REVSTAT – Statistical Journal
Volume 18, Number 5, October 2020, 593–612

# THE XGAMMA FAMILY: CENSORED REGRESSION MODELLING AND APPLICATIONS

Authors: Gauss M. Cordeiro
– Department of Statistics, Universidade Federal de Pernambuco, Recife, Brazil
gausscordeiro@gmail.com

Emrah Altun
– Department of Mathematics, Bartin University, Bartin, Turkey
emrahaltun@bartin.edu.tr

Mustafa Ç. Korkmaz
– Department of Measurement and Evaluation, Artvin Çoruh University, Artvin, Turkey
mcagatay@artvin.edu.tr

Rodrigo R. Pescim
– Department of Statistics, Universidade Estadual de Londrina, Londrina, Brazil
rrpescim@gmail.com

Ahmed Z. Afify
– Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt
ahmed.afify@fcom.bu.edu.eg

Haitham M. Yousof
– Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt
haitham.yousof@fcom.bu.edu.eg

Received: May 2019 Revised: December 2019 Accepted: December 2019

Abstract:

- In this paper, a new family of distributions with one extra shape parameter, called the xgamma-G, is proposed. comprehensive treatment of some of its mathematical properties including ordinary and incomplete moments and quantile and generating functions are derived. The unknown model parameters are estimated by the maximum likelihood method and the performance of the maximum likelihood estimators are assessed via two extensive simulation studies. Additionally, the log-location-scale regression model for censored data based on a special member of the family is introduced. The usefulness of the proposed models is illustrated utilizing three real data sets.

Key-Words:

- *censored data; maximum likelihood estimation; moment; regression model; xgamma distribution.*

AMS Subject Classification:

- 60E05, 62F10, 62N01.

## 1. INTRODUCTION

Statistical distributions are important tools to model the characteristics of data sets such as right or left skewness, bi-modality or multi-modality observed in different applied sciences such as engineering, medicine and finance and among others. The well-known distributions such as normal, Weibull, gamma, Lindley are extensively used because of theirs simple forms and identifiability properties. However, in the last decade, researchers have focused on the more complex and flexible distributions to increase the modeling ability of these distributions by adding one or more shape parameters. The well-known family of distributions can be cited as follows: Marshall-Olkin-G (Marshall and Olkin [19], 1997), beta-G (Eugene *et al.* [11], 2002), gamma-G (Zografos and Balakrishnan [33], 2009), type I half-logistic-G (Cordeiro *et al.* [8], 2016), Burr X-G (Yousof *et al.* [32], 2016), generalized transmuted-G (Nofal *et al.* [23], 2017) and exponentiated Weibull-H (Cordeiro *et al.* [7], 2017), among others.

Recently, Sen *et al.* [29] (2016) proposed and studied the xgamma (XG) distribution with cumulative distribution function (cdf) and probability density function (pdf) (for $\theta > 0$) given by

$$G(x;\theta) = 1 - \frac{1+\theta+\theta x + \frac{1}{2}\theta^2 x^2}{1+\theta}\exp(-\theta x), \; x > 0 \tag{1.1}$$

and

$$g(x;\theta) = \frac{\theta^2}{1+\theta}\left(1+\frac{\theta}{2}x^2\right)\exp(-\theta x), \tag{1.2}$$

respectively. During the recent years, the xgamma distribution has been shown great interest by researchers. Altun and Hamedani [2] (2018) introduced a new bounded distribution using the transformation $Y = \exp(-X)$ as an alternative to the beta distribution based on the xgamma distribution. Biçer [4] (2019) introduced the transmuted-xgamma distribution and studied its statistically properties comprehensively. The another generalization of xgamma distribution was provided by Sen *et al.* [27] (2018a) on the basis of special mixture of exponential and gamma distributions. Sen *et al.* [28] (2018b) studied the parameter estimation of xgamma distribution under progressively type-II right censoring scheme by maximum likelihood and Bayesian estimation methods. Sen and Chandra [25] (2017) introduced the quasi-xgamma distribution by using the xgamma distribution as a baseline distribution. The weighted generalization of xgamma distribution, using $w(x) = x^r$ as a weighting function, was studied by Sen *et al.* [26] (2017).

In this paper, we introduce and study a new class of distributions called the *xgamma-G* (XG-G) family. The idea is to incorporate any distribution into a larger family through an application of the XG cdf. In fact, based on the T–X transform defined by Alzaatreh *et al.* [3] (2013) and the XG cdf, we construct the XG-G family. The some of its mathematical properties are provided comprehensively. The new family has flexible shapes to model various lifetime data sets. Additionally, its special models produce better fits than other well-known families.

To this end, we define the cdf of the XG-G family with one extra shape parameter $\theta > 0$ by

$$\begin{aligned} F(x;\theta,\boldsymbol{\xi}) &= \frac{\theta^2}{1+\theta}\int_0^{-\log \overline{G}(x;\boldsymbol{\xi})}\left(1+\frac{\theta}{2}t^2\right)\exp\left(-\theta t\right)\,dt \\ &= 1-\frac{1+\theta-\theta\log\overline{G}(x;\boldsymbol{\xi})+\frac{1}{2}\theta^2\left[\log\overline{G}(x;\boldsymbol{\xi})\right]^2}{1+\theta}\overline{G}(x;\boldsymbol{\xi})^{\theta}, \end{aligned} \tag{1.3}$$

where $\overline{G}(x;\boldsymbol{\xi}) = 1 - G(x;\boldsymbol{\xi})$ and $G\left(x;\boldsymbol{\xi}\right)$ is a baseline cdf with a parameter vector $\boldsymbol{\xi}$.

The pdf corresponding to (1.3) reduces to

$$f(x;\theta,\boldsymbol{\xi}) = \frac{\theta}{1+\theta}g(x;\boldsymbol{\xi})\overline{G}(x;\boldsymbol{\xi})^{\theta-1}\left\{\theta+\frac{1}{2}\theta^2\left[\log\overline{G}(x;\boldsymbol{\xi})\right]^2\right\}, \tag{1.4}$$

where $g(x;\boldsymbol{\xi}) = dG(x;\boldsymbol{\xi})/dx$. If the random variable (rv) $T$ has the xgamma distribution (1), then $X = G^{-1}\left[1-\exp\left(-T\right)\right]$ follows the XG-G family (4). Henceforth, we denote by $X \sim$XG-G$(\theta,\boldsymbol{\xi})$ a rv having density (1.4). The hazard rate function (hrf) of $X$ is given by

$$\tau\left(x;\theta,\boldsymbol{\xi}\right) = \frac{\theta r(x;\boldsymbol{\xi})\left\{\theta+\frac{1}{2}\theta^2\left[\log\overline{G}(x;\boldsymbol{\xi})\right]^2\right\}}{\left\{1+\theta-\theta\log\overline{G}(x;\boldsymbol{\xi})+\frac{1}{2}\theta^2\left[\log\overline{G}(x;\boldsymbol{\xi})\right]^2\right\}}.$$

The identifiability is an important property of the statistical distributions to satisfy the precise inference for the model parameters. The following theorem is given to prove the identifiability property of XG-G family.

**Theorem 1.1.** *The cdf (1.3) is identifiable.*

**Proof:** Assume that the baseline cdf $G\left(x;\boldsymbol{\xi}\right)$ is identifiable. The cdf (1.3) is identifiable once $F\left(x;\theta_1\right) = F\left(x;\theta_2\right)$ is valid if and only if $\theta_1 = \theta_2$. Using (1.3), we have

$$\begin{aligned} &F\left(x;\theta_1\right) = F\left(x;\theta_2\right) \\ &1-\left(1+\theta_1-\theta_1 A+\tfrac{1}{2}\theta_1^2A^2\right)\left(1+\theta_1\right)^{-1}\exp\left(A\theta_1\right) \\ &= 1-\left(1+\theta_2-\theta_2 A+\tfrac{1}{2}\theta_2^2A^2\right)\left(1+\theta_2\right)^{-1}\exp\left(A\theta_2\right) \end{aligned} \tag{1.5}$$

where $A = \log\bar{G}\left(x\right)$. (1.5) can be simplified as follows

$$\begin{aligned} &\left[\frac{\exp\left(A\theta_2\right)}{\left(1+\theta_2\right)}-\frac{\exp\left(A\theta_1\right)}{\left(1+\theta_1\right)}\right]+\left[\frac{\exp\left(A\theta_2\right)\theta_2}{\left(1+\theta_2\right)}-\frac{\exp\left(A\theta_1\right)\theta_1}{\left(1+\theta_1\right)}\right] \\ &-\left[\frac{\exp\left(A\theta_2\right)\theta_2 A}{\left(1+\theta_2\right)}-\frac{\exp\left(A\theta_1\right)\theta_1}{\left(1+\theta 1\right)}\right]+\left[\frac{\exp\left(A\theta_2\right)\theta_2^2A^2}{2\left(1+\theta_2\right)}-\frac{\exp\left(A\theta_1\right)\theta_1^2A^2}{2\left(1+\theta_1\right)}\right]=0 \end{aligned} \tag{1.6}$$

The expression (1.6) is equal to zero for all $x$ only when the parameters $\theta_1 = \theta_2$. Since the parameter $\theta > 0$, it is concluded that the model is identifiable: $F\left(x;\theta_1\right) = F\left(x;\theta_2\right) \Leftrightarrow \theta_1 = \theta_2$. □

The purpose of the generation of the XG-G family is to provide new opportunities to model the different characteristics of the data sets such as left skewness, excess kurtosis and bathtub failure rate. The well-known distributions are insufficient to model these kinds of data sets. The special members of the XG-G family can be used to model skewed and long-tailed data sets to improve the modeling accuracy of interested data set with only one extra shape parameter. Moreover, the proposed family is highly effective in modeling the censored lifetimes of individuals with some covariates in a location-scale regression framework.

The remaining part of the paper is organized as follows. In Section 2, three special cases of the XG-G family are given. In Section 3, a linear representation of the XG-G density is provided. The comprehensive mathematical properties of the XG-G density are obtained and reported in Section 4. Section 5 is devoted to the maximum likelihood estimation of the model parameters for uncensored and censored data. In Section 6, we present a new log-location-scale regression model based on the log XG-Weibull distribution. Section 7 deals with simulation studies to evaluate the maximum likelihood estimators of the parameters of proposed models. In Section 8, three applications to the real data sets are given to prove empirically the importance of XG-G family. Section 9 contains the concluding remarks of the study.

## 2. SOME SPECIAL XG-G MODELS

### 2.1. The XG-Lindley (XG-Li) model

Consider the cdf $G(x) = 1 - \frac{1+a+ax}{1+a} \exp(-ax)$ of the Li distribution with scale parameter $a > 0$. The XG-Li density (for $x > 0$) can be determined from (1.4). Some plots of the XG-Li density and hazard functions for selected parameter values are displayed in Figure 1. These plots reveal that the pdf of the XG-Li model can be reversed J-shape, right skewed or unimodal. The hrf can be unimodal or bathtub.

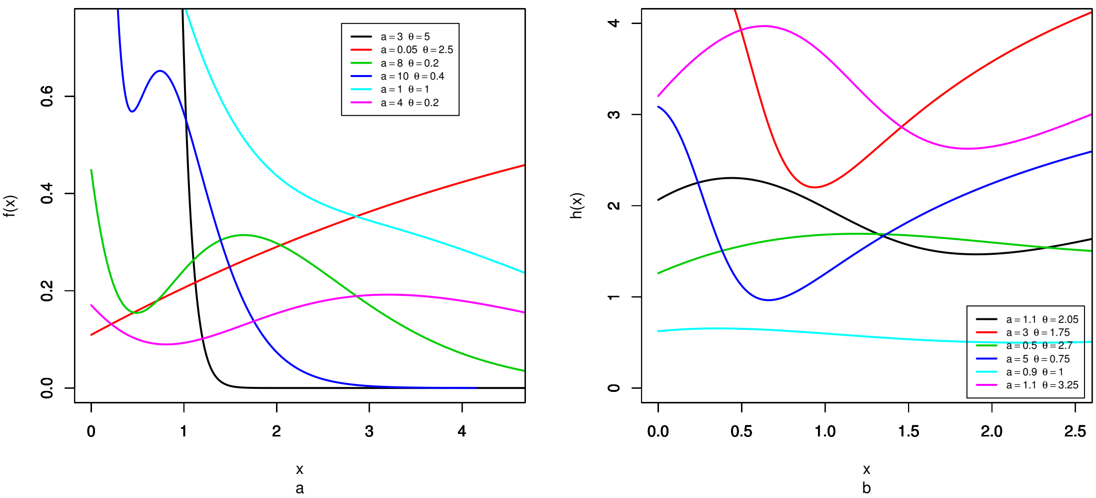



**Figure 1**: Plots of the XG-Li pdf (left) and hrf (right) for some parameter values.

## 2.2. The XG-Weibull (XG-W) model

Consider the cdf $G(x) = 1 - \exp[-(ax)^b]$ of the W distribution with scale $a > 0$ and shape $b > 0$. The pdf of the XG-W model (for $x > 0$) follows from (1.4). Some plots of the XG-W pdf and hrf for selected parameter values are displayed in Figure 2. Figure 2 reveals that the XG-W density can be concave down, left skewed or right skewed. The hrf of the XG-W model can be increasing, decreasing, bathtub or unimodal.

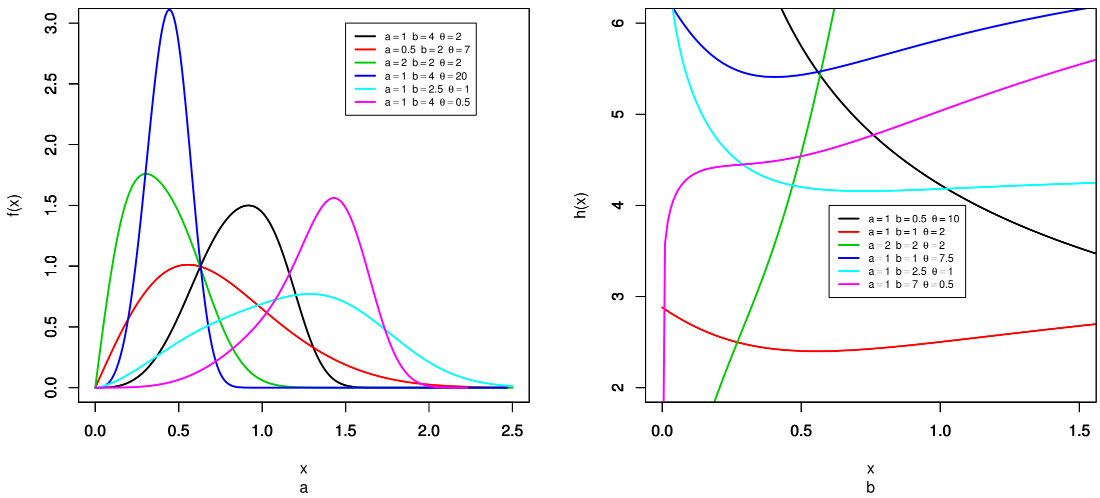



**Figure 2**: Plots of the XG-W pdf (left) and hrf (right) for some parameter values.

## 2.3. The XG-BurrXII (XG-BXII) model

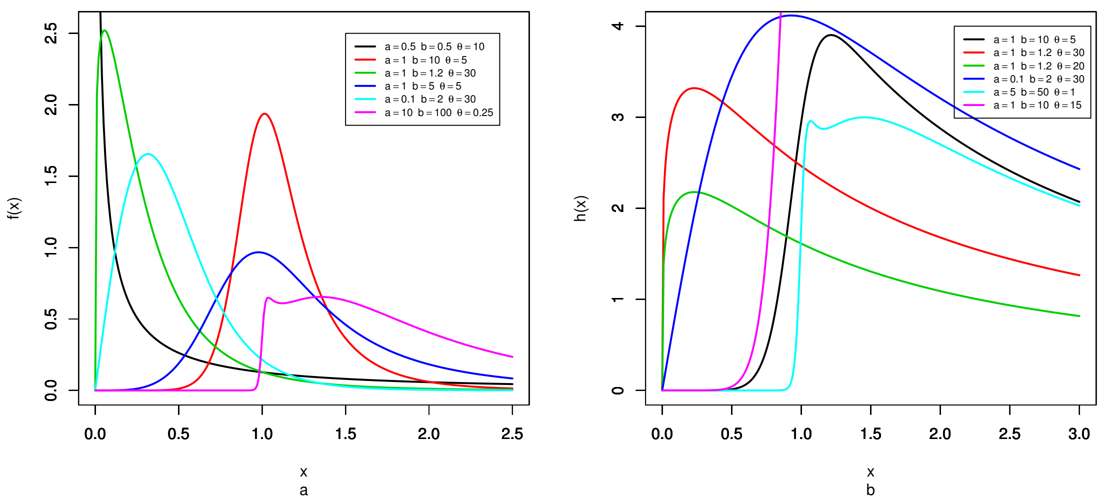



**Figure 3**: Plots of the XG-BXII pdf (left) and hrf (right) for some parameter values.

Consider the cdf $G(x) = 1 - (1 + x^a)^{-b}$ of the BXII distribution with parameters $a > 0$ and $b > 0$. The pdf of the XG-BXII model (for $x > 0$) can be obtained from (1.4). Some plots of the XG-BXII pdf and hrf for selected parameter values are displayed in Figure 3. These plots reveal that the pdf of the XG-BXII model can be reversed J-shape, concave down or right skewed. Its hrf can be increasing or unimodal.

## 3. USEFUL REPRESENTATION OF PDF AND CDF

The XG-G family density in (1.4) can be expressed as

$$f\,(x) = \frac{\theta^2 g(x)}{1+\theta}\overline{G}(x)^{\theta-1} + \frac{\theta^3 g(x)}{2\,(1+\theta)}\overline{G}(x)^{\theta-1}\underbrace{[\log \overline{G}(x)]}_{A}^{2}.$$

Consider

$$\log(1-z) = -\sum_{i=0}^{\infty}\frac{z^{i+1}}{i+1}, \quad |z| \; < 1, \tag{3.1}$$

and the power series raised to a positive integer $n$ (Gradshteyn and Ryzhik [14, Section 0.314], 2002)

$$\left(\sum_{j=0}^{\infty} a_j\, u^j\right)^n = \sum_{j=0}^{\infty} c_{n,j}\, u^j, \tag{3.2}$$

where the coefficients $c_{n,j}$ (for $j = 1, 2, ...$) can be easily determined from the recurrence equation

$$c_{n,j} = (j a_0)^{-1}\sum_{m=1}^{j}\,[m\,(n+1) - j]\; a_m\, c_{n,j-m} \quad \text{and} \quad c_{n,0} = a_0^n.$$

The coefficient $c_{n,j}$ can be calculated from $c_{n,0}, ..., c_{n,j-1}$ and hence from the quantities $a_0, ..., a_j$. For $|z| < 1$ and $b > 0$, the power series holds

$$(1-z)^{b-1} = \sum_{k=0}^{\infty}\frac{(-1)^k\,\Gamma(b)}{k!\,\Gamma(b-k)}\, z^k. \tag{3.3}$$

Applying (3.1) to the quantity $A$ gives

$$f(x) = \frac{\theta^2 g(x)}{1+\theta}\overline{G}(x)^{\theta-1} + \frac{\theta^3 g(x)}{2(1+\theta)}\overline{G}(x)^{\theta-1} G(x)^2 \underbrace{\left[\sum_{i=0}^{\infty}\frac{G(x)^i}{i+1}\right]^2}_{B}.$$

Next, the quantity $B$ follows using (3.2) as

$$f(x) = \frac{\theta^2 g(x)}{1+\theta}\;\underbrace{\overline{G}(x)^{\theta-1}}_{C} + \frac{\theta^3 g(x)}{2(1+\theta)}\sum_{i=0}^{\infty} c_{2,i}\, G(x)^{i+2}\;\underbrace{\overline{G}(x)^{\theta-1}}_{C},$$

where $a_i = 1/(i+1)$.

Applying the power series (3.3) to the quantity $C$, we obtain

$$f(x) = \sum_{k=0}^{\infty}\left[b_k\ \pi_{k+1}(x) + \sum_{i=0}^{\infty} b_{i,k}\ \pi_{i+k+3}(x)\right], \tag{3.4}$$

where

$$b_k = \frac{(-1)^k\,\theta^2\Gamma(\theta)}{(k+1)\,(1+\theta)\,\Gamma(\theta-k)}, \quad b_{i,k} = \frac{(-1)^k\,\theta^3\Gamma(\theta)\ c_{2,i}}{2(1+\theta)(i+k+3)\,k!\,\Gamma(\theta-k)},$$

and $\pi_\alpha(x) = \alpha\,g(x)\,G(x)^{\alpha-1}$ is the exponentiated-G (Exp-G) density function with power parameter $\alpha > 0$. So, the density of $X$ is a linear combination of Exp-G densities.

The properties of Exp-G distributions have been studied by many authors in recent years, see, for example, Mudholkar and Srivastava [20] (1993) and Mudholkar *et al.* [21] (1995) for exponentiated Weibull (EW), Gupta and Kundu [16] (1999) for exponentiated exponential and Nadarajah and Gupta [22] (2007) for exponentiated gamma, among others.

The cdf of $X$ follows by integrating (3.4) as

$$F(x) = \sum_{k=0}^{\infty}\left[b_k\ \Pi_{k+1}(x) + \sum_{i=0}^{\infty} b_{i,k}\ \Pi_{i+k+3}(x)\right], \tag{3.5}$$

where $\Pi_\alpha(x) = G(x)^\alpha$ is the Exp-G cdf with power parameter $\alpha$. Equations (3.4) and (3.5) are the main results of this section.

## 4. PROPERTIES

In this section, we investigate some mathematical properties of the XG-G family.

### 4.1. Quantile function

The quantile function (qf) of $X$ can be determined by inverting $F(x) = u$ in (1.3). We require numerical methods to obtain the quantiles. For given $u$, we solve numerically for $z = z(u)$ in the equation

$$\left[1+\theta-\theta\log(z)+0.5\,\theta^2\,\log^2(z)\right]\ z^\theta = (1+\theta)(1-u),$$

and then $x = Q(u) = G^{-1}(1-z)$ is a variate from the XG-G family (1.4).

### 4.2. Moments

Let $Y_\alpha$ be a rv having density $\pi_\alpha(x)$. The $r$-th ordinary moment of $X$, say $\mu'_r$, follows from (3.4) as

$$\mu'_r = E(X^r) = \sum_{k=0}^{\infty}\left[b_k\ E\left(Y_{k+1}^r\right) + \sum_{i=0}^{\infty} b_{i,k}\ E\left(Y_{i+k+3}^r\right)\right], \tag{4.1}$$

where $E(Y_\alpha^r) = \alpha \int_{-\infty}^{\infty} x^r\ g(x)\, G(x)^{\alpha-1}\ dx$ can be evaluated numerically in terms of the baseline qf $Q_G(u) = G^{-1}(u)$ as $E(Y_\alpha^n) = \alpha \int_0^1 Q_G(u)^n\ u^{\alpha-1} du$. Setting $r = 1$ in (4.1) gives the mean of $X$. Table 1 lists the first three ordinary moments of XG-W distribution. The results given in this table show that when the parameter $\theta$ increases, the ordinary moments of XG-W decrease for fixed $a$ and $b$ parameters.

**Table 1**: Moments of XG-W distribution for several parameter values.

| Parameters | | | $\mu_1'$ | $\mu_2'$ | $\mu_3'$ |
|---|---|---|---|---|---|
| $\theta$ | $a$ | $b$ | | | |
| 2 | 2 | 2 | 1.619 | 3.333 | 7.990 |
| 2 | 2 | 1 | 0.809 | 0.833 | 0.999 |
| 2 | 2 | 0.5 | 0.405 | 0.208 | 0.125 |
| 2 | 1 | 0.5 | 0.417 | 0.333 | 0.375 |
| 2 | 0.5 | 0.5 | 0.667 | 2.125 | 14.035 |
| 1 | 0.5 | 0.5 | 3.500 | 48.000 | 1304.865 |
| 0.5 | 0.5 | 0.5 | 17.231 | 953.497 | 94367.230 |

## 4.3. Incomplete moments

The $r$-th incomplete moment of $X$ is given by

$$m_r(y) = \int_{-\infty}^{y} x^r\, f(x) dx. \tag{4.2}$$

Using (3.4), the $r$-th incomplete moment of XG-G family is

$$m_r(y) = \sum_{k=0}^{\infty} \left[ b_k\ m_{r,k+1}(y) + \sum_{i=0}^{\infty} b_{i,k}\ m_{r,i+k+3}(y) \right],$$

where $m_{r,\alpha}(y) = \int_0^{G(y)} Q_G^r(u)\, u^{\alpha-1}\ du$. The $m_{r,\alpha}(y)$ can be calculated numerically by using the software such as **Matlab**, **R**, **Mathematica** etc. The incomplete moments of the XG-W distribution are given in Table 2. As seen from the results given in Table 2, the incomplete moments of XG-W distribution increases for fixed $a$ and $b$ parameters when the parameter $\theta$ increases.

**Table 2**: Incomplete moments of XG-W distribution for several parameter values.

| Parameters | | | $\mu_1(0.5)$ | $\mu_1(1)$ | $\mu_1(2)$ |
|---|---|---|---|---|---|
| $\theta$ | $a$ | $b$ | | | |
| 2 | 2 | 2 | 0.026 | 0.170 | 0.790 |
| 2 | 2 | 1 | 0.085 | 0.395 | 0.799 |
| 2 | 2 | 0.5 | 0.197 | 0.400 | 0.405 |
| 2 | 1 | 0.5 | 0.135 | 0.293 | 0.406 |
| 2 | 0.5 | 0.5 | 0.080 | 0.168 | 0.313 |
| 1 | 0.5 | 0.5 | 0.051 | 0.129 | 0.320 |
| 0.5 | 0.5 | 0.5 | 0.022 | 0.060 | 0.166 |

### 4.4. Moment generating function

The moment generating function (mgf) of $X$, say $M(t) = E(\mathrm{e}^{t\,X})$, is obtained from (3.4) as

$$M(t) = \sum_{k=0}^{\infty}\left[b_k\; M_{k+1}\,(t) + \sum_{i=0}^{\infty} b_{i,k}\; M_{i+k+3}\,(t)\right],$$

where $M_\alpha(t)$ is the generating function of $Y_\alpha$ given by

$$M_\alpha(t) = \alpha \int_{-\infty}^{\infty} \mathrm{e}^{t\,x} G(x) g(x)^{\alpha-1}\,dx = \alpha \int_0^1 \exp[t\,Q_G(u;\alpha\,)]\,u^{\alpha-1}du.$$

The last two integrals can be computed numerically for most parent distributions.

## 5. ESTIMATION

This section deals with the maximum likelihood estimation of the unknown model parameters.

### 5.1. Maximum likelihood estimation

Let $x_1, \cdots, x_n$ be a random sample from the XG-G models with a parameter vector $\mathbf{\Phi} = (\theta, \xi^\intercal)^\intercal$. The log-likelihood function is given by

$$\ell_n(\mathbf{\Phi}) \;=\; n\log\theta - n\log(1+\theta) + \sum_{i=1}^{n}\log g(x_i;\xi) + (\theta-1)\sum_{i=1}^{n}\log\overline{G}(x_i;\xi) + \sum_{i=1}^{n}\log\left\{\theta + \frac{1}{2}\theta^2\left[\log\overline{G}(x_i;\xi)\right]^2\right\}.$$

Taking the partial derivatives of the log-likelihood function concerning the parameters, we obtain the score vectors. The simultaneous solution of these equations for zero gives the maximum likelihood estimate of $\mathbf{\Phi}$. Since it is not possible to obtain closed-form expressions of the maximum likelihood estimators of the parameters of XG-G family, direct maximization of the log-likelihood is needed. In this study, the **optim** function of **R** software is used to minimize the minus of the log-likelihood function which is equivalent to the maximization of log-likelihood.

## 5.2. Multi-censored maximum likelihood estimation

Censored data are often encountered in survival analysis and reliability studies. Here, the general case of multi-censored data is considered. Assume that $m_0$ subjects of $m$ are failed at the times $x_1, \cdots, x_{m_0}$, $m_1$ subjects of $m$ are failed in $(s_{j-1}, s_j)$ interval where $j = 1, ..., m_1$ and $m_2$ subjects of $m$ survived until a time $r_j$, $j = 1, ..., m_2$. Note that $m_0 + m_1 + m_2 = m$. The log-likelihood function for $\mathbf{\Phi}$ is

$$\begin{aligned}
\ell_m(\mathbf{\Phi}) \;=\; & m_0 \log\theta - m_0 \log(1+\theta) + \sum_{i=1}^{m_0} \log g(x_i, \xi) \\
& + (\theta - 1) \sum_{i=1}^{m_0} \log \overline{G}(x_i, \xi) + \sum_{i=1}^{m_0} \log \left\{ \theta + \frac{1}{2}\theta^2 \left[\log \overline{G}(x_i, \xi)\right]^2 \right\} \\
& + \sum_{i=1}^{m_2} \log \left\{ \frac{1}{1+\theta} \left[ 1 + \theta - \theta \log t_{r_i} + \frac{(\log t_{r_i})^2}{2\theta^{-2}} \right] t_{r_i}^{\theta} \right\} \\
& + \sum_{i=1}^{m_1} \log \left( \left\{ 1 - \frac{1}{1+\theta} \left[ 1 + \theta - \theta \log t_{s_i} + \frac{(\log t_{s_i})^2}{2\theta^{-2}} \right] t_{s_i}^{\theta} \right\} \right. \\
& \left. - \left\{ 1 - \frac{1}{1+\theta} \left[ 1 + \theta - \theta \log t_{s_{i-1}} + \frac{(\log t_{s_{i-1}})^2}{2\theta^{-2}} \right] t_{s_{i-1}}^{\theta} \right\} \right),
\end{aligned}$$

where $t_{r_i} = \overline{G}(r_i, \xi)$, $t_{s_i} = \overline{G}(s_i, \xi)$, $t_{s_{i-1}} = \overline{G}(s_{i-1}, \xi)$ and the normal equations are available before.

# 6. THE LXG-W REGRESSION MODEL FOR CENSORED DATA

Let $X$ be a rv having the XG-W density function. The rv $Y = \log(X)$ defines the *log-xgamma Weibull* (LXG-W) distribution. Let $a = \mathrm{e}^{-\mu}$ and $b = \sigma^{-1}$. Then, the pdf of $Y$ (for $y \in \Re$) is given by

$$f(y) = \frac{\theta}{\sigma(1+\theta)} \exp\left[(1-\theta)\left(\frac{y-\mu}{\sigma}\right)\right] \left\{ \theta + \frac{\theta^2}{2} \left[ -\exp\left(\frac{y-\mu}{\sigma}\right)\right]^2 \right\}, \tag{6.1}$$

where $\mu \in \Re$, $\sigma > 0$ and $\theta > 0$. If $Y$ is a rv having density function (6.1), we can write $Y \sim$ LXG-W$(\theta, \mu, \sigma)$. For $\sigma = 1$, the LXG-W distribution reduces to the log-xgamma-exponential (LXG-E) distribution. The survival function (sf) corresponding to (6.1) is given by

$$S(y) = \frac{1}{1+\theta} \left[ (1+\theta) + \left(\theta - \frac{\theta^2}{2}\right) \exp\left(\frac{y-\mu}{\sigma}\right) \right] \left\{ \exp\left[ -\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\theta}.$$

We define the standardized rv $Z = (Y - \mu)/\sigma$ with pdf (for $z \in \Re$) given by

$$f(z) = \frac{\theta}{(1+\theta)} \exp\left[(1-\theta)z\right] \left[\theta + \frac{\theta^2}{2} \exp(2z)\right]. \tag{6.2}$$

Regression models are widely used to model dependent variable with some covariates. The lifetimes of individuals are generally effected by some explanatory variables such as gender, age, alcohol abuse or smoking. To model these kind of data sets, we propose a new log-location-scale regression model based on the LXG-W density. Let $yi$ be the response variable and $\mathbf{v}_i^T = (v_{i1}, ..., v_{ip})$ is the explanatory variable vector, we consider the following regression model

$$y_i = \mathbf{v}_i^T \beta + \sigma\, z_i, \;\; i = 1, ..., n. \tag{6.3}$$

where $y_i$ follows the LXG-W density with unknown parameters $\mu_i \in \Re$, $\theta > 0$, and $\sigma > 0$. The location of $y_i$, $\mu_i$, is modeled by using the identity link function, $\mu_i = \mathbf{v}_i^T \beta$. The vector $\mu = (\mu_1, ..., \mu_n)^T$ is defined as $\mu = \mathbf{V}\beta$, where $\mathbf{V} = (\mathbf{v}_1, ..., \mathbf{v}_n)^T$ is a known model matrix.

Let the random sample $(y_1, \mathbf{v}_1), ..., (y_n, \mathbf{v}_n)$ are independent and the response variable is defined as $y_i = \min\{\log(t_i), \log(c_i)\}$. Assume that the lifetimes and censoring times are independent. $F$ and $C$ represent the sets of individuals for the log-lifetime and log-censoring, respectively. The log-likelihood function for the vector of parameters $\eta = (\beta^T, \theta, \sigma)^T$ is given by

$$\begin{aligned} l(\eta) \;=\; & r \log\left[\frac{\theta}{\sigma\,(1+\theta)}\right] + (1-\theta) \sum_{i \in F} \frac{y_i - \mathbf{v}_i^T \beta}{\sigma} \\ & + \sum_{i \in F} \log\left\{\theta + \frac{\theta^2}{2}\left[-\exp\left(\frac{y_i - \mathbf{v}_i^T\beta}{\sigma}\right)\right]^2\right\} \\ & + c \log\left(\frac{1}{1+\theta}\right) - \theta \sum_{i \in C} \exp\left(\frac{y_i - \mathbf{v}_i^T\beta}{\sigma}\right) \\ & + \sum_{i \in C} \log\left[(1+\theta) + \left(\theta - \frac{\theta^2}{2}\right) \exp\left(\frac{y_i - \mathbf{v}_i^T\beta}{\sigma}\right)\right], \end{aligned} \tag{6.4}$$

where $r$ and $c$ are the number of uncensored (failures) and censored observations. The parameter vector, $\eta$, of the LXG-W regression model is estimated by minimizing the minus of log-likelihood function, given in (6.4). To do this, the **optim** function of R software is used. The inverse of the observed information matrix is used to obtain corresponding standard errors and construct 95% asymptotic confidence intervals of the parameters. The observed information matrix is evaluated numerically at $\hat{\eta}$ by **hessian** function of R software.

## 7. SIMULATION STUDIES

In this section, three simulation studies are given to evaluate the finite sample performance of the parameters of proposed models.

### 7.1. Simulations of XG-W and XG-N distributions

Here, we perform two simulation studies using the XG-W and XG-normal (XG-N) distributions. To verify the performance of the MLEs for these distributions, we generate 1,000 samples of sizes 20, 50 and 100 from their qfs by inverting the cdfs. The simulation results are reported in Tables 3 and 4. These results reveal that the mean estimates become closer to the true parameter values when the sample size increases, whereas the standard errors of the estimates decrease.

The cdfs of the XG-W and XG-N distributions are given here for convenience

$$F(x) = 1 - \frac{1+\theta+\theta\,(ax)^{\,b} + \frac{\theta^2}{2}(ax)^{2b}}{1+\theta}\exp\left[-\theta(ax)^b\right]$$

and

$$F(x) \;=\; 1 - \frac{1+\theta-\theta\log\left[1-\boldsymbol{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]+\frac{1}{2}\theta^2\left\{\log\left[1-\boldsymbol{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]\right\}^2}{1+\theta} \times\left[1-\boldsymbol{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{\theta},$$

respectively, where $x\,,\mu\in\Re,\theta,\sigma>0$.

**Table 3**: Empirical means and standard errors (in parentheses) for different values of the XG-W parameters.

| Parameters | | $n=20$ | | | $n=50$ | | | $n=100$ | |
|---|---|---|---|---|---|---|---|---|---|
| $a,b,\theta$ | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\theta}$ | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\theta}$ | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\theta}$ |
| 5, 5, 5 | 5.3196 | 5.2006 | 4.7679 | 5.2329 | 4.8741 | 4.9432 | 5.0876 | 4.9086 | 5.0471 |
| | (0.5927) | (0.9124) | (1.5578) | (0.5339) | (0.5449) | (1.5381) | (0.1677) | (0.3679) | (0.3703) |
| 50,3,3 | 50.6839 | 3.0894 | 3.2746 | 49.5001 | 2.9022 | 3.1821 | 49.9190 | 2.9504 | 3.1318 |
| | (1.9717) | (0.6324) | (5.0238) | (1.9357) | (0.3407) | (0.5332) | (1.8658) | (0.2525) | (0.4192) |
| 3,3,50 | 3.3019 | 3.2521 | 50.0152 | 3.0971 | 3.0684 | 50.0124 | 3.0622 | 3.0433 | 49.9866 |
| | (0.7011) | (0.6395) | (0.1291) | (0.4194) | (0.3396) | (0.1232) | (0.2885) | (0.2412) | (0.3244) |
| 3,10,3 | 3.0714 | 10.3185 | 2.9319 | 3.0328 | 9.8262 | 3.0470 | 3.0203 | 9.8692 | 3.0153 |
| | (0.1308) | (1.3464) | (0.6916) | (0.0526) | (0.6161) | (0.1897) | (0.0460) | (0.5184) | (0.1689) |
| 50,10,50 | 51.0354 | 10.6994 | 50.1005 | 50.4395 | 10.2804 | 50.0442 | 50.3066 | 10.1571 | 50.0298 |
| | (3.2573) | (2.0583) | (0.3469) | (2.0441) | (1.1738) | (0.2420) | (1.4197) | (0.7910) | (0.1551) |
| 0.01,2,5 | 0.0107 | 1.9949 | 4.9998 | 0.0106 | 1.9961 | 4.9999 | 0.0105 | 1.9970 | 5.0001 |
| | (0.0013) | (0.1175) | (0.0013) | (0.0008) | (0.0391) | (0.0003) | (0.0005) | (0.0264) | (0.0004) |
| 1,1,1 | 0.9338 | 1.1141 | 1.2913 | 0.9526 | 1.0424 | 1.1839 | 1.0430 | 1.0327 | 1.0584 |
| | (0.5201) | (0.2313) | (0.5837) | (0.3619) | (0.1177) | (0.4090) | (0.3041) | (0.1023) | (0.3542) |
| 1,2,3 | 1.1736 | 2.0477 | 2.9188 | 1.1361 | 1.9665 | 3.0585 | 1.0643 | 1.9385 | 3.0133 |
| | (0.4524) | (0.3743) | (0.7855) | (0.3183) | (0.2056) | (0.7672) | (0.1287) | (0.1538) | (0.3812) |
| 2,2,2 | 2.2538 | 1.9168 | 2.4294 | 2.0430 | 1.9369 | 2.3574 | 1.9826 | 1.9603 | 2.3360 |
| | (0.8596) | (0.3450) | (1.2688) | (0.5113) | (0.2314) | (0.7653) | (0.4837) | (0.1494) | (0.7579) |
| 5,0.9,5 | 5.6190 | 0.9244 | 5.5789 | 5.2117 | 0.8744 | 5.2745 | 5.1839 | 0.9059 | 5.1761 |
| | (0.8374) | (0.1629) | (1.0001) | (0.5833) | (0.0824) | (0.5583) | (0.4753) | (0.0792) | (0.4106) |
| 0.025,0.9,1 | 0.0271 | 0.9142 | 1.0044 | 0.0254 | 0.9081 | 0.9965 | 0.0253 | 0.8999 | 0.9968 |
| | (0.0131) | (0.1036) | (0.0730) | (0.0041) | (0.0767) | (0.0365) | (0.0040) | (0.0485) | (0.0540) |

**Table 4**: Empirical means and standard errors (in parentheses) for different values of the XG-N parameters.

| Parameters | | $n=20$ | | | $n=50$ | | | $n=100$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta,\mu,\sigma$ | $\widehat{\theta}$ | $\widehat{\mu}$ | $\widehat{\sigma}$ | $\widehat{\theta}$ | $\widehat{\mu}$ | $\widehat{\sigma}$ | $\widehat{\theta}$ | $\widehat{\mu}$ | $\widehat{\sigma}$ |
| 5,0,1 | 4.9748 | -0.0626 | 0.9531 | 5.0029 | -0.0163 | 0.9829 | 5.0023 | -0.0111 | 0.9918 |
| | (0.7832) | (0.2480) | (0.1611) | (0.4507) | (0.1539) | (0.0968) | (0.2156) | (0.0966) | (0.0674) |
| 1,0,1 | 0.9254 | 0.0846 | 0.9605 | 1.0915 | 0.0519 | 0.9791 | 1.0126 | 0.1369 | 1.0126 |
| | (0.3459) | (0.6632) | (0.1990) | (0.5446) | (0.5433) | (0.1819) | (0.1570) | (0.4313) | (0.1570) |
| 5,-1,1 | 4.8906 | -1.0657 | 0.9416 | 5.0272 | -1.0172 | 0.9850 | 4.9931 | -1.0043 | 0.9922 |
| | (0.7691) | (0.2425) | (0.1598) | (1.0933) | (0.1696) | (0.1023) | (0.1220) | (0.0951) | (0.0685) |
| 5,-1,2 | 4.5540 | -1.4028 | 1.8114 | 4.8770 | -1.1455 | 1.9313 | 5.0085 | -1.0687 | 1.9668 |
| | (1.5270) | (0.7732) | (0.3485) | (1.1746) | (0.5544) | (0.2326) | (1.2758) | (0.5046) | (0.1903) |
| 1,0,2 | 1.1759 | 0.2696 | 1.9480 | 1.0623 | 0.0886 | 1.9207 | 1.0115 | 0.0078 | 2.0198 |
| | (0.6231)) | (1.0898) | (0.4291) | (0.5966) | (1.0551) | (0.3179) | (0.3250) | (0.1924) | (0.2135) |
| 5,0.25,0.5 | 4.9820 | 0.2220 | 0.4792 | 5.0115 | 0.2426 | 0.4914 | 4.9943 | 0.2457 | 0.4948 |
| | (0.2254) | (0.0996) | (0.0719) | (0.2712) | (0.0716) | (0.0476) | (0.0751) | (0.0444) | (0.0340) |
| 1,1,1 | 1.2206 | 1.1946 | 0.9452 | 1.1730 | 1.1117 | 0.9606 | 0.9647 | 1.0068 | 0.9741 |
| | (0.5303) | (0.5923) | (0.1978) | (0.6568) | (0.6751) | (0.1619) | (0.1463) | (0.0056) | (0.1575) |
| 50,5,5 | 50.0425 | 4.5608 | 4.8112 | 49.9274 | 4.7581 | 4.9854 | 49.9130 | 4.9392 | 4.9764 |
| | (0.5333) | (1.6910) | (0.8088) | (0.9864) | (1.0880) | 4.9854 | 49.9130 | 4.9392 | 4.9764 |
| 4,-50,10 | 4.2033 | -50.6667 | 9.4225 | 4.0616 | -50.5333 | 9.6902 | 4.0208 | -49.8745 | 10.0017 |
| | (1.4918) | (2.5538) | (1.4512) | (1.5078) | (2.4856) | (1.1384) | (1.0730) | (2.2228) | (0.8002) |
| 0.9,0,0.01 | 0.8998 | 0.00042 | 0.009603 | 0.9001 | 0.0001 | 0.0098 | 0.9000 | 0.0000 | 0.0099 |
| | (0.0000) | (0.0032) | (0.0014) | (0.0000) | (0.0019) | (0.0009) | (0.0000) | (0.0014) | (0.0006) |
| 0.9,50,10 | 0.9594 | 50.2913 | 10.2145 | 0.9038 | 50.0927 | 9.8318 | 0.9012 | 50.0459 | 9.9480 |
| | (0.3181) | (1.5910) | (1.3660) | (0.1473) | (1.1424) | (1.2016) | (0.1046) | (0.9813) | (0.7806) |

## 7.2. Simulation of the LXG-W regression model

**Table 5**: Simulation results of LXG-W regression model.

| Censoring rate=0.10 | | n=50 | | | n=200 | | | n=500 | |
|---|---|---|---|---|---|---|---|---|---|
| Parameters | AE | Bias | MSE | AE | Bias | MSE | AE | Bias | MSE |
| $\theta$ | 2.4958 | 0.4958 | 0.7937 | 2.2815 | 0.2815 | 0.4029 | 2.0743 | 0.0743 | 0.1866 |
| $\sigma$ | 0.5077 | 0.0077 | 0.0061 | 0.5090 | 0.0090 | 0.0028 | 0.5048 | 0.0048 | 0.0010 |
| $\beta_0$ | 1.9266 | -0.0734 | 0.4430 | 1.9563 | -0.0437 | 0.2712 | 1.9870 | -0.0130 | 0.1184 |
| $\beta_1$ | 1.9992 | -0.0008 | 0.0215 | 1.9997 | -0.0003 | 0.0054 | 2.0010 | 0.0010 | 0.0020 |
| Censoring rate=0.20 | | n=50 | | | n=200 | | | n=500 | |
| Parameters | AE | Bias | MSE | AE | Bias | MSE | AE | Bias | MSE |
| $\theta$ | 2.3648 | 0.3648 | 0.4329 | 2.1059 | 0.1059 | 0.3109 | 2.0528 | 0.0528 | 0.0722 |
| $\sigma$ | 0.5059 | 0.0059 | 0.0030 | 0.5120 | 0.0120 | 0.0008 | 0.5147 | 0.0147 | 0.0001 |
| $\beta_0$ | 1.9669 | -0.0331 | 0.2197 | 1.9776 | -0.0224 | 0.0840 | 1.8440 | -0.1560 | 0.0080 |
| $\beta_1$ | 2.0047 | 0.0047 | 0.0135 | 2.0037 | 0.0037 | 0.0020 | 1.9994 | -0.0006 | 0.0001 |
| Censoring rate=0.30 | | n=50 | | | n=200 | | | n=500 | |
| Parameters | AE | Bias | MSE | AE | Bias | MSE | AE | Bias | MSE |
| $\theta$ | 2.2911 | 0.2911 | 0.6217 | 2.1508 | 0.1508 | 0.1559 | 1.9377 | -0.0623 | 0.0262 |
| $\sigma$ | 0.5011 | 0.0011 | 0.0018 | 0.5077 | 0.0077 | 0.0026 | 0.5084 | 0.0084 | 0.0014 |
| $\beta_0$ | 2.0236 | 0.0236 | 0.1682 | 2.0620 | 0.0620 | 0.3145 | 1.9769 | -0.0231 | 0.1693 |
| $\beta_1$ | 2.0030 | 0.0030 | 0.0100 | 1.9962 | -0.0038 | 0.0061 | 2.0009 | 0.0009 | 0.0027 |

The simulation study is given to evaluate the MLEs of the parameters of LXG-W regression model. The three censoring rates (10%, 20%, 30%) and sample sizes ($n = 50, 200, 500$) are used. The simulation replication is $N = 1,000$. The lifetimes are generated by using the quantile function of the LXG-W distribution. The following parameter vector is used: $(\theta = 2, \sigma = 0.5, \beta_0 = 2, \beta_1 = 2)$. For each generated sample sizes, the biases, average of estimates (AEs) and MSEs are calculated. The simulation results are reported in Table 5. As seen from the results, the estimated biases and MSEs are near the desired value, zero. Moreover, the estimated AEs are closer the nominal values which indicates that the estimates are stable. The similar results can be also obtained for different parameter vector.

## 8. DATA ANALYSIS

In this section, we provide three applications to real data to illustrate the importance and flexibility of the XG-W, XG-N and LXG-W distributions. The Akaike Information Criteria (AIC), Bayesian information criterion (BIC) and Kolmogorov-Smirnov (K-S) statistic are used to compare the fitted distributions. All computations are performed using the **maxLik** routine in the R software.

### 8.1. Application 1: Glass fibres data

The first data set represents the strength of 1.5 cm glass fibres measured at National physical laboratory, England (Smith and Naylor [30], 1987). These data have been analyzed by Korkmaz and Genç [18] (2017). We shall compare the fits of the XG-W, Kumaraswamy-Weibull (Kw-W) (Cordeiro and de Castro [9], 2011), beta-Weibull (BW) (Famoye *et al.* [12], 2005), Lindley-Weibull (LW) (Cakmakyapan and Ozel [6], 2016), EW (Mudholkar and Srivastava [20], 1993) and odd log-logistic-Weibull (OLL-W) (Gleaton and Lynch [13], 2010; da Cruz *et al.* [10], 2016) distributions to the glass fibres data. The cdfs of the Kw-W, BW, LW, EW and OLL-W models (for $x > 0$) are given by

$$\begin{aligned}
F(x) &= 1-\left(1-\left\{1-\exp\left[-(ax)^b\right]\right\}^\gamma\right)^\eta,\\
F(x) &= \frac{1}{B(\gamma,\eta)}B\left(1-\exp\left[-(ax)^b\right],\gamma,\eta\right),\\
F(x) &= 1-\exp\left[-\theta\,(ax)^\beta\right]\left[1+\frac{\theta}{\theta+1}\,(ax)^\beta\right],\\
F(x) &= \left\{1-\exp\left[-(\alpha x)^\beta\right]\right\}^\theta,\\
F(x) &= \frac{\left\{1-\exp\left[-(ax)^b\right]\right\}^\theta}{\left\{1-\exp\left[-(ax)^b\right]\right\}^\theta+\exp\left[-\theta(ax)^b\right]},
\end{aligned}$$

respectively, where $B(\gamma, \eta)$ is the complete beta function and the parameters of the above densities are all positive real numbers. The MLEs (and their corresponding standard errors in parentheses) of the parameters, AIC, BIC and K-S statistics for the above fitted models are displayed in Table 6. The values in this table indicate that the XG-W model provides a better fit than the other fitted models because proposed model has the smallest values of the AIC, BIC and K-S statistics and has the largest p-value of the K-S statistic.

**Table 6**: The MLEs (standard errors in parentheses), AIC, BIC and K-S (with p-values in $\{\cdot\}$) statistics for glass fibres data.

| Model | $\widehat{\gamma}$ | $\widehat{\eta}$ | $\widehat{\theta}$ | $\widehat{a}$ | $\widehat{b}$ | AIC | BIC | K-S |
|---|---|---|---|---|---|---|---|---|
| XG-W | - | - | 0.4392 | 0.8952 | 4.6911 | 31.4342 | 37.8636 | 0.1210 |
| | - | - | (0.0421) | (0.0250) | (0.1330) | | | {0.3151} |
| Kw-W | 0.7910 | 112.6514 | - | 0.2702 | 7.2790 | 38.3943 | 46.9669 | 0.1524 |
| | (0.1088) | (15.9962) | - | (0.0226) | (0.5671) | | | {0.1073} |
| BW | 0.6207 | 120.6149 | - | 0.3051 | 7.7653 | 37.1752 | 45.7477 | 0.1455 |
| | (0.0947) | (0.3733) | - | (0.0088) | (0.1522) | | | {0.1388} |
| LW | - | - | 117.0336 | 0.2698 | 5.7804 | 36.4135 | 42.8429 | 0.1522 |
| | - | - | (9.4031) | (0.0215) | (0.5800) | | | {0.1078} |
| EW | - | - | 0.6713 | 0.5821 | 7.2841 | 35.3510 | 41.7804 | 0.1462 |
| | - | - | (0.2876) | (0.0332) | (2.0252) | | | {0.1351} |
| OLL-W | - | - | 0.9438 | 0.6159 | 6.0252 | 36.3736 | 42.8030 | 0.1537 |
| | - | - | (0.2655) | (0.0163) | (1.3273) | | | {0.1018} |

## 8.2. Application 2: Leukemia data

The second data set represents the lifetimes in days of 40 patients suffering from leukemia from one of the Ministry of Health Hospitals in Saudi Arabia (Abouammoh *et al.* [1], 1994). The data have been analyzed by Sarhan *et al.* [24] (2013). We compare the XG-N distribution with the Kumaraswamy-normal (Kw-N) (Cordeiro and de Castro [9], 2011), power-normal (PN) (Gupta and Gupta [15], 2008), logistic-normal (L-N) (Tahir *et al.* [31], 2016) and odd log-logistic-normal (OLL-N) (Braga *et al.* [5], 2016) distributions. The cdfs of the Kw-N, PN, L-N, and OLL-N models are given by

$$\begin{aligned} F(x) &= 1-\left\{1-\left[\mathbf{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{\gamma}\right\}^{\eta}, \\ F(x) &= \left[\mathbf{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{\theta}, \\ F(x) &= \left\{1+\left[1-\mathbf{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{-\theta}\right\}^{-1}, \\ F(x) &= \frac{\left[\mathbf{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{\theta}}{\left[\mathbf{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{\theta}+\left[1-\mathbf{\Phi}\left(\frac{x-\mu}{\sigma}\right)\right]^{\theta}}, \end{aligned}$$

respectively, where $x, \mu \in \Re, \gamma, \eta, \sigma > 0$ and $\mathbf{\Phi}(\cdot)$ is the cdf of the standard normal distribution.

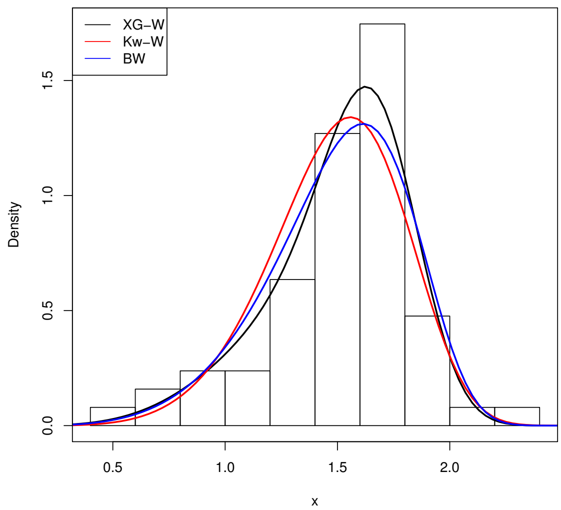
Table 7 lists the MLEs (and their standard errors) of the parameters and the K-S statistic for the fitted models. The figures in this table reveal that the XG-N distribution has the smallest values of the AIC, BIC and K-S statistics and has the largest p-value of the

K-S statistic. Therefore, we can conclude that the XG-N distribution could be chosen as the most adequate model for this data set.

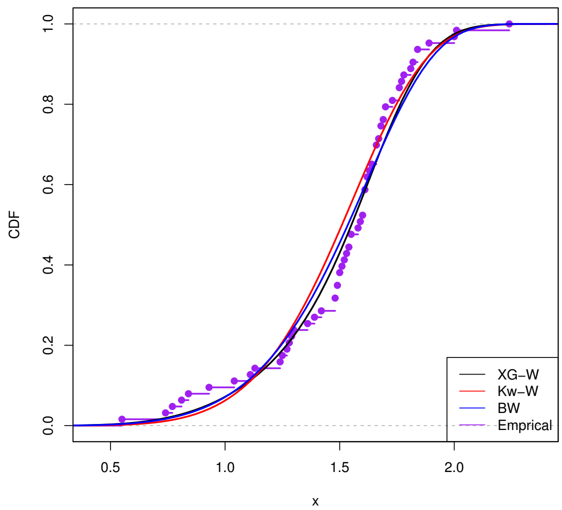
**Table 7**: The MLEs (standard errors in parentheses) AIC, BIC and K-S (with p-values in $\{\cdot\}$) statistic for leukemia data.

| Model | $\widehat{\gamma}$ | $\widehat{\eta}$ | $\widehat{\theta}$ | $\widehat{\mu}$ | $\widehat{\sigma}$ | AIC | BIC | K-S |
|---|---|---|---|---|---|---|---|---|
| XG-N | - | - | 0.6892 | 662.6324 | 609.7157 | 609.7157 | 614.7824 | 0.0825 |
| | - | - | (0.0717) | (0.9018) | (1.5488) | | | {0.9484} |
| Kw-N | 0.8320 | 0.2217 | - | 614.6680 | 294.0911 | 616.3179 | 623.0734 | 0.1314 |
| | (1.0023) | (0.2217) | - | (0.5583) | (1.7209) | | | {0.4942} |
| PN | - | - | 5.5078 | 189.8173 | 776.1012 | 615.2070 | 620.2736 | 0.1196 |
| | - | - | (0.8745) | (4.2845) | (0.8745) | | | {0.6163} |
| L-N | - | - | 4.4833 | 719.6505 | 1329.4302 | 617.0567 | 622.1234 | 0.1022 |
| | - | - | (0.5873) | (0.5873) | (4.1943) | | | {0.7976} |
| OLL-N | - | - | 36.6070 | 1169.5520 | 16331.9508 | 614.8475 | 619.9142 | 0.0869 |
| | - | - | (5.9316) | (4.1943) | (7.2647) | | | {0.9228} |

The histogram of both data sets and the estimated pdfs and cdfs of the XG-W and XG-N models and their competitive models are displayed in Figures 4 and 5, respectively. It is clear from these plots that the XG-W and XG-N models provide the best fits to both data sets.

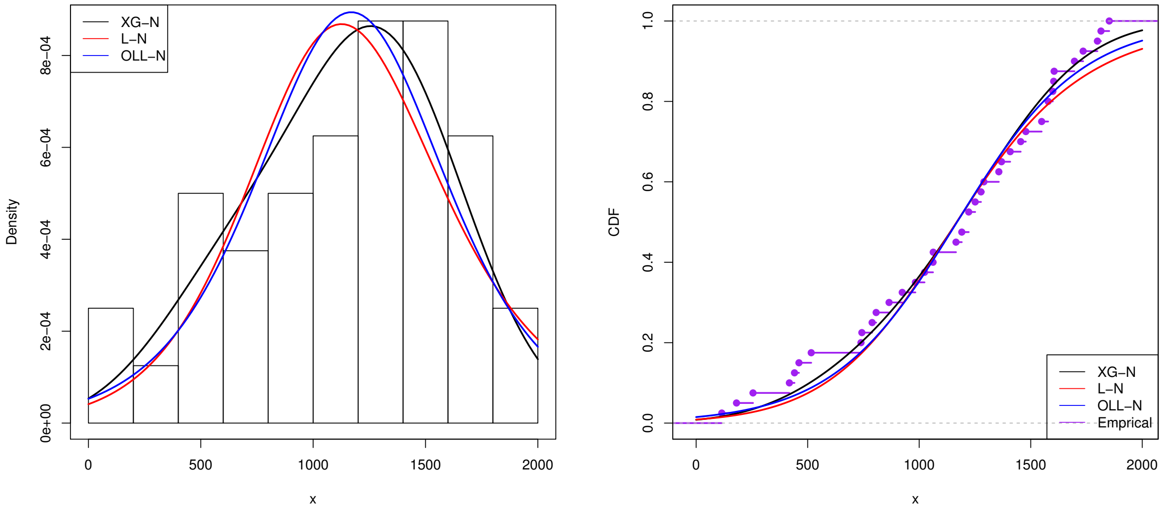



**(a)** Fitted pdfs for data set I.




**(b)** Fitted cdfs for data set I.

**Figure 4**: Plots of the estimated pdfs and cdfs of the XG-W distribution and other competitive models.




**(a)** Fitted pdfs for data set II. **(b)** Fitted cdfs for data set II.

**Figure 5**: Plots of the estimated pdfs and cdfs of the XG-N distribution and other competitive models.

## 8.3. Application 3: Diabetic retinopathy study

We consider a data set analyzed by Huster *et al.* [17] (1989) which represents patients with diabetic retinopathy in both eyes and 20/100 or better visual acuity for both eyes were eligible for the study. The patients were followed for two consecutively completed 4 month follow-ups and the endpoint was the occurrence of visual acuity less than 5/200. We choose only the treatment time. A 50% sample of the high-risk patients defined by diabetic retinopathy criteria was taken for the data set ($n$ =197) and the percentage of censored observations was 72.4%. The variables involved in the study are: $t_i$ – failure time for the treatment (in min); censoring indicator (0 = censoring, 1 = lifetime observed); $x_{i1}$ – age (0 = patient is an adult diabetic, 1 = patient is a juvenile diabetic). The below regression structure is fitted by LXG-W regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma\, z_i,$$

where the rv $Y_i$ has the LXG-W distribution (6.1) for $i = 1, ..., 197$. The statistical software **R** is used to estimate the unknown model parameters by MLE approach. The **optim** function of **R** software is used to minimize the minus of log-likelihood function, given in (6.4). The initial values of the parameters are taken from the fitted LXG-E regression model (with $\sigma = 1$). The MLEs of the parameters of LXG-W regression model (approximate standard errors and p-values in parentheses) are: $\hat{\theta} = 1.7187\,(1.8739)$, $\hat{\sigma} = 1.2085\,(0.1518)$, $\hat{\beta}_0 = 4.2902\,(1.9308)\,(0.0068)$ and $\hat{\beta}_1 = 0.6474\,(0.3755)\,(0.0215)$. The explanatory variable $x_1$ is found statistically significant at the 5% significance level. In order to assess the validity of the fitted regression model, the estimated survival functions of the LXG-W regression model and empirical one are displayed in Figure 6. As seen from this figure, the LXG-W regression model provides substantial fit to these data.

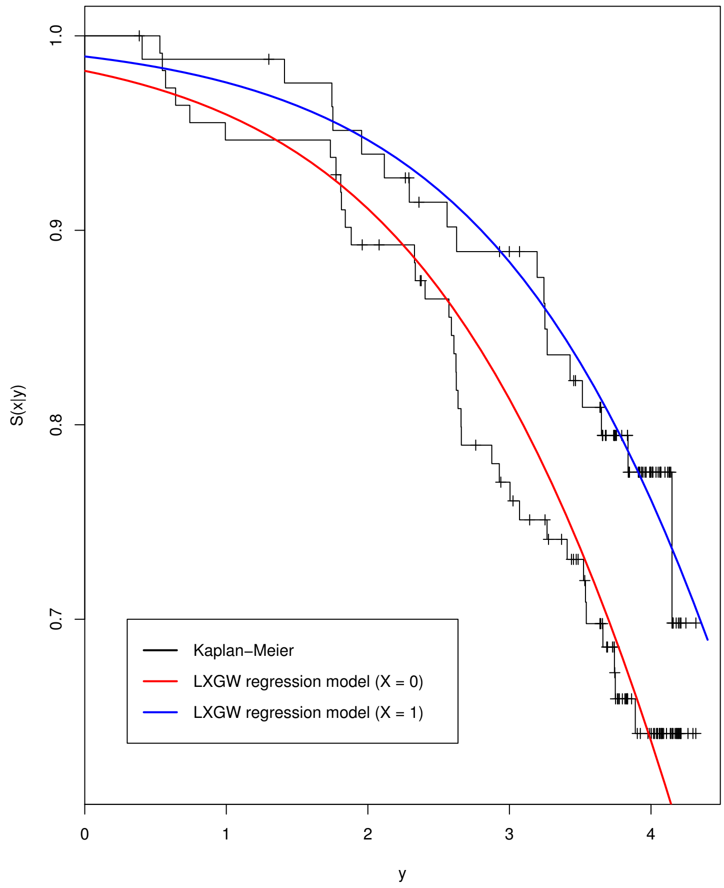



**Figure 6**: Estimated survival function by fitting the LXG-W regression model and the empirical survival for each level of the diabetic retinopathy study.

## 9. CONCLUSIONS

A new class of distributions called the xgamma-G family with one extra positive parameter is introduced and studied. We provide some mathematical properties of the new family including ordinary and incomplete moments, quantile and generating functions and mean deviations. The maximum likelihood method is used for estimating the model parameters. We assess of the performance of the maximum likelihood estimators in terms of biases and mean squared errors by means of two simulation studies. We also introduced a new linear regression model based on the logarithm of the xgamma random variable for uncensored and censored data. We prove that the special models of the proposed family provide consistently better fits than other competitive models by means of three real data sets.

## REFERENCES


[1] Abouammoh, A.M.; Abdulghani, S.A. and Qamber, I.S. (1994). On partial orderings and testing of new better than renewal used classes, *Reliability Engineering & System Safety*, **43**, 37–41.

[2] Altun, E. and Hamedani, G.G. (2018). The log-xgamma distribution with inference and application, *Journal de la Société Française de Statistique*, **159**, 40–55.

[3] Alzaatreh, A.; Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions, *Metron*, **71**, 63–79.

[4] Biçer, H.D. (2019). Properties and inference for a new class of XGamma distributions with an application, *Mathematical Sciences*, **4**, 335–346.

[5] Braga, A.S.; Cordeiro, G.M.; Ortega, E.M.M. and da Cruz, J.N. (2016). The odd log–logistic normal distribution: theory and applications in analysis of experiments, *Journal of Statistical Theory and Practice*, **10**, 311–335.

[6] Cakmakyapan, S. and Ozel, G. (2016). The Lindley family of distributions: properties and applications, *Hacettepe Journal of Mathematics and Statistics*, **46**, 1113–1137.

[7] Cordeiro, G.M.; Afify, A.Z.; Yousof, H.M.; Pescim, R.R. and Aryal, G.R. (2017). The exponentiated Weibull-H family of distributions: theory and applications, *Mediterranean Journal of Mathematics*, **14**, 155.

[8] Cordeiro, G.M.; Alizadeh, M. and Diniz Marinho, P.R. (2016). The type I half-logistic family of distributions, *J. Stat. Comput. Simul.*, **86**, 707–728.

[9] Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions, *J. Stat. Comput. Simul.*, **81**, 883–898.

[10] da Cruz, J.N.; Ortega, E.M.M. and Cordeiro, G.M. (2016). The log-odd log-logistic Weibull regression model: modelling, estimation, influence diagnostics and residual analysis, *J. Stat. Comput. Simul.*, **86**, 1516–1538.

[11] Eugene, N.; Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, *Commun. Stat. Theory Methods*, **31**, 497–512.

[12] Famoye, F.; Lee, C. and Olumolade, O. (2005). The beta-Weibull distribution, *Journal of Statistical Theory and Applications*, **4**, 121–136.

[13] Gleaton, J.U. and Lynch (2010). Extended generalized log-logistic families of lifetime distributions with an application, *Journal of Probability and Statistical Science*, **8**, 1–17.

[14] Gradshteyn, I.S. and Ryzhik, I.M. (2002). *Table of Integrals, Series, and Products*, Academic Press, San Diego, CA.

[15] Gupta, R.D. and Gupta. R.C. (2008). Analyzing skewed data by power normal model, *Test*, **17**, 197–210.

[16] Gupta, R.D. and Kundu, D. (1999). Generalized Exponential distributions, *Australian and New Zealand Journal of Statistics*, **41**, 173–188.

[17] Huster, W.J.; Brookmeyer, R. and Self, S.G. (1989). Modelling paired survival data with covariates, *Biometrics*, **45**, 145–156.

[18] Korkmaz, M.Ç. and Genç, A.I. (2017). A new generalized two-sided class of distributions with an emphasis on two-sided generalized normal distribution, *Comm. Stat. Simulation and Computation*, **46**, 1441–1460.

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

[20] Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analysing bathtub failure rate data, *IEEE Transactions on Reliability*, **42**, 299–302.

[21] Mudholkar, G.S.; Srivastava, D.K. and Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data, *Technometrics*, **37**, 436–445.

[22] Nadarajah, S. and Gupta, A.K. (2007). The exponentiated gamma distribution with application to drought data, *Calcutta Statistical Association Bulletin*, **59**, 29–54.

[23] Nofal, Z.M.; Afify, A.Z.; Yousof, H.M. and Cordeiro, G.M. (2017). The generalized transmuted-G family of distributions, *Commun. Stat. Theory Methods*, **46**, 4119–4136.

[24] Sarhan, A.M.; Ahmad, A.A. and Alasbahi, A.I. (2013). Exponentiated generalized linear exponential distribution, *Applied Mathematical Modelling*, **37**, 2838–2849.

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

[26] Sen, S.; Chandra, N. and Maiti, S.S. (2017). The weighted xgamma distribution: properties and application, *Journal of Reliability and Statistical Studies*, **10**, 43–58.

[27] Sen, S.; Chandra, N. and Maiti, S.S. (2018a). On properties and applications of a two-parameter xgamma distribution, *Journal of Statistical Theory and Applications*, **4**, 674–685.

[28] Sen, S.; Chandra, N. and Maiti, S.S. (2018b). Survival estimation in xgamma distribution under progressively type-II right censored scheme, *Model Assisted Statistics and Applications*, **2**, 107–121.

[29] Sen, S.; Maiti, S.S. and Chandra, N. (2016). The xgamma distribution: statistical properties and application, *Journal of Modern Applied Statistical Methods*, **15**, 774–788.

[30] Smith, R.L. and Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution, *Appl. Statist.*, **36**, 358–369.

[31] Tahir, M.H.; Cordeiro, G.M.; Alzaatreh, A.; Mansoor, M. and Zubair, M. (2016). The logistic-X family of distributions and its applications, *Comm. Stat. Theory and Methods*, **45**, 7326–7349.

[32] Yousof, H.M.; Afify, A.Z.; Hamedani, G.G. and Aryal, G. (2016). The Burr X generator of distributions for lifetime data, *Journal of Statistical Theory and Applications*, **16**, 1–19.

[33] Zografos, K. and Balakrishnan, N. (2009). On families of beta and generalized gamma-generated distributions and associated inference, *Statistical Methodology*, **6**, 344–362.