Journal de la Société Française de Statistique
Vol. 159 No. 3 (2018)

The log-xgamma distribution with inference and application

Titre: La distribution log-gamma : inférence et application

Emrah Altun[1] and GG Hamedani[2]

Abstract: In this paper, we introduce a new one-parameter distribution, called *log-xgamma* distribution, defined on the unit interval. Some of the statistical properties of the proposed distribution including moments, the incomplete moments and mean residual life function are obtained. Some useful characterization results of proposed distribution are presented. The maximum likelihood method, method of moments and least square estimation method are used to estimate the unknown parameter of the proposed model and finite sample performance of estimation methods are evaluated by means of Monte-Carlo simulation study. An application to the real data set is given to demonstrate the usefulness of the proposed distribution against the beta, the Kumaraswamy and the Topp-Leone distributions.

Résumé : Nous introduisons une nouvelle distribution à un paramètre sur l'intervalle [0, 1]. Ces principales caractéristiques (moments, moments censurés, fonction de survie) sont données, ainsi que d'autres caractérisations utiles. Les méthodes du maximum de vraisemblance, des moments et des moindres carrés sont présentés pour l'estimation de son paramètre. Les performances de ces estimateurs sont évaluées par des simulations de Monte-Carlo pour des échantillons de taille réduite. Une application à des données réelles est réalisée pour montrer l'intérêt de cette distribution par rapport aux distributions beta, de Kumaraswamy et de Topp-Leone.

Keywords: Bounded distributions, Xgamma distribution, Characterization, Simulation
Mots-clés : Distributions bornés, Distribution Xgamma, Caractérisation, Simulation
AMS 2000 subject classifications: 62E10, 62N05

1. Introduction

In the last decade, researchers have shown a great interest in introducing new extended distributions by adding extra shape parameters to the baseline distributions. The main motivation of these studies is to increase the modeling ability of the distributions and to open the new opportunities to model different characteristics of the data sets. Generally, researchers have paid attention to unbounded support. However, there are many real-life situations in which the observations can take values only in a bounded range, such as percentages, and proportions. It is possible to face this situation in economic variables such as industry market shares and the proportion of income spent on non-durable consumptions (see, Papke and Wooldridge, 1996 for details).

The most well-known distribution defined on the unit interval is the Beta distribution which is widely used in various areas of sciences such as economics, biology and medical sciences since it has great flexibility regarding the shapes of the hazard rate function. The main drawback

[1] Department of Statistics, Bartin University, Bartin 74100, Turkey.
E-mail: emrahaltun@bartin.edu.tr
[2] Department of Mathematics, Statistics and Computer Science, Marquette University, USA.
E-mail: g.hamedani@mu.edu

of the Beta distribution is that its cumulative distribution function (cdf) cannot be expressed in closed form and contains the beta function. After the pioneer work of Nadarajah and Kotz (2003), Topp-Leone distribution, proposed by Topp and Leone (1955), has increased its popularity. The Topp-Leone distribution has some advantages over the Beta distribution. The most important advantage is that its distribution function is simple and does not contain any special function such as beta and gamma functions. The other widely used distribution on the unit interval is the Kumaraswamy distribution, introduced by Kumaraswamy (1980). The Kumaraswamy distribution has increased its popularity after the work of Cordeiro and Castro (2011). More recently, the unit-Birnbaum-Saunders distribution was introduced by Mazucheli et al. (2018). The unit-Birnbaum-Saunders distribution can be viewed as an alternative unit distribution to the two parameter Beta and Kumaraswamy distributions.

The goal of this paper is to introduce an alternative distribution for modeling data sets on the interval $[0,1]$. To achieve this goal, xgamma distribution is used to generate a new distribution defined on the unit interval. The proposed distribution has important advantages over the well-known distributions defined on unit interval such as Beta, Kumaraswamy and Topp-Leone distributions. The superiority of the proposed distribution comes from its simple form and its flexibility via hazard rate function. The statistical properties of the log-xgamma distribution can be obtained in explicit forms for its probability density and cumulative distribution functions, moments, skewness and kurtosis measures. Since the log-xgamma distribution has only one parameter, the estimation of the model parameter is easier than those of Beta and Kumaraswamy distributions.

The rest of the paper is organized as follows: In Section 2 we obtain the mathematical properties of the proposed distribution comprehensively. Section 3 provides certain characterizations of the proposed distribution. In Section 4, the parameter estimation of the model is discussed via maximum likelihood method, method of moments and least square estimation method. In Section 5, the Monte-Carlo simulation study is conducted to evaluate the finite sample performance of the parameter estimation methods. In Section 6, two real data sets are analysed to demonstrate the flexibility of the log-xgamma distribution against the well-known distributions defined on the unit interval. Conclusion is given in Section 7.

2. The log-xgamma distribution

Let the random variable X follow a Lindley distribution with probability density function (pdf)

$$f(x;\theta)=\frac{\theta^2}{1+\theta}(1+x)\exp(-\theta x),x>0, \tag{1}$$

where $\theta>0$ is the scale parameter. As seen from (1), the Lindley distribution is a mixture of the exponential (θ) and gamma $(2,\theta)$ distributions. The corresponding cdf is

$$F(x)=1-\frac{\theta+1+\theta x}{\theta+1}\exp(-\theta x),x\geq 0. \tag{2}$$

Most of the statistical properties of the Lindley distribution such as moments, stochastic ordering, entropies were obtained by Ghitany et al. (2008). Recently, Sen et al. (2016) introduced the xgamma distribution following the idea of Lindley distribution. The extensions of xgamma distribution was studied by Sen and Chandra (2017) and Sen et al. (2017). The pdf of xgamma distribution is given by

$$f(x;\theta)=\frac{\theta^2}{1+\theta}\left(1+\frac{\theta}{2}x^2\right)\exp(-\theta x),x>0,\theta>0. \tag{3}$$

As seen from (3), the xgamma distribution is a mixture of exponential (θ) and gamma $(3,\theta)$ distributions. The corresponding cdf is

$$F(x)=1-\frac{1+\theta+\theta x+\frac{\theta^2x^2}{2}}{\theta+1}\exp(-\theta x),x\geq 0. \tag{4}$$

Proposition 1. *Let random variable $Y=\exp(-X)$, then the pdf of Y is*

$$f(y;\theta)=\frac{\theta^2}{1+\theta}\left(1+\frac{\theta}{2}\ln(y)^2\right)y^{\theta-1},0<y<1, \tag{5}$$

where $\theta>0$ is the shape parameter. Hereafter, the random variable Y is denoted by $Y\sim$ log-xgamma (θ). The corresponding cdf is

$$F(y)=y^{\theta}(\theta+1)^{-1}\left(1+\theta-\theta\ln(y)+\frac{\theta^2\ln(y)^2}{2}\right),0\leq y\leq 1. \tag{6}$$

Figure 1 displays the possible pdf shapes of the log-xgamma distribution. The log-xgamma distribution can be a good choice to model extremely left or right skewed data sets.

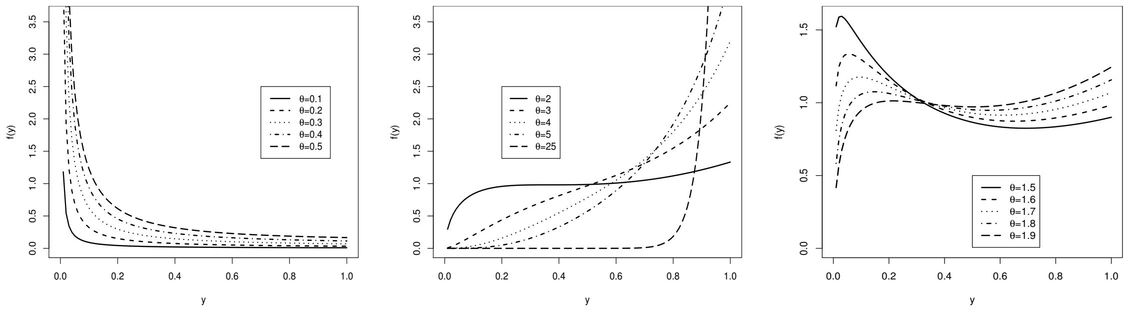

FIGURE 1. *The pdf plots of log-xgamma distribution for selected parameter values.*

The survival function (sf) of Y is

$$S(y)=1-y^{\theta}(\theta+1)^{-1}\left(1+\theta-\theta\ln(y)+\frac{\theta^2\ln(y)^2}{2}\right),0\leq y\leq 1, \tag{7}$$

and the hazard rate function (hrf) of Y is

$$h(y)=\frac{\theta^2\left(1+\frac{\theta}{2}\ln(y)^2\right)y^{\theta-1}}{(1+\theta)\left\{1-y^{\theta}(\theta+1)^{-1}\left(1+\theta-\theta\ln(y)+\frac{\theta^2\ln(y)^2}{2}\right)\right\}},0<y<1. \tag{8}$$

Figure 2 displays the possible hrf shapes of the log-xgamma distribution. The log-xgamma distribution has the following hrf shapes: increasing and bathtub.

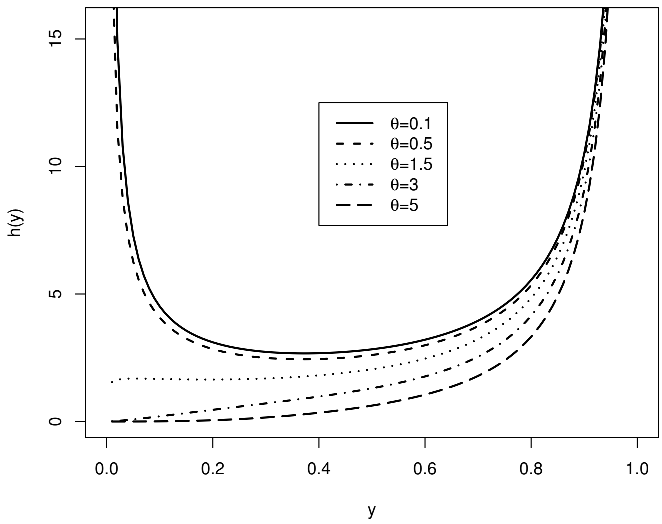

FIGURE 2. *The hrf plots of log-xgamma distribution for selected parameter values.*

Proposition 2. *Let $F(y)$ denote the cdf of log-xgamma distribution. The following algorithm can be used to generate random observations from $Y \sim$ log-xgamma (θ).*

1. *Generate $u \sim uniform(0,1)$,*
2. *Solve the following non-linear equation for a given parameter value,*

$$y^{\theta}(\theta+1)^{-1}\left(1+\theta-\theta\ln(y)+\frac{\theta^2\ln(y)^2}{2}\right)-u=0. \tag{9}$$

The ***uniroot*** *function of R software can be used to solve non-linear equation given in step 2. Note that the above algorithm is a very well-known simulation method, called as inverse transform method.*

Proposition 3. *Let the random variable Y follow a log-xgamma disribution. The raw moments of Y are given by*

$$E(Y^r)=\frac{\theta^2\left[\theta^2+(2r+1)\theta+r^2\right]}{(\theta+1)(\theta+r)^3}. \tag{10}$$

The first four raw moments of log-xgamma distribution are given by

$$
\begin{aligned}
E(Y) = \mu &= \frac{\theta^2\left(\theta^2+3\theta+1\right)}{(\theta+1)^4},\\
E\left(Y^2\right) &= \frac{\theta^2\left(\theta^2+5\theta+4\right)}{(\theta+1)(\theta+2)^3},\\
E\left(Y^3\right) &= \frac{\theta^2\left(\theta^2+7\theta+9\right)}{(\theta+1)(\theta+3)^3},\\
E\left(Y^4\right) &= \frac{\theta^2\left(\theta^2+9\theta+16\right)}{(\theta+1)(\theta+4)^3}.
\end{aligned} \tag{11}
$$

The variance of the log-xgamma distribution is

$$
\sigma^2 = \frac{\theta^2\left(\theta^2+5\theta+4\right)}{(\theta+1)(\theta+2)^3} - \frac{\theta^4\left(\theta^2+3\theta+1\right)^2}{(\theta+1)^8}. \tag{12}
$$

Let γ_1 and γ_2 denote the skewness and kurtosis values of the the log-xgamma distribution, respectively. The measures γ_1 and γ_2 can be obtained from

$$
\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E\left(X^3\right)-3\mu\sigma^2-\mu^3}{\sigma^3}
$$

$$
\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E\left(X^4\right)-4\mu E\left(X^3\right)+6\mu^2\sigma^2+3\mu^4}{\sigma^4}
$$

Figure 3 displays the skewness and kurtosis values of log-xgamma distribution. As seen from Figure 3, the log-xgamma distribution can be symmetric, left or right skewed.

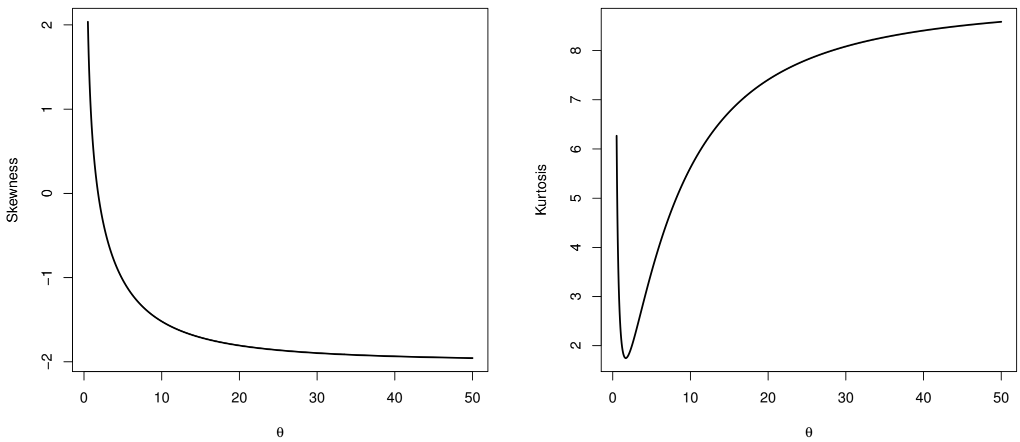

FIGURE 3. *The skewness and kurtosis plots of log-xgamma distribution.*

Proposition 4. *Let the random variable Y follow a log-xgamma disribution. The incomplete moments of Y are given by*

$$T_r(t) = E\left[Y^r \mathrm{I}_{\{Y<t\}}\right] = \frac{\theta^2}{1+\theta} \int_0^t \left(1 + \frac{\theta}{2} \ln(y)^2\right) y^{\theta+r-1} dy. \tag{13}$$

The above integration can not be carried out analytically. In particular, for $r = 1$, $T_1(t)$ *is*

$$T_1(t) = \frac{t^{\theta+1}}{2(\theta+1)} + \frac{t^{\theta+1}\left[\ln(t)^2(\theta+1)^2 - 2\ln(t)(\theta+1) + 2\right]}{4(\theta+1)^3}. \tag{14}$$

Proposition 5. *Let the random variable Y follow a log-xgamma disribution. The mean residual life function of Y is given by*

$$m(t) = E(Y - t|Y > t) = \frac{1}{1 - F(t)} \int_t^1 [1 - F(y)]\, dy. \tag{15}$$

The analytical solution of the above integral is

$$m(t) = \frac{1}{1 - t^{\theta}(\theta+1)^{-1}\left(1 + \theta - \theta\ln(t) + \frac{\theta^2 \ln(t)^2}{2}\right)} \left[\begin{array}{l} 1 - t + \frac{t^{\theta+1}-1}{(\theta+1)^2} + \frac{\theta\left(t^{\theta+1}-1\right)}{(\theta+1)^2} \\ - \frac{\theta\left(\frac{1}{(\theta+1)^2} + \frac{t^{\theta+1}\left(\log(t) - \frac{1}{\theta+1}\right)}{\theta+1}\right)}{\theta+1} \\ - \frac{\theta^2\left(\frac{2}{(\theta+1)^3} - \frac{t^{\theta+1}\left(\log(t)^2(\theta+1)^2 - 2\log(t)(\theta+1) + 2\right)}{(\theta+1)^3}\right)}{2(\theta+1)} \end{array}\right]. \tag{16}$$

3. Characterizations

This section deals with the characterizations of the log-xGamma distribution in two directions: (i) based on the ratio of two truncated moments and (ii) in terms of reverse (reversed) hazard function. Note that for the characterization (i) the cdf need not have a closed form and depends on the solution of a first order differential equation, which provides a bridge between probability and differential equation. We like to also mention that characterization (i) is stable in the sense of weak convergence (Glänzel , 1990). We present our characterizations $(i) - (ii)$ in two subsections.

3.1. Characterizations based on two truncated moments

In this subsection we present the characterizations of log-xgamma distribution based on the ratio of two truncated moments. Our first characterization employs a theorem due to Glänzel (1987), see Theorem 1 of Appendix A . The result, however, holds also when the interval H is not closed, since the condition of the Theorem is on the interior of H.

Proposition 6. *Let $Y:\Omega\to(0,1)$ be a continuous random variable and let* $q_1(y)=\left\{1+\frac{\theta(\ln y)^2}{2}\right\}^{-1}$ *and* $q_2(y)=q_1(y)\,y^{\theta}$ *for* $y\in(0,1)$. *The random variable Y has pdf* (5) *if and only if the function* ξ *defined in Theorem 1 is of the form*

$$\xi(y)=\frac{1}{2}\left(1+y^{\theta}\right),\quad y\in(0,1).$$

Proof. Suppose the random variable Y has pdf (5), then

$$(1-F(y))\,E\left[q_1(Y)\mid Y\geq y\right]=\frac{\theta}{\theta+1}\left(1-y^{\theta}\right),\ \ y\in(0,1)$$

and

$$(1-F(y))\,E\left[q_2(Y)\mid Y\geq y\right]=\frac{\theta}{2(\theta+1)}\left(1-y^{2\theta}\right),\ \ y\in(0,1).$$

Further,

$$\xi(y)\,q_1(y)-q_2(y)=\frac{q_1(y)}{2}\left(1-y^{\theta}\right)>0,\quad for\ \ y\in(0,1).$$

Conversely, if ξ is of the above form, then

$$s'(y)=\frac{\xi'(y)\,q_1(y)}{\xi(y)\,q_1(y)-q_2(y)}=\frac{\theta y^{\theta-1}}{1-y^{\theta}},\quad y\in(0,1),$$

and consequently

$$s(y)=-\ln\left(1-y^{\theta}\right),\quad y\in(0,1).$$

Now, according to Theorem 1, Y has density (5).

□

Corollary 1. *Let $Y:\Omega\to(0,1)$ be a continuous random variable and let $q_1(y)$ be as in Proposition 6. The random variable Y has pdf (5) if and only if there exist functions q_2 and ξ defined in Theorem 1 satisfying the following differential equation*

$$\frac{\xi'(y)\,q_1(y)}{\xi(y)\,q_1(y)-q_2(y)}=\frac{\theta y^{\theta-1}}{1-y^{\theta}},\quad y\in(0,1).$$

Corollary 2. *The general solution of the differential equation in Corollary 1 is*

$$\xi(y)=\left(1-y^{\theta}\right)^{-1}\left[-\int\theta x^{\theta-1}\left(q_1(x)\right)^{-1}q_2(x)\,dx+D\right],$$

where D is a constant. We like to point out that one set of functions satisfying the above differential equation is given in Proposition 3.1 with $D=\frac{1}{2}$. Clearly, there are other triplets (q_1,q_2,ξ) which satisfy conditions of Theorem 1.

3.2. Characterization in terms of the reverse hazard function

The reverse hazard function, r_F, of a twice differentiable distribution function, F , is defined as

$$r_F\left(y\right)=\frac{f\left(y\right)}{F\left(y\right)},\ \ y\in support\ of\ F.$$

In this subsection we present a characterization of log-xgamma distribution in terms of the reverse hazard function.

Proposition 7. *Let $Y:\Omega\rightarrow(0,1)$ be a continuous random variable. The random variable Y has pdf (5), if and only if its reverse hazard function $r_F\left(y\right)$ satisfies the following differential equation*

$$r_F'\left(y\right)+y^{-1}r_F\left(y\right)=\theta^2y^{-1}\frac{d}{dy}\left\{\frac{1+\frac{\theta(\ln y)^2}{2}}{1+\theta-\theta\ln y+\frac{\theta^2(\ln y)^2}{2}}\right\},\ \ y\in(0,1).$$

Proof. If Y has density (5), then clearly the above differential equation holds. Now, if the differential equation holds, then

$$\frac{d}{dy}\left\{yr_F\left(y\right)\right\}=\theta^2\frac{d}{dy}\left\{\frac{1+\frac{\theta(\ln y)^2}{2}}{1+\theta-\theta\ln y+\frac{\theta^2(\ln y)^2}{2}}\right\},\ \ y\in(0,1),$$

from which we obtain the reverse hazard function of (5).

□

4. Estimation

In this section, the maximum likelihood estimation, method of moments and least squares estimation method are considered to estimate the unknown parameter of the log-xgamma distribution.

4.1. Maximum likelihood estimation

Let $y_1, y_2, \ldots, y_n$ be a random sample from the log-xgamma distribution. The log-likelihood function of log-xgamma distribution is given by

$$\ell = n \ln\left(\frac{\theta^2}{1+\theta}\right) + \sum_{i=1}^{n} \ln\left(1 + \frac{\theta}{2} \ln(y_i)^2\right) + (\theta - 1) \sum_{i=1}^{n} \ln(y_i). \tag{17}$$

Taking partial derivative from (17) with respect to θ, the following normal equation is obtained

$$\frac{\partial \ell}{\partial \theta} = n \frac{(\theta+2)}{\theta(\theta+1)} + \sum_{i=1}^{n} \frac{\ln(y_i)^2}{\theta \ln(y_i)^2 + 2} + \sum_{i=1}^{n} \ln(y_i).$$

The maximum likelihood estimate (MLE) of θ, say $\hat{\theta}$, is the solution of the following equation: $\frac{\partial \ell}{\partial \theta} = 0$. Since the likelihood equation contains non-linear functions, it is not possible to obtain explicit form of the MLE. Therefore, it needs to be solved using numerical methods. S-Plus, R or MATLAB can be used to obtain the MLE of the parameter. It is well-known that under the regularity conditions that are fulfilled for the parameter, the asymptotic distribution of $\widehat{\theta}$, as $n \to \infty$, is a normal distribution with mean θ and variance $\mathbf{I}_F^{-1}(\theta)$ where $\mathbf{I}_F(\theta) = E\left(-\partial^2 \ell / \partial \theta^2\right)$. Therefore, the asymptotic equi-tailed $100(1-p)\%$ confidence interval (CI) for the parameter θ is given by

$$\widehat{\theta} \pm z_{p/2} \sqrt{\widehat{Var(\widehat{\theta})}},$$

where $z_{p/2}$ is the upper $p/2$ quantile of the standard normal distribution.

4.2. Method of moments

The method of moments (MM) estimate of the the parameter θ can be obtained by equating the theoretical moment to the corresponding sample moment as given below

$$\bar{y} = \frac{\theta^2\left(\theta^2 + 3\theta + 1\right)}{(\theta+1)^4}, \tag{18}$$

where $\bar{y}$ is the sample mean. The MM estimate of the parameter θ can be obtained by solving the following equation

$$\bar{y} - \frac{\theta^2\left(\theta^2 + 3\theta + 1\right)}{(\theta+1)^4} = 0. \tag{19}$$

4.3. Least squares estimation

Let $y_{(1)}, y_{(2)}, y_{(3)}, \cdots, y_{(n)}$ denote the ordered sample of the random sample of size n from the log-xgamma distribution. The least square estimator (LSE) of θ can be obtained by minimizing the following equation

$$\sum_{i=1}^{n} \left(F\left(Y_{(i)}\right) - \frac{i}{n+1}\right)^2, \tag{20}$$

where $F\left(Y_{(i)}\right)$ is the cdf of log-xgamma distribution. Inserting (6) in Equation (20), the following equation is obtained.

$$\sum_{i=1}^{n}\left[y_{(i)}^{\theta}(\theta+1)^{-1}\left(1+\theta-\theta\ln\left(y_{(i)}\right)+\frac{\theta^2\ln\left(y_{(i)}\right)^2}{2}\right)-\frac{i}{n+1}\right]^2. \tag{21}$$

5. Simulation

In this section, MLE, LSE and MM methods are considered to estimate the unknown parameter of log-xgamma distribution. We compare the parameter estimation efficiency of MLE, LSE and MM methods for the parameter of log-xgamma distribution by means of Monte Carlo simulation. The following simulation procedure is implemented:

1. Set the sample size n and the parameter θ,
2. Generate random observations from the log-xgamma (θ) distribution with size n using the Equation (9),
3. Using the generated random observations in Step 2, estimate θ by means of MLE, LSE and MM methods,
4. Repeat steps 2 and 3 N times,
5. Using $\hat{\theta}$ and θ compute the mean relative estimates (MREs) and mean square errors (MSEs) via the following equations:

$$Bias=\sum_{j=1}^{N}\frac{\hat{\theta}_j-\theta}{N},$$
$$MRE=\sum_{j=1}^{N}\frac{\hat{\theta}_j/\theta}{N},$$
$$MSE=\sum_{j=1}^{N}\frac{\left(\hat{\theta}_j-\theta\right)^2}{N}.$$

.

The simulation results are computed with software R. The chosen parameter of simulation study is $\theta=2.5$, $N=1.000$ and $n=(20,25,30,...,500)$. We expect that MREs are closer to one when the MSEs are near zero. Figure 4 represents estimated biases, MSEs and MREs obtained by MLE, LSE and MM methods. Based on Figure 4, the biases and MSE of all estimates tend to zero for large n values, and also as expected, the values of MREs tend to one. As seen from Figure 4, the MLE approaches nominal value of the MSEs faster than the LSE and MM estimates. Therefore, the MLE method can be chosen as the more suitable method than LSE and MM methods for estimating the parameter of the log-xgamma distribution.

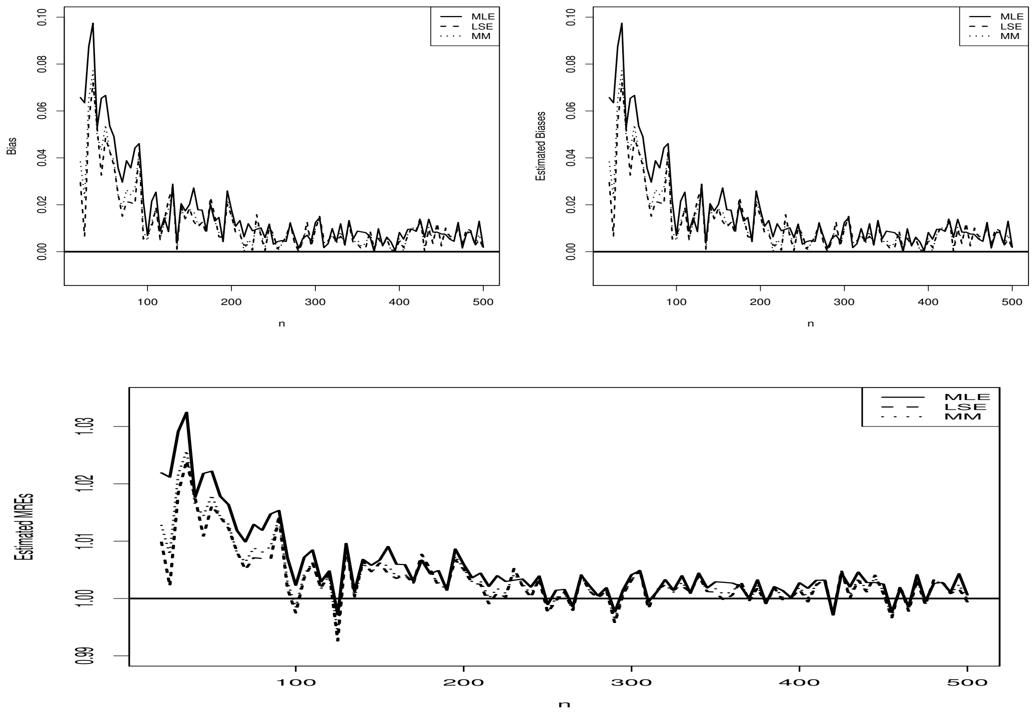

FIGURE 4. *Estimated biases, MSEs and MREs for the parameter of the log-xgamma distribution.*

6. Empirical study

In this section, two real data sets are used to compare the log-xgamma model with Beta, Kumaraswamy and Topp-Leone distributions. The optim function is used to estimate the unknown model parameters. The MLEs and corresponding standard errors, estimated $-\ell$, Kolmogorov-Smirnov (K-S) statistic and corresponding p value, Cramér-von Mises (W*), Anderson-Darling (A*) statistics, Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) are reported in Tables 1 and 2. The lowest values of these criteria show the best fitted model on the data set, except p-value. The log-xgamma distribution is compared with the following distributions defined on the unit interval.

1. Beta distribution

$$f(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},\alpha>0,\beta>0,\ \ 0\leq x\leq 1. \tag{22}$$

2. Kumaraswamy distribution

$$f(x;\alpha,\beta)=\alpha\beta x^{\alpha-1}(1-x^{\alpha})^{\beta-1},\alpha>0,\beta>0,\ \ 0\leq x\leq 1. \tag{23}$$

3. Topp-Leone distribution

$$f(x;\theta)=\theta\,(2-2x)\left(2x-x^{2}\right)^{\theta-1},\theta>0,\ \ 0\leq x\leq 1. \tag{24}$$

The following data is from Caramanis et al. (1983) and Mazumdar and Gaver (1984), where they compare the two different algorithms called SC16 and P3 for estimating unit capacity factors. The values resulted from the algorithm SC16 is 0.853, 0.759, 0.866, 0.809, 0.717, 0.544, 0.492, 0.403, 0.344, 0.213, 0.116, 0.116, 0.092, 0.070, 0.059, 0.048, 0.036, 0.029, 0.021, 0.014, 0.011, 0.008, 0.006. The information about the hazard shape can be helpful in selecting the suitable model. For this purpose, a device called the total time on test (TTT) plot (Aarset, 1987) can be used. The TTT plot is obtained by plotting

$$G(r/n) = \left[\left(\sum_{i=1}^{r} y_{(i)} \right) + (n-r) y_{(r)} \right] / \sum_{i=1}^{n} y_{(i)},$$

where $r = 1, \ldots, n$ and $y_{(i)}$ $(i = 1, \ldots, n)$ are the order statistics of the sample, against r/n. If the shape of TTT plot is a straight diagonal, the hazard is constant. The TTT plot has convex shape for decreasing hazards and concave shape for increasing hazards. The bathtub-shaped hazard is obtained when first is convex and then concave. As seen from Figure 5, the hazard shape of SC16 data set is bathtub-shaped. Therefore, log-xgamma distribution can be good a choice to model this data set.

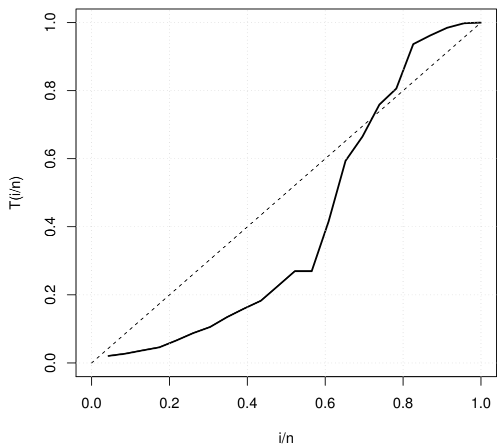

FIGURE 5. *The TTT plot of SC16 data set.*

Table 1 lists the estimated parameters of the models, corresponding standard errors and goodness-of-fit statistics for the SC16 data set. As seen from Table 1, log-xgamma distribution has the lowest values of the goodness-of-fit statistics. Figure 6 displays the fitted pdfs of the models on the histogram of the SC16 data set and fitted functions of the log-xgamma distribution. The right panel of the Figure 6 shows that the log-xgamma distribution provides an adequate fit to the SC16 data set. As seen from application to real data set, one-parameter log-xgamma distribution provides better fits than two-parameter Beta and Kumaraswamy distributions as well as Topp-Leone distribution. The log-xgamma distribution opens new opportunity for modeling the bathtub hazard shape.

Figure 7 displays the fitted hrfs of Beta, Kumaraswamy and Topp-Leone distributions. As seen from Figure 7, all fitted hrfs have bathtub shape.

TABLE 1. *MLEs and their SEs (on second line) of the fitted models and goodness-of-fit statistics for SC16 data*

Models	Parameter estimations		$-\ell$	AIC	BIC	A*	W*	K-S	p-value
Beta(α,β)	0.4869	1.1679	-9.6075	-15.2149	-12.9439	0.6901	0.1099	0.1836	0.4202
	(0.1208)	(0.3578)							
Kumaraswamy(α,β)	0.5044	1.1861	-9.6708	-15.3416	-13.0706	0.6816	0.1084	0.1790	0.4526
	(0.1288)	(0.3264)							
Topp-Leone(θ)	0.5943		-8.1151	-14.2303	-13.0948	0.7456	0.1197	0.1690	0.5273
	(0.1239)								
Log-xgamma(θ)	0.9319		-10.7977	-19.5955	-18.4600	0.4982	0.0761	0.1512	0.6693
	(0.1318)								

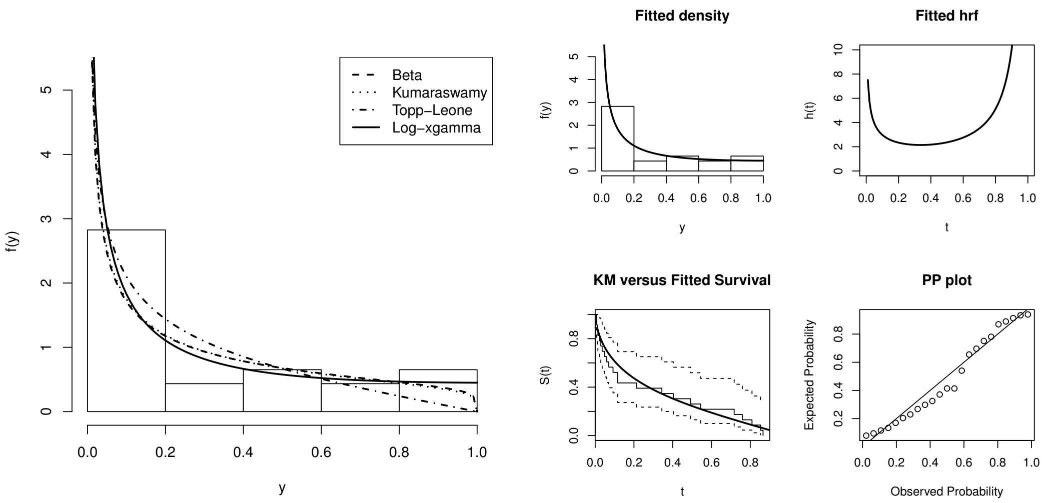

FIGURE 6. *The fitted pdfs of the models for SC16 data (left-panel), the fitted functions of log-xgamma distribution (right-panel).*

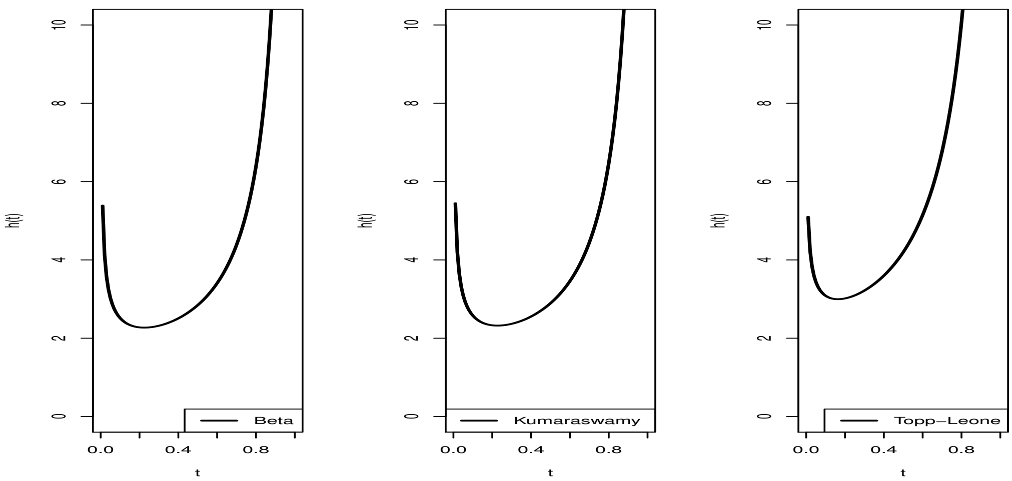

FIGURE 7. *The fitted hrfs of Beta, Kumaraswamy and Topp-Leone distributions*

7. Conclusion

A new one-parameter lifetime distribution called, "log-xgamma" is introduced for modeling lifetime data sets. Some of the statistical properties including the moments, shapes of the hazard rate function, incomplete moments and mean residual life function are obtained. The maximum likelihood, least square and method of moments methods are discussed for estimating the unknown parameter of the log-xgamma distribution via simulation study. Two real data sets are analysed to demonstrate the flexibility and comparability of the log-xgamma distribution against the Beta, Kumaraswamy and Topp-Leone lifetime distributions. The log-xgamma regression model as an alternative to beta regression model can be viewed as a future work of this study. We hope that the results given here will be helpful to the researchers dealing with distribution theory.

References

[1] Caramanis, M., Stremel, J., Fleck, W. and Daniel, S. (1983). Probabilistic production costing: an investigation of alternative algorithms. International Journal of Electrical Power and Energy Systems, 5(2), 75-86.

[2] Cordeiro, G. M. and de Castro, M. (2011). A new family of generalized distributions. Journal of statistical computation and simulation, 81(7), 883-898.

[3] Glänzel, W., A characterization theorem based on truncated moments and its application to some distribution families, Mathematical Statistics and Probability Theory (Bad Tatzmannsdorf, 1986), Vol. B, Reidel, Dordrecht, 1987, 75–84.

[4] Glänzel, W. (1990). Some consequences of a characterization theorem based on truncated moments. Statistics: A Journal of Theoretical and Applied Statistics, 21(4), 613-618.

[5] Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application. Mathematics and computers in simulation, 78(4), 493-506.

[6] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. Journal of Hydrology, 46(1-2), 79-88.

[7] Mazucheli, J., Menezes, A. F. and Dey, S. (2018). The unit-Birnbaum-Saunders distribution with applications. Chilean Journal of Statistics (ChJS), 9(1), 47-57.

[8] Mazumdar, M. and Gaver, D. P. (1984). On the computation of power-generating system reliability indexes. Technometrics, 26(2), 173-185.

[9] Nadarajah, S. and Kotz, S. (2003). Moments of some J-shaped distributions. Journal of Applied Statistics, 30(3), 311-317.

[10] Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of applied econometrics, 11(6), 619-632.

[11] Sen, S., Maiti, S. S. and Chandra, N. (2016). The Xgamma distribution: statistical properties and application. Journal of Modern Applied Statistical Methods, 15(1), 38.

[12] Sen, S., Chandra, N. and Maiti, S. S. (2017). The weighted xgamma distribution: properties and application. Journal of Reliability and Statistical Studies, 10(1).

[13] Sen, S. and Chandra, N (2017). The quasi xgamma distribution with application in bladder cancer data. Journal of Data Science, 15, 61-76.

[14] Topp, C. W. and Leone, F. C. (1955). A family of J-shaped frequency functions. Journal of the American Statistical Association, 50(269), 209-219.

Appendix A

Theorem 1. Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a given probability space and let $H = [a, b]$ be an interval for some $a < b$ $(a = -\infty,\ b = \infty$ might as well be allowed). Let $Y : \Omega \to H$ be a continuous random variable with the distribution function F and let q_1 and q_2 be two real functions defined on H such that

$$\mathbf{E}[q_2(Y) \mid Y \geq y] = \mathbf{E}[q_1(Y) \mid Y \geq y]\,\xi(y), \quad y \in H,$$

is defined with some real function ξ. Assume that $q_1, q_2 \in C^1(H)$, $\xi \in C^2(H)$ and F is twice continuously differentiable and strictly monotone function on the set H. Finally, assume that the equation $\xi q_1 = q_2$ has no real solution in the interior of H. Then F is uniquely determined by the functions q_1, q_2 and ξ , particularly

$$F(y) = \int_a^y C \left| \frac{\xi'(u)}{\xi(u)\, q_1(u) - q_2(u)} \right| \exp(-s(u))\ du\,,$$

where the function s is a solution of the differential equation $s' = \frac{\xi'\, q_1}{\xi q_1 - q_2}$ and C is the normalization constant, such that $\int_H dF = 1$.

We like to mention that this kind of characterization based on the ratio of truncated moments is stable in the sense of weak convergence (see, Glänzel, 1990), in particular, let us assume that there is a sequence $\{Y_n\}$ of random variables with distribution functions $\{F_n\}$ such that the functions q_{1n}, q_{2n} and ξ_n $(n \in \mathbb{N})$ satisfy the conditions of Theorem 1 and let $q_{1n} \to q_1$, $q_{2n} \to q_2$ for some continuously differentiable real functions q_1 and q_2 . Let, finally, Y be a random variable with distribution F . Under the condition that $q_{1n}(Y)$ and $q_{2n}(Y)$ are uniformly integrable and the family $\{F_n\}$ is relatively compact, the sequence Y_n converges to Y in distribution if and only if ξ_n converges to ξ , where

$$\xi(y) = \frac{E[q_2(Y) \mid Y \geq y]}{E[q_1(Y) \mid Y \geq y]}.$$

This stability theorem makes sure that the convergence of distribution functions is reflected by corresponding convergence of the functions q_1 , q_2 and ξ , respectively. It guarantees, for instance, the 'convergence' of characterization of the Wald distribution to that of the Lévy-Smirnov distribution if $\alpha \to \infty$.

A further consequence of the stability property of Theorem 1 is the application of this theorem to special tasks in statistical practice such as the estimation of the parameters of discrete distributions. For such purpose, the functions q_1, q_2 and, specially, ξ should be as simple as possible. Since the function triplet is not uniquely determined it is often possible to choose ξ as a linear function. Therefore, it is worth analyzing some special cases which helps to find new characterizations reflecting the relationship between individual continuous univariate distributions and appropriate in other areas of statistics.