A New Generalization of Exponential Distribution for Modelling reliability Data

V. Jilesh, Aifoona Ahammed P.M

•

Department of Statistics, Government Arts and Science College, Kozhikode. Kerala, India.
jileshstat@gmail.com
aifoonaahammed@gmail.com

Abstract

In this paper, a new generalization of the exponential distribution is proposed. Different properties, important reliability measures and special cases of this distribution are investigated. Unknown parameters are estimated using the maximum likelihood method of estimation. A simulation study is carried out to assess the accuracy of the maximum likelihood estimates. Two real data sets are successfully modelled with the proposed distribution.

Keywords: Exponential distribution, Reliability, Maximum likelihood estimation.

1. Introduction

Recently, researchers are more interested in developing new probability distributions and generalizations from the existing family of distributions. The aim of more realistic modelling of complex datasets can be attained through such generalizations. Such newly formed distributions are also showing better flexibilty and properties than the baseline distribution, becomes more suitable in reliability studies and other related fields. For instance, see Eugene et al [3], Bourguignon et al. [4], Cordeiro and Castro [3], Marshall and Olkin[9], Zografos and Balakrishnan [15], Silva et al. [14], Jayakumar and Mathew [7], Nadarajah and Kotz [10], Nadarajah and Kotz [12], Nadarajah and Gupta [11]. One important family of distribution, which is the basis of many other well studied probability distributions is the exponential distribution. One main limitation of this distribution in using the reliability modelling is it's constant hazard rate. So we can find different generalizations of the exponential distribution to overcome this problem. Some important generalizations are gamma, Weibull, Rayleigh and Generalized exponential distribution of Gupta and Kundu [5]. But, in some of the datasets, we can observe, a sudden drop in the frequency of observations after some specific data points. But the available generalizations are not appropriate to model that kind of datasets, which demands a breakage in the flexibility of the probability density function. Here we introduce a new distribution to model such data sets. The family of distrbution under study is derived using the exponential distribution and two sided power distribution of René Van Dorp and Kotz [13]. The probability density function of the two sided power distribution is given by,

$$g(x) = \begin{cases} \alpha \left(\frac{x}{\theta}\right)^{\alpha-1}, & if \quad 0 < x \leq \theta \\ \alpha \left(\frac{1-x}{1-\theta}\right)^{\alpha-1} & ,if \quad \theta \leq x < 1. \end{cases} \quad \alpha > 0, 0 < \theta < 1. \tag{1}$$

This paper is organized as follows. In section 2, we propose the new distribution and discuss it's basic properties, Reliability properties are studied in Section 3. The estimation of unknown parameters in the proposed distribution is done using maximum likelihood method and is discussed in Section 4. A simulation study to check the estimation procedure is conducted in section 5. In section 6, we successsfully modelled the fatigue failure data and aircraft failure time dataset using the proposed distribution.

2. Definition and Properties

A random variable X is said to follow a generalized exponential distribution, denoted by G(α, θ), if it's probability density function (pdf) is of the form

$$f(x) = \begin{cases} \alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1}, & if \quad 0 < x \le -ln(1-\theta) \\ \alpha e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1}, & if \quad -ln(1-\theta) \le x < \infty, \end{cases} \tag{2}$$

where $0 < \theta < 1$ and $\alpha > 0$, α is not necessarily be an integer. When α=1, the above probability density function is the probability density function of exponential random variable with mean unity. The above probability denssity function can be derived by mixing the two sided power distribution and the exponential distribution. The shape of the probability density curves given by (2) for different values of parameters are shown in Figure (1). Note that for the values $\alpha > 1$ the density curve first increases and reach a maximum value corresponding to $x = log(\alpha)$ and then decreases, there is a cutting point at $-ln(1-\theta)$. For $0 < \alpha < 1$, the curve is always decreases with a cutting point at $-ln(1-\theta)$. The values taken by the parameter θ doesnot effect the monotone behaviour of the curve. Mode of the distribution is $log(\alpha)$. To derive the cumulative distribution function (CDF) of the proposed distribution, we have to consider two cases.

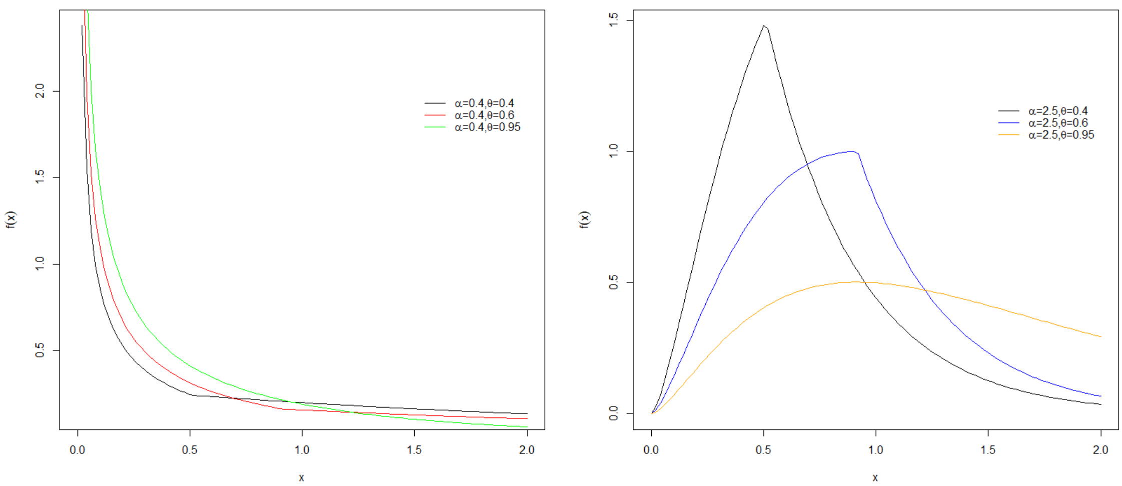

Figure 1: *Density plot of G(α,θ) for different values of θ*

Case 1: $0 < t \le -ln(1-\theta)$,
Then

$$\begin{aligned} F(t) &= \int_0^t \alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx \\ &= \frac{\alpha}{\theta^{\alpha-1}} \int_0^t e^{-x}(1-e^{-x})^{\alpha-1} dx \end{aligned}$$

by substitution for $1-e^{-x}$ by u, we get

$$\begin{aligned} F(t) &= \frac{\alpha}{\theta^{\alpha-1}} \int_0^{1-e^{-t}} u^{\alpha-1} du \\ &= \theta\left(\frac{1-e^{-t}}{\theta}\right)^{\alpha}. \end{aligned}$$

Case 2: $-ln(1-\theta) < t < \infty$

$$F(t) = \int_0^{-ln(1-\theta)} \alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx + \int_{-ln(1-\theta)}^{t} \alpha e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1} dx$$
$$= 1-(1-\theta)\frac{e^{-\alpha t}}{(1-\theta)^{\alpha}}.$$

Thus, the cumulative distribution function is given by

$$F(x) = \begin{cases} \theta\left(\frac{1-e^{x}}{\theta}\right)^{\alpha}, & if \quad 0 < x \leq -ln(1-\theta) \\ 1-(1-\theta)\frac{e^{-\alpha x}}{(1-\theta)^{\alpha}}, & if \quad -ln(1-\theta) \leq x < \infty. \end{cases} \tag{3}$$

Similarly, the ***quantile function*** can be derived by inverting the distribution function. Thus, we obtain

$$Q(u) = F^{-1}(u), \quad 0 < u < 1 \tag{4}$$
$$= \begin{cases} -ln\left[1-\theta\left(\frac{u}{\theta}\right)^{\frac{1}{\alpha}}\right] & 0 < u \leq \theta \\ -\frac{1}{\alpha}\left[ln(1-u)+(\alpha-1)ln(1-\theta)\right] & \theta < u < 1. \end{cases}$$

The simulation of random sample having G(α,θ) distribution can be done using the above equation, where U=u is a realization from a uniform random variable, that is $U \longrightarrow U(0,1)$.
Moment generating function of a random variable X with G(α,θ) distribution, is given by,

$$M_X(t) = \int_0^{-ln(1-\theta)} e^{tx}\alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx + \int_{-ln(1-\theta)}^{\infty} e^{tx}\alpha e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1} dx \tag{5}$$
$$= \frac{\alpha}{\theta^{\alpha-1}}\int_0^{-ln(1-\theta)} e^{-x(1-t)}(1-e^{-x})^{\alpha-1}dx + \frac{\alpha}{(1-\theta)^{\alpha-1}}\int_{-ln(1-\theta)}^{\infty} e^{-x(\alpha-t)}dx$$
$$= \frac{\alpha}{\theta^{\alpha-1}}B(\theta,\alpha,1-t) + \frac{\alpha}{(1-\theta)^{\alpha-1}}\frac{(1-\theta)^{\alpha-t}}{(\alpha-t)},$$

where

$$B(\theta,m,n) = \int_0^{\theta} u^{m-1}(1-u)^{n-1}du$$

is the incomplete beta function or the distribution function of beta 1^{st} kind.
The k^{th} order raw moment for G(α,θ) distribution can be written as

$$E(X^k) = \int_0^{\infty} x^k f(x)dx \tag{6}$$
$$= \int_0^{-ln(1-\theta)} \alpha x^k e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx + \int_{-ln(1-\theta)}^{\infty} \alpha x^k e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1} dx.$$

But, it is difficult to get a good expression for the above integral. Hence, we have calculated the moments numerically. Table (1) gives the the mean and variance of G(α,θ), for different values taken by the parameters α and θ.

An ***entropy*** is a measure of variation or uncertainty. The Rényi entropy of a random variable with probability density function f(·) is defined as

$$I_R(x) = \frac{1}{1-\gamma}log\int_0^{\infty} f^{\gamma}(x)dx, \gamma > 0, \gamma \neq 1.$$

Table 1: *mean and variance of G(α,θ), for different values α and θ*

	$\theta = 0.3$		$\theta = 0.5$		$\theta = 0.8$	
α	E(X)	V(X)	E(X)	V(X)	E(X)	V(X)
0.5	1.6828	3.8627	1.4467	3.5725	1.0269	2.0617
1	1	1	1	1	1	1
1.5	0.7772	0.4552	0.8669	0.4825	1.0480	0.6123
2	0.6677	0.2606	0.8068	0.2892	1.0976	0.4406
2.5	0.6029	0.1691	0.7740	0.1946	1.1410	0.3429
3	0.5603	0.1185	0.7539	0.1408	1.1780	0.2794
3.5	0.5301	0.0879	0.7406	0.1071	1.2096	0.2346
4	0.5076	0.0678	0.7313	0.0844	1.2368	0.2012
4.5	0.4902	0.0539	0.7245	0.0685	1.2604	0.1755
5	0.4764	0.0439	0.7194	0.0567	1.2812	0.1546

The Shannon entropy of a random variable X is defined by $E\left[-logf(x)\right]$.
First derive the Rényi entropy for the corresponding to G(α,θ) distribution. We have

$$\begin{aligned}\int_0^\infty f^\gamma(x) &= \left(\frac{\alpha}{\theta^{(\alpha-1)}}\right)^\gamma \int_0^{-ln(1-\theta)} e^{-\gamma x}(1-e^{-x})^{\gamma(\alpha-1)}dx \\ &+ \left(\frac{\alpha}{(1-\theta)^{(\alpha-1)}}\right)^\gamma \int_{-ln(1-\theta)}^\infty e^{-\alpha\gamma x}dx \end{aligned} \tag{7}$$

$$= \left(\frac{\alpha}{\theta^{(\alpha-1)}}\right)^\gamma \int_0^{-ln(1-\theta)} e^{-\gamma x}(1-e^{-x})^{\gamma(\alpha-1)}dx + \frac{\alpha^{\gamma-1}(1-\theta)^\gamma}{\gamma}$$

But, the above expresson is difficult to be express in an explicit form.
Consider the Shannon entropy, we have

$$\begin{aligned} H(X) &= E(-logf(x)) \\ &= \int_0^\infty -logf(x)f(x)dx \\ &= \int_0^{-ln(1-\theta)} -logf(x)f(x)dx + \int_{-ln(1-\theta)}^\infty -logf(x)f(x)dx \\ &= -2log(\alpha) + (\alpha+1)E(X) - (\alpha-1)\left[\theta\left(log(\theta) - \frac{1}{\alpha}\right) - 2log(\theta)\right]. \end{aligned} \tag{8}$$

Order statistics refers the ranking a sample from a distribution. Let $X_1, X_2, .., X_k$ be k independent and identically distributed random variables, each with cumulative distribution function F(x) given in equation (3). We denote $X_{(r)}$ as the r^{th} order statistic, r=1,2,...,k. Then $f_r(x)$, the probability density function of $X_{(r)}$ for G(α,θ) distribution is given by

$$f_r(x) = \frac{1}{B(r,k-r+1)}F^{r-1}(x)(1-F(x))^{k-r}f(x) \tag{9}$$

$$= \frac{k!}{(r-1)!(k-r)!}\begin{cases} \frac{\alpha e^{-x}(1-e^{-x})^{\alpha r-1}\left[\theta^{(\alpha-1)}-(1-e^{-x})^\alpha\right]^{k-r}}{\theta^{(\alpha-1)k}}, 0 < x < -ln(1-\theta) \\ \frac{\alpha e^{-\alpha x(k-r+1)}\left[(1-\theta)^{\alpha-1}-e^{-\alpha x}\right]^{r-1}}{(1-\theta)^{(\alpha-1)k}}, -ln(1-\theta) < x < \infty. \end{cases}$$

3. Reliability measures of the G(α, θ) distribution

From equation (3), the ***reliability function*** is

$$R(t) = 1 - F(t) \tag{10}$$

$$= 1 - \begin{cases} \theta \left(\frac{1-e^{x}}{\theta}\right)^{\alpha}, & if \quad 0 < x \leq -ln(1-\theta) \\ 1-(1-\theta)\frac{e^{-\alpha x}}{(1-\theta)^{\alpha}}, & if \quad -ln(1-\theta) \leq x < \infty \end{cases}$$

$$= \begin{cases} 1 - \frac{(1-e^{-x})^{\alpha}}{\theta^{\alpha-1}} & 0 < x \leq -ln(1-\theta) \\ \frac{e^{-\alpha x}}{(1-\theta)^{\alpha-1}} & -ln(1-\theta) \leq x < \infty. \end{cases}$$

To identify the applicability of the introduced distribution in reliability studies, we have to famililarize with the shape characteristics of the reliability and hazard curves with the changes in parameter values. For the G(α,θ) distribution the ***hazard function*** is given by

$$h(x) = \frac{f(x)}{R(x)} \tag{11}$$

$$= \begin{cases} \frac{\alpha e^{-x}(1-e^{-x})^{\alpha-1}}{\theta^{\alpha-1}-(1-e^{-x})^{\alpha}} & 0 < x \leq -ln(1-\theta) \\ \alpha & -ln(1-\theta) \leq x < \infty. \end{cases}$$

The plot of hazard function is given in Figure (2). For any fixed θ, The G(α,θ) distribution has an increasing hazard function for $\alpha > 1$ and it has decreasing hazard function for $\alpha < 1$. For α=1 the hazard function becomes 1, independent of x. These results are not very difficult to prove, it simply follows from the fact that the G(α,θ) distribution has a log-concave density for $\alpha > 1$ and it is log-convex for $\alpha \leq 1$. The hazard function of the G(α,θ) distribution behaves exactly the same way as the hazard functions of the Weibull distribution distribution.

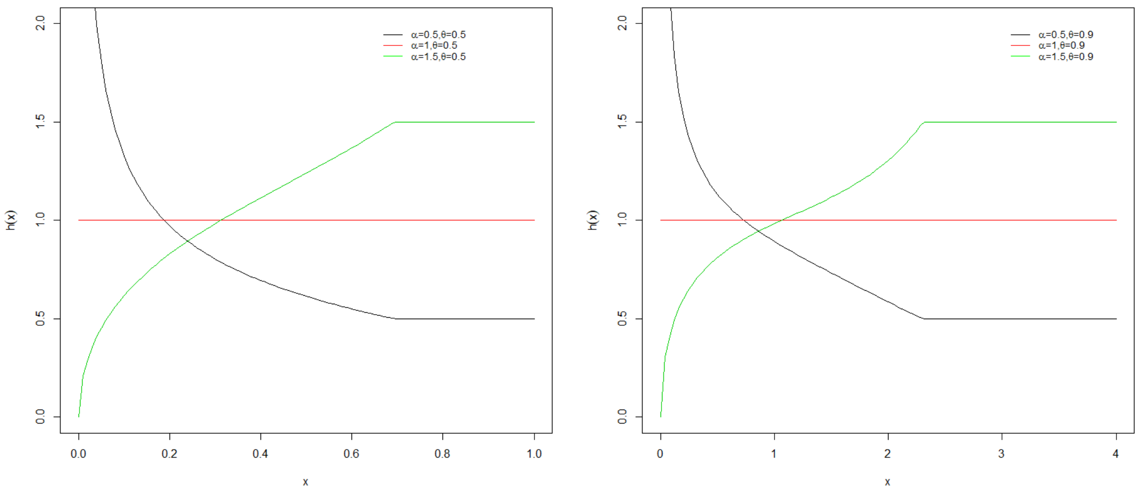

Figure 2: *Plot of hazard function for different values of parameters α and θ*

Reversed hazard function Reversed hazard has been using for the analysis of right- truncated and left-censored data and is applicable in such areas as Forensic Science. The formula of reversed hazard rate of a random life is defined as the ratio between the life probability density to its distribution function. The reversed hazard function for the G(α,θ) distribution is given by,

$$r(x, \alpha, \theta) = \begin{cases} \frac{\alpha}{e^{x}-1}, & 0 < x \leq -ln(1-\theta) \\ \frac{\alpha}{e^{\alpha x}(1-\theta)^{(\alpha-1)}-1}, & -ln(1-\theta) \leq x < \infty. \end{cases} \tag{12}$$

It has observed that, for all values of α and θ, the reversed hazard function is a decreasing function of x. Further, there is no non negative random variable have an increasing reversed hazard rate. Distributions showing the same behaviour in the reverse hazard function are weibull, lognormal.

Elasticity of a distribution express the change that, the distribution function undergoes when faced with the variation in the random variable. It is one of the most important concept in economics theory. In economics, Elasticity measures how sensitive an *output* variable is to change in an *input* variable. This classical concept of elasticity to an economic function can be extended to the cumulative distribution function of a random variable. The elasticity function e(x) is defined as

$$e(x) = \frac{dlnF(x)}{dlnx} = \frac{F'(x)/F(x)}{1/|x|} = \frac{|x|f(x)}{F(x)}, F(x) > 0. \tag{13}$$

In the case of G(α,θ) distribution, Elasticity function is given by

$$\begin{aligned} e(x) &= |x| r(x, \alpha, t) \qquad (14) \\ &= |x| \begin{cases} \frac{\alpha}{e^x - 1}, & 0 < x \leq -ln(1-\theta) \\ \frac{\alpha}{e^{\alpha x}(1-\theta)^{(\alpha-1)} - 1}, & -ln(1-\theta) \leq x < \infty, \end{cases} \end{aligned}$$

which shows the close relationship that exists between the reversed hazard function and elasticity.

Mean residual life function (MRLF) is the average lifetime remaining for a component or an individual which had survived at time. For a continuous, non-negative random variable X, representing lifetime of a component or an individual is the residual life random variable at age t, denoted by $X_t = X - t|X > t$, is simply the remaining lifetime beyond that age. Then the MRLF denoted by μ(t) is defined as

$$\begin{aligned} \mu(t) &= E(X - t|X > t) \qquad (15) \\ &= \frac{1}{\overline{F}(t)} \int_t^\infty \overline{F}(x) dx \\ &= \frac{1}{R(t)} \int_t^\infty R(x) dx \end{aligned}$$

Then for G(α,θ) distribution we have

$$\begin{aligned} \mu(t) &= \frac{\theta^{(\alpha-1)}}{\theta^{(\alpha-1)} - (1-e^{-t})^\alpha} \int_t^{-ln(1-\theta)} 1 - \frac{(1-e^{-x})^\alpha}{\theta^{(\alpha-1)}} dx \\ &+ \frac{(1-\theta)^{(\alpha-1)}}{e^{-\alpha t}} \int_{-ln(1-\theta)}^\infty \frac{e^{-\alpha t}}{(1-\theta)^{(\alpha-1)}} \end{aligned}$$

substituting for $(1 - e^{-x})$ by u in first part of the integral we get

$$\mu(t) = \frac{1}{\theta^{(\alpha-1)} - (1-e^{-t})^\alpha} \int_{1-e^{-t}}^\theta u^\alpha (1-u)^{-1} du + \frac{1}{e^{-\alpha t}} \left[\frac{(1-\theta)^\alpha}{\alpha} \right]. \tag{16}$$

But, the above expression does not have an explicit form and thus we need to calculate it numerically.

4. Estimation of Parameters; Maximum Likelihood Estimation

The proposed derivation of the maximum likelihood estimation procedure of a G(α,θ) distribution is quite instructive. Let for a sample $\underline{X} = (X_1, X_2, .., X_s)$, the order statistics be $X_{(1)} < X_{(2)} < .. < X_{(s)}$. By definition, the likelihood function for $\underline{X}$ is

$$L(\underline{X}; \theta, \alpha) = \alpha^s \prod_{i=1}^s e^{-X_{(i)}} \left[\frac{\prod_{i=1}^r (1 - e^{-X(i)}) \prod_{i=r+1}^s e^{-X_{(i)}}}{\theta^r (1-\theta)^{s-r}} \right]^{\alpha-1}, \tag{17}$$

where, $X_{(r)} \leq -ln(1-\theta) < X_{(r+1)}$, with $X_{(0)} \equiv 0$
Then, the MLE estimators of the parameters are given by

$$\hat{\theta} = 1 - e^{-X_{(\hat{r})}} \tag{18}$$

$$\hat{\alpha} = -\frac{s}{logM(\hat{r})}, \tag{19}$$

where, $\hat{r}$ = arg $max_{r\in\{1,2,..,s\}} M(r)$ and

$$M(r) = \prod_{i=1}^{r-1} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r)}}} \prod_{i=r+1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(r)}}}. \tag{20}$$

To maximize the likelihood (17), we set

$$\max_{\alpha>0,0<\theta<1} L(\underline{X};\theta,\alpha) = \max_{\alpha>0}\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right], \tag{21}$$

where, $\hat{M}$ is given by

$$\hat{M} = \max_{0<\theta<1}\left[\frac{\prod_{i=1}^{r}(1-e^{-X(i)})\prod_{i=r+1}^{s} e^{-X_{(i)}}}{\theta^r(1-\theta)^{s-r}}\right], \tag{22}$$

and as above $X_{(r)} \leq -ln(1-\theta) < X_{(r+1)}$, with $X_{(0)} \equiv 0$.
Using the properties of Pitman family, $X_{(\hat{r})}$ would be the estimate of $\hat{\hat{\theta}}$= -ln(1-θ). Then by inverting we have,

$$\hat{\theta} = 1 - e^{-X_{(\hat{r})}}$$

Now,

$$log\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right] = slog(\alpha) - \sum_{i=1}^{s} X_{(i)} + (\alpha-1)log(\hat{M}) \tag{23}$$

and

$$\frac{\partial}{\partial\alpha} log\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right] = \frac{s}{\alpha} + log(\hat{M}) \tag{24}$$

equating to 0 yields,

$$\hat{\alpha} = -\frac{s}{log(\hat{M})}.$$

From equation (24), it follows that

$$\frac{\partial}{\partial\alpha} log\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right] > 0 \Leftrightarrow \hat{\alpha} < -\frac{s}{log(\hat{M})}. \tag{25}$$

Hence, $\hat{\alpha}$ corresponds to a global maximum of both (23) and (21). Note that for $i < r$, it follows that $0 < \frac{1-e^{-X_{(i)}}}{\theta} < 1$ and for $i > r$, it follows that $0 < \frac{e^{-X_{(i)}}}{1-\theta} < 1$. Hence $0 < \hat{M} < 1$ and thus $\hat{\alpha} > 0$. Using equation (22), we may write $\hat{M} = \max_{r\in\{0,..,s\}} H(r)$, where,

$$H(r) = \max_{X_{(r)}\leq\hat{\hat{\theta}}\leq X_{(r+1)}}\left[\frac{\prod_{i=1}^{r}(1-e^{-X(i)})\prod_{i=r+1}^{s} e^{-X_{(i)}}}{\theta^r(1-\theta)^{s-r}}\right]. \tag{26}$$

We shall discuss three cases: $r \in \{1,2,...,s-1\}$, r = 0 and r = s.

Case 1: $r \in \{1,2,...,s-1\}$: Here, $X_{(r)} \leq \hat{\hat{\theta}} \leq X_{(r+1)}$. From (26),

$$H(r) = \max_{r'\in\{r,r+1\}} \prod_{i=1}^{r'-1} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r')}}} \prod_{i=r'+1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(r')}}} \tag{27}$$

Case 2: r = 0 : Here, $0 \leq \hat{\hat{\theta}} \leq X_{(1)}$. From (26) it follows that in this case

$$H(0) = \max_{0 \leq \hat{\hat{\theta}} \leq X_{(1)}} \left[\prod_{i=1}^{s} \frac{e^{-X(i)}}{1-\theta} \right] \tag{28}$$

Hence,

$$H(0) = \prod_{i=1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(1)}}} = \prod_{i=2}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(1)}}} \tag{29}$$

Case 3: r = s : Here, $X_{(s)} \leq \hat{\hat{\theta}} < \infty$. From (26) it follows that in this case

$$H(s) = \max_{X_{(s)} \leq \hat{\hat{\theta}} < \infty} \left[\frac{1 - e^{X_{(i)}}}{\theta} \right] \tag{30}$$

Hence

$$H(s) = \prod_{i=1}^{s} \frac{1 - e^{X_{(i)}}}{1 - e^{-X_{(s)}}} = \prod_{i=1}^{s-1} \frac{1 - e^{X_{(i)}}}{1 - e^{-X_{(s)}}} \tag{31}$$

From (27), (29)and (31) we obtain that $\hat{M} = \max_{r \in \{1,..s\}} M(r)$, where

$$M(r) = \prod_{i=1}^{r-1} \frac{1 - e^{-X_{(i)}}}{1 - e^{-X_{(r)}}} \prod_{i=r+1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(r)}}} \tag{32}$$

Note that $\hat{M}$ attained at $\hat{\hat{\theta}} = X_{(\hat{r})}$ where $\hat{r}$ = arg $\max_{r \in \{1,2,...,s-1\}} M(r)$.

The estimates given in (18) and (19) are quite intuitive. In particular the estimator of the parameter θ is in term of a specific order statistic. Note that the approach for determining the MLE estimate $\hat{\theta}$ for the G(α,θ) distribution is similar to the approach for determining the MLE estimate $\hat{\theta}$ for a triangular distribution and STSP (θ, n) distribution (see René Van Dorp and Kotz [13]). In order to find the estimate, we use a quite different method.

Consider the matrix **A** = $[a_{i,r}]$ where

$$a_{i,r} = \begin{cases} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r)}}}, & \text{if} \quad i < r \\ \\ \frac{e^{-X_{(i)}}}{e^{-X_{(r)}}}, & \text{if} \quad i \geq r. \end{cases} \tag{33}$$

Then **A** will be a real matrix with unit diagonal entries. Then, we find the product of the matrix elements in the r^{th} column which are equal to the values of M(r) given by equation (32). identify the maximum value of M(r) and the corresponding r^{th} order statistic $X_{(r)}$ is taken as the estimate of $\hat{\hat{\theta}}$ and by inverting we get the maximum likelihood estimate of the parameter θ. The maximum likelihood estimate of second parameter α can be evaluated using the equation (19), where s denotes the total number of observations.

5. Simulation

To verify the estimation procedure, We have considered a simulation study. We generated samples of different sizes using the quantile function of the proposed distribution and the parameters are estimated with the above discussed procedure. It is repeated 500 times and the mean of estimates are taken as the estimate of the parameters. Table (2) gives the estimates of the parameters, the values in brackets indicates the mean squared error. Note that, as the sample size increases, the estimate becomes more close to the true value of the parameter.

Table 2: *Estimated values of the parameters and mean squared error.*

	θ			α		
Sample size	0.3	0.5	0.9	1.5	2	3
100	0.3107 (0.0135)	0.5036 (0.0031)	0.8985 (0.0003)	1.5469 (0.0229)	2.0444 (0.0416)	3.0467 (0.0873)
150	0.3081 (0.0085)	0.5012 (0.0022)	0.8980 (0.0003)	1.5354 (0.0162)	2.0326 (0.0285)	3.0422 (0.0636)
200	0.3072 (0.0056)	0.5007 (0.0014)	0.8982 (0.0001)	1.5186 (0.0104)	2.0161 (0.0196)	3.0323 (0.0442)
500	0.3018 (0.0021)	0.5004 (0.0005)	0.8998 (5.213×10^{-05})	1.5093 (0.0048)	2.009 (0.0091)	3.0228 (0.0190)

6. Data Analysis

In this section, we provide an application of the G(α,θ) distribution by modeling two real data sets.

Data set I: The first data set represents the life of fatigue fracture of Kevlar 373/epoxy that are subject to constant pressure at the 90 percentage stress level until all had failed, so we have complete data with the exact times of failure. The same datasets were used in the literature by Alizadeh et. all [1]. The estimation of the parameters is done using the maximum likelihood method as discussed in the previous section. Firstly, we identified the matrix **A** with entries defined as in (33). Then we find the product of the matrix elements in the rth column which gives the values of M(r) given by equation (32). The maximum value of M(r) is identified and the corresponding r^{th} order statistic is taken as the maximum likelihood estimate of $\hat{\hat{\theta}}$ and by inverting we obtained the estimate of the parameter θ. Similarly, the maximum likelihood estimate $\hat{\alpha}$ can be obtained using equation (19). Here, we obtained the estimate of θ as 0.9998879 and the estimate of α as 2.4003. To verify, goodness of fit of the proposed distribution, we have performed Kolmogorov-Smirnov test and the corresponding p-value is 0.07278, which indicates that the G(α,θ) is a suitable model for the data. The histogram of the data together with the fitted probability density curve is given in Figure (3).

Data set II: The second real data set represent the failure times of 84 aircraft windshield. This data is taken from Ijaz et all. [6]. The estimation was carried out as done in the previous data set. The value of M(r) and corresponding order statistic are noted and we obtain the estimte of θ and α. The parameters for the data set II are estimated as $\hat{\theta}$ = 0.9961512, $\hat{\alpha}$ = 7.827007. The p-value for Kolmogorov-Smirnov test is 0.1958. The histogram of the data set II with fitted probability curve is given in Figure (4).

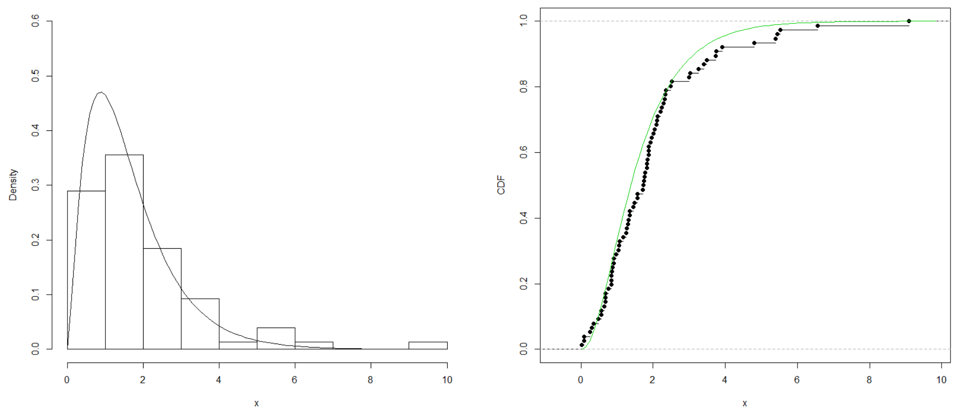

Figure 3: *(a) Histogram and fitted G(α,θ) probability density function for dataset I. (b) Theoretical and fitted distribution function for the dataset I.*

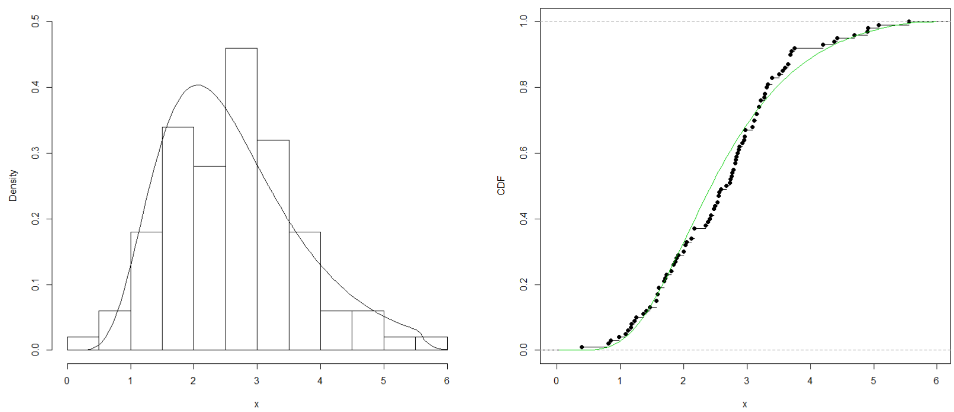

Figure 4: *(a) Histogram and fitted G(α,θ) probability density function for dataset II.*
(b) Theoretical and fitted distribution function for the dataset II.

References

[1] Alizadeh, M. Merovci, F. and Hamedani, G. G. (2000). Generalized Transmuted Family of Distributions: Properties and Applications. *Hacetteppe University Bulletin of Natural science and Engineering Series B: Mathematics and Statistics*, 46: 645-668.

[2] Bourguignon, M, Silva and R B, Cordeiro, G M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12: 53–68.

[3] Cordeiro, G M and Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81: 883–898.

[4] Eugene, N. Lee, C and Famoye, F. (2002). Beta-normal Distribution and its applications. *Communications in Statistics – Theory and Methods*, 31: 497-512.

[5] Gupta, R. D. and Kundu, D. (1999). Generalized Exponential Distributions. *Australian and New Zealand Journal of Statistics*, 41(2): 173-188.

[6] Ijaz,M. , Asim, S. M. and Alamgir. (2019). Lomax exponential distribution with an application to real-life data, *PLOS ONE*, 14(12): 1-16.

[7] Jayakumar, K and Mathew, T. (2008). On a generalization to Marshall-Olkin scheme and its application to Burr type XII distribution. *Statistical Papers*, 49: 421–439.

[8] Lai, C.D., Zhang, L. and Xie, M. (2004). Mean Residual Life and Other Properties of Weibull Related Bathtub Shape Failure Rate Distribution. *International Journal of Reliability, Quality and Safety Engineering*, 11(2): 113-132.

[9] Marshall, A W and Olkin, I. A. (1997). new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84: 641–652.

[10] Nadarajah, S and S. Kotz. (2004). The beta Gumbel distribution. *Mathematical Problems in Engineering*, 10: 323–332.

[11] Nadarajah, S and A. K. Gupta. (2004). The beta Frechet distribution. *Far East Journal of Theoretical Statistics*, 14: 15–24.

[12] Nadarajah S and S. Kotz. (2006). The beta exponential distribution. *Reliability Engineering and System Safety* 91: 689–697.

[13] René Van Dorp, J. and Kotz, S. (2002). The Standard Two-Sided Power Distribution and it's Properties. *The American Statistician*, 56(2): 90-99.

[14] Silva, R B., Bourguignon, M., Dias, C R B, Cordeiro, G M. (2013). The compound class of extended Weibull power series distributions. *Computational Statistics and Data Analysis*, 58: 352-367.

[15] Zogarafos, K and Balakrishnan, N. (2009). On the families of beta-and generalized gamma-generated distribution and associated inference, *Statistical Methodology*, 6: 344–362.