Chilean Journal of Statistics
Vol. 7, No. 2, September 2016, 29-50

Research Paper

The Exponentiated Gompertz Generated Family of Distributions: Properties and Applications

Gauss M. Cordeiro[1], Morad Alizadeh[2,*], Abraão D. C. Nascimento[1], Mahdi Rasekhi[3]

[1]Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Brazil,
[2]Department of Statistics, Persian Gulf University, Bushehr, Iran,
[3]Department of Statistics, Malayer University, Malayer, Iran,

(Received: June 21, 2016 · Accepted in final form: August 31, 2016)

Abstract

The proposal of more flexible distributions is an activity often required in practical contexts. In particular, adding a positive real parameter to a probability distribution by exponentiation of its cumulative distribution function has provided flexible generated distributions having interesting statistical properties. In this paper, we study general mathematical properties of a new generator of continuous distributions with three extra parameters called the exponentiated Gompertz generated (EGG) family. We present some of its special models as well as an essay on its physical motivation. From mathematical point of view, we derive explicit expressions of the EGG family: the ordinary and incomplete moments, quantile and generating functions, Bonferroni and Lorenz curves, Shannon and Rényi entropies and order statistics, which are valid for any baseline model. We also provide a bivariate EGG extension. The estimation procedure by maximum likelihood of the new class is elaborated and discussed. In order to quantify and to assess the asymptotic behavior of this procedure, we perform a simulation study. Finally, two applications to real data are performed. Results furnish evidence in favor of the use of the EGG beta distribution as a good proposal to these data sets.

Keywords: Reliability and life testing · Applications to social sciences · Gompertz distribution · New generator · New Distribution for proportion.

1. Introduction

In many practical situations, classical distributions do not provide adequate fits to real data. Thus, several generators of introducing one or more parameters to generate new distributions have been proposed in the statistical literature. Some well known generators are

- Marshal-Olkin generated family (MO-G) (Marshall and Olkin, 1997),
- the beta-G by Eugene et al. (2002) and Jones (2004), Kumaraswamy-G (Kw-G for short) by Cordeiro and de Castro (2011) and McDonald-G (Mc-G) by Alexander et al. (2012),
- gamma-G (type 1) by Zografos and Balakrishnan (2009), gamma-G (type 2) by Ristić and Balakrishnan (2012), gamma-G (type 3) by Torabi and Hedesh (2012) and log-gamma-G by Amini et al. (2012),

*Corresponding author. Email: moradalizadeh78@gmail.com

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

- logistic-G by Torabi and Montazeri (2012),
- exponentiated generalized-G by Cordeiro et al. (2011),
- Transformed-Transformer (T-X) by Alzaatreh et al. (2013) and exponentiated (T-X) by Alzaghal et al. (2013),
- Weibull-G by Bourguignon et al. (2014) and
- Exponentiated half logistic generated family by Cordeiro et al. (2014).

In general, all above classes can be expressed within one formulation as follows.

Let $r(t)$ be the probability density function (pdf) of the random variable $T \in [a, b]$ for $-\infty < a < b < \infty$ and $W[G(x)]$ be the cumulative distribution function (cdf) of the random variable X, satisfying the following conditions:

(i) $W[G(x)] \in [a, b]$,
(ii) $W[G(x)]$ is differentiable and monotonically non-decreasing, and
(iii) $W[G(x)] \to a$ as $x \to -\infty$ and $W[G(x)] \to b$ as $x \to \infty$.

Recently, Alzaatreh et al. (2013) defined the *T-X family* of distributions by

$$F(x) = \int_a^{W[G(x)]} r(t)\,\mathrm{d}t, \tag{1}$$

where $W[G(x)]$ satisfies the conditions (i)–(iii). The pdf obtained from (1) is given by

$$f(x) = \left\{\frac{\mathrm{d}}{\mathrm{d}x} W[G(x)]\right\} r\{W[G(x)]\}. \tag{2}$$

In the remainder of this paper, we introduce a new class from (1) and study some of its statistical properties.

El-Gohary et al. (2013) defined the three-parameter *generalized Gompertz* (GG) distribution having pdf and cdf given by

$$\pi(x\,;\theta,\gamma,\alpha) = \alpha\,\theta\,\mathrm{e}^{\gamma x}\,\mathrm{e}^{-\frac{\theta}{\gamma}(\mathrm{e}^{\gamma x}-1)}\left[1-\mathrm{e}^{-\frac{\theta}{\gamma}(\mathrm{e}^{\gamma x}-1)}\right]^{\alpha-1}$$

and

$$\Pi(x\,;\theta,\gamma,\alpha) = \left[1-\mathrm{e}^{-\frac{\theta}{\gamma}(\mathrm{e}^{\gamma x}-1)}\right]^{\alpha},$$

respectively, where $x, \theta, \gamma, \alpha > 0$. Notice that the GG model colapses in the Gompertz (G) distribution when $\alpha = 1$. Now, applying (i) $W[G(x)] = -\log[1 - G(x)]$ and (ii)

$$r(t) = \alpha\,\theta\,\mathrm{e}^{\gamma t}\,\mathrm{e}^{-\frac{\theta}{\gamma}(\mathrm{e}^{\gamma t}-1)}\left[1-\mathrm{e}^{-\frac{\theta}{\gamma}(\mathrm{e}^{\gamma t}-1)}\right]^{\alpha-1}, \quad t > 0,$$

on (2), we define a new family of distributions–called the *exponentiated Gompertz generated* ("EGG" for short) family– having cdf expressed in the form of the next theorem.

Theorem 1.1 The cdf of the EGG family is given by

$$F(x\,;\theta,\gamma,\alpha,\boldsymbol{\xi}) = \int_0^{-\log[1-G(x;\boldsymbol{\xi})]} \pi(w\,;\theta,\gamma,\alpha)\,\mathrm{d}w = \left\{1-\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G(x;\boldsymbol{\xi})]^{-\gamma}\}}\right\}^{\alpha}, \tag{3}$$

where $\theta > 0$, $\gamma > 0$ and $\alpha > 0$ are three extra shape parameters.

Equation (3) is a wider family of continuous distributions. It includes several classes such as the *proportional hazard rate distributions* (Gupta and Gupta, 2007). For each baseline G, the EGG-G distribution is defined by the cdf (3). Henceforth, $X \sim$ EGG-G$(\theta, \gamma, \alpha, \boldsymbol{\xi})$ denotes a random variable having the cdf (3) and $G(x\,;\boldsymbol{\xi})$ denotes the baseline cdf depending on a parameter vector $\boldsymbol{\xi}$. For simplicity, we can omit the dependence on $\boldsymbol{\xi}$ and write simply $G(x) = G(x;\boldsymbol{\xi})$.

A probabilistic deduction and physical interpretation of the EGG family can be addressed as follows. Let Y be a random variable having the Gompertz distribution (with positive parameters θ and γ representing the shape and scale, respectively) and Z be any random variable with baseline cdf $G(z;\boldsymbol{\xi})$. Thus, if we consider α systems, $T_1, \ldots, T_\alpha$ working independently as a random sample from $T^\bullet = G^{-1}(1 - \mathrm{e}^{-Y};\boldsymbol{\xi})$, then the cdf of $X = \max\{T_1, \ldots, T_\alpha\}$ given by

$$\begin{aligned}
\Pr(X \leq x) &= \Pr(\cap_{i=1}^{\alpha} \{X_i \leq x\}) = \{\Pr(\{X_1 \leq x\})\}^{\alpha} \\
&= \{\Pr(G^{-1}(1 - \mathrm{e}^{-Y};\boldsymbol{\xi}) \leq x)\}^{\alpha} \\
&= \{\Pr(Y \leq -\log[1 - G(x;\boldsymbol{\xi})])\}^{\alpha} \\
&= F(x\,;\theta, \gamma, \alpha, \boldsymbol{\xi}),
\end{aligned}$$

follows the proposed family. In other words, this property characterizes the distribution of the maximum of α random variables having the G distribution (with parameter vector $\boldsymbol{\xi}$) such that their cdf's are defined by $1 - \mathrm{e}^{-Y}$, where Y has the Gompertz(θ, γ) distribution. Applications in several areas can be associated to the above discussion, such those in reliability and anthropological contexts. Consider a system formed by α independent components following specific transformed distribution involving the Gompertz and a baseline G model. Suppose the system fails if all of the α components fail. Let $T_1, \ldots, T_\alpha$ denote the lifetimes. Thus, X describes the maximum lifetime under the influence of the parameters α, θ, γ and $\boldsymbol{\xi}$.

This paper is organized as follows. In Section 2, we provide the EGG density and quantile functions. Three special cases of the proposed family are addressed in Section 3. Some useful expansions are derived in Section 4. A power series expansion for the quantile function (qf) of X is derived in Section 5. In Section 6, we provide explicit expressions for the moments and generating function of X. In Section 7, we present general expressions for the Rényi and Shannon entropies. In Section 8, the order statistics are investigated. In Section 9, we introduce a bivariate extension of the new family. Estimation of the model parameters by maximum likelihood is performed in Section 10. A simulation study and two applications to real data illustrate the usefulness of the EGG family in Section 12. The paper is concluded in Section 13.

2. The new family

The pdf and hrf corresponding to (3) are provided in the corollary:

Corollary 2.1 The pdf and hrf of X are given by

$$f(x\,;\theta, \gamma, \alpha, \boldsymbol{\xi}) = \frac{\alpha\,\theta\, g(x\,;\boldsymbol{\xi})\,\mathrm{e}^{\frac{\theta}{\gamma}\{1 - [1 - G(x\,;\boldsymbol{\xi})]^{-\gamma}\}}}{[1 - G(x\,;\boldsymbol{\xi})]^{1+\gamma}} \left\{1 - \mathrm{e}^{\frac{\theta}{\gamma}\{1 - [1 - G(x\,;\boldsymbol{\xi})]^{-\gamma}\}}\right\}^{\alpha - 1}, \tag{4}$$

and

$$h(x\,;\theta,\gamma,\alpha,\boldsymbol{\xi}) \;=\; \frac{\theta\; g(x\,;\,\boldsymbol{\xi})}{[\,1\,-\,G(x\,;\,\boldsymbol{\xi})\,]^{\gamma+1}}\;\frac{\left\{1\,-\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G(x\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\right\}^{\alpha-1}}{\left[1\,-\,\left\{1\,-\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G(x\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\right\}^{\alpha}\right]}, \qquad (5)$$

respectively, where $x \in \mathcal{X}$ and $g(x\,;\,\boldsymbol{\xi})$ is the baseline pdf.

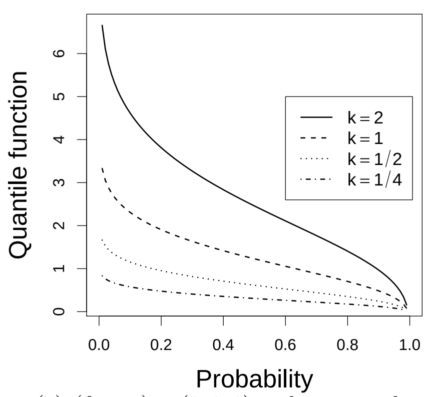
A simple pdf (4) arises when one considers simple expressions for $G(x)$ and $g(x)$. As the EGG family can be understood as a "dictionary" of special models, an other practical issue is "how to select the more appropriate baseline for a data set?". In practice, the following criteria are often used: (i) take into account the variation range of real data and/or (ii) choose one parent distribution among many others models based on the smaller value assumed by the sum of squablack errors between the histogram and the fitted density $f(x\,;\widehat{\theta},\widehat{\gamma},\widehat{\alpha},\widehat{\boldsymbol{\xi}})$ as discussed by Cintra et al. (2013), where $\widehat{\theta}$, $\widehat{\gamma}$, $\widehat{\alpha}$ and $\widehat{\boldsymbol{\xi}}$ are obtained by any estimation method, usually maximum likelihood.

The EGG family of distributions is easily simulated by inverting (3) as follows: if u is an outcome from the uniform $U(0,1)$ distribution, then

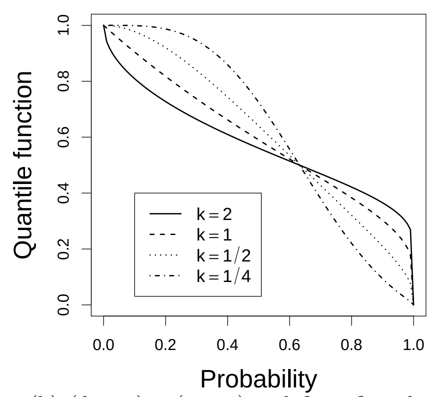
$$x \;=\; Q_G\left\{1-\left[1-\frac{\gamma}{\theta}\log(1-u^{\frac{1}{\alpha}})\right]^{\frac{-1}{\gamma}}\right\}, \qquad (6)$$

is an outcome of a random variable X having pdf (4), where $Q_G(x) = G^{-1}(x)$ is the qf of G.

Figure 1 displays the qf's of X taking the normal $\mathrm{N}(\mu,\sigma^2)$ and beta $\mathrm{B}(\beta_1,\beta_2)$ models as the parent distributions. These situations are detailed in the next section.



(a) $(\theta,\gamma,\alpha) = (1,1,1)$ and $\mu = \sigma = k$



(b) $(\theta,\gamma,\alpha) = (1,1,1)$ and $\beta_1 = \beta_2 = k$

Figure 1. Plots of the EGG qf's for the $N(\mu,\sigma^2)$ and $B(\beta_1,\beta_2)$ baselines.

3. Some Special EGG distributions

For $\alpha = 1$, we obtain as a special case of (4) the Gompertz-G family of distributions (Alizadeh et al., 2016). The exponentiated class of distributions proposed by Cordeiro et al. (2016) arises as another special case when $\gamma \downarrow 0$. For $\gamma \downarrow 0$ and $\alpha = 1$, equation (4) reduces to the Gupta and Gupta's proportional hazard rate class (Gupta and Gupta, 2007). For $\gamma \downarrow 0$ and $\theta = 1$, we have the reversed hazard rate class (Gupta and Gupta, 2007), which provides greater flexibility of its tails and can be widely applied in many areas of engineering and biology. Some special models are listed in Table 1.

This table indicates that the hypotheses $H_0: \alpha = 1$ and $\gamma \downarrow 0$, $H_0: \theta = 1$ and $\gamma \downarrow 0$ and $H_0: \theta = \alpha = 1$ and $\gamma \downarrow 0$ can be used when one wishes to assess an extension of the class proposed by Alizadeh et al. (2016), a change of the proportional hazard rate model in the form of the parameter θ and the influence of the new class on the baseline distribution, respectively. Next, we present three special cases of the EGG family, where $\gamma > 0$, $\theta > 0$ and $\alpha > 0$.

Table 1. Some special models

α	θ	γ	$G(x)$	family
1	-	-	-	Gompertz-G family of distributions (Alizadeh et al., 2016)
-	-	$\downarrow 0$	-	Exponentiated class of distributions (Cordeiro et al., 2016)
1	-	$\downarrow 0$	-	Proportional hazard rate model (Gupta and Gupta, 2007)
-	1	$\downarrow 0$	-	Proportional reversed hazard rate model (Gupta and Gupta, 2007)
1	1	$\downarrow 0$	-	$G(x)$
1	-	-	$1 - e^{-x}$	Gompertz distribution
-	-	-	$1 - e^{-x}$	Generalized Gompertz distribution (El-Gohary et al., 2013)
1	-	$\downarrow 0$	x^v	Kumaraswamy distribution (Kumaraswamy, 1980; Jones, 2009)
-	-	$\downarrow 0$	x^v	Exponentiated Kumaraswamy distribution and its log transform (Lemonte et al., 2013)

3.1 The EGGN distribution

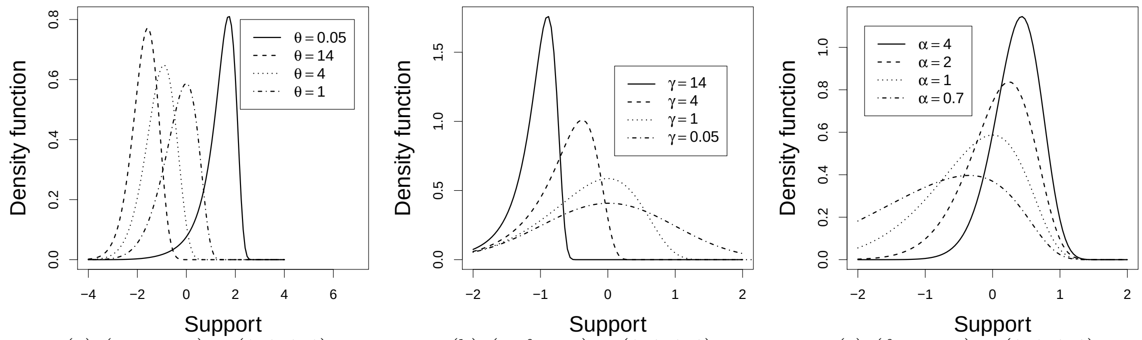
We define the *exponentiated Gompertz generated normal* (EGGN) distribution from (4) by taking $G(x\,;\boldsymbol{\xi}) = \Phi(\frac{x-\mu}{\sigma})$ and $g(x\,;\boldsymbol{\xi}) = \sigma^{-1}\,\phi(\frac{x-\mu}{\sigma})$ to be the cdf and pdf of the normal distribution with parameters μ and σ^2, respectively, $\boldsymbol{\xi} = (\mu, \sigma^2)$, and where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. Then, the EGGN pdf is given by

$$
\begin{aligned}
f(x\,;\theta,\gamma,\alpha,\mu,\sigma) \;=\; & \frac{\alpha\,\theta}{\sigma}\,\phi\Big(\frac{x-\mu}{\sigma}\Big)\Big[1 - \Phi\Big(\frac{x-\mu}{\sigma}\Big)\Big]^{-\gamma-1} e^{\frac{\theta}{\gamma}\{1-[1-\Phi(\frac{x-\mu}{\sigma})]^{-\gamma}\}} \\
& \times \Big\{1 - e^{\frac{\theta}{\gamma}\{1-[1-\Phi(\frac{x-\mu}{\sigma})]^{-\gamma}\}}\Big\}^{\alpha-1}, \qquad (7)
\end{aligned}
$$

where $x \in \mathrm{R}$, $\mu \in \mathrm{R}$ is a location parameter and $\sigma > 0$ is a scale parameter. The random variable with pdf (7) is denoted by $X \sim \text{EGGN}\ (\theta, \gamma, \alpha, \mu, \sigma^2)$. For $\mu = 0$, $\sigma = 1$ and $\gamma \to 0$, we have the power-normal (PN) distribution (Gupta and Gupta, 2008). Further, the basic exemplar when $\theta = 1$ and $\gamma \to 0$ is the normal distribution. Plots of the EGGN density function for some shape parameter values are displayed in Figure 2. These plots indicate that decreasing α and γ causes a flattening of the density curves, whereas this behavior happens under increasing θ (for $\theta > 1$).

3.2 The EGGa distribution

We also present as a special model the *exponentiated Gompertz generated gamma* (EGGGa) distribution from the gamma distribution with shape parameter $a > 0$ and scale parameter



(a) $(\alpha, \gamma, \mu, \sigma) = (1, 1, 1, 1)$ (b) $(\alpha, \theta, \mu, \sigma) = (1, 1, 1, 1)$ (c) $(\theta, \gamma, \mu, \sigma) = (1, 1, 1, 1)$

Figure 2. Plots of the EGGN density function for some parameter values.

$b > 0$. In this case, the baseline pdf and cdf (for $x > 0$) are given by

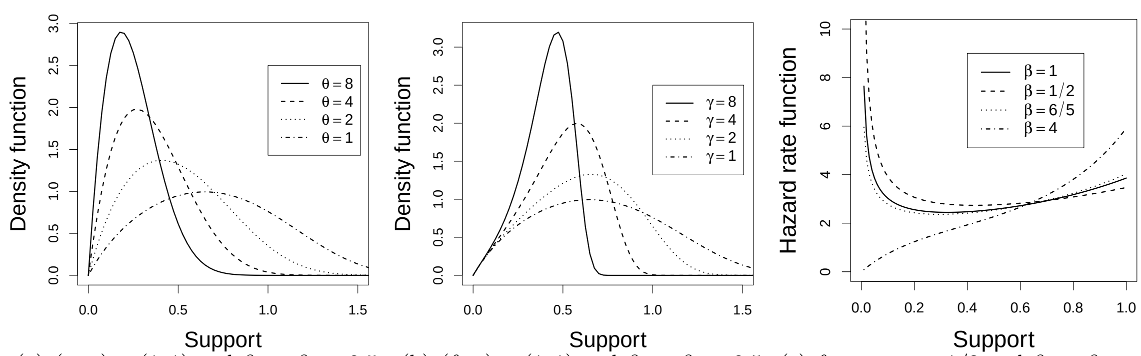
$$g(x\,;\,\beta_1, \beta_2) \;=\; \frac{\beta_2^{\beta_1}}{\Gamma(\beta_1)}\, x^{\beta_1 - 1}\, \mathrm{e}^{-\beta_2 x} \text{ and } G(x\,;\,\beta_1, \beta_2) \;=\; \gamma_1(\beta_1, \beta_2\, x),$$

where $\gamma(a, x) = \int_0^x t^{a-1}\,\mathrm{e}^{-t}\,\mathrm{d}t$ is the incomplete gamma function and $\gamma_1(a, x) = \gamma(a, x)/\Gamma(a)$ is the incomplete gamma function ratio.

Inserting these expressions in (4) leads to the EGGGa density function

$$f(x\,;\,\theta, \gamma, \alpha, \beta_1, \beta_2) = \alpha\,\theta\, \frac{\beta_2^{\beta_1}}{\Gamma(\beta_1)}\, x^{\beta_1 - 1}\, \mathrm{e}^{-b\,x} \Big[1 \,-\, \gamma_1(\beta_1, \beta_2\, x)\Big]^{-\gamma - 1} \mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1 - [\,1 - \gamma_1(\beta_1, \beta_2\, x)\,]^{-\gamma}\,\}}$$
$$\times \Big\{1 \,-\, \mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1 - [\,1 - \gamma_1(\beta_1, \beta_2\, x)\,]^{-\gamma}\,\}}\Big\}^{\alpha - 1}.$$

The exponentiated-gamma (EG) distribution comes from this model when $\gamma \to 0$. Plots of the EGGGa pdf and hrf for selected parameter values are displayed in Figure 3.



(a) $(\gamma, \alpha) = (1, 1)$ and $\beta_1 = \beta_2 = 0.5$ (b) $(\theta, \alpha) = (1, 1)$ and $\beta_1 = \beta_2 = 0.5$ (c) $\theta = \gamma = \alpha = 1/2$ and $\beta_1 = \beta_2 = \beta$

Figure 3. Plots of the EGGGa density and hazard rate functions for some parameter values.

3.3 The EGGB distribution

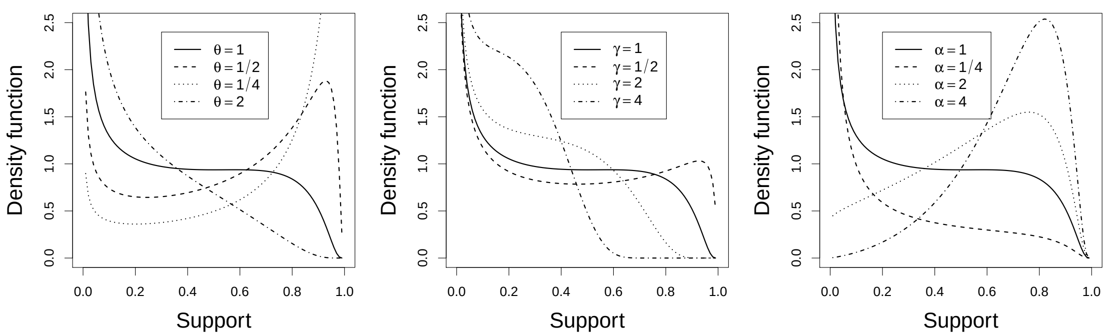
We define the *exponentiated Gompertz generated beta* (EGGB) distribution from the beta distribution with positive shape parameters β_1 and β_2. Its pdf and cdf (for $0 < x < 1$) are given by

$$g(x\,;\,\beta_1,\beta_2) \;=\; \frac{1}{B(\beta_1,\beta_2)}\,x^{\beta_1\,-\,1}\,(1\,-\,x)^{\beta_2\,-\,1} \text{ and } G(x\,;\,\beta_1,\beta_2) \;=\; I_x(\beta_1,\beta_2),$$

where $I_x(\beta_1,\beta_2) \,=\, B(\beta_1,\beta_2)^{-1}\,\int_0^x\, w^{\beta_1\,-\,1}\,(1-w)^{\beta_2\,-\,1}\mathrm{d}w$ is the incomplete beta function ratio and $B(\beta_1,\beta_2) \,=\, \int_0^1\, w^{\beta_1-1}\,(1\,-\,w)^{\beta_2\,-\,1}\,\mathrm{d}w \,=\, \Gamma(\beta_1)\Gamma(\beta_2)/\Gamma(\beta_1\,+\,\beta_2)$ is the beta function. Inserting these expressions in (4) yields the EGGB density function

$$\begin{aligned} f(x\,;\,\theta,\gamma,\alpha,\beta_1,\beta_2) \;=\; &\frac{\alpha\,\theta}{B(\beta_1,\beta_2)}\,x^{\beta_1\,-\,1}\,(1\,-\,x)^{\beta_2\,-\,1}\Big[1\,-\,I_x(\beta_1,\beta_2)\Big]^{-\gamma\,-\,1} \\ &\times\,\mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1\,-\,[\,1\,-\,I_x(\beta_1,\beta_2)\,]^{-\gamma}\,\}}\Big\{\,1\,-\,\mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1\,-\,[\,1\,-\,I_x(\beta_1,\beta_2)\,]^{-\gamma}\,\}}\,\Big\}^{\alpha\,-\,1}. \end{aligned}$$

The beta distribution arises as a special case when $\gamma \to 0$, $\alpha = 1$ and $\theta = 1$. Plots of the EGGB density function for some parameter values are displayed in Figure 4.



(a) $(\gamma,\alpha) = (1,1)$ and $\beta_1 = \beta_2 = 1/2$ (b) $(\theta,\alpha) = (1,1)$ and $\beta_1 = \beta_2 = 1/2$ (c) $(\theta,\gamma) = (1,1)$ and $\beta_1 = \beta_2 = 1/2$

Figure 4. Plots of the EGGB density function for some parameter values.

4. Useful expansions

For an arbitrary baseline cdf $G(x)$, a random variable Z has the *exponentiated-G* ("exp-G" for short) distribution with power parameter $c > 0$, say $Z \sim$exp-G(c), if its pdf and cdf are given by

$$h_c(x) \;=\; c\,G(x)^{c-1}\,g(x) \qquad \text{and} \qquad H_c(x) \;=\; G(x)^c,$$

respectively. Some structural properties of the exp-G distributions are studied by Mudholkar et al. (1995), Gupta and Kundu (1999) and Nadarajah and Kotz (2006), among several others.

Nadarajah et al. (2013) demonstrated the following lemma.

Lemma 4.1 Let X have the cdf (3). Then, its cdf admits the expansion

$$F(x) \;=\; \sum_{k=0}^{\infty} b_k\,H_k(x), \tag{8}$$

where $H_k(x)$ denotes the exp-G cdf with power parameter k and

$$b_k = \mathrm{I}_0(k) + \underbrace{\sum_{i=1}^{\infty}\sum_{j=0}^{\infty} w_{i,j,k}}_{=w_{+,+,k}} = \mathrm{I}_0(k) + w_{+,+,k}.$$

Here, $\mathrm{I}_0(k)$ is an indicator function which takes one if $k = 0$ and the coefficient $w_{i,j,k}$ is given by

$$w_{i,j,k} = \frac{(-1)^{i+j+k}}{j!}\binom{\alpha}{i}\binom{-j\,\gamma}{k}\left(\frac{i\,\theta}{\gamma}\right)^{j} \mathrm{e}^{\frac{i\,\theta}{\gamma}}.$$

Based on Lemma 4.1, it follows the Corollary 3.

Corollary 4.2 The pdf of X can be expressed as an infinite linear combination of exp-G density functions

$$f(x;\alpha,\lambda,p,\boldsymbol{\xi}) = \sum_{k=0}^{\infty} b_{k+1}\, h_{k+1}(x;\boldsymbol{\xi}), \tag{9}$$

where $h_{k+1}(x;\boldsymbol{\xi}) = (k+1)\, g(x;\boldsymbol{\xi})\, G(x;\boldsymbol{\xi})^k$ denotes the exp-G pdf with power parameter $k+1$. A random variable having this density function will be denoted from now on by $Y_{k+1} \sim$ exp-G$(k+1)$.

Thus, some mathematical properties of the proposed family can be obtained directly from those properties of the exp-G distribution.

5. Quantile power series

In this section, we propose alternative explicit expressions for the moments and generating function of the EGG family using a power series for the qf (6). This result is given in the following theorem.

Theorem 5.1 The qf of X can be expanded as

$$Q(u) = \sum_{m=0}^{\infty} e_p\, u^p, \tag{10}$$

where $e_p = \sum_{i=0}^{\infty} a_i\, q_{i,p}$, and for $i \geq 0$, $q_{i,0} = \delta_0^i$ and (for $p > 1$)

$$q_{i,p} = (p\,\delta_0)^{-1} \sum_{n=1}^{p} [n(i+1) - p]\, \delta_n\, q_{i,p-n}. \tag{11}$$

The definition of the quantities are in Theorem 2 and the a_i's related to the baseline qf are given in Appendix A.

Equation (10) is the main result of this section. It allows to obtain various mathematical quantities for the EGG family as proved in the next sections.

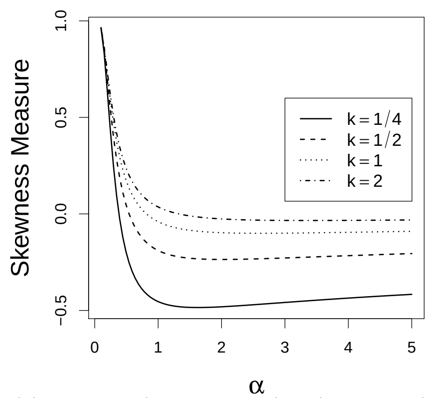
The effects of the shape parameters on the skewness and kurtosis can be based on quantile measures. The shortcomings of the classical kurtosis measure are well-known. The Bowley skewness (Kenney and Keeping, 1962, pp. 101–102) is one of the earliest skewness measures defined by the average of the quartiles minus the median, divided by half the interquartile range, namely

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

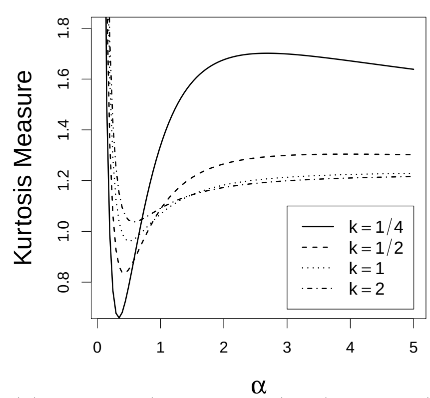
Since only the middle two quartiles are consideblack and the outer two quartiles are ignoblack, this adds robustness to the measure. The Moors kurtosis (Moors, 1998) is based on octiles

$$M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}.$$

From the last two equations, the skewness and kurtosis measures can be determined as functions of the qf of X in equation (6). These measures are less sensitive to outliers and they exist even for distributions without moments. Figure 5 displays the plots of the measures B and M for the EGGB distribution discussed in Section 3. These plots indicate that both measures B and M depend very much on the shape parameters.



(a) B_{EGGB}: $(\theta, \gamma, \alpha, \beta_1, \beta_2) = (k, k, \bullet, 1, 1)$



(b) M_{EGGN}: $(\theta, \gamma, \alpha, \beta_1, \beta_2) = (k, k, \bullet, 1, 1)$

Figure 5. Plots of the skewness and kurtosis measures for the EGGB distribution.

6. Moments and generating function

Let Y_{k+1} be a random variable having the exp-G distribution with power parameter $k+1$, i.e., with density $h_{k+1}(x)$. A first formula for the nth moment of X follows from (9) as

$$\mathrm{E}(X^n) = \sum_{k=0}^{\infty} b_{k+1}\,\mathrm{E}(Y_{k+1}^n). \tag{12}$$

Expressions for moments of several exp-G distributions are given by Nadarajah and Kotz (2006), which can be used to obtain $\mathrm{E}(X^n)$.

A second formula for $\mathrm{E}(X^n)$ can be expressed from (12) in terms of the G qf as

$$\mathrm{E}(X^n) = \sum_{k=0}^{\infty} (k+1)\,b_{k+1}\,\tau(n,k), \tag{13}$$

where $\tau(n,k)=\int_{-\infty}^{\infty}x^n\,G(x)^k\,g(x)\,dx=\int_0^1 Q_G(u)^n\,u^a du$. Cordeiro and Nadarajah (2011) obtained the quantity $\tau(n,k)$ for some well-known models such as the normal, beta, gamma and Weibull distributions.

We now move to the nth incomplete moment of X defined by $m_n(y)=\mathrm{E}(X^n|X<y)=\int_{-\infty}^{y}x^r\,f(x)\,\mathrm{d}x$. For empirical purposes, the shape of many distributions can be usefully described by what we call the incomplete moments. These types of moments play an important role for measuring inequality, for example, income quantiles and Lorenz and Bonferroni curves, which depend upon the incomplete moments. The quantity $m_n(y)$ can be expressed as

$$m_n(y)=\mathrm{E}(X^n|X<y)=\sum_{k=0}^{\infty}(k+1)\,b_{k+1}\int_0^{G(y;\boldsymbol{\xi})}Q_G(u)^n\,u^k du. \tag{14}$$

The integral in equation (14) can be computed at least numerically for most baseline distributions. A second method to obtain the incomplete moments of X follows from (14) using equations (A5) and (A6):

Corollary 6.1 The nth incomplete of X is given by

$$m_n(y)=\sum_{k,m=0}^{\infty}\frac{(k+1)\,b_{k+1}\,c_{n,m}}{(m+k+1)}\,G(y;\boldsymbol{\xi})^{m+k+1}, \tag{15}$$

where the coefficients $c_{n,m}$ can be determined by (A6).

Let $M_X(t)=\mathrm{E}(\mathrm{e}^{\mathrm{t\,X}})$ be the moment generating function (mgf) of X. A first expression for $M_X(t)$ comes from (9) as

$$M_X(t)=\sum_{k=0}^{\infty}b_{k+1}\,M_{k+1}(t), \tag{16}$$

where $M_{k+1}(t)$ is the mgf of Y_{k+1}. Hence, $M_X(t)$ can be determined from the exp-G generating function.

A second formula for $M(t)$ can be derived from (9) as

$$M(t)=\sum_{i=0}^{\infty}(k+1)\,b_{k+1}\,\rho(t,k), \tag{17}$$

where $\rho(t,k)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{t\,x}}\,\mathrm{G(x)^k\,g(x)dx}=\int_0^1\exp[\mathrm{t\,Q_G(u)}]\,\mathrm{u^k\,du}$.

We can obtain the mgfs of several distributions directly from equations (16) and (17).

7. Entropies

An entropy is a measure of variation or uncertainty of a random variable X. Two popular entropy measures are the Rényi and Shannon entropies (Shannon, 1948; Rényi, 1961). The Rényi entropy of a random variable with pdf $f(x)$ is defined as

$$I_R(c)=\frac{1}{1-c}\log\left(\int_0^{\infty}f^c(x)\,\mathrm{d}x\right),$$

for $c > 0$ and $c \neq 1$. The Shannon entropy of a random variable X is defined by $E\{-\log[f(X)]\}$. It is the special case of the Rényi entropy when $c \uparrow 1$. Direct calculation yields:

COROLLARY 7.1 The Shannon entropy of X is given by

$$
\begin{aligned}
&\mathrm{E}\{-\log[f(X)]\} = -\log(\alpha\theta) - \mathrm{E}\{\log[g(X;\boldsymbol{\xi})]\}\\
&\quad + \frac{(\gamma+1)\alpha\theta}{\gamma^2} \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j+1}}{(j+1)\,(j+1)!} \binom{\alpha-1}{i} \left[\frac{(i+1)\theta}{\gamma}\right]^j \mathrm{e}^{\frac{(i+1)\theta}{\gamma}}\\
&-\frac{\theta}{\gamma}\left\{1 - \frac{\alpha\theta}{\gamma^2} \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}}{(j+2)\,j!} \binom{\alpha-1}{i} \left[\frac{(i+1)\theta}{\gamma}\right]^j\right\} \mathrm{e}^{\frac{(i+1)\theta}{\gamma}}\\
&-\frac{\alpha(\alpha-1)\theta}{\gamma} \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j+1}}{(j+1)!} \left[\frac{(i+1)\theta}{\gamma}\right]^j \mathrm{e}^{\frac{(i+1)\theta}{\gamma}} \left[\frac{\partial}{\partial t}\binom{t+\alpha-1}{i}\Big|_{t=0}\right].
\end{aligned}
$$

After some algebraic developments, we can obtain an alternative expression for $I_R(c)$.

COROLLARY 7.2 The Rényi's entropy of X is given by

$$
I_R(c) = \frac{c}{1-c}\log(\alpha\theta) + \frac{1}{1-c}\log\left[\sum_{i,j,k}^{\infty} m_{i,j,k}\, I(c,k)\right] \tag{18}
$$

where

$$
m_{i,j,k} = \frac{(-1)^{i+j+k}\binom{c(\alpha-1)}{i}\mathrm{e}^{\frac{\theta(c+i)}{\gamma}}\,(\theta(c+i))^j\binom{-c(\gamma+1)\,-\,\gamma\,j}{k}}{j!\,\gamma^j}
$$

and

$$
I(c,k) = \int_{-\infty}^{\infty} g(x)^c\, G(x)^k\,\mathrm{d}x
$$

8. Order statistics

Order statistics make their appearance in many areas of statistical theory and practice. Suppose $X_1,\ldots,X_n$ is a random sample from the EGG-G family of distributions. Let $X_{i:n}$ denote the ith order statistic. The pdf of $X_{i:n}$ can be expressed as

$$
\begin{aligned}
f_{i:n}(x) &= K\, f(x)\, F^{i-1}(x)\ \{1-F(x)\}^{n-i}\\
&= K\sum_{j=0}^{n-i}(-1)^j\binom{n-i}{j} f(x)\, F(x)^{j+i-1},
\end{aligned}
$$

where $K = n!/[(i-1)!\,(n-i)!]$.

We can demonstrate that the density function of the ith order statistic of any EGG-G distribution follows Corollary 6.

COROLLARY 8.1 Let $X_1,\ldots,X_n$ be a random sample from $X \sim \mathrm{EGG}(\theta,\gamma,\alpha,\boldsymbol{\xi})$, The pdf of $X_{i:n}$ can be expressed as

$$f_{i:n}(x) = \sum_{r,k=0}^{\infty} \sum_{j=0}^{n-i} m_{j,r,k}\, h_{r+k+1}(x), \tag{19}$$

where $h_{r+k}(x)$ denotes the exp-G density function with power parameter $r+k$,

$$m_{j,r,k} = \frac{(-1)^j\, n!}{(i-1)!\,(n-i-j)!\,j!}\;\frac{(r+1)\,b_{r+1}\,f_{j+i-1,k}}{[r+k+1]}, \tag{20}$$

and b_k is given by (8). Here, the quantity $f_{j+i-1,k}$ is obtained recursively from $f_{j+i-1,0} = b_0^{j+i-1}$ and (for $k \geq 1$)

$$f_{j+i-1,k} = (k\,b_0)^{-1} \sum_{m=1}^{k} [m(j+i)-k]\, b_m\, f_{j+i-1,k-m}.$$

So, we can easily obtain the ordinary and incomplete moments and generating function for the EGG-G order statistics (based on any G distribution) from equation (19).

9. Bivariate extension

A bivariate extension is presented in the following theorem.

Theorem 9.1

$$F(x,y\,;\boldsymbol{\xi}) \;=\; \left\{1 \;-\; \mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1-[\,1-G(x,y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\right\}^{\alpha},$$

where $G(x,y\,;\boldsymbol{\xi})$ is a bivariate continuous distribution with marginal cdf's $G_1(x\,;\boldsymbol{\xi})$ and $G_2(y;\boldsymbol{\xi})$. The marginal cdf's are given by

$$F_X(x\,;\boldsymbol{\xi}) \;=\; \left\{1 \;-\; \mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1-[\,1-G(x\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\right\}^{\alpha}$$

and

$$F_Y(y\,;\boldsymbol{\xi}) \;=\; \left\{1 \;-\; \mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1-[\,1-G(y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\right\}^{\alpha}.$$

The joint pdf of (X,Y) is easily determined by $f_{X,Y}(x,y) = \partial^2 F_{X,Y}(x,y) \big/ \partial x\, \partial y$:

Corollary 9.2

$$\begin{aligned} f_{X,Y}(x,y) \;&= \alpha\,\theta\,[\,1 \;-\; G(x,y\,;\boldsymbol{\xi})\,]^{-\gamma-1}\,\mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1-[\,1-G(x,y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}} \\ &\quad\times\; A(x,y\,;\boldsymbol{\xi})\,\left\{1 \;-\; \mathrm{e}^{\frac{\theta}{\gamma}\,\{\,1-[\,1-G(x,y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\right\}^{\alpha-1}, \end{aligned}$$

where

$$A(x,y\,;\,\boldsymbol{\xi}) \;=\; g(x,y\,;\,\boldsymbol{\xi}) \;-\; \left[\frac{\gamma\,+\,1}{1\,-\,G(x,y\,;\,\boldsymbol{\xi})}\right]\left[\frac{\partial G(x,y\,;\,\boldsymbol{\xi})}{\partial x}\right]\;\left[\frac{\partial G(x,y\,;\,\boldsymbol{\xi})}{\partial y}\right]$$
$$-\;\frac{\theta}{[\,1\,-\,G(x,y\,;\,\boldsymbol{\xi})\,]^{\gamma+1}}\left[\frac{\partial G(x,y\,;\,\boldsymbol{\xi})}{\partial x}\right]\;\left[\frac{\partial G(x,y\,;\,\boldsymbol{\xi})}{\partial y}\right]$$
$$+\;\frac{\theta\,(\alpha\,-\,1)}{[\,1\,-\,G(x,y\,;\,\boldsymbol{\xi})\,]^{\gamma+1}}\;\frac{\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G(x,y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}}{1\,-\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G(x,y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}}\;\left[\frac{\partial G(x,y\,;\,\boldsymbol{\xi})}{\partial x}\right]\;\left[\frac{\partial G(x,y\,;\,\boldsymbol{\xi})}{\partial y}\right].$$

The marginal pdf's are

$$\begin{aligned} f_X(x\,;\,\theta,\gamma,\alpha,\boldsymbol{\xi}) \;&=\alpha\,\theta\,g_1(x\,;\,\boldsymbol{\xi})\,[\,1\,-\,G_1(x;\boldsymbol{\xi})]^{-\gamma-1}\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G_1(x\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}} \\ &\times\left\{\,1\,-\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G_1(x\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\,\right\}^{\alpha-1} \end{aligned}$$

and

$$\begin{aligned} f_Y(y\,;\,\theta,\gamma,\alpha,\boldsymbol{\xi}) \;&=\alpha\,\theta\,g_2(y\,;\,\boldsymbol{\xi})\,[\,1\,-\,G_2(y\,;\,\boldsymbol{\xi})\,]^{-\gamma-1}\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G_2(y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}} \\ &\times\left\{\,1\,-\,\mathrm{e}^{\frac{\theta}{\gamma}\{\,1-[\,1-G_2(y\,;\boldsymbol{\xi})\,]^{-\gamma}\,\}}\,\right\}^{\alpha-1}. \end{aligned}$$

The conditional density functions are

$$f(x\mid y)=\frac{[1-G(x,y;\boldsymbol{\xi})]^{-\gamma-1}\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G(x,y;\boldsymbol{\xi})]^{-\gamma}\}}A(x,y;\boldsymbol{\xi})\left\{1-\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G(x,y;\boldsymbol{\xi})]^{-\gamma}\}}\right\}^{\alpha-1}}{g_2(y;\boldsymbol{\xi})[1-G_2(y;\boldsymbol{\xi})]^{-\gamma-1}\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G_2(y;\boldsymbol{\xi})]^{-\gamma}\}}\left\{1-\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G_2(y;\boldsymbol{\xi})]^{-\gamma}\}}\right\}^{\alpha-1}}$$

and

$$f(y\mid x)=\frac{[1-G(x,y;\boldsymbol{\xi})]^{-\gamma-1}\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G(x,y;\boldsymbol{\xi})]^{-\gamma}\}}A(x,y;\boldsymbol{\xi})\left\{1-\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G(x,y;\boldsymbol{\xi})]^{-\gamma}\}}\right\}^{\alpha-1}}{g_1(x;\boldsymbol{\xi})[1-G_1(x;\boldsymbol{\xi})]^{-\gamma-1}\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G_1(x;\boldsymbol{\xi})]^{-\gamma}\}}\left\{1-\mathrm{e}^{\frac{\theta}{\gamma}\{1-[1-G_1(x;\boldsymbol{\xi})]^{-\gamma}\}}\right\}^{\alpha-1}}.$$

10. Estimation

We determine the maximum likelihood estimates (MLEs) of the parameters of the new family from complete samples only. Let $x_1,\ldots,x_n$ be the observed values from the EGG-G distribution with parameters θ,γ,α and $\boldsymbol{\xi}$. Let $\Theta=(\theta,\gamma,\alpha,\boldsymbol{\xi})^\top$ be the $r\times 1$ parameter vector. The total log-likelihood function for Θ is given by

$$\begin{aligned} \ell_n=\ell_n(\Theta)&=n\,\log(\alpha\,\theta)+\sum_{i=1}^{n}\log\left[g(x_i;\boldsymbol{\xi})\right]-(\gamma+1)\sum_{i=1}^{n}\log\left[1-G(x\,;\,\boldsymbol{\xi})\right] \\ &+\sum_{i=1}^{n}\log(1-t_i)+(\alpha-1)\sum_{i=1}^{n}\log(t_i), \end{aligned} \tag{21}$$

where $t_i = 1 - e^{\frac{\theta}{\gamma}\{1-[1-G(x_i;\boldsymbol{\xi})]^{-\gamma}\}}$.

The log-likelihood can be maximized either directly or by solving the nonlinear likelihood equations obtained by differentiating (21). The components of the score function $U_n(\Theta) = \left(\partial\ell_n/\partial\theta,\, \partial\ell_n/\partial\gamma,\, \partial\ell_n/\partial\alpha,\, \partial\ell_n/\partial\boldsymbol{\xi}\right)^\top$ are

$$\frac{\partial\ell_n}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n}\frac{t_i^{(\theta)}}{1-t_i} + (\alpha-1)\sum_{i=1}^{n}\frac{t_i^{(\theta)}}{t_i},$$

$$\frac{\partial\ell_n}{\partial\gamma} = -\sum_{i=1}^{n}\log[1-G(x_i;\boldsymbol{\xi})] - \sum_{i=1}^{n}\frac{t_i^{(\gamma)}}{1-t_i} + (\alpha-1)\sum_{i=1}^{n}\frac{t_i^{(\gamma)}}{t_i},$$

$$\frac{\partial\ell_n}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log(t_i)$$

and

$$\frac{\partial\ell_n}{\partial\boldsymbol{\xi}} = \sum_{i=1}^{n}\frac{g^{(\xi)}(x;\boldsymbol{\xi})}{g(x;\boldsymbol{\xi})} + (\gamma+1)\sum_{i=1}^{n}\frac{g^{(\xi)}(x;\boldsymbol{\xi})}{1-G(x;\boldsymbol{\xi})} - \sum_{i=1}^{n}\frac{t_i^{(\xi)}}{1-t_i} + (\alpha-1)\sum_{i=1}^{n}\frac{t_i^{(\xi)}}{t_i},$$

where $t_i^{(\zeta)} = C_\zeta\, e^{\frac{\theta}{\gamma}\{1-[1-G(x;\boldsymbol{\xi})]^{-\gamma}\}}$,

$$C_\zeta = \begin{cases} -\frac{1}{\gamma},\ \text{if}\ \zeta = \theta, \\ \frac{\theta}{\gamma^2},\ \text{if}\ \zeta = \gamma, \\ \frac{\gamma\, G^{(\boldsymbol{\xi})}(x;\boldsymbol{\xi})}{[1-G(x;\boldsymbol{\xi})]^{\gamma-1}},\ \text{if}\ \zeta = \boldsymbol{\xi} \end{cases}$$

and $h^{(\boldsymbol{\xi})}(\cdot)$ denotes the derivative of the function h with respect to $\boldsymbol{\xi}$. Often with lifetime data and reliability studies, one encounters censoring. A very simple random censoring mechanism very often realistic is one in which each individual i is assumed to have a lifetime X_i and a censoring time C_i, where X_i and C_i are independent random variables. Suppose that the data consist of n independent observations $x_i = \min(X_i, C_i)$ and $\delta_i = I(X_i \le C_i)$ is such that $\delta_i = 1$ if X_i is a time to event and $\delta_i = 0$ if it is right censoblack for $i = 1, \ldots, n$. The censoblack likelihood $L(\Theta)$ for the model parameters blackuces to

$$L(\Theta) \propto \prod_{i=1}^{n}\left[f(x_i;\theta,\gamma,\alpha,\boldsymbol{\xi})\right]^{\delta_i}\left[S(x_i;\theta,\gamma,\alpha,\boldsymbol{\xi})\right]^{1-\delta_i},$$

where $f(x;\theta,\gamma,\alpha,\boldsymbol{\xi})$ is given by (4) and $S(x;\theta,\gamma,\alpha,\boldsymbol{\xi})$ is the survival function computed by (3).

11. Simulation study

Here, we give a simulation study in order to assess the MLEs described in Section 13. One of advantages of the EGG family of distributions is that its cdf has tractable analytical form. This fact implies in a simple random number generator (RNG). We use the algorithm.

(1) Generate $U \sim U(0, 1)$.

(2) Specify a function $G(\cdot;\cdot)$ as defined in Section 3.
(3) Obtain an outcome of X by

$$X = G^{-1}\left(1 - \left[1 - \frac{\gamma}{\theta}\log(1 - U^{\frac{1}{\alpha}})\right]^{\frac{-1}{\gamma}}; \boldsymbol{\xi}\right).$$

The EGGB fitted density is plotted in Figure 6.

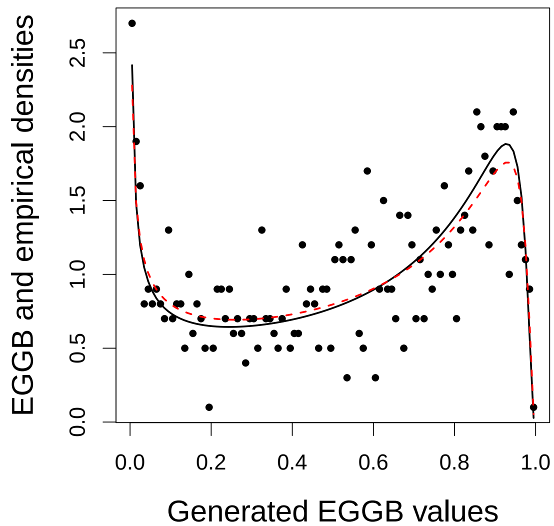


Figure 6. Plots of the theoretical (black and solid), fitted (black and long dashes) and empirical (points) density function for the EGGB(1/2, 1, 1, 1/2, 1/2) distribution.

We perform a simulation study in order to assess the influence of the additional parameters (α, θ and γ) on the difference between the theoretical and fitted curves associated with the EGGB distribution. To that end, we consider 1,000 Monte Carlo's replications and, on each replication, the mean square error (MSE) between the fitted and empirical densities is quantified such as a goodness-of-fit criterion. The current simulation process is conducted following the steps:

(1) Simulated EGGB distributed data of $N \in \{50, 100, 150\}$ are obtained by means of the EGG RNG.
(2) Fixing the baseline vector of parameters as $\boldsymbol{\xi} = (\beta_1, \beta_2) = (1/2, 1/2)$ for the EGGB distribution, respectively. Three scenarios are consideblack: (a) $\theta = 1/2$, $\gamma = 1$ and $\alpha \in \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$; (b) $\theta = 1/2$, $\gamma \in \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ and $\alpha = 1$; and (c) $\theta \in \{1,\ 1.5,\ 2,\ 2.5,\ 3,\ 3.5,\ 4,\ 4.5,\ 5\}$, $\gamma = 1$ and $\alpha = 1$.
(3) The generated data is submitted to the ML estimation to obtain the parameter estimates and the estimated EGGB pdf $\hat{f}_{\text{EGGB}}(\cdot)$.
(4) The MSEs between the exact and estimated pdfs are computed.

Figure 7 displays the relationship between the shape parameters and the MSEs. Based on the asymptotic properties of the MLEs, the MSEs decrease when the sample size increases (as expected). It is noticeable that the estimation of the parameter θ is the most hard situation, whereas the estimation of α is the most tractable.

In general, for a given sample size, the estimation of the EGGB extra parameters γ and α tends not to be influenced by increasing the parameter values.

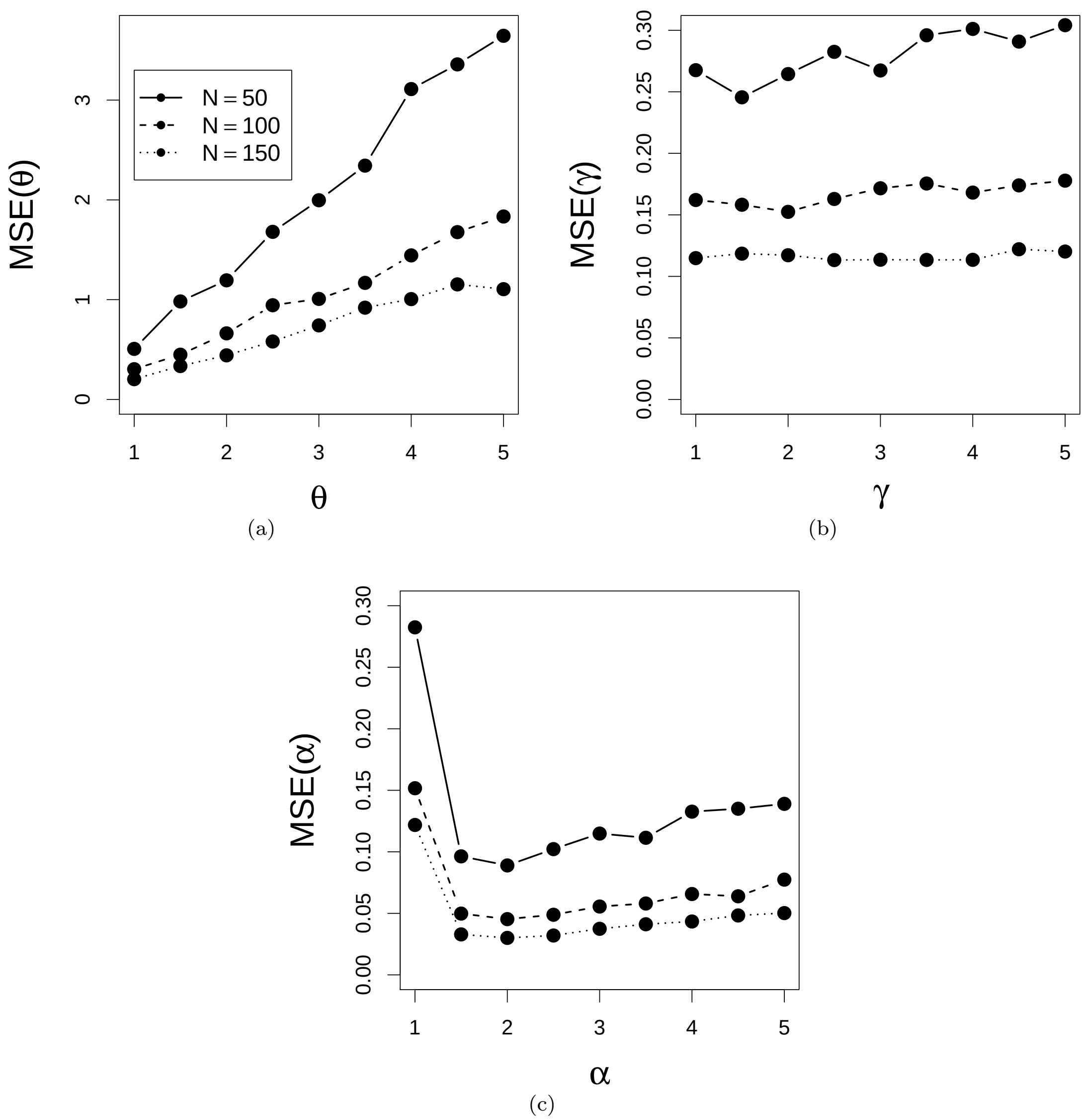


Figure 7. MSEs for some EGGB parameter values.

12. Applications

In this section, two applications to real data are performed in order to illustrate the potentiality of the EGGB distribution.

The first application consists the total milk production in one hundred seven SINDI race cows on the first birth after to calve.These cows are property of the Carnauba farm which belongs to the Agropecuaria Manoel Dantas Ltda (AMDA), located in the Taperoa City, Paraiba (Brazil). The original data are not in the interval $(0, 1)$ and it was made the transformation given by $x_i = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)}$, for $i = 1, ..., 107$. These data are presented in Cordeiro and Brito (2012). These data have already been used in Cordeiro and Brito (2012)presented evidence that such set of data can be well described by the beta power distribution.

As a second application, we describe the proportion of crude oil converted to gasoline after distillation and fractionation, discussed by Prater (1956). Ferrari and Cribari-Neto (2004) used this data as the response variable of a quantile regression. These data can be found in the "betareg" package of R statistical software with name "GasolineYield".

Table 2. Parameter estimates (standard errors) and $-\hat{\ell}$ values of GoF for the first application

Models	Estimates (SEs)	$-\hat{\ell}$	W^*	A^*
$EGGB$ $(\theta, \gamma, \alpha, \beta_1, \beta_2)$	0.44(0.02), 1.22(0.04), 3.76(0.36), 0.22(0.01), 0.64(0.02)	-28.10	0.067	0.431
B (β_1, β_2)	2.41(0.17), 2.82(0.20)	-23.77	0.228	1.385
EB $(\alpha, \beta_1, \beta_2)$	0.04(4×10^{-3}), 40.10(2.62), 10.61(1.17)	-26.35	0.141	0.857
BB (a, b, β_1, β_2)	0.42(0.03), 86.11(12.75), 3.94(0.21), 0.07(0.01)	-27.96	0.078	0.503
KwB (a, b, β_1, β_2)	6.88(0.12), 223.83(21.63), 0.23(6×10^{-3}), 0.18(4×10^{-3})	-27.85	0.078	0.512
McB $(a, b, c, \beta_1, \beta_2)$	0.31(0.02), 25.76(4.34), 0.29(0.01), 18.45(1.06), 0.03(0.01)	-27.83	0.084	0.536
GoB $(\theta, \gamma, \beta_1, \beta_2)$	22.80(2.20), 7.01(2.07), 1.70(0.08), 0.10(7×10^{-3})	-27.38	0.097	0.613

In the both examples, we shall compare the $EGGB$ distribution with the following models: the beta (B), the beta beta (BB) (Zografos and Balakrishnan, 2009), the exponentiated beta (EB) (Nadarajah, 2005) Kumaraswamy beta (KwB) (Cordeiro and de Castro, 2011), McDonald beta (Alexander et al., 2012) and Gompertz beta (Cordeiro et al., 2014) distributions.

The measures of goodness-of-fit including the AndersonDarling (A^*) and Cramervon Mises(W^*) statistics are computed to compare the fitted models. The statistics A^* and W^* are described in Evans et al. (2008). They showed W^* and A^* can be calculated as

$$W^* = \sum_{i=1}^{n} \left(\hat{F}(x_{(i)}) - \frac{i - 0.5}{n} \right)^2 + \frac{1}{12n}$$

and

$$A^* = -\sum_{i=1}^{n} \frac{2i-1}{n} \left(\ln\left(\hat{F}(x_{(i)})\right) + \ln\left(1 - \hat{F}(x_{(n+1-i)})\right)\right) - n.$$

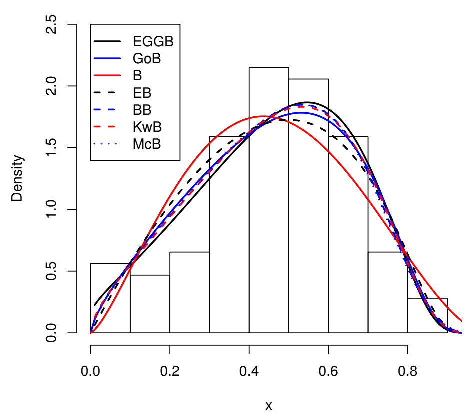
where n is the sample size. In general, smaller values of these statistics indicate better fits to the data sets.

Tables 2 and 3 present the MLEs and their corresponding standard errors (in parentheses) of the model parameters as well as values for $-$ maximized log likelihood ($\hat{\ell}$), A^* and W^* .

Figure 8 shows empirical and fitted densities associated with the two set of data. The $EGGB$ model was indicated as the best fit for both data set by means of all GoFs. Tables 2 and 3 present used goodness-of-fit (GoF) values for both applications.

Table 3. Parameter estimates (standard errors) and $-\hat{\ell}$ values of GoF for the second application

Models	Estimates (SEs)	$-\hat{\ell}$	W^*	A^*
$EGGB$ $(\theta,\gamma,\alpha,\beta_1,\beta_2)$	0.10(0.02), 0.08(0.03), 0.53(0.09), 6.28(0.91), 49.67(4.79)	-29.11	0.027	0.187
B (β_1,β_2)	2.46(0.27), 10.11(1.21)	-28.38	0.044	0.283
EB (α,β_1,β_2)	0.06(0.01), 27.91(2.49), 44.27(5.44)	-28.75	0.035	0.226
BB (a,b,β_1,β_2)	0.09(0.01), 1.29(0.55), 21.03(1.85), 32.63(3.96)	-28.76	0.035	0.224
KwB (a,b,β_1,β_2)	0.25(0.01), 10.23(1.80), 7.41(0.48), 1.98(0.42)	-28.51	0.042	0.263
McB (a,b,c,β_1,β_2)	0.07(0.01), 2.83(1.57), 10.74(1.69), 2.57(0.25), 7.86(0.83)	-28.83	0.033	0.218
GoB $(\theta,\gamma,\beta_1,\beta_2)$	0.09(0.01), 0.04(0.01), 3.33(0.46), 52.77(0.60)	-28.96	0.034	0.219



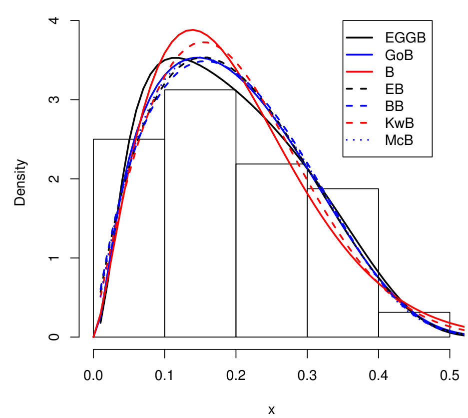


Figure 8. (Left panel): histogram of Example 1 and fitted distributions, (Right panel): histogram of Example 2 and fitted distributions.

13. Conclusions

The distributions for modeling data with any support in the statistical literature are numerous. We define the exponentiated Gompertz generated family in order to provide great flexibility to any continuous distribution by adding three extra shape parameters. Some special models are briefly discussed. We investigate general structural properties of the new family including shapes, ordinary and incomplete moments, quantile and generating functions, Bonferroni and Lorenz curves, Shannon and Rényi entropies and order statistics. The model parameters are estimated by maximum likelihood. A bivariate extension is proposed. A simulation study is performed to assess the influence of the additional parameters on the difference between the theoretical and fitted curves of the new family. Its usefulness are illustrated by means of two applications to real data. The results indicate that the exponentiated Gompertz generated beta (EGGB) model is a good distribution for describing these data according to six goodness-of-fit measures.

Appendix

Appendix A. An expansion for the EGG quantile function

If the G qf, say $Q_G(u)$, does not have a closed-form expression, this function can usually be expressed in terms of a power series

$$Q_G(u) = \sum_{i=0}^{\infty} a_i\, u^i, \tag{A1}$$

where the coefficients a_i are suitably chosen real numbers which depend on the parameters of the G distribution. For several important distributions, such as the normal, Student t, gamma and beta distributions, $Q_G(u)$ does not have explicit expressions but it can be expanded as in equation (A1).

Next, we derive an expansion for the argument of $Q_G(\cdot)$ in (6)

$$A = 1 - \left[1 - \frac{\gamma}{\theta}\log(1 - u^{\frac{1}{\alpha}})\right]^{\frac{-1}{\gamma}}.$$

First, for $z \in (0,1)$ and any real non-integer α, we can write

$$z^{\alpha} = \sum_{r=0}^{\infty} s_r(\alpha)\, z^r, \tag{A2}$$

where $s_r(\alpha) = \sum_{m=r}^{\infty}(-1)^{m+r}\binom{\alpha}{m}\binom{m}{r}$. Second, using (A2) and expanding the binomial term, we obtain

$$A = 1 - \sum_{k=0}^{\infty} s_k(-\gamma^{-1}) \sum_{j=0}^{k}(-1)^j \binom{k}{j}\left(\frac{\gamma}{\theta}\right)^j \log^j[1 - u^{1/\alpha}]. \tag{A3}$$

Further, we use the expansion

$$[\log(1 - u^{1/\alpha})]^j = \left[-u\sum_{r=0}^{\infty}\frac{u^{r/\alpha}}{(r+1)}\right]^j. \tag{A4}$$

Now, we use throughout the paper a result of Gradshteyn and Ryzhik (2000, Section 0.314) for a power series raised to a positive integer n (for $n \geq 1$)

$$Q_G(u)^n = \left(\sum_{i=0}^{\infty} a_i\, u^i\right)^n = \sum_{i=0}^{\infty} c_{n,i}\, u^i, \tag{A5}$$

where the coefficients $c_{n,i}$ (for $i = 1, 2, \ldots$) are obtained from the recurrence equation (with $c_{n,0} = a_0^n$)

$$c_{n,i} = (i\, a_0)^{-1}\sum_{m=1}^{i}[m(n+1) - i]\, a_m\, c_{n,i-m}. \tag{A6}$$

Clearly, $c_{n,i}$ can be determined from $c_{n,0}, \ldots, c_{n,i-1}$ and then from the quantities $a_0, \ldots, a_i$. So, we can write equation (A4) as

$$[\log(1-u^{1/\alpha})]^j = \sum_{r=0}^{\infty}(-1)^r\, d_{j,r}\, u^{r/\alpha+j}, \tag{A7}$$

where the coefficients $d_{j,r}$ come from equations (A5) and (A6) as $d_{j,r} = r^{-1}\sum_{m=1}^{r}\frac{[m(j+1)-r]}{(m+1)}\, d_{j,r-m}$ for $r \geq 0$ and $d_{j,0} = 1$. Combining equations (A3) and (A4), we obtain

$$A = 1 - \sum_{r=0}^{\infty}\sum_{j=0}^{k} t_{r,j}\, u^{r/\alpha+j},$$

where $t_{r,j} = \sum_{k=0}^{\infty}(-1)^{j+r}\, s_k(-\gamma^{-1})\binom{k}{j}\left(\frac{\gamma}{\theta}\right)^j d_{j,r}$. Using again (A2), we have $u^{\frac{r}{\alpha}-j} = \sum_{p=0}^{\infty} s_p(\frac{r}{\alpha}+j)\, u^p$ and inserting in the last equation gives

$$A = \sum_{p=0}^{\infty}\delta_p\, u^p,$$

where $\delta_p = \sum_{r=0}^{\infty}\sum_{j=0}^{k} t_{r,j}\; s_p(r/\alpha+j)$ for $p \geq 1$ and $\delta_0 = 1-\sum_{r=0}^{\infty}\sum_{j=0}^{k} t_{r,j}\; s_0(r/\alpha-j)$. Then, for any baseline G distribution, we obtain the EGG qf

$$Q(u) = Q_G\left(\sum_{p=0}^{\infty}\delta_p\, u^p\right) = \sum_{i=0}^{\infty} a_i\left(\sum_{p=0}^{\infty}\delta_p\, u^p\right)^i,$$

and using (A5) and (A6), we obtain the expansion of Theorem 2

$$Q(u) = \sum_{m=0}^{\infty} e_p\, u^p,$$

where $e_p = \sum_{i=0}^{\infty} a_i\, q_{i,p}$, and for $i \geq 0$, $q_{i,0} = \delta_0^i$ and (for $p > 1$)

$$q_{i,p} = (p\,\delta_0)^{-1}\sum_{n=1}^{p}[n(i+1)-p]\,\delta_n\, q_{i,p-n}.$$

References

Alexander, C., Cordeiro, G.M., Ortega, E.M.M., and Sarabia, J.M. 2012. Generalized beta-generated distributions. Computational Statistics & Data Analysis, 56, 1880-1897.

Alizadeh, M., Cordeiro, G.M., Pinho, L.G.B. and Ghosh, I. 2016. The Gompertz-G family of distributions. Submitted.

Alzaatreh, A., Lee, C. and Famoye, F. 2013. A new method for generating families of continuous distributions. Metron, 71, 63-79.

Alzaghal, A., Famoye, F. and Lee, C. 2013. Exponentiated T-X family of distributions with some applications. International Journal of Statistics and Probability, 2, 1-31.

Amini, M., MirMostafaee, S.M.T.K. and Ahmadi, J. 2012. Log-gamma-generated families of distributions. Statistics, 1, 1-20.

Bourguignon, M., Silva, R.B. and Cordeiro, G.M. 2016. The Weibull-G family of probability distributions. Journal of Data Science, 12, 53-68.

Cintra, R.J., Frery, A.C. and Nascimento, A.D.C. 2013. Parametric and nonparametric tests for speckled imagery. Pattern Analysis and Applications, 16, 141-161.

Cordeiro, G. M., Alizadeh, M., and Ortega, E.M. 2014. The exponentiated half-logistic family of distributions: Properties and applications. Journal of Probability and Statistics, 2014, Article ID 864396. doi: 10.1155/2014/864396

Cordeiro, G.M., Alizadeh, M., Silva, R.B. and Ramires T.G. 2016. A new wider family of continuous models: The extended Cordeiro and De Castro family. Under review.

Cordeiro, G.M. and Brito, R.D.S. 2012. The beta power distribution. Brazilian Journal of Probability and Statistics, 26, 88-112.

Cordeiro, G.M. and de Castro, M. 2011. A new family of generalized distributions. Journal of Statistical Computation and Simulation, 81, 883-898.

Cordeiro, G.M., Ortega, E.M.M. and Silva, G.O. 2011. The exponentiated generalized gamma distribution with application to lifetime data. Journal of statistical computation and simulation, 81, 827-842.

Evans, D.L., Drew, J.H. and Leemis, L.M. 2008. The distribution of the kolmogorovsmirnov, cramervon mises, and andersondarling test statistics for exponential populations with estimated parameters. Communications in Statistics - Simulation and Computation, 37, 1396-1421.

El-Gohary, A., Alshamrani, A. and Al-Otaibi, A.N. 2013. The generalized Gompertz distribution. Applied Mathematical Modelling, 37, 13-24.

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

Ferrari, S.L.P., Cribari-Neto, F. 2004. Beta regression for modeling rates and proportions. Journal of Applied Statistics, 31, 799-815.

Gradshteyn, I.S. and Ryzhik, I.M. 2000. Table of Integrals, Series, and Products. Academic Press.

Gupta, R.C. and Gupta, R.D. (2007). Proportional reversed hazard rate model and its applications. Journal of Statistical Planning and Inference, 137, 3525-3536.

Gupta, R.D. and Gupta, R.C. 2008. Analyzing skewed data by power normal model. Test, 17, 197-210.

Gupta, R.D. and Kundu, D. 1999. Generalized exponential distributions. Australian and New Zealand Journal of Statistics, 41, 173-188.

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

Jones, M.C. 2009. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. Statistical Methodology, 6, 70-81.

Kenney, J.F. and Keeping, E.S. 1962. Kurtosis. Mathematics of Statistics, 3, 102-103.

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

Lemonte, A.J., Barreto-Souza, W. and Cordeiro, G.M. 2013. The exponentiated Kumaraswamy distribution and its log-transform. Brazilian Journal of Probability and Statistics, 27, 31-53.

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

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

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

Nadarajah, S. 2005. Exponentiated beta distributions. Computers & Mathematics with Applications, 49, 1029-1035.

Nadarajah, S., Cordeiro, G.M. and Ortega, E.M.M. 2013. The exponentiated Weibull distribution: A survey. Statistical Papers, 54, 839-877.

Nadarajah, S. and Kotz, S. 2006. The exponentiated type distributions. Acta Applicandae Mathematica, 92, 97-111.

Prater, N.H. 1956. Estimate gasoline yields from crudes. Petroleum Refiner, 35, 236-238.

Rényi, A. 1961. On measures of entropy and information. In 4th Berkeley Symposium on Mathematical Statistics and Probability, 1, 547-561.

Ristić, M.M. and Balakrishnan, N. 2012. The gamma-exponentiated exponential distribution. Journal of Statistical Computation and Simulation, 82, 1191-1206.

Shannon, C.E. 1948. A mathematical theory of communication. Bell System Technical Journal, 27, 379-423.

Torabi, H. and Hedesh, N.M. 2012. The gamma-uniform distribution and its applications. Kybernetika, 1, 16-30.

Torabi, H. and Montazeri, N.H. 2014. The logistic-uniform distribution and its applications. Communications in Statistics - Simulation and Computation, 43, 2551-2569.

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