Journal of Statistical Computation & Simulation
Vol. 00, No. 00, August 2009, 1–17

A new family of generalized distributions

Gauss M. Cordeiro[a] and Mário de Castro[b]*

[a] *Universidade Federal Rural de Pernambuco, Departamento de Estatística e Informática, 52171-900, Recife-PE, Brasil*; [b] *Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, Caixa Postal 668, 13560-970, São Carlos-SP, Brasil*

(*Received December 7, 2009*)

Kumaraswamy [1] introduced a distribution for double bounded random processes with hydrological applications. For the first time, based on this distribution, we describe a new family of generalized distributions (denoted with the prefix "Kw") to extend the normal, Weibull, gamma, Gumbel, inverse Gaussian distributions, among several well-known distributions. Some special distributions in the new family such as the Kw-normal, Kw-Weibull, Kw-gamma, Kw-Gumbel and Kw-inverse Gaussian distribution are discussed. We express the ordinary moments of any Kw generalized distribution as linear functions of probability weighted moments of the parent distribution. We also obtain the ordinary moments of order statistics as functions of probability weighted moments of the baseline distribution. We use the method of maximum likelihood to fit the distributions in the new class and illustrate the potentiality of the new model with an application to real data.

Keywords: gamma distribution; Kumaraswamy distribution; moments; normal distribution; order statistics; Weibull distribution

AMS Subject Classification: 62E10; 62F03; 62F05; 62F10

1. Introduction

Beta distributions are very versatile and a variety of uncertainties can be usefully modeled by them. Many of the finite range distributions encountered in practice can be easily transformed into the standard beta distribution. In econometrics, many times the data are modeled by finite range distributions. Generalized beta distributions have been widely studied in statistics and numerous authors have developed various classes of these distributions. Eugene *et al.* [2] proposed a general class of distributions for a random variable defined from the logit of the beta random variable by employing two parameters whose role is to introduce skewness and to vary tail weight. Following the work of Eugene *et al.* [2], who defined the beta normal distribution, Nadarajah and Kotz [3] introduced the beta Gumbel distribution, Nadarajah and Gupta [4] proposed the beta Fréchet distribution and Nadarajah and Kotz [5] worked with the beta exponential distribution. However, all these works lead to some mathematical difficulties because the beta distribution is not fairly tractable and, in particular, its cumulative distribution function (cdf) involves the incomplete beta function ratio.

The paper by Kumaraswamy [1] proposed a new probability distribution for double bounded random processes with hydrological applications. The Kumaraswamy's distribution appears to have received considerable interest in hydrology and related areas, see [6–9].

*Corresponding author. Email: mcastro@icmc.usp.br

ISSN: 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/0094965YYxxxxxxxx
http://www.informaworld.com

In reliability and life testing experiments, many times the data are modeled by finite range distributions. See, for example, [10]. We start with the Kumaraswamy's distribution (called from now on the Kw distribution) on the interval $(0, 1)$, having the probability density function (pdf) and the cdf with two shape parameters $a > 0$ and $b > 0$ defined by

$$f(x) = a\,b\,x^{a-1}(1 - x^a)^{b-1} \quad \text{and} \quad F(x) = 1 - (1 - x^a)^b. \tag{1}$$

The density function in (1) has many of the same properties as the beta distribution but has some advantages in terms of tractability.

The Kw distribution does not seem to be very familiar to statisticians and has not been investigated systematically in much detail before, nor has its relative interchangeability with the beta distribution has been widely appreciated. However, in a very recent paper, Jones [11] explored the background and genesis of the Kw distribution and, more importantly, made clear some similarities and differences between the beta and Kw distributions. For example, the Kw densities are also unimodal, uniantimodal, increasing, decreasing or constant depending in the same way as the beta distribution on the values of its parameters. He highlighted several advantages of the Kw distribution over the beta distribution: the normalizing constant is very simple; simple explicit formulae for the distribution and quantile functions which do not involve any special functions; a simple formula for random variate generation; explicit formulae for L-moments and simpler formulae for moments of order statistics. Further, according to Jones [11], the beta distribution has the following advantages over the Kw distribution: simpler formulae for moments and moment generating function; a one-parameter sub-family of symmetric distributions; simpler moment estimation and more ways of generating the distribution via physical processes.

Consider starting from a parent continuous distribution function $G(x)$. A natural way of generating families of distributions on some other support from a simple starting parent distribution with pdf $g(x) = dG(x)/dx$ is to apply the quantile function to a family of distributions on the interval $(0, 1)$. We now combine the works of Eugene *et al.* [2] and Jones [11] (see also [12]) to construct a new class of Kw generalized (Kw-G) distributions. From an arbitrary parent cdf $G(x)$, the cdf $F(x)$ of the Kw-G distribution is defined by

$$F(x) = 1 - \{1 - G(x)^a\}^b, \tag{2}$$

where $a > 0$ and $b > 0$ are two additional parameters whose role is to introduce skewness and to vary tail weights. Because of its tractable distribution function (2), the Kw-G distribution can be used quite effectively even if the data are censored.

Correspondingly, the density function of this family of distributions has a very simple form

$$f(x) = a\,b\,g(x)\,G(x)^{a-1}\{1 - G(x)^a\}^{b-1}, \tag{3}$$

whereas the density of the beta-G distribution is given by

$$f(x) = \frac{1}{B(a, b)} g(x) G(x)^{a-1} \left\{1 - G(x)\right\}^{b-1}, \tag{4}$$

where $B(\cdot, \cdot)$ denotes the beta function. The basic difference (except for a scale multiplier) between (3) and (4) is the power of $G(x)$ inside the braces. Clearly, for $b = 1$ both densities are identical.

The new density (3) has an advantage over the class of generalized beta distributions due to Eugene *et al.* [2], since it does not involve any special function. For each continuous *name* distribution (here *name* denotes the name of the parent distribution), we can associate the *Kw-name* distribution with two extra parameters a and b from the cdf $G(x)$ and pdf $g(x)$ of the *name* distribution whose density function is defined by formula (3).

Special Kw generalized distributions can be generated as follows: the Kw-normal (KwN) distribution is obtained by taking $G(x)$ in formula (3) to be the distribution function of the normal distribution. Analogously, the Kw-Weibull (KwW), Kw-gamma ($KwGa$) and Kw-Gumbel ($KwGu$) distributions are obtained by taking $G(x)$ to be the cdf of the Weibull, gamma and Gumbel distributions, respectively, among several others. Hence, each new Kw-G distribution can be obtained from a specified G distribution. The Kw distribution is a special case of the Kw-G distribution with G being the uniform distribution on $[0,1]$, whereas the G distribution is the distribution corresponding to $a = b = 1$. With $a = 1$, the Kw-G distribution coincides with the beta-G distribution generated by the beta$(1, b)$ distribution. Furthermore, for $b = 1$ and a being an integer, the Kw-G is the distribution of the maximum of a random sample of size a from G. One major benefit of the Kw family of generalized distributions is its ability of fitting skewed data that can not be properly fitted by existing distributions.

In this article we deal with formula (3) in some generality. The mathematical properties of the Kw generalized family are usually much simpler to derive than those of the class of generalized beta distributions proposed by Eugene *et al.* [2]. Even if $g(x)$ is a symmetric function around 0, then $f(x)$ will not be a symmetric distribution even when $a = b$. From (1), if u is sampled from the uniform (0,1) distribution, then $G^{-1}(\{1-(1-u)^{1/b}\}^{1/a})$ is drawn from the Kw-G distribution.

The paper is outlined as follows. Section 2 provides some special distributions in the Kw generalized family. In Section 3, we derive general expansions for the density of the Kw-G distribution as a function of the parent density $g(x)$ multiplied by power series in $G(x)$ depending if a is integer or real non-integer. We can easily apply these expansions to several Kw-G distributions. Probability weighted moments (PWMs) are expectations of certain functions of a random variable and they can be defined for any random variable whose ordinary moments exist. In Section 4, we derive two simple expansions for the moments of any Kw-G distribution as linear functions of PWMs of the G distribution which are valid if a is integer or real non-integer. We derive in Section 5 some expansions for the density of order statistics of the class of Kw-G distributions. In Section 6, PWMs are obtained for this class. Section 7 provides an alternative formula for moments of order statistics of the Kw-G distribution. The L-moments are also given in this section. Some inferential tools are discussed in Section 8. A real data set is analyzed by some distributions in the Kw-G family in Section 9. Section 10 ends with some conclusions.

2. Special Kw generalized distributions

The Kw-G family of densities (3) allows for greater flexibility of its tails and can be widely applied in many areas of engineering and biology. We will study in Section 3 some mathematical properties of this new class of distributions because it extends several widely-known distributions in the literature. The density (3) will be most tractable when the cdf $G(x)$ and the pdf $g(x)$ have simple analytic expressions. We now discuss some special Kw generalized distributions.

2.1. *Kw-normal*

The KN density is obtained from (3) by taking $G(\cdot)$ and $g(\cdot)$ to be the cdf and pdf of the normal $N(\mu, \sigma^2)$ distribution, so that

$$f(x) = \frac{ab}{\sigma}\,\phi\left(\frac{x-\mu}{\sigma}\right)\left\{\Phi\left(\frac{x-\mu}{\sigma}\right)\right\}^{a-1}\left\{1-\Phi\left(\frac{x-\mu}{\sigma}\right)^{a}\right\}^{b-1},$$

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter, $a, b > 0$ are shape parameters, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. A random variable with density $f(x)$ above is denoted by $X \sim KwN(a, b, \mu, \sigma^2)$. For $\mu = 0$ and $\sigma = 1$ we obtain the standard KwN distribution. Further, the KwN distribution with $a = 2$ and $b = 1$ coincides with the skew normal distribution with shape parameter equal to one [13].

2.2. *Kw-Weibull*

The cdf of the Weibull distribution with parameters $\beta > 0$ and $c > 0$ is $G(x) = 1-\exp\{-(\beta x)^c\}$ for $x > 0$. Correspondingly, the density of the Kw-Weibull distribution, say $KwW(a, b, c, \beta)$, reduces to

$$\begin{aligned} f(x) &= a\,b\,c\,\beta^c x^{c-1}\exp\{-(\beta x)^c\}[1-\exp\{-(\beta x)^c\}]^{a-1} \\ &\quad \times \{1-[1-\exp\{-(\beta x)^c\}]^a\}^{b-1}, \quad x, a, b, c, \beta > 0. \end{aligned}$$

If $c = 1$ we obtain the Kw-exponential distribution. The $KwW(1, b, 1, \beta)$ distribution corresponds to the exponential distribution with parameter $\beta^\star = b\beta$.

2.3. *Kw-gamma*

Let Y be a gamma random variable with cdf $G(y) = \Gamma_{\beta y}(\alpha)/\Gamma(\alpha)$ for y, α, $\beta > 0$, where $\Gamma(\cdot)$ is the gamma function and $\Gamma_z(\alpha) = \int_0^z t^{\alpha-1}e^{-t}dt$ is the incomplete gamma function. The density of a random variable X following a $KwGa$ distribution, say $X \sim KwGa(a, b, \beta, \alpha)$, can be expressed as

$$f(x) = \frac{a\,b\,\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)^{ab}}\Gamma_{\beta x}(\alpha)^{a-1}\left\{\Gamma(\alpha)^a - \Gamma_{\beta x}(\alpha)^a\right\}^{b-1}, \quad x, \beta, \alpha, a, b > 0.$$

For $\alpha = 1$, we obtain the Kw-exponential distribution. Note that $KwGa(1, b, \beta, 1)$ means the exponential distribution with parameter $\beta^\star = b\beta$.

2.4. *Kw-Gumbel*

The density and distribution functions of the Gumbel distribution with location parameter $\mu > 0$ and scale parameter $\sigma > 0$ are given by

$$g(x) = \frac{1}{\sigma}\exp\left\{\frac{x-\mu}{\sigma} - \exp\left(\frac{x-\mu}{\sigma}\right)\right\}, \quad x > 0,$$

and

$$G(x) = 1 - \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}.$$

The mean and variance are equal to $\mu - \gamma\sigma$ and $\pi^2\sigma^2/6$, respectively, where γ is the Euler's constant ($\gamma \approx 0.57722$). Inserting these expressions into (3) yields the $KwGu$ distribution, say $KwGu(a, b, \mu, \sigma)$.

2.5. *Kw-inverse Gaussian*

Adopting the parametrization in Stasinopoulos and Rigby [14], the pdf and cdf of the inverse Gaussian distribution are

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2 x^3}} \exp\left\{-\frac{1}{2\mu^2\sigma^2 x}(x-\mu)^2\right\}, \quad x, \mu, \sigma > 0$$

and

$$G(x) = \Phi\left(\frac{1}{\sqrt{\sigma^2 x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2}{\mu\sigma^2}\right)\Phi\left(-\frac{1}{\sqrt{\sigma^2 x}}\left(\frac{x}{\mu} + 1\right)\right).$$

The expectation and variance are equal to μ and $\sigma^2\mu^3$, respectively. Replacing these expressions into (3) leads to the Kw-inverse Gaussian distribution, say $KwIG(a, b, \mu, \sigma^2)$.

Fig. 1 illustrates some of the possible shapes of the density functions for some Kw-G distributions. Fig. 2 does the same for the hazard functions defined by $h(x) = f(x)/\{1 - F(x)\}$. These plots illustrate the great flexibility achieved with the Kw-G distributions.

3. A general expansion for the density function

For $b > 0$ real non-integer, we use the series representation

$$\{1 - G(x)^a\}^{b-1} = \sum_{i=0}^{\infty} (-1)^i \binom{b-1}{i} G(x)^{a\,i},$$

where the binomial coefficient is defined for any real. From the above expansion and formula (3), we can write the Kw-G density as

$$f(x) = g(x) \sum_{i=0}^{\infty} w_i G(x)^{a(i+1)-1}, \tag{5}$$

where the coefficients are

$$w_i = w_i(a, b) = (-1)^i a\, b \binom{b-1}{i}$$

and $\sum_{i=0}^{\infty} w_i = 1$.

If b is an integer, the index i in the previous sum stops at $b - 1$. If a is an integer, formula (5) shows that the density of the Kw-G distribution is just equal to the density of the G distribution multiplied by an infinite weighted power series of cdfs of the G distribution. Otherwise, if a is real non-integer, we can expand

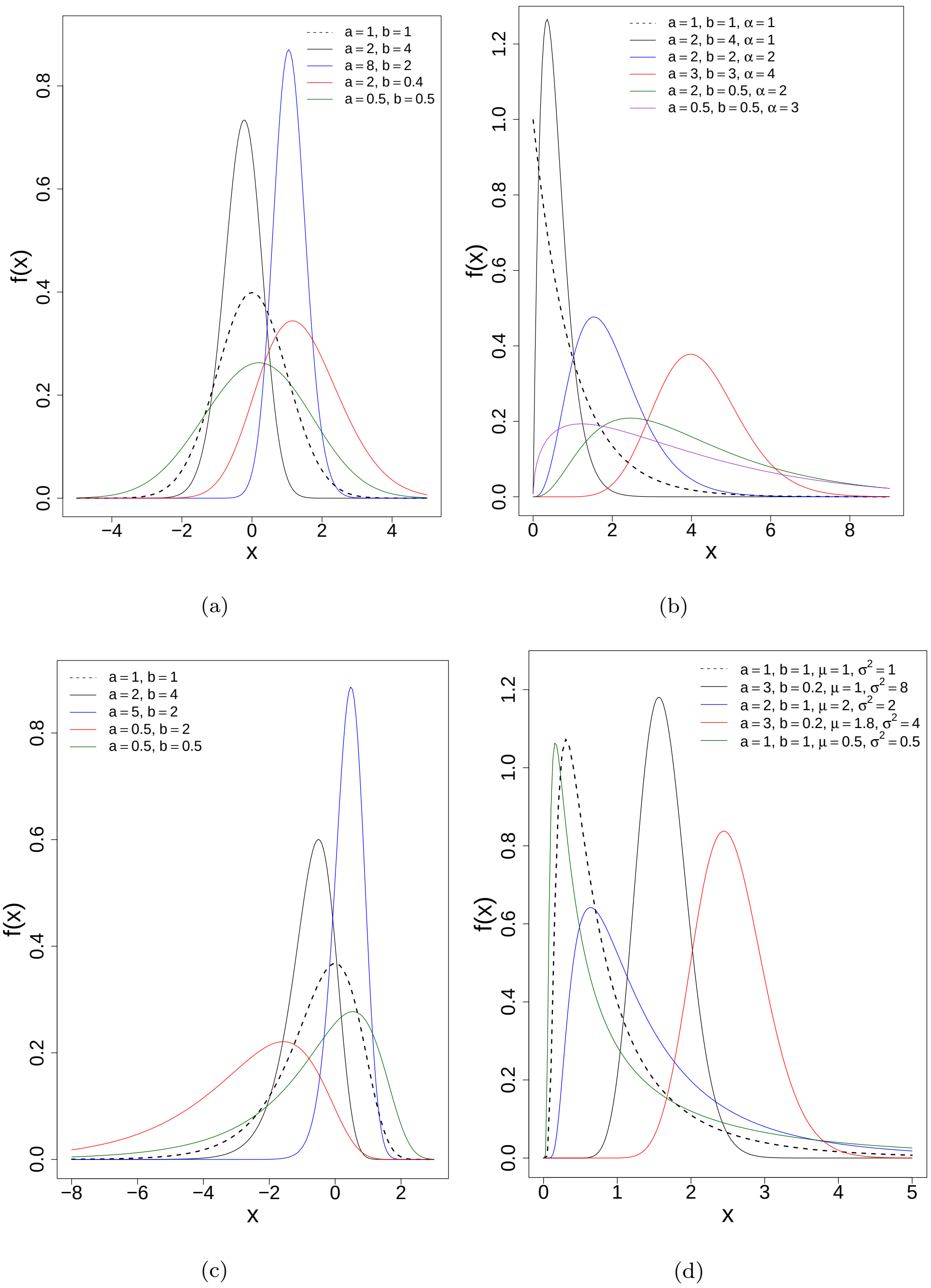

Figure 1. (a) Kw-normal$(a, b, 0, 1)$, (b) Kw-gamma$(a, b, 1, \alpha)$, (c) Kw-Gumbel$(a, b, 0, 1)$ and (d) Kw-inverse Gaussian(a, b, μ, σ^2) density functions (the dashed lines represent the parent distributions).

$G(x)^{a(i+1)-1}$ as follows

$$G(x)^{a(i+1)-1} = [1 - \{1 - G(x)\}]^{a(i+1)-1} = \sum_{j=0}^{\infty} (-1)^j \binom{a(i+1)-1}{j} \{1 - G(x)\}^j$$

and then

$$G(x)^{a(i+1)-1} = \sum_{j=0}^{\infty} \sum_{r=0}^{j} (-1)^{j+r} \binom{a(i+1)-1}{j} \binom{j}{r} G(x)^r.$$

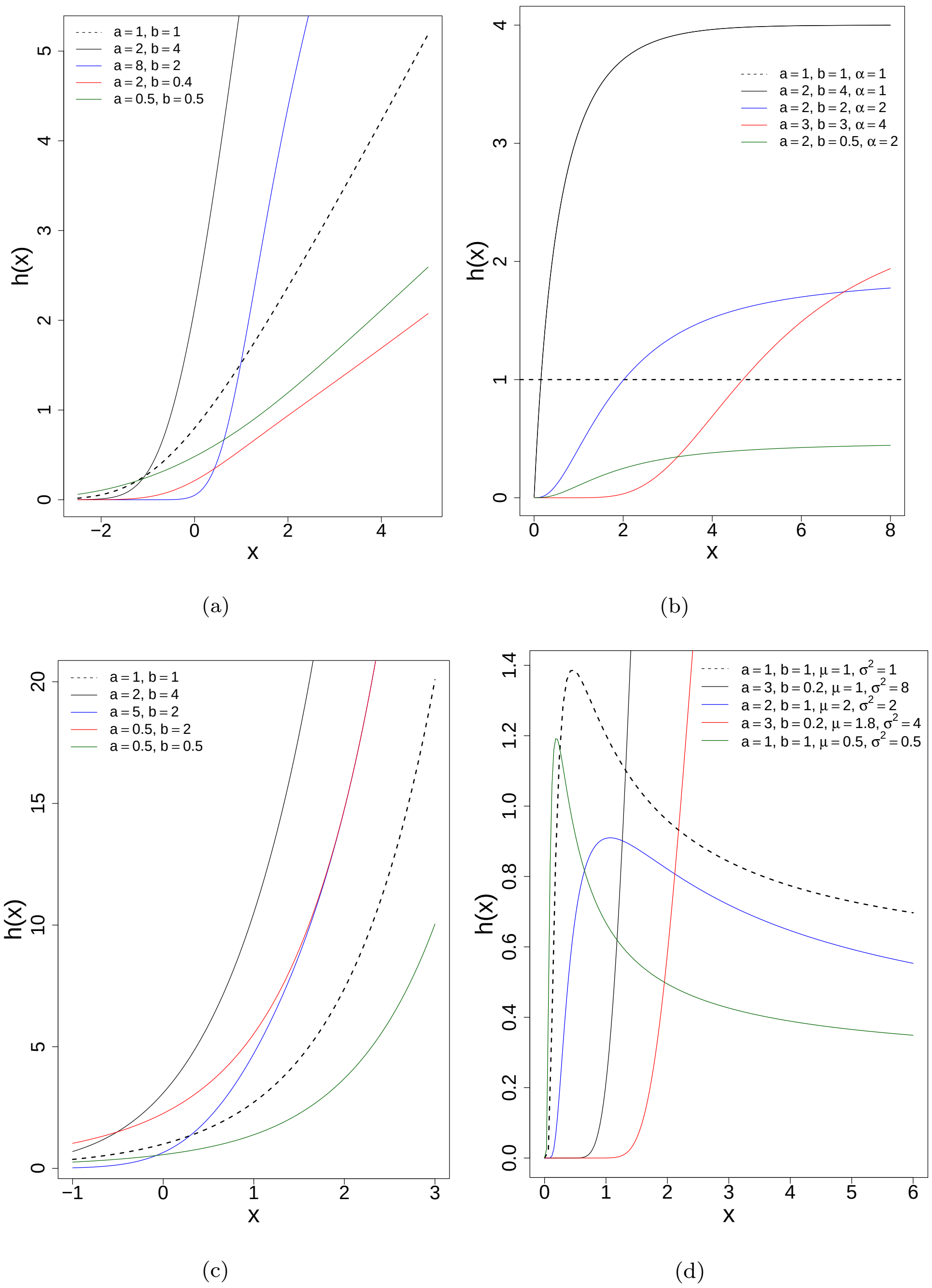

Figure 2. (a) Kw-normal$(a, b, 0, 1)$, (b) Kw-gamma$(a, b, 1, \alpha)$, (c) Kw-Gumbel$(a, b, 0, 1)$ and (d) Kw-inverse Gaussian(a, b, μ, σ^2) hazard functions (the dashed lines represent the parent distributions).

Further, the density $f(x)$ in (3) can be rearranged in the form

$$f(x) = g(x) \sum_{i,j=0}^{\infty} \sum_{r=0}^{j} w_{i,j,r} G(x)^r, \tag{6}$$

where the coefficients

$$w_{i,j,r} = w_{i,j,r}(a, b) = (-1)^{i+j+r}\, a\, b \binom{a(i+1)-1}{j} \binom{b-1}{i} \binom{j}{r} \tag{7}$$

are constants satisfying $\sum_{i,j=0}^{\infty}\sum_{r=0}^{j} w_{i,j,r} = 1$.

Expansion (6), which holds for any real non-integer a, gives the pdf of the Kw-G distribution as an infinite weighted power series of cdfs of the G distribution. If b is an integer, the index i in equation (6) stops at $b-1$. Hence, for any a real non-integer, the pdf of the Kw-G distribution is given by three (two infinite and one finite) weighted power series sums of the baseline cdf $G(x)$. The constants $w_{i,j,r}$ in formula (7) are readily computed numerically using existing software. Recall that, if a is an integer, the density of the Kw-G distribution in (5) is given by only one infinite weighted power series sum of the baseline distribution function $G(x)$.

For a real non-integer, we can derive an alternative expansion for double checking, although more expensive computationally, with three infinite sums instead of two infinite sums and one finite sum as in formula (6). First of all, we calculate a power series expansion for $G(x)^q$ which holds for any $q > 0$ real non-integer. We have

$$G(x)^q = [1 - \{1 - G(x)\}]^q = \sum_{j=0}^{\infty} \binom{q}{j} (-1)^j \{1 - G(x)\}^j$$

and then

$$G(x)^q = \sum_{j=0}^{\infty}\sum_{r=0}^{j} (-1)^{j+r} \binom{q}{j}\binom{j}{r} G(x)^r.$$

Replacing $\sum_{j=0}^{\infty}\sum_{r=0}^{j}$ by $\sum_{r=0}^{\infty}\sum_{j=r}^{\infty}$ we obtain

$$G(x)^q = \sum_{r=0}^{\infty}\sum_{j=r}^{\infty} (-1)^{j+r} \binom{q}{j}\binom{j}{r} G(x)^r$$

and

$$G(x)^q = \sum_{r=0}^{\infty} s_r(q) G(x)^r, \tag{8}$$

where the coefficients are

$$s_r(q) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{q}{j}\binom{j}{r}, \tag{9}$$

for $r = 0, 1, \ldots$ From the density (3) by expanding $G(x)^{a(j+1)}$ as in formula (8), we immediately have the Kw-G density to be

$$f(x) = g(x) \sum_{j,r=0}^{\infty} t_j(a,b) G(x)^r, \tag{10}$$

where the coefficients $t_j(a,b)$ are defined by

$$t_j(a,b) = (-1)^j\, a\, b \binom{b-1}{j} s_r(a(j+1)-1). \tag{11}$$

Hence, for a real non-integer, the pdf of the Kw-G distribution is now given by three infinite weighted power series sums of the baseline distribution function $G(x)$, i.e., two sums in equation (10) and one sum for the coefficients $t_j(a,b)$ defined in (11) which come from equation (9). The coefficients $t_j(a,b)$ are readily computed numerically using standard statistical software. Equations (5), (6) and (10) are the main results of this section and play an important role in the paper. In the numerical calculations using these equations, infinity should be substituted by a large integer number.

We conclude this section with an additional result involving the beta-G density function. For a integer, the expansion of density function in (4) is

$$f(x) = g(x) \sum_{i=0}^{\infty} w_i G(x)^{a+i-1}, \tag{12}$$

where $w_i = w_i(a,b) = (-1)^i \binom{b-1}{i} / B(a,b)$. We note that the main difference between the mixture forms in (5) and (12) is basically the power of the cdf $g(x)$. For the Kw-G distribution the power is $a(i+1)-1$, whereas for the beta-G distribution is $a+i-1$. The weights of both representations are also different. For a real non-integer, the main difference of the density expansions is given by the weights.

4. General formulae for the moments

The s-th moment of the Kw-G distribution can be expressed as an infinite weighted sum of PWMs of order (s,r) of the parent distribution G from equation (5) for a integer and from (6) (or (10)) for a real non-integer. We assume Y and X following the baseline G and Kw-G distribution, respectively. The s-th moment of X, say μ'_s, can be expressed in terms of the (s,r)-th PWMs $\tau_{s,r} = E\{Y^s G(Y)^r\}$ of Y for $r = 0, 1, \ldots$, as defined by Greenwood *et al.* [15]. For a integer, we obtain

$$\mu'_s = \sum_{r=0}^{\infty} w_r \tau_{s,\, a(r+1)-1}, \tag{13}$$

whereas for a real non-integer we write from formula (6)

$$\mu'_s = \sum_{i,j=0}^{\infty} \sum_{r=0}^{j} w_{i,j,r} \tau_{s,r}. \tag{14}$$

Formulae (13) and (14) are of very simple forms and constitute the main results of this section. We can calculate the moments of the Kw-G distribution in terms of infinite weighted sums of PWMs of the G distribution. Established power series expansions to calculate the moments of any Kw-G distribution can be more efficient than computing these moments directly by numerical integration of the expression

$$\mu'_s = ab \int x^s g(x) G(x)^{a-1} \{1 - G(x)^a\}^{b-1} dx,$$

which can be prone to rounding off errors among others.

5. Order statistics

Order statistics make their appearance in many areas of statistical theory and practice. The density $f_{i:n}(x)$ of the i-th order statistic, for $i = 1, \ldots, n$, from i.i.d. random variables $X_1, \ldots, X_n$ following any Kw-G distribution, is simply given by

$$\begin{aligned} f_{i:n}(x) &= \frac{f(x)}{B(i, n-i+1)} F(x)^{i-1}\{1 - F(x)\}^{n-i} \\ &= \frac{ab}{B(i, n-i+1)} g(x) G(x)^{i-1} [1 - \{1 - G(x)^a\}^b] \{1 - G(x)^a\}^{b(n-i+1)-1}, \end{aligned}$$

where $B(\cdot, \cdot)$ denotes the beta function, and then

$$f_{i:n}(x) = \frac{f(x)}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} F(x)^{i+j-1}. \tag{15}$$

We now present an expression for the density of order statistics of the Kw-G distribution as a function of the baseline density multiplied by infinite weighted sums of powers of $G(x)$. This result enables us to derive the ordinary moments of order statistics of the Kw-G distribution as infinite weighted sums of PWMs of the G distribution. In Section 7 we offer a simple alternative formula for the moments of order statistics of the Kw-G distribution.

From equation (2), we obtain an expansion for $F(x)^{i+j-1}$

$$F(x)^{i+j-1} = \sum_{k=0}^{i+j-1} \binom{i+j-1}{k} (-1)^k \{1 - G(x)^a\}^{kb}.$$

Using the series expansion for $\{1 - G(x)^a\}^{kb}$

$$\{1 - G(x)^a\}^{kb} = \sum_{m=0}^{\infty} (-1)^m \binom{kb}{m} G(x)^{ma}$$

and then from (8), we obtain

$$F(x)^{i+j-1} = \sum_{k=0}^{i+j-1} \binom{i+j-1}{k} (-1)^k \sum_{r=0}^{\infty} v_r(a, b, k) G(x)^r, \tag{16}$$

where the coefficients $v_r(a, b, k)$ are defined by

$$v_r(a, b, k) = \sum_{m=0}^{\infty} (-1)^m \binom{kb}{m} s_r(ma)$$

and the quantities $s_r(ma)$ come easily from (9). Interchanging the sums in formula (16), we have

$$F(x)^{i+j-1} = \sum_{r=0}^{\infty} p_{r,i+j-1}(a, b) G(x)^r,$$

where the coefficients $p_{r,u}(a,b)$ can be calculated as

$$p_{r,u}(a,b) = \sum_{k=0}^{u} \binom{u}{k} (-1)^k \sum_{m=0}^{\infty} \sum_{l=r}^{\infty} (-1)^{mr+l} \binom{kb}{m} \binom{ma}{l} \binom{l}{r} \tag{17}$$

for $r, u = 0, 1, \ldots$.

If a is real non-integer, inserting (6) and (16) into (15) and changing indices, we can rewrite the density $f_{i:n}(x)$ in the form

$$f_{i:n}(x) = \frac{g(x)}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \sum_{r,u,v=0}^{\infty} \sum_{t=0}^{v} w_{u,v,t}\, p_{r,i+j-1}(a,b)\, G(x)^{r+t}. \tag{18}$$

If a is integer, we can obtain from formulae (5), (15) and (16)

$$f_{i:n}(x) = \frac{g(x)}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \sum_{r,u=0}^{\infty} w_u\, p_{r,i+j-1}(a,b)\, G(x)^{a(u+1)+r-1}. \tag{19}$$

Formulae (18) and (19) immediately yield the density of order statistics of the Kw-G distribution as a function of the density of the baseline distribution multiplied by infinite weighted sums of powers of $G(x)$. Hence, the ordinary moments of order statistics of the Kw-G distribution can be written as infinite weighted sums of PWMs of the G distribution. These generalized moments for some baseline distributions can be accurate computationally by numerical integration as mentioned before.

6. Probability weighted moments

A general theory for PWMs covers the summarization and description of theoretical probability distributions, the summarization and description of observed data samples, nonparametric estimation of the underlying distribution of an observed sample, estimation of parameters and quantiles of probability distributions and hypothesis testing for probability distributions. The PWM method can generally be used for estimating parameters of a distribution whose inverse form cannot be expressed explicitly.

The (s,r)-th PWM of X following the Kw-G distribution, say $\tau_{s,r}^{Kw}$, is formally defined by

$$\tau_{s,r}^{Kw} = E\{X^s F(X)^r\} = \int_{-\infty}^{\infty} x^s F(x)^r f(x) dx.$$

From equations (6) and (16) we can write

$$\tau_{s,r}^{Kw} = \sum_{m,u,v=0}^{\infty} \sum_{l=0}^{v} p_{r,m}(a,b) w_{u,v,l} \tau_{s,m+l}, \tag{20}$$

where $\tau_{s,m+l} = \int_{-\infty}^{\infty} x^s\, G(x)^{m+l}\, g(x) dx$ is the $(s, m+l)$-th PWM of the G distribution and the coefficients $p_{r,m}(a,b)$ are just defined in equation (17).

Formula (20) shows that any PWM of the Kw-G distribution can be calculated from an infinite weighted linear combination of PWMs of the G distribution.

Clearly, the generalized moments $\tau_{s,r}^{Kw}$ can be obtained numerically in many existing software by taking a large number to substitute infinity in equation (20). PWMs of the baseline distributions can be evaluated by numerical integration as discussed before.

In estimation problems we use frequently the moments of order $(1, r)$. For example, for the Gumbel and Weibull distributions [15], we have

$$\tau_{1,r} = \frac{\mu + \sigma\{\log(1+r) + \epsilon\}}{1+r} \quad \text{and} \quad \tau_{1,r} = \sum_{k=0}^{r} \binom{r}{k} (-1)^k \frac{\Gamma(1+1/c)}{\beta(1+k)^{1+1/c}},$$

respectively. Thus, the quantities $\tau_{1,r}^{Kw}$ for the $KwGu$ and KwW distributions are easily computed from (20).

7. Alternative formula for moments of order statistics

We now offer an alternative formula for the moments of order statistics of the Kw-G distribution based on PWMs of the G distribution. We use the formula for the s-th moment due to Barakat and Abdelkader [16] applied to the independent and identically distributed case, subject to existence,

$$E(X_{i:n}^s) = s \sum_{j=n-i+1}^{n} (-1)^{j-n+i-1} \binom{j-1}{n-i} \binom{n}{j} I_j(s), \tag{21}$$

where $I_j(s)$ denotes the integral

$$I_j(s) = \int_{-\infty}^{\infty} x^{s-1} \{1 - F(x)\}^j dx.$$

Using the binomial expansion and interchanging terms, the last integral becomes

$$I_j(s) = \sum_{m=0}^{j} (-1)^m \binom{j}{m} \tau_{s-1,m}^{Kw},$$

where $\tau_{s-1,m}^{Kw} = \int_{-\infty}^{\infty} x^{s-1} F(x)^m dx$.

Inserting the expression for $I_j(s)$ in formula (21) yields

$$E(X_{i:n}^s) = s \sum_{j=n-i+1}^{n} \sum_{m=0}^{j} (-1)^{j-n+i+m-1} \binom{j-1}{n-i} \binom{n}{j} \binom{j}{m} \tau_{s-1,m}^{Kw}, \tag{22}$$

where the PWMs $\tau_{s-1,m}^{Kw}$ of the Kw-G distribution are immediately obtained from equation (20) as linear functions of PWMs of the G distribution. Thus, we show that the moments of order statistics of the Kw-G distribution can be expressed explicitly in terms of infinite weighted sums of PWMs of the G distribution. Formula (22) is the main result of this section.

The L-moments are analogous to the ordinary moments but can be estimated by linear combinations of order statistics. The L-moments have several theoretical advantages over the ordinary moments. They exist whenever the mean of the distribution exists, even though some higher moments may not exist. They are able to characterize a wider range of distributions and, when estimated from a sample, are more robust to the effects of outliers in the data. Unlike usual moment estimates, the parameter estimates obtained from L-moments are sometimes more accurate in small samples than even the maximum likelihood estimates (MLEs). The L-moments are linear functions of expected order statistics defined as

$$\lambda_{r+1} = (r+1)^{-1} \sum_{k=0}^{r} (-1)^k \binom{r}{k} E(X_{r+1-k:r+1}), \quad r = 0, 1, \ldots,$$

see [17]. The first four L-moments are $\lambda_1 = E(X_{1:1})$, $\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2})$, $\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3})$ and $\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})$.

From equation (22) applied to the means ($s = 1$) of order statistics, we can easily obtain expansions for the L-moments of the Kw-G distribution. The L-moments can also be calculated in terms of PWMs given in (20) as

$$\lambda_{r+1} = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \tau_{1,k}^{Kw}, \quad r = 0, 1, \ldots$$

In particular, $\lambda_1 = \tau_{1,0}^{Kw}$, $\lambda_2 = 2\tau_{1,1}^{Kw} - \tau_{1,0}^{Kw}$, $\lambda_3 = 6\tau_{1,2}^{Kw} - 6\tau_{1,1}^{Kw} + \tau_{1,0}^{Kw}$ and $\lambda_4 = 20\tau_{1,3}^{Kw} - 30\tau_{1,2}^{Kw} + 12\tau_{1,1}^{Kw} - \tau_{1,0}^{Kw}$.

8. Inference

Henceforth, let γ be the p-dimensional parameter vector of the baseline distribution in equations (2) and (3). We consider independent random variables $X_1, \ldots, X_n$, each X_i following a Kw-G distribution with parameter vector $\theta = (a, b, \gamma)$. The log-likelihood function $\ell = \ell(\theta)$ for the model parameters obtained from (3) is

$$\ell(\theta) = n\{\log(a) + \log(b)\} + \sum_{i=1}^{n} \log\{g(x_i;\gamma)\} + (a-1)\sum_{i=1}^{n} \log\{G(x_i;\gamma)\} + (b-1)\sum_{i=1}^{n} \log\{1 - G(x_i;\gamma)^a\}.$$

The elements of the score vector are given by

$$\frac{\partial \ell(\theta)}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \log\{G(x_i;\gamma)\} \left\{1 - \frac{(b-1)G(x_i;\gamma)^a}{1 - G(x_i;\gamma)^a}\right\},$$

$$\frac{\partial \ell(\theta)}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} \log\{1 - G(x_i;\gamma)^a\}$$

and

$$\frac{\partial \ell(\theta)}{\partial \gamma_j} = \sum_{i=1}^{n} \left[\frac{1}{g(x_i;\gamma)} \frac{\partial g(x_i;\gamma)}{\partial \gamma_j} + \frac{1}{G(x_i;\gamma)} \frac{\partial G(x_i;\gamma)}{\partial \gamma_j} \left\{1 - \frac{a(b-1)}{G(x_i;\gamma)^{-a} - 1}\right\}\right],$$

Table 1. *AIC* in increasing order, parameter estimates and standard errors for the adjusted distributions.

Distribution	AIC	Parameter estimate (Standard error)			
		a	b		
Beta normal	7176.9	18.20	0.25	$\mu = 12.69$	$\sigma = 42.76$
		(12.7)	(0.0749)	(23.4)	(8.89)
Kw-normal	7177.4	14.86	0.27	$\mu = 25.52$	$\sigma = 42.24$
		(1.74)	(0.0531)	(0.592)	(3.86)
Kw-exponential	7180.3	15.54	1.34	$\beta = 46.22$	
		(3.61)	(0.325)	(6.98)	
Kw-gamma	7180.9	1.85	0.67	$\alpha = 7.37$	$\beta = 14.21$
		(1.36)	(0.476)	(6.31)	(11.7)
Gamma	7183.9			$\alpha = 8.99$	$\beta = 15.56$
				(0.474)	(0.843)

for $j = 1, \ldots, p$. These partial derivatives depend on the specified baseline distribution. Numerical maximization of the log-likelihood above is accomplished by using the RS method [18] available in the gamlss package [14] in R [19]. Since numerically the maximum likelihood estimation of the parameters of the Kw-G distributions is much simpler than the estimation of the parameters of the generalized beta distributions, we recommend to use Kw-G distributions in place of the second family of distributions. Under suitable regularity conditions, the asymptotic distribution of the maximum likelihood estimator $\widehat{\theta}$ is multivariate normal with mean vector θ and covariance matrix that can be estimated by $\left\{-\partial^2\ell(\theta)/\partial\theta\partial\theta^\top\right\}^{-1}$ evaluated at $\theta = \widehat{\theta}$. The required second derivatives are computed numerically.

Consider two nested Kw-G distributions: a Kw-G_A distribution with corresponding parameters $\theta_1, \ldots, \theta_r$ and maximized log-likelihood $-2\ell(\widehat{\theta}_A)$, and a Kw-G_B distribution containing the same parameters $\theta_1, \ldots, \theta_r$ plus additional parameters $\theta_{r+1}, \ldots, \theta_p$ and maximized log-likelihood $-2\ell(\widehat{\theta}_B)$, the models otherwise being identical. For testing the Kw-G_A distribution against the Kw-G_B distribution, the likelihood ratio statistic (LR) is simply equal to the difference $-2\{\ell(\widehat{\theta}_A) - \ell(\widehat{\theta}_B)\}$ and has an asymptotic χ^2_{p-r} distribution.

We can compare non-nested Kw-G distributions by penalizing over-fitting using the Akaike information criterion given by $AIC = -2\ell(\widehat{\theta}) + 2p^\star$, where $p^\star$ is the number of model parameters. The distribution with the smallest value of AIC (among all distributions considered) is usually taken as the best model for describing the given data set. This comparison is based on the consideration of a model that shows a lack of fit with one that does not.

9. Application

In this section we present an example with data from adult numbers of *T. confusum* cultured at 29°C presented by Eugene *et al.* [2]. Table 1 gives AIC values in increasing order for some fitted distributions and the MLEs of the parameters together with its standard errors. According to AIC, the beta normal and Kw-normal distributions yield slightly different fittings, outperforming the remaining selected distributions. Notice that for the beta normal distribution the variability in the estimates of a, μ and σ is appreciably greater.

The fitted distributions superimposed to the histogram of the data in Figure 3 reinforce the result in Table 1 for the gamma distribution. The beta normal and the Kw-normal distributions are almost indistinguishable. This claim is further strengthened by the comparison between observed and expected frequencies in Table 2. The mean absolute deviation between expected and observed frequencies

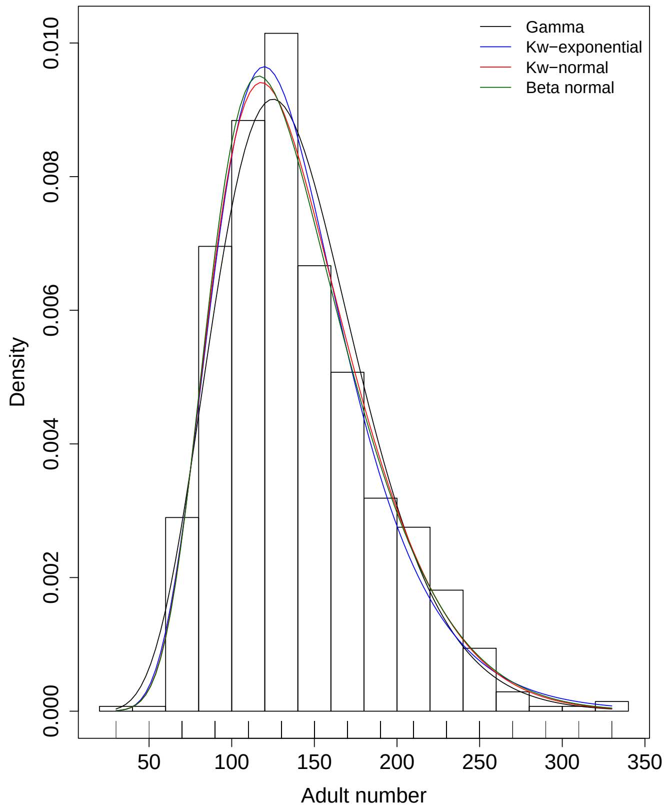

Figure 3. Histogram of adult number and fitted probability density functions.

Table 2. Observed and expected frequencies of adult numbers for *T. confusum* cultured at 29°C and mean absolute deviation (MAD) between the frequencies.

Adult		Expected			
number	Observed	Gamma	Kw-exponential	Kw-normal	Beta normal
30	1	0.75	0.19	0.21	0.22
50	1	9.85	6.32	5.77	5.67
70	40	39.72	37.82	37.43	37.39
90	96	83.59	90.75	91.85	93.70
110	122	117.15	127.14	125.53	127.54
130	140	124.56	127.81	123.95	123.73
150	92	108.51	104.22	102.41	100.77
170	70	81.35	74.35	75.95	74.30
190	44	54.24	48.68	52.00	50.97
210	38	32.93	30.16	33.18	32.80
230	25	18.51	18.03	19.80	19.86
250	13	9.76	10.53	11.06	11.32
270	4	4.87	6.06	5.79	6.08
290	1	2.32	3.46	2.84	3.08
310	1	1.06	1.96	1.31	1.47
330	2	0.47	1.10	0.56	0.66
Total	690	689.7	688.6	689.6	689.5
MAD		6.17	4.74	4.60	4.39

reaches the minimum value for the Kw-normal distribution.

Based on the values of the LR statistic (Section 8), the Kw-gamma and the Kw-exponential distributions are not significantly different yielding $LR = 1.542$ (1 d.f., p-value $= 0.214$). Comparing the Kw-gamma and the gamma distributions, we find a significant difference ($LR = 6.681$, 2 d.f., p-value $= 0.035$).

10. Conclusions

Following the idea of the class of beta generalized distributions [2] and the distribution by Kumaraswamy [1], we define a new family of Kw generalized (Kw-G) distributions to extend several widely-known distributions such as the normal, Weibull, gamma and Gumbel distributions. For each distribution G, we can define the corresponding Kw-G distribution using simple formulae.

We show how some mathematical properties of the Kw-G distributions are readily obtained from those of the parent distributions. The moments of the Kw-G distribution can be expressed explicitly in terms of infinite weighted sums of probability weighted moments (PWMs) of the G distribution. The same happens for the moments of order statistics and PWMs of the Kw-G distributions.

We discuss maximum likelihood estimation and inference on the parameters. The maximum likelihood estimation in Kw-G distributions is much simpler than the estimation in beta generalized distributions. Further, we can easily compute the maximum values of the unrestricted and restricted log-likelihoods to construct likelihood ratio statistics for testing nested models in the new family of distributions. An application of the new family to real data is given to show the feasibility of our proposal. We hope this generalization may attract wider applications in statistics.

Acknowledgements

We would like to thank an anonymous referee for helpful comments and suggestions which improved the paper. The work of the first author is partially supported by CNPq, Brazil.

References

[1] P. Kumaraswamy. Generalized probability density-function for double-bounded random-processes. *Journal of Hydrology* 462, 79–88, 1980.
[2] N. Eugene, C. Lee, and F. Famoye. Beta-normal distribution and its applications. *Communications in Statistics. Theory and Methods* 31, 497–512, 2002.
[3] S. Nadarajah and S. Kotz. The beta Gumbel distribution. *Mathematical Problems in Engineering* 10, 323–332, 2004.
[4] S. Nadarajah and A. K. Gupta. The beta Fréchet distribution. *Far East Journal of Theoretical Statistics* 14, 15–24, 2004.
[5] S. Nadarajah and S. Kotz. The beta exponential distribution. *Reliability Engineering & System Safety* 91, 689–697, 2006.
[6] V. Sundar and K. Subbiah. Application of double bounded probability density-function for analysis of ocean waves. *Ocean Engineering* 16, 193–200, 1989.
[7] S. C. Fletcher and K. Ponnambalam. Estimation of reservoir yield and storage distribution using moments analysis. *Journal of Hydrology* 182, 259–275, 1996.
[8] A. Seifi, K. Ponnambalam, and J. Vlach. Maximization of manufacturing yield of systems with arbitrary distributions of component values. *Annals of Operations Research* 99, 373–383, 2000.
[9] A. Ganji, K. Ponnambalam, D. Khalili, and M. Karamouz. Grain yield reliability analysis with crop water demand uncertainty. *Stochastic Environmental Research and Risk Assessment*, 20, 259–277, 2006.
[10] R. E. Barlow and F. Proschan. *Statistical theory of reliability and life testing: probability models.* Holt, Rinehart and Winston, New York; London, 1975.
[11] M. C. Jones. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology* 6, 70–81, 2008.
[12] M. C. Jones. Families of distributions arising from distributions of order statistics (with discussion). *Test* 13, 1–43, 2004.
[13] A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178, 1985.
[14] D. M. Stasinopoulos and R. A. Rigby. Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software* 23, 1–46, 2007.
[15] J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis. Probability weighted moments – definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research* 15, 1049–1054, 1979.

[16] H. M. Barakat and Y. H. Abdelkader. Computing the moments of order statistics from nonidentical random variables. *Statistical Methods and Applications* 13, 15–26, 2004.

[17] J. R. M. Hosking. *L*-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B.* 52, 105–124, 1990.

[18] R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape (with discussion). *Applied Statistics* 54, 507–554, 2005.

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