Generalized Beta-Generated Distributions

Carol Alexander,[a] Gauss M. Cordeiro,[b] Edwin M. M. Ortega[c] and José María Sarabia[d]

Abstract

This article introduces generalized beta-generated (GBG) distributions. Sub-models include all classical beta-generated, Kumaraswamy-generated and exponentiated distributions. They are maximum entropy distributions under three intuitive conditions, which show that the classical beta generator skewness parameters only control tail entropy and an additional shape parameter is needed to add entropy to the centre of the parent distribution. This parameter controls skewness without necessarily differentiating tail weights. The GBG class also has tractable properties: we present various expansions for moments, generating function and quantiles. The model parameters are estimated by maximum likelihood and the usefulness of the new class is illustrated by means of some real data sets.

Key Words: Entropy, Exponentiated, Kumaraswamy, Kurtosis, McDonald, Minimax, Skewness.

[a] ICMA Centre, Henley Business School at Reading, UK. e-mail:*c.alexander@icmacentre.ac.uk*
[b] Departamento de Estatística, Universidade Federal de Pernambuco, Brazil.
e-mail:*gausscordeiro@uol.com.br*
[c] Departamento de Ciências Exatas, Universidade de São Paulo, Brazil. e-mail:*edwin@esalq.usp.br*
[d] Department of Economics, University of Cantabria, Spain. e-mail:*sarabiaj@unican.es*

1 Introduction

The statistics literature is filled with hundreds of continuous univariate distributions: see Johnson *et al.* (1994, 1995). Recent developments focus on new techniques for building meaningful distributions, including the two-piece approach introduced by Hansen (1994), the perturbation approach of Azzalini and Capitanio (2003), and the generator approach pioneered by Eugene *et al.* (2002) and Jones (2004). Many subsequent articles apply these techniques to induce a skew into well-known symmetric distributions such as the Student t: see Aas and Haff (2006) for a review. Using the two-piece approach with a view to finance applications, Zhu and Galbraith (2010) argued that, in addition to Student t parameters, three shape parameters are required: one parameter to control asymmetry in the centre of a distribution and two parameters to control the left and right tail behavior.

This paper addresses similar issues to Zhu and Galbraith but takes a different approach. We introduce a class of generalized beta generated (GBG) distributions that have three shape parameters in the generator. By considering quantile-based measures of skewness and kurtosis and by decomposing the entropy, we demonstrate that two parameters control skewness and kurtosis through altering only the tail entropy and one controls skewness and kurtosis through adding entropy to the centre of the parent distribution as well.

Denote the parent distribution and density by $F(\cdot)$ and $f(\cdot)$, respectively, and let $X = F^{-1}(U)$ with $U \sim \mathcal{B}(a,b)$, the classical beta distribution. Then the random variable X is said to have a beta generated (BG) distribution. This may be characterized by its density

$$f_{\mathcal{BG}}(x) = B(a,b)^{-1}\, f(x)\, F(x)^{a-1}\, [1 - F(x)]^{b-1}, \qquad x \in \mathcal{I}. \tag{1}$$

The general class of BG distributions were introduced by Jones (2004), who concentrates on cases where F is symmetric about zero with no free parameters other than location and scale and where $\mathcal{I}$ is the whole real line.

The first distribution of the BG class to be studied in depth was the beta normal distribution, introduced by Eugene *et al.* (2002). Denote the standard normal distribution and density by $\Phi(.)$ and $\phi(.)$, respectively, and let $X = \Phi^{-1}(U)$ with $U \sim \mathcal{B}(a,b)$, the classical beta distribution. Then X has a beta normal distribution $\mathcal{BN}(a,b,0,1)$ with density

$$f_{\mathcal{BN}}(x; a, b, 0, 1) = B(a,b)^{-1}\, \phi(x)\, [\Phi(x)]^{a-1}\, [1 - \Phi(x)]^{b-1}, \qquad -\infty < x < \infty. \tag{2}$$

Location and scale parameters are redundant in the generator since if $X \sim \mathcal{BN}(a,b,0,1)$ then $Y = \sigma X + \mu \sim \mathcal{BN}(a,b,\mu,\sigma)$ has the non-standard beta normal distribution with $N(\mu,\sigma^2)$ parent. The parameters a and b control skewness through the relative tail weights. The beta normal density is symmetric if $a = b$, it has negative skewness when $a < b$ and positive skewness when $a > b$. When $a = b > 1$ the beta-normal distribution has positive excess kurtosis and when $a = b < 1$ it has negative excess kurtosis, as demonstrated by Eugene *et al.* (2002). However, both skewness and kurtosis are very limited and the only way to gain even a modest degree of excess kurtosis is to skew the distribution as far as possible. Eugene *et al.* (2002) tabulated the mean, variance, skewness and kurtosis of $\mathcal{BN}(a,b,0,1)$ for some particular values of a and b between 0.05 and 100. The skewness always lies in the interval $(-1, 1)$ and the largest kurtosis value found is 4.1825, for $a = 100$ and $b = 0.1$ and vice versa.

The BG class encompasses many other types of distributions, including skewed t and log F. Other specific BG distributions have been studied by Nadarajah and Kotz (2004, 2005), Akinsete *et al.* (2008), Zografos and Balakrishnan (2009) and Barreto-Souza *et al.* (2010). Jones and Larsen (2004) and Arnold *et al.* (2006) introduced the multivariate BG class. Some practical applications have been considered: e.g. Jones and Larsen (2004) fitted skewed t and log F to temperature data; Akinsete *et al.* (2008) fitted the beta Pareto distribution to flood data; and Razzaghi (2009) applied the beta normal distribution to dose-response modeling. However, the classical beta generator has only two parameters, so it can add only a limited structure to the parent distribution. For many choices of parent the computations of quantiles and moments of a BG distribution can become rather complex. Also, when $a = b$ (so the skewness is zero if F is symmetric) the beta generator typically induces *meso*kurtosis, in that the BG distribution has a lower kurtosis than the parent. For example, using a Student t parent and $a = b > 1$ we find that the kurtosis converges rapidly to 3 as a and b increase, and for $a = b < 1$ the kurtosis is infinite.

Jones (2009) advocated replacing the beta generator by the Kumaraswamy (1980) distribution, commonly termed the "minimax" distribution. It has tractable properties especially for simulation, as its quantile function takes a simple form. However, Kumaraswamy-generated (KwG) distributions still introduce only two extra shape parameters, whereas three may be required to control both tail weights and the distribution of weight in the centre. Therefore, we propose the use of a more flexible generator distribution: the generalized beta distribution of the first kind. It has one more shape parameter than the classical beta and Kumaraswamy distributions, and we shall demonstrate that this parameter gives additional control over both skewness and kurtosis. Special cases of GBG distributions include BG and KwG distributions and the class of exponentiated distributions.

The rest of the paper is organized as follows. Section 2 describes the distribution, density and hazard functions of the GBG distribution. Section 3 investigates the role of the generator parameters and relates this to the skewness of the GBG distribution and the decomposition of the GBG entropy. In Section 4, we present some special models. A variety of theoretical properties are considered in Section 5. Estimation by maximum likelihood (ML) is described in Section 6. We present a simulation study in Section 7. In Section 8, we provide some empirical applications. Finally, conclusions are noted in Section 9.

2 The GBG distribution

The generalized beta distribution of the first kind (or, beta type I) was introduced by McDonald (1984). It may be characterized by its density

$$f_{\mathcal{GB}}(u;a,b,c) = c\,B(a,b)^{-1}\,u^{ac-1}\,(1-u^c)^{b-1},\ \ 0<u<1, \tag{3}$$

where $a>0$, $b>0$ and $c>0$. Two important special cases are the classical beta distribution ($c=1$), and the Kumaraswamy distribution ($a=1$).

Given a parent distribution $F(x;\boldsymbol{\tau})$, $x\in\mathcal{I}$ with parameter vector $\boldsymbol{\tau}$ and density $f(x;\boldsymbol{\tau})$, the GBG distribution may be characterized by its density:

$$f_{\mathcal{GBG}}(x;\boldsymbol{\tau},a,b,c) = c\,B(a,b)^{-1}\,f(x;\boldsymbol{\tau})\ F(x;\boldsymbol{\tau})^{ac-1}\ [1-F(x;\boldsymbol{\tau})^c]^{b-1}, x\in\mathcal{I}. \tag{4}$$

Now, a,b and c are shape parameters, in addition to those in $\boldsymbol{\tau}$. If X is a random variable with density (4), we write $X\sim\mathcal{GBG}(F;\boldsymbol{\tau},a,b,c)$. Two important special sub-models are the BG distribution ($c=1$) proposed by Jones (2004), and the Kumaraswamy generated (KwG) distribution ($a=1$) recently proposed by Cordeiro and de Castro (2011). Of course, the beta type I density function itself arises immediately if $F(x;\boldsymbol{\tau})$ is taken to be the uniform distribution.

We remark that Zografos (2011) studied a closely-related class of distributions that are generated by the distributions defined by McDonald and Xu (1995). These have a scale parameter that is unnecessary for GBG distributions, because they admit a straightforward translation of location and scale. If $Y=\mu+\sigma X$ then Y has density

$$f_{\mathcal{GBG}}(y;\boldsymbol{\tau},a,b,c,\mu,\sigma) = K(a,b,c,\sigma)\,f(\sigma^{-1}(y-\mu))F(\sigma^{-1}(y-\mu))^{ac-1}\left[1-F(\sigma^{-1}(y-\mu))^c\right]^{b-1},$$

where $K(a,b,c,\sigma) = c\,\sigma^{-1}\,B(a,b)^{-1}$. Assuming the mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$ of X exist, then Y has mean $\mu+\sigma\,\tilde{\mu}$ and standard deviation $\sigma\tilde{\sigma}$. Hence, it is redundant to use scale or location parameters in the generator itself, only shape parameters are needed.

The random variable X admits the simple stochastic representation

$$X = F^{-1}(U^{1/c}), \quad U\sim\mathcal{B}(a,b). \tag{5}$$

Using the transformation (5), the distribution of (4) may be written:

$$F_{\mathcal{GBG}}(x;\boldsymbol{\tau},a,b,c) = I(F(x;\boldsymbol{\tau})^c;a,b) = B\,(a,b)^{-1}\int_0^{F(x;\boldsymbol{\tau})^c}\omega^{a-1}(1-\omega)^{b-1}d\omega, \tag{6}$$

where $I(x;a,b)$ denotes the incomplete beta ratio function. The GBG hazard rate function (hrf) reduces to

$$h(x;\boldsymbol{\tau},a,b,c) = \frac{c\,f(x;\boldsymbol{\tau})\,F(x;\boldsymbol{\tau})^{ac-1}\,\{1-F(x;\boldsymbol{\tau})^c\}^{b-1}}{B\,(a,b)\,\{1-I(F(x;\boldsymbol{\tau})^c;a,b)\}}. \tag{7}$$

It follows immediately from (4) that the GBG with parent distribution F is a standard BG distribution with exponentiated parent distribution F^c. The properties of exponentiated distributions have been studied by several authors: see Mudholkar *et al.* (1995) for exponentiated Weibull, Gupta *et al.* (1998) for exponentiated Pareto, Gupta and Kundu (2001) for exponentiated exponential, and Nadarajah and Gupta (2007) for the exponentiated gamma distribution.

From now on, for an arbitrary parent distribution $F(x;\boldsymbol{\tau})$, we write $X\sim\text{Exp}^cF$ if X has distribution and density functions given by

$$G_c(x) = F(x;\boldsymbol{\tau})^c \qquad \text{and} \qquad g_c(x) = cf(x;\boldsymbol{\tau})\,F(x;\boldsymbol{\tau})^{c-1},$$

respectively. This is called the Lehmann type I distribution. Clearly, the double construction beta-Exp$^c F$ yields the GBG distribution. The derivations of several properties of GBG distributions will be facilitated by this simple transformation. Note that the dual transformation Exp$^c(1-F)$, referred to as the Lehmann type II distribution, corresponds to the parent $[1 - F(x; \boldsymbol{\tau})]^c$. Thus, in addition to the classical BG and KwG distributions and the parent distribution F itself, the GBG distribution encompasses the Exp$F$ (for $b = 1$) and Exp$(1 - F)$ (for $a = c = 1$) distributions. The classes of distributions that are included as special sub-models of the GBG class are displayed in Figure 1.

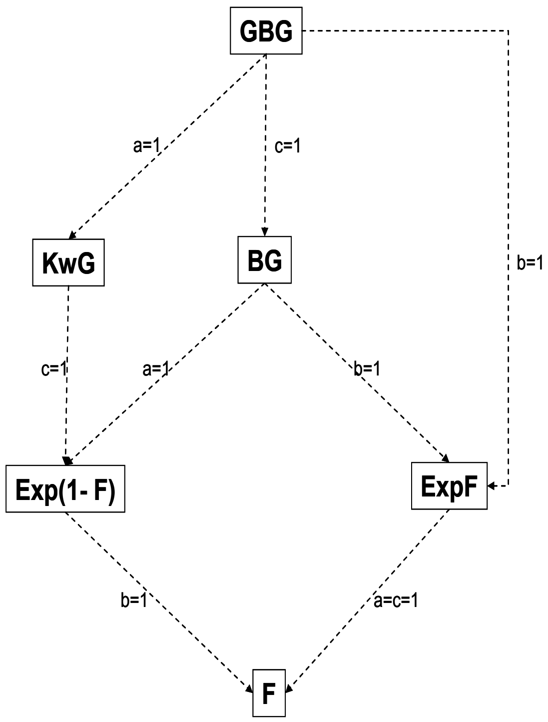

Figure 1: Relationships of the GBG sub-models.

We now state some useful expansions for the GBG distribution and density. From henceforth we shall drop the explicit reference to the parent parameters $\boldsymbol{\tau}$, unless otherwise stated, for brevity of notation. Also note that the binomial coefficient generalized to real arguments is defined as $\binom{x}{y} = \Gamma(x+1)/[\Gamma(y+1)\Gamma(x-y+1)]$.

First, when b is an integer (6) may be written:

$$F_{\mathcal{GBG}}(x; a, b, c) = \sum_{r=0}^{b-1} \binom{a+b-1}{r} F(x)^{c(a+b-r-1)} \left[1 - F(x)^c\right]^r.$$

Alternatively, if a is an integer:

$$F_{\mathcal{GBG}}(x; a, b, c) = 1 - \sum_{r=0}^{a-1} \binom{a+b-r}{r} F(x)^{cr} \left[1 - F(x)^c\right]^{a+b-r-1}.$$

Other expansions for (4) and (6) may be derived using properties of exponentiated distributions. Expanding the binomial in (4) yields:

$$f_{\mathcal{GBG}}(x; a, b, c) = c\, B(a, b+1)^{-1} f(x) \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} F(x)^{c\,(a+i)-1}. \tag{8}$$

Hence, the GBG is a linear combination of exponentiated-F distributions. Its density may be written:

$$f_{\mathcal{GBG}}(x;a,b,c) = \sum_{i=0}^{\infty} w_i\, g_{a\,(i+c)}(x), \tag{9}$$

and its distribution is

$$F_{\mathcal{GBG}}(x;a,b,c) = \sum_{i=0}^{\infty} w_i\, G_{a\,(i+c)}(x), \tag{10}$$

where $g_{a\,(i+c)}(x)$ and $G_{a\,(i+c)}(x)$ denote the density and distribution of $\mathrm{Exp}^{a\,(i+c)}F$ and

$$w_i = \frac{(-1)^i \binom{b}{i}}{(a+i)\,B(a,b+1)}. \tag{11}$$

The linear combinations (9) and (10) allow certain GBG properties to be derived from the corresponding properties of exponentiated distributions.

Finally, we remark that the distribution (6) may also be expressed in terms of the hypergeometric function as

$$F_{\mathcal{GBG}}(x;a,b,c) = \frac{c\,F(x)^a}{a\,B(a,b)}\, {}_2F_1\left(F(x)^a; a, 1-b, a+1\right),$$

where

$${}_2F_1(x;a,b,c) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\,\Gamma(b+j)}{\Gamma(c+j)}\,\frac{x^j}{j!}.$$

Thus, the GBG properties could, in principle, be obtained from well-established properties of the hypergeometric function. See Gradshteyn and Ryzhik (2000, Section 9.1).

3 Entropy and higher moments

In this section we interpret the effect of each parameter a, b and c by examining the skewness and kurtosis of the generated distribution and the decomposition of the GBG Shannon entropy.

3.1 Skewness and Kurtosis

To illustrate the effect of the shape parameters a, b and c on skewness and kurtosis we consider measures based on quantiles. The shortcomings of the classical kurtosis measure are well-known; see Seier and Bonett (2003) and Brys *et al.* (2006). There are many heavy-tailed distributions for which this measure is infinite, so it becomes uninformative precisely when it needs to be. Indeed, our motivation to use quantile-based measures stemmed from the non-existence of classical kurtosis for many of the GBG distributions we studied.

The Bowley skewness (see Kenney and Keeping, 1962) is based on quartiles:

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

and the Moors kurtosis (see Moors, 1998) is based on octiles:

$$M = \frac{Q(7/8) - Q(5/8) - Q(3/8) + Q(1/8)}{Q(6/8) - Q(2/8)},$$

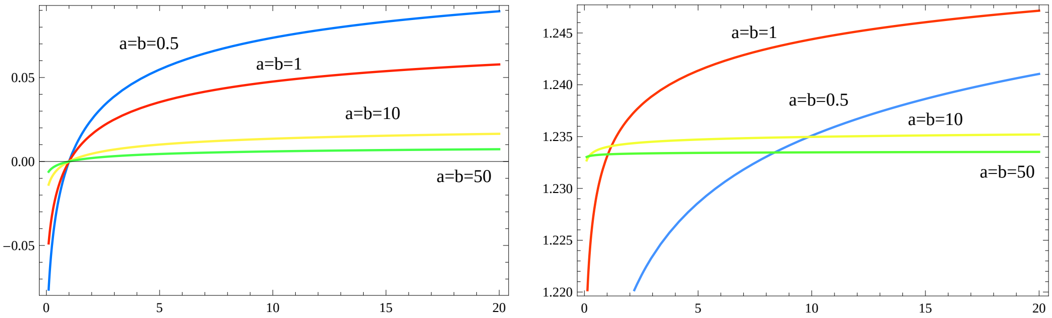

Figure 2: The Bowley skewness (left) and Moors kurtosis (right) coefficients for the GBN distribution as a function of c, for $a = b = 0.5, 1, 10$ and 50.

where $Q(\cdot)$ represents the quantile function. These measures are less sensitive to outliers and they exist even for distributions without moments. For the standard normal distribution, these measures are 0 (Bowley) and 1.2331 (Moors). For the classical Student t distribution with 10 degrees of freedom, these measures are 0 (Bowley) and 1.27705 (Moors). For the classical Student t distribution with 5 degrees of freedom, they are zero (Bowley) and 1.32688 (Moors). Figures 2, 3 and 4 depict the Bowley and Moors measures for the GB-normal (GBN) distribution and for the GB-Student distribution with 10 and 5 degrees of freedom, respectively. These are easily derived from the GBG quantile function given in Section 5.2. Since our primary purpose is to investigate the effect that c has on skewness and kurtosis, we have represented these measures as a function of c assuming that the other two shape parameters a and b are equal and take the values 0.5, 1, 10 and 50. For all three distributions, the Bowley skewness coefficient is zero for $c = 1$, negative for $c < 1$ and positive for $c > 1$. Whilst the three skewness plots are similar, they have different vertical and horizontal scales, and it is clear that when $c \neq 1$ the degree of skewness that can be induced by the generator increases with the kurtosis of the parent, and decreases with the values of a and b.

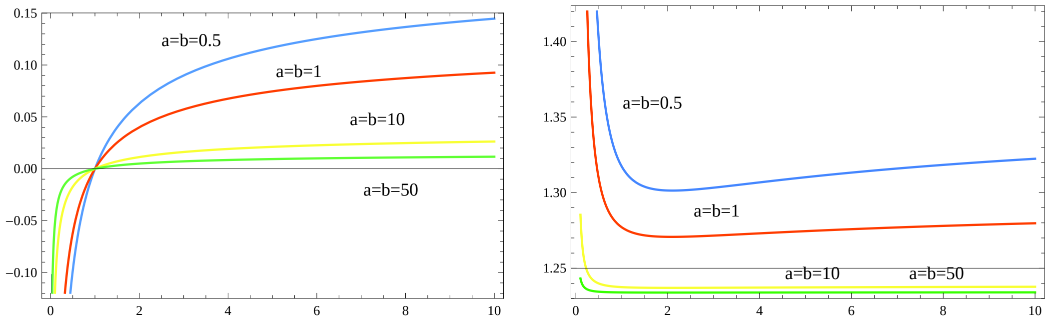

Figure 3: The Bowley skewness (left) and Moors kurtosis (right) coefficients for the GB-Student t distribution with 10 degrees of freedom as a function of c, for $a = b = 0.5, 1, 10$ and 50.

A different picture emerges on consideration of the Moors kurtosis, which may be fairly limited for GBG distributions. With a normal parent, the right hand graph in Figure 2 shows that the Moors coefficient is increasing in c, for all values of a and b, and is largest when $a = b = 1$, except when c is very small. It is equal to the parent kurtosis for $c = 1$, less than the parent kurtosis for $c < 1$ and greater for $c > 1$. However, as c increases the Moors kurtosis converges to a limit only a little greater than 1.2331, i.e. the normal Moors kurtosis. Therefore we conclude that when the parent is normal the additional parameter c acts primarily as another skewness parameter rather than a kurtosis parameter.

The right hand graphs of Figures 3 and 4 depict the Moors kurtosis of GBG distributions

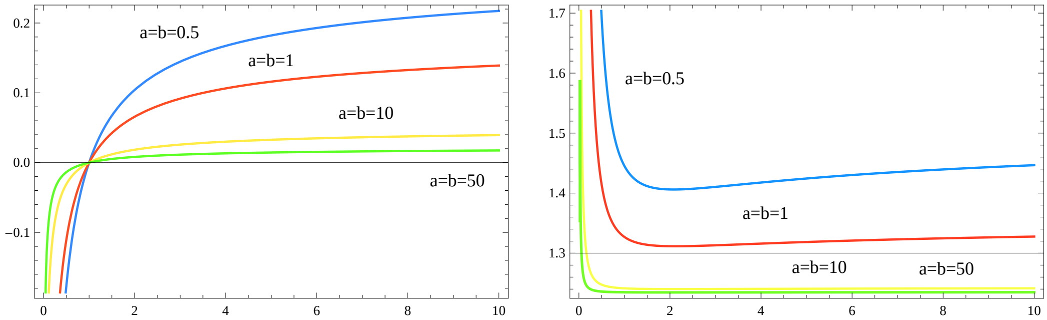

Figure 4: The Bowley skewness (left) and Moors kurtosis (right) coefficients for the GB-Student t distribution with 5 degrees of freedom as a function of c, for $a = b = 0.5, 1, 10$ and 50.

with Student t parents. Here the Moors measure *can* become large and positive, but only for small values of c. Initially, as c increases from small values the measure decreases sharply, presenting a minimum around $c = 2$. When $a = b \leq 1$ the Moors kurtosis of the GB Student t distribution is always greater than the Moors kurtosis of the classical Student t distribution (marked on the graph by the horizontal line). The Moors measures are very close to zero when $a = b = 10$, and when $a = b = 50$, except when c takes extremely small values. For small values of c the Moors kurtosis can become quite large, but as this is accompanied by large negative values of Bowley skewness we conclude that c is again primarily a skewness parameter.

3.2 Entropy

The Shannon (1948) entropy $H(g) = -\mathrm{E}_g[\log g(x)] = -\int g(x)\log g(x)dx$ of a density $g(x)$ (henceforth called simply entropy) is a measure of the uncertainty in a probability distribution and its negative is a measure of information. Ebrahimi *et al.* (1999) showed that there is no universal relationship between variance and entropy and where their orderings differ entropy is the superior measure of information. Because the generalized beta is a classical BG distribution with parent $G(x) = x^c$, we may write (3) as:

$$f_{\mathcal{GB}}(x; a, b, c) = B(a, b)^{-1} g(x)[G(x)]^{a-1}[1 - G(x)]^{b-1}, \tag{12}$$

where $G(x) = x^c$, $0 < x < 1$ and $g(x) = G'(x)$. Hence,

$$\begin{aligned}
-\mathrm{E}_{\mathcal{GB}}[\log f_{\mathcal{GB}}(x)] &= \log B(a,b) - \mathrm{E}_{f_{\mathcal{GB}}}[\log g(x)] \\
&\quad -(a-1)\mathrm{E}_{f_{\mathcal{GB}}}[\log G(x)] - (b-1)\mathrm{E}_{f_{\mathcal{GB}}}[\log(1 - G(x))], \\
-\mathrm{E}_{\mathcal{GB}}[\log G(x)] &= \zeta(a,b), \\
-\mathrm{E}_{\mathcal{GB}}[\log(1 - G(x))] &= \zeta(b,a), \\
\mathrm{E}_{\mathcal{GB}}[\log g(x)] &= \int_0^1 [\log c + (c-1)\log x]\, f_{\mathcal{GB}}(x)dx \\
&= \log c - c^{-1}(c-1)\zeta(a,b),
\end{aligned}$$

where $\zeta(a, b) = \psi(a + b) - \psi(a)$ and $\psi(\cdot)$ represents the digamma function. Thus, the entropy of a BG distribution is:

$$\begin{aligned}
-\mathrm{E}_{\mathcal{GB}}[\log f_{\mathcal{GB}}(x)] &= \log B(a,b) - \log c + c^{-1}(c-1)\zeta(a,b) \\
&\quad +(a-1)\,\zeta(a,b) + (b-1)\,\zeta(b,a).
\end{aligned} \tag{13}$$

Setting a, b and c in (13) equal to certain values gives the entropy of sub-model distributions. For instance, for $c = 1$ we obtain the entropy of the classical beta distribution (Nadarajah and Zografos, 2003) and for $a = 1$ we obtain the Kumaraswamy entropy:

$$-\mathrm{E}_{Kw}[\log f_{Kw}(x)] = -\log(c\,b) + c^{-1}(c-1)\zeta(1,b) + (b-1)\zeta(b,1), \tag{14}$$

which follows on noting that $B(1,b)=b^{-1}$.

Appendix A proves that the GBG density (4) with parent distribution F statisfies

$$-\mathrm{E}_{\mathcal{GBG}}[\log F(X)^c] = \zeta(a,b), \quad (15)$$
$$-\mathrm{E}_{\mathcal{GBG}}[1-\log F(X)^c] = \zeta(b,a), \quad (16)$$
$$\mathrm{E}_{\mathcal{GBG}}[\log f(X)] - \mathrm{E}_U[\log f(F^{-1}(U^{1/c}))] = c^{-1}(c-1)\zeta(a,b), \quad (17)$$

where $f(x)=F'(x)$ and $U\sim\mathcal{B}(a,b)$. Furthermore, the GBG has the maximum entropy of all distributions satisfying the information constraints (15) – (17) and its entropy is:

$$\begin{aligned}-\mathrm{E}_{\mathcal{GBG}}[\log f_{\mathcal{GBG}}(x)] &= \log B(a,b) - \log c + c^{-1}(c-1)\zeta(a,b)\\ &\quad +(a-1)\zeta(a,b)+(b-1)\zeta(b,a) - \mathrm{E}_U[\log f(F^{-1}(U^{1/c}))]. \quad (18)\end{aligned}$$

Thus, the GBG entropy is the sum of the entropy of the generalized beta generator (13), which is independent of the parent, and another term $-\mathrm{E}_U[\log f(F^{-1}(U^{1/c}))]$ that is related to the entropy of the parent. Furthermore, the constraints (15) and (16) reflect information only about the generalized beta generator. Note that (13) takes its maximum value of zero when $a=b=c=1$; otherwise, the structure in the generator adds information to the GBG distribution. From

$$\mathrm{E}_{\mathcal{GBG}}[\log F(X)^c] = \mathrm{E}_{\mathcal{BG}}[\log G(x)] = \mathrm{E}_{\mathcal{B}}[\log U],$$

$$\mathrm{E}_{\mathcal{GBG}}[\log F(X)] = \mathrm{E}_{\mathcal{BG}}[\log(1-G(x))] = \mathrm{E}_{\mathcal{B}}[\log(1-U)],$$

it follows that (15) is related to the information in the left tail and (16) is related to the information in the right tail, and this information is the same when $a=b$. Clearly, the parameters a and b influence skewness only though taking different values, i.e. having differential weights in the two tails. By contrast, the additional parameter c influences skewness through (17) by adding entropy to the centre of the distribution.

4 Special cases of GBG distributions

This section introduces some of the many distributions which can arise as special sub-models within the GBG class of distributions. We have considered eleven different parents: normal, log-normal, skewed Student-t, Laplace, exponential, Weibull, Gumbel, Birnbaum-Saunders, gamma, Pareto and logistic distributions. In each case the distribution of the parent and the density of the corresponding GBG distribution is stated in Table 1. The GBG distributions can be applied to the same areas as their corresponding parent distributions, to offer an improve fit to the data.

Numerous distributions that have previously been studied are special cases of these GBG distributions. For instance, consider the generalized beta Ssewed-t distribution which is obtained when F is a scaled Student-t distribution on two degrees of freedom with scale factor $\sqrt{\lambda/2}$. When $c=1$ we have the skewed $t(a,b)$ distribution proposed by Jones and Faddy (2003) and when in addition $a=b$ we obtain a Student-t distribution with $2p$ degrees of freedom. Even some distributions that have only been introduced very recently are encompassed by the GBG class. For instance, the Beta-Birnbaum Saunders distribution introduced by Cordeiro and Lemonte (2011) is a special case of the generalized beta Birnbaum-Saunders distribution, which has the classical Birnbaum and Saunders (1969) distribution as parent.

For brevity, in the remainder of this section we shall only comment in detail on the properties of two of the most important GBG distributions: the generalized beta normal (GBN) and generalized beta Weibull (GBW) distributions. The GBN density is obtained from (4) by taking $F(\cdot)$ and $f(\cdot)$ to be the $N(\mu,\sigma^2)$ distribution and density functions, so that

$$f_{\mathcal{GBN}}(x;a,b,c) = \frac{c}{B(a,b)\sigma}\,\phi\left(\sigma^{-1}(x-\mu)\right)\left\{\Phi\left(\sigma^{-1}(x-\mu)\right)\right\}^{ac-1}\left\{1-\Phi\left(\sigma^{-1}(x-\mu)\right)^c\right\}^{b-1},$$

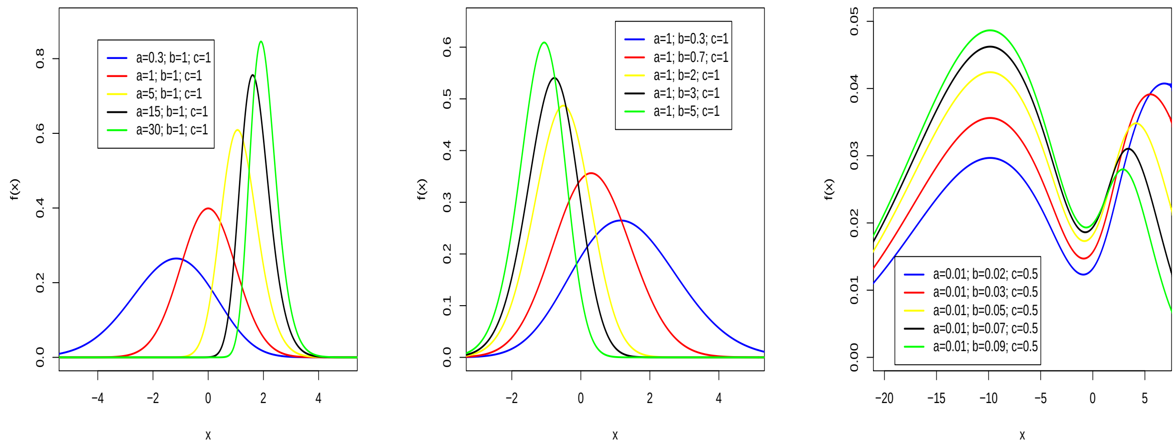

Figure 5: The GBN density for some parameter values and $\mu = 0$ and $\sigma = 1$.

where $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ is a location parameter and $\sigma > 0$ is the scale parameter. Plots of the GBN density for selected parameter values are given in Figure 5.

If we set $c = 1$ in the GBN density we obtain the standard beta normal distribution proposed by Eugene *et al.* (2002). The GBN distribution with $\mu = 0$, $\sigma = 1$, $b = 1$ and $ac = 2$ coincides with the skew normal distribution with shape parameter equal to one (Azzalini, 1985). Using the methodology proposed by Nadarajah (2008) the moments of the GBN distribution may be expressed as a finite sum of the Lauricella functions of type A (Exton, 1978), when a, b and c are integers. This also applies to the moments of the GBlogN distribution when a, b and c are integers.

Taking F to be a Weibull distribution with scale parameter $\lambda > 0$ and shape parameter $\gamma > 0$ we obtain the GBW distribution, with density function as stated in Table 1. We focus on this distribution because it extends many distributions previously considered in the literature. The generalized beta exponential (GBE) distribution corresponds to the choice $\gamma = 1$. Other sub-models, apart from the Weibull distribution itself, include: the beta Weibull (BW) (Lee *et al.*, 2007) and Kumaraswamy Weibull (KwW) (Cordeiro *et al.*, 2010) distributions for $c = 1$ and $a = 1$, respectively; the exponentiated Weibull (EW) (Mudholkar *et al.*, 1995; Mudholkar *et al.*, 1996; Nassar and Eissa, 2003; Nadarajah and Gupta, 2005; and Choudhury, 2005) and exponentiated exponential (EE) (Gupta and Kundu, 2001) distributions for $b = c = 1$ and $b = c = \lambda = 1$, respectively; and the beta exponential (BE) (Nadarajah and Kotz, 2006) distribution for $b = c = \gamma = 1$.

The distribution and hazard rate functions corresponding to the GBW distribution are

$$F_{\mathcal{GBW}}(x) = I\left(\{1-\exp[-(\lambda x)^{\gamma}]\}^{c}; a, b\right) = \frac{1}{B(a,b)} \int_{0}^{\{1-\exp[-(\lambda x)^{\gamma}]\}^{c}} \omega^{a-1}\,(1-\omega)^{b-1} d\omega$$

and

$$h_{\mathcal{GBW}}(x) = \frac{c\,\gamma\,\lambda^{\gamma}\,x^{\gamma-1}\,\exp\{-(\lambda\,x)^{\gamma}\}\,\{1-\exp[-(\lambda x)^{\gamma}]\}^{ac-1}\,\{1-[1-\exp\{-(\lambda x)^{\gamma}\}]^{c}\}^{b-1}}{B(a,b)\left\{1-I\left(\{1-\exp[-(\lambda x)^{\gamma}]\}^{c}; a, b\right)\right\}}, \quad (19)$$

respectively. A characteristic of the GBW distribution is that its hazard rate function can be bathtub shaped, monotonically increasing or decreasing and upside-down bathtub depending basically on the parameter values. Some possible shapes for the GMW density function for selected parameter values, including those of some well-known distributions, are illustrated in Figure 6.

Distribution	Distribution of the Parent	Density of the Generalized Beta Distribution
Normal	$F(x) = \Phi(x)$	$f(x;a,b,c) = \frac{c\,\phi(x)\,\Phi(x)^{ac-1}\,[1-\Phi(x)^c]^{b-1}}{B(a,b)}$
Log-Normal	$F(x) = \Phi(\log x)$	$f(x;a,b,c) = \frac{c\phi(\log x)\Phi(\log x)^{ac-1}[1-\Phi(\log x)^c]^{b-1}}{xB(a,b)}$
Student-t	$F(x) = \frac{1}{2}\left(1+\frac{x}{\sqrt{\lambda+x^2}}\right)$	$f(x;a,b,c) = \frac{\lambda c}{2^{ac}B(a,b)}\frac{1}{(\lambda+x^2)^{3/2}}\left(1+\frac{x}{\sqrt{\lambda+x^2}}\right)^{ac-1}\left[1-\frac{1}{2^c}\left(1+\frac{x}{\sqrt{\lambda+x^2}}\right)^c\right]^{b-1}$
Laplace	$F(x) = \begin{cases} \frac{1}{2}\exp(x/\lambda), & x<0, \\ 1-\frac{1}{2}\exp(-x/\lambda), & x>0 \end{cases}$	$f(x;a,b,c) = \frac{cf(x)[F(x)]^{ac-1}[1-F(x)^c]^{b-1}}{B(a,b)}$
Exponential	$F(x) = 1-\exp(-\lambda x), \lambda>0$	$f(x;a,b,c) = \frac{c\lambda e^{-\lambda x}[1-e^{-\lambda x}]^{ac-1}[1-(1-e^{-\lambda x})^c]^{b-1}}{B(a,b)}$
Weibull	$F(x) = 1-\exp(-(\lambda x)^\gamma), \gamma>0$	$f(x;a,b,c) = \frac{c\gamma\lambda^\gamma e^{-(\lambda x)^\gamma}[1-e^{-(\lambda x)^\gamma}]^{ac-1}[1-(1-e^{-(\lambda x)^\gamma})^c]^{b-1}}{B(a,b)}$
Gumbel	$F(x) = 1-\exp(-e^x)$	$f(x;a,b,c) = \frac{c\exp(x-e^x)[1-\exp(-e^x)]^{ac-1}[1-(1-\exp(-e^x))^c]^{b-1}}{B(a,b)}$

Distribution	Distribution of the Parent	Density of the Generalized Beta Distribution
Birnbaum-Saunders	$F(x)=\Phi\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}}-\sqrt{\frac{\beta}{t}}\right)\right],\ \alpha,\beta>0$	$f(x;a,b,c)=\frac{cf(x)[F(x)]^{ac-1}[1-F(x)^c]^{b-1}}{B(a,b)}$
Gamma	$F(x)=\gamma_1(\alpha,\beta x)=\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)},\ \alpha,\beta>0$	$f(x;a,b,c)=\frac{c\,\beta^\alpha x^{\alpha-1}e^{-\beta x}\gamma_1(\alpha,\beta x)^{ac-1}\,[1-\gamma_1(\alpha,\beta x)^c]^{b-1}}{B(a,b)\Gamma(\alpha)}$
Pareto	$F(x)=1-\frac{1}{(1+x)^\nu},\ \nu>0$	$f(x;a,b,c)=\frac{c\nu\left[1-\frac{1}{(1+x)^\nu}\right]^{ac-1}\left[1-\left(1-\frac{1}{(1+x)^\nu}\right)^c\right]^{b-1}}{(1+x)^{\nu+1}B(a,b)}$
Logistic	$F(x)=\frac{1}{1+e^{-x}}$	$f(x;a,b,c)=\frac{ce^{-x}\left[1-(1+e^{-x})^{-c}\right]^{b-1}}{(1+e^{-x})^{ac+1}B(a,b)}$

Table 1: Special GBG Distributions. $\Phi(x)$ and $\phi(x)$ denote the standard normal distribution and density functions. In the Student-t distribution, $\lambda=a+b$. $\gamma(\alpha,z)=\int_0^z t^{\alpha-1}\,e^{-t}dt$ is the incomplete gamma function.

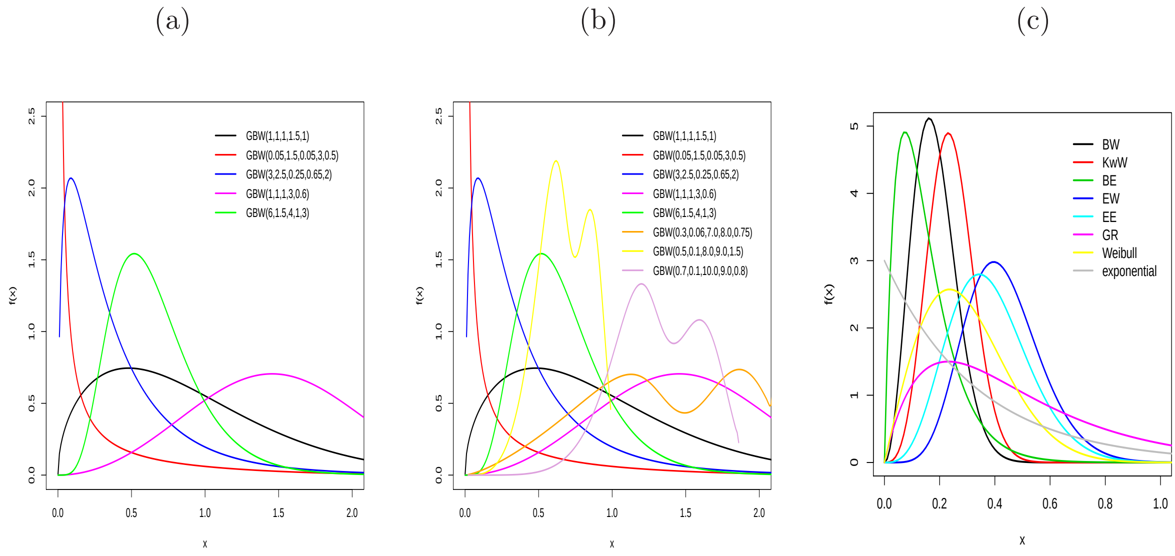

Figure 6: The GBW density: (a) and (b) for some parameter values; (c) for some sub-models. The GBW density function (given in Table 1) is denoted by GBW$(a, b, c, \lambda, \gamma)$.

5 Theoretical properties

In this section, we derive the moments, the moment generating function (mgf) and the quantile function of GBG distributions. General formulae for their deviations and reliability have also been derived but these are not included for brevity. They are available from the authors on request.

5.1 Moments and generating function

We derive several representations for the sth moment $\mu'_s = E(X^s)$ and mgf $M(t) = E[\exp(t\,X)]$ of $X \sim \mathcal{GBG}(F; \boldsymbol{\tau}, a, b, c)$. Note that other kinds of moments related to the L-moments of Hosking (1990) may also be obtained in closed form, but we confine ourselves here to μ'_s for brevity.

By definition, when $X \sim \mathcal{GBG}(F; \boldsymbol{\tau}, a, b, c)$, $\mu'_s = E\{[F^{-1}(U^{1/c})]^s\}$, where $U \sim \mathcal{B}(a, b)$. Then, we obtain

$$\mu'_s = c\,B(a,b)^{-1} \int_0^1 [F^{-1}(x)]^s\, x^{ac-1}(1-x^c)^{b-1} dx. \tag{20}$$

Our first representation is an approximation to μ'_s obtained upon expanding $F^{-1}(x)$ in (20) using a Taylor series around the point $E(X_F) = \mu_F$:

$$\mu'_s \approx \sum_{k=0}^{s} \binom{s}{k} [F^{-1}(\mu_F)]^{s-k}\, [F^{-1(1)}(\mu_F)]^k \sum_{i=0}^{k} (-1)^i \binom{k}{i} \mu_F^i \frac{B\left(a + \frac{k-i}{c}, b\right)}{B(a,b)},$$

where $F^{-1(1)}(x) = \frac{d}{dx} F^{-1}(x) = [f(F^{-1}(x))]^{-1}$.

An alternative series expansion for μ'_s is given in terms of $\tau(s,r) = E\{Y^s F(Y)^r\}$, where Y follows the parent distribution F, for $s, r = 0, 1, \ldots$ The probability weighted moments $\tau(s,k)$ can be derived for most parent distributions. When both a and c are integers, it follows immediately from equation (8) that

$$\mu'_s = c\,B(a, b+1)^{-1} \sum_{i=0}^{\infty} (-1)^i \binom{b}{i} \tau(s, a\,(i+c) - 1).$$

Otherwise, if at least one of these parameters is a real non-integer, we can use the following expansion for $F(x)^\alpha$ for real non-integer α:

$$F(x)^\alpha = \sum_{k=0}^{\infty} t_k(\alpha)\, F(x)^k, \tag{21}$$

where

$$t_k(\alpha) = \sum_{j=k}^{\infty} (-1)^{k+j} \binom{\alpha}{j} \binom{j}{k}.$$

Substituting (21) in equation (8) yields:

$$\mu'_s = \sum_{k=0}^{\infty} v_k\, \tau(s,k), \tag{22}$$

where

$$v_k = c\, B(a, b+1)^{-1} \sum_{i=0}^{\infty}\sum_{j=k}^{\infty} (-1)^{k+i+j} \binom{a\,(i+c)-1}{j} \binom{j}{k} \binom{b}{i}. \tag{23}$$

A third representation for μ'_s is also derived from (8) as

$$E(X^s) = c\, B(a, b+1)^{-1} \sum_{i=0}^{\infty} (-1)^i \binom{b}{i}\, \tau(s, c\,(i+a)-1), \tag{24}$$

where, setting $Q_F(u) = F^{-1}(u)$:

$$\tau(s,r) = \int_{-\infty}^{\infty} x^s\, F(x)^r\, f(x) dx = \int_0^1 Q_F(u)^s\, u^r du. \tag{25}$$

Similarly, the mgf of X may be written:

$$M(t) = c\, B(a, b+1)^{-1} \sum_{i=0}^{\infty} (-1)^i \binom{b}{i}\, \rho(t, a\,(i+c)-1), \tag{26}$$

where

$$\rho(t,r) = \int_{-\infty}^{\infty} \exp(t\,x)\, F(x)^r\, f(x) dx = \int_0^1 \exp\{t\, Q_F(u)\}\, u^r du. \tag{27}$$

If b is an integer, the index i in (26) stops at $b-1$. An application of these expansions to three specific GBG distributions is given in Table 2.

A fourth representation for μ'_s can be obtained from (9) as

$$\mu'_s = \sum_{k=0}^{\infty} w_k\, E(Y_k^s), \qquad Y_k \sim \mathrm{Exp}^{c(k+a)} F. \tag{28}$$

Similarly, equation (9) yields an alternative expression for the mgf of a GBG distribution:

$$M(t) = \sum_{k=0}^{\infty} w_k\, M_k(t),$$

where $M_k(t)$ is the mgf of Y_k. Again, if b is an integer, the index k in the previous sum stops at $b-1$. This way, GBG moments and generating functions may be inferred from moments and generating functions of exponentiated F distributions, such as those derived by Nadarajah and

Kotz (2006). For example:

(i) If Y has the distribution $F(x) = (1 - e^{-x})^\alpha$ of the exponentiated unit exponential with parameter $\alpha > 0$, its mgf is $M_Y(t) = \Gamma(\alpha+1)\,\Gamma(1-t)/\Gamma(\alpha - t + 1)$ for $-1 < t < 1$. Then, the mgf of the GB-exponential with unit parameter is

$$M(t) = \sum_{i=0}^{\infty} w_i \, \frac{\Gamma(a\,(i+c)+1)\,\Gamma(1-t)}{\Gamma(a\,(i+c)-t+1)}, \qquad -1 < t < 1.$$

(ii) For the GBW distribution introduced in Section 4, Y has the Exp-Weibull(a) distribution whose moments are known:

$$E(Y^s) = \frac{a}{\lambda^s}\Gamma\left(\frac{s}{\gamma}+1\right)\sum_{i=0}^{\infty}\frac{(1-a)_i}{i!\,(i+1)^{(s+\gamma)/\gamma}},$$

where $(a)_i = a\,(a+1)\,\ldots\,(a+i-1)$ denotes the ascending factorial (also called the Pochhammer function). From this expectation and (28), the sth moment of the GBW distribution reduces to

$$\mu_s' = \lambda^{-s}\Gamma\left(\frac{s}{\gamma}+1\right)\sum_{k,i=0}^{\infty}\frac{w_k\, c\,(k+a)\,[1-c(k+a)_i]}{i!\,(i+1)^{(s+\gamma)/\gamma}}.$$

(iii) For the generalized beta Gumbel (GBGu) distribution, say $F(x) = 1 - \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}$, the moments of Y having the exponentiated-Gumbel(a) can be obtained from Nadarajah and Kotz (2006) as

$$E(Y^s) = a\sum_{i=0}^{s}\binom{s}{i}\mu^{s-i}\,(-\sigma)^i\left(\frac{\partial}{\partial p}\right)^i\left[a^{-p}\,\Gamma(p)\right]\Bigg|_{p=1}.$$

From the last equation and (28), the sth moment of the GBGu distribution becomes

$$\mu_s' = c\sum_{k=0}^{\infty} w_k\,(k+a)\sum_{i=0}^{s}\binom{s}{i}\mu^{s-i}\,(-c\,\sigma)^i\left(\frac{\partial}{\partial p}\right)^i\left[(k+a)^{-p}\,\Gamma(p)\right]\Bigg|_{p=1}.$$

Finally, we mention that another alternative for the mgf of X is evidently

$$M(t) = \sum_{s=0}^{\infty}\frac{\mu_s'}{s!}\,t^s,$$

where μ_s' is calculated (for both a and c integers and for at least one of the parameters real non-integers) from any of the moment expressions derived before.

Distribution	GB-exponential	GB-Pareto	GB-standard logistic
Parent	$1-\exp(-\lambda x), \lambda>0$	$1-(1+x)^{-\nu}, \nu>0$	$[1+\exp(-x)]^{-1}$
$E(X^s)$	$s!\,\lambda^s \sum_{k,j=0}^{\infty} v_k\,(-1)^{s+j}\binom{c\,(k+a)-1}{j}(j+1)^{-(s+1)}$	$\sum_{k,j=0}^{\infty} v_k\,(-1)^{s+j}\binom{s}{j} B\left(c\,(k+a), 1-\frac{j}{\nu}\right)$	$\sum_{k=0}^{\infty} v_k\,\left(\frac{\partial}{\partial t}\right)^s B(t+c\,(k+a), 1-t)\Big\vert_{t=0}$
M(t)	$B(a,b+1)^{-1}c\sum_{i=0}^{\infty}(-1)^i\binom{b}{i} B(a\,(i+c), 1-\lambda t)$	$\exp(-t)\sum_{i,r=0}^{\infty} B(a\,(i+c), 1-r\nu^{-1})\frac{w_i\,t^r}{r!}$	$\sum_{i=0}^{\infty} w_i\,B(t+a\,(i+c), 1-t)$

Table 2: Moments and generating functions of some specific GBG distributions. The parent distribution is given in the first row, the second row applies equations (24) and (25) to calculate the moments and the third row uses (26) and (27) to find the generating function. The coefficients w_i and v_k are defined in (11) and (23), respectively.

5.2 Quantile function

The GBG quantile function is obtained by inverting the parent distribution F. We have

$$Q_{\mathcal{GBG}}(u;\boldsymbol{\tau},a,b,c) = F^{-1}\left(\left[Q_{\beta(a,b)}(u)\right]^{1/c}\right), \tag{29}$$

where $Q_{\beta(a,b)}(u) = I^{-1}(u;a,b)$ is the ordinary beta quantile function. It is possible to obtain some expansions for the beta quantile function with positive parameters a and b. One of them can be found on the Wolfram website (http://functions.wolfram.com/06.23.06.0004.01) as

$$z = Q_{\beta(a,b)}(u) = a_1 v + a_2 v^2 + a_3 v^3 + a_4 v^4 + O(v^{5/a}),$$

where $v = [a\,B(a,b)\,u]^{1/a}$ for $a > 0$ and $a_0 = 0$, $a_1 = 1$, $a_2 = (b-1)/(a+1)$,

$$a_3 = (b-1)[a^2 + (3b-1)a + 5b - 4]/[2(a+1)^2(a+2)],$$

$$\begin{aligned} a_4 &= (b-1)[a^4 + (6b-1)a^3 + (b+2)(8b-5)a^2 + (33b^2 - 30b + 4)a \\ &+ b(31b-47) + 18]/[3(a+1)^3(a+2)(a+3)], \ldots \end{aligned}$$

The coefficients a_i for $i \geq 2$ can be derived from a cubic recursion of the form

$$\begin{aligned} a_i &= \frac{1}{[i^2 + (a-2)i - (a-1)]}\Big\{(1-\delta_{i,2})\sum_{r=2}^{i-1} a_r\, a_{i+1-r}\,\big[r(1-a)(i-r) \\ &- r(r-1)\big] + \sum_{r=1}^{i-1}\sum_{s=1}^{i-r} a_r\, a_s\, a_{i+1-r-s}\,\big[r(r-a) + s(a+b-2)(i+1-r-s)\big]\Big\}, \end{aligned}$$

where $\delta_{i,2} = 1$ if $i = 2$ and $\delta_{i,2} = 0$ if $i \neq 2$. In the last equation, we note that the quadratic term only contributes for $i \geq 3$.

6 Estimation

Here we derive the ML estimators of the parameters of the GBG family. Let $x_1, \ldots, x_n$ be a random sample of size n on $X \sim \mathcal{GBG}(F;\boldsymbol{\tau},a,b,c)$, where $\boldsymbol{\tau}$ is a $p \times 1$ vector of unknown parameters in the parent distribution $F(x;\boldsymbol{\tau})$. The log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\tau},a,b,c)$ may be written:

$$\begin{aligned} l(\boldsymbol{\theta}) &= n\log(c) - n\log[B(a,b)] + \sum_{i=1}^{n}\log[f(x_i;\boldsymbol{\tau})] + (ac-1)\sum_{i=1}^{n}\log[F(x_i;\boldsymbol{\tau})] \\ &\quad + (b-1)\sum_{i=1}^{n}\log[1 - F^c(x_i;\boldsymbol{\tau})]. \end{aligned} \tag{30}$$

This may be maximized either directly, e.g. using SAS (Proc NLMixed) or the Ox (sub-routine MaxBFGS) (Doornik, 2007), or by solving the nonlinear likelihood equations obtained by differentiating (30).

Initial estimates of the parameters a and b may be inferred from estimates of c and $\boldsymbol{\tau}$. To see why, note that $X \sim \mathcal{GBG}(F;\boldsymbol{\tau},a,b,c)$ implies $F(X;\boldsymbol{\tau}) \sim \mathcal{GB}(a,b,c)$. Then, knowing the moments of a GBG distribution, we have

$$E[F(X)^s] = \frac{B\left(a + \frac{s}{c}, b\right)}{B(a,b)}, \tag{31}$$

$$E\{[1 - F(X)^c]^r\} = \frac{B(a, b+r)}{B(a,b)}, \tag{32}$$

$$E\{F(X)^s[1 - F(X)^c]^r\} = \frac{B\left(a + \frac{s}{c}, b+r\right)}{B(a,b)}. \tag{33}$$

Now considering (31) with $s = c$ and (33) with $s = c$ and $r = 1$ we have:

$$\begin{aligned} E[F(X)^c] &= \frac{a}{a+b}, \\ E\{F(X)^c[1 - F(X)^c]\} &= \frac{ab}{(a+b)(a+b+1)}. \end{aligned}$$

Solving the last two equations for a and b yields

$$a = \frac{uv}{u(1-u)-v} \qquad \text{and} \qquad b = \frac{v(1-u)}{u(1-u)-v}, \tag{34}$$

where $u = E[F(X)^c]$ and $v = E\{F(X)^c[1 - F(X)^c]\}$.

The components of the score vector $U(\boldsymbol{\theta})$ are given by

$$\begin{aligned} U_a(\boldsymbol{\theta}) &= -n\big[\psi(a) - \psi(a+b)\big] + c\sum_{i=1}^{n} \log[F(x_i;\boldsymbol{\tau})], \\ U_b(\boldsymbol{\theta}) &= -n\big[\psi(b) - \psi(a+b)\big] + \sum_{i=1}^{n} \log[1 - F^c(x_i;\boldsymbol{\tau})], \\ U_c(\boldsymbol{\theta}) &= \frac{n}{c} + \sum_{i=1}^{n} \log[f(x_i;\boldsymbol{\tau})] - (b-1)\sum_{i=1}^{n} \frac{F^c(x_i;\boldsymbol{\tau})\log[F(x_i;\boldsymbol{\tau})]}{[1 - F^c(x_i;\boldsymbol{\tau})]}, \\ U_{\boldsymbol{\tau}}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \frac{\dot{f}(x_i)_{\boldsymbol{\tau}}}{f(x_i;\boldsymbol{\tau})} + (ac-1)\sum_{i=1}^{n} \frac{\dot{F}(x_i)_{\boldsymbol{\tau}}}{F(x_i;\boldsymbol{\tau})} + c(b-1)\sum_{i=1}^{n} \frac{F^{c-1}(x_i;\boldsymbol{\tau})\dot{F}(x_i)_{\boldsymbol{\tau}}}{[1 - F^c(x_i;\boldsymbol{\tau})]}, \end{aligned}$$

where $\dot{f}(x_i)_{\boldsymbol{\tau}} = \partial f(x_i;\boldsymbol{\tau})/\partial\boldsymbol{\tau}$ and $\dot{F}(x_i)_{\boldsymbol{\tau}} = \partial F(x_i;\boldsymbol{\tau})/\partial\boldsymbol{\tau}$ are $p \times 1$ vectors.

For interval estimation and hypothesis tests on the model parameters, we require the observed information matrix. The $(p+3)\times(p+3)$ unit observed information matrix $J = J(\boldsymbol{\theta})$ is given in Appendix B. Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal, $\mathrm{N}_{p+3}(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. We can substitute $I(\boldsymbol{\theta})$ by $J(\widehat{\boldsymbol{\theta}})$, i.e., the observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$. The multivariate normal $\mathrm{N}_{p+3}(0, J(\widehat{\boldsymbol{\theta}})^{-1})$ distribution can be used to construct approximate confidence intervals for the individual parameters and for the hazard rate and survival functions.

To construct the likelihood ratio (LR) statistics for testing some sub-models of the GBG distribution we can compute the maximum values of the unrestricted and restricted log-likelihoods. For example, we may use the LR statistic to check if the fit using a GBG sub-model is statistically superior to a fit using the BG distribution for a given data set. In any case, hypothesis tests of the type $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, can be performed using LR statistics. For example, a comparison of the GBG and BG distributions is equivalent to testing $H_0 : c = 1$ versus $H_1 : c \neq 1$. The LR statistic reduces to

$$w = 2\,\{\ell(\widehat{a}, \widehat{b}, \widehat{c}, \widehat{\lambda}, \widehat{\gamma}) - \ell(\widetilde{a}, \widetilde{b}, 1, \widetilde{\lambda}, \widetilde{\gamma})\}, \tag{35}$$

where $\widehat{a}$, $\widehat{b}$, $\widehat{c}$, $\widehat{\lambda}$, and $\widehat{\gamma}$ are the ML estimates under H_1 and $\widetilde{a}$, $\widetilde{b}$, $\widetilde{\lambda}$ and $\widetilde{\gamma}$ are the estimates under H_0.

7 A simulation study

Here we assess the finite sample behavior of the ML estimators of a, b, c, μ and σ in a GBN model using a simulation study. We perform 5000 Monte Carlo replications where the simulations are carried out using the MaxBFGS subroutine in the matrix programming language available in Doornik (2007). In each replication a random sample of size n is drawn from the

GBN(a, b, c, μ, σ) distribution and the parameters are re-estimated from the sample by maximum likelihood. The GBN random number generation was performed using the inversion method. The true parameter values used in the data generating process are $a = 1.5$, $b = 0.5$, $c = 1.8$, $\mu = 0.5$ and $\sigma = 1$.

Table 3: Monte Carlo simulation results: mean estimates and RMSEs of $\widehat{a}$, $\widehat{b}$, $\widehat{c}$, $\widehat{\mu}$ and $\widehat{\sigma}$ for the GBN model.

n	Parameter	Mean	RMSE
100	a	1.4531	1.3699
	b	1.0597	1.6082
	c	2.1398	1.7623
	μ	0.5792	0.9904
	σ	1.0865	0.6497
300	a	1.4106	1.2719
	b	0.6696	0.5698
	c	2.2425	1.6291
	μ	0.4780	0.8240
	σ	0.9810	0.4287
500	a	1.4696	1.0787
	b	0.5799	0.3311
	c	2.2004	1.2339
	μ	0.5096	0.6965
	σ	0.9297	0.3238

Table 3 lists the means of the ML estimates of the model parameters and the corresponding root mean squared errors (RMSEs) for sample sizes $n = 100$, $n = 300$ and $n = 500$. These indicate that the biases of the ML estimators and RMSEs of the ML estimates of the model parameters decay toward zero as the sample size increases, as expected. Note that there is a small sample bias in the estimation of the GBN parameters. Further research could be conducted to obtain bias corrections for these estimators, thus reducing their systematic errors in finite samples.

8 Empirical applications

In this section we first illustrate the entropy decomposition derived in Section 3.2 by fitting a GBG distribution to some financial data. Then we use three other real data sets to compare the fits of a GBG distribution with those of three sub-models, i.e., the BG, KwG and the parent distribution itself.

In each case, the parameters are estimated via maximum likelihood, as described in Section 6, using the subroutine NLMixed in SAS. First, we describe the data sets. Then we report the ML estimates (and the corresponding standard errors in parentheses) of the parameters and the values of the Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Bayesian Information Criterion (BIC) statistics. The lower the value of these criteria, the better the fit. Note that over-parameterization is penalized in these criteria so that the additional parameter in the GBG distribution does not necessarily lead to a lower value of the AIC, BIC or CAIC statistics. Next, we perform the LR tests, described in Section 6, for a formal test of the need for a third skewness parameter. Finally, we provide the histograms of the data sets to have a visual comparison of the fitted densities. The respiratory data are taken from a study by the University of São Paulo, ESALQ (Laboratory of Physiology and Post-Colheita Biochemistry) and these data cannot be published since we do not have permission from the owner. However, the other data sets and all the computational code may be obtained from the website http://www.lce.esalq.usp.br/edwin/ .

(i) Financial data
To illustrate the concepts introduced in Section 3.2 we consider two samples of daily returns on the S&P 500 index. Sample 1 (January 2004 - December 2006) and Sample 2 (January 2007 to December 2009) are chosen to represent two very different market regimes: in the first sample markets were relatively stable with a low volatility; the second sample encompasses the credit crunch and the banking crisis, when market risk factors were extremely volatile. These observations are verified by the sample statistics shown in the upper section of Table 4. Our purpose here is to compare $\hat{a}$, $\hat{b}$ and $\hat{c}$ and hence also the entropy decomposition (18) for the two samples. However, the size of the GBG parameters depends on the mean and variance, which are quite different in these two samples. So, we estimate the GBG parameters on standardized samples $Z = (X - \hat{\mu})/\hat{\sigma}$, where $\hat{\mu}$ and $\hat{\sigma}$ are the mean and standard deviation of each sample. The parameters a, b and c, estimated using the method outlined in Section 6, are reported in the middle section of Table 4. Note that afterwards, if required, we could translate the $\mathcal{GBN}(\hat{a}, \hat{b}, \hat{c}; 0, 1)$ distribution to $\mathcal{GBN}(\hat{a}, \hat{b}, \hat{c}; \hat{\mu}, \hat{\sigma}^2)$ using $X = Z\hat{\sigma} + \hat{\mu}$. The bottom section of the table reports the lower and upper tail components of the generalized beta entropy, i.e. $\zeta(\hat{a}, \hat{b})$ and $\zeta(\hat{b}, \hat{a})$ in (15) and (16), and the term $-\log \hat{c} + \hat{c}^{-1}(\hat{c} - 1)\zeta(\hat{a}, \hat{b})$ in (18) which is zero for the classical BG distribution but which more generally controls the information in the centre of a GBG distribution.

Table 4: S&P 500 daily returns: descriptive statistics, GBN parameter estimates and entropy decomposition.

	Sample 1 **Jan 2004 - Dec 2006**	**Sample 2** **Jan 2007 - Dec 2009**
Mean	0.00034	-0.00014
Std. Dev.	0.00659	0.01885
Skewness	0.00259	0.05921
Kurtosis	3.25954	9.35811
$\hat{a}$	0.33261	0.44417
$\hat{b}$	1.00898	1.05067
$\hat{c}$	3.02177	2.32358
$\zeta(\hat{a}, \hat{b})$	3.01637	2.30010
$\zeta(\hat{b}, \hat{a})$	0.43951	0.52847
$-\log \hat{c} + \hat{c}^{-1}(\hat{c} - 1)\zeta(\hat{a}, \hat{b})$	0.91232	0.46710

Since both samples have positive skewness, $\hat{a} < \hat{b}$ in both cases. The lower tail entropy $\zeta(\hat{a}, \hat{b})$ is greater in sample 1 and the upper tail entropy $\zeta(\hat{b}, \hat{a})$ is greater in sample 2. Thus, sample 1, where markets were stable and trending, has less uncertainty (i.e. more information) about positive returns than sample 2; but sample 2, which is during the banking crisis, has more information than sample 1 does about negative returns. Finally, considering the estimates of the parameter c which determines the entropy in the centre of the GBN distribution, it is greater for sample 1 than it is for sample 2. Moreover, the last row of the table indicates that there is greater entropy (i.e. less information) in the centre of sample 1 than there is in sample 2.

The empirical results for the other three data sets will be analyzed collectively.

(ii) Voltage data
This data set was previously studied by Meeker and Escobar (1998, p. 383). These data repre-

sent the times of failure and running times for a sample of devices from a field-tracking study of a larger system. At a certain point in time 30 units were installed in normal service conditions. Two causes of failure were observed for each unit that failed: the failure caused by an accumulation of randomly occurring damage from power-line voltage spikes during electric storms and failure caused by normal product wear. Following Meeker and Escobar, we assume the data are generated by a Weibull distribution, and now use this for the parent of our family of generated distributions.

(iii) Respiratory data
These data were taken from a study by the University of São Paulo, ESALQ (Laboratory of Physiology and Post-Colheita Biochemistry) which evaluated the effects of mechanical damage on banana fruits (genus Musa spp.), see Saavedra del Aguila *et al.* (2010) for more details. The major problem affecting bananas during and after harvest is the susceptibility of the mature fruit to physical damage caused during transport and marketing. The extent of the damage is measured by the respiratory rate. We use 630 data points on respiratory rate and assume a normal parent distribution.

(iv) Skew data
A random sample from a skew normal distribution with parameters $(0, 1, 5)$ is used to compare the fits of the GBN, BN, KwN and normal distributions. The particular sample we use was previously considered in Section 6.3 of Azzalini and Capitanio (2003) and in Sartori (2006), also available at http://azzalini.stat.unipd.it/SN/index.html. It is of interest because it leads to an infinite estimate for the skewness parameter λ in the skew normal model. Hence, we shall here use the normal parent for a comparison of the GBG fits; even though the fit of the parent is not expected to be good, what is of interest to us is the comparison between the fits of the GBN, BN and KwN distributions.

Table 5 gives a descriptive summary of each sample. The voltage data has negative skewness and kurtosis, reflecting the bathtub-shaped empirical density. The respiratory and skew data have positive skewness and kurtosis, larger values of these sample moments being apparent in the respiratory data.

We now compute the ML estimates and the AIC, BIC and CAIC statistics for each data set. For the voltage data we compare the GBW distribution, which has parameters a, b, c and $\boldsymbol{\tau}$, where $\boldsymbol{\tau} = (\lambda, \gamma)^T$, with the BW, KwW and Weibull distributions. The ML estimates of λ and γ for the Weibull distribution are taken as starting values for the numerical iterative procedure. For the other two data sets, the GBN distribution (with parameters a, b, c, μ and σ) was fitted and compared with the fitted BN, KwN and normal distributions. In this case, the estimates of μ and σ for the normal distribution are taken as starting values.

Table 5: Descriptive statistics.

Data	Mean	Median	Mode	Std. Dev.	Variance	Skewness	Kurtosis	Min.	Max.
Voltage	177.03	196.5	300	114.99	13223.2	-0.29	-1.61	2	300
Respiratory	34.31	19.85	13.46	27.77	771.29	1.73	3.78	10.71	209.63
Skew	0.88	0.70	-0.1	0.76	0.57	0.93	0.44	-0.1	3.04

The results are reported in Table 6. Note that the three information criteria agree on the model's ranking in every case. For the respiratory data and the skew data, the lowest values of the information criteria are obtained when fitting the GBN distribution. The second lowest values are for the BN distribution, the third lowest values are for the KwN distribution and the worst model is the normal. Clearly, the GBN model having three skewness parameters is preferred, for both the respiratory and the skew data. However, from the values of the information criteria when fitting the voltage data, we can infer that the best model is the BW distribution and the

Table 6: ML estimates and information criteria.

Voltage	a	b	c	λ	γ	AIC	BIC	CAIC
GBW	0.0727	0.0625	1.0613	0.0050	7.9634	349.9	356.9	352.4
	(0.0168)	(0.0137)	(0.0521)	(0.00006)	(0.2336)			
BW	0.0879	0.0987	1	0.0047	7.8425	348.3	353.9	349.9
	(0.0261)	(0.0707)	-	(0.0004)	(0.2461)			
KwW	1	0.2269	0.0484	0.0043	7.7495	352.7	358.3	354.3
	-	(0.0966)	(0.0236)	(0.0003)	(0.2387)			
Weibull	1	1	1	0.0053	1.2650	372.6	375.4	373.1
	-	-	-	(0.00079)	(0.2044)			
Respiratory	a	b	c	μ	σ	AIC	CAIC	BIC
GBN	10021.0	0.4681	4.6369	-186.04	47.9945	5638.3	5638.4	5660.5
	(8.8561)	(0.0305)	(0.6311)	(7.9203)	(1.7718)			
BN	50.9335	0.4135	1	-56.1790	32.2426	5709.9	5710.0	5727.7
	(2.5794)	(0.0296)	-	(2.1684)	(0.9699)			
KwN	1	0.4520	13.4721	-32.7704	29.4031	5775.1	5775.2	5792.9
	-	(0.0329)	(1.4283)	(2.5507)	(0.8140)			
Normal	1	1	1	34.3166	27.7500	5979.3	5979.4	5988.2
	-	-	-	(1.1056)	(0.7818)			
Skew	a	b	c	μ	σ	AIC	CAIC	BIC
GBN	0.0889	0.1186	15.9969	0.2813	0.3366	109.2	110.6	118.8
	(0.0351)	(0.0246)	(7.1536)	(0.0045)	(0.0022)			
BN	2.9920	0.1405	1	-0.4438	0.4244	112.2	113.1	119.9
	(1.0429)	(0.0223)	-	(0.1614)	(0.0216)			
KwN	1	0.1508	1.9342	-0.1948	0.4001	113.0	113.9	120.7
	-	(0.0243)	(0.8905)	(0.1657)	(0.0212)			
Normal	1	1	1	0.8849	0.7489	117.0	117.2	120.8
	-	-	-	(0.1059)	(0.0749)			

Table 7: LR tests.

Voltage	Hypotheses	Statistic w	p-value
GBW vs BW	$H_0: c=1$ vs $H_1: H_0$ is false	0.40	0.5271
GBW vs KwW	$H_0: a=1$ vs $H_1: H_0$ is false	4.80	0.0285
GBW vs Weibull	$H_0: a=b=c=1$ vs $H_1: H_0$ is false	28.70	<0.0001
Respiratory	Hypotheses	Statistic w	p-value
GBN vs BN	$H_0: c=1$ vs $H_1: H_0$ is false	73.6	<0.00001
GBN vs KwN	$H_0: a=1$ vs $H_1: H_0$ is false	138.8	<0.00001
GBN vs Normal	$H_0: a=b=c=1$ vs $H_1: H_0$ is false	347.0	<0.00001
Skew	Hypotheses	Statistic w	p-value
GBN vs BN	$H_0: c=1$ vs $H_1: H_0$ is false	5.0	0.0253
GBN vs KwN	$H_0: a=1$ vs $H_1: H_0$ is false	5.8	0.0160
GBN vs Normal	$H_0: a=b=c=1$ vs $H_1: H_0$ is false	13.8	0.0032

GBW is only the second best – but it is still better than the KwW distribution.

To further investigate the fit of the GBW distributions to the voltage data we applied two formal goodness-of-fit criteria, i.e. the Cramér-von Mises (W^*) and Anderson-Darling (A^*) statistics (Chen and Balakrishnan, 1995). The smaller the values of W^* and A^*, the better the fit to the sample data. We obtain $W^* = 0.1892$ and $A^* = 0.8572$ for the GBW distribution, $W^* = 0.1901$ and $A^* = 0.8492$ for the BW distribution, $W^* = 0.2328$ and $A^* = 1.1043$ for the KwW distribution and $W^* = 0.3853$ and $A^* = 1.8206$ for the Weibull distribution. Thus, according to the Cramér-von Mises statistic, the GBW distribution yields the best fit, but the Anderson-Darling test suggests that the BW gives the best fit.

A formal test of the need for the third skewness parameter in GBG distributions is based on the LR statistics described in Section 6. Applying these to our three data sets, the results are shown in Table 7. For the voltage data, the additional parameter of the GBW distribution may not, in fact, be necessary because the LR test provides no indications against the BW model when compared with the GBW model. However, for the respiratory and skew data, we reject the null hypotheses of all three LR tests in favor of the GBN distribution. The rejection is extremely highly significant for the respiratory data, and highly or very highly significant for the skew data. This gives clear evidence of the potential need for three skewness parameters when modeling real data.

More information is provided by a visual comparison of the histograms of the data sets with the fitted densities. In Figure 7, we plot the histogram of the voltage data and the estimated densities of the GBW, BW, KwW and Weibull distributions. In each case, both the GBW and BW distributions provide reasonable fits, but it is clear that the GBW distribution provides a more adequate fit to the histogram of the data and better captures its extreme bathtub shape. The plots of the fitted GBN, BN, KwN and normal density functions are given in Figure 8 for the respiratory data and Figure 9 for the skew data. In both cases, the new distribution appears to provide a closer fit to the histogram than the other three sub-models.

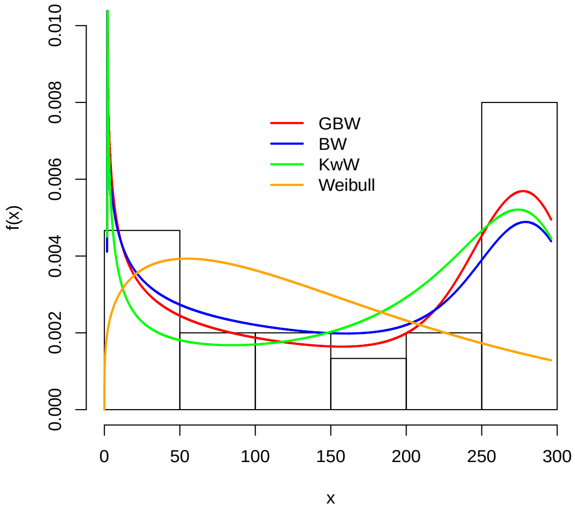

Figure 7: Estimated densities of the GBW, BW, KwW and Weibull models for the voltage data.

9 Conclusions

We propose a new class of generalized beta-generated (GBG) distributions which includes as special cases the beta-generated, Kumaraswamy-generated and exponentiated distributions. The GBG class extends several common distributions such as normal, log-normal, Laplace, Weibull, gamma and Gumbel distributions. Indeed, for any parent distribution F, we can define the corresponding generalized beta F distribution. The characteristics of GBG distributions, such as quantiles and moments and their generating functions, have tractable mathematical properties. The role of the generator parameters has been investigated and related to the skewness of the GBG distribution and the decomposition of its entropy. We have discussed maximum likelihood estimation and inference on the parameters based on likelihood ratio statistics for testing nested

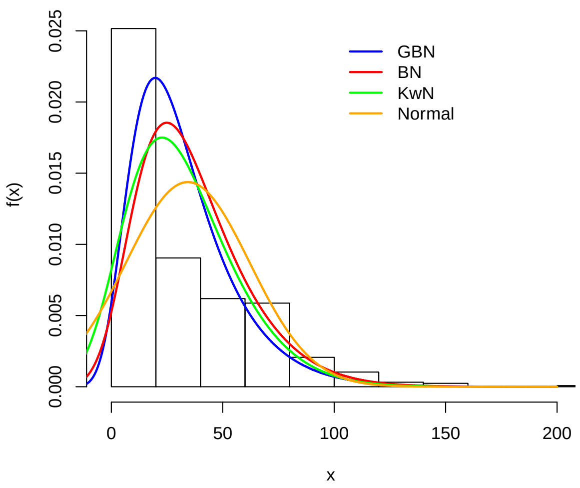

Figure 8: Fitted GBN, BN, KwN and normal densities for the respiratory rate data.

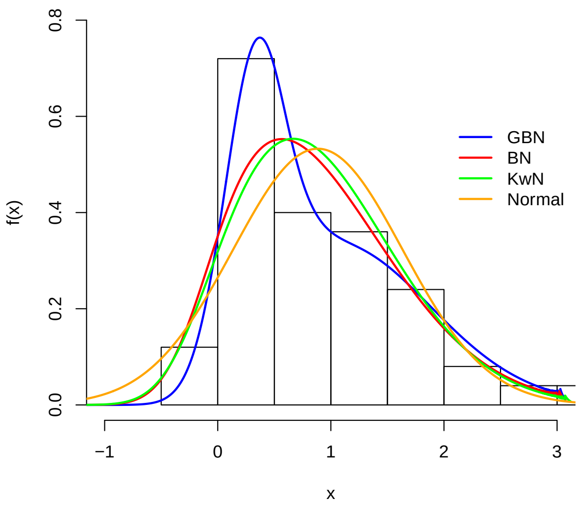

Figure 9: Fitted GBN, BN, KwN and normal densities for the skew data.

models. A simulation study is performed. Several applications to real data have shown the usefulness of the GBG class for applied statistical research. SAS code and data are available from the fourth author's homepage (http://www.lce.esalq.usp.br/edwin/).

Acknowledgments

Many thanks to Prof. M.C. Jones of the Open University, UK, for very insightful comments on the first draft of this paper. Also, José María Sarabia acknowledges support from the Ministerio de Educación of Spain (PR2009-0200, Programa Nacional de Movilidad de Recursos Humanos, and ECO2010-15455) and would like to thank the ICMA Centre for its hospitality. We gratefully acknowledge the grants from CNPq-Brazil. The authors would like to thank an associate editor and two referees for several useful comments that considerably improved the paper.

Appendix A: GBG Entropy

The conditions (15) and (16) follow immediately from the first two conditions in Lemma 1 of Zografos and Balakrishnan (2009), on noting that the GBG density is a standard beta density with parent $G(x) = x^c$. The third condition may be written:

$$\mathrm{E}_{\mathcal{BG}}[\log g(X)] = \mathrm{E}_U[\log g(G^{-1}(U))], \qquad U \sim \mathcal{B}(a,b), \tag{36}$$

where $g(x) = G'(x) = c\,F(x)^{c-1} f(x)$. Now $G^{-1}(U) = F^{-1}(U^{1/c})$ and

$$\log g(X) = \log(c) + (c-1)\log F(X) + \log f(X),$$

and since

$$\mathrm{E}_{\mathcal{B}}[\log U] = B(a,b)^{-1} \int_0^1 u^{a-1}\,(1-u^p)^{b-1}\,\log(u)du = -\zeta(a,b),$$

we have

$$\begin{aligned} \mathrm{E}_{\mathcal{B}}[\log g(G^{-1}(U))] &= \log(c) + (c-1)\mathrm{E}_{\mathcal{B}}[\log U^{1/c}] + \mathrm{E}_{\mathcal{B}}[\log f(F^{-1}(U^{1/c}))], \\ &= \log(c) - c^{-1}(c-1)\zeta(a,b) + \mathrm{E}_{\mathcal{B}}[\log f(F^{-1}(U^{1/c}))]. \end{aligned} \tag{37}$$

By Lemma 1 of Zografos and Balakrishnan (2009), $\mathrm{E}_{\mathcal{BG}}[\log F(X)] = -\zeta(a,b)$. Hence (36) becomes:

$$(c-1)\zeta(a,b) - \mathrm{E}_{\mathcal{GBG}}[\log f(X)] = c^{-1}(c-1)\zeta(a,b) - \mathrm{E}_{\mathcal{B}}[\log f(F^{-1}(U^{1/c}))],$$

which may be rewritten as (17). The entropy (18) follows from (37), and corollary 1 and proposition 1 of Zografos and Balakrishnan (2009).

Appendix B: Information Matrix

The elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters $(a, b, c, \boldsymbol{\tau})$ are:

$$J_{aa} = -n[\psi'(a) - \psi'(a+b)], \qquad J_{ab} = n\psi'(a+b), \qquad J_{ac} = \sum_{i=1}^n \log[F(x_i;\boldsymbol{\tau})],$$

$$J_{a\boldsymbol{\tau}} = c\sum_{i=1}^n \frac{\dot{F}(x_i)_{\boldsymbol{\tau}}}{F(x_i,\boldsymbol{\tau})},\; J_{bb} = -n[\psi'(b) - \psi'(a+b)],\; J_{bc} = -\sum_{i=1}^n \frac{F^c(x_i;\boldsymbol{\tau})\log[F(x_i;\boldsymbol{\tau})]}{[1-F^c(x_i;\boldsymbol{\tau})]},$$

$$J_{b\boldsymbol{\tau}} = -\sum_{i=1}^n \frac{cF^{c-1}(x_i;\boldsymbol{\tau})\dot{F}(x_i)_{\boldsymbol{\tau}}}{[1-F^c(x_i;\boldsymbol{\tau})]},\; J_{cc} = -\frac{n}{c^2} - (b-1)\sum_{i=1}^n \frac{F^c(x_i,\boldsymbol{\tau})\log^2[F(x_i;\boldsymbol{\tau})]}{[1-F^c(x_i;\boldsymbol{\tau})]^2},$$

$$\begin{aligned} J_{c\boldsymbol{\tau}} \;=\; & \sum_{i=1}^n \frac{\dot{f}(x_i)_{\boldsymbol{\tau}}}{f(x_i;\boldsymbol{\tau})} - (b-1)\sum_{i=1}^n \frac{F^{c-1}(x_i;\boldsymbol{\tau})\dot{F}(x_i)_{\boldsymbol{\tau}}}{[1-F^c(x_i;\boldsymbol{\tau})]^2}\Big\{\big[c\,\log[F(x_i;\boldsymbol{\tau})]+1\big]\big[1-F^c(x_i;\boldsymbol{\tau})\big] \\ & + c\,F^c(x_i;\boldsymbol{\tau})\log[F(x_i;\boldsymbol{\tau})]\Big\}, \end{aligned}$$

$$\begin{aligned} J_{\boldsymbol{\tau\tau}} \;=\; & \sum_{i=1}^n \frac{\ddot{f}(x_i)_{\boldsymbol{\tau\tau}} f(x_i;\boldsymbol{\tau}) - [\dot{f}(x_i)_{\boldsymbol{\tau}}]^2}{f^2(x_i;\boldsymbol{\tau})} + (ac-1)\sum_{i=1}^n \frac{\ddot{F}(x_i)_{\boldsymbol{\tau\tau}} F(x_i;\boldsymbol{\tau}) - [\dot{F}(x_i)_{\boldsymbol{\tau}}]^2}{F^2(x_i;\boldsymbol{\tau})} + \\ & (b-1)\sum_{i=1}^n \frac{cF^{c-1}(x_i;\boldsymbol{\tau})}{[1-F^c(x_i;\boldsymbol{\tau})]^2}\Big\{\big[(c-1)F^{-1}(x_i;\boldsymbol{\tau})[\dot{F}(x_i)_{\boldsymbol{\tau}}]^2 + \ddot{F}(x_i)_{\boldsymbol{\tau\tau}}\big] \\ & \times\big[1-F^c(x_i;\boldsymbol{\tau})\big] + cF^{c-1}(x_i;\boldsymbol{\tau})[\dot{F}(x_i)_{\boldsymbol{\tau}}]^2\Big\}, \end{aligned}$$

where $\psi'(\cdot)$ is the derivative of the digamma function, $\dot{f}(x_i)_{\boldsymbol{\tau}} = \partial f(x_i;\boldsymbol{\tau})/\partial\boldsymbol{\tau}$, $\dot{F}(x_i)_{\boldsymbol{\tau}} = \partial F(x_i;\boldsymbol{\tau})/\partial\boldsymbol{\tau}$, $\ddot{f}(x_i)_{\boldsymbol{\tau\tau}} = \partial^2 f(x_i;\boldsymbol{\tau})/\partial\boldsymbol{\tau}\boldsymbol{\tau}^T$ and $\ddot{F}(x_i)_{\boldsymbol{\tau\tau}} = \partial^2 F(x_i;\boldsymbol{\tau})/\partial\boldsymbol{\tau}\boldsymbol{\tau}^T$.

References

Aas, K., Haff, I.H., 2006. The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, **4**, 275-309.

Akinsete, A., Famoye, F., Lee, C., 2008. The beta-Pareto distribution. *Statistics*, **42**, 547-563.

Arnold, B.C., Castillo, E., Sarabia, J.M., 2006. Families of multivariate distributions involving the Rosenblatt construction. *Journal of the American Statistical Association*, **101**, 1652-1662.

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

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society B*, **65**, 367-389.

Barreto-Souza, W., Santos, A.H.S., Cordeiro, G.M., 2010. The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, **80**, 159-172.

Birnbaum, Z.W., Saunders, S.C., 1969. A new family of life distributions. *Journal of Applied Probability*, **6**, 319-327.

Brys, G., Hubert, M., Struyf, A., 2006. Robust measures of tail weight. *Computational Statistics and Data Analysis*, **50**, 733-759.

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

Choudhury, A., 2005. A simple derivation of moments of the exponentiated Weibull distribution. *Metrika*, **62**, 17–22.

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

Cordeiro, G.M., Lemonte, A.J., 2011. The β-Birnbaum-Saunders distribution: an improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, **55**, 1445-1461.

Cordeiro, G.M., Ortega, E.M.M., Nadarajah, S., 2010. The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, **347**, 1399-1429.

Doornik, J.A., 2007. *An Object-Oriented Matrix Language Ox 5.* Timberlake Consultants Press, London.

Ebrahimi, N., Maasoumi, E., Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics*, **90**, 317-336.

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

Exton, H., 1978. *Handbook of Hypergeometric Integrals: Theory, Applications, Tables, Computer Programs.* Halsted Press, New York.

Gradshteyn, I.S., Ryzhik, I.M., 2000. *Table of Integrals, Series, and Products*, sixth edition. Academic Press, San Diego.

Gupta, R.C., Gupta, R.D., Gupta, P.L., 1998. Modeling failure time data by Lehmann alternatives. *Communications in Statistics—Theory and Methods*, **27**, 887-904.

Gupta, R.D., Kundu, D., 2001. Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal*, **43**, 117-130.

Hansen, B.E., 1994. Autoregressive conditional density estimation. *International Economic Review*, **35**, 705-730.

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

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994.*Continuous Univariate Distributions I.* Wiley, New York.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1995.*Continuous Univariate Distributions II.* Wiley, New York.

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

Jones, M.C., 2004. Families of distributions arising from distributions of order statistics. *Test*, **13**, 1-43.

Jones, M.C., Larsen, P.V., 2004. Multivariate distributions with support above the diagonal. *Biometrika*, **91**, 975-986.

Jones, M.C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tracta bility advantages. *Statistical Methodology*, **6**, 70-81.

Kenney, J.F., Keeping, E.S., 1962. *Mathematics of statistics.* Princeton, New Jersey.

Kumaraswamy, P., 1980. A generalized probability density for double-bounded random processes. *Journal of Hydrology*, **46**, 79-88.

Lee, C., Famoye, F., Olumolade, O., 2007. Beta-Weibull Distribution: Some Properties and Applications to Censored Data. *Journal of Modern Applied Statistical Methods*, **6**, 173-186.

Meeker, W.Q., Escobar, L.A., 1998. *Statistical Methods for Reliability Data.* John Wiley, New York.

McDonald, J.B., 1984. Some generalized functions for the size distribution of income. *Econometrica*, **52**, 647-663.

McDonald J.B., Xu, Y.J., 1995. A generalization of the beta distribution with application. *Journal of Econometrics*, **66**, 133-152.

Moors, J.J.A., 1998. A quantile alternative for kurtosis. *Journal of the Royal Statistical Society D, The Statistician*, **37**, 25-32.

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

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

Nadarajah, S., 2008. Explicit expressions for moments of order statistics. *Statistics and Probability Letters*, **78**, 196-205.

Nadarajah S., Zografos, K., 2003. Formulas for Renyi information and related measures for univariate distributions. *Information Sciences*, **155**, 199-138.

Nadarajah, S., Kotz, S., 2004. The beta Gumbel distribution. *Math. Probab. Eng.*, **10**, 323-332.

Nadarajah, S., Gupta, A.K., 2005. On the moments of the exponentiated Weibull distribution. *Communication in Statistics - Theory and Methods*, **35**, 253-256.

Nadarajah, S., Kotz, S., 2006. The beta exponential distribution. *Reliability Engineering and System Safety*, **91**, 689-697.

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

Nassar, M.M., Eissa, F.H., 2003. On the exponentiated Weibull distribution. *Communication in Statistics - Theory and Methods*, **32**, 1317-1336.

Razzaghi, M., 2009. Beta-normal distribution in dose-response modeling and risk assessment for quantitative responses. *Environmental and Ecological Science*, **16**, 21-36.

Saavedra del Aguila, J., Heiffig-del Aguila, L.S., Sasaki, F.F., Tsumanuma, G.M., Ongarelli, M.G., Spoto, M.H.F., Jacomino, A.P., Ortega, E.M.M., Kluge, R.A., 2010. Postharvest modifications of mechanically injured bananas. *Revista Iberoamericana de Tecnologia Postcosecha*, **10**, 73–85.

Sartori, N., 2006. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, **136**, 4259-4275.

Shannon, C.E., 1948. The mathematical theory of communication. *Bell System Technical Journal* July-Oct. Reprinted in: C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Illinois University Press, Urbana.

Seier, E., Bonett, D., 2003. Two families of kurtosis measures. *Metrika*, **58**, 59-70.

Zhu, D., Galbraith, J.W., 2010. A generalized asymmetric Student-t distribution with application to financial econometrics. *Journal of Econometrics*, **157**, 297-305.

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

Zografos, K., 2011. *Generalized beta generated-II distributions. In: Modern Mathematical Tools and Techniques in Capturing Complexity.* L. Pardo, N. Balakrishnan and Maria Angeles Gil, Editors. Springer.