

THE EXPONENTIATED GENERALIZED EXTENDED PARETO DISTRIBUTION

Thiago A. N. De Andrade*, Luz M. Zea²

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

²Universidade Federal do Rio Grande do Norte Departamento de Estatística, Lagoa Nova, 59078-970, Natal, RN, Brazil

ABSTRACT

We define and study a three-parameter model with positive real support called the exponentiated generalized extended Pareto distribution. We provide a comprehensive mathematical treatment and prove that the formulas related to the new model are simple and manageable. We study the behaviour of the maximum likelihood estimates for the model parameters using Monte Carlo simulation. We take advantage of applied studies and offer two applications to real data sets that proves empirically the power of adjustment of the new model when compared to another twelve lifetime distributions.

Keywords: Applied studies, Exponentiated generalized model, Extended Pareto distribution, Lehmann's alternatives, Monte Carlo simulation.

* Corresponding author . Thiago A. N. De Andrade
Email: thiagoan.andrade@gmail.com

1 Introduction

The extended Pareto (EP) distribution, also known as generalized Pareto, is a very popular statistical model due to its wide use in practical situations. Its flexibility to model real phenomena has become an attraction for applied researchers. This distribution was introduced by Pickands [24] in the seventies, who studied the distribution of samples that exceeded a certain upper limit [13]. Since then, many works considering the EP model have been published. Here, we refer to some recent works: [13], [12], [17], [28] and [22] to mention a few. The cumulative distribution function (cdf) of the EP distribution is given by (for $x \in \mathbb{R}$)

$$G(x; \xi) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}}, & \text{if } \xi > 0 \\ 1 - \exp(-x), & \text{if } \xi = 0 \end{cases} \quad (1)$$

Its probability density function (pdf) becomes

$$g(x; \xi) = \begin{cases} (1 + \xi x)^{-\frac{1}{\xi}-1}, & \text{if } \xi > 0 \\ \exp(-x), & \text{if } \xi = 0 \end{cases} \quad (2)$$

where $\xi \in \mathbb{R}$ is the shape parameter.

The recent statistical literature offers a broad arsenal of univariate continuous distributions that can be (and indeed are) widely used in several data modeling. However, this large number of available distributions does not seem to deal with the huge variety of data arising from several fields such as medicine, engineering, demography, biology, actuarial, economics, finance, reliability, among others [5]. Indeed, statistical and applied researchers have shown great interest in building new extended continuous distributions, which are more flexible for data modeling [2]. There are several ways to extend well-known distributions. Probably, one of the most popular ways is to consider distribution generators. Here, we refer to the following papers: [18] for the Marshall-Olkin class, [14] for the beta class, [30], [25] and [21] for the gamma class and [1] for the Kumaraswamy class of distributions.

More recently, for any baseline cdf $G(x)$, and $x \in \mathbb{R}$, [7] defined the exponentiated generalized (EG) class with two extra shape parameters. The EG family with baseline cdf $G(x)$ and two extra parameters $a > 0$ and $b > 0$ has cdf and pdf given by

$$F(x) = \{1 - [1 - G(x)]^a\}^b \quad (3)$$

and

$$f(x) = ab[1 - G(x)]^{a-1} \{1 - [1 - G(x)]^a\}^{b-1} g(x). \quad (4)$$

We have some reasons to believe that the EG model may be more appropriate than some generators to extend well-known distributions. Next, we will provide some properties of the EG class that, in our opinion, make this one of the first choices among available generators. First, it is important to highlight the simplicity of equations (3) and (4). They do not involve any sophisticated function and the generated distribution will be simple whenever $G(x)$ is

simple. This is not the case of the beta-G class, which, although quite popular, does not have, for example, cdf in closed-form. In addition, another important aspect is that the EG model contains as special cases the two classes of Lehmann's alternatives. In fact, for $a = 1$, (3) reduces to $F(x) = G(x)^b$ and, for $b = 1$, we obtain $F(x) = 1 - [1 - G(x)]^a$, which correspond to the cdf's of the Lehmann type I and II families [16], respectively. The additional shape parameters in equation (4) have the important role of controlling the tails of the distribution, while still allowing entropy to be added at the center of the distribution. As we shall see later, these parameters have a considerable impact on model asymmetry and kurtosis.

The above properties and many others have been discussed and explored in recent works for the EG class. Here, we refer to the papers and baseline distributions: [4] for the Birnbaum-Saunders distribution, [8] for the generalized gamma model, [27] for the Dagum distribution, [10] for the Gumbel model, [6] for the arcsine distribution, [9] for an extended exponential model, [23] for the exponential distribution, [26] for the half-normal model.

In the context of generators, two important extensions are the beta EP (BEP) and Kumaraswamy EP (KwEP) distributions investigated by [17] and [22], respectively. Based on a similar motivation, we study another promising extension of the EP distribution using the EG generator called the exponentiated generalized extended Pareto (EGEP for short) distribution, by inserting equation (1) in equation (3). Setting $\xi = 0$ in equation (1), we have the exponential distribution, which was studied by many authors in recent years, whereas setting $\xi < 0$ leads to a support for X that depends on unknown parameters. For these reasons, we consider only the EP model with $\xi > 0$ and positive support.

Two points here need to be clear. The first important point to be highlighted is that the model we are considering here is not exactly new. It appears as special cases of several important distributions. However, no author has studied this particular case in detail, so little reporter this model as "The Exponentiated Generalized Extended Pareto distribution". Here, we follow a specific methodology, namely the one proposed by [7], to give rise to the EGEP model. The second important point to be considered is that our study of the EGEP model has very clear and forceful motivations:

1. We study the structural properties of the EGEP model and we verify that all the formulas associated to the proposed model are simple and manageable using computational resources.
2. The EGEP model we are studying has only three parameters and does not have any complicated form for the density, cumulative and likelihood functions, among others. This represents a gain, since it makes it easier to obtain analytical and numerical results.
3. The simulated study showed that the EGEP model has attractive properties for the maximum likelihood estimators.
4. Application studies with real data were performed. The EGEP model was compared to more than ten major models and proved superior to all of them in terms of fit. This indicates

that the EGEP distribution deserves special attention in the literature, due to its great potential for adjustment to real data.

For the reasons listed above, we strongly believe that it is important to study in detail the EGEP distribution.

The paper is organized as follows. In Section 2, we define the EGEP model. In Section 3, we study the forms of the pdf and hazard rate function (hrf) by means of their first and second derivatives. In Section 4, we determine the quantile function (qf) and investigate its behavior. Some properties, such as linear representations for the pdf and cdf, ordinary and incomplete moments, moment generating function (mgf), Rényi entropy and order statistics are presented in Section 5. Section 6 deals with estimation and inference. Two applications to real data and a simulation study are given in Section 7. In Section 8, we offer some conclusions.

2 The EGEP distribution

By taking the EP distribution as the baseline in equation (3), we obtain the EGEP model with cdf (for $x > 0$) given by

$$F(x) = [1 - (1 + \xi x)^{-\frac{a}{\xi}}]^b, \quad (5)$$

where $a > 0$, $b > 0$ and $\xi > 0$. The EGEP density reduces to

$$f(x) = ab(1 + \xi x)^{-\frac{(a+\xi)}{\xi}} [1 - (1 + \xi x)^{-\frac{a}{\xi}}]^{b-1}. \quad (6)$$

Henceforth, we denote by X a random variable having pdf (6). For $a = b = 1$, equation (6) reduces to the EP density. The EGEP model also includes the Lehmann type I and type II transformations of the EP distribution, denoted by EEPI and EEPII, when $a = 1$ and $b = 1$, respectively. Some plots of the density (6) are displayed in Figure 1. They reveal that the pdf of X is quite flexible and can take asymmetric forms, among others. In summary, they reinforce the importance of the proposed model.

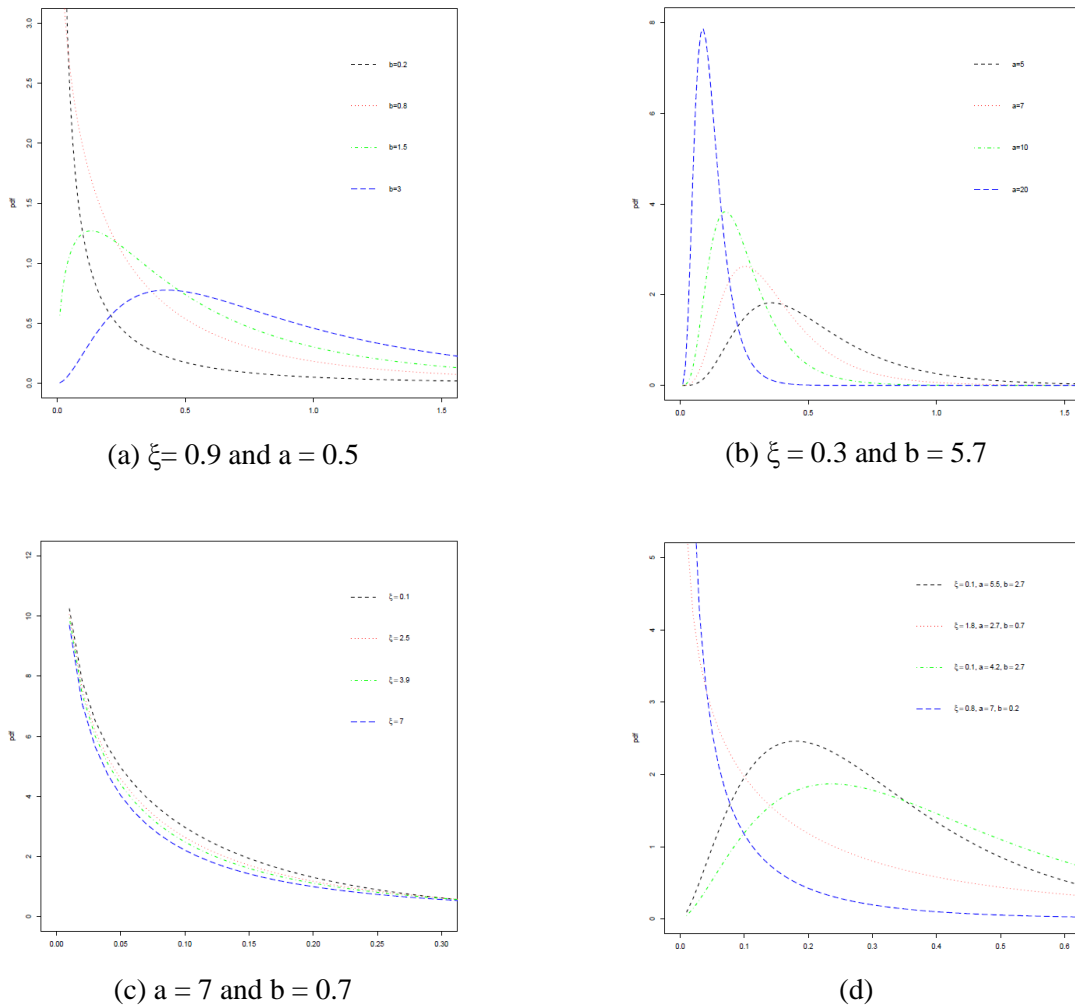


Figure 1: Plots of the EGEP density function for some parameter values.

Besides the cdf (5) and pdf (6), other functions can be used to feature the EGEP model as the survival function and hrf. These are particularly important to analyze survival data that involve the time associated to an event of interest such as the time until failure of a certain component, until the death of a patient or a disease relapse. The hrf of X is given by

$$h(x) = \frac{ab(1 + \xi x)^{-\frac{(a+\xi)}{\varepsilon}} [1 - (1 + \xi x)^{-\frac{a}{\xi}}]^{b-1}}{1 - [1 - (1 + \xi x)^{-\frac{a}{\varepsilon}}]^b}. \tag{7}$$

Some plots of the hrf (7) are displayed in Figure 2. Besides monotone forms, the hrf of X can take an inverted bathtub shape. This non-monotone form is particularly important because of its great practical applicability.

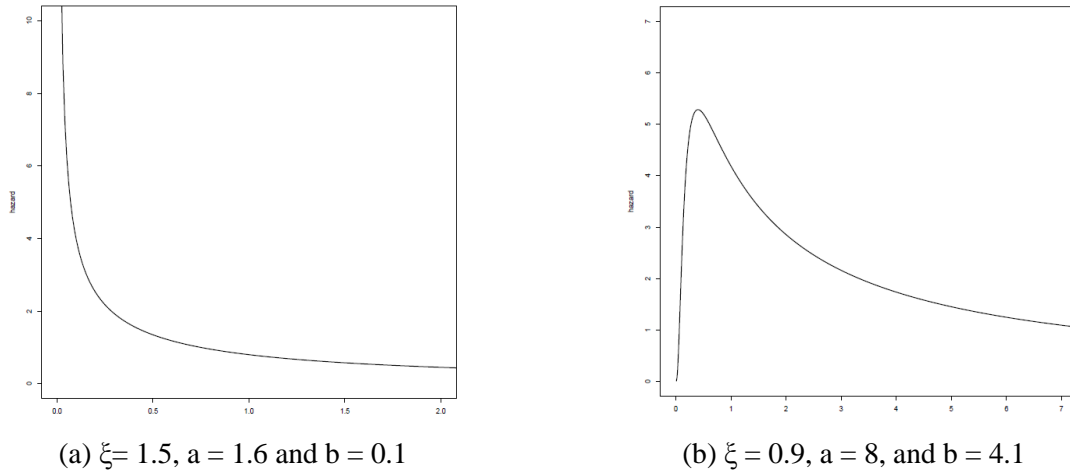


Figure 2: Plots of the EGEP hazard function for some parameter values.

3 Shapes

The different forms of the pdf and hrf of X can be investigated through their first and second derivatives. After some algebraic calculations, the first and second derivatives of $\log[f(x)]$ can be expressed as

$$\frac{d \log[f(x)]}{dx} = -\frac{a + \xi}{1 + \xi x} + \frac{a(b-1)z(x)^{-\frac{(a+\xi)}{\xi}}}{1 - z(x)^{-\frac{a}{\xi}}},$$

where $z(x) = 1 + \xi x$, and

$$\frac{d^2 \log[f(x)]}{dx^2} = -\frac{\xi(a + \xi)}{(a + \xi x)^2} - \frac{a(b-1)z(x)^{-\frac{(a+\xi)}{\xi}}}{1 - z(x)^{-\frac{a}{\xi}}} \left[(a + \xi) + \frac{a z(x)^{-\frac{a}{\xi}}}{1 - z(x)^{-\frac{a}{\xi}}} \right].$$

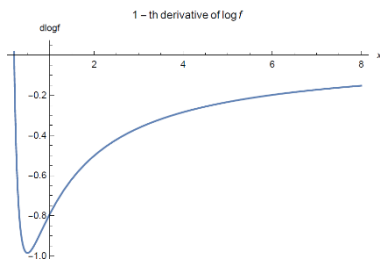
It is often difficult to obtain an analytical solution for the critical points of this function. Therefore, it is common to obtain numerical solutions with high accuracy through optimization routines in most mathematical and statistical platforms. Some plots of the first derivative of $\log[f(x)]$ for selected parameter values are display in Figure 3.

Similarly, the first and second derivatives of $\log[h(x)]$ can be expressed as

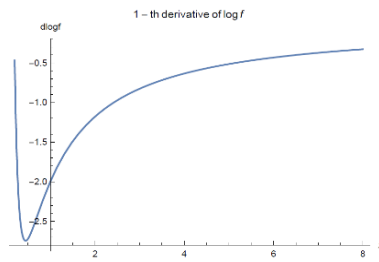
$$\frac{d \log[h(x)]}{dx} = -\frac{(a + \xi)}{z(x)} + \frac{a(b-1)z(x)^{-\frac{(a+\xi)}{\xi}}}{1 - z(x)^{-\frac{a}{\xi}}} - \frac{ab(1 - z(x)^{-\frac{a}{\xi}})^{b-1}}{1 - (1 - z(x)^{-\frac{a}{\xi}})^b},$$

and

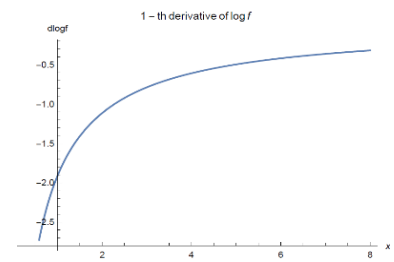
$$\frac{d^2 \log[h(x)]}{dx^2} = -\frac{\xi(a + \xi)}{z(x)^2} + \frac{a(b - 1)z(x)^{-\frac{(a+2\xi)}{\xi}}}{1 - z(x)^{-\frac{a}{\xi}}} \left[(a + \xi) - \frac{a z(x)^{-\frac{a}{\xi}}}{1 - z(x)^{-\frac{a}{\xi}}} \right] - \frac{ab(1 - z(x)^{-\frac{a}{\xi}})^{b-2}}{1 - (1 - z(x)^{-\frac{a}{\xi}})^b} \left[a(b - 1)z(x)^{-\frac{(a+\xi)}{\xi}} \right] + \frac{ab(1 - z(x)^{-\frac{a}{\xi}})^b z(x)^{-\frac{(a+\xi)}{\xi}}}{1 - (1 - z(x)^{-\frac{a}{\xi}})^b}$$



(a) a = 1.81, b = 2.05 and $\xi = 4.69$



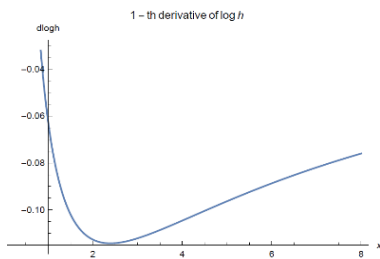
(b) a = 8, b = 4.08 and $\xi = 4.69$



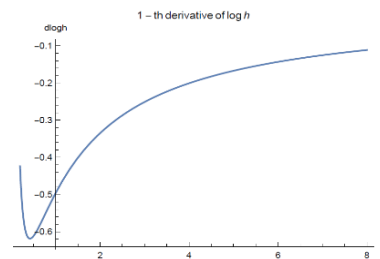
(c) a = 3.95, b = 0.85 and $\xi = 2.34$

Figure 3: Plots of the first derivative of $\log[f(x)]$.

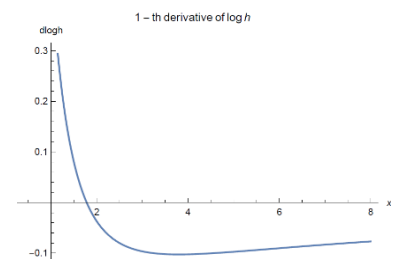
Some plots of the first derivative of $\log[h(x)]$ for selected parameter values are displayed in Figure 4.



(a) a = 0.8, b = 1.27 and $\xi = 0.2$



(b) a = 5.44, b = 1.27 and $\xi = 1.02$



(c) a = 3.46, b = 8 and $\xi = 8$

Figure 4: Plots of the first derivative of $\log[h(x)]$.

4 Quantile function

The qf of X can be expressed as

$$Q(u) = [(1 - u^b)^{-\frac{\xi}{a}} - 1]/\xi.$$

We use the qf of X to determine the Bowley's skewness [15] (B) and Moors' kurtosis [19] (M). These measures are given by

$$B = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{1}{2}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)} \quad \text{and} \quad M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}$$

In Figures 5, 6, 7 and 8, we present 3D plots of the B and M measures for selected baseline parameter values. These plots are obtained using the Wolfram Mathematica software. Based on these plots, it is possible to conclude that, for fixed baseline parameters values, changes in the additional parameters a and b have a considerable impact on the skewness and kurtosis of the EGEP model, thus corroborating for its greater flexibility. So, these plots reinforce the importance of the additional parameters.

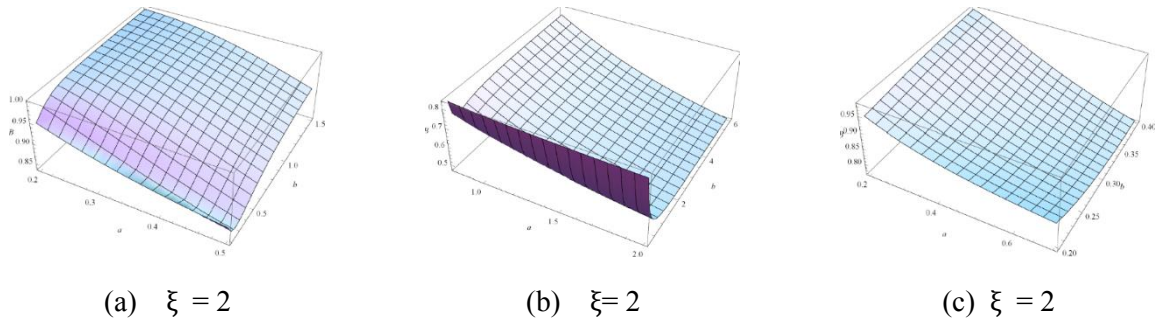


Figure 5: Plots of the Bowley's skewness for the EGEP model with $\xi = 2$.

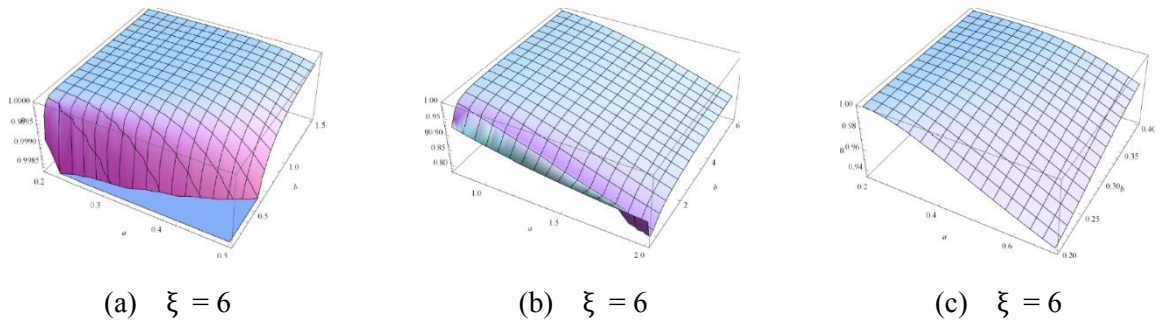


Figure 6: Plots of the Bowley's skewness for the EGEP model with $\xi = 6$.

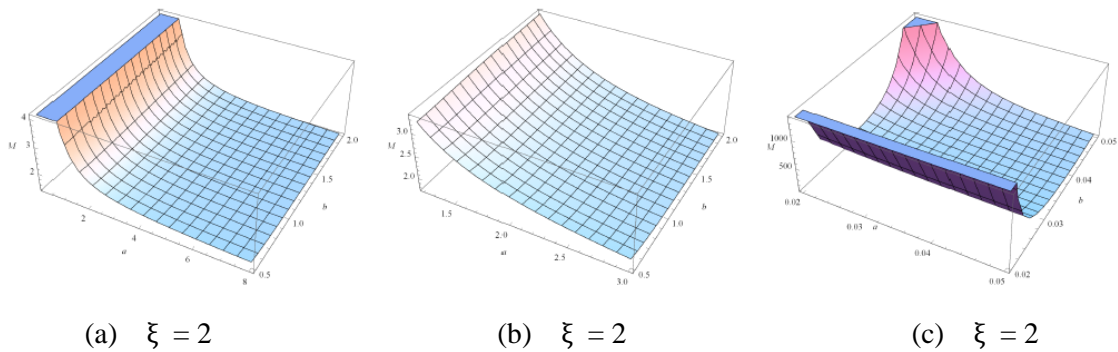


Figure 7: Plots of the Moors' kurtosis for the EGEP model with $\xi = 2$.

5 Properties

5.1. Linear representations

Often it may be difficult or even impossible to obtain explicit analytical solutions for certain mathematical quantities using directly equation (6). For this reason, it is very common to adopt expansions that allow to obtain simpler expressions for these quantities. In this section, we present simple expressions for equations (3) and (4) to facilitate the derivation of some structural properties for the EGEP model. Our starting point is the class of exponentiated-G (exp-G) distributions, also called Lehmann type I transforms. This class has been vigorously studied since the 1950s, and from that time, several exp-G distributions were investigated.

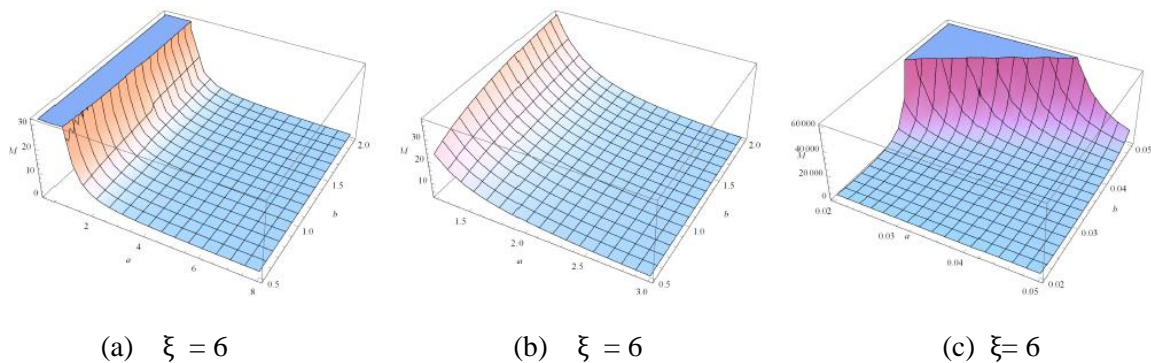


Figure 8: Plots of the Moors' kurtosis for the EGEP model with $\xi = 6$.

For an arbitrary baseline cdf $G(x)$, a random variable Y_c has the exp-G distribution with power parameter $c > 0$, say $Y_c \text{ exp-G}(c)$, if its cdf and pdf are given by $H_c(x) = G(x)^c$ and $h_c(x) = cg(x)G(x)^{c-1}$, respectively. For a comprehensive discussion about the exponentiated distributions, see a recent paper by [29]. The density $h_{j+1}(x)$ of the exponentiated extended Pareto (exp-EP) with power parameter $j + 1$ (for $j \geq 0$) is given by

$$h_{j+1}(x) = (j + 1)(1 + \xi x)^{-\frac{1}{\xi}-1} [1 - (1 + \xi x)^{-\frac{1}{\xi}}]^j.$$

Based on some results in [4], we can express the EGEP pdf (6) (for $\xi > 0$) as

$$f(x) = \sum_{j=0}^{\infty} w_{j+1} h_{j+1}(x), \tag{8}$$

where $w_{j+1} = \sum_{m=1}^{\infty} (-1)^{j+m+1} \binom{b}{m} \binom{ma}{j+1}$.

Equation (8) is the main result of this section. It gives the EGEP density as a linear combination of Exp-EP densities. So, some mathematical properties of the EGEP distribution can be easily obtained from those of the Exp-EP model.

5.2. Moments

The n th moment of X can be determined from (8) as

$$E(X^n) = \frac{1}{\xi^{n+1}} \sum_{k=0}^{\infty} \rho_k B\left(n+1, \frac{(k+1-n\xi)}{\xi}\right), \quad (9)$$

where

$$\rho_k = \sum_{j=k}^{\infty} (-1)^k (j+1) \omega_{j+1} \binom{j}{k}, \quad (10)$$

and $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$ is the beta function. Equation (9) holds for $n < 1/\xi$.

5.3. Incomplete moments

The n th incomplete moment of X follows from (8) as

$$T_n(z) = z^{n+1} n! \sum_{k=0}^{\infty} \rho_k {}_2F_1\left(n+1, \frac{\xi+k+1}{\xi}; n+2; -z\xi\right),$$

where ρ_k is given in (10), and ${}_2F_1(a, b, c; z)/\Gamma(c)$ is the regularized hypergeometric function. In the Mathematica platform, we can write this function as `Hypergeometric2F1Regularized[a, b; c; z]`. See the Wolfram website for more details about this function: <http://mathworld.wolfram.com/RegularizedHypergeometricFunction.html>.

5.4. Generating function

An important characterization of X is given by its mgf, say $M(t) = E(e^{tx})$. Using the linear representation (8), we obtain (for $t < 0$)

$$M(t) = \frac{e^{-\frac{t}{\xi}}}{\xi} \sum_{k=0}^{\infty} \rho_k \mathbb{E}\left[\xi + k + \frac{1}{\xi}, -\frac{t}{\xi}\right],$$

where ρ_k is given in equation (10) and $\mathbb{E}[\delta, z] = \int_1^{\infty} t^{-\delta} e^{-zt} dt$ is the exponential integral function. See Wolfram website: <http://mathworld.wolfram.com/En-Function.html>. On the Mathematica platform, we can write `ExpIntegralE[δ , z]`.

5.5. Rényi entropy

Given a certain random phenomenon under study, it is important to quantify the uncertainty associated with the random variable of interest. One of the most popular measures used to quantify the variability of X is the Rényi entropy. The Rényi entropy of X with density (6), say $I_R(\rho)$, is given by

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

where $\rho > 0$ and $\rho \neq 1$.

Next, we consider the generalized binomial expansion

$$(1 - z)^b = \sum_{k=0}^\infty (-1)^k \binom{b}{k} z^k, \tag{11}$$

which holds for any real non-integer b and $|z| < 1$. Using (11) twice in equation (4), we can write

$$f(x)^\rho = (ab)^\rho \sum_{k,l=0}^\infty (-1)^{k+l} \binom{\rho(b-1)}{k} \binom{ak + \rho(a-1)}{l} g(x)^\rho G(x)^l \tag{12}$$

Substituting (1) and (2) in equation (12), and after some algebra, we have

$$I_R(\rho) = \frac{1}{(1 - \rho)} \log \left[(ab)^\rho \sum_{k,l=0}^\infty \sum_{m=0}^l (-1)^{k+l+m} \binom{l}{m} \binom{\rho(b-1)}{k} \binom{ak + \rho(a-1)}{l} \times \int_0^\infty (1 + \epsilon x)^{-\frac{[\rho(\epsilon+1)+m]}{\epsilon}} dx \right]$$

and then

$$I_R(\rho) = \frac{1}{(1 - \rho)} \log \left[\sum_{k,l=0}^\infty \sum_{m=0}^l \frac{(-1)^{k+l+m} (ab)^\rho}{m + \rho + \epsilon(\rho - 1)} \binom{l}{m} \binom{\rho(b-1)}{k} \binom{ak + \rho(a-1)}{l} \right].$$

5.6. Order statistics

The importance of the order statistics and their applications are widely explored in the literature. The density function $f_{i:n}(x)$ of the i th order statistic, say $X_{i:n}$, based on a random sample X_1, \dots, X_n can be expressed as (for $i = 1, \dots, n$)

$$f_{i:n}(x) = \frac{f(x)}{B(i, n - i + 1)} F(x)^{i-1} [1 - F(x)]^{n-i},$$

where $f(x)$ is the pdf (4) and $F(x)$ is the cdf (3).

It was demonstrated by [4] that the density of the EG order statistics can written as

$$f_{i:n}(x) = \sum_{l=0}^\infty q_l h_{l+1}(x), \tag{13}$$

where q_l is given by

$$q_l = \frac{ab}{(l+1)B(i, n-i+1)} \sum_{j=0}^{n-i} \sum_{k=0}^{\infty} (-1)^{j+k+l} \binom{n-i}{j} \binom{b(i+j)-1}{k} \binom{a(k+1)-1}{l},$$

and $h_{l+1}(x)$ denotes the exp-G density function with power parameter $l+1$ (for $l \geq 0$).

Equation (13) reveals that the density function of the EG order statistic is a linear combination of exp-G densities.

6 Estimation and inference

Several approaches for parameter estimation are available in the literature but the maximum likelihood method is the most commonly employed. The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used when constructing confidence intervals and also in test statistics. Let x_1, \dots, x_n be a sample of size n from EGEP distribution. The log-likelihood function for the vector of parameters $\theta = (a, b, \xi)^T$ can be expressed as

$$l(\theta) = n \log(ab) + \left(\frac{a+\xi}{\xi}\right) \sum_{i=1}^n \log(z_i) + (b-1) \sum_{i=1}^n \log\left(1 - z_i^{-\frac{a}{\xi}}\right)$$

where $z_i = 1 + \xi x_i$

The elements of the score vector are given by

$$U_a = \frac{\partial l(\theta)}{\partial a} = \frac{n}{a} + \frac{1}{\xi} \sum_{i=1}^n \log(z_i) + (b-1) \sum_{i=1}^n \frac{z_i^{-a/\xi} \log(z_i)}{\xi(1 - z_i^{-a/\xi})}$$

$$U_b = \frac{\partial l(\theta)}{\partial b} = \frac{n}{b} + \sum_{i=1}^n \log(1 - z_i^{-a/\xi})$$

$$U_\xi = \frac{\partial l(\theta)}{\partial \xi} = \frac{1}{\xi} \sum_{i=1}^n \log(z_i) - \frac{(a+\xi)}{\xi^2} \sum_{i=1}^n \log(z_i) + \frac{(a+\xi)}{\xi} \sum_{i=1}^n \frac{x_i}{z_i} - \frac{a(b-1)}{\xi^2} \sum_{i=1}^n \frac{z_i^{-\frac{a}{\xi}} (\log(z_i) - \xi x_i / z_i)}{1 - z_i^{-a/\xi}}$$

For interval estimation and hypothesis tests on the parameters a , b and ξ , we determine the 3×3 observed information matrix given by $J(\theta) = \{-U_{r,s}\}$, whose elements $U_{r,s} = \frac{\partial^2 l(\theta)}{\partial_r \partial_s}$ for $r, s \in \{a, b, \xi\}$ can be obtained from the authors upon request.

7 Applied results

7.1. Simulation study

In this section we present a simulation study that investigate the behavior of the MLEs of a , b and ε parameters. We simulate 10.000 Monte Carlo Samples. We consider sample sizes $n = 100, 200, 300, 500$ and $n = 800$. For the generating process we take $a = 0.5, b = 0.5$ and $\varepsilon = 0.8$.

7.2. Application

In this section, we fit the EGEP model to a two real data sets. First, we investigate the quality of adjustment of the EGEP distributions when compared to other generalizations of the EP model. For our initial comparative study, we consider four-parameter models, namely the beta extended Pareto model (BEP) proposed by [17] and the kumaraswamy extended Pareto model (KwEP) pioneered by [22]. We also fit the three-parameter model that arises by inserting the EP baseline distribution in the gamma-G family. So, we consider the gamma extended Pareto (GEP) distribution (by [11]).

The first data set was obtained from [20], and consists of the times between failures for repairable items: 1.43, 0.11, 0.71, 0.77, 2.63, 1.49, 3.46, 2.46, 0.59, 0.74, 1.23, 0.94, 4.36, 0.40, 1.74, 4.73, 2.23, 0.45, 0.70, 1.06, 1.46, 0.30, 1.82, 2.37, 0.63, 1.23, 1.24, 1.97, 1.86, 1.17. Table 2 gives some descriptive statistics for these data.

In Table 3, we present the MLEs (and the corresponding standard errors in parentheses) of the fitted models and also the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) statistics. Based on these results, it is noted that the EGEP model has the lowest values of the AIC, BIC and CAIC statistics, so it could be chosen as the best model for the time failure data.

In Table 4, we present formal goodness-of-fit tests to the times between failures. The values of the Kolmogorov-Smirnov (K-S), Cramér-von Mises (W^*) and Anderson-Darling (A^*) statistics indicate that the fitted models are competitive models to the time failure data, although the EGEP model has the advantage of less parameters.

Table 1: Means, standard error estimates, Biases and RMSEs of \hat{a} , \hat{b} and $\hat{\xi}$ for the EGEP model.

n	Parameter	Mean	Biases	StandardError	RMSE
100	a	0.6045	0.1045	0.3994	0.1705
	b	0.5262	0.0262	0.1094	0.0127
	ξ	1.0047	0.2047	0.8952	0.8433
200	a	0.5396	0.0396	0.1951	0.0396
	b	0.5111	0.0111	0.0660	0.0045
	ξ	0.8731	0.0731	0.4057	0.1700
300	a	0.5258	0.0258	0.1502	0.0232
	b	0.5073	0.0073	0.0522	0.0028
	ξ	0.8473	0.0473	0.3126	0.1000
500	a	0.5157	0.0157	0.1113	0.0126
	b	0.5048	0.0048	0.0393	0.0016
	ξ	0.8289	0.0289	0.2305	0.0540
800	a	0.5094	0.0094	0.0856	0.0074
	b	0.5028	0.0028	0.0308	0.0010
	ξ	0.8172	0.0172	0.1761	0.0313

Table 2: Descriptive statistics for the time failure data.

Statistic	
Mean	1.5430
Median	1.2350
Variance	1.2717
Minimum	0.1100
Maximum	4.7300

Table 3: MLEs (standard errors in parentheses) and the AIC, BIC and CAIC statistics for the time failure data.

Distribution	$\hat{\xi}$	$\hat{\sigma}$	\hat{a}	\hat{b}	AIC	BIC	CAIC
GEP	0.010 (0.168)	0.756 (0.448)	1.989 (0.721)	-	85.258	89.461	86.181
KwEP	1.119 (2.371)	4.972 (4.817)	1.917 (0.837)	14.432 (12.687)	87.283	92.888	88.883
EGEP	0.001 (0.235)	-	1.005 (0.482)	2.107 (0.894)	85.232	89.436	86.155
BEP	0.156 (1.960)	1.061 (3.852)	1.979 (0.460)	13.850 (1.983)	87.257	92.862	88.857

Table 4: Goodness-of-fit statistics for the GEP, BEP, KwEP and EGEP models for the time failure

Model	data		
	Statistics		
	K-S	W*	A*
GEP	0.018	0.018	0.140
BEP	0.018	0.018	0.140
KwEP	0.017	0.017	0.128
EGEP	0.018	0.018	0.133

Next, we consider an uncensored data set corresponding to remission times (in months) of a random sample of 128 bladder cancer patients. These data were previously studied by [3]. Some descriptive statistics are presented in Table 5.

Table 5: Descriptive statistics for the remission times data.

Statistic	
Mean	9.366
Median	6.395
Variance	110.425
Minimum	0.080
Maximum	79..050

For the second comparative study, we consider EGEP, GGP, BGP, KwGP, exponentiated Weibull (EW), Kumaraswamy Burr XII (KwBXII), exponentiated exponential (EE), Kumaraswamy Weibull (KwW), beta Weibull (BW), gamma Weibull (GW), Kumaraswamy Gamma (KwGamma), beta Gamma (BGamma) and gamma Gamma (GGamma) models.

Table 6 provides the MLEs (and the corresponding standard errors in parentheses) of the fitted models and also the values of the following statistics: AIC, BIC and CAIC. Based on these measures, we note that the EGEP model has the lowest values of the AIC, BIC and CAIC statistics. This suggests that the EGEP model could be chosen as the best model for the remission times data.

The goodness-of-fit measures for the current data are given in Table 7. We use the values of the K-S, W* and A* statistics. In general, the lowest values of these statistics indicate the best fitted model to the data. Hence, we can conclude from the figures in Table 7 that the fitted EGEP model is better than the other fitted models.

In summary, we have three important conclusions from these applications. First, the proposed model is quite adequate to fit the current data sets. Second, the EGEP model is superior to other extensions of the EP model using other generators. In other words, the EG class provides a generalization of the EP model more flexible in terms of adjustment than

those obtained with the Kw-G, beta-G and gamma-G generators. Finally, the third important conclusion is that even considering models that do not take the EP distribution as the base model, our distribution has superior adjustment power.

Table 6: MLEs (standard errors in parentheses) and the AIC, BIC and CAIC statistics for the remission times data.

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\xi}$	$\hat{\theta}$	$\hat{\theta}$	AIC	BIC	CAIC
EGEP	0.193 (0.043)	1.647 (0.292)	0.041 (0.026)	-	-	826.290	834.846	826.484
GGP	4.463 (1.424)	1.569 (0.285)	0.196 (0.088)	-	-	826.189	834.745	826.383
BGP	2.521 (0.511)	14.218 (0.592)	5.939 (2.182)	23.876 (12.403)	-	835.569	846.977	835.894
KwGP	1.578 (0.211)	8.912 (4.161)	2.027 (1.215)	23.535 (3.106)	-	827.941	839.349	828.266
EW	2.854 (1.488)	0.457 (0.273)	0.654 (0.151)	-	-	827.386	835.942	827.579
KwBXII	4.344 (3.904)	12.632 (10.730)	0.556 (0.341)	0.850 (1.344)	3.751 (2.898)	832.015	846.275	832.507
EE	-	-	1.221 (0.149)	0.121 (0.014)	-	830.156	835.860	830.252
KwW	8.180 (2.584)	8.805 (5.268)	0.274 (0.064)	2.348 (1.743)	-	829.290	840.698	829.615
BW	0.685 (0.149)	2.913 (1.918)	2.631 (1.164)	0.800 (0.499)	-	829.398	840.806	829.723
GW	3.508 (0.776)	-	0.538 (0.063)	0.743 (0.509)	-	827.719	836.275	827.913
KwGamma	25.045 (0.535)	0.618 (0.084)	0.058 (0.008)	7.113 (0.581)	-	830.371	841.779	830.696
BGamma	0.065 (0.011)	4.236 (0.011)	19.655 (0.755)	0.384 (0.039)	-	830.275	841.683	830.601
GGamma	1.090 (0.718)	-	1.078 (0.710)	8.004 (1.107)	-	832.719	841.275	832.912

8 Conclusions

In this article we propose and study a new three-parameter distribution with positive support called the exponentiated generalized extended Pareto (EGEP) distribution. This model extends the generalized Pareto distribution introduced by Pickands [24] in the seventies. Our proposal includes as special cases both Lehmann's type I and II transformations, which not only makes the model more flexible but also facilitates the achievement of its mathematical properties. We study the main mathematical properties of the new model, such as, linear representations for the pdf and cdf, ordinary and in- complete moments, moment generating function, Rényi entropy and order statistics. The model parameters are estimated by the maximum likelihood method. A simulation study reveals that the estimators have desirable properties such as small biases and variances even in moderate sample sizes. We prove empirically that the new distribution provides a better fit to a real data set than other competitive models.

Table 7: Goodness-of-fit statistics for the GEP, BEP, KwEP and EGEP models for the air

Model	conditioning data		
	Statistics		
	W*	K-S	A*
EGEP	0.028	0.038	0.188
GGP	0.030	0.040	0.199
BGP	0.096	0.068	0.659
KGP	0.025	0.035	0.168
EW	0.043	0.043	0.288
KwBXII	0.041	0.040	0.293
EE	0.112	0.073	0.673
KwW	0.043	0.048	0.281
BW	0.044	0.047	0.289
GW	0.049	0.047	0.317
KwGama	0.060	0.049	0.375
BGama	0.056	0.050	0.358
GGama	0.120	0.074	0.717

References

We thank two anonymous referees and the associate editor for their valuable suggestions, which certainly contributed to the improvement of this paper. Additionally, Thiago A. N. de Andrade is grateful for the financial support from CAPES (Brazil).

- [1] Cordeiro, G. M., and de Castro, M. A new family of generalized distributions. *Journal of Statistical Computation and Simulation* 81 (2011), 883–893.
- [2] Cordeiro, G. M., Hashimoto, E. M., and Ortega, E. M. The McDonald Weibull model. *Statistics: A Journal of Theoretical and Applied Statistics* 48 (2014), 256–278.
- [3] Cordeiro, G. M., and Lemonte, A. J. The β -Birnbau-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis* 55 (2011), 1445–1461.
- [4] Cordeiro, G. M., and Lemonte, A. J. The exponentiated generalized Birnbau-Saunders distribution. *Applied Mathematics and Computation* 247 (2014), 762–779.
- [5] Cordeiro, G. M., Lemonte, A. J., and Ortega, E. M. An extended fatigue life distribution. *Statistics: A Journal of Theoretical and Applied Statistics* 47 (2013), 626–653.
- [6] Cordeiro, G. M., and Lemonte, A. J. and Campelo, A. K. Extended arcsine distribution to proportional data: Properties and applications. *Studia Scientiarum Mathematicarum Hungarica* 53 (2016), 440–466.
- [7] Cordeiro, G. M., Ortega, E. M. M., and Cunha, D. C. C. The exponentiated generalized class of distributions. *Journal of Data Science* 11 (2013), 1–27.
- [8] da Silva, R. V., Gomes-Silva, F., Ramos, M. W. A., and Cordeiro, G. M. A new extended gamma generalized model. *International Journal of Pure and Applied Mathematics* 100 (2015), 309–335.
- [9] De Andrade, T. A. N., Bourguignon, M., and Cordeiro, G. M. The exponentiated generalized extended exponential distribution. *Journal of Data Science* 14 (2016), 393–414.
- [10] De Andrade, T. A. N., Rodrigues, H., Bourguignon, M., and Cordeiro, G. M. The exponentiated generalized gumbel distribution. *Revista Colombiana de Estadística* 38 (2015), 123–143.

-
- [11] De Andrade, T. A. N., Zea, L. M., Gomes-Silva, S., and Cordeiro, G. M. The gamma generalized pareto distribution with applications in survival analysis. *International Journal of Statistics and Probability* 6 (2017), 141–156.
- [12] de Zea Bermudez, P., and Kotz, S. Parameter estimation of the generalized pareto distribution—part ii. *Journal of Statistical Planning and Inference* 140 (2010), 1374–1388.
- [13] de Zea Bermudez, P., and Kotz, S. Parameter estimation of the generalized Pareto distribution-Part I. *Journal of Statistical Planning and Inference* 140 (2010), 1353–1373.
- [14] Eugene, N., Lee, C., and Famoye, F. Beta-normal distribution and its applications. *Communications in Statistics - Theory and Methods* 31 (2002), 497–512.
- [15] Kenney, J. F., and Keeping, E. S. *Mathematics of Statistics*. 3rd ed., Part 1, New Jersey, 1962.
- [16] Lehmann, E. L. The power of rank tests. *Annals of Mathematical Statistics* 24 (1953), 23–43.
- [17] Mahmoudi, E. The beta generalized Pareto distribution with application to lifetime data. *Mathematics and Computers in Simulation* 81 (2011), 2414–2430.
- [18] Marshall, A. N., and Olkin, I. A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrika* 84 (1997), 641–652.
- [19] Moors, J. J. A quantile alternative for kurtosis. *J. Royal Statist. Soc. D* 37 (1988), 25–32.
- [20] Murthy, D., Xie, M., and Jiang, R. *Weibull models*. Wiley series in probability and statistics. Hoboken (NJ): JohnWiley and Sons, 2004.
- [21] Nadarajah, S., Cordeiro, G. M., and Ortega, E. M. M. The Zografos- Balakrishnan-G Family of Distributions: Mathematical Properties and Applications. *Communications in Statistics - Theory and Methods* 44 (2015), 186–215.
- [22] Nadarajah, S., and Eljabri, S. The Kumaraswamy GP Distribution. *Journal of Data Science* 11 (2013), 739–766.
- [23] Oguntunde, P. E., Adejumo, A. O., and Adepoju, K. A. Assessing the flexibility of the exponentiated generalized exponential distribution. *The Pacific Journal of Science and Technology* 17 (2016), 49–57.

- [24] Pickands, J. Statistical inference using extreme order statistics. *Annals of Statistics* 3 (1975), 119–131.
- [25] Ristic, M. M., and Balakrishnan, N. The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, doi: 10.1080/00949655.2011.574633 (2011).
- [26] Sanchez, J. J. D., Freitas, W. L., and Cordeiro, G. M. The extended generalized half-normal distribution. *Brazilian Journal of Probability and Statistics* 30 (2016), 366–384.
- [27] Silva, A. O., da Silva, L. C. M., and Cordeiro, G. M. The extended dagum distribution: Properties and application. *Journal of Data Science* 13 (2015), 53–72.
- [28] Song, J., and Song, S. A quantile estimation for massive data with generalized Pareto distribution. *Computational Statistics & Data Analysis* 56 (2012), 143–150.
- [29] Tahir, M. H., and Nadarajah, S. Parameter induction in continuous univariate distributions: Well-established G families. *Anais da Academia Brasileira de Ciências* 87 (2015), 539–568.
- [30] Zografos, K., and Balakrishnan, N. On families of Beta- and generalized Gamma-generated distributions and associated inference. *Statistical Methodology* 6 (2009), 344–362.