

The Zografos-Balakrishnan odd log-logistic family of distributions: Properties and Applications

Gauss M. Cordeiro* Morad Alizadeh^{†‡} Edwin M.M.Ortega[§] and Luis H. Valdivieso Serrano[¶]

Abstract

We study some mathematical properties of a new generator of continuous distributions with two additional shape parameters called the *Zografos-Balakrishnan odd log-logistic* family. We present some special models and investigate the asymptotes and shapes. The density function of the new family can be expressed as a mixture of exponentiated densities based on the same baseline distribution. We derive a power series for its quantile function. Explicit expressions for the ordinary and incomplete moments, quantile and generating functions, Shannon and Rényi entropies and order statistics, which hold for any baseline model, are determined. We estimate the model parameters by maximum likelihood. Two real data sets are used to illustrate the potentiality of the proposed family.

Keywords: Estimation, Gamma distribution, Generated family, Maximum likelihood, Mean deviation, Moment, Quantile function.

Received : 30.06.2015 *Accepted :* 25.10.2015 *Doi :* 10.15672/HJMS.20159714145

*Departamento de Estatística, Universidade Federal de Pernambuco, 50740-540, Recife, PE, Brazil,

Email: gausscordeiro@uol.com.br

[†]Department of Statistics, Faculty of Sciences, Persian Gulf University, Bushehr, 75169, Iran, Email: moradalizadeh78@gmail.com

[‡]Corresponding Author.

[§]Departamento de Ciências Exatas, Universidade de São Paulo, 13418-900, Piracicaba, SP, Brazil, Email: edwin@usp.br

[¶]Departamento de Ciencias, Pontificia Universidad Catolica del Perú, Lima, 32, Perú, Email: lvaldiv@pucp.edu.pe

1. Introduction

The statistics literature is filled with hundreds of continuous univariate distributions: see Johnson *et al.* (1994, 1995). Recent developments have been focused to define new families by adding shape parameters to control skewness, kurtosis and tail weights thus providing great flexibility in modeling skewed data in practice, including the two-piece approach introduced by Hansen (1994) and the generators pioneered by Eugene *et al.* (2002), Cordeiro and de Castro (2011), Alexander *et al.* (2012) and Cordeiro *et al.* (2013). Many subsequent articles apply these techniques to induce skewness into well-known symmetric distributions such as the symmetric Student *t*. For a review, see Aas and Haff (2006).

We study several mathematical properties of a new family of distributions called the *Zografos-Balakrishnan odd log-logistic-G* (“ZBOLL-G” for short) family with two additional shape parameters. These parameters can provide great flexibility to model the skewness and kurtosis of the generated distribution. Indeed, for any baseline G distribution, the new family can extend several common models such as the normal, Weibull and Gumbel distributions by adding these parameters to a parent G. The proposed family is an extension of that one introduced recently by Zografos and Balakrishnan (“ZB”) (2009) and Ristic and Balakrishnan (2012), although both are based on the same gamma generator.

Let W be any continuous distribution defined on a finite or an infinite interval. The ZB family is defined from the cumulative distribution function (cdf) (for $\beta > 0$)

$$(1.1) \quad F(x) = \frac{\gamma(\beta, -\log[1 - W(x)])}{\Gamma(\beta)}, \quad x \in \mathbb{R},$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ and $\gamma(\beta, z) = \int_0^z t^{\beta-1} e^{-t} dt$ are the gamma function and lower incomplete gamma function, respectively.

Further, we define $W(x)$ from any baseline cdf $G(x; \boldsymbol{\tau})$ ($x \in \mathbb{R}$), where $\boldsymbol{\tau}$ denotes the parameters in the parent G, as

$$(1.2) \quad W(x) = \frac{G^\alpha(x; \boldsymbol{\tau})}{G^\alpha(x; \boldsymbol{\tau}) + \bar{G}^\alpha(x; \boldsymbol{\tau})},$$

where $\alpha > 0$ and $\bar{G}(x; \boldsymbol{\tau}) = 1 - G(x; \boldsymbol{\tau})$ is the baseline survival function. According to Marshall and Olkin (2007, equation (21)), the function $W(x)$ in (1.2) is the odd log-logistic-G (OLL-G) cdf. By inserting (1.2) in equation (1.1), we have

$$(1.3) \quad F(x) = \frac{1}{\Gamma(\beta)} \gamma \left\{ \beta, -\log \left[1 - \frac{G^\alpha(x; \boldsymbol{\tau})}{G^\alpha(x; \boldsymbol{\tau}) + \bar{G}^\alpha(x; \boldsymbol{\tau})} \right] \right\}.$$

The model (1.3) is called the ZBOLL-G distribution with parameters α and β . Let $g(x; \boldsymbol{\tau}) = dG(x; \boldsymbol{\tau})/dx$ be the baseline probability density function (pdf). The density function corresponding to (1.3) is given by

$$(1.4) \quad f(x) = \frac{\alpha g(x; \boldsymbol{\tau}) G^{\alpha-1}(x; \boldsymbol{\tau}) \bar{G}^{\alpha-1}(x; \boldsymbol{\tau})}{\Gamma(\beta) [G^\alpha(x; \boldsymbol{\tau}) + \bar{G}^\alpha(x; \boldsymbol{\tau})]^2} \left\{ -\log \left[\frac{\bar{G}^\alpha(x; \boldsymbol{\tau})}{G^\alpha(x; \boldsymbol{\tau}) + \bar{G}^\alpha(x; \boldsymbol{\tau})} \right] \right\}^{\beta-1}.$$

Henceforth, a random variable X with density function (1.4) is denoted by $X \sim \text{ZBOLL-G}(\alpha, \beta, \boldsymbol{\tau})$. The ZBOLL-G family has the same parameters of the parent G plus the parameters α and β . For $\alpha = \beta = 1$, it reduces to the baseline G distribution. For $\alpha = 1$, we obtain the gamma-G (G-G) family and, for $\beta = 1$, we have the OLL-G family. The

hazard rate function (hrf) of X is given by

$$(1.5) \quad h(x) = \frac{\alpha g(x; \boldsymbol{\tau}) G^{\alpha-1}(x; \boldsymbol{\tau}) \bar{G}^{\alpha-1}(x; \boldsymbol{\tau})}{[G^{\alpha}(x; \boldsymbol{\tau}) + \bar{G}^{\alpha}(x; \boldsymbol{\tau})]^2} \times \frac{\left\{ -\log \left[\frac{\bar{G}^{\alpha}(x; \boldsymbol{\tau})}{G^{\alpha}(x; \boldsymbol{\tau}) + \bar{G}^{\alpha}(x; \boldsymbol{\tau})} \right] \right\}^{\beta-1}}{\Gamma(\beta) - \gamma \left\{ \beta, -\log \left[1 - \frac{G^{\alpha}(x; \boldsymbol{\tau})}{G^{\alpha}(x; \boldsymbol{\tau}) + \bar{G}^{\alpha}(x; \boldsymbol{\tau})} \right] \right\}}.$$

Each new ZBOLL-G distribution can be defined from a specified G distribution. The ZBOLL family is easily simulated by inverting (1.3) as follows: if V has the $\gamma(\beta, 1)$ distribution, then the solution of the nonlinear equation

$$(1.6) \quad X = G^{-1} \left\{ \frac{(1 - e^{-V})^{\frac{1}{\alpha}}}{(1 - e^{-V})^{\frac{1}{\alpha}} + e^{-\frac{V}{\alpha}}} \right\}$$

has density (1.4).

The parameters α and β have a clear interpretation. Following the key idea of Zografos and Balakrishnan (2009) and Ristic and Balakrishnan (2012), we can also interpret (1.4) in this way: if $X_{U(1)}, X_{U(2)}, \dots, X_{U(n)}$ are upper record values from a sequence of independent random variables with common pdf

$$w(x) = W'(x) = \frac{\alpha g(x; \boldsymbol{\tau}) \{G(x; \boldsymbol{\tau})[1 - G(x; \boldsymbol{\tau})]\}^{\alpha-1}}{\{G^{\alpha}(x; \boldsymbol{\tau}) + \bar{G}^{\alpha}(x; \boldsymbol{\tau})\}^2},$$

then the pdf of the n th upper record value has the pdf (1.4).

It is important to mention that the results presented in this paper follow similar lines of those developed by Nadarajah et al. (2015), although their model is completely different from that one discussed in this paper.

The rest of the paper is organized as follows. In Section 2, we present some new distributions. In Section 3, we introduce the asymptotic properties of equations (1.3), (1.4) and (1.5). Section 4 deals with two useful representations for (1.3) and (1.4). In Section 5, we derive a power series for the quantile function (qf) of X . In Sections 6 and 7, we obtain the entropies and order statistics. Estimation of the model parameters by maximum likelihood and the observed information matrix are presented in Section 8. Two applications to real data prove empirically the importance of the new family in Section 9. Finally, some conclusions and future work are noted in Section 10.

2. Special ZOBLL-G distributions

The ZOBLL-G family of density functions (1.4) allows for greater flexibility of its tails and can be widely applied in many areas of engineering and biology. In this section, we present and study some special cases of this family because it extends several widely-known distributions in the literature. The density function (1.4) will be most tractable when $G(x; \boldsymbol{\tau})$ and $g(x; \boldsymbol{\tau})$ have simple analytic expressions.

2.1. Zografos-Balakrishnan odd log-logistic Weibull (ZBOLL-W) model. If $G(x; \boldsymbol{\tau})$ is the Weibull cdf with scale parameter $\kappa > 0$ and shape parameter $\lambda > 0$, where $\boldsymbol{\tau} = (\lambda, \kappa)^T$, say $G(x; \boldsymbol{\tau}) = 1 - \exp\{-(x/\lambda)^{\kappa}\}$, the ZOBLL-W density function

(for $x > 0$) is given by

$$\begin{aligned}
 f_{\text{ZBOLL-W}}(x) &= \frac{\alpha \kappa \lambda^{-\kappa} x^{\kappa-1} \exp[-(x/\lambda)^\kappa] \{1 - \exp[-(x/\lambda)^\kappa]\}^{\alpha-1}}{\Gamma(\beta) \{ \{1 - \exp[-(x/\lambda)^\kappa]\}^\alpha + \exp[-\alpha(x/\lambda)^\kappa] \}^2} \\
 &\times \exp[-(\alpha - 1)(x/\lambda)^\kappa] \left\{ -\log \left[\frac{\exp[-\alpha(x/\lambda)^\kappa]}{\{1 - \exp[-(x/\lambda)^\kappa]\}^\alpha + \exp[-\alpha(x/\lambda)^\kappa]} \right] \right\}^{\beta-1}.
 \end{aligned}
 \tag{2.1}$$

Figure 1 displays some possible shapes of the ZBOLL-W density function.

2.2. Zografos-Balakrishnan odd log-logistic normal (ZBOLL-N) model. The ZBOLL-N distribution is defined from (1.4) by taking $G(x; \tau) = \Phi(\frac{x-\mu}{\sigma})$ and $g(x; \tau) = \sigma^{-1} \phi(\frac{x-\mu}{\sigma})$ to be the cdf and pdf of the normal $N(\mu, \sigma^2)$ distribution, where $\tau = (\mu, \sigma)^T$. Its density function is given by

$$\begin{aligned}
 f_{\text{ZBOLL-N}}(x) &= \frac{\alpha \phi(z) \Phi^{\alpha-1}(z) [1 - \Phi(z)]^{\alpha-1}}{\sigma \Gamma(\beta) \{ \Phi^\alpha(z) + [1 - \Phi(z)]^\alpha \}^2} \\
 &\times \left\{ -\log \left[\frac{[1 - \Phi(z)]^\alpha}{\Phi^\alpha(z) + [1 - \Phi(z)]^\alpha} \right] \right\}^{\beta-1},
 \end{aligned}
 \tag{2.2}$$

where $z = (x - \mu)/\sigma$, $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, α and β are shape and scale parameters, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. For $\mu = 0$ and $\sigma = 1$, we obtain the ZBOLL-standard normal (ZBOLL-SN) distribution. Plots of the ZBOLL-N density function for selected parameter values are displayed in Figure 2.

2.3. Zografos-Balakrishnan odd log-logistic Gumbel (ZBOLL-Gu) model. Consider the Gumbel distribution with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$, $\tau = (\mu, \sigma)^T$, and the pdf and cdf (for $x \in \mathbb{R}$) given by

$$g(x; \tau) = \frac{1}{\sigma} \exp \left[\left(\frac{x - \mu}{\sigma} \right) - \exp \left(\frac{x - \mu}{\sigma} \right) \right]$$

and

$$G(x; \tau) = 1 - \exp \left[-\exp \left(\frac{x - \mu}{\sigma} \right) \right],$$

respectively. The mean and variance are equal to $\mu - \gamma\sigma$ and $\pi^2\sigma^2/6$, respectively, where γ is the Euler's constant ($\gamma \approx 0.57722$). Inserting these expressions in (1.4) gives the ZBOLL-Gu density function

$$\begin{aligned}
 f_{\text{ZBOLL-Gu}}(x) &= \frac{\alpha \exp[z - \exp(z)] \{1 - \exp[-\exp(z)]\}^{\alpha-1} \exp[-(\alpha - 1)\exp(z)]}{\sigma \Gamma(\beta) \{ \{1 - \exp[-\exp(z)]\}^\alpha + \exp[-\alpha \exp(z)] \}^2} \\
 &\times \left\{ -\log \left(\frac{\exp[-\alpha(\exp(z))]}{\{1 - \exp[-\exp(z)]\}^\alpha + \exp[-\alpha \exp(z)]} \right) \right\}^{\beta-1},
 \end{aligned}
 \tag{2.3}$$

where $z = (x - \mu)/\sigma$, $x, \mu \in \mathbb{R}$ and $\alpha, \beta, \sigma > 0$. Plots of (2.3) for selected parameter values are displayed in Figure 3.

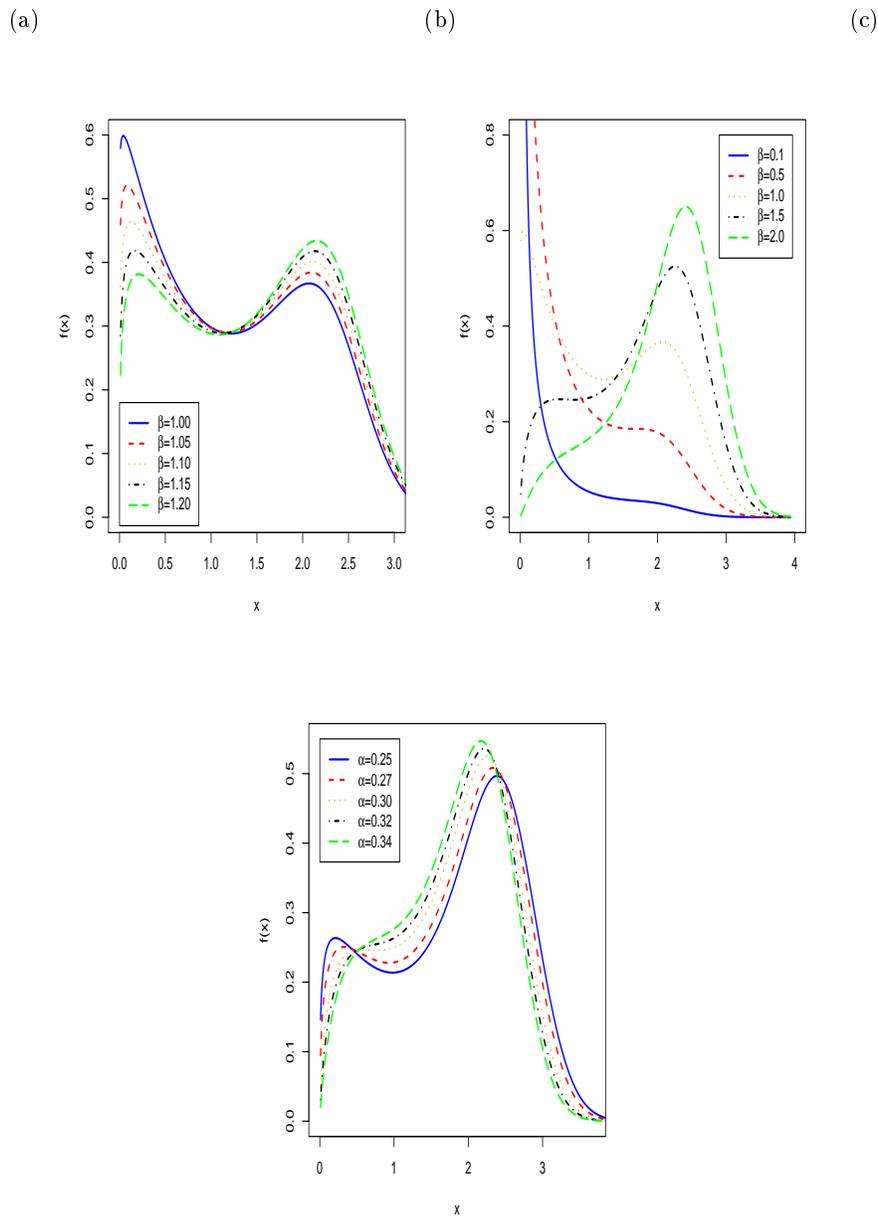


Figure 1. Plots of the ZBOLL-W density function for some parameter values. (a) For different values of β , with $\alpha = 0.3$, $\kappa = 3.5$ and $\lambda = 1.4$. (b) For different values of β with $\alpha = 0.3$, $\kappa = 3.5$ and $\lambda = 1.4$. (c) For different values of α with $\beta = 1.5$, $\kappa = 3.5$ and $\lambda = 1.4$.

3. Asymptotics

Let $c = \inf\{x|G(x) > 0\}$, then the asymptotics of equations (1.3), (1.4) and (1.5) when $x \rightarrow c$ are given by

$$\begin{aligned}
 F(x) &\sim \frac{G(x)^{\alpha\beta}}{\Gamma(\beta + 1)} \quad \text{as } x \rightarrow c, \\
 f(x) &\sim \frac{\alpha g(x) G(x)^{\alpha\beta-1}}{\Gamma(\beta)} \quad \text{as } x \rightarrow c, \\
 h(x) &\sim \frac{\alpha g(x) G(x)^{\alpha\beta-1}}{\Gamma(\beta)} \quad \text{as } x \rightarrow c.
 \end{aligned}$$

(a) (b) (c)

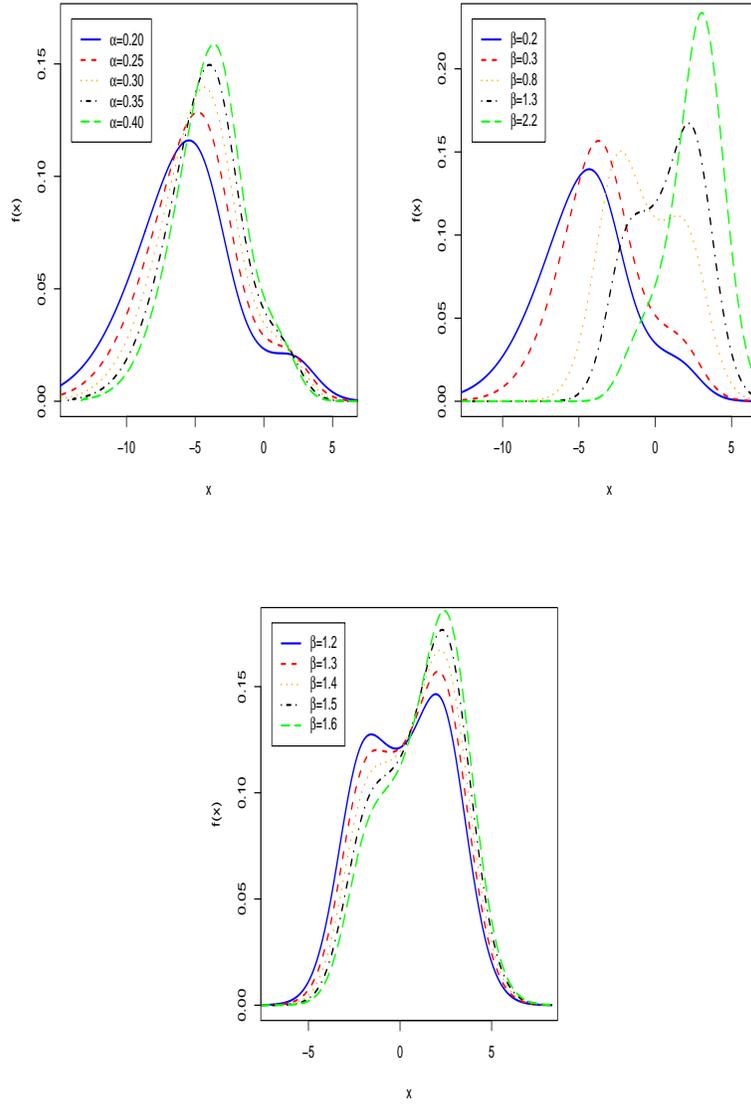


Figure 2. Plots of the ZBOLL-N density function for some parameter values. (a) For different values of α with $\beta = 0.2$, $\mu = 0$ and $\sigma = 1$. (b) For different values of β with $\alpha = 0.3$, $\mu = 0$ and $\sigma = 1.0$. (c) For different values of β with $\alpha = 0.3$, $\mu = 0$ and $\sigma = 0.1$.

The asymptotics of equations (1.3), (1.4) and (1.5) when $x \rightarrow \infty$ are given by

$$1 - F(x) \sim \frac{1}{\Gamma(\beta)} \{-\alpha \log [\bar{G}(x)]\}^{\beta-1} \bar{G}(x)^\alpha \text{ as } x \rightarrow \infty,$$

$$f(x) \sim \frac{\alpha g(x) \bar{G}(x)^{\alpha-1} \{-\alpha \log [\bar{G}(x)]\}^{\beta-1}}{\Gamma(\beta)} \text{ as } x \rightarrow \infty,$$

$$h(x) \sim \frac{\alpha g(x)}{\bar{G}(x)} \text{ as } x \rightarrow \infty.$$

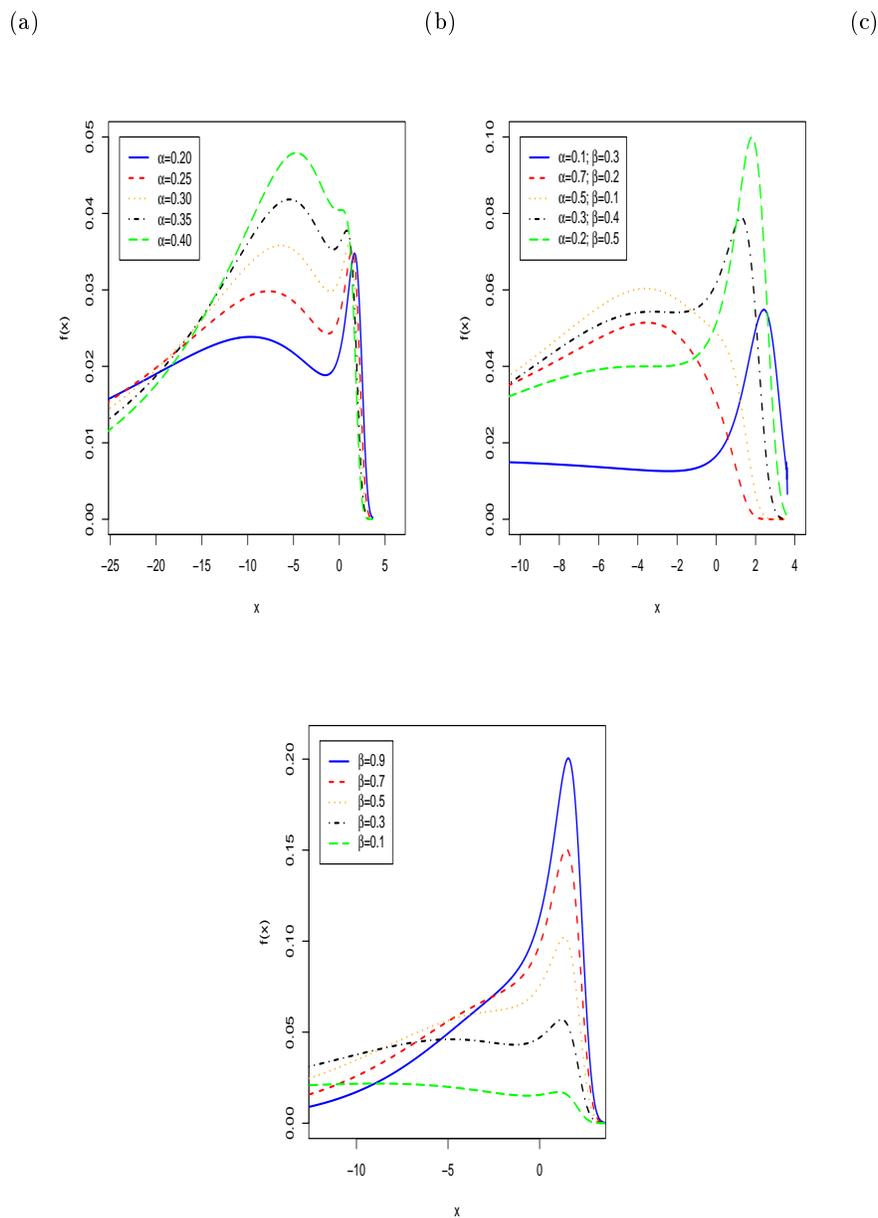


Figure 3. Plots of the ZBOLL-Gu density function for some parameter values. (a) For different values of α with $\beta = 0.2$, $\mu = 0$ $\sigma = 1$. (b) For different values of α and β with $\mu = 0$ and $\sigma = 1.0$. (c) For different values of β with $\alpha = 0.3$, $\mu = 0$ and $\sigma = 0$.

4. Two useful representations

Two useful linear representations for (1.3) and (1.4) can be derived using the concept of exponentiated distributions. For an arbitrary baseline cdf $G(x)$, a random variable

is said to have the exponentiated-G (exp-G) distribution with power parameter $a > 0$, say $Z \sim \text{exp-G}(a)$, if its pdf and cdf are $h_a(x) = aG^{a-1}(x)g(x)$ and $H_a(x) = G^a(x)$, respectively. The properties of exponentiated distributions have been studied by many authors in recent years, see Mudholkar and Srivastava (1993) for exponentiated Weibull, Gupta *et al.* (1998) for exponentiated Pareto, Gupta and Kundu (1999) for exponentiated exponential, Nadarajah (2005) for exponentiated Gumbel, Kakde and Shirke (2006) for exponentiated lognormal, and Nadarajah and Gupta (2007) for exponentiated gamma.

The generalized binomial coefficient for real arguments is given by $\binom{x}{y} = \Gamma(x + 1)/[\Gamma(y + 1)\Gamma(x - y + 1)]$. By using the incomplete gamma function expansion, we can write

$$F(x) = \frac{1}{\Gamma(\beta)} \sum_{i=0}^{\infty} \frac{(-1)^i}{i!(\beta + i)} \left\{ -\log \left[1 - \frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right] \right\}^{\beta+i}.$$

For any real positive power parameter, the formula below holds (<http://functions.wolfram.com/ElementaryFunctions/Log/06/01/04/03/>)

$$\begin{aligned} \left\{ -\log \left[1 - \frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right] \right\}^{\beta+i} &= (\beta + i) \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k - \beta - i}{k} \binom{k}{j}}{(\beta + i - j)} \\ &\times p_{j,k} \left[\frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right]^{\beta+i+k}, \end{aligned} \tag{4.1}$$

where the constants $p_{j,k}$ can be determined recursively by

$$p_{j,k} = k^{-1} \sum_{m=1}^k [k - m(j + 1)] c_m p_{j,k-m} \tag{4.2}$$

for $k = 1, 2, \dots$, $c_k = (-1)^{k+1}/(k + 1)$ and $p_{j,0} = 1$.

Further,

$$\left[\frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right]^{\beta+i+k} = \frac{\sum_{r=0}^{\infty} \lambda_r G(x)^r}{\sum_{r=0}^{\infty} \rho_r G(x)^r} = \sum_{r=0}^{\infty} a_r G(x)^r,$$

where

$$\lambda_r = \sum_{l=r}^{\infty} (-1)^{l+r} \binom{\alpha(\beta + i + k)}{l} \binom{l}{r}, \quad \rho_r = h_r(\alpha, \beta + i + k),$$

and (for $r \geq 1$)

$$a_r = a_r(\alpha, \beta, i, k) = \frac{1}{\rho_0} \left(\rho_r - \frac{1}{\rho_0} \sum_{s=1}^r \rho_s a_{r-s} \right),$$

$a_0 = \lambda_0/\rho_0$ and $h_r(\alpha, \beta + i + k)$ is defined in Appendix A.

Then, equation (1.3) can be expressed as

$$F(x) = \sum_{r=0}^{\infty} b_r H_r(x), \tag{4.3}$$

where

$$b_r = \frac{1}{\Gamma(\beta)} \sum_{i,k=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{i+j+k} p_{j,k} a_r(\alpha, \beta, i, k)}{(\beta + i - j) i!} \binom{k - \beta - i}{k} \binom{k}{j},$$

and $H_r(x)$ denotes the cdf of the exp- $G(r)$ distribution. The pdf (1.4) reduces to

$$(4.4) \quad f(x) = \sum_{r=0}^{\infty} b_{r+1} h_{r+1}(x),$$

where $h_{r+1}(x)$ denotes the pdf of the exp- $G(r+1)$ distribution. So, several mathematical properties of the proposed family can be obtained by knowing those of the exp- G distribution, see, for example, Mudholkar *et al.* (1996), Gupta and Kundu (2001) and Nadarajah and Kotz (2006), among others.

5. Quantile function

The gamma regularized function is defined by $Q(\beta, z) = \int_z^{\infty} x^{\beta-1} e^{-x} / \Gamma(\beta)$. The inverse gamma regularized function $Q^{-1}(\beta, u)$ admits a power series expansion given by (<http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/06/01/03/>)

$$Q^{-1}(\beta, u) = u \sum_{i=0}^{\infty} m_i u^i,$$

where $w = [\Gamma(\beta+1)(1-u)]^{1/\beta}$, $m_0 = 1$, $m_1 = 1/(\beta+1)$, $m_2 = (3\beta+5)/[2(\beta+1)^2(\beta+2)]$, $m_3 = [\beta(8\beta+3)+31]/[2(\beta+1)^3(\beta+2)(\beta+3)]$, etc.

First,

$$B = \frac{(1 - e^{-v})^{\frac{1}{\beta}}}{(1 - e^{-v})^{\frac{1}{\beta}} + e^{-\frac{v}{\beta}}} = \frac{1}{1 + e^{-\frac{v}{\beta}} (1 - e^{-v})^{-\frac{1}{\beta}}}$$

By using Taylor expansion and generalized binomial expansion, we have obtain

$$e^{-\frac{v}{\beta}} (1 - e^{-v})^{-\frac{1}{\beta}} = \sum_{k=0}^{\infty} b_k^* v^k,$$

where $b_0^* = 1$ and, for $k \geq 1$, $b_k^* = \frac{(-1)^{j+k} (j+\beta-1)^k}{k!} \binom{-1/\beta}{j}$.

Then,

$$B = \frac{1}{\sum_{k=0}^{\infty} b_k^* v^k} = \sum_{k=0}^{\infty} c_k^* v^k$$

where $c_0^* = 1/b_0^*$ and c_k^* (for $k \geq 1$) is obtained from the last equation as

$$c_k^* = -\frac{1}{b_0^*} \sum_{r=1}^k b_r^* c_{k-r}^*.$$

Further, we can write

$$(5.1) \quad \begin{aligned} A &= \frac{(1 - e^{-Q^{-1}(\beta, u)})^{\frac{1}{\beta}}}{(1 - e^{-Q^{-1}(\beta, u)})^{\frac{1}{\beta}} + e^{-\frac{Q^{-1}(\beta, u)}{\beta}}} = \sum_{k=0}^{\infty} c_k^* [Q^{-1}(\beta, u)]^k \\ &= \sum_{k=0}^{\infty} c_k^* \left(u \sum_{i=0}^{\infty} m_i u^i \right)^k. \end{aligned}$$

We use an equation by Gradshteyn and Ryzhik (2000, Section 0.314) for a power series raised to a positive integer j

$$(5.2) \quad \left(\sum_{i=0}^{\infty} a_i u^i \right)^j = \sum_{i=0}^{\infty} c_{j,i} u^i.$$

Here, for $j \geq 0$, $c_{j,0} = a_0^j$, and the coefficients $c_{j,i}$ (for $i = 1, 2, \dots$) are determined from the recurrence equation

$$(5.3) \quad c_{j,i} = (i a_0)^{-1} \sum_{p=1}^i [p(j+1) - i] a_p c_{j,i-p},$$

So, the coefficient $c_{j,i}$ follows from $c_{j,0}, \dots, c_{j,i-1}$ and then from a_0, \dots, a_i .

Based on equations (5.2) and (5.3), we can rewrite (5.1) as

$$A = \sum_{i,k=0}^{\infty} c_k^* v_{k,i} u^{i+k} = \sum_{l=0}^{\infty} d_l^* u^l,$$

where, for $k \geq 0$, the coefficients $v_{k,i}$ (for $i = 1, 2, \dots$) are determined from the recurrence equation

$$v_{k,i} = (i m_0)^{-1} \sum_{p=1}^i [p(j+1) - i] m_p v_{k,i-p},$$

with $v_{k,0} = m_0^k$ and $d_l^* = \sum_{(i,k) \in I_l} c_k^* v_{k,i}$ and $I_l = \{(i, k) | i + k = l; i, k = 0, 1, 2, \dots\}$.

Then, the qf of X reduces to

$$(5.4) \quad Q(u) = Q_G \left(\sum_{l=0}^{\infty} d_l^* u^l \right).$$

In general, even when $Q_G(u)$ does not have a closed-form expression, this function can usually be expressed in terms of a power series

$$(5.5) \quad Q_G(u) = \sum_{i=0}^{\infty} s_i u^i,$$

where the coefficients s_i 's are suitably chosen real numbers. For several important distributions, such as the normal, Student t, gamma and beta distributions, $Q_G(u)$ does not have a closed-form expression but it can be expanded as in equation (5.5).

By combining (5.4) and (5.5) and using again (5.2) and (5.3), we obtain

$$(5.6) \quad Q(u) = \sum_{l=0}^{\infty} h_l u^l,$$

where $h_l = \sum_{i=0}^{\infty} s_i h_{i,l}$ (for $i \geq 0$ and $l \geq 0$), $h_{i,l} = (l d_0^*)^{-1} \sum_{p=1}^l [p(i+1) - l] d_p^* h_{i,l-p}$, for $l \geq 1$, and $h_{i,0} = d_0^*$.

Hence, equation (5.6) reveals that the qf of the ZBOLL-G distribution can be expressed as a power series. For practical purposes, we can adopt ten terms in this power series.

Let $W(\cdot)$ be any integrable function in the positive real line. We can write

$$(5.7) \quad \int_{-\infty}^{\infty} W(x) f(x) dx = \int_0^1 W \left(\sum_{l=0}^{\infty} h_l u^l \right) du.$$

Equations (5.6) and (5.7) are the main results of this section. We can obtain from them various ZBOLL-G mathematical properties using integrals over $(0, 1)$, which are usually more simple than if they are based on the left integral. For example, an alternative formula for the n th ordinary moment of X follows from (5.7) combined with (5.2) and (5.3) as

$$\mu'_n = \int_0^1 \left(\sum_{l=0}^{\infty} h_l u^l \right)^n du = \sum_{l=0}^{\infty} \frac{f_{n,l}}{(l+1)},$$

where (for $n \geq 0$) $f_{n,0} = h_0^n$ and, for $n \geq 1$,

$$f_{n,l} = (lh_0)^{-1} \sum_{r=1}^l [r(n+1) - l] h_r f_{n,l-r}.$$

6. Entropies

An entropy is a measure of variation or uncertainty of a random variable X . Two popular entropy measures are the Rényi and Shannon entropies. The Rényi entropy of a random variable with pdf $f(\cdot)$ 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 defined by $E\{-\log[f(X)]\}$. It is a special case of the Rényi entropy when $\gamma \uparrow 1$.

Here, we derive expressions for the Rényi and Shannon entropies of the ZBOLL-G family. By using (4.1), we can write

$$\begin{aligned} \left\{ -\log \left[1 - \frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right] \right\}^{\gamma\beta-\gamma} &= (\gamma\beta - \gamma) \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k-\gamma\beta+\gamma}{k} \binom{k}{j} p_{j,k}}{[\gamma(\beta-1) - j]} \\ &\times \left[\frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right]^{\lceil \gamma(\beta-1)+k \rceil}. \end{aligned}$$

Hence,

$$\begin{aligned} &\left\{ -\log \left[1 - \frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right] \right\}^{\gamma\beta-\gamma} \left[\frac{\alpha g(x) G(x)^{\alpha-1} \bar{G}(x)^{\alpha-1}}{\Gamma(\beta) [G(x)^\alpha + \bar{G}(x)^\alpha]^2} \right]^\gamma = \\ &\alpha^\gamma (\gamma\beta - \gamma) \sum_{k,s=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k-\gamma\beta+\gamma}{k} \binom{k}{j} \binom{\gamma(\alpha-1)}{s} p_{j,k}}{\Gamma(\beta)^{-\gamma} [\gamma(\beta-1) - j]} \\ &\times \frac{G(x)^{\alpha\gamma(\beta-1)+k\alpha+\gamma(\alpha-1)+s}}{[G(x)^\alpha + \bar{G}(x)^\alpha]^{\gamma(\beta-1)+k+2\gamma}}. \end{aligned}$$

Further,

$$\frac{G(x)^{\alpha\gamma(\beta-1)+k\alpha+\gamma(\alpha-1)+s}}{[G(x)^\alpha + \bar{G}(x)^\alpha]^{\gamma(\beta-1)+k+2\gamma}} = \frac{\sum_{r=0}^{\infty} \lambda'_r G(x)^r}{\sum_{r=0}^{\infty} \rho'_r G(x)^r} = \sum_{r=0}^{\infty} a'_r G(x)^r,$$

where

$$\begin{aligned} \lambda'_r &= \sum_{l=r}^{\infty} (-1)^{l+r} \binom{\alpha\gamma(\beta-1) + k\alpha + \gamma(\alpha-1) + s}{l} \binom{l}{r} \\ \rho'_r &= h_r(\alpha, \gamma(\beta-1) + k + 2\gamma) \\ a'_r &= a'_r(\alpha, \beta, i, k) = \frac{1}{\rho'_0} \left(\rho'_r - \frac{1}{\rho'_0} \sum_{s=1}^r \rho'_s a'_{r-s} \right), \text{ for } r \geq 1, \end{aligned}$$

$a_0 = \lambda'_0/\rho'_0$ and $h_r(\alpha, \beta + i + k)$ is defined in the Appendix. Then,

$$\begin{aligned} \int_0^\infty g^\gamma(x) dx &= \frac{1}{\Gamma(\beta)^\gamma} \int_0^\infty \left\{ -\log \left[1 - \frac{G(x)^\alpha}{G(x)^\alpha + \bar{G}(x)^\alpha} \right] \right\}^{\gamma\alpha-\gamma} \times \\ &\left\{ \frac{\alpha g(x) G(x)^{\alpha-1} \bar{G}(x)^{\alpha-1}}{[G(x)^\alpha + \bar{G}(x)^\alpha]^2} \right\}^\gamma dx \\ &= \frac{\alpha^\gamma (\gamma\beta - \gamma)}{\Gamma(\beta)^\gamma} \sum_{k,r,s=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k-\gamma\beta+\gamma}{k} \binom{k}{j} \binom{\gamma(\alpha-1)}{s} p_{j,k} a'_r}{[\gamma(\beta-1) - j]} S_r, \end{aligned}$$

where S_r can be evaluated from the baseline distribution as

$$S_r = \int_0^\infty G(x)^r g^\gamma(x) dx.$$

Hence, the Rényi entropy of X is given by

$$I_R(\gamma) = \frac{\gamma}{1-\gamma} \log(\alpha) - \frac{\gamma}{1-\gamma} \log[\Gamma(\beta)] + \frac{1}{1-\gamma} \log(\gamma\beta - \gamma) + \frac{1}{1-\gamma} \log \left\{ \sum_{k,r,s=0}^\infty \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k-\gamma\beta+\gamma}{k} \binom{k}{j} \binom{\gamma(\alpha-1)}{s} p_{j,k} a'_r}{[\gamma(\beta-1)-j]} I_r \right\}.$$

The Shannon entropy can be obtained by limiting $\gamma \uparrow 1$ in $I_R(\gamma)$. However, it is easier to derive an expression for it from first principles. Using the power series for $\log(1-z)$, we can write

$$E\{-\log[f(X)]\} = -\log(\alpha) + \log[\Gamma(\beta)] - E\{\log[g(X)]\} + (1-\alpha)E\{\log[G(X)]\} + (1-\alpha)E\{\log[\bar{G}(X)]\} + 2E\{\log[G^\alpha(X) + \bar{G}^\alpha(X)]\} + (1-\beta)E\left\{-\log\left[1 - \frac{G^\alpha(X)}{G^\alpha(X) + \bar{G}^\alpha(X)}\right]\right\}.$$

First, we define and compute

$$A(a_1, a_2, a_3, a_4; \alpha) = \int_0^1 \frac{u^{a_1}(1-u)^{a_2}}{[u^\alpha + (1-u)^\alpha]^{a_3}} \left\{-\log\left[1 - \frac{u^\alpha}{u^\alpha + (1-u)^\alpha}\right]\right\}^{a_4} du.$$

Along the same lines of the derivation of the Rényi entropy, we obtain

$$A(a_1, a_2, a_3, a_4; \alpha) = a_4 \sum_{k,s=0}^\infty \sum_{j=0}^k \frac{(-1)^{j+k+s} \binom{k-a_4}{k} \binom{k}{j} \binom{a_2}{s} p_{j,k}}{a_4 - j} \times \int_0^1 \frac{u^{\alpha(a_4+k)+a_1+s}}{[u^\alpha + (1-u)^\alpha]^{a_4+k+a_3}} du.$$

Also,

$$\frac{u^{\alpha(a_4+k)+a_1+s}}{[u^\alpha + (1-u)^\alpha]^{a_4+k+a_3}} = \frac{\sum_{r=0}^\infty \lambda_r'' u^r}{\sum_{r=0}^\infty \rho_r'' u^r} = \sum_{r=0}^\infty a_r'' u^r,$$

where (for $r \geq 1$)

$$\lambda_r'' = \sum_{l=r}^\infty (-1)^{l+r} \binom{\alpha(a_4+k) + a_1 + s}{l} \binom{l}{r},$$

$$\rho_r'' = h_r(\alpha, a_4 + k + a_3),$$

$$a_r'' = a_r''(\alpha, \beta, i, k) = \frac{1}{\rho_0''} \left(\rho_r' - \frac{1}{\rho_0''} \sum_{s=1}^r \rho_s'' a_{r-s}'' \right),$$

$a_0'' = \lambda_0''/\rho_0''$ and $h_r(\alpha, a_4 + k + a_3)$ is defined in the Appendix. Then,

$$A(a_1, a_2, a_3, a_4; \alpha) = a_4 \sum_{k,s,r=0}^\infty \sum_{j=0}^k \frac{(-1)^{j+k+s} \binom{k-a_4}{k} \binom{k}{j} \binom{\alpha-1}{s}}{(a_4-j)(r+1)} p_{j,k} a_r''(\alpha, \beta, i, k).$$

Hence,

$$E \{ \log [G(X)] \} = \frac{\alpha}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha + t - 1, \alpha - 1, 2, \beta - 1; \alpha) \Big|_{t=0},$$

$$E \{ \log [\bar{G}(X)] \} = \frac{\alpha}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha - 1, \alpha + t - 1, 2, \beta - 1; \alpha) \Big|_{t=0},$$

$$E \{ \log [G(X)^\alpha + \bar{G}(X)^\alpha] \} = \frac{\alpha}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha - 1, \alpha - 1, 2 - t, \beta - 1; \alpha) \Big|_{t=0}$$

and

$$E \left\{ -\log \left[1 - \frac{G^\alpha(X)}{G^\alpha(X) + \bar{G}^\alpha(X)} \right] \right\} = \frac{\alpha}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha - 1, \alpha - 1, 2, \beta + t - 1; \alpha) \Big|_{t=0}.$$

The simplest formula for the Shannon entropy of X is given by

$$\begin{aligned} E \{ -\log[f(X)] \} &= -\log(\alpha) + \log[\Gamma(\beta)] - E \{ \log[g(X; \tau)] \} \\ &+ \frac{\alpha(1-\alpha)}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha + t - 1, \alpha - 1, 2, \beta - 1; \alpha) \Big|_{t=0} \\ &+ \frac{\alpha(1-\alpha)}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha - 1, \alpha + t - 1, 2, \beta - 1; \alpha) \Big|_{t=0} \\ &+ \frac{2\alpha}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha - 1, \alpha - 1, 2 - t, \beta - 1; \alpha) \Big|_{t=0} \\ &+ \frac{\alpha(1-\beta)}{\Gamma(\beta)} \frac{\partial}{\partial t} A(\alpha - 1, \alpha - 1, 2, \beta + t - 1; \alpha) \Big|_{t=0}. \end{aligned}$$

We provide in Figures 4a-b a numerical investigation to identify how the parameter values change the shapes of the Rényi entropy of X for some parameter ranges. To evaluate the values of $I_R(\gamma)$ we consider the random variable X having the ZBOLL-W distribution given in equation (2.1).

7. Order statistics

Suppose X_1, \dots, X_n is a random sample from the ZBOLL-G family. Denote the random variables in ascending order as $X_{1:n} \leq \dots \leq X_{n:n}$. The pdf of $X_{i:n}$ is given by (David and Nagarajah, 2003)

$$\begin{aligned} f_{i:n}(x) &= K f(x) F^{i-1}(x) \{1 - F(x)\}^{n-i} = K \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x) F(x)^{j+i-1} \\ (7.1) \quad &= \sum_{r,k=0}^{\infty} \sum_{j=0}^{n-i} m_{j,r,k} h_{r+k+1}(x), \end{aligned}$$

where $K = n! / [(i-1)!(n-i)!]$, $h_{r+k+1}(x)$ denotes the exp-G density function with power parameter $r+k+1$ and

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

where b_k is defined by (4.3). Here, the quantities $f_{j+i-1,k}$ are obtained recursively by $f_{j+i-1,0} = b_0^{j+i-1}$ and (for $k \geq 1$)

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

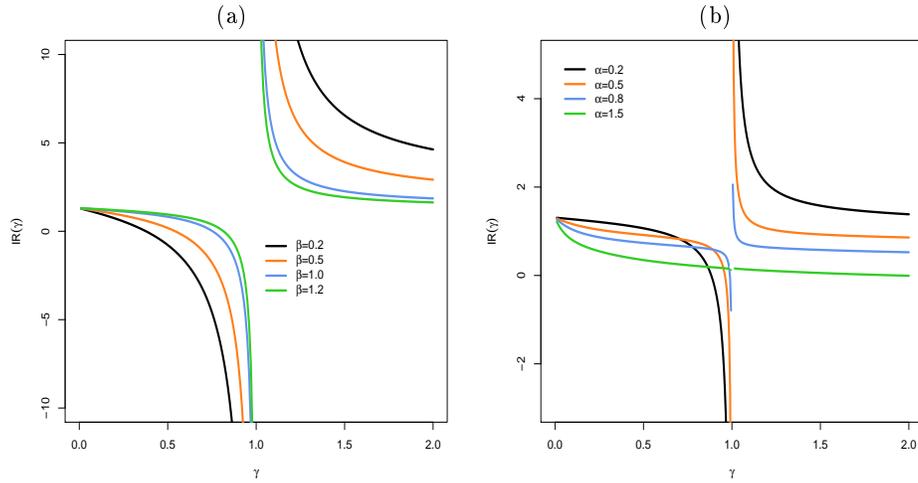


Figure 4. The Rényi entropy of X as function of γ for $\lambda = 1.5, \kappa = 3.5$ and: (a) $\alpha = 0.2$ for some values of β ; (b) $\beta = 1.5$ for some values of α .

Thus, one can easily obtain ordinary and incomplete moments and generating function of ZBOLL-G order statistics from (7.1) for any parent G .

8. Maximum likelihood estimation

In this section, we determine the maximum likelihood estimates (MLEs) of the model parameters of the new family from complete samples only. Let x_1, \dots, x_n be observed values from the ZBOLL-G family with parameters α, β and τ . Let $\theta = (\alpha, \beta, \tau^\top)^\top$ be the $r \times 1$ parameter vector. The total log-likelihood function for θ is given by

$$\begin{aligned}
 \ell_n(\theta) = \ell_n &= n \log(\alpha) - n \log[\Gamma(\beta)] + \sum_{i=1}^n \log[g(x_i; \tau)] \\
 &+ (\alpha - 1) \sum_{i=1}^n \log[G(x_i; \tau)] + (\alpha - 1) \sum_{i=1}^n \log[1 - G(x_i; \tau)] \\
 &- 2 \sum_{i=1}^n \log\{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha\} \\
 (8.1) \quad &+ (\beta - 1) \sum_{i=1}^n \log \left\{ -\log \left[\frac{[1 - G(x_i; \tau)]^\alpha}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha} \right] \right\}.
 \end{aligned}$$

The log-likelihood function can be maximized either directly by using the SAS (PROC NLMIXED) or by solving the nonlinear likelihood equations obtained by differentiating (8.1). The components of the score function

$U_n(\theta) = (\partial \ell_n / \partial \alpha, \partial \ell_n / \partial \beta, \partial \ell_n / \partial \tau)^\top$ are given by

$$\begin{aligned} \frac{\partial \ell_n}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log[G(x_i; \tau)] + \sum_{i=1}^n \log[1 - G(x_i; \tau)] \\ &\quad - 2 \sum_{i=1}^n \frac{G^\alpha(x_i; \tau) \log[G(x_i; \tau)] + [1 - G(x_i; \tau)]^\alpha \log[1 - G(x_i; \tau)]}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha} \\ &\quad + (\beta - 1) \sum_{i=1}^n \frac{G^\alpha(x_i; \tau) \log \left\{ \frac{[1 - G(x_i; \tau)]}{G(x_i; \tau)} \right\}}{[G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha] \log \left\{ \frac{[1 - G(x_i; \tau)]^\alpha}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha} \right\}}, \\ \frac{\partial \ell_n}{\partial \beta} &= -n\psi(\beta) + \sum_{i=1}^n \log \left\{ -\log \left[\frac{[1 - G(x_i; \tau)]^\alpha}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha} \right] \right\}, \\ \frac{\partial \ell_n}{\partial \tau} &= \sum_{i=1}^n \frac{[\dot{g}(x_i; \tau)]_\tau}{g(x_i; \tau)} - (\alpha - 1) \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_\tau}{G(x_i; \tau)} \\ &\quad - 2\alpha \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_\tau \{G^\alpha(x_i; \tau) - [1 - G(x_i; \tau)]^{\alpha-1}\}}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha} \\ &\quad + \alpha(\beta - 1) \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_\tau [G^\alpha(x_i; \tau) + G^{\alpha-1}(x_i; \tau)]}{[G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha] \log \left\{ \frac{[1 - G(x_i; \tau)]^\alpha}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha} \right\}}, \end{aligned}$$

where

$$\begin{aligned} [\dot{g}(x_i; \tau)]_\alpha &= \frac{dg(x_i; \tau)}{d\alpha}, & [\dot{G}(x_i; \tau)]_\alpha &= \frac{dG(x_i; \tau)}{d\alpha}, \\ [\dot{g}(x_i; \tau)]_\beta &= \frac{dg(x_i; \tau)}{d\beta}, & [\dot{G}(x_i; \tau)]_\beta &= \frac{dG(x_i; \tau)}{d\beta}, \\ [\dot{g}(x_i; \tau)]_\tau &= \frac{dg(x_i; \tau)}{d\tau}, & [\dot{G}(x_i; \tau)]_\tau &= \frac{dG(x_i; \tau)}{d\tau}, \end{aligned}$$

and the functions $g(\cdot)$ and $G(\cdot)$ are defined in Section 1 and $\psi(\cdot)$ is the digamma function.

The MLE $\hat{\theta}$ of θ is obtained by solving the nonlinear likelihood equations $U_\alpha(\theta) = 0$, $U_\beta(\theta) = 0$ and $U_\tau(\theta) = \mathbf{0}$. These equations cannot be solved analytically and statistical software can be used to solve them numerically. We can use iterative techniques such as a Newton-Raphson type algorithm to obtain the estimate $\hat{\theta}$. We employ the numerical procedure NLMixed in SAS.

For interval estimation of (α, β, τ) and hypothesis tests on these parameters, we obtain the observed information matrix since the expected information matrix is very complicated and requires numerical integration. The $(p + 2) \times (p + 2)$ observed information matrix $J(\theta)$, where p is the dimension of the vector τ , becomes

$$J(\theta) = - \begin{pmatrix} \mathbf{L}_{\alpha\alpha} & \mathbf{L}_{\alpha\beta} & \mathbf{L}_{\alpha\tau} \\ \cdot & \mathbf{L}_{\beta\beta} & \mathbf{L}_{\beta\tau} \\ \cdot & \cdot & \mathbf{L}_{\tau\tau} \end{pmatrix},$$

whose elements are given in Appendix B.

Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the asymptotic distribution of $(\hat{\theta} - \theta)$ is $N_{p+2}(\mathbf{0}, I(\theta)^{-1})$, where $I(\theta)$ is the expected information matrix. The multivariate normal $N_{p+2}(\mathbf{0}, J(\hat{\theta})^{-1})$ distribution, where $I(\theta)$ is replaced by $J(\hat{\theta})$, i.e., the observed information matrix evaluated at $\hat{\theta}$, can be used to construct approximate confidence intervals for the individual parameters.

We can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain likelihood ratio (LR) statistics for testing some special models of the proposed family. Tests of the hypotheses of the type $H_0 : \psi = \psi_0$ versus $H : \psi \neq \psi_0$, where ψ is a subset of parameters of θ , can be performed through LR statistics in the usual way.

9. Applications

In this section, we use two real data sets to compare the fits of the ZBOLL-G family with others commonly used lifetime models. In each case, the parameters are estimated by maximum likelihood (Section 8) using the subroutine NLMixed in SAS. First, we describe the data sets and give the MLEs (the corresponding standard errors and 95% confidence intervals) of the model parameters and the values of the Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC) and Kolmogorov-Smirnov (K-S) statistics. The lower the values of these criteria, the better the fit. Note that over-parametrization is penalized in these criteria, so that the two additional parameters in the proposed family do not necessarily lead to smaller values of these statistics. Next, we perform LR tests for testing some special models. Finally, we provide the histograms of the data sets to have a visual comparison of the fitted density functions.

9.1. Application 1: Zootechnics data. The data come from the zootechnics records of a Brazilian company engaged in raising beef cattle, where the farms stocked with the Nelore breed are located in the States of Bahia and São Paulo. In the analysis, only data on females born in 2000 were used and the age at first calving was the reproductive characteristic analyzed. In this case, the response variable is the logarithm of the age of the cows at first calving (measured in days). The first calving age is an important characteristic for beef cattle breeders because the faster cows reach reproductive maturity, the more calves they will produce during their breeding cycle and the greater the breeder's return on investment will be. Further, this trait is easy and inexpensive to measure. The sample size in this study is $n = 897$.

First, we describe the descriptive statistics of the data in Table 1. They suggest negatively skewed distributions with different degrees of variability, skewness and kurtosis. Then, we report the MLEs (and the corresponding standard errors in parentheses) of the

Table 1. Descriptive statistics.

Mean	Median	Mode	Variance	Skewness	Kurtosis	Min.	Max.	n
1004.32	1053.0	1074.0	13838.9	-0.405	-0.139	722	1453	897

parameters in Table 2. Additionally, we compare the models using the AIC, CAIC, BIC and K-S statistics (see Table 3). The figures in this table indicate that the ZBOLL-W model gives the best fit among the fitted models.

A comparison of the proposed distribution with some of its sub-models using LR statistics is performed in Table 4. The figures in this table, specially the p-values, reveal that the ZBOLL-W model gives a better fit to these data than the other three sub-models.

More information is provided by a visual comparison of the histogram of the data with the fitted density functions. The plots of the fitted ZBOLL-W, OLL-W, gamma-W and Weibull density functions are displayed in Figure 5. We also conclude that the ZBOLL-W distribution provides an adequate fit to these data.

Table 2. Estimates of the parameters, standard errors in [·] and 95% confidence intervals in (·) for the zootechnics data.

Model	α	β	κ	λ
ZBOLL-W	86.8186 [4.4833] (78.0196, 95.6175)	0.1612 [0.0136] (0.1346, 0.1879)	0.4420 [0.035] (0.3735, 0.5106)	2582.45 [168.38] (2251.98, 2912.91)
OLL-W	1.3982 [0.1074] (1.1874, 1.6089)	1 - -	7.5136 [0.4975] (6.5372, 8.4899)	1068.10 [5.1663] (1057.97, 1078.24)
Gamma-W	1 - -	2.0976 [0.3808] (1.3503, 2.8449)	6.4239 [0.6520] (5.1443, 7.7035)	924.76 [40.0309] (846.20, 1003.33)
Weibull	1 - -	1 - -	9.4418 [0.2198] (9.0104, 9.8732)	1054.36 [3.9260] (1046.07, 1062.07)

Table 3. The AIC, CAIC, BIC and K-S statistics for the zootechnics data.

Model	AIC	CAIC	BIC	K-S	<i>p</i> -value
ZBOLL-W	10838	10839	10857	0.1381	<0.001
OLL-W	11081	11082	11095	0.1519	<0.001
Gamma-W	11078	11079	11092	0.1717	<0.001
Weibull	11103	11104	11113	0.1595	<0.001

Table 4. LR statistics for the zootechnics data.

Model	Hypotheses	Statistic w	<i>p</i> -value
ZBOLL-W vs OLL-W	$H_0 : \beta = 1$ vs $H_1 : H_0$ is false	244.0	<0.00001
ZBOLL-W vs Gamma-W	$H_0 : \alpha = 1$ vs $H_1 : H_0$ is false	241.0	<0.00001
ZBOLL-W vs Weibull	$H_0 : \alpha = \beta = 1$ vs $H_1 : H_0$ is false	268.0	<0.00001

9.2. Application 2: Temperature data. The variable temperature ($^{\circ}\text{C}$) corresponding to daily data for the period from January 1 to December 31, 2011, obtained from the weather station of the Department of Biosystem Engineering of the Luiz de Queiroz School of Agriculture (ESALQ) of the University of São Paulo (USP), located in the City of Piracicaba, at latitude $22^{\circ}42'30''\text{S}$, longitude $47^{\circ}38'30''\text{W}$ and altitude of 546 meters. First, we describe the data set in Table 5.

Table 5. Descriptive statistics.

Mean	Median	Mode	Variance	Skewness	Kurtosis	Min.	Max.	n
22.32	22.90	19.25	8.71	-0.50	-0.73	14.68	27.25	365

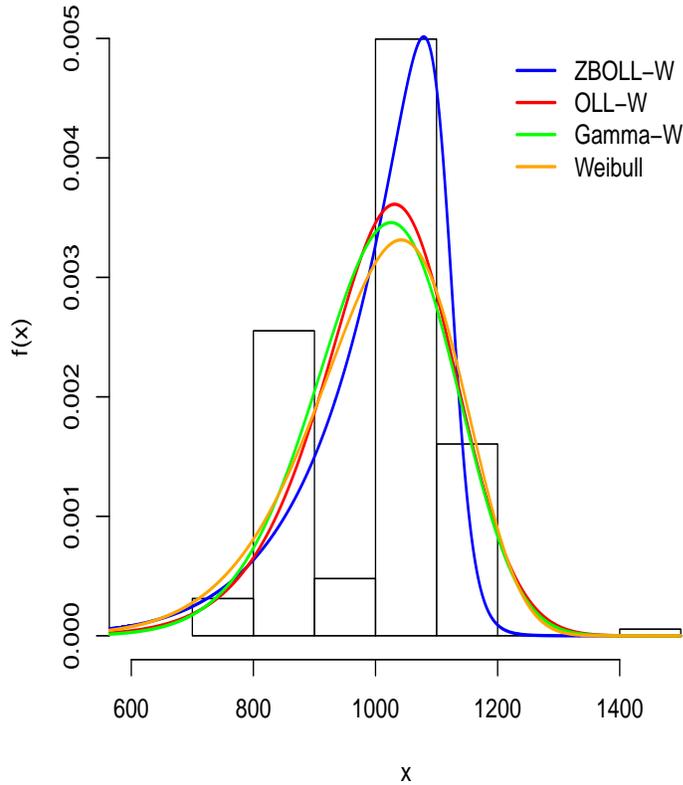


Figure 5. (a) Fitted ZBOLL-W, OLL-W, gamma-W and Weibull densities for the zootechnics data.

For these data, we compare the fitted ZBOLL-N, OLL-N, gamma-N and normal distributions. The MLEs of μ and σ for the normal distribution are taking as starting values for the iterative procedure to fit the ZBOLL-N, OLL-N and gamma-N models. The MLEs of the parameters, standard errors and 95% confidence intervals for the parameters are given in Table 6. Additionally, we compare the models using the AIC, CAIC, BIC and K-S statistics (see Table 7). Since the values of these statistics are smaller for the ZBOLL-N distribution compared to those values of the other models (see Table 6), the new distribution produces a fit to the current data quite better than its special models.

A comparison of the proposed distribution with some of its sub-models using LR statistics is performed in Table 8. The figures in this table, specially the p-values, indicate that the ZBOLL-N model gives a better fit to these data than the other three sub-models.

Table 6. Estimates of the parameters, standard errors in [·] and 95% confidence intervals in (·) for the ZBOLL-N model and its special models and three criteria for the temperature data.

Model	α	β	μ	σ
ZBOLL-N	0.1783 [0.0262] (0.1183, 0.2382)	1.3744 [0.1355] (1.1579, 1.5907)	21.0200 [0.3640] (20.4648, 21.5757)	0.9293 [0.0729] (0.7422, 1.1163)
OLL-N	0.1861 [0.0448] (0.0958, 0.2763)	1 - -	21.9071 [0.1220] (21.6668, 22.1474)	0.8915 [0.1281] (0.6332, 1.1498)
Gamma-N	1 - -	0.1246 [0.0069] (0.1109, 0.1383)	26.698 [0.1910] (26.1933, 26.9446)	1.4409 [0.0319] (1.3782, 1.5037)
Normal	1 - -	1 - -	22.3271 [0.1542] ()	2.9463 [0.1090] ()

Table 7. AIC, CAIC, BIC and K-S statistics for the temperature data.

Model	AIC	CAIC	BIC	K-S	p -values
ZBOLL-N	1777.9	1778.1	1793.5	0.0617	0.0731
OLL-N	1790.4	1791.4	1802.1	0.1108	0.0002
Gamma-N	1797.6	1797.7	1809.3	0.0818	0.0151
Normal	1828.7	1829.7	1836.5	0.1029	0.0005

Table 8. LR statistics for the temperature data.

Model	Hypotheses	Statistic w	p -value
ZBOLL-N vs OLL-N	$H_0 : \beta = 1$ vs $H_1 : H_0$ is false	14.5	0.00014
ZBOLL-N vs Gamma-N	$H_0 : \alpha = 1$ vs $H_1 : H_0$ is false	22.0	<0.00001
ZBOLL-N vs Normal	$H_0 : \alpha = \beta = 1$ vs $H_1 : H_0$ is false	54.8	<0.00001

More information is provided by a visual comparison of the histogram of the data and the fitted density functions. The plots of the fitted ZBOLL-N, OLL-N, gamma-N and normal densities are displayed in Figure 6. We conclude that the ZBOLL-N distribution provides the best fit to these data.

10. Conclusions

In this paper, we propose a new family of distributions with two extra generator parameters, which includes as special cases all classical continuous distributions. For any parent continuous distribution G , we define the so-called *Zografos-Balakrishnan odd log-logistic-G* family with two extra positive parameters. The new family extends several widely known distributions and some of its special models are discussed. We demonstrate that the new family density function is a linear mixture of exponentiated- G densities.

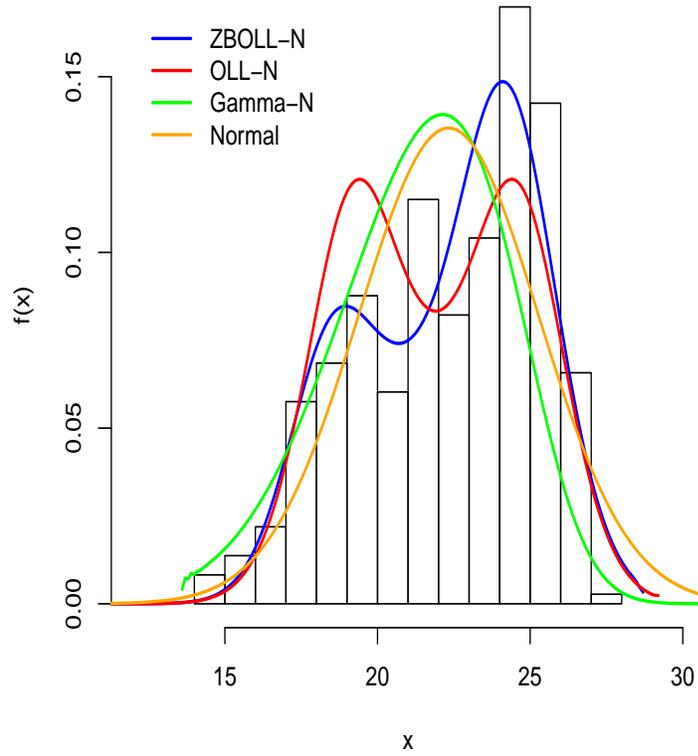


Figure 6. (a) Fitted ZBOLL-N, OLL-N, gamma-N and normal densities for the zootechnics data.

We obtain some of its mathematical properties, which include ordinary and incomplete moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves, two types of entropies and order statistics. The application of the new family is straightforward. The model parameters are estimated by maximum likelihood. Two real examples are used for illustration, where the new family does fit well both data sets.

Appendix A: Three useful power series

We present three power series required for the algebraic developments in Section 3 and 6. First, for $b > 0$ real non-integer and $-1 < u < 1$, we have the binomial expansion

$$(10.1) \quad (1 - u)^a = \sum_{j=0}^{\infty} (-1)^j \binom{a}{j} u^j,$$

where the binomial coefficient is defined for any real.

Second, expanding z^λ in Taylor series, we can write

$$(10.2) \quad z^\lambda = \sum_{k=0}^{\infty} (\lambda)_k (z-1)^k / k! = \sum_{i=0}^{\infty} f_i z^i$$

where

$$(10.3) \quad f_i = f_i(\lambda) = \sum_{k=i}^{\infty} \frac{(-1)^{k-i}}{k!} \binom{k}{i} (\lambda)_k$$

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

Third, we obtain an expansion for $[G(x)^a + \bar{G}(x)^a]^c$. We can write from equation (10.2) and (10.1)

$$(10.4) \quad [G(x)^a + \bar{G}(x)^a] = \sum_{j=0}^{\infty} t_j G(x)^j,$$

where $t_j = t_j(a) = a_j(a) + (-1)^j \binom{a}{j}$ and $a_j(a)$ is defined by (10.2). Then, using (10.2), we have

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

where $f_i = f_i(c)$.

Finally, using again equations (10.3) and (10.4), we have

$$(10.5) \quad [G(x)^a + \bar{G}(x)^a]^c = \sum_{j=0}^{\infty} h_j(a, c) G(x)^j,$$

where $h_j(a, c) = \sum_{i=0}^{\infty} f_i m_{i,j}$ and (for $i \geq 0$) $m_{i,j} = (j t_0)^{-1} \sum_{m=1}^j [m(j+1) - j] t_m m_{i,j-m}$ (for $j \geq 1$) and $m_{i,0} = t_0^i$.

Appendix B

The elements of the observed information matrix $J(\theta)$ for the parameters (α, β, τ) are given by

$$\begin{aligned} J_{\alpha\alpha} &= \frac{-n}{\alpha^2} \\ &- 2 \sum_{i=1}^n \frac{G^\alpha(x_i; \tau) [1 - G(x_i; \tau)]^\alpha \left\{ \log[G(x_i; \tau)] \log\left[\frac{G(x_i; \tau)}{1 - G(x_i; \tau)}\right] \right\}}{[G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha]^2} \\ &- 2 \sum_{i=1}^n \frac{G^\alpha(x_i; \tau) [1 - G(x_i; \tau)]^\alpha \left\{ \log[1 - G(x_i; \tau)] \log\left[\frac{1 - G(x_i; \tau)}{G(x_i; \tau)}\right] \right\}}{[G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha]^2}, \\ J_{\alpha\beta} &= - \sum_{i=1}^n \frac{G^\alpha(x_i; \tau) \log\left[\frac{G(x_i; \tau)}{G(x_i; \tau)}\right]}{[G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha] \log\left[1 - \frac{G^\alpha(x_i; \tau)}{G^\alpha(x_i; \tau) + [1 - G(x_i; \tau)]^\alpha}\right]}, \end{aligned}$$

$$\begin{aligned}
J_{\alpha\tau} &= \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_{\tau}}{G(x_i; \tau)} - \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_{\tau}}{1 - G(x_i; \tau)} \\
&- 2 \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_{\tau} [G^{\alpha-1}(x_i; \tau) - [1 - G(x_i; \tau)]^{\alpha-1}]}{G^{\alpha}(x_i; \tau) + [1 - G(x_i; \tau)]^{\alpha}} \\
&- 2\alpha \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_{\tau} G^{\alpha}(x_i; \tau) [1 - G(x_i; \tau)]^{\alpha} \log \left[\frac{G(x_i; \tau)}{1 - G(x_i; \tau)} \right]}{[G^{\alpha}(x_i; \tau) + [1 - G(x_i; \tau)]^{\alpha}]^2}
\end{aligned}$$

$$J_{\beta\beta} = -n\psi'(\beta)$$

$$J_{\beta\tau} = -\alpha \sum_{i=1}^n \frac{[\dot{G}(x_i; \tau)]_{\tau} G^{\alpha-1}(x_i; \tau)}{[1 - G(x_i; \tau)] [G^{\alpha}(x_i; \tau) + [1 - G(x_i; \tau)]^{\alpha}] \log \left[1 - \frac{G^{\alpha}(x_i; \tau)}{G^{\alpha}(x_i; \tau) + [1 - G(x_i; \tau)]^{\alpha}} \right]}$$

$$\begin{aligned}
J_{\tau\tau} &= (\alpha - 1) \sum_{i=1}^n \left\{ \frac{[\ddot{G}(x_i; \tau)]_{\tau\tau}}{G(x_i; \tau)} - \frac{[\dot{G}(x_i; \tau)]_{\tau}^2}{[G(x_i; \tau)]^2} \right\} \\
&+ (\alpha - 1) \sum_{i=1}^n \left\{ \frac{[\ddot{G}(x_i; \tau)]_{\tau\tau}}{[1 - G(x_i; \tau)]} + \frac{[\dot{G}(x_i; \tau)]_{\tau}^2}{[1 - G(x_i; \tau)]^2} \right\} \\
&- \beta \sum_{i=1}^n \left\{ \frac{[\ddot{G}(x_i; \tau)]_{\tau\tau}}{[1 - G(x_i; \tau)]^2} + \frac{2[\dot{G}(x_i; \tau)]_{\tau}^2}{[1 - G(x_i; \tau)]^3} \right\} + \sum_{i=1}^n \left\{ \frac{[\dot{g}(x_i; \tau)]_{\tau\tau}}{g(x_i; \tau)} - \frac{[\dot{g}(x_i; \tau)]_{\tau}^2}{[g(x_i; \tau)]^2} \right\},
\end{aligned}$$

where

$$[\dot{g}(x_i; \tau)]_{\tau} = \frac{dg(x_i; \tau)}{d\tau}, \quad [\dot{G}(x_i; \tau)]_{\tau} = \frac{dG(x_i; \tau)}{d\tau}, \quad [G(x_i; \tau)]^2 = \frac{dG(x_i; \tau)}{d\tau} \left(\frac{dG(x_i; \tau)}{d\tau} \right)^T,$$

$$[\dot{g}(x_i; \tau)]_{\tau\tau} = \frac{d^2g(x_i; \tau)}{d\tau^2}, \quad [\ddot{G}(x_i; \tau)]_{\tau\tau} = \frac{d^2G(x_i; \tau)}{d\tau^2},$$

and $g(\cdot)$ and $G(\cdot)$ are defined in Section 1.

References

- [1] Aas, K., and I. Haff (2006). The generalized hyperbolic skew student's t-distribution. *Journal of Financial Econometrics*, **4**, 275-309.
- [2] Alexander, C., Cordeiro, G.M., Ortega, E.M.M. and Sarabia, J.M. (2012). Generalized beta-generated distributions. *Computational Statistics and Data Analysis*, **56**, 1880-1897.
- [3] Cordeiro, G.M. and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, **81**, 883-898.
- [4] Cordeiro, G.M., Nadarajah, S. and Ortega, E.M.M. (2013). General results for the beta Weibull distribution. *Journal of Statistical Computation and Simulation*, **83**, 1082-1114.
- [5] David, H.A., Nagaraja. (2003). *Order Statistics*, third edition. John Wiley and Sons, New Jersey.
- [6] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics - Theory and Methods*, **31**, 497-512.
- [7] Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Table of Integrals, Series, and Products*, sixth edition. Academic Press, San Diego.
- [8] Gupta, R.C., Gupta, R.D., Gupta, P.L. (1998). Modeling failure time data by Lehmann alternatives. *Communications in Statistics, Theory and Methods*, **27**, 887-904.
- [9] Gupta, R.D., Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, **41**, 173-188.

- [10] Gupta, R.D., Kundu, D. (2001). Exponentiated exponential distribution: an alternative to gamma and Weibull distributions. *Biometrical Journal*, **43**, 117-130.
- [11] Hansen, B.E. (1994). Autoregressive conditional density estimation. *International Economic Review*, **35**, 705-730.
- [12] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Volume 1, 2nd edition. John Wiley and Sons, New York.
- [13] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Volume 2, 2nd edition. John Wiley and Sons, New York.
- [14] Kakde, C.S and Shirke, D.T. (2006). On Exponentiated Lognormal distribution. *International Journal of Agricultural and Statistics Sciences*, **2**, 319-326.
- [15] Marshall, A.W., Olkin, I. (2007). *Life Distributions: Structure of Nonparametric, Semi-parametric, and Parametric Families*. Spinger-Verlag, New York.
- [16] Mudholkar, G.S., Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-real data. *IEEE Transaction on Reliability*, **42**, 299-302.
- [17] Mudholkar, G.S., Srivastava, D.K., Kollia, G.D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of American Statistical Association*, **91**, 1575-1583.
- [18] Nadarajah, S., Cordeiro, G.M., Ortega, E.M.M. (2015) The Zografos-Balakrishnan-G family of distributions: Mathematical properties and applications. *Communications in Statistics - Theory and Methods*, **44**, 186-215.
- [19] Nadarajah, S. (2005). The exponentiated Gumbel distribution with climate application. *Environmetrics*, **17**, 13-23.
- [20] Nadarajah, S., Gupta, A.K. (2007). The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin*, **59**, 29-54.
- [21] Nadarajah, S., Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97-111.
- [22] Ristic, M.M. and Balakrishnan, N. (2012). The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, , 1191-1206.
- [23] Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, **6**, 344-362.

