Computational Statistics and Data Analysis 56 (2012) 1880–1897

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Generalized beta-generated distributions

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

[a] *ICMA Centre, Henley Business School at Reading, UK*
[b] *Departamento de Estatística, Universidade Federal de Pernambuco, Brazil*
[c] *Departamento de Ciências Exatas, Universidade de São Paulo, Brazil*
[d] *Department of Economics, University of Cantabria, Spain*

A R T I C L E   I N F O

Article history:
Received 31 January 2011
Received in revised form 21 November 2011
Accepted 21 November 2011
Available online 27 November 2011

Keywords:
Entropy
Exponentiated
Kumaraswamy
Kurtosis
McDonald
Minimax
Skewness

A B S T R A C T

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.

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 behaviour.

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

* Corresponding author. Tel.: +55 11 19 34294127; fax: +55 11 19 34294468.
E-mail addresses: c.alexander@icmacentre.ac.uk (C. Alexander), gausscordeiro@uol.com.br (G.M. Cordeiro), edwin@esalq.usp.br (E.M.M. Ortega), sarabiaj@unican.es (J.M. Sarabia).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.11.015

characterized by its density

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

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

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

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 modelling. 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

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

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:

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

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.