Hacettepe Journal of Mathematics and Statistics
Volume 45 (4) (2016), 1175 – 1202

The beta odd log-logistic generalized family of distributions

Gauss M. Cordeiro[*], Morad Alizadeh[†], M. H. Tahir[‡,§], M. Mansoor[¶], Marcelo Bourguignon[‖] and G. G. Hamedani[**]

Abstract

We introduce a new family of continuous models called the *beta odd log-logistic generalized family* of distributions. We study some of its mathematical properties. Its density function can be symmetrical, left-skewed, right-skewed, reversed-J, unimodal and bimodal shaped, and has constant, increasing, decreasing, upside-down bathtub and J-shaped hazard rates. Five special models are discussed. We obtain explicit expressions for the moments, quantile function, moment generating function, mean deviations, order statistics, Rényi entropy and Shannon entropy. We discuss simulation issues, estimation by the method of maximum likelihood, and the method of minimum spacing distance estimator. We illustrate the importance of the family by means of two applications to real data sets.

Keywords: Beta-G family, characterizations, exponential distribution, generalized family, log-logistic distribution, maximum likelihood, method of minimum spacing distance.

2000 AMS Classification: 60E05; 62E10; 62N05.

Received : 13.12.2014 *Accepted* : 30.06.2015 *Doi* : 10.15672/HJMS.20157311545

[*]Department of Statistics, Federal University of Pernambuco, 50740-540, Recife, PE, Brazil, Email: gauss@de.ufpe.br, gausscordeiro@gmail.com

[†]Department of Statistics, Persian Gulf University of Bushehr, Bushehr 751691-3798, Iran, Email: moradalizadeh78@gmail.com

[‡]Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan, Email: mtahir.stat@gmail.com, mht@iub.edu.pk

[§]Corresponding Author.

[¶]Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan, Email: mansoor.abbasi143@gmail.com

[‖]Department of Statistics, Federal University of Pernambuco, 50740-540, Recife, PE, Brazil, Email: m.p.bourguignon@gmail.com

[**]Department of Mathematics, Statistics and Computer Science, Marquette University, WI 53201-1881, Milwaukee, USA, Email: g.hamedani@mu.edu

1. Introduction

There has been an increased interest in defining new generators or generalized (G) classes of univariate continuous distributions by adding shape parameter(s) to a baseline model. The extended distributions have attracted several statisticians to develop new models because the computational and analytical facilities available in programming softwares like R, Maple and Mathematica can easily tackle the problems involved in computing special functions in these extended models. Several mathematical properties of the extended distributions may be easily explored using mixture forms of the exponentiated-G ("exp-G" for short) distributions. The addition of parameter(s) has been proved useful in exploring skewness and tail properties, and also for improving the goodness-of-fit of the generated family. The well-known generators are the following: beta-G by Eugene et al. [15] and Jones [29], Kumaraswamy-G (Kw-G) by Cordeiro and de Castro [10], McDonald-G (Mc-G) by Alexander et al. [1], gamma-G type 1 by Zografos and Balakrishnan [53] and Amini et al. [6], gamma-G type 2 by Ristić and Balakrishnan [44], odd-gamma-G type 3 by Torabi and Montazari [50], logistic-G by Torabi and Montazari [51], odd exponentiated generalized (odd exp-G) by Cordeiro et al. [12], transformed-transformer (T-X) (Weibull-X and gamma-X) by Alzaatreh et al. [3], exponentiated T-X by Alzaghal et al. [5], odd Weibull-G by Bourguignon et al. [7], exponentiated half-logistic by Cordeiro et al. [13], logistic-X by Tahir et al. [47], T-X{Y}-quantile based approach by Aljarrah et al. [2] and T-R{Y} by Alzaatreh et al. [4].

This paper is organized as follows. In Section 2, we define the *beta odd log-logistic generalized* (BOLL-G) family. Some of its special cases are presented in Section 3. In Section 4, we derive some of its mathematical properties such as the asymptotics, shapes of the density and hazard rate functions, mixture representation for the density, quantile function (qf), moments, moment generating function (mgf), mean deviations, explicit expressions for the Rényi and Shannon entropies and order statistics. Section 5 deals with some characterizations of the new family. Estimation of the model parameters and simulation using maximum likelihood and the method of minimum spacing distance are discussed in Section 6. In Section 7, we illustrate the importance of the new family by means of two applications to real data. The paper is concluded in Section 8.

2. The odd log-logistic and beta odd log-logistic families

The log-logistic (LL) distribution is widely used in practice and it is an alternative to the log-normal model since it presents a hazard rate function (hrf) that increases, reaches a peak after some finite period and then declines gradually. Its properties make the distribution an attractive alternative to the log-normal and Weibull models in the analysis of survival data. If T has a logistic distribution, then $Z = \mathrm{e}^{\mathrm{T}}$ has the LL distribution. Unlike the more commonly used Weibull distribution, the LL distribution has a non-monotonic hrf which makes it suitable for modeling cancer survival data.

The *odd log-logistic (OLL) family* of distributions was originally developed by Gleaton and Lynch [18, 19]; they called this family the *generalized log-logistic (GLL) family*. They showed that:

– the set of GLL transformations form an Abelian group with the binary operation of composition;

– the transformation group partitions the set of all lifetime distributions into equivalence classes, so that any two distributions in an equivalence class are related through a GLL transformation;

– either every distribution in an equivalence class has a moment generating function, or none does;

– every distribution in an equivalence class has the same number of moments;

– each equivalence class is linearly ordered according to the transformation parameter, with larger values of this parameter corresponding to smaller dispersion of the distribution about the common class median; and

– within an equivalence class, the Kullback-Leibler information is an increasing function of the ratio of the transformation parameters.

In addition, Gleaton and Rahman obtained results about the distributions of the MLE's of the parameters of the distribution. Gleaton and Rahman [20, 21] showed that for distributions generated from either a 2-parameter Weibull distribution or a 2-parameter inverse Gaussian distribution by a GLL transformation, the joint maximum likelihood estimators of the parameters are asymptotically normal and efficient, provided the GLL transformation parameter exceeds 3.

Given a continuous baseline cumulative distribution function (cdf) $G(x;\boldsymbol{\xi})$ with a parameter vector $\boldsymbol{\xi}$, the cdf of the OLL-G family (by integrating the LL density function with an additional shape parameter $c>0$) is given by

$$F_{\text{OLL-G}}(x)=\int_0^{G(x;\xi)/\overline{G}(x;\xi)}\frac{c\,t^{c-1}}{(1+t^c)^2}\,dt=\frac{G(x;\boldsymbol{\xi})^c}{G(x;\boldsymbol{\xi})^c+\overline{G}(x;\boldsymbol{\xi})^c}. \tag{2.1}$$

If $c>1$, the hrf of the OLL-G random variable is unimodal and when $c=1$ it decreases monotonically. The fact that its cdf has closed-form is particularly important for analysis of survival data with censoring.

We can write

$$c=\frac{\log\left[F(x;\boldsymbol{\xi})/\overline{F}(x;\boldsymbol{\xi})\right]}{\log\left[G(x;\boldsymbol{\xi})/\overline{G}(x;\boldsymbol{\xi})\right]}\qquad\text{and}\qquad \overline{G}(x;\boldsymbol{\xi})=1-G(x;\boldsymbol{\xi}).$$

Here, the parameter c represents the quotient of the log-odds ratio for the generated and baseline distributions.

The probability density function (pdf) corresponding to (2.1) is

$$f_{\text{OLL-G}}(x)=\frac{c\,g(x;\boldsymbol{\xi})\,\left\{G(x;\boldsymbol{\xi})\,\overline{G}(x;\boldsymbol{\xi})\right\}^{c-1}}{\left\{G(x;\boldsymbol{\xi})^c+\overline{G}(x;\boldsymbol{\xi})^c\right\}^2}. \tag{2.2}$$

In this paper, we propose a new extension of the OLL-G family. Based on a baseline cdf $G(x;\boldsymbol{\xi})$ depending on a parameter vector $\boldsymbol{\xi}$, survival function $\overline{G}(x;\boldsymbol{\xi})=1-G(x;\boldsymbol{\xi})$ and pdf $g(x;\boldsymbol{\xi})$, we define the cdf of the BOLL-G family of distributions (for $x\in\mathbb{R}$) by

$$F(x)=F(x;a,b,c,\boldsymbol{\xi})=\frac{1}{B(a,b)}\,B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right), \tag{2.3}$$

where $a>0$, $b>0$ and $c>0$ are three additional shape parameters, $B(z;a,b)=\int_0^z w^{a-1}(1-w)^{b-1}\mathrm{d}w$ is the incomplete beta function, $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function and $\Gamma(a)=\int_0^\infty t^{a-1}\,\mathrm{e}^{-t}\,\mathrm{d}t$ is the gamma function. We also adopt the notation $I_z(a,b)=B(z;a,b)/B(a,b)$.

The pdf and hrf corresponding to (2.3) are, respectively, given by

$$f(x)\;=\;f(x;a,b,c,\boldsymbol{\xi})=\frac{c\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{ac-1}\overline{G}(x;\boldsymbol{\xi})^{bc-1}}{B(a,b)\left\{G(x;\boldsymbol{\xi})^c+\overline{G}(x;\boldsymbol{\xi})^c\right\}^{a+b}} \tag{2.4}$$

and

$$h(x)\;=\;\frac{c\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{ac-1}\overline{G}(x;\boldsymbol{\xi})^{bc-1}}{\left\{G(x;\boldsymbol{\xi})^c+\overline{G}(x;\boldsymbol{\xi})^c\right\}^{a+b}\left\{B(a,b)-B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right)\right\}}. \tag{2.5}$$

Clearly, if we take $G(x) = x/(1+x)$, equation (2.3) becomes the beta log-logistic distribution. The family (2.4) contains some sub-families listed in Table 1. The baseline G distribution is a basic exemplar of (2.4) when $a = b = c = 1$. Hereafter, $X \sim$ BOLL-G$(a, b, c, \boldsymbol{\xi})$ denotes a random variable having density function (2.4). We can omit the parameters in the pdf's and cdf's.

Table 1: Some special models of the BOLL-G family.

a	b	c	$G(x)$	Reduced distribution
-	-	1	$G(x)$	Beta-G family (Eugene et al. [15])
1	1	-	$G(x)$	Odd log-logistic family (Gleaton and Lynch[19])
1	-	1	$G(x)$	Proportional hazard rate family (Gupta et al. [26])
-	1	1	$G(x)$	Proportional reversed hazard rate family (Gupta and Gupta [25])
1	1	1	$G(x)$	$G(x)$

The BOLL-G family can easily be simulated by inverting (2.3) as follows: if V has a beta (a, b) distribution, then the random variable X can be obtained from the baseline qf, say $Q_G(u) = G^{-1}(u)$. In fact, the random variable

$$X = Q_G\Big[\frac{V^{\frac{1}{c}}}{V^{\frac{1}{c}} + (1-V)^{\frac{1}{c}}}\Big] \tag{2.6}$$

has density function (2.4).

3. Some special models

Here, we present some special models of the BOLL-G family.

3.1. The BOLL-exponential (BOLL-E) distribution. The pdf and cdf of the exponential distribution with scale parameter $\alpha > 0$ are given by $g(x;\alpha) = \alpha\,\mathrm{e}^{-\alpha\,x}$ and $G(x;\alpha) = 1 - \mathrm{e}^{-\alpha\,x}$, respectively. Inserting these expressions in (2.4) gives the BOLL-E pdf

$$f(x;a,b,c,\alpha) = \frac{c\,\alpha\;\mathrm{e}^{-\alpha\,b\,x}\left\{1-\mathrm{e}^{-\alpha\,x}\right\}^{ac-1}}{B(a,b)\,[\{1-\mathrm{e}^{-\alpha\,x}\}^{c} + \mathrm{e}^{-c\alpha\,x}]^{a+b}}.$$

3.2. The BOLL-normal (BOLL-N) distribution. The BOLL-N distribution is defined from (2.4) by taking $G(x;\boldsymbol{\xi}) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ and $g(x;\boldsymbol{\xi}) = \sigma^{-1}\,\phi\left(\frac{x-\mu}{\sigma}\right)$ for the cdf and pdf of the normal distribution with parameters μ and σ^2, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively, and $\boldsymbol{\xi} = (\mu, \sigma^2)$. The BOLL-N pdf is given by

$$f(x;a,b,c,\mu,\sigma^2) = \frac{c\,\phi(\frac{x-\mu}{\sigma})\;\left\{\Phi\left(\frac{x-\mu}{\sigma}\right)\right\}^{ac-1}\;\left\{1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right\}^{bc-1}}{\sigma B(a,b)\left[\left\{\Phi\left(\frac{x-\mu}{\sigma}\right)\right\}^{c} + \left\{1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right\}^{c}\right]^{a+b}}, \tag{3.1}$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is a location parameter and $\sigma > 0$ is a scale parameter.

We can denote by $X \sim$ BOLL-N(a, b, c, μ, σ^2) a random variable having pdf (3.1).

3.3. The BOLL-Lomax (BOLL-Lx) distribution. The pdf and cdf of the Lomax distribution with scale parameter $\beta > 0$ and shape parameter $\alpha > 0$ are given by

$g(x;\alpha,\beta) = (\alpha/\beta)\,[1+(x/\beta)]^{-(\alpha+1)}$ and $G(x;\alpha,\beta) = 1-[1+(x/\beta)]^{-\alpha}$, respectively. The BOLL-Lx pdf follows by inserting these expressions in (2.4) as

$$f(x;a,b,c,\alpha,\beta) = \frac{\frac{c\,\alpha}{\beta}\left\{1+\left(\frac{x}{\beta}\right)\right\}^{-(\alpha+1)}\left\{1+\left(\frac{x}{\beta}\right)\right\}^{-\alpha(ac-1)}}{B(a,b)\left[\left\{1-\left[1+\left(\frac{x}{\beta}\right)\right]^{-\alpha}\right\}^{c}+\left\{1+\left(\frac{x}{\beta}\right)\right\}^{-\alpha\,c}\right]^{a+b}}.$$

3.4. The BOLL-Weibull (BOLL-W) distribution. The pdf and cdf of the Weibull distribution with scale parameter $\alpha > 0$ and shape parameter $\beta > 0$ are given by $g(x;\alpha,\beta) = \alpha\beta x^{\beta-1}\,\mathrm{e}^{-\alpha x^{\beta}}$ and $G(x;\alpha,\beta) = 1-\mathrm{e}^{-\alpha x^{\beta}}$, respectively. Inserting these expressions in (2.4) yields the BOLL-W pdf

$$f(x;a,b,c,\alpha,\beta) = \frac{c\,\alpha\,\beta\,x^{\beta-1}\;\mathrm{e}^{-b\,c\,\alpha\,x^{\beta}}\left\{1-\mathrm{e}^{-\alpha\,x^{\beta}}\right\}^{ac-1}}{B(a,b)\left[\left\{1-\mathrm{e}^{-\alpha\,x^{\beta}}\right\}^{c}+\left\{\mathrm{e}^{-\alpha\,x^{\beta}}\right\}^{c}\right]^{a+b}}.$$

3.5. The BOLL-Gamma (BOLL-Ga) distribution. Consider the gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, where the pdf and cdf (for $x > 0$) are given by

$$g(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\,x^{\alpha-1}\,\mathrm{e}^{-\beta x} \quad \text{and} \quad G(x;\alpha,\beta) = \frac{\gamma(\alpha,\beta\,x)}{\Gamma(\alpha)},$$

where $\gamma(\alpha,\beta\,x) = \int_0^{\beta\,x} t^{\alpha-1}\,\mathrm{e}^{-t}\,\mathrm{d}t$ is the incomplete gamma function. Inserting these expressions in equation (2.4), the BOLL-Ga density function follows as

$$f(x;a,b,c,\alpha,\beta) = \frac{c\,\beta^{\alpha}\,x^{\alpha-1}\,\mathrm{e}^{-\beta x}\left\{\frac{\gamma(\alpha,\beta\,x)}{\Gamma(\alpha)}\right\}^{ac-1}\left\{1-\frac{\gamma(\alpha,\beta\,x)}{\Gamma(\alpha)}\right\}^{bc-1}}{\Gamma(\alpha)\,B(a,b)\left[\left\{\frac{\gamma(\alpha,\beta\,x)}{\Gamma(\alpha)}\right\}^{c}+\left\{1-\frac{\gamma(\alpha,\beta\,x)}{\Gamma(\alpha)}\right\}^{c}\right]^{a+b}}.$$

In Figures 1 and 2, we display some plots of the pdf and hrf of the BOLL-E, BOLL-N and BOLL-Lx distributions for selected parameter values. Figure 1 reveals that the BOLL-E, BOLL-N and BOLL-Lx densities generate various shapes such as symmetrical, left-skewed, right-skewed, reversed-J, unimodal and bimodal. Also, Figure 2 shows that these models can produce hazard rate shapes such as constant, increasing, decreasing, J and upside-down bathtub. This fact implies that the BOLL-G family can be very useful for fitting data sets with various shapes.

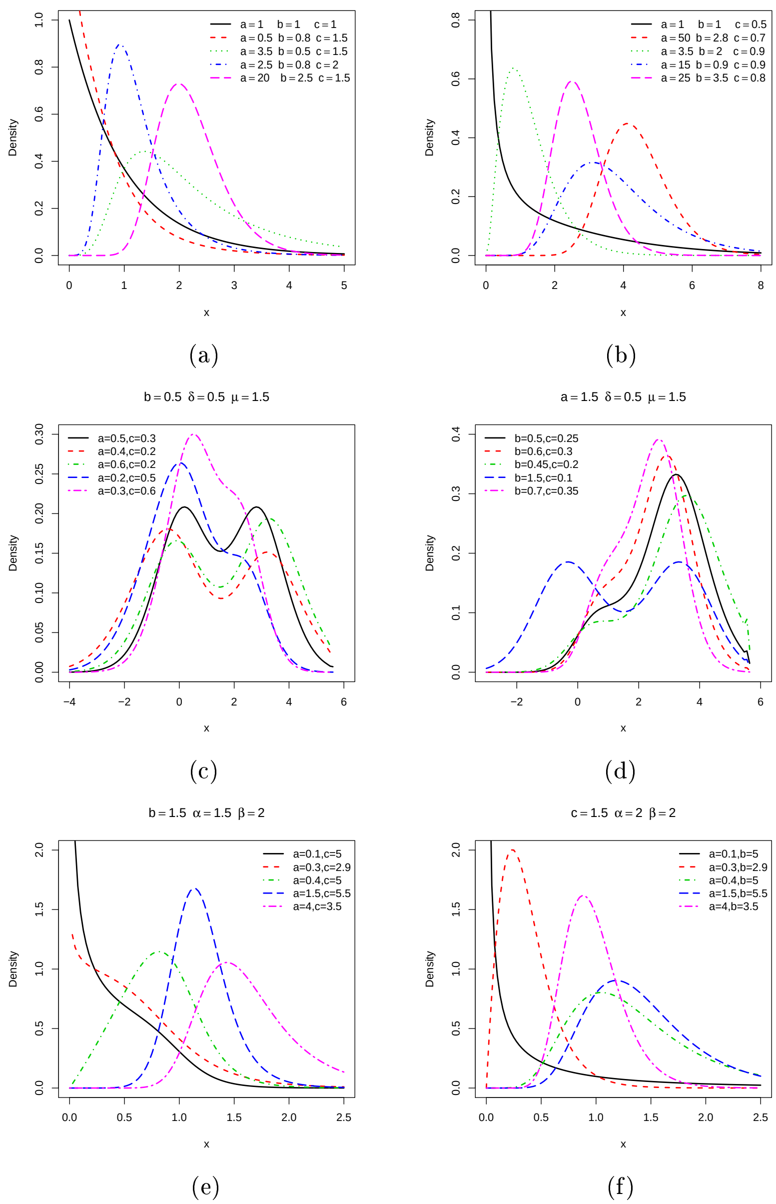


Figure 1. Density plots (a)-(b) of the BOLL-E model, (c)-(d) of the BOLL-N model and (e)-(f) of the BOLL-Lx model.

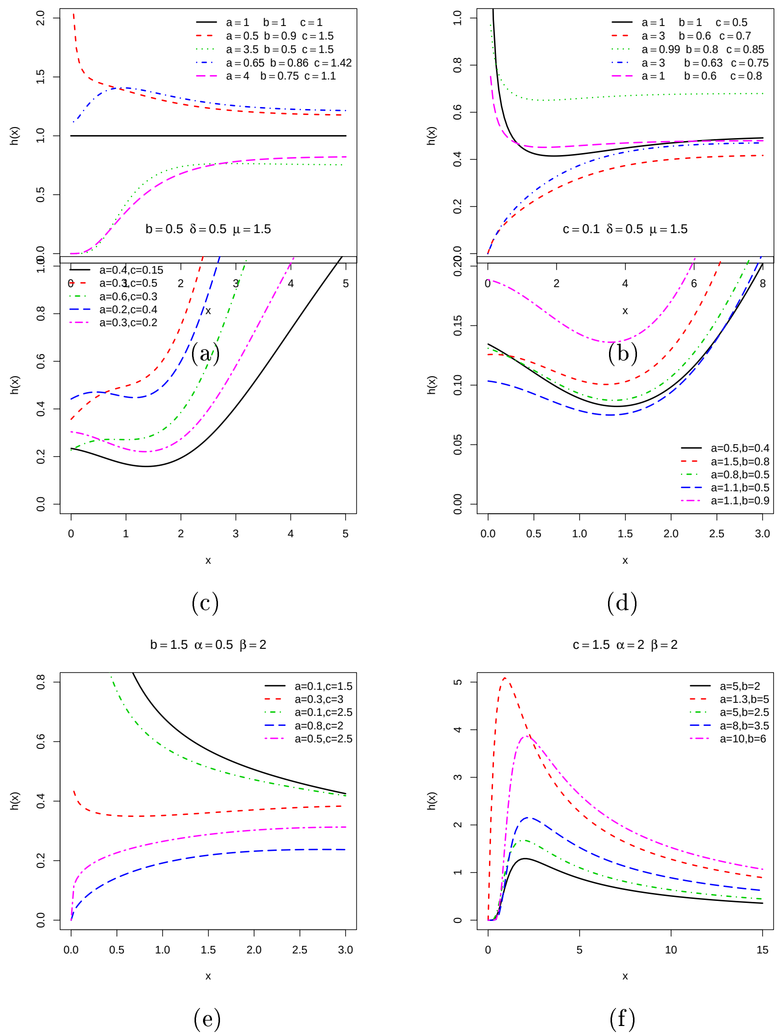


Figure 2. Hazard rate plots (a)-(b) of the BOLL-E model, (c)-(d) of the BOLL-N model and (e)-(f) of the BOLL-Lx model.

4. Mathematical properties

Here, we present some mathematical properties of the new family of distributions.

4.1. Asymptotics and shapes. The asymptotes of equations (2.3), (2.4) and (2.5) as $x \to 0$ and $x \to \infty$ are given by

$$
\begin{aligned}
&F(x) \sim I_{G(x)^c}(a,b) \quad \text{as } x \to 0,\\
&1-F(x) \sim I_{\bar{G}(x)^c}(b,a) \quad as\ x \to \infty,\\
&f(x) \sim \frac{c}{B(a,b)}\, g(x) G(x)^{a\,c-1} \quad \text{as } x \to 0,\\
&f(x) \sim \frac{c}{B(a,b)}\, g(x) \bar{G}(x)^{b\,c-1} \quad \text{as } x \to \infty,\\
&h(x) \sim \frac{c\, g(x) G(x)^{a\,c-1}}{1-I_{G(x)^c}(a,b)} \quad \text{as } x \to 0,\\
&h(x) \sim \frac{c\, g(x) \bar{G}(x)^{b\,c-1}}{I_{\bar{G}(x)^c}(b,a)} \quad \text{as } x \to \infty.
\end{aligned}
$$

The shapes of the density and hazard rate functions can be described analytically. The critical points of the BOLL-G density function are the roots of the equation:

$$
\frac{g'(x)}{g(x)} + (ac-1)\frac{g(x)}{G(x)} + (1-bc)\frac{g(x)}{\overline{G}(x)} - c(a+b)g(x)\frac{G(x)^{c-1}-\overline{G}(x)^{c-1}}{G(x)^c+\overline{G}(x)^c} = 0. \tag{4.1}
$$

There may be more than one root to (4.1). Let $\lambda(x) = d^2 \log[f(x)]/d\,x^2$. We have

$$
\begin{aligned}
\lambda(x) \;=\;& \frac{g''(x)g(x)-[g'(x)]^2}{g(x)^2} + (ac-1)\frac{g'(x)G(x)-g(x)^2}{G(x)^2}\\
+\;& (1-bc)\frac{g'(x)\overline{G}(x)+g(x)^2}{\overline{G}(x)^2} - c(a+b)g'(x)\frac{G(x)^{c-1}-\overline{G}(x)^{c-1}}{G(x)^c+\overline{G}(x)^c}\\
-\;& c(c-1)(a+b)g(x)^2\,\frac{G(x)^{c-2}+\overline{G}(x)^{c-2}}{G(x)^c+\overline{G}(x)^c}\\
-\;& (a+b)\left\{cg(x)\frac{G(x)^{c-1}-\overline{G}(x)^{c-1}}{G(x)^c+\overline{G}(x)^c}\right\}^2.
\end{aligned}
$$

If $x = x_0$ is a root of (4.1) then it corresponds to a local maximum if $\lambda(x) > 0$ for all $x < x_0$ and $\lambda(x) < 0$ for all $x > x_0$. It corresponds to a local minimum if $\lambda(x) < 0$ for all $x < x_0$ and $\lambda(x) > 0$ for all $x > x_0$. It refers to a point of inflexion if either $\lambda(x) > 0$ for all $x \neq x_0$ or $\lambda(x) < 0$ for all $x \neq x_0$.

The critical points of the hrf $h(x)$ are obtained from the equation

$$
\begin{aligned}
&\frac{g'(x)}{g(x)} + (ac-1)\frac{g(x)}{G(x)} + (1-bc)\frac{g(x)}{\overline{G}(x)} - c(a+b)g(x)\frac{G(x)^{c-1}-\overline{G}(x)^{c-1}}{G(x)^c+\overline{G}(x)^c}\\
&+ \frac{cg(x)G(x)^{ac-1}\overline{G}(x)^{bc-1}}{B(a,b)\left\{G(x)^c+\overline{G}(x)^c\right\}^{a+b}\left\{1-I_{\frac{G(x)^c}{\overline{G}(x)^c+G(x)^c}}(a,b)\right\}} = 0.
\end{aligned} \tag{4.2}
$$

There may be more than one root to (4.2). Let $\tau(x)=d^2\log[h(x)]/dx^2$. We have

$$
\begin{aligned}
\tau(x) &= \frac{g''(x)g(x)-[g'(x)]^2}{g(x)^2}+(ac-1)\frac{g'(x)G(x)-g(x)^2}{G(x)^2}\\
&+ (1-bc)\frac{g'(x)\overline{G}(x)+g(x)^2}{\overline{G}(x)^2}-c(a+b)g'(x)\frac{G(x)^{c-1}-\overline{G}(x)^{c-1}}{G(x)^c+\overline{G}(x)^c}\\
&+ c(c-1)(a+b)g(x)^2\,\frac{G(x)^{c-2}+\overline{G}(x)^{c-2}}{G(x)^c+\overline{G}(x)^c}\\
&- (a+b)\left\{cg(x)\frac{G(x)^{c-1}-\overline{G}(x)^{c-1}}{G(x)^c+\overline{G}(x)^c}\right\}^2\\
&+ \frac{cg'(x)G(x)^{ac-1}\overline{G}(x)^{bc-1}}{\left\{G(x)^c+\overline{G}(x)^c\right\}^{a+b}\left\{B(a,b)-B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right)\right\}}\\
&+ \frac{c(ac-1)g(x)^2G(x)^{ac-2}\overline{G}(x)^{bc-1}}{\left\{G(x)^c+\overline{G}(x)^c\right\}^{a+b}\left\{B(a,b)-B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right)\right\}}
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{c(bc-1)g(x)^2G(x)^{ac-1}\overline{G}(x)^{bc-2}}{\left\{G(x)^c+\overline{G}(x)^c\right\}^{a+b}\left\{B(a,b)-B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right)\right\}}\\
&- \frac{c^2(a+b)^2g(x)G(x)^{ac-1}\overline{G}(x)^{bc-1}\left\{G(x)^{c-1}-\overline{G}(x)^{c-1}\right\}}{\left\{G(x)^c+\overline{G}(x)^c\right\}^{a+b+1}\left\{B(a,b)-B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right)\right\}}\\
&+ \left\{\frac{cg(x)G(x)^{ac-1}\overline{G}(x)^{bc-1}}{\left\{G(x)^c+\overline{G}(x)^c\right\}^{a+b}\left\{B(a,b)-B\left(\frac{G(x;\xi)^c}{G(x;\xi)^c+\overline{G}(x;\xi)^c};a,b\right)\right\}}\right\}^2.
\end{aligned}
$$

If $x=x_0$ is a root of (4.2) then it refers to a local maximum if $\tau(x)>0$ for all $x<x_0$ and $\tau(x)<0$ for all $x>x_0$. It corresponds to a local minimum if $\tau(x)<0$ for all $x<x_0$ and $\tau(x)>0$ for all $x>x_0$. It gives an inflexion point if either $\tau(x)>0$ for all $x\neq x_0$ or $\tau(x)<0$ for all $x\neq x_0$.

4.2. Useful expansions. For an arbitrary baseline cdf $G(x)$, a random variable Z has the exp-G distribution (see Section 1) with power parameter $c>0$, say $Z\sim$exp-G(c), if its pdf and cdf are given by $h_c(x)=c\,G(x)^{c-1}\,g(x)$ and $H_c(x)=G(x)^c$, respectively. Some structural properties of the exp-G distributions are studied by Mudholkar and Srivastava [35], Mudholkar et al. [36], Mudholkar and Hutson [34], Gupta et al. [26], Gupta and Kundu [27, 28], Nadarajah and Kotz [39], Nadarajah and Gupta [40, 41] and Nadarajah [37].

We can prove that the cdf (2.3) admits the expansion

$$
\begin{aligned}
F(x) &= \sum_{l=0}^{\infty}\frac{(-1)^l}{B(a,b)(a+l)}\binom{b-1}{l}\frac{G(x)^{c(a+l)}}{[G(x)^c+\bar{G}(x)^c]^{a+l}}\\
&= \sum_{l=0}^{\infty}\frac{(-1)^l}{B(a,b)(a+l)}\binom{b-1}{l}\frac{\sum_{k=0}^{\infty}\alpha_k^{(l)}\,G(x)^k}{\sum_{k=0}^{\infty}\beta_k^{(l)}\,G(x)^k}.
\end{aligned}
$$

Using the power series for the ratio of two power series, we have

$$
F(x) = \sum_{l=0}^{\infty}\frac{(-1)^l}{B(a,b)(a+l)}\binom{b-1}{l}\sum_{k=0}^{\infty}\gamma_k^{(l)}\,G(x)^k,
$$

where (for each l) $\alpha_k^{(l)} = a_k(c(a+l))$, $\beta_k^{(l)} = h_k(c, a+l)$, $a_k(c(a+l))$ and $h_k(c, a+l)$ are defined in the Appendix A and $\gamma_k^{(l)}$ is determined recursively as

$$\gamma_k^{(l)} = \gamma_k(a,c) = \frac{1}{\beta_0^{(l)}} \left(\alpha_k^{(l)} - \frac{1}{\beta_0^{(l)}} \sum_{r=1}^{k} \beta_r^{(l)} \, \gamma_{k-r}^{(l)} \right).$$

Then, we have

$$F(x) \;=\; \sum_{k=0}^{\infty} b_k \, H_k(x),$$

where

$$b_k = \sum_{l=0}^{\infty} \frac{(-1)^l \, \gamma_k^{(l)}}{B(a,b)\,(a+l)} \binom{b-1}{l}, \tag{4.3}$$

and $H_k(x) = G(x)^k$ denotes the exp-G cdf with power parameter k. So, the density function of X can be expressed as

$$f(x) = f(x; a, b, c, \boldsymbol{\xi}) = \sum_{k=0}^{\infty} b_{k+1} \, h_{k+1}(x; \boldsymbol{\xi}), \tag{4.4}$$

where $h_{k+1}(x) = h_{k+1}(x; \boldsymbol{\xi}) = (k+1)\, g(x; \boldsymbol{\xi})\, G(x; \boldsymbol{\xi})^k$ denotes the exp-G density function with power parameter $k+1$. Hereafter, a random variable having density function $h_{k+1}(x)$ is denoted by $Y_{k+1} \sim$ exp-G$(k+1)$. Equation (4.4) reveals that the BOLL-G density function is an infinite mixture of exp-G densities. Thus, some mathematical properties of the new model can be obtained directly from those exp-G properties. For example, the ordinary and incomplete moments, and mgf of X can be determined from those quantities of the exp-G distribution.

The formulae derived throughout the paper can be easily handled in most symbolic computation software platforms such as Maple, Mathematica and Matlab. These platforms have currently the ability to deal with analytic expressions of formidable size and complexity. Established explicit expressions to calculate statistical measures can be more efficient than computing them directly by numerical integration. The infinity limit in these sums can be substituted by a large positive integer such as 20 or 30 for most practical purposes.

4.3. Quantile function. The qf of X, say $x = Q(u) = F^{-1}(u)$, can be obtained by inverting (2.3). Let $z = Q_{a,b}(u)$ be the beta qf. Then,

$$x = Q(u) = Q_G \left\{ \frac{[Q_{a,b}(u)]^{\frac{1}{c}}}{[Q_{a,b}(u)]^{\frac{1}{c}} + [1 - Q_{a,b}(u)]^{\frac{1}{c}}} \right\}.$$

It is possible to obtain some expansions for $Q_{a,b}(u)$ from the Wolfram website `http://functions.wolfram.com/06.23.06.0004.01` such as

$$z = Q_{a,b}(u) = \sum_{i=0}^{\infty} e_i \, u^{i/a},$$

where $e_i = [a\,B(a,b)]^{1/a}\, d_i$ and $d_0 = 0$, $d_1 = 1$, $d_2 = (b-1)/(a+1)$,

$$d_3 = \frac{(b-1)\,(a^2 + 3ab - a + 5b - 4)}{2(a+1)^2(a+2)},$$

$$\begin{aligned} d_4 &= (b-1)[a^4 + (6b-1)a^3 + (b+2)(8b-5)a^2 + (33b^2 - 30b + 4)a \\ &+ b(31b-47) + 18]/[3(a+1)^3(a+2)(a+3)], \ldots \end{aligned}$$

The effects of the shape parameters a, b and c on the skewness and kurtosis of X can be based on quantile measures. The Bowley skewness (Kenney and Keeping [30]) is one of the earliest skewness measures defined by the average of the quartiles minus the median, divided by half the interquartile range, namely

$$B = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{1}{2}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}.$$

Since only the middle two quartiles are considered and the outer two quartiles are ignored, this adds robustness to the measure. The Moors kurtosis (Moors [33]) is based on octiles

$$M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}.$$

These measures are less sensitive to outliers and they exist even for distributions without moments.

In Figure 3, we plot the measures B and M for the BOLL-N and BOLL-Lx distributions. The plots indicate the variability of these measures on the shape parameters.

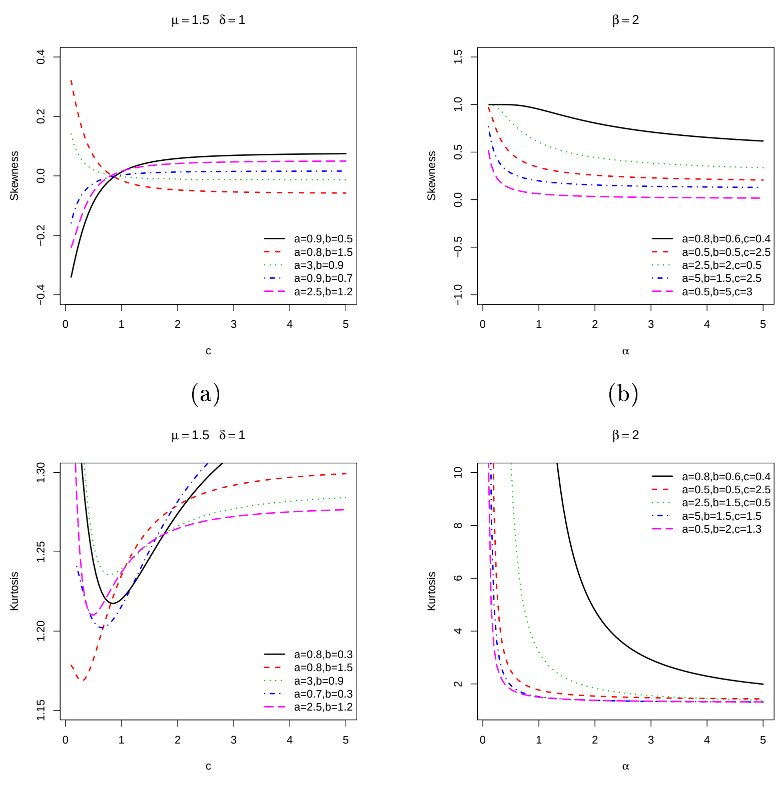


Figure 3. Skewness (a) and (b) and kurtosis (c) and (d) of X based on the quantiles for the BOLL-N and BOLL-Lx distributions, respectively.

4.4. Moments. We assume that Y is a random variable having the baseline cdf $G(x)$. The moments of X can be obtained from the (r,k)th probability weighted moment (PWM) of Y defined by Greenwood et al. [23] as

$$\tau_{r,k} = \mathrm{E}[Y^r\, G(Y)^k] = \int_{-\infty}^{\infty} x^r\, G(x)^k\, g(x)dx.$$

The PWMs are used to derive estimators of the parameters and quantiles of generalized distributions. The method of estimation is formulated by equating the population and sample PWMs. These moments have low variance and no severe biases, and they compare favorably with estimators obtained by maximum likelihood. The maximum likelihood method is adopted in Section 6.1 since it is easier to estimate the BOLL-G parameters because of several computer routines available in widely known softwares. The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used when constructing confidence intervals and regions and also in test statistics.

We can write from equation (4.4)

$$\mu_r' = \mathrm{E}(X^r) = \sum_{k=0}^{\infty}(k+1)\, b_{k+1}\, \tau_{r,k}, \tag{4.5}$$

where $\tau_{r,k} = \int_0^1 Q_G(u)^r\, u^k du$ can be computed at least numerically from any baseline qf.

Thus, the moments of any BOLL-G distribution can be expressed as an infinite weighted sum of the baseline PWMs. We now provide the PWMs for three distributions discussed in Section 3. For the BOLL-N and BOLL-Ga distributions discussed in subsections 3.2 and 3.5, the quantities $\tau_{r,k}$ can be expressed in terms of the Lauricella functions of type A (see Exton [16] and Trott [52]) defined by

$$F_A^{(n)}(a; b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n) =$$
$$\sum_{m1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\ldots+m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_n)_{m_n}}\,\frac{x_1^{m_1}\ldots x_n^{m_n}}{m_1!\ldots m_n!},$$

where $(a)_i = a(a+1)\ldots(a+i-1)$ is the ascending factorial (with the convention that $(a)_0 = 1$).

In fact, Cordeiro and Nadarajah [11] determined $\tau_{r,k}$ for the standard normal distribution as

$$\tau_{r,k} = 2^{r/2}\, \pi^{-(k+1/2)} \sum_{\substack{l=0 \\ (r+k-l)\text{ even}}}^{k} \binom{k}{l} 2^{-l}\, \pi^l\, \Gamma\left(\frac{r+k-l+1}{2}\right) \times$$
$$F_A^{(k-l)}\left(\frac{r+k-l+1}{2}; \frac{1}{2}, \ldots, \frac{1}{2}; \frac{3}{2}, \ldots, \frac{3}{2}; -1, \ldots, -1\right).$$

This equation holds when $r+k-l$ is even and it vanishes when $r+k-l$ is odd. So, any BOLL-N moment can be expressed as an infinite weighted linear combination of Lauricella functions of type A.

For the gamma distribution, the quantities $\tau_{r,k}$ can be expressed from equation (9) of Cordeiro and Nadarajah [11] as

$$\tau_{r,k} = \frac{\Gamma(r+(k+1)\alpha)}{\alpha^k\, \beta^r\, \Gamma(\alpha)^{k+1}}\, F_A^{(k)}(r+(k+1)\alpha; \alpha, \ldots, \alpha; \alpha+1, \ldots, \alpha+1, -1, \ldots, -1).$$

Finally, for the BOLL-W distribution, the quantities $\tau_{r,k}$ are given by

$$\tau_{r,k} = \frac{\Gamma(r/\beta+1)}{\alpha^{r/\beta}} \sum_{s=0}^{k} \frac{(-1)^s}{(s+1)^{r/\beta+1}} \binom{k}{s}.$$

4.5. Generating function. Here, we provide two formulae for the mgf $M(s) = E(e^{s\,X})$ of X. The first formula for $M(s)$ comes from equation (4.4) as

$$M(s) = \sum_{k=0}^{\infty} b_{k+1}\, M_{k+1}(s), \tag{4.6}$$

where $M_{k+1}(s)$ is the exp-G generating function with power parameter $k+1$.

Equation (4.6) can also be expressed as

$$M(s) = \sum_{k=0}^{\infty} (k+1)\; b_{k+1}\, \rho_k(s), \tag{4.7}$$

where the quantity $\rho_k(s) = \int_0^1 \exp\left[s\, Q_G(u)\right] u^k du$ can be computed numerically.

4.6. Mean deviations. Incomplete moments are useful for measuring inequality, for example, the Lorenz and Bonferroni curves and Pietra and Gini measures of inequality all depend upon the incomplete moments of the distribution. The nth incomplete moment of X is defined by $m_n(y) = \int_{-\infty}^{y} x^n\, f(x)dx$. Here, we propose two methods to determine the incomplete moments of the new family. First, the nth incomplete moment of X can be expressed as

$$m_n(y) = \sum_{k=0}^{\infty} b_{k+1} \int_0^{G(y;\,\xi)} Q_G(u)^n\, u^k\, du. \tag{4.8}$$

The integral in (4.8) can be computed at least numerically for most baseline distributions.

The mean deviations about the mean ($\delta_1 = E(|X - \mu_1'|)$) and about the median ($\delta_2 = E(|X - M|)$) of X are given by

$$\delta_1 = 2\mu_1'\, F\left(\mu_1'\right) - 2m_1\left(\mu_1'\right) \qquad \text{and} \qquad \delta_2 = \mu_1' - 2m_1(M), \tag{4.9}$$

respectively, where $M = Q(0.5)$ is the median of X, $\mu_1' = \mathrm{E}(X)$ comes from equation (4.5), $F(\mu_1')$ can easily be calculated from (2.3) and $m_1(z) = \int_{-\infty}^{z} x\, f(x)dx$ is the first incomplete moment.

Next, we provide two alternative ways to compute δ_1 and δ_2. A general equation for $m_1(z)$ can be derived from equation (4.4) as

$$m_1(z) = \sum_{k=0}^{\infty} b_{k+1}\, J_{k+1}(z), \tag{4.10}$$

where

$$J_{k+1}(z) = \int_{-\infty}^{z} x\, h_{k+1}(x)dx.$$

Equation (4.10) is the basic quantity to compute the mean deviations in (4.9). A simple application of (4.10) refers to the BOLL-W model. The exponentiated Weibull density function (for $x > 0$) with power parameter $k+1$, shape parameter α and scale parameter β, is given by

$$h_{k+1}(x) = (k+1)\, \alpha\, \beta^{\alpha}\, x^{\alpha-1}\, \exp\left\{-(\beta x)^{\alpha}\right\}\, \left[1 - \exp\left\{-(\beta x)^{\alpha}\right\}\right]^k,$$

and then

$$J_{k+1}(z) = c\,(k+1)\,\beta^{\alpha}\, \sum_{r=0}^{\infty} (-1)^r \binom{k}{r} \int_0^z x^{\alpha}\, \exp\left\{-(r+1)(\beta x)^{\alpha}\right\} dx.$$

The last integral reduces to the incomplete gamma function and then

$$J_{k+1}(z) = \beta^{-1} \sum_{r=0}^{\infty} \frac{(-1)^r\,(k+1)\,\binom{k}{r}}{(r+1)^{1+\alpha^{-1}}}\, \gamma\left(1 + \alpha^{-1}, (r+1)(\beta z)^{\alpha}\right).$$

A second general formula for $m_1(z)$ can be derived by setting $u = G(x)$ in (4.4)

$$m_1(z) = \sum_{k=0}^{\infty} (k+1)\, b_{k+1}\; T_k(z),$$

where $T_k(z) = \int_0^{G(z)} Q_G(u)\, u^k du$.

The main application of the first incomplete moment refers to the Bonferroni and Lorenz curves which are very useful in economics, reliability, demography, insurance and medicine. For a given probability π, applications of these equations can be addressed to obtain these curves defined by $B(\pi) = m_1(q)/(\pi\, \mu_1')$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $q = Q(\pi)$ is calculated from the parent qf.

4.7. Entropies. An entropy is a measure of variation or uncertainty of a random variable X. Two popular entropy measures are the Rényi [43] and Shannon [45]. The Rényi entropy of a random variable with pdf $f(x)$ is defined by

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left(\int_0^{\infty} f^{\gamma}(x) dx \right),$$

for $\gamma > 0$ and $\gamma \neq 1$. The Shannon entropy of a random variable X is given by $I_S = E\left\{-\log\left[f(X)\right]\right\}$. It is the special case of the Rényi entropy when $\gamma \uparrow 1$. Direct calculation yields

$$\begin{aligned} I_S \quad &= \quad -\log\left[\frac{c}{B(a,b)}\right] - \mathrm{E}\left\{\log\left[g(X;\boldsymbol{\xi})\right]\right\} + (1-ac)\,\mathrm{E}\left\{\log\left[G(x;\boldsymbol{\xi})\right]\right\} \\ &+ \quad (1-bc)\,\mathrm{E}\left\{\log\left[\bar{G}(x;\boldsymbol{\xi})\right]\right\} + (a+b)\,\mathrm{E}\left\{\log\left[G(x;\boldsymbol{\xi})^c + \bar{G}(x;\boldsymbol{\xi})^c\right]\right\}. \end{aligned}$$

First, we define and compute

$$\begin{aligned} A(a_1,a_2,a_3;a) \quad &= \quad \int_0^1 \frac{u^{a_1}(1-u)^{a_2}}{\left[u^a + (1-u)^a\right]^{a_3}} du \\ &= \quad \sum_{i=0}^{\infty} (-1)^i \binom{a_2}{i} \int_0^1 \frac{u^{a_1+i}}{\left[u^a + (1-u)^a\right]^{a_3}} du \\ &= \quad \sum_{i=0}^{\infty} (-1)^i \binom{a_2}{i} \int_0^1 \frac{\sum_{k=0}^{\infty} \delta_{1,k} u^k}{\sum_{k=0}^{\infty} \delta_{2,k}\, u^k} du \\ &= \quad \sum_{i=0}^{\infty} (-1)^i \binom{a_2}{i} \int_0^1 \sum_{k=0}^{\infty} \delta_{3,k} u^k \\ &= \quad \sum_{i=0}^{\infty} \frac{(-1)^i\, \delta_{3,k}}{(k+1)} \binom{a_2}{i}, \end{aligned}$$

where $\delta_{1,k} = a_k(a_1+i)$, $\delta_{2,k} = h_k(a, a_3)$ and

$$\delta_{3,k} = \frac{1}{\delta_{2,0}} \left(\delta_{1,k} - \frac{1}{\delta_{2,0}} \sum_{r=1}^{k} \delta_{2,r}\, \delta_{3,k-r} \right).$$

After some algebraic manipulations, we obtain:

4.1. Theorem. *Let X be a random variable with pdf* (2.4)*. Then,*

$$\mathrm{E}\left\{\log\left[G(X)\right]\right\} = \frac{c}{B(a,b)} \frac{\partial}{\partial t} A(a\,c + t - 1, b\,c - 1, a+b; c)\Big|_{t=0},$$

$$\mathrm{E}\left\{\log\left[\bar{G}(X)\right]\right\} = \frac{c}{B(a,b)} \frac{\partial}{\partial t} A(a\,c - 1, b\,c + t - 1, a+b; c)\Big|_{t=0},$$

$$\mathrm{E}\left\{G(x;\boldsymbol{\xi})^{a}+\bar{G}(X;\boldsymbol{\xi})^{a}\right\}=\frac{c}{B(a,b)}\,\frac{\partial}{\partial t}A(a\,c-1,b\,c-1,a+b-t;c)\Big|_{t=0}.$$

The simplest formula for the entropy of X is given by

$$\begin{aligned}I_S=&-\log\left[\frac{c}{B(a,b)}\right]-\mathrm{E}\left\{\log\left[g(X;\boldsymbol{\xi})\right]\right\}\\&+\frac{c\,(1-a\,c)}{B(a,b)}\,\frac{\partial}{\partial t}A(a\,c+t-1,b\,c-1,a+b;c)\Big|_{t=0}\\&+\frac{c\,(1-b\,c)}{B(a,b)}\,\frac{\partial}{\partial t}A(a\,c-1,b\,c+t-1,a+b;c)\Big|_{t=0}\\&+\frac{c\,(a+b)}{B(a,b)}\,\frac{\partial}{\partial t}A(a\,c-1,b\,c-1,a+b-t;c)\Big|_{t=0}.\end{aligned}$$

After some algebraic developments, we have an alternative expression for $I_R(\gamma)$:

$$I_R(\gamma)=\frac{\gamma}{1-\gamma}\log\left[\frac{c}{B(a,b)}\right]+\frac{1}{1-\gamma}\log\left[\sum_{i,k=0}^{\infty}t_{i,k}\;\mathrm{E}_{V_k}(g^{\gamma-1}[G^{-1}(Y)])\right].$$

Here, V_k has a beta distribution with parameters $k+1$ and one,

$$t_{i,k}=\frac{(-1)^i\,\gamma_{3,k}(a,b,c,i)}{(k+1)}\binom{c(a-1)}{i},$$

$$\gamma_{1,k}=a_k\Big((a\,c-1)\gamma+i\Big),\ \gamma_{2,k}=h_k\Big(c,(a+b)\gamma\Big)$$

and

$$\gamma_{3,k}=\frac{1}{\gamma_{2,0}}\left(\gamma_{1,k}-\frac{1}{\gamma_{2,0}}\sum_{r=1}^{k}\gamma_{2,r}\gamma_{3,k-r}\right),$$

where $a_k((a\,c-1)\gamma+i)$ and $h_k\Big(c,(a+b)\gamma\Big)$ are defined in equation (8.6) given in Appendix A.

4.8. Order statistics. Order statistics make their appearance in many areas of statistical theory and practice. Suppose $X_1,\ldots,X_n$ is a random sample from the BOLL-G family of distributions. We can write the density of the ith order statistic, say $X_{i:n}$, as

$$f_{i:n}(x)=\;K\,f(x)\,F^{i-1}(x)\;\{1-F(x)\}^{n-i}=K\sum_{j=0}^{n-i}(-1)^j\binom{n-i}{j}f(x)\,F(x)^{j+i-1},$$

where $K=n!/[(i-1)!\,(n-i)!]$.

Following similar algebraic developments of Nadarajah et al. [38], we can write the density function of $X_{i:n}$ as

$$f_{i:n}(x)=\sum_{r,k=0}^{\infty}m_{r,k}\,h_{r+k+1}(x),\tag{4.11}$$

where $h_{r+k}(x)$ denotes the exp-G density function with power parameter $r+k+1$ (for $r,k\geq 0$)

$$m_{r,k}=\frac{n!\,(r+1)\,(i-1)!\,b_{r+1}}{(r+k+1)}\sum_{j=0}^{n-i}\frac{(-1)^j\,f_{j+i-1,k}}{(n-i-j)!\,j!},$$

and b_k is defined in equation (4.3). The quantities $f_{j+i-1,k}$ can be obtained recursively by $f_{j+i-1,0} = b_0^{j+i-1}$ and

$$f_{j+i-1,k} = (k\, b_0)^{-1} \sum_{m=1}^{k} [m(j+i) - k]\, b_m\, f_{j+i-1,k-m}, \ \ k \geq 1.$$

Equation (4.11) is the main result of this section. It reveals that the pdf of the BOLL-G order statistics is a linear combination of exp-G density functions. So, several mathematical quantities of the BOLL-G order statistics such as ordinary, incomplete and factorial moments, mgf, mean deviations and several others can be determined from those quantities of the exp-G distribution.

5. Characterizations of the new family based on two truncated moments

The problem of characterizing distributions is an important problem which has attracted the attention of many researchers recently. An investigator will, generally, be interested to know if their chosen model fits the requirements of a particular distribution. Hence, one will depend on the characterizations of this distribution which provide conditions under which one can check to see if the underlying distribution is indeed that particular distribution. Various characterizations of distributions have been established in many different directions. In this section, we present characterizations of the BOLL-G distribution based on a simple relationship between two truncated moments. Our characterization results will employ a theorem due to Glänzel [24] (Theorem 5.1, below). The advantage of the characterizations given here is that the cdf F is not required to have a closed-form and is given in terms of an integral whose integrand depends on the solution of a first order differential equation, which can serve as a bridge between probability and differential equation. We believe that other characterizations of the BOLL-G family may not be possible.

5.1. Theorem. *Let $(\Omega, \Sigma, \mathbf{P})$ 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 $X : \Omega \to H$ be a continuous random variable with distribution function $F(x)$ and let q_1 and q_2 be two real functions defined on H such that*

$$\mathbb{E}\left[q_1(X) \,|\, X \geq x\right] = \mathbb{E}\left[q_2(X) \,|\, X \geq x\right] \, \eta(x), \quad x \in H,$$

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

$$F(x) = \int_a^x C \left| \frac{\eta'(u)}{\eta(u)\ q_2(u) - q_1(u)} \right| \ \mathrm{e}^{-s(u)}\, du\,,$$

where the function s is a solution of the differential equation $s' = \frac{\eta'\, q_2}{\eta\, q_2 - q_1}$ and C is a constant chosen to make $\int_H dF = 1$.

We have to mention that this kind of characterization based on the ratio of truncated moments is stable in the sense of weak convergence. In particular, let us assume that there is a sequence $\{X_n\}$ of random variables with distribution functions $\{F_n\}$ such that the functions $q_{1,n}$, $q_{2,n}$ and η_n ($n \in \mathbb{N}$) satisfy the conditions of Theorem 5.1 and let $q_{1,n} \to q_1$, $q_{2,n} \to q_2$ for some continuously differentiable real functions q_1 and q_2. Finally, let X be a random variable with distribution F. Under the condition that $q_{1,n}(X)$ and $q_{2,n}(X)$ are uniformly integrable and the family $\{F_n\}$ is relatively compact, the sequence X_n converges to X in distribution if and only if η_n converges to η, where

$$\eta(x)=\frac{E\left[q_1(X)\,|X\geq x\right]}{E\left[q_2(X)\,|X\geq x\right]}.$$

5.2. Remark. **(a)** In Theorem 5.1, the interval H need not be closed since the condition is only on the interior of H.
(b) Clearly, Theorem 5.1 can be stated in terms of two functions q_1 and η by taking $q_2(x)=1$, which will reduce the condition in Theorem 5.1 to $E\left[q_1(X)\,|X\geq x\right]=\eta(x)$. However, adding an extra function will give a lot more flexibility, as far as its application is concerned.

5.3. Proposition. *Let* $X:\Omega\to\mathbb{R}$ *be a continuous random variable and let* $q_1(x)=q_2(x)\;G(x;\xi)^{ac}$ *and* $q_2(x)=\left\{G(x;\xi)^c+\overline{G}(x;\xi)^c\right\}^{-(a+b)}\overline{G}(x;\xi)^{1-bc}$ *for* $x\in\mathbb{R}$. *The pdf of* X *is* (2.4) *if and only if the function* η *defined in Theorem 5.1 has the form*

$$\eta(x)=\frac{1}{2}\left[1+G(x;\xi)^{ac}\right],\quad x\in\mathbb{R}.$$

Proof. If X has pdf (2.4), then

$$\left[1-F(x)\right]\mathbf{E}\left[q_2(X)\,|X\geq x\right]=\frac{1}{aB(a,b)}\left[1-G(x;\xi)^{ac}\right],\quad x\in\mathbb{R}$$

and

$$\left[1-F(x)\right]\mathbf{E}\left[q_1(X)\,|X\geq x\right]=\frac{1}{2aB(a,b)}\left[1-G(x;\xi)^{2ac}\right],\quad x\in\mathbb{R}.$$

Finally,

$$\eta(x)\;q_2(x)-q_1(x)=\frac{1}{2}q_2(x)\left[1-G(x;\xi)^{ac}\right]>0,\quad \text{for}\quad x\in\mathbb{R}.$$

Conversely, if η is given as above, then

$$s'(x)=\frac{\eta'(x)\;q_2(x)}{\eta(x)\;q_2(x)-q_1(x)}=\frac{a\,c\;g(x)\;G(x;\xi)^{ac-1}}{\left[1-G(x;\xi)^{ac}\right]},\quad x\in\mathbb{R},$$

and hence

$$s(x)=-\log\left[1-G(x;\xi)^{ac}\right],\quad x\in\mathbb{R}.$$

Now, in view of Theorem 5.1, X has pdf (2.4). □

5.4. Corollary. *Let* $X:\Omega\to\mathbb{R}$ *be a continuous random variable and let* $q_2(x)$ *be as in Proposition 5.3. The pdf of* X *is (2.4) if and only if there exist functions* q_1 *and* η *defined in Theorem 5.1 satisfying the differential equation*

$$\frac{\eta'(x)\,q_2(x)}{\eta(x)\,q_2(x)-q_1(x)}=\frac{a\,c\;g(x)\,G(x;\xi)^{ac-1}}{\left[1-G(x;\xi)^{ac}\right]},\quad x\in\mathbb{R}.$$

5.5. Remark. **(a)** The general solution of the differential equation in Corollary 5.4 is

$$\eta(x)=\left[1-G(x;\xi)^{ac}\right]^{-1}\left[-\int a\,c\;g(x)\,G(x;\xi)^{ac-)}\,q_1(x)\,q_2(x)^{-1}\,dx+D\right],$$

for $x\in\mathbb{R}$, where D is a constant. One set of appropriate functions is given in Proposition 5.3 with $D=1/2$.

(b) Clearly there are other triplets of functions (q_1,q_2,η) satisfying the conditions of Theorem 5.1, e.g.,

$$q_1(x)=q_2(x)\,\overline{G}(x;\xi)^{bc}$$

and

$$q_2(x) = \left[G(x;\xi)^c + \overline{G}(x;\xi)^c\right]^{-(a+b)}\ G(x;\xi)^{1-ac}, \quad x \in \mathbb{R}.$$

Then, $\eta(x) = \frac{1}{2}\ \overline{G}(x;\xi)^{bc}$ and $s'(x) = \frac{\eta'(x)\, q_2(x)}{\eta(x)\, q_2(x) - q_1(x)} = b\, c\ g(x)\, \overline{G}(x)^{-1}, \quad x \in \mathbb{R}.$

6. Different methods of estimation

Here, we discuss parameter estimation using the methods of maximum likelihood and of minimum spacing distance estimator proposed by Torabi [48].

6.1. Maximum likelihood estimation. We consider the estimation of the unknown parameters of this family from complete samples only by the method of maximum likelihood. Let $x_1, \ldots, x_n$ be observed values from the BOLL-G distribution with parameters a, b, c and $\boldsymbol{\xi}$. Let $\Theta = (a, b, c, \boldsymbol{\xi})^\top$ be the $r \times 1$ parameter vector. The total log-likelihood function for Θ is given by

$$\begin{aligned}\ell_n \quad = \quad & n\,\log(c) - n\log[B(a,b)] + \sum_{i=1}^n \log[g(x_i;\boldsymbol{\xi})] + (ac-1)\sum_{i=1}^n \log[G(x_i;\boldsymbol{\xi})] \\ & + (bc-1)\sum_{i=1}^n \log[\bar{G}(x_i;\boldsymbol{\xi})] - (a+b)\sum_{i=1}^n \log\left\{G(x_i;\boldsymbol{\xi})^c + \bar{G}(x_i;\boldsymbol{\xi})^c\right\}. \end{aligned} \tag{6.1}$$

The log-likelihood function can be maximized either directly by using the R (AdequacyModel or Maxlik) (see R Development Core Team [42]), SAS (PROC NLMIXED), Ox program (sub-routine MaxBFGS) (see Doornik [14]), Limited-Memory quasi-Newton code for bound-constrained optimization (L-BFGS-B) or by solving the nonlinear likelihood equations obtained by differentiating (6.1).

Let $U_n(\Theta) = (\partial\ell_n/\partial a, \partial\ell_n/\partial b, \partial\ell_n/\partial c, \partial\ell_n/\partial\boldsymbol{\xi})^\top$ be the score function. Its components are given by

$$\begin{aligned}
\frac{\partial \ell_n}{\partial a} &= -n\psi(a) + n\psi(a+b) + c\sum_{i=1}^n \log[G(x_i;\boldsymbol{\xi})] - \sum_{i=1}^n \log\left\{G(x_i;\boldsymbol{\xi})^c + \bar{G}(x_i;\boldsymbol{\xi})^c\right\}, \\
\frac{\partial \ell_n}{\partial b} &= -n\psi(b) + n\psi(a+b) + c\sum_{i=1}^n \log[\bar{G}(x_i;\boldsymbol{\xi})] - \sum_{i=1}^n \log\left\{G(x_i;\boldsymbol{\xi})^c + \bar{G}(x_i;\boldsymbol{\xi})^c\right\}, \\
\frac{\partial \ell_n}{\partial c} &= \frac{n}{c} + a\sum_{i=1}^n \log[G(x_i;\boldsymbol{\xi})] + b\sum_{i=1}^n \log[\bar{G}(x_i;\boldsymbol{\xi})] \\
&\quad -(a+b)\sum_{i=1}^n \frac{G(x_i;\boldsymbol{\xi})^c \log[G(x_i;\boldsymbol{\xi})] + \bar{G}(x_i;\boldsymbol{\xi})^c \log[\bar{G}(x_i;\boldsymbol{\xi})]}{G(x_i;\boldsymbol{\xi})^c + \bar{G}(x_i;\boldsymbol{\xi})^c}, \\
\frac{\partial \ell_n}{\partial \boldsymbol{\xi}} &= \sum_{i=1}^n \frac{g(x_i;\boldsymbol{\xi})^{(\xi)}}{g(x_i;\boldsymbol{\xi})} + (ac-1)\sum_{i=1}^n \frac{G(x_i;\boldsymbol{\xi})^{(\xi)}}{G(x_i;\boldsymbol{\xi})} + (1-bc)\sum_{i=1}^n \frac{G(x_i;\boldsymbol{\xi})^{(\xi)}}{\bar{G}(x_i;\boldsymbol{\xi})} \\
&\quad -c(a+b)\sum_{i=1}^n G(x_i;\boldsymbol{\xi})^{(\xi)}\, \frac{G(x_i;\boldsymbol{\xi})^{c-1} - \bar{G}(x_i;\boldsymbol{\xi})^{c-1}}{G(x_i;\boldsymbol{\xi})^c + \bar{G}(x_i;\boldsymbol{\xi})^c},
\end{aligned}$$

where $h^{(\xi)}(\cdot)$ means the derivative of the function h with respect to ξ.

For interval estimation and hypothesis tests, we can use standard likelihood techniques based on the observed information matrix, which can be obtained from the authors upon request.

6.2. Minimum spacing distance estimator (MSDE). Torabi [48] introduced a general method for estimating parameters through spacing called maximum spacing distance estimator (MSDE). Torabi and Bagheri [49] and Torabi and Montazeri [51] used different MSDEs to compare with the MLEs. Here, we used two MSDEs, "minimum spacing absolute distance estimator" (MSADE) and "minimum spacing absolute-log distance estimator" (MSALDE) and compared them with the MLEs of the BOLL-E distribution. For mathematical details, the reader is referred to Torabi and Bagheri [49] and Torabi and Montazeri [51].

Table 2: The AEs, biases and MSEs of the MLEs, MSADEs and MSALDEs of the parameters based on 1,000 simulations of the BOLL-E(2, 1.5, 0.5, 1) distribution for $n = 100$, 200, 300 and 400.

		MLE			MSADE			MSALDE		
n		AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
100	a	3.158	1.158	5.743	2.271	0.271	5.404	2.361	0.361	14.717
	b	2.826	1.326	5.933	1.870	0.370	5.206	2.053	0.553	14.854
	c	0.587	2.658	0.301	0.509	1.771	0.027	0.582	1.861	0.133
	α	1.203	0.203	0.817	1.074	0.074	0.303	1.145	0.145	0.485
200	a	2.862	0.862	3.915	2.179	0.179	2.771	2.072	0.072	2.715
	b	2.461	0.961	3.758	1.750	0.250	2.855	1.651	0.151	2.837
	c	0.539	2.362	0.126	0.535	1.679	0.048	0.582	1.572	0.081
	α	1.114	0.114	0.440	1.078	0.078	0.245	1.141	0.141	0.334
300	a	2.112	0.112	2.492	2.666	0.666	2.609	2.133	0.133	3.709
	b	1.695	0.195	2.331	2.217	0.717	2.475	1.695	0.195	3.368
	c	0.554	1.612	0.072	0.519	2.166	0.080	0.583	1.633	0.080
	α	1.051	0.051	0.176	1.097	0.097	0.310	1.130	0.130	0.248
400	a	2.587	0.587	1.956	2.048	0.048	0.956	2.143	0.143	3.588
	b	2.109	0.609	1.869	1.602	0.102	0.970	1.669	0.169	3.383
	c	0.498	2.087	0.049	0.534	1.548	0.026	0.558	1.643	0.039
	α	1.080	0.080	0.232	1.062	0.062	0.161	1.135	0.135	0.220

We simulate the BOLL-E distribution for n=100, 200, 300 and 400 with $a = 2$, $b = 1.5$, $c = 0.5$ and $\alpha = 1$. For each sample size, we compute the MLEs, MSADEs and MSALDEs of the parameters. We repeat this process 1,000 times and obtain the average estimates (AEs), biases and mean square error (MSEs). The results are reported in Table 2. From the figures in this table, we note that the performances of the MLEs and MSADEs are better than MSALDEs.

7. Applications

In this section, we provide two applications to real data to illustrate the importance of the BOLL-G family through the special models: BOLL-E, BOLL-N and BOLL-Lx. The MLEs of the parameters are computed and the goodness-of-fit statistics for these models are compared with other competing models.

7.1. Data set 1: Strength of glass fibres. The first data set represents the strength of 1.5 cm glass fibres, measured at National physical laboratory, England (see, Smith and Naylor [46]). The data are: 0.55, 0.93, 1.25, 1.36, 1.49, 1.52, 1.58, 1.61, 1.64, 1.68, 1.73, 1.81, 2.00, 0.74, 1.04, 1.27, 1.39, 1.49, 1.53, 1.59, 1.61, 1.66, 1.68, 1.76, 1.82, 2.01, 0.77, 1.11, 1.28, 1.42, 1.50, 1.54, 1.60, 1.62, 1.66, 1.69, 1.76, 1.84, 2.24, 0.81, 1.13, 1.29, 1.48, 1.50, 1.55, 1.61, 1.62, 1.66, 1.70, 1.77, 1.84, 0.84, 1.24, 1.30, 1.48, 1.51, 1.55, 1.61, 1.63, 1.67, 1.70, 1.78, 1.89.

We fit the BOLL-E, BOLL-N, McDonald-Normal (McN) (Cordeiro et al. [9]), beta-normal (BN) (Famoye et al. [17]) and beta-exponential (BE) (Nadarajah and Kotz [39]) models to data set 1 and also compare them through seven goodness-of-fit statistics. The densities of the McN, BN and BE models are, respectively, given by:

$$\text{McN}: \; f_{\text{McN}}(x;a,b,c,\mu,\sigma) = \frac{c}{\sigma\, B(a,b)}\, \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right)^{ac-1} \left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)^{c}\right]^{b-1}, \quad \mu \in \Re, \quad a,b,c,\sigma > 0,$$

$$\text{BN}: \; f_{\text{BN}}(x;a,b,\mu,\sigma) = \frac{1}{\sigma\, B(a,b)} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right)^{a-1} \left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{b-1}, \quad \mu \in \Re, \quad a,b,\sigma > 0,$$

$$\text{BE}: \; f_{\text{BE}}(x;a,b,\alpha) = \frac{\alpha}{B(a,b)}\, \mathrm{e}^{-\alpha\, b\, x} \, \left(1-\mathrm{e}^{-\alpha\, x}\right)^{a-1}, \quad a,b,\alpha > 0.$$

7.2. Data set 2: Bladder cancer patients. The second data set represents the uncensored remission times (in months) of a random sample of 128 bladder cancer patients reported in Lee and Wang [31]. The data are: 0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33,5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

We fit the BOLL-E, BOLL-Lx, McDonald-Lomax (McLx) and beta-Lomax (BLx) (Lemonte and Cordeiro [32]) and BE models to these data and also compare their goodness-of-fit statistics. The densities of the McLx and BLx models are, respectively, given by

$$\begin{aligned}\text{McLx}: \; f_{\text{McLx}}(x;a,b,c,\alpha,\beta) &= \frac{c\,\alpha}{\beta\, B(a,b)} \left[1+\left(\frac{x}{\beta}\right)\right]^{-(\alpha+1)} \\ &\times \left\{1-\left[1+\left(\frac{x}{\beta}\right)\right]^{-\alpha}\right\}^{ac-1} \left[1-\left\{1-\left[1+\left(\frac{x}{\beta}\right)\right]^{-\alpha}\right\}^{c}\right]^{b-1}, \\ & a,b,c,\alpha,\beta > 0,\end{aligned}$$

$$\begin{aligned}\text{BLx}: \; f_{\text{BLx}}(x;a,b,\alpha,\beta) &= \frac{\alpha}{\beta\, B(a,b)} \left[1+\left(\frac{x}{\beta}\right)\right]^{-(\alpha b+1)} \left\{1-\left[1+\left(\frac{x}{\beta}\right)\right]^{-\alpha}\right\}^{a-1} \\ &\times \left[1-\left\{1-\left[1+\left(\frac{x}{\beta}\right)\right]^{-\alpha}\right\}^{a}\right]^{b-1}, \quad a,b,\alpha,\beta > 0.\end{aligned}$$

For all models, the MLEs are computed using the Limited-Memory Quasi-Newton Code for Bound-Constrained Optimization (L-BFGS-B). Further, the log-likelihood function evaluated at the MLEs ($\hat{\ell}$), Akaike information criterion (AIC), consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), Anderson-Darling (A^*), Cramér–von Mises (W^*) and Kolmogorov-Smirnov (K-S) statistics are calculated to compare the fitted models. The statistics A^* and W^* are defined by Chen and Balakrishnan [8]. In general, the smaller the values of these statistics, the better the fit to the data. The required computations are carried out in R-language.

Table 3: MLEs and their standard errors (in parentheses) for the first data set.

Distribution	a	b	c	μ	σ	α
BOLL-E	0.0698	0.1834	50.4548	-	-	0.4118
	(0.0931)	(0.2712)	(66.9766	-	-	(0.0125)
BOLL-N	0.0358	0.0764	34.7642	1.6597	0.6056	-
	(0.0660)	(0.1384)	(65.6410)	(0.0381)	(0.5323)	-
McN	0.5298	17.2226	1.2924	2.3850	0.4773	-
	(0.5249)	(48.8078)	(6.2595)	(1.8112)	(0.9820)	-
BN	0.5836	21.9402	-	2.5679	0.4658	-
	(0.6444)	(79.8234)	-	(1.3451)	(0.4546)	-
BE	17.4548	38.3856	-	-	-	0.2514
	(3.1323)	(65.8297)	-	-	-	(0.3684)

Table 4: The statistics $\hat{\ell}$, AIC, CAIC , BIC , HQIC, A^* and W^* for the first data set.

Distribution	$\hat{\ell}$	AIC	CAIC	BIC	HQIC	A^*	W^*
BOLL-E	$\mathbf{-10.4852}$	**28.9703**	**29.6599**	**37.5429**	**32.3419**	**0.3923**	**0.0681**
BOLL-N	$\mathbf{-9.9976}$	**29.9953**	**31.0479**	**40.7110**	**34.2098**	2.0245	**0.2858**
McN	-14.0577	38.1154	39.1680	48.8311	42.3299	0.9289	0.1659
BN	-14.0560	36.1119	36.8016	44.6845	39.4836	0.9179	0.1637
BE	-24.0256	54.0511	54.4579	60.4805	56.5798	3.1307	0.5708

Table 5: The K-S statistics and p-values for the first data set.

Distribution	K-S	p-value (K-S)
BOLL-E	**0.1126**	**0.4013**
BOLL-N	**0.0928**	**0.6496**
McN	0.1369	0.1886
BN	0.1356	0.1973
BE	0.2168	0.0053

Table 6: MLEs and their standard errors (in parentheses) for the second data set.

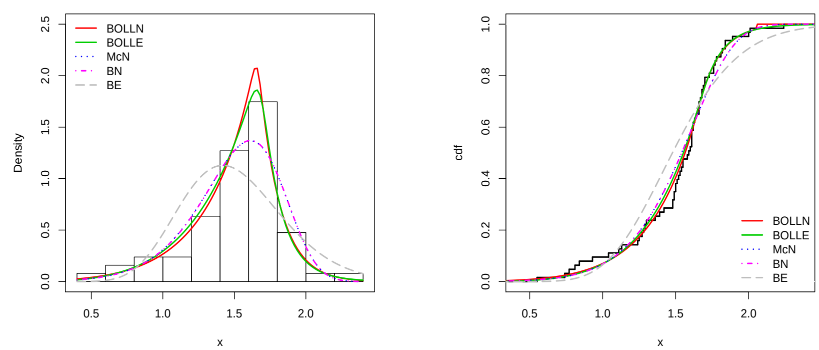
Distribution	a	b	c	α	β
BOLL-E	0.2772	0.1548	3.7895	0.1563	-
	(0.2529)	(0.1441)	(3.1996)	(0.0413)	-
BOLL-Lx	0.4507	0.3046	2.5267	8.5700	57.6246
	(0.4279)	(0.3573)	(2.0183)	(14.4135)	(88.4252)
McLx	1.5052	5.9638	2.0608	0.7177	10.9267
	(0.2831)	(30.1616)	(2.9944)	(3.0698)	(16.6896)
BLx	1.5882	12.0014	-	0.3859	20.4693
	(0.2830)	(319.2372)	-	(10.0697)	(14.0657)
BE	1.3781	0.2543	-	0.4595	-
	(0.2162)	(0.0251)	-	(0.0028)	-

Table 7: The statistics $\hat{\ell}$, AIC, CAIC , BIC , HQIC, A^* and W^* for the second data set.

Distribution	$\hat{\ell}$	AIC	CAIC	BIC	HQIC	A^*	W^*
BOLL-E	**−409.8323**	**827.6646**	**827.9898**	**839.0727**	**832.2998**	**1.5745**	**0.2022**
BOLL-Lx	**−409.2256**	**828.4513**	**828.9431**	**842.7115**	**834.2453**	**0.0800**	**0.0126**
McLx	−409.9128	829.8256	830.3174	844.0858	835.6196	0.1688	0.0254
BLx	−410.0813	828.1626	828.4878	839.5708	832.7978	0.1917	0.0285
BE	−412.1016	830.2033	830.3968	838.7594	833.6797	0.5475	0.0896

Table 8: The K-S statistics and p-values for the second data set.

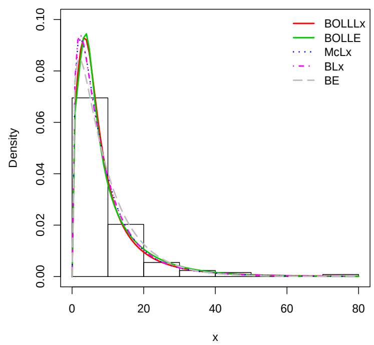
Distribution	K-S	p-value (K-S)
BOLL-E	**0.0295**	**0.9999**
BOLL-Lx	**0.0341**	**0.9984**
McLx	0.0391	0.9896
BLx	0.0407	0.9840
BE	0.0688	0.5793



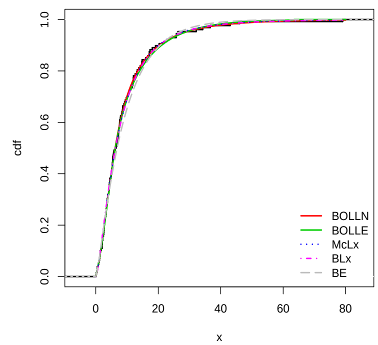
(a) Estimated pdfs (b) Estimated cdfs

Figure 4. Plots (a) and (b) of the estimated pdfs and cdfs of the BOLL-E and BOLL-N and other competitive models.

Tables 3 and 6 list the MLEs and their corresponding standard errors (in parentheses) of the parameters. The values of the model selection statistics AIC, CAIC, BIC, HQIC, A^*, W^* and K-S are listed in Tables 4-5 and 7-8. We note from Tables 4 and 5 that the BOLL-E and BOLL-N models have the lowest values of the AIC, CAIC, BIC, HQIC, W^* and K-S statistics (for the first data set) among the fitted McN, BN and BE models, thus suggesting that the BOLL-E and BOLL-N models provide the best fits, and therefore could be chosen as the most adequate models for the first data set. The histogram of these data and the estimated pdfs and cdfs of the BOLL-E and BOLL-N models and their competitive models are displayed in Figure 4. Similarly, it is also evident from the results in Tables 7 and 8 that the BOLL-E and BOLL-Lx models give the lowest values for the $\hat{\ell}$, AIC, CAIC, BIC, HQIC, A^*, W^* and K-S statistics (for the second data set) among the fitted McLx, BLx, KwLx and Lx distributions. Thus, the BOLL-E and BOLL-Lx models can be chosen as the best models. The histogram of the second data set and the estimated pdfs and cdfs of the BOLL-E and BOLL-Lx models and other competitive models are displayed in Figure 5.



(a) Estimated pdfs



(b) Estimated cdfs

Figure 5. Plots (a) and (b) of the estimated pdfs and cdfs of the BOLL-E and BOLL-Lx models and other competitive models.

It is clear from the figures in Tables 4-5 and 7-8, and Figures 4 and 5 that the BOLL-E, BOLL-N and BOLL-Lx models provide the best fits to these two data sets as compared to other models.

8. Concluding remarks

The generalized continuous univariate distributions have been widely studied in the literature. We propose a new class of distributions called the *beta odd log-logistic-G* family. We study some of its structural properties including an expansion for its density function and explicit expressions for the moments, generating function, mean deviations, quantile function and order statistics. The maximum likelihood method and the method of minimum spacing distance are employed to estimate the model parameters. We fit three special models of the proposed family to two real data sets to demonstrate its usefulness. We use some goodness-of-fit statistics in order to determine which distribution fits better to these data. We conclude that these special models provide consistently better fits than other competing models. We hope that the new family and its generated models will attract wider applications in several areas such as reliability engineering, insurance, hydrology, economics and survival analysis.

Appendix A

We present four power series expansions required for the proof of the general result in Section 4. First, for $a > 0$ real non-integer, we have the binomial expansion

$$(8.1) \qquad (1-u)^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} u^j,$$

where the binomial coefficient is defined for any real a as $a(a-1)(a-2), \ldots, (a-j+1)/j!$.

Second, the following expansion holds for any $\alpha > 0$ real non-integer

$$(8.2) \qquad G(x)^\alpha = \sum_{r=0}^{\infty} a_r(\alpha)\, G(x)^r,$$

where $a_r(\alpha) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{\alpha}{j} \binom{j}{r}$. The proof of (8.2) follows from $G(x)^\alpha = \{1 - [1 - G(x)]\}^\alpha$ by applying (8.1) twice.

Third, by expanding z^λ in Taylor series (when k is a positive integer), we have

$$z^\lambda = \sum_{k=0}^{\infty} (\lambda)_k \, (z-1)^k / k! = \sum_{i=0}^{\infty} f_i \, z^i, \tag{8.3}$$

where

$$f_i = f_i(\lambda) = \sum_{k=0}^{\infty} \frac{(-1)^{k-i}}{k!} \binom{k}{i} (\lambda)_k$$

and $(\lambda)_k = \lambda(\lambda - 1)\ldots(\lambda - k + 1)$ is the descending factorial.

Fourth, we use throughout an equation of Gradshteyn and Ryzhik [22] for a power series raised to a positive integer i given by

$$\left(\sum_{j=0}^{\infty} a_j \, v^j \right)^i = \sum_{j=0}^{\infty} c_{i,j} \, v^j, \tag{8.4}$$

where the coefficients $c_{i,j}$ (for $j = 1, 2, \ldots$) are obtained from the recurrence equation (for $j \geq 1$)

$$c_{i,j} = (j a_0)^{-1} \sum_{m=1}^{j} [m(j+1) - j] \, a_m \, c_{i,j-m} \tag{8.5}$$

and $c_{i,0} = a_0^i$. Hence, $c_{i,j}$ can be calculated directly from $c_{i,0}, \ldots, c_{i,j-1}$ and, therefore, from $a_0, \ldots, a_j$.

We now obtain an expansion for $[G(x)^c + \bar{G}(x)^c]^a$. We can write from equations (8.1) and (8.2)

$$[G(x)^c + \bar{G}(x)^c] = \sum_{j=0}^{\infty} t_j \, G(x)^j,$$

where

$$t_j = (-1)^j \left[\binom{c}{j} + \sum_{i=j}^{\infty} (-1)^i \binom{c}{i} \binom{c}{j} \right].$$

Then, using (8.3), we have

$$[G(x)^c + \bar{G}(x)^c]^a = \sum_{i=0}^{\infty} f_i \, \left(\sum_{j=0}^{\infty} t_j \, G(x)^j \right)^i,$$

where $f_i = f_i(a)$ is defined before.

Finally, using equations (8.4) and (8.5), we obtain

$$[G(x)^c + \bar{G}(x)^c]^a = \sum_{j=0}^{\infty} h_j \, G(x)^j, \tag{8.6}$$

where

$$h_j = h_j(c, a) = \sum_{i=0}^{\infty} f_i \, m_{i,j},$$

$$m_{i,j} = (j \, t_0)^{-1} \sum_{m=1}^{j} [m(j+1) - j] \, t_m \, m_{i,j-m} \quad (\text{for} \quad j \geq 1)$$

and $m_{i,0} = t_0^i$.

Acknowledgments

The authors gratefully acknowledge the help of Professor Hamzeh Torabi in developing R-code for the MSDEs. The authors are also grateful to the Editor-in-Chief (Professor Cem Kadilar) and two referees for their helpful comments and suggestions.

References

[1] Alexander, C., Cordeiro, G.M., Ortega, E.M.M. and Sarabia, J.M. *Generalized beta-generated distributions*, Computational Statistics and Data Analysis **56**, 1880–1897, 2012.

[2] Aljarrah, M.A., Lee, C. and Famoye, F. *On generating T-X family of distributions using quantile functions*, Journal of Statistical Distributions and Applications **1**, Article 2, 2014.

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

[4] Alzaatreh, A., Lee, C. and Famoye, F. *T-normal family of distributions: A new approach to generalize the normal distribution*, Journal of Statistical Distributions and Applications **1**, Article 16, 2014.

[5] Alzaghal, A., Lee, C. and Famoye, F. *Exponentiated T-X family of distributions with some applications*, International Journal of Probability and Statistics **2**, 31–49, 2013.

[6] Amini, M., MirMostafaee, S.M.T.K. and Ahmadi, J. *Log-gamma-generated families of distributions*, Statistics **48**, 913–932, 2014.

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

[8] Chen, G. and Balakrishnan, N. *A general purpose approximate goodness-of-fit test*, Journal of Quality Technology **27**, 154–161, 1995.

[9] Cordeiro, G.M., Cintra, R.J., Rego, L.C. and Ortega, E.M.M. *The McDonald normal distribution*, Pakistan Journal of Statistics and Operations Research **8**, 301–329, 2012.

[10] Cordeiro, G.M. and de Castro, M. *A new family of generalized distributions*, Journal of Statistical Computation and Simulation **81**, 883–893, 2011.

[11] Cordeiro, G.M. and Nadarajah, S. *Closed-form expressions for moments of a class of beta generalized distributions*, Brazilian Journal of Probability and Statistics **25**, 14–33, 2011.

[12] Cordeiro, G.M., Ortega, E.M.M. and da Cunha, D.C.C. *The exponentiated generalized class of distributions*, Journal of Data Science **11**, 1–27, 2013.

[13] Cordeiro, G.M., Alizadeh, M. and Ortega, E.M.M. *The exponentiated half-logistic family of distributions: Properties and applications*, Journal of Probability and Statistics Article ID 864396, 21 pages, 2014.

[14] Doornik, J.A. *Ox 5: An Object-Oriented Matrix Programming Language*, Fifth edition (Timberlake Consultants, London, 2007)

[15] Eugene, N., Lee, C. and Famoye, F. *Beta-normal distribution and its applications*, Communications in Statistics–Theory and Methods **31**, 497–512, 2002.

[16] Exton, H. *Handbook of Hypergeometric Integrals: Theory, Applications, Tables, Computer Programs* (Ellis Horwood, New York, 1978)

[17] Famoye, F., Lee, C. and Eugene, N. *Beta-normal distribution: Biomodality properties and application*, Journal of Modern Applied Statistical Methods **3**, 85–103, 2004.

[18] Gleaton, J.U. and Lynch, J.D. *On the distribution of the breaking strain of a bundle of brittle elastic fibers*, Advances in Applied Probability **36**, 98–115, 2004.

[19] Gleaton, J.U. and Lynch, J.D. *Properties of generalized log-logistic families of lifetime distributions*, Journal of Probability and Statistical Science **4**, 51–64, 2006.

[20] Gleaton, J.U. and Rahman, M. *Asymptotic properties of MLE's for distributions generated from a 2-parameter Weibull distribution by a generalized log-logistic transformation*, Journal of Probability and Statistical Science **8**, 199–214, 2010.

[21] Gleaton, J.U. and Rahman, M. *Asymptotic properties of MLE's for distributions generated from a 2-parameter inverse Gaussian distribution by a generalized log-logistic transformation*, Journal of Probability and Statistical Science **12**, 85–99, 2014.

[22] Gradshteyn, I.S. and Ryzhik, I.M. *Table of Integrals, Series, and Products*, Sixth edition (Academic Press, San Diego, 2000)

[23] Greenwood, J.A., Landwehr, J.M. and Matalas, N. C. *Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form*, Water Resources Research **15**, 1049–1054, 1979.

[24] Glänzel, W. *A characterization theorem based on truncated moments and its application to some distribution families*, In: Mathematical Statistics and Probability Theory, Volume B, pp. 75–84 (Reidel, Dordrecht, 1987)

[25] Gupta, R.C. and Gupta, R.D. *Proportional revered hazard rate model and its applications*, Journal of Statistical Planning and Inference **137**, 3525–3536, 2007.

[26] Gupta, R. C., Gupta, P. I. and Gupta, R. D. *Modeling failure time data by Lehmann alternatives*, Communications in statistics–Theory and Methods **27**, 887–904, 1998.

[27] Gupta, R.D. and Kundu, D. *Generalized exponential distribution*, Australian and New Zealand Journal of Statistics **41**, 173–188, 1999.

[28] Gupta, R.D. and Kundu, D. *Generalized exponential distribution: An alternative to Gamma and Weibull distributions*, Biometrical Journal **43**, 117–130, 2001.

[29] Jones, M.C. *Families of distributions arising from the distributions of order statistics*, Test **13**, 1–43, 2004.

[30] Kenney, J. and Keeping, E. *Mathematics of Statistics*, Volume 1, Third edition (Van Nostrand, Princeton, 1962)

[31] Lee, E.T. and Wang, J.W. *Statistical Methods for Survival Data Analysis*, Third edition (Wiley, New York, 2003)

[32] Lemonte, A.J. and Cordeiro, G.M. *An extended Lomax distribution*, Statistics **47**, 800–816, 2013.

[33] Moors, J.J.A. *A quantile alternative for kurtosis*, The Statistician **37**, 25–32, 1998.

[34] Mudholkar, G.S. and Hutson, A.D. *The exponentiated Weibull family: Some properties and a flood data application*, Communications in Statistics–Theory and Methods **25**, 3059–3083, 1996.

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

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

[37] Nadarajah, S. *The exponentiated exponential distribution: a survey*, AStA Advances in Statistical Analysis **95**, 219–251, 2011.

[38] Nadarajah, S., Cordeiro, G.M., Ortega and E.M.M. *The Zografos-Balakrishnan–G family of distributions: Mathematical properties and applications*, Communications in Statistics–Theory and Methods **44**, 186–215, 2015.

[39] Nadarajah, S. and Kotz, S. *The exponentiated-type distributions*, Acta Applicandae Mathematicae **92**, 97–111, 2006.

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

[41] Nadarajah, S. and Gupta, A.K. *A generalized gamma distribution with application to drought data*, Mathematics in Computer and Simulation **74**, 1–7, 2007.

[42] R Development Core Team *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing (Vienna, Austria, 2009)

[43] Rényi, A. *On measures of entropy and information*, In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability–I, University of California Press, Berkeley, pp. 547–561, 1961.

[44] Ristić, M.M. and Balakrishnan, N. *The gamma-exponentiated exponential distribution*, Journal of Statistical Computation and Simulation **82**, 1191–1206, 2012.

[45] Shannon, C.E. *A mathematical theory of communication*, Bell System Technical Journal **27**, 379–432, 1948.

[46] Smith, R.L. and Naylor, J.C. *A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution*, Applied Statistics **36**, 358–369, 1987.

[47] Tahir, M.H., Cordeiro, G.M., Alzaatreh, A., Zubair, M. and Mansoor, M. *The Logistic-X family of distributions and its applications*, Communications in Statistics–Theory and Methods **45**, 7326–7349, 2016.

[48] Torabi, H. *A general method for estimating and hypotheses testing using spacings*, Journal of Statistical Theory and Applications **8**, 163–168, 2008.

[49] Torabi, H. and Bagheri, F.L. *Estimation of parameters for an extended generalized half-logistic distribution based on complete and densored data*, Journal of the Iranian Statistical Society **9**, 171–195, 2010.

[50] Torabi, H. and Montazari, N.H. *The gamma-uniform distribution and its application*, Kybernetika **48**, 16–30, 2012.

[51] Torabi, H. and Montazari, N.H. *The logistic-uniform distribution and its application*, Communications in Statistics–Simulation and Computation **43**, 2551–2569, 2014.

[52] Trott, M. *The Mathematica Guidebook for Symbolics* (Springer, New York, 2006)

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