liyanage GW, Oluyede BO (2015) A new class of generalized Power Lindley distribution with applications to lifetime data. Theoretical Mathematics & Applications 5(1): 53-96.
22. Sankaran, M. (1970). 275. note: The discrete poisson-lindley distribution. *Biometrics*, 145-149.
23. Shanker, R., (2016a). Sujatha distribution and Its Applications, Statistics in Transition new series, 17 (3), pp. 391–410.
24. Shanker, R., (2016 b). Amarendra distribution and Its Applications, American Journal of Mathematics and Statistics, 6 (1), pp. 44–56.

25. Shanker, R., (2016 c). Devya distribution and Its Applications, International Journal of Statistics and Applications, 6 (4), pp. 189–202.
26. Shanker, R., (2016 d). Shambhu Distribution and Its Applications, International Journal of Probability and Statistics, 5 (2), pp. 48–63.
27. Shanker R, Mishra A (2013a) A two-parameter Lindley distribution. Statistics in Transition-new series 14 (1): 45-56.
28. Shanker R, Mishra A (2013b) A quasi Lindley distribution. African Journal of Mathematics and Computer Science Research 6(4): 64-71.
29. Shanker, R.; Sharma, S. and Shanker, R. (2013c). "A Two-Parameter Lindley Distribution for Modeling Waiting and Survival Times Data". Applied Mathematics. 4: 363-368.
30. Sharma, Vikas Kumar, et al. 2015. "The generalized inverse Lindley distribution: A new inverse. Publishing Contemporary Mathematics and Statistics, Vol. (2) No. 1,47-54.
31. Shibu D. S., Irshad M. R.. 2016. EXTENDED NEW GENERALIZED LINDLEY DISTRIBUTION. *STATISTICA, anno LXXVI, n. 1, 41-56.*
32. Singh, S. K., Singh, U., & Sharma, V. K. (2014). The truncated lindley distribution: Inference and application. Journal of Statistics Applications & Probability, 3(2), 219.
33. Sharma, V. K., Singh, S. K., Singh, U., & Merovci, F. (2016). The generalized inverse Lindley distribution: A new inverse statistical model for the study of upside-down bathtub data. Communications in Statistics-Theory and Methods, 45(19), 5709-5729.
34. Zakerzadeh H., Dolati A. (2009). Generalized Lindley distribution. Journal of Mathematical Extension, 3(2): 13–25.
35. Zamani, H. and Ismail, N., 2010. Negative binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics*, *6*(1), pp.4-9.
36. Zakerzadeh, H., & Dolati, A. (2009). Generalized Lindley Distribution. *Journal of Mathematical extension*.