Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 1, 84–109
DOI: 10.1214/11-BJPS157

The exp-*G* family of probability distributions

Wagner Barreto-Souza[a] **and Alexandre B. Simas**[b]

[a]*Universidade de São Paulo*
[b]*Universidade Federal da Paraíba*

Abstract. In this paper we introduce a new method to add a parameter to a family of distributions. The additional parameter is completely studied and a full description of its behaviour in the distribution is given. We obtain several mathematical properties of the new class of distributions such as Kullback–Leibler divergence, Shannon entropy, moments, order statistics, estimation of the parameters and inference for large sample. Further, we show that the new distribution has the reference distribution as special case, and that the usual inference procedures also hold in this case. We present a comprehensive study of two special cases of the exp-*G* class: exp-Weibull and exp-beta distributions. Further, an application to the real data set is presented. This family also opens a wide variety of research, as the authors may develop its special cases in full detail.

1 Introduction

The present work is an enhanced and extended version of the pioneering manuscript presented at Estância de São Pedro, São Paulo, Brazil, in the 18° SINAPE, 2008; see Barreto-Souza et al. (2008). One year after the primary version of this work was published in the annals of the SINAPE, Nadarajah et al. (2009) published a few similar results in a technical report.

In many practical situations, usual probability distributions do not provide an adequate fit. For example, if the data are asymmetric, normal distribution will not be a good choice. With this, several methods of introducing a parameter to expand a family of distributions have been studied.

Marshall and Olkin (1997) introduced a new way to expand probability distributions and applied to yield a two-parameter extension of the exponential distribution which can serve as a competitor to such commonly-used two-parameter distributions as the Weibull, gamma and lognormal distributions. Furthermore, this method was used to obtain a three-parameter extension of the Weibull distribution. Moreover, Mudholkar et al. (1996) introduced a three-parameter distribution alternative to the Weibull distribution, that has the Weibull as limiting distribution.

Some methods of introducing of parameters to symmetric distributions have been studied in order to add skewness. For instance, Azzalini (1985) introduced

Key words and phrases. Exp-*G* distribution, order statistics, Kullback–Leibler divergence, Fisher's information matrix.

Received August 2010; accepted May 2011.

and studied the well-known skew-normal distribution, which is obtained by adding a shape parameter to the normal distribution. Another symmetric distribution that was extended by adding a skewness parameter was the Student's t-distribution by Jones and Faddy (2003). Finally, Ma and Genton (2004) introduced a general class of skew-symmetric distributions, whereas Ferreira and Steel (2006) provided a general perspective on the introduction of skewness into symmetric distributions.

Jones (2004) introduced a class of distributions that adds two parameters to a reference distribution. Further, Jones and Pewsey (2009) advanced a four-parameter family that has both symmetric and skewed members and allows for tail weights that are both heavier and lighter than those of the generating distribution.

Recently, Morais and Barreto-Souza (2011) introduced a three-parameter class of distributions, so-called Weibull-power series (WPS) distributions. The WPS distributions are very useful in the modeling of lifetime data since they have a flexible hazard function that can be increasing, decreasing and upside-down bathtub shaped, among others.

In this article, we introduce a new method to add a parameter to some reference distribution. The resulting distribution exhibits the remarkable reciprocal property. We study this parameter in detail, and we give a full description of its behaviour in the distribution. The augmented distribution has several connections with the reference distribution, for instance, the Kullback–Leibler divergence of the augmented distribution with respect to the original distribution is finite and only depends on the new parameter. Several others properties in this direction are also given. The inferential aspects of this distribution are studied in detail.

Special attention must be given to the fact that it is not straightforward that this new distribution contains the reference distribution as a special case. We show that this is the case if we enlarge the parameter space, and, also, that this enlargement is good, in the sense that all the standard inferential procedures work if this new value in the parameter space is considered to be the true value of the parameter.

The remainder of the article unfolds as follows: in Section 2 the new class of distributions is introduced. Several properties are given in Section 3, including a characterization of the new class and a full study with respect to the new parameter. In Section 4 we discuss estimation of the parameters and inferential aspects are carefully studied. Two special cases of the exp-G class are studied in Sections 5 and 6. In Section 7 we present an application to the real data set. The paper is concluded in Section 8. The Appendix contains the proofs of the results presented in the article.

2 The new class of distributions

The c.d.f. of a random variable with truncated exponential distribution in the interval [0, 1] with parameter λ is given by

$$F_\lambda^*(x) = \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda}}, \tag{2.1}$$

where $\lambda > 0$ and $x \in [0, 1]$. We now observe that $F^*_\lambda(\cdot)$ is a c.d.f. for $\lambda \in \mathbb{R} \setminus \{0\}$, and that

$$\lim_{\lambda \to 0} F^*_\lambda(x) = x, \qquad x \in [0, 1].$$

Therefore, we extend the parameter space of the distribution above for the entire line:

$$F_\lambda(x) = \begin{cases} F^*_\lambda(x), & \text{if } \lambda \neq 0, \\ x, & \text{if } \lambda = 0. \end{cases}$$

We now define the new class as follows. Let $G(x;\theta)$ be the c.d.f. of a continuous or discrete random variable with θ being the parameters related to G, then the class of distributions exp-G, indexed by λ, is defined by

$$F^G_\lambda(x) = F_\lambda(G(x;\theta)). \tag{2.2}$$

From now on, we will denote a random variable X with c.d.f. (2.2) by $X \sim$ exp-$G(\Theta)$, where $\Theta = (\lambda, \theta)^T$.

If $G(x;\theta)$ is a c.d.f. of a continuous random variable, then the exp-G distribution is absolutely continuous for every $\lambda \neq 0$, and its probability density function (p.d.f.), which is the derivative of the c.d.f. (2.2) with respect to x, is given by

$$f(x) \equiv f_\lambda(x) = \frac{\lambda}{1 - e^{-\lambda}} g(x;\theta) \exp\{-\lambda G(x;\theta)\}, \tag{2.3}$$

where $g(\cdot;\theta)$ is the p.d.f. associated to the c.d.f. $G(\cdot;\theta)$. If $\lambda = 0$, we obtain that $f_0(x) = \lim_{\lambda\to 0} f_\lambda(x) = g(x;\theta)$. Let $G(x;\theta)$ be a c.d.f. of a discrete random variable taking values on the set $\{x_1, x_2, \ldots\}$, where $x_1 < x_2 < \cdots$, then the corresponding exp-G distribution is also discrete, taking values on the same set for every $\lambda \neq 0$, and its probability function is given by

$$P_\lambda(x_i) = \frac{\exp\{-\lambda G(x_{i-1};\theta)\} - \exp\{-\lambda G(x_i;\theta)\}}{1 - e^{-\lambda}}, \tag{2.4}$$

where $G(x_0) = 0$. For $\lambda = 0$, we have that $P_0(x_i) = \lim_{\lambda\to 0} P_\lambda(x_i) = G(x_i;\theta) - G(x_{i-1};\theta)$.

If $G(x;\theta)$ is an absolutely continuous c.d.f., then its hazard function is given by

$$h(x;\theta) \equiv h_\lambda(x;\theta) = \frac{\lambda g(x;\theta)}{1 - \exp\{-\lambda S(x;\theta)\}}, \tag{2.5}$$

where $S(x;\theta) = 1 - G(x;\theta)$ is the survival function of a random variable with c.d.f. $G(\cdot,\theta)$.

3 Properties

3.1 Relationship between exp-G and G distributions

We now state several results regarding the relationship between the exp-G and G distributions, where proofs can be found in the Appendix.

Proposition 3.1. *Let X and X_λ have G distribution and exp-G distribution with parameter λ, respectively. Let also μ be the law of X, and μ_λ be the law of X_λ. Then,*

(i) *X and X_λ have the same support for all $\lambda \in \mathbb{R}$;*
(ii) *if X is continuous, singular (see Billingsley, 1995, page 409, for a definition) or discrete, then X_λ is continuous, singular or discrete, respectively, for all $\lambda \in \mathbb{R}$;*
(iii) *$\mu \ll \mu_\lambda$, that is, μ_λ is absolutely continuous with respect to μ. Moreover, the Radon–Nikodym derivative of μ_λ with respect to μ is, almost surely,*

$$\frac{d\mu_\lambda}{d\mu}(x) = \frac{1}{1-e^{-\lambda}} \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \frac{\exp\{-\lambda G(x-\varepsilon)\} - \exp\{-\lambda G(x)\}}{G(x) - G(x-\varepsilon)};$$

(iv) *if X is continuous and $\lambda \neq 0$, the relative entropy (Kullback–Leibler divergence) between X and X_λ is*

$$D_{\mathrm{KL}}(\mu \parallel \mu_\lambda) = -\int \log \frac{d\mu_\lambda}{d\mu}\, d\mu = 1 - \frac{\lambda}{e^\lambda - 1} - \log\left(\frac{\lambda}{1-e^{-\lambda}}\right).$$

If $\lambda = 0$, we have that $D_{\mathrm{KL}}(\mu \parallel \mu_0) = 0$;
(v) *if $E(|X|^r) < \infty$, then $E(|X_\lambda|^r) < \infty$, and, moreover, if $\lambda > 0$,*

$$E(|X|^r) \geq \frac{\lambda}{1-e^{-\lambda}} E(|X_\lambda|^r),$$

and if $\lambda < 0$,

$$\frac{\lambda}{1-e^{-\lambda}} E(|X|^r) \leq E(|X_\lambda|^r).$$

For $\lambda = 0$, it follows that $E(|X|^r) = E(|X_0|^r)$.

3.2 A characterization

We now give a characterization for our class of distributions through Shannon entropy. Such entropy were introduced by Shannon (1948) and, for a random variable X with density $f(\cdot)$, with respect to a σ-finite measure μ, usually the Lebesgue or counting measure, is given by

$$\mathbb{H}_S(f) = -\int_{\mathbb{R}} f(x) \log f(x)\, d\mu. \tag{3.1}$$

Jaynes (1957) introduced one of the most powerful techniques employed in the field of probability and statistics called maximum entropy method. This method is closely related to the Shannon entropy and considers a class of density functions

$$\mathbb{F} = \{f(x) : E_f\{T_i(X)\} = \alpha_i,\ i = 0, \dots, m\}, \tag{3.2}$$

where $T_i(X)$, $i = 1, \dots, m$, are absolutely integrable functions with respect to $f\,d\mu$, and $T_0(X) = \alpha_0 = 1$. In the continuous case, the maximum entropy principle suggests to derive the unknown density function of the random variable X by the model that maximizes the Shannon entropy in (3.1), subject to the information constraints defined in the class $\mathbb{F}$.

The maximum entropy distribution is the density of the class $\mathbb{F}$, denoted by f^{ME}, which is obtained as the solution of the optimization problem

$$f^{\mathrm{ME}} = \arg\max_{f \in \mathbb{F}} \mathbb{H}_S(f).$$

Jaynes (1957), in page 623, states that the maximum entropy distribution f^{ME}, obtained by the constrained maximization problem described above, "is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have." It is the distribution which should not incorporate additional exterior information other than which is specified by the constraints.

In order to obtain a maximum entropy characterization for our class of distributions, we now derive suitable constraints. For this, the next result plays an important role. We will assume in Propositions 3.2 and 3.3 that the reference measure, μ, is the Lebesgue measure, and that all the random variables involved are continuous.

Proposition 3.2. *Let G be the distribution of a continuous random variable, with p.d.f., $g(\cdot)$, and let X be a random variable with p.d.f., $f(\cdot)$, given by (2.3). Then, we have that*

$$\text{(C1)}\quad E\{\log g(X;\theta)\} = E\{\log g(G^{-1}(U;\theta))\} = \frac{\lambda}{1-e^{-\lambda}} \int_0^1 \log g(G^{-1}(u;\theta)) e^{-\lambda u}\,du,$$

$$\text{(C2)}\quad E\{G(X;\theta)\} = \frac{1}{\lambda} - \frac{1}{e^{\lambda}-1},$$

and the Shannon entropy of $f(\cdot)$ is given by

$$\mathbb{H}_S(f) = 1 - \frac{\lambda}{e^{\lambda}-1} - \log\left(\frac{\lambda}{1-e^{-\lambda}}\right) - E\{\log g(G^{-1}(U;\theta))\}, \tag{3.3}$$

where U follows truncated exponential distribution with parameter λ and c.d.f. given by (2.1).

The next proposition shows that the class exp-G of distributions has maximum entropy in the class of all probability distributions specified by the constraints stated therein.

Proposition 3.3. *The p.d.f. $f(\cdot)$ of a random variable X, given by* (2.3), *is the unique solution of the optimization problem*

$$f = \arg\max_{h \in \mathbb{F}} \mathbb{H}_S(h),$$

under the constraints (C1) *and* (C2) *presented in Proposition* 3.2.

3.3 λ as a concentration parameter

We provide two asymptotic results of this class, by making the parameter λ tend to $\pm\infty$. These results will allow us to give an interpretation for this parameter. Since $F_\lambda(x) \to x$ as $\lambda \to 0$, we have, trivially, that if $X_\lambda^G \sim$ exp-G and $X^G \sim G$, then

$$X_\lambda^G \xrightarrow{d} X^G,$$

as $\lambda \to 0$, where $\xrightarrow{d}$ stands for convergence in distribution.

Therefore, the definition of the family exp-G by using (2.2) with $\lambda \in \mathbb{R}$ is good. This fact plays an important role in our paper because this makes the family exp-G contain G as a particular case. The following result is very important since regular distributions in Statistics enjoy many desirable properties.

Proposition 3.4. *If G is a parametric regular probability distribution, with parametric space Θ, then so is the exp-G distribution, with respect to the parametric space $\mathbb{R} \times \Theta$.*

Proof. The proof follows from a simple verification of the conditions given in Chapter 6 (Section 5) from Lehmann and Casella (2003). □

The distribution may present very different behaviour for large absolute values of λ, thus showing that this is a rich class of distributions.

Going further on the discussion of what happens when the absolute value of λ is large, we begin by noting that F_λ^G will tend to one if λ tends to infinity, whenever x is such that $G(x) > 0$, and will be zero otherwise. Therefore, if X_λ follows a exp-G distribution, where G is any c.d.f., then

$$X_\lambda \xrightarrow{v} \delta_a,$$

as $\lambda \to \infty$, where $a = \inf\{x; G(x) > 0\}$, '$\xrightarrow{v}$' stands for vague convergence, and δ_a is the Dirac's measure concentrated on a, that is, $\delta_a(\{a\}) = 1$. Note that we needed to consider the vague convergence instead of convergence in distribution to allow $a = -\infty$. If $a = -\infty$, then

$$X_\lambda \xrightarrow{v} 1,$$

where 1 is the function identically equal to one, which is not a probability measure. However, we may interpret this case as a "probability measure" concentrated at $-\infty$, that is, if a random variable would follow 1, then $\mathrm{pr}(X \leq x) = 1$ for all $x \in \mathbb{R}$.

We now obtain the asymptotic behaviour of $\lambda \to -\infty$. For this case, a simple calculus argument allows us to conclude that F_λ^G will tend to zero, whenever x is such that $G(x) < 1$, and will be 1 otherwise. Therefore, if X_λ follows a exp-G distribution, where G is any c.d.f., then

$$X_\lambda \xrightarrow{v} \delta_b,$$

as $\lambda \to -\infty$, where $b = \sup\{x; G(x) < 1\}$. Note that we also needed to use the vague convergence to include the case where $b = \infty$. In this case

$$X_\lambda \xrightarrow{v} 0,$$

where 0 is the function identically equal to zero, which, again, is not a probability measure. However, we may, accordingly, interpret this case as a "probability measure" concentrated at ∞, that is, if a random variable would follow 0, then $\mathrm{pr}(X \leq x) = 0$ for all $x \in \mathbb{R}$.

We see from this result that the parameter λ can be interpreted as a concentration parameter, because it moves the exp-G distribution to a degenerated distribution in a (if a is finite), when it varies from zero to infinity, and to a degenerated distribution in b (if b is finite) when it varies from 0 to minus infinity. Furthermore, if a equals minus infinity, the distribution moves towards the left side of the axis until the mass escapes entirely, when λ tends to infinity. Analogously, when b equals infinity, the distribution moves towards the right side of the axis until the mass escapes entirely, when λ tends to minus infinity.

3.4 Reciprocal property

This family of distributions enjoys a very interesting reciprocal property. We begin by introducing some notation, let $X^G \sim G$, and $1/X^G \sim S$, where G is continuous. Therefore, we have that if $X_\lambda^G \sim$ exp-G, then $1/X_\lambda^G \sim$ exp-S. To see this, observe that, for $\lambda \neq 0$,

$$\mathrm{pr}(1/X_\lambda^G \leq x) = \mathrm{pr}(X_\lambda^G \geq 1/x) = \frac{1 - \exp\{\lambda(1 - G(1/x))\}}{1 - \exp\{\lambda\}} = F_{-\lambda}^*(S(1/x)).$$

We also would like to remark that the reciprocal of X_λ^G has a corresponding exp-S distribution with $-\lambda$, that is, X_λ^G has c.d.f. $F_\lambda^G(x)$ and $1/X_\lambda^G$ has c.d.f. $F_{-\lambda}^S(x)$.

This means that whenever we study a special case of the exp-G distribution, we may easily study the reciprocal case.

3.5 Expansions, order statistics and moments

We now give an useful expansion for the p.d.f. (2.3). With this expansion, we can obtain mathematical properties such as ordinary moments, factorial moments and moment generating function of the exp-G distribution. Expanding the term $e^{-\lambda G(x;\theta)}$ in (2.3), it follows

$$f(x) = \frac{\lambda}{1-e^{-\lambda}} g(x;\theta) \sum_{j=0}^{\infty} \frac{(-\lambda)^j}{j!} G(x;\theta)^j. \tag{3.4}$$

Let now $X_1, \ldots, X_n$ be a random sample with p.d.f. in the form (2.3) and define $X_{i:n}$ the ith order statistic. The p.d.f. of the $X_{i:n}$, say, $f_{i:n}$, is given by

$$\begin{aligned} f_{i:n}(x) &= \frac{1}{B(i, n-i+1)} f(x) F(x)^{i-1} \{1 - F(x)\}^{n-i} \\ &= \frac{\lambda g(x;\theta) e^{-\lambda G(x;\theta)}}{B(i, n-i+1)(1-e^{-\lambda})^n} \left\{1 - e^{-\lambda G(x;\theta)}\right\}^{i-1} \left\{e^{-\lambda G(x;\theta)} - e^{-\lambda}\right\}^{n-i}. \end{aligned} \tag{3.5}$$

By using binomial expansion for the terms $\{1 - e^{-\lambda G(x;\theta)}\}^{i-1}$ and $\{e^{-\lambda G(x;\theta)} - e^{-\lambda}\}^{n-i}$ in (3.5), it follows

$$\begin{aligned} f_{i:n}(x) = \frac{(1-e^{-\lambda})^{-n}}{B(i, n-i+1)} \sum_{j=0}^{i-1} \sum_{k=0}^{n-i} \frac{(-1)^{n+j-k-i}}{j+k+1} \binom{i-1}{j} \binom{n-i}{k} e^{-\lambda(n-k-i)} \\ \times \left(1 - e^{-\lambda(j+k+1)}\right) f_{j,k}(x), \end{aligned} \tag{3.6}$$

where $f_{j,k}(\cdot)$ denotes the p.d.f. of a random variable with exp-$G(\lambda(j+k+1), \theta)$ distribution. Therefore, the p.d.f. of $X_{i:n}$ can be written as a linear combination of pdfs in the form (2.3) and, hence, the mathematical properties of the order statistics can be obtained from associated exp-G distribution.

We now give general expressions for the moments of the family exp-G of distributions. Consider X and Y as random variables with exp-$G(\lambda, \theta)$ and G distributions, respectively. An useful expression for the rth moment of the exp-G distributions follows from (3.4) and it is given in function of the probability weighted moments of the Y:

$$E(X^r) = \frac{\lambda}{1-e^{-\lambda}} \sum_{j=0}^{\infty} \frac{(-\lambda)^j}{j!} E\{Y^r G(Y;\theta)^j\}. \tag{3.7}$$

In particular, formula (3.7) provides us another proof of condition (v) in Proposition 3.1. Finally, with the result (3.6) the rth moment of the ith order statistic is given by

$$\begin{aligned} E(X_{i:n}^r) = \frac{(1-e^{-\lambda})^{-n}}{B(i, n-i+1)} \sum_{j=0}^{i-1} \sum_{k=0}^{n-i} \frac{(-1)^{n+j-k-i}}{j+k+1} \binom{i-1}{j} \binom{n-i}{k} e^{-\lambda(n-k-i)} \\ \times \left(1 - e^{-\lambda(j+k+1)}\right) E(Z_{j,k}^r), \end{aligned} \tag{3.8}$$

where $Z_{j,k}$ has exp-$G(\lambda(j+k+1),\theta)$ distribution.

4 Maximum likelihood estimation and inference

Let X be a random variable with exp-$G(\lambda,\theta)$ distribution, with $\lambda \neq 0$. The log-density of X with observed value x is given by

$$\ell = \ell(\lambda,\theta) = \log\{\lambda/(1-e^{-\lambda})\} + \log g(x;\theta) - \lambda G(x;\theta)$$

and the associated score function is $U = (\partial\ell/\partial\lambda, \partial\ell/\partial\theta)^\top$, where

$$\frac{\partial\ell}{\partial\lambda} = \frac{1}{\lambda} - \frac{1}{e^{\lambda}-1} - G(x;\theta), \qquad \frac{\partial\ell}{\partial\theta} = U^*(\theta) - \lambda\frac{\partial G(x;\theta)}{\partial\theta},$$

with $U^*(\theta)$ being the associated score function of the log-density of a random variable with p.d.f. $g(\cdot,\theta)$. From regularity conditions, we have $E\{G(X;\theta)\} = \lambda^{-1} - (e^{\lambda}-1)^{-1}$ and $E\{\partial G(X;\theta)/\partial\theta\} = \lambda^{-1}E\{U^*(\theta)\}$.

The information matrix $K = K((\lambda,\theta)^\top)$ is

$$K = \begin{pmatrix} \kappa_{\lambda,\lambda} & \kappa_{\lambda,\theta} \\ \kappa_{\theta,\lambda} & \kappa_{\theta,\theta} \end{pmatrix},$$

where

$$\kappa_{\lambda,\lambda} = \frac{1}{\lambda^2} - \frac{e^{\lambda}}{(e^{\lambda}-1)^2}, \qquad \kappa_{\theta,\theta} = \lambda E\left\{\frac{\partial^2 G(X;\theta)}{\partial\theta\,\partial\theta^\top}\right\} - E\left\{\frac{\partial U^*(\theta)}{\partial\theta}\right\},$$
$$\kappa_{\theta,\lambda} = \lambda^{-1}E\{U^*(\theta)\}.$$

For a random sample $x = (x_1,\ldots,x_n)$ of size n from X and $\Theta = (\lambda,\theta)^T$, the total log-likelihood is

$$\ell_n = \ell_n(\Theta) = \sum_{i=1}^{n} \ell^{(i)},$$

where $\ell^{(i)}$ is the log-likelihood for the ith observation $(i = 1,\ldots,n)$ as given before. The total score function is $U_n = U_n(\Theta) = \sum_{i=1}^{n} U^{(i)}$, where $U^{(i)}$ for $i = 1,\ldots,n$ has the form given earlier and the total information matrix is $K_n(\theta) = nK(\Theta)$.

The maximum likelihood estimator (MLE) $\hat{\Theta}$ of Θ is obtained numerically from the solution of the nonlinear system of equations $U_n = 0$. Under conditions that are fulfilled for the parameter Θ in the interior of the parameter space but not on the boundary, the asymptotic distribution of

$$\sqrt{n}(\hat{\Theta} - \Theta) \overset{A}{\sim} N_{k+1}(0, K(\Theta)^{-1}),$$

where '$\overset{A}{\sim}$' stands for the asymptotic distribution. The asymptotic multivariate normal $N_{k+1}(0, K_n(\hat{\Theta})^{-1})$ distribution of $\hat{\Theta}$ can be used to construct approximate confidence regions for some parameters and for the hazard and survival functions. In fact, an $100(1-\gamma)\%$ asymptotic confidence interval for each parameter Θ_i is given by

$$ACI_i = \left(\hat{\Theta}_i - z_{\gamma/2}\sqrt{\hat{\kappa}^{\Theta_i,\Theta_i}},\, \hat{\Theta}_i + z_{\gamma/2}\sqrt{\hat{\kappa}^{\Theta_i,\Theta_i}}\right),$$

where $\hat{\kappa}^{\Theta_i,\Theta_i}$ denotes the ith diagonal element of $K_n(\hat{\Theta})^{-1}$ for $i=1,\ldots,k+1$ and $z_{\gamma/2}$ is the quantile $1-\gamma/2$ of the standard normal distribution. The asymptotic normality is also useful for testing goodness of fit of the exp-G distribution and for comparing this distribution with some of its special submodels using one of the three well-known asymptotically equivalent test statistics—namely, the likelihood ratio (LR) statistic, Rao (S_R) and Wald (W) statistics.

4.1 Modified profile likelihood estimator

Since λ is a parameter added to some distribution, it can be seen as a nuisance parameter. With this in mind, we will advance a modified profile estimator for θ. From the last subsection, we have that

$$\frac{\partial \ell}{\partial \theta} = U^*(\theta) - \lambda \frac{\partial G(x;\theta)}{\partial \theta},$$

with

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = 0.$$

Therefore, if $\partial G(x;\theta)/\partial\theta^\top\, \partial G(x;\theta)/\partial\theta$, that belongs to $\mathbb{R}$, does not vanish for all values in some open neighbourhood of the true value of θ, let

$$\breve{\lambda} = \left\{\frac{\partial G(x;\theta)}{\partial \theta^\top}\frac{\partial G(x;\theta)}{\partial \theta}\right\}^{-1}\frac{\partial G(x;\theta)}{\partial \theta^\top}U^*(\theta),$$

with

$$\begin{aligned}\frac{\partial \breve{\lambda}}{\partial \theta} = &-2\left\{\frac{\partial G(x;\theta)}{\partial \theta^\top}\frac{\partial G(x;\theta)}{\partial \theta}\right\}^{-2}\left\{\frac{\partial^2 G(x;\theta)}{\partial \theta^\top \partial \theta^\top}\frac{\partial G(x;\theta)}{\partial \theta}\right\}\frac{\partial G(x;\theta)}{\partial \theta^\top}U^*(\theta)\\ &+\left\{\frac{\partial G(x;\theta)}{\partial \theta^\top}\frac{\partial G(x;\theta)}{\partial \theta}\right\}^{-1}\left\{\frac{\partial^2 G(x;\theta)}{\partial \theta^\top \partial \theta^\top}U^*(\theta) + \frac{\partial G(x;\theta)}{\partial \theta^\top}\frac{\partial U^*(\theta)}{\partial \theta}\right\},\end{aligned}$$

where $\partial^2 G(x;\theta)/\partial\theta^\top\partial\theta^\top$ stands for the row vector containing the diagonal elements of the Hessian matrix of G, $\partial^2 G(x;\theta)/\partial\theta\,\partial\theta^\top$, and $\partial U^*(\theta)/\partial\theta$ stands for the column vector $(\partial U_1^*(\theta)/\partial\theta_1,\ldots,\partial U_k^*(\theta)/\partial\theta_k)^\top$.

We, therefore, obtain the modified profile likelihood function:

$$\breve{\ell} = \breve{\ell}(\theta) = \log\breve{\lambda} - \log(1-e^{-\breve{\lambda}}) + \log g(x;\theta) - \breve{\lambda}G(x;\theta).$$

The modified profile estimator for θ can be obtained by maximizing $\breve{\ell}$. Let V be the estimating equation given by

$$V(\theta)=\frac{\partial\breve{\ell}}{\partial\theta}=\frac{1}{\breve{\lambda}}\frac{\partial\breve{\lambda}}{\partial\theta}-\frac{1}{e^{\breve{\lambda}}-1}\frac{\partial\breve{\lambda}}{\partial\theta}+U^*(\theta)-\frac{\partial\breve{\lambda}}{\partial\theta}G(x;\theta)-\breve{\lambda}\frac{\partial G(x;\theta)}{\partial\theta};$$

one may also obtain the profile likelihood estimator by solving the equation $V_n(\theta)=\sum_{i=1}^n V^{(i)}(\theta)=0$.

4.2 Interest case $\lambda = 0$

We now discuss estimation and inference when $\lambda = 0$. It is very important to discuss this case because we are interested in testing the hypotheses $H_0:\lambda=0$ versus $H_1:\lambda\neq 0$, that is, to test if the exp-G fit is significantly better than the G fit. The next result plays a important role in this paper.

Theorem 4.1. *Let $F_\lambda(\cdot)$ and $f_\lambda(\cdot)$ be the c.d.f. and p.d.f. defined by* (2.2) *and* (2.3), *respectively. The following conditions are true*:

(i) *If G is continuous, then $F_\lambda \to G$ uniformly when $\lambda\to 0$;*
(ii) *$f_\lambda \to g$ uniformly when $\lambda \to 0$, consequently, $\ell_n(\lambda,\theta)\to\ell_n^*(\theta)$, where $\ell_n^*(\theta)$ is the log-likelihood associated to G;*
(iii) *$\partial\ell/\partial\lambda \to 1/2 - G(x;\theta)$ and $\partial\ell/\partial\theta\to U^*(\theta)$, when $\lambda\to 0$;*
(iv) *$\kappa_{\lambda,\lambda}\to 1/12$, $\kappa_{\theta,\theta}\to\int\partial U^*(\theta)/\partial\theta\,\partial\theta^\top\,dG$ and $\kappa_{\lambda,\theta}\to -(1/2,1/2,1/2)^\top$, when $\lambda\to 0$, with $\int\partial U^*(\theta)/\partial\theta\,\partial\theta^\top\,dG$ being the information matrix with respect to G;*
(v) *If G is regular and $(\lambda_0,\theta_0)\in\Theta$, then*

$$\sqrt{n}\{(\hat{\lambda},\hat{\theta}^\top)^\top-(\lambda_0,\theta_0^\top)^\top\}\xrightarrow{d} N_{p+1}(0_{q+1},K(\lambda_0,\theta_0)^{-1});$$

(vi) *If G is regular and $(\lambda_0,\theta_0)\in\Theta$, then the likelihood ratio, Wald and Score statistics have null asymptotic distribution χ_q^2, where q is the number of parameters estimated in the alternative hypothesis minus the number of parameters estimated in the null hypothesis.*

5 The exp-Weibull distribution

We now move to the exp-G class of distributions. When G is the c.d.f. of the Weibull distribution, we will call this class of distributions by exp-Weibull. More precisely, to obtain the exp-Weibull distribution, we put in (2.2) the c.d.f. of the Weibull distribution $G(x)=1-\exp\{-(x/\beta)^\alpha\}$, where $\beta>0$, $\alpha>0$ and $x>0$. Therefore, the c.d.f. of the exp-Weibull distribution given by

$$F(x)=\frac{1-\exp\{-\lambda(1-e^{-(x/\beta)^\alpha})\}}{1-e^{-\lambda}},\qquad x>0.$$

From the general expressions (2.3) and (2.5) we obtain that the p.d.f. and hazard functions are given by

$$f(x) = \frac{\lambda\alpha\beta^{-\alpha}}{1-e^{-\lambda}} x^{\alpha-1} \exp\{-\lambda(1-e^{-(x/\beta)^{\alpha}}) - (x/\beta)^{\alpha}\}, \qquad x > 0, \tag{5.1}$$

and

$$h(x) = \frac{\lambda\alpha\beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^{\alpha}}}{1-\exp\{-\lambda e^{-(x/\beta)^{\alpha}}\}}, \qquad x > 0,$$

respectively.

We now illustrate the flexibility of this class of distributions by presenting some graphics of both the p.d.f. and hazard functions. Figure 1 shows the plots

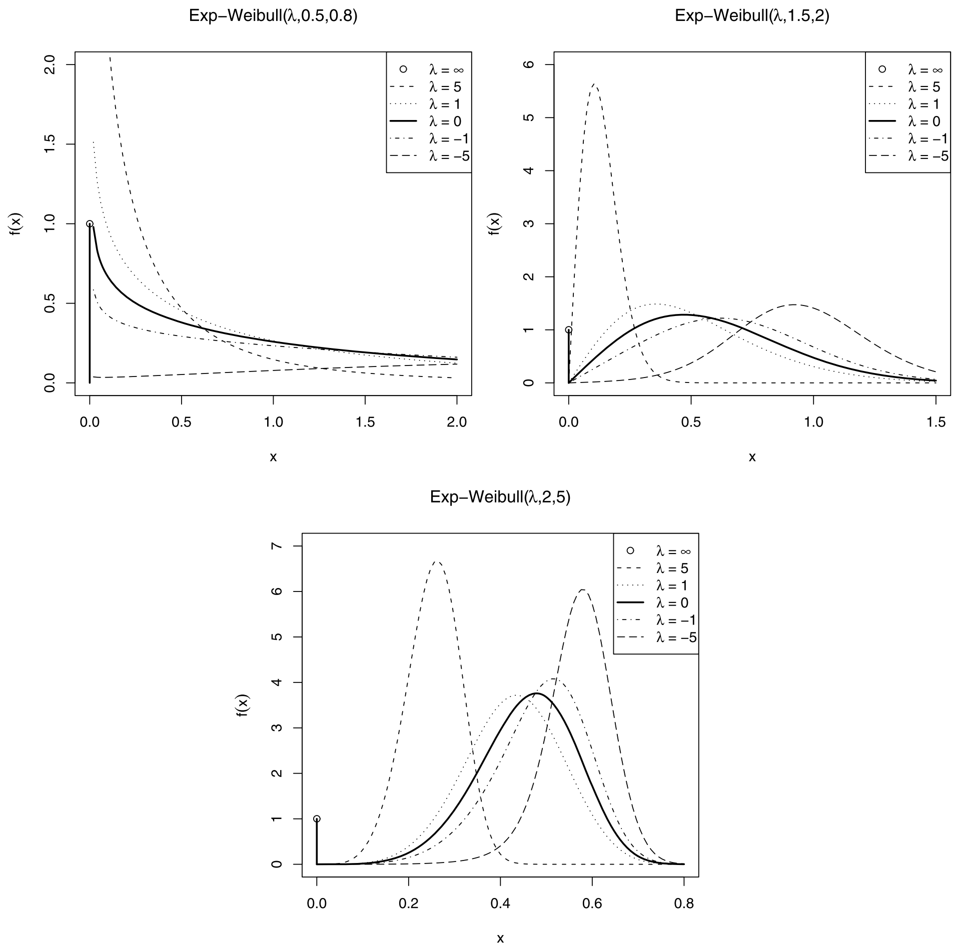

Figure 1 *Graphics of the p.d.f. of the exp-Weibull distribution for some values of the parameters.*

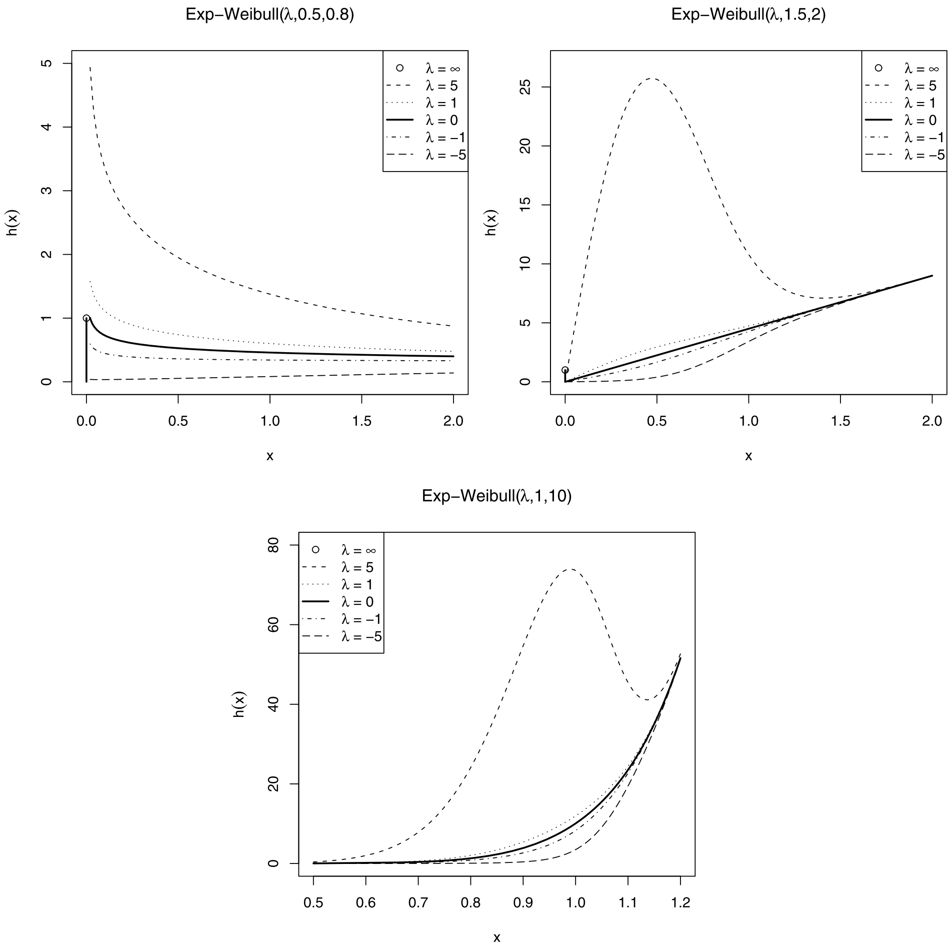

Figure 2 *Graphics of the hazard function of the exp-Weibull distribution for some values of the parameters.*

of the p.d.f. of the exp-Weibull distribution for some values of α and β, and for $\lambda = -5, -1, 0, 1, 5, \infty$. We note that when the value of λ increases the p.d.f. becomes more 'peaked.' Figure 2 contains the plots of the hazard function of the exp-Weibull distribution for different values of α and β and $\lambda = -5, -1, 0, 1, 5, \infty$. We note that the behaviour of the hazard function of the Weibull distribution is close to the behaviour of the graphics with $\lambda = 1.0$, and as the value of λ increases, the behaviour of the hazard function of the exp-Weibull becomes very different from the behaviour of the hazard function of the Weibull distribution, showing that as the value of λ gets larger the exp-Weibull "moves away" from the Weibull distribution, and gets closer to the Dirac mass at zero, as remarked on the end of the last section.

Order statistics and moments

The p.d.f. of the ith order statistic of a random sample from the exp-$W(\lambda, \beta, \alpha)$ distribution is given by

$$f_{i:n}(x) = \frac{\lambda\alpha\beta^{-\alpha}x^{\alpha-1}\exp\{-\lambda(1-e^{-(x/\beta)^{\alpha}}) - (x/\beta)^{\alpha}\}}{B(i, n-i+1)(1-e^{-\lambda})^{n}} \times \{1-e^{-\lambda(1-e^{-(x/\beta)^{\alpha}})}\}^{i-1}\{e^{-\lambda(1-e^{-(x/\beta)^{\alpha}})} - e^{-\lambda}\}^{n-i}, \qquad x > 0.$$

We will now obtain series representation for the moments of the exp-Weibull distribution and of the order statistics. To this end, let X be a random variable following a exp-Weibull distribution with parameters $\beta > 0$, $\alpha > 0$ and $\lambda > 0$. From now on we will use the notation $X \sim$ exp-Weibull(λ, β, α) to indicate this fact.

We have that the probability weighted moment of a random variable Y following a Weibull distribution with parameter vector $\theta = (\beta, \alpha)^{\top}$ can be written as $E\{Y^{r}G(Y;\theta)^{j}\} = \beta^{r}\int_{0}^{1}(-\log u)^{r/\alpha}(1-u)^{j}\,du$. Therefore, from (3.7) it follows that the rth moment of X is

$$E(X^{r}) = \frac{\lambda\beta^{r}}{1-e^{-\lambda}}\sum_{k=0}^{\infty}\frac{(-\lambda)^{k}}{k!}\int_{0}^{1}(1-u)^{k}(-\log u)^{r/\alpha}\,du. \tag{5.2}$$

We now give an alternative expression to (5.2) more simply. The rth moment of X is

$$\begin{aligned} E(X^{r}) &= \int_{0}^{\infty} x^{r} f(x)\,dx \\ &= \int_{0}^{\infty}\frac{\lambda\alpha\beta^{-\alpha}}{1-e^{-\lambda}}x^{r+\alpha-1}\exp\{-\lambda(1-e^{-(x/\beta)^{\alpha}}) - (x/\beta)^{\alpha}\}\,dx. \end{aligned}$$

Now, expanding $\exp\{\lambda e^{-(x/\beta)^{\alpha}}\}$ in Taylor's series, we get

$$\begin{aligned} E(X^{r}) &= \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\sum_{k=0}^{\infty}\frac{\lambda^{k}}{k!}\int_{0}^{\infty} x^{r}\alpha\beta^{-\alpha}x^{\alpha-1}e^{-\{(k+1)^{1/\alpha}x/\beta\}^{\alpha}}\,dx \\ &= \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\sum_{k=0}^{\infty}\frac{\lambda^{k}E(Y_{k}^{r})}{(k+1)!}, \end{aligned}$$

where Y_k follows the Weibull distribution with parameters $(k+1)^{1/\alpha}/\beta$ and α, and the interchange between the series and integral being possible due to Fubini's theorem together with the fact that we are dealing with the positive integrand. Hence, we have that the rth moment of a exp-Weibull distribution can be written as

$$E(X^{r}) = \lambda\beta^{r}\frac{\Gamma(r/\alpha+1)}{e^{\lambda}-1}\sum_{k=0}^{\infty}\frac{\lambda^{k}}{k!(k+1)^{r/\alpha+1}}. \tag{5.3}$$

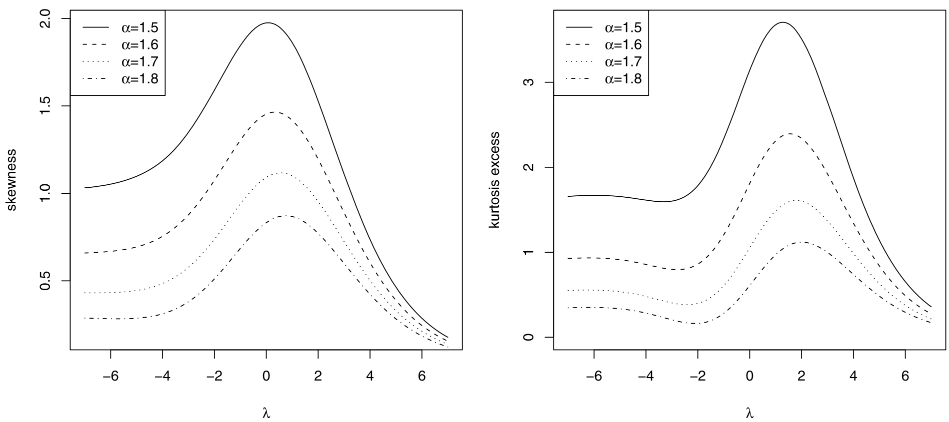

Figure 3 *Skewness and kurtosis of the exp-Weibull distribution for some values of the parameters.*

Figure 3 shows skewness and kurtosis of the exp-Weibull distribution, obtained from application of the formula of the moments above, for $\beta = 0.5$ and some values of α as a function of λ. We now note from (5.3) that all moments of the exp-Weibull distribution tend to zero as λ increases to infinity, which is a very remarkable fact. So, as we can note from Figure 3, as λ increases, the skewness tends to zero, as well as the kurtosis, one more time reflecting the expected behaviour of the limiting distribution as $\lambda \to \infty$.

An expression for the rth moment of the ith order statistic of the exp-Weibull distribution, say, $X_{i:n}$, follows from (3.8) and (5.3):

$$E(X_{i:n}^r) = \frac{\lambda\beta^r\Gamma(r/\alpha+1)}{B(i,n-i+1)(1-e^{-\lambda})^n} \times \sum_{l=0}^{\infty}\sum_{j=0}^{i-1}\sum_{k=0}^{n-i}(-1)^{n+j-k-i}\binom{i-1}{j}\binom{n-i}{k} \times e^{-\lambda(n+j-i+1)}\frac{\{\lambda(j+k+1)\}^l}{l!(l+1)^{r/\alpha+1}}. \tag{5.4}$$

Expressions (5.2) and (5.4) show the importance of the expansions given in Section 3.5. Furthermore, result (5.3) shows that alternative expressions to (3.7) can be obtained depending on the G distribution.

Order statistics and moments of the exp-Fréchet distribution

Here, we use the reciprocal property of the exp-G distributions to obtain expressions for the moments and order statistics of the exp-Fréchet distribution.

Let $Y \sim$ exp-Fr(λ, β, α) and $Y_{i:n}$ be the ith order statistics from a random sample, of size n, of the exp-Fréchet distribution. From formulae (5.3) and (5.4), we have that the moments of Y and $Y_{i:n}$ are

$$E(Y^r) = -\lambda \frac{\Gamma(1-r/\alpha)}{\beta^r(e^{-\lambda}-1)} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!(k+1)^{1-r/\alpha}}$$

and

$$E(X_{i:n}^r) = \frac{-\lambda\Gamma(1-r/\alpha)}{\beta^r B(i,n-i+1)(1-e^{\lambda})^n} \times \sum_{l=0}^{\infty}\sum_{j=0}^{i-1}\sum_{k=0}^{n-i}(-1)^{n+j-k-i}\binom{i-1}{j}\binom{n-i}{k} \times e^{\lambda(n+j-i+1)}\frac{\{-\lambda(j+k+1)\}^l}{l!(l+1)^{1-r/\alpha}},$$

respectively, for $r < \alpha$.

Score function and information matrix

Let $\theta = (\lambda, \beta, \alpha)^T$ be the parameter vector and X random variable with exp-Weibull(λ, β, α) distribution. The log-density $\ell = \ell(\theta)$ for the random variable X with observed value x is given by

$$\ell = -\alpha \log\beta + \log(\alpha\lambda) - \left(\frac{x}{\beta}\right)^{\alpha} - \lambda\{1 - e^{-(x/\beta)^{\alpha}}\} - \log(1-e^{-\lambda}) + (\alpha-1)\log x, \qquad x > 0.$$

The score function is given by

$$\frac{\partial l}{\partial\lambda} = -1 + e^{-(x/\beta)^{\alpha}} + \frac{1}{1-e^{\lambda}} + \frac{1}{\lambda},$$

$$\frac{\partial l}{\partial\beta} = \alpha\beta^{-1}\left[-1 + \left(\frac{x}{\beta}\right)^{\alpha}\{1+\lambda e^{-(x/\beta)^{\alpha}}\}\right],$$

$$\frac{\partial l}{\partial\alpha} = \frac{1}{\alpha} + \log(x) - \log(\beta) - \left(\frac{x}{\beta}\right)^{\alpha}\log\left(\frac{x}{\beta}\right)\{1+\lambda e^{-(x/\beta)^{\alpha}}\}.$$

From the regularity conditions one obtains the following closed-form expressions:

$$E[e^{-(X/\beta)^{\alpha}}] = 1 - \frac{1}{1-e^{\lambda}} - \frac{1}{\lambda},$$

$$E\left[\left(\frac{X}{\beta}\right)^{\alpha}\{1+\lambda e^{-(X/\beta)^{\alpha}}\}\right] = 1$$

and

$$E\left[\left(\frac{X}{\beta}\right)^{\alpha}\log\left(\frac{X}{\beta}\right)\{1+\lambda e^{-(X/\beta)^{\alpha}}\}\right] = \frac{1}{\alpha} - \log(\beta) + E\{\log(X)\}.$$

For interval estimation and hypothesis tests on the model parameters, we require the information matrix. We will, therefore, use some of the expressions above to obtain the Fisher's information matrix. The 3×3 unit information matrix $K = K((\lambda, \beta, \alpha)^T)$ is

$$K = \begin{pmatrix} \kappa_{\lambda,\lambda} & \kappa_{\lambda,\beta} & \kappa_{\lambda,\alpha} \\ \kappa_{\lambda,\beta} & \kappa_{\beta,\beta} & \kappa_{\beta,\alpha} \\ \kappa_{\lambda,\alpha} & \kappa_{\beta,\alpha} & \kappa_{\alpha,\alpha} \end{pmatrix},$$

whose elements are

$$\kappa_{\lambda,\lambda} = \lambda^{-2} - \frac{e^{\lambda}}{(-1+e^{\lambda})^2}, \qquad \kappa_{\lambda,\beta} = \frac{\alpha}{\beta\lambda}\left[E\left\{\left(\frac{X}{\beta}\right)^{\alpha}\right\} - 1\right],$$

$$\kappa_{\lambda,\alpha} = \frac{1}{\lambda}\left[E\left\{\left(\frac{X}{\beta}\right)^{\alpha}\log\left(\frac{X}{\beta}\right)\right\} - \frac{1}{\alpha} + \log(\beta) - E\{\log(X)\}\right],$$

$$\kappa_{\beta,\beta} = \frac{\alpha}{\beta^3}\left[1 - \alpha\lambda E\left\{e^{-(X/\beta)^{\alpha}}\left(\frac{X}{\beta}\right)^{2\alpha}\right\}\right],$$

$$\kappa_{\beta,\alpha} = \alpha\left[\lambda E\left\{e^{-(X/\beta)^{\alpha}}\left(\frac{X}{\beta}\right)^{2\alpha}\log\left(\frac{X}{\beta}\right)\right\} - \frac{1}{\alpha} + \log(\beta) - E\{\log(X)\}\right]$$

and

$$\kappa_{\alpha,\alpha} = \frac{1}{\alpha^2} + E\left[\left(\frac{X}{\beta}\right)^{\alpha}\log\left(\frac{X}{\beta}\right)^2\left\{1 - \lambda e^{-(X/\beta)^{\alpha}}\left(\frac{X}{\beta}\right)^{\alpha} + \lambda e^{-(X/\beta)^{\alpha}}\right\}\right].$$

These elements of the information matrix depend on some expectations that can be easily obtained through numerical integration.

6 The exp-beta distribution

Let Y be a random variable following standard beta distribution with parameters $a > 0$ and $b > 0$. The c.d.f. of Y is given by $G(x; (a, b)^{\top}) = I_x(a, b)$, where $I_x(a, b) = B(a, b)^{-1}\int_0^x t^{a-1}(1-t)^{b-1}\,dt$ denotes the incomplete beta function and $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ is the beta function. The exp-beta distribution is introduced by taking G as being the c.d.f. of Y in (2.2). We will denote a random variable X with exp-beta distribution by $X \sim$ exp-beta(λ, a, b).

The p.d.f. and c.d.f. of the exp-beta distribution are given by

$$f(x) = \frac{\lambda}{B(a,b)(1-e^{-\lambda})}x^{a-1}(1-x)^{b-1}e^{-\lambda I_x(a,b)}, \qquad x \in (0,1),$$

and

$$F(x) = \frac{1-e^{-\lambda I_x(a,b)}}{1-e^{-\lambda}}, \qquad x \in (0,1),$$

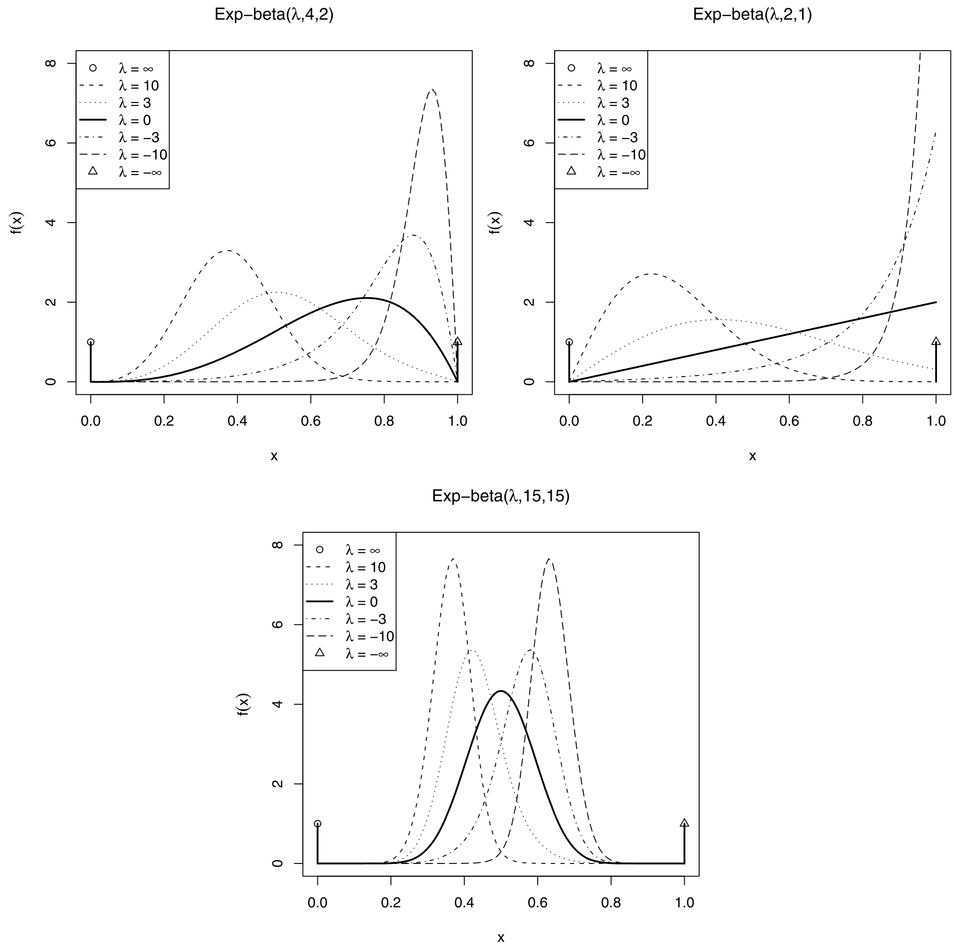

Figure 4 *Graphics of the p.d.f. of the exp-beta distribution for some values of the parameters.*

respectively.

Figure 4 shows the plots of the p.d.f. of the exp-beta distribution for some values of a and b, and for $\lambda = -\infty, -10, -3, 0, 3, 10, \infty$. Observe that for the exp-beta$(\lambda, 2, 1)$ distribution, the density of the beta(2, 1) distribution is very close to a straight line, whereas the densities of the exp-beta$(\lambda, 2, 1)$ distributions may assume various shapes, such as unimodal, and strictly increasing.

Figure 5 shows skewness and kurtosis of the exp-beta distribution for $a = 2$ and some values of b as a function of λ.

Score function and information matrix

Let X be a random variable with exp-beta(λ, a, b) distribution and $\theta = (\lambda, \beta, \alpha)^T$ be the parameter vector, with $\lambda \neq 0$. The log-density $\ell = \ell(\theta)$ for the random

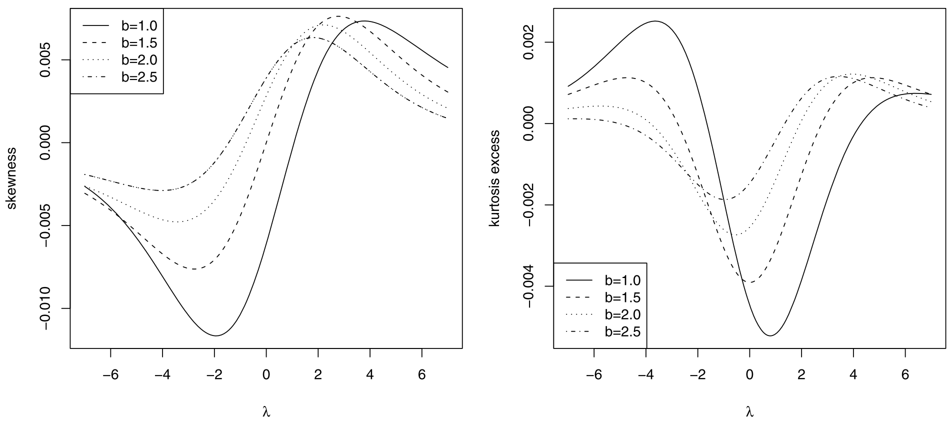

Figure 5 *Skewness and kurtosis of the exp-beta distribution for some values of the parameters.*

variable X with observed value x is given by

$$\ell = \log\lambda - \log B(a,b) - \log(1-e^{-\lambda}) + (a-1)\log x + (b-1)\log(1-x) - \lambda I_x(a,b)$$

for $x \in (0,1)$.

The score function is given by

$$\frac{\partial l}{\partial\lambda} = \frac{1}{\lambda} - \frac{1}{e^{\lambda}-1} - I_x(a,b),$$
$$\frac{\partial l}{\partial a} = -\Psi(a) + \Psi(a+b) + \log x - \lambda\frac{\partial I_x(a,b)}{\partial a},$$
$$\frac{\partial l}{\partial b} = -\Psi(b) + \Psi(a+b) + \log(1-x) - \lambda\frac{\partial I_x(a,b)}{\partial b},$$

where $\Psi(y) = d\log\Gamma(y)/dy$.

Under the usual regularity conditions, the expected value of the score function vanishes. Hence, we obtain

$$E\{I_X(a,b)\} = \frac{1}{\lambda} - \frac{1}{e^{\lambda}-1},$$
$$E\left\{\frac{\partial I_X(a,b)}{\partial a}\right\} = \lambda^{-1}\{\Psi(a+b) - \Psi(a) + E(\log X)\}$$

and

$$E\left\{\frac{\partial I_X(a,b)}{\partial b}\right\} = \lambda^{-1}[\Psi(a+b) - \Psi(b) + E\{\log(1-X)\}].$$

The Fisher's information matrix $K = K((\lambda, a, b)^T)$ is

$$K = \begin{pmatrix} \kappa_{\lambda,\lambda} & \kappa_{\lambda,a} & \kappa_{\lambda,b} \\ \kappa_{\lambda,a} & \kappa_{a,a} & \kappa_{a,b} \\ \kappa_{\lambda,b} & \kappa_{a,b} & \kappa_{b,b} \end{pmatrix},$$

whose elements are

$$\kappa_{\lambda,\lambda} = \lambda^{-2} - \frac{e^{\lambda}}{(-1+e^{\lambda})^2}, \qquad \kappa_{\lambda,a} = \lambda^{-1}\{\Psi(a+b) - \Psi(a) + E(\log X)\},$$

$$\kappa_{\lambda,b} = \lambda^{-1}[\Psi(a+b) - \Psi(b) + E\{\log(1-X)\}],$$

$$\kappa_{a,a} = \Psi'(a) - \Psi'(a+b) + \lambda E\left\{\frac{\partial^2 I_X(a,b)}{\partial a^2}\right\},$$

$$\kappa_{a,b} = \Psi'(b) - \Psi'(a+b) + \lambda E\left\{\frac{\partial^2 I_X(a,b)}{\partial b^2}\right\}$$

and

$$\kappa_{b,b} = -\Psi'(a+b) + \lambda E\left\{\frac{\partial^2 I_X(a,b)}{\partial a\,\partial b}\right\}.$$

These elements of the information matrix depend on some expectations that can be easily obtained through numerical integration.

7 Application

Our aim in this section is to motivate the use of the class exp-G of distributions by showing a successful application to one real data set. We here will fit exp-Weibull distribution to the data set given by Birnbaum and Saunders (1969) on the fatigue life of 6061-T6 aluminium coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set consists of 101 observations with maximum stress per cycle 31,000 psi. The data are as follows: 70, 90, 96, 97, 99, 100, 103, 104, 104, 105, 107, 108, 108, 108, 109, 109, 112, 112, 113, 114, 114, 114, 116, 119, 120, 120, 120, 121, 121, 123, 124, 124, 124, 124, 124, 128, 128, 129, 129, 130, 130, 130, 131, 131, 131, 131, 131, 132, 132, 132, 133, 134, 134, 134, 134, 136, 136, 137, 138, 138, 138, 139, 139, 141, 141, 142, 142, 142, 142, 142, 142, 144, 144, 145, 146, 148, 148, 149, 151, 151, 152, 155, 156, 157, 157, 157, 157, 158, 159, 162, 163, 163, 164, 166, 166, 168, 170, 174, 201, 212.

The MLEs and the maximized log-likelihood determined by fitting the exp-Weibull and Weibull distributions are

$$\widehat{\beta} = 55.670932, \qquad \widehat{\lambda} = -41.645738,$$

$$\widehat{\alpha} = 1.642486, \qquad \widehat{\ell}_{\text{exp-Weibull}} = -454.3272$$

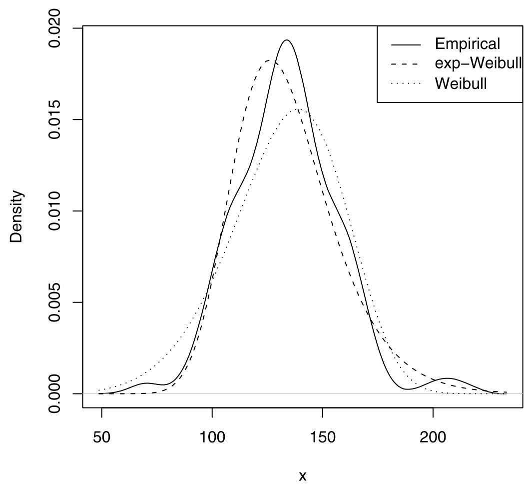

Figure 6 *Empirical density and fitted exp-Weibull and Weibull densities for the Birnbaum and Saunders's* (1969) *data set.*

and

$$\widehat{\beta} = 143.3150, \qquad \widehat{\alpha} = 5.9790, \qquad \widehat{\ell}_{\text{Weibull}} = -459.0999,$$

respectively.

We test the null hypothesis H_0 : Weibull model against the alternative hypothesis H_1 : exp-Weibullmodel. The LR statistic is 9.5453 and the p-value $= 2 \times 10^{-3}$. Hence, for any usual significance level, we reject the null model (Weibull) in favour of the alternative exp-Weibull model. In Figure 6 are displayed the empirical density, fitted exp-Weibull and Weibull densities. Hence, we see that the exp-Weibull distribution yields a better fit than the Weibull distribution.

8 Conclusion

We defined a family of distributions that provides a rather general and flexible framework for statistical analysis. It also provides a rather flexible mechanism for fitting a wide spectrum of real world data sets. Several properties of this class of distributions were obtained, such as the Kullback–Leibler divergence between G and exp-G distributions, characterization based on Shannon entropy, moments, order statistics, estimation of the parameters and inference.

With this, we moved to two special distributions, the exp-Weibull and exp-beta distributions, which were studied with some details. The article was motivated by a successful application to fatigue life data.

Appendix

In this appendix we prove Propositions 3.1 and 3.3, and Theorem 4.1.

Proof of Proposition 3.1

The case $\lambda = 0$ will not be treated since $X_0 = X$, and thus the result is trivial.

(i) and (ii) follow directly from equation (2.2).

For (iii) suppose that $\lambda > 0$, the calculation being analogous for $\lambda < 0$. Then,

$$\begin{aligned}\frac{\exp\{-\lambda G(x-\varepsilon)\} - \exp\{-\lambda G(x)\}}{(1-e^{-\lambda})\{G(x) - G(x-\varepsilon)\}} &= \frac{1}{(1-e^{-\lambda})\{G(x) - G(x-\varepsilon)\}} \int_{G(x-\varepsilon)}^{G(x)} \lambda e^{-\lambda x}\, dx \\ &\leq \frac{\lambda}{1-e^{-\lambda}} \exp\{-\lambda G(x-\varepsilon)\}.\end{aligned} \tag{A.1}$$

A similar computation shows that

$$\frac{\exp\{-\lambda G(x-\varepsilon)\} - \exp\{-\lambda G(x)\}}{(1-e^{-\lambda})\{G(x) - G(x-\varepsilon)\}} \geq \frac{\lambda}{1-e^{-\lambda}} \exp\{-\lambda G(x)\}.$$

Therefore, if x is a continuous point of G, we have that

$$\begin{aligned}\frac{1}{1-e^{-\lambda}} \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \frac{\exp\{-\lambda G(x-\varepsilon)\} - \exp\{-\lambda G(x)\}}{G(x) - G(x-\varepsilon)} &\\ = \frac{\lambda}{1-e^{-\lambda}} \exp\{-\lambda G(x)\}.&\end{aligned}$$

Nevertheless, if x is a discontinuity point of G, we have that

$$\begin{aligned}\frac{1}{1-e^{-\lambda}} \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \frac{\exp\{-\lambda G(x-\varepsilon)\} - \exp\{-\lambda G(x)\}}{G(x) - G(x-\varepsilon)} &\\ = \frac{\exp\{-\lambda G(x-)\} - \exp\{-\lambda G(x)\}}{(1-e^{-\lambda})\{G(x) - G(x-)\}}.&\end{aligned}$$

Suppose that G is discontinuous at the points $\{x_1, \ldots\}$, and let G_c be the part of G with continuity points (the sum of the continuous and singular parts of G), then it is easy to observe that

$$F_\lambda^G(x) = F_\lambda(G_c(x)) + \sum_{i=1}^{\infty} 1_{\{x_i \leq x\}} \frac{\exp\{-\lambda G(x_{i-1})\} - \exp\{-\lambda G(x_i)\}}{1-e^{-\lambda}},$$

where $1_A(x)$ is the indicator function of the set A. We now have that

$$\begin{aligned}
&\int_{(-\infty,x]} \frac{1}{1-e^{-\lambda}} \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \frac{\exp\{-\lambda G(x-\varepsilon)\} - \exp\{-\lambda G(x)\}}{G(x) - G(x-\varepsilon)}\, dG(x)\\
&\quad = \int_{(-\infty,x]} \frac{\lambda}{1-e^{-\lambda}} \exp\{-\lambda G_c(x)\}\, dG_c(x)\\
&\qquad + \sum_{i=1}^{\infty} 1_{\{x_i \le x\}} \frac{\exp\{-\lambda G(x_{i-1})\} - \exp\{-\lambda G(x_i)\}}{1-e^{-\lambda}}\\
&\quad = \frac{1-\exp\{-\lambda G_c(x)\}}{1-e^{-\lambda}} + \sum_{i=1}^{\infty} 1_{\{x_i \le x\}} \frac{\exp\{-\lambda G(x_{i-1})\} - \exp\{-\lambda G(x_i)\}}{1-e^{-\lambda}},
\end{aligned}$$

which concludes the proof of (iii). The proof of (iv) is a simple application of (iii), and to prove (v), one uses (iii) and the inequality in equation (A.1), for $\lambda > 0$, and a similar inequality for $\lambda < 0$.

Proof of Proposition 3.3

Let $z(\cdot)$ be a p.d.f. which satisfies the constraints (C1) and (C2). The Kullback–Leibler divergence between z and f is

$$D_{\mathrm{KL}}(z \parallel f) = \int_{\mathbb{R}} z \log\left(\frac{z}{f}\right) dx.$$

With this, we follow Cover and Thomas (1991) and obtain

$$\begin{aligned}
0 \le D_{\mathrm{KL}}(z \parallel f) &= \int_{\mathbb{R}} z \log z\, dx - \int_{\mathbb{R}} z \log f\, dx\\
&= -\mathbb{H}_S(z) - \int_{\mathbb{R}} z \log f\, dx.
\end{aligned}$$

With the definition of f and based on the constraints (C1) and (C2), it is easy to see that

$$\begin{aligned}
\int_{\mathbb{R}} z \log f\, dx &= -1 + \frac{\lambda}{e^{\lambda}-1} + \log\left(\frac{\lambda}{1-e^{-\lambda}}\right) + E\{\log g(G^{-1}(U;\theta))\}\\
&= \int_{\mathbb{R}} f \log f\, dx = -\mathbb{H}_S(f),
\end{aligned}$$

where U is defined as before. With this, we have

$$\mathbb{H}_S(z) \le \mathbb{H}_S(f),$$

with equality if and only if $z(x) = f(x)$ Lebesgue—almost everywhere, thus proving the uniqueness.

Proof of Theorem 4.1

(i) It is well known from real analysis that if f_n is a sequence of bounded and right-continuous functions that converge in all points for a continuous function, then this convergence is uniform. Therefore, since F_λ satisfies the above conditions and converges for G, which is a continuous function, the proof of (i) is completed.

(ii)–(iv) The proofs are easily checked.

(v) For $(\lambda_0, \theta_0) \in \Theta$, with $\lambda_0 \in \mathbb{R} \setminus \{0\}$, the result follows from Proposition 3.2 and Theorem 6.5.1 from Lehmann and Casella (2003). For $\lambda_0 = 0$, we have to use the results in (ii)–(iv) to adapt the proof given in Lehmann and Casella (2003) to our case. We begin by showing that

$$\sqrt{n}\hat{\lambda} \xrightarrow{d} N(0, 12),$$

when $n \to \infty$. The following lemma will be useful in this proof.

Lemma A.1. *Let $f : [a, b) \to \mathbb{R}$ (b can be ∞). If f admits n derivatives to right around the point a, then*

$$f(x) = f(a) + f^{(1)}(a)(x - a) + \cdots + \frac{f^{(n)}(a)}{n!}(\tilde{x} - a)^n,$$

where $f^{(k)}$ represents the kth derivative of f to the right and $\tilde{x} \in (a, x)$.

The proof in the above lemma is similar to the usual proof, but replacing the derivative by the derivative from the right. It is clear that the same result holds for the derivative from the left.

Applying the lemma to the log-likelihood with respect to p.d.f. $f(x; \lambda) = \lambda e^{-\lambda x}/(1 - e^{-\lambda})$, that is, $G(x; \theta) = x$, it follows that

$$\begin{aligned}\ell'(\hat{\lambda}) &= \ell'(0) + \hat{\lambda}\ell''(0) + \frac{1}{2}\hat{\lambda}^2\ell'''(\tilde{\lambda}) \\ &= \frac{n}{2} - \sum_{j=1}^{n} x_j - \frac{n}{12}\hat{\lambda} + \frac{1}{2}\hat{\lambda}^2\ell'''(\tilde{\lambda}),\end{aligned}$$

where $\ell'''(\tilde{\lambda}) = -\frac{1}{6}\tilde{\lambda}^{-3} - d^3(e^{\tilde{\lambda}} - 1)^{-1}/d\tilde{\lambda}$.

By supposition, $\ell'(\hat{\lambda}) = 0$, thus,

$$\sqrt{n}\hat{\lambda} = n^{-1/2}\frac{n/2 - \sum_{j=1}^{n} x_j}{12^{-1} - (2n)^{-1}\hat{\lambda}\ell'''(\tilde{\lambda})}.$$

As $n^{-1}\hat{\lambda}\ell'''(\tilde{\lambda}) \to 0$ in probability and $n^{-1/2}(n/2 - \sum_{j=1}^{n} x_j) \xrightarrow{d} N(0, 1)$, when $n \to \infty$, then $\sqrt{n}\hat{\lambda}$ converges in distribution to a normal distribution with mean 0 and variance 12.

The rest of the proof is analogous to the one given in Lehmann, where one may use the results in (ii)–(iv) to ensure that all the arguments hold true.

(vi) It follows from asymptotic normality of $\hat{\Theta}$; see Lehmann and Romano (2008) for more details.

Acknowledgments

We gratefully acknowledge financial support from CNPq (Brazil). We also thank two anonymous referees for their comments.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Barreto-Souza, W., Simas, A. B. and Santos, A. H. S. (2008). A new family of distributions based on the trucanted exponential distribution. Article presented as oral communication in Sinape—Simpósio Nacional de Probabilidade e Estatística, 2008, Estância de São Pedro, SP, Brazil.

Billingsley, P. (1995). *Probability and Measure*. New York: Wiley. MR1324786

Birnbaum, Z. W. and Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability* **6**, 328–347. MR0251807

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley. MR1122806

Ferreira, J. T. A. S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association* **101**, 823–829. MR2256188

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Reviews* **106**, 620–630. MR0087305

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics (with discussion). *Test* **13**, 1–43. MR2065642

Jones, M. C. and Faddy, M. J. (2003). A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society, Ser. B* **65**, 159–174. MR1959820

Jones, M. C. and Pewsey, A. (2009). Sinh–arcsinh distributions. *Biometrika* **96**, 761–780. MR2564489

Lehmann, E. L. and Casella, G. (2003). *Theory of Point Estimation*. New York: Springer.

Lehmann, E. L. and Romano, J. P. (2008). *Testing Statistical Hypotheses*. New York: Springer.

Ma, Y. and Genton, M. G. (2004). A flexible class of skew-symmetric distributions. *Scandinavian Journal of Statistics* **31**, 459–468. MR2087837

Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika* **84**, 641–652. MR1603936

Morais, A. L. and Barreto-Souza, W. (2011). A compound class of Weibull and power series distributions. *Computational Statistics and Data Analysis* **55**, 1410–1425.

Mudholkar, G. S., Srivastava, D. K. and Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association* **91**, 1575–1583. MR1439097

Nadarajah, S., Nassiri, V. and Mohammadpour, A. (2009). Truncated-exponential skew-symmetric distributions, Technical report, School of Mathematics, Univ. Manchester.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–432. MR0026286

Departamento de Estatística
Universidade de São Paulo
Rua do Matão, 1010
São Paulo/SP, 05508-090
Brazil
E-mail: wagnerbs85@gmail.com

Departamento de Matemática
Universidade Federal da Paraíba
João Pessoa/PB, 58051-900
Brazil
E-mail: alexandre@mat.ufpb.br