

RESEARCH

Open Access



A useful extension of the Burr III distribution

Gauss M. Cordeiro¹, Antonio E. Gomes^{2*} , Cibele Q. da-Silva² and Edwin M. M. Ortega³

*Correspondence:

aegomes@unb.br

²Departamento de Estatística,
Universidade de Brasília, 70910-900
Brasília, DF, Brazil

Full list of author information is
available at the end of the article

Abstract

For any continuous baseline G distribution, Zografos and Balakrishnan (Statistical Methodology 6:344–362, 2009) introduced the gamma-generated family of distributions with an extra shape parameter. Based on this family, we define a new four-parameter extension of the Burr III distribution. It can have decreasing, unimodal and decreasing-increasing-decreasing hazard rate function. We provide a comprehensive account of some of its structural properties. We propose a new log-gamma Burr III regression model, which is a feasible alternative for modeling the four existing types of failure rates. Two applications to real data sets and a simulation study illustrate the performance of the new models.

Keywords: Estimation, Gamma distribution, Gamma-G family, Moment, Regression model

Introduction

Adding new shape parameters to expand a parent distribution plays a fundamental role to generate a larger family with a wide range of skewness and light or heavy tails. Several mathematical properties of the extended family may be easily explored using linear combination of exponentiated-G (“exp-G” for short) distributions. Further, this induction of parameters has been proved useful in investigating tail properties and also for improving the goodness-of-fit of the generator family. The well-known generators are the following ones: beta-G by Eugene et al. (2002), Kumaraswamy-G (Kw-G) by Cordeiro and de Castro (2011), McDonald-G (Mc-G) by Alexander et al. (2012), *gamma-G* by Zografos and Balakrishnan (2009), among others. Recently, several distribution generators have been proposed, for example, Alzaatreh et al. (2013) proposed the the T-X family of distributions, Cordeiro et al. (2014) introduced the the Lomax generator, Cordeiro et al. (2015) defined a new generalized Weibull family, Tahir et al. (2015) studied the odd generalized exponential family, Nofal et al. (2016) proposed the generalized transmuted-G family and Cordeiro et al. (2017) investigated the generalized odd log-logistic family.

Zografos and Balakrishnan (2009) proposed a family of univariate distributions generated by gamma random variables with an additional shape parameter to a parent model. For any baseline cumulative distribution function (cdf) $G(x)$, and $x \in \mathbb{R}$, they defined the

gamma-G family by the probability density function (pdf) $f(x)$ and cdf $F(x)$ (for $c > 0$) given by

$$f(x) = \frac{1}{\Gamma(c)} \{-\log[1 - G(x)]\}^{c-1} g(x) \tag{1}$$

and

$$F(x) = \frac{\gamma(c, -\log[1 - G(x)])}{\Gamma(c)} = \frac{1}{\Gamma(c)} \int_0^{-\log[1 - G(x)]} t^{c-1} e^{-t} dt, \tag{2}$$

respectively, where $g(x) = dG(x)/dx$, $\Gamma(c) = \int_0^\infty t^{c-1} e^{-t} dt$ denotes the gamma function, and $\gamma(c, z) = \int_0^z t^{c-1} e^{-t} dt$ denotes the incomplete gamma function. The *gamma-G* family has the same parameters of the *G* distribution plus an extra shape parameter c . The increase of one parameter is the price to pay for adding more flexibility to the generated model compared to *G*. For $c = 1$, Eq. (1) becomes the density function $g(x)$, which is a positive point. The parameter c can provide greater flexibility in the form of the generated distribution and, consequently, it can be a very useful model for fitting positive data.

For a random variable X with pdf (1), we have $X \stackrel{d}{=} G^{-1}(1 - e^{-Z})$, where $Z \sim \text{Gamma}(c, 1)$. If $c = 1$, then $Z \sim \text{exp}(1)$ and $X \stackrel{d}{=} G^{-1}(U)$, where $U \sim U(0, 1)$.

The Burr III (BIII) cumulative distribution is given by

$$G_{\alpha,\beta,s}(x) = \left[1 + \left(\frac{x}{s}\right)^{-\alpha}\right]^{-\beta} = \left[\frac{(x/s)^\alpha}{1 + (x/s)^\alpha}\right]^\beta, \tag{3}$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters and $s > 0$ is a scale parameter. The pdf corresponding to (3) is given by

$$g_{\alpha,\beta,s}(x) = \frac{\alpha \beta}{s (x/s)^{\alpha+1}} \left[\frac{(x/s)^\alpha}{1 + (x/s)^\alpha}\right]^{\beta+1}. \tag{4}$$

The BIII distribution has been used in various fields of sciences and its features extensively analyzed. It appeared under the name of the Dagum (1977) distribution in studies of income, wage and wealth distributions. For an excellent survey on its genesis and empirical applications, see Kleiber and Kotz (2003) and Kleiber (2008). It is known as the *inverse Burr* distribution (see, e.g., Klugman et al. 1998) in the actuarial literature and as the *kappa* distribution in the meteorological area (Mielke, 1973). This distribution has also been employed in finance, environmental studies, survival analysis and reliability theory (see Lindsay et al. 1996; Gove et al. 2008).

In this paper, we define and study the four-parameter *gamma Burr III* (GBIII) distribution by inserting (3) and (4) in the generator density (1). The GBIII density becomes

$$f(x) = \frac{\alpha \beta}{s (x/s)^{\alpha+1} \Gamma(c)} \left[\frac{(x/s)^\alpha}{1 + (x/s)^\alpha}\right]^{\beta+1} \left\{-\log\left(1 - \left[\frac{(x/s)^\alpha}{1 + (x/s)^\alpha}\right]^\beta\right)\right\}^{c-1}. \tag{5}$$

The BIII pdf is a special case of (5) when $c = 1$. The cdf of the GBIII distribution reduces to

$$F(x) = \frac{1}{\Gamma(c)} \gamma\left(c, -\log\left\{1 - \left[\frac{(x/s)^\alpha}{1 + (x/s)^\alpha}\right]^\beta\right\}\right). \tag{6}$$

Hereafter, we denote the GBIII random variable having pdf (5) by $X \sim GBIII(c, \alpha, \beta, s)$. The hazard rate function (hrf) of X is given by

$$h(x) = \frac{\alpha \beta s^{-1} (x/s)^{-(\alpha+1)} [1 + (x/s)^{-\alpha}]^{-(\beta+1)} \left\{ -\log \left(1 - [1 + (x/s)^{-\alpha}]^{-\beta} \right) \right\}^{c-1}}{\Gamma(c) - \gamma \left(c, -\log \left\{ 1 - [1 + (x/s)^{-\alpha}]^{-\beta} \right\} \right)} \quad (7)$$

Figure 1 displays some plots of the density and hrf of X for selected parameter values, respectively. The hrf of X can be decreasing, unimodal and decreasing-increasing-decreasing. The GBIII model can have either positive or negative skewness.

Inverting $F(x) = u$, we obtain the quantile function (qf) of X as

$$Q(u) = s \left[\left[1 - \exp(-Q^{-1}(c, 1 - u)) \right]^{-1/\beta} - 1 \right]^{-1/\alpha} \quad (8)$$

for $0 < u < 1$, where $Q^{-1}(c, u)$ is the inverse function of $Q(c, x) = 1 - \gamma(c, x) / \Gamma(c)$; see, for details, <http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/>. One can also use (8) for simulating GBIII variates: if U is a uniform random variable on the unit interval $(0, 1)$, then $X = Q(U)$ will be a GBIII random variable.

The rest of the paper is outlined as follows. In Section Structural properties of the GBIII distribution, we obtain some structural properties of the GBIII distribution and estimate the model parameters by maximum likelihood. We propose a new regression model based on the logarithm of this distribution in Section The log-gamma Burr III regression model. Two applications to real data and a simulation study are addressed in

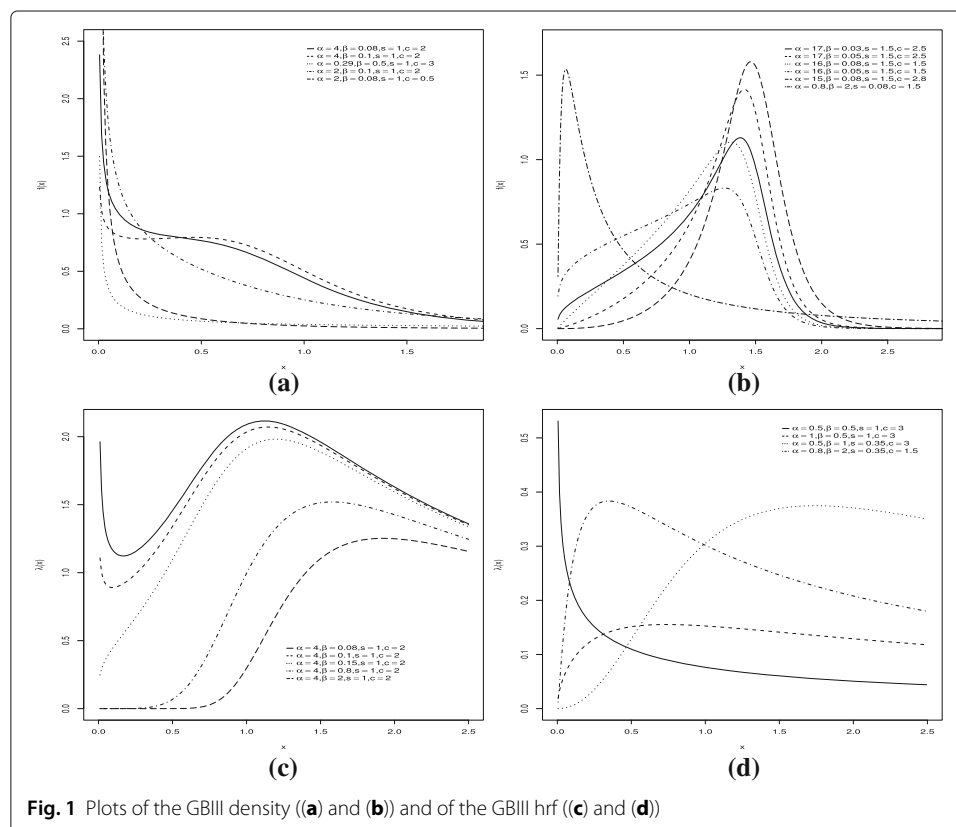


Fig. 1 Plots of the GBIII density ((a) and (b)) and of the GBIII hrf ((c) and (d))

Section Applications and simulation to prove empirically the flexibility of the new models. Finally, some conclusions are offered in Section Conclusions.

Structural properties of the GBIII distribution

In the following subsections, we obtain a linear representation for the density function and a power series for the qf of the new distribution and estimate its parameters. These expressions can be computed numerically in platforms such as MAPLE, MATHEMATICA, Ox and R using a large number in the upper limit instead of infinity.

Linear representation

Here, we demonstrate that the GBIII density can be expressed as a linear combination of BIII densities. The binomial coefficient generalized to real arguments is given by $\binom{x}{y} = \Gamma(x + 1)/[\Gamma(y + 1)\Gamma(x - y + 1)]$. For any real parameter $c > 0$, the convergent series holds (<http://functions.wolfram.com/ElementaryFunctions/Log/06/01/04/03/>)

$$\{-\log [1 - G_{\alpha,\beta,s}(x)]\}^{c-1} = (c-1) \sum_{k=0}^{\infty} \binom{k+1-c}{k} \sum_{j=0}^k \frac{(-1)^{j+k} \binom{k}{j} p_{j,k}}{(c-1-j)} G_{\alpha,\beta,s}(x)^{c+k-1}, \tag{9}$$

where $G_{\alpha,\beta,s}(x)^{c+k-1} = G_{\alpha,(c+k-1)\beta,s}(x)$ and the constants $p_{j,k}$ can be determined recursively (for $j \geq 0$) as

$$p_{j,k} = k^{-1} \sum_{m=1}^k \frac{(-1)^k [m(j+1) - k]}{(m+1)} p_{j,k-m}$$

for $k = 1, 2, \dots$ and $p_{j,0} = 1$.

For a real parameter $c > 0$, we define

$$b_k = \frac{\binom{k+1-c}{k}}{(c+k)\Gamma(c-1)} \sum_{j=0}^k \frac{(-1)^{j+k} p_{j,k}}{(c-1-j)} \binom{k}{j}.$$

By using this result, the pdf $f(x)$ can be expressed as a linear combination

$$f(x) = \sum_{k=0}^{\infty} b_k g_{\alpha,(c+k)\beta,s}(x), \tag{10}$$

where $g_{\alpha,(c+k)\beta,s}(x)$ denotes the BIII pdf in (4) with parameters α , $(c+k)\beta$ and s . So, several mathematical properties of the GBIII distribution can be obtained from those of the BIII distribution using (10) in platforms such as MAPLE and MATHEMATICA.

Equation (10) holds for any real parameter $c > 0$ and then some mathematical properties of the new model are valid in the same parameter space, where those properties of the BIII model hold. Evidently, the integrals for the ordinary and incomplete moments and generating function of X can also be computed numerically in Ox and R.

Quantile expansion

We use throughout the paper an equation by Gradshteyn and Ryzhik (2000, Section 0.314) for a power series raised to a positive integer n

$$\left(\sum_{i=0}^{\infty} a_i u^i\right)^n = \sum_{i=0}^{\infty} c_{n,i} u^i, \tag{11}$$

where the coefficients $c_{n,i}$ (for $n \geq 0$ and $i = 1, 2, \dots$) are determined from the recurrence equation using any algebraic or numerical software

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

and $c_{n,0} = a_0^n$.

If $V \sim \text{Gamma}(c, 1)$, the qf $Q_V(u)$ of V admits the power series

$$Q_V(u) = \sum_{i=0}^{\infty} m_i [\Gamma(c+1) u]^{i/c},$$

where $m_0 = 0, m_1 = 1$ and any coefficient m_{i+1} (for $i \geq 1$) can be determined by the cubic recurrence equation

$$m_{i+1} = \frac{1}{i(c+i)} \left\{ \sum_{r=1}^i \sum_{s=1}^{i-s+1} s(i-r-s+2) m_r m_s m_{i-r-s+2} - \Delta(i) \sum_{r=2}^i r[r-c-(1-c)(i+2-r)] m_r m_{i-r+2} \right\},$$

where $\Delta(i) = 0$ if $i < 2$ and $\Delta(i) = 1$ if $i \geq 2$. The first few coefficients are $m_2 = 1/(c+1), m_3 = (3c+5)/[2(c+1)^2(c+2)], \dots$. Let $m_0 = 0$ and define $q_i = m_{i+1} \Gamma(c+1)^{(i+1)/2}$ (for $i = 0, 1, 2, \dots$).

For $m \geq 1$, we define $J_m = \{(i, k), i+k = m; i, k = 0, 1, 2, \dots\}$. Then, we can rewrite the GBIII qf from (8) as

$$Q(u) = s \left\{ \left[1 - \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left(\sum_{i=0}^{\infty} q_i u^{(i+1)/c} \right)^k \right]^{-1/\beta} - 1 \right\}^{-1/\alpha}.$$

Next, using Eq. (11), we obtain

$$Q(u) = s \left\{ \left[1 - \sum_{k,i=0}^{\infty} \frac{(-1)^k d_{k,i}}{k!} u^{(i+k)/c} \right]^{-1/\beta} - 1 \right\}^{-1/\alpha}, \tag{12}$$

where (for $k \geq 0$) $d_{k,0} = q_0^k$ and, for $i = 1, 2, \dots, d_{k,i} = (i q_0)^{-1} \sum_{j=1}^i [j(k+1) - i] q_j d_{k,i-j}$.

Further, we have

$$1 - \sum_{k,i=0}^{\infty} \frac{(-1)^k d_{k,i}}{k!} u^{(i+k)/c} = \sum_{m=1}^{\infty} v_m u^{m/c},$$

and then we rewrite Eq. (12) as

$$Q(u) = s \left\{ \left[\sum_{m=1}^{\infty} v_m u^{m/c} \right]^{-1/\beta} - 1 \right\}^{-1/\alpha},$$

where (for $m \geq 1$)

$$v_m = \sum_{\substack{k,i \geq 0 \\ (i,k) \in J_m}}^{\infty} \frac{(-1)^{k+1} d_{k,i}}{k!}.$$

We now consider the delta expansion given by

$$H(z)^\gamma = 1 + \sum_{n=1}^{\infty} \frac{\tau_n}{n!} [H(z) - 1]^n,$$

where $\tau_n = \prod_{j=0}^{n-1} (\gamma - j)$. Based on this expansion, we can write $Q(u)$ as

$$Q(u) = s \left\{ \sum_{n=1}^{\infty} \frac{\tau_n}{n!} \left(\sum_{m=0}^{\infty} v_m u^{m/c} \right)^n \right\}^{-1/\alpha},$$

where $v_0 = -1$. Then, using again (11), we obtain

$$Q(u) = s \left[\sum_{m=0}^{\infty} s_m u^{m/c} \right]^{-1/\alpha}, \tag{13}$$

where (for $m \geq 0$) $s_m = \sum_{n=1}^{\infty} \frac{\tau_n}{n!} \delta_{n,m}$, and (for $n \geq 1$) $\delta_{n,m} = (m v_0)^{-1} \sum_{p=1}^m [p(n+1) - m] v_m \delta_{n,m-p}$, and $\delta_{n,0} = v_0^n$.

Hence, Eq. (13) reveals that the GBIII qf can be expressed as a power series raised to $-1/\alpha$. Let $W(\cdot)$ be any integrable function in a positive real line. We can write

$$\int_{-\infty}^{\infty} W(x) f(x) dx = \int_0^1 W \left(s \left[\sum_{m=0}^{\infty} s_m u^{m/c} \right]^{-1/\alpha} \right) du. \tag{14}$$

Equations (13) and (14) are the main results of this section since we can obtain from them various GBIII mathematical quantities (moments, generating function, etc). In fact, some of them follow by using the right integral for special $W(\cdot)$ functions, which are sometimes simpler than if they are based on the left integral.

Maximum likelihood estimation

The maximum likelihood estimates (MLEs) enjoy desirable properties and can be used when constructing confidence intervals. First-order asymptotic theory for these estimates delivers simple approximations that may work well in finite samples. In this section, we consider the estimation of the unknown parameters of the GBIII distribution by maximum likelihood. Let $x_1 \dots, x_n$ be a random sample from (5) and $\theta = (c, \alpha, \beta, s)^T$ be the parameter vector. The log-likelihood function $\ell(\theta) = \log L(\theta)$ for θ is given by

$$\begin{aligned} \ell(\theta) = & -n \log \Gamma(c) + n \log \beta + n \log \alpha + n \alpha \log s \\ & -(\alpha + 1) \sum_{i=1}^n \log x_i - (\beta + 1) \sum_{i=1}^n \left[1 + \left(\frac{x_i}{s} \right)^{-\alpha} \right] \\ & + (c - 1) \sum_{i=1}^n \log \left\{ -\log \left[1 - \left[1 - \left(\frac{x_i}{s} \right)^{-\alpha} \right]^{-\beta} \right] \right\}. \end{aligned} \tag{15}$$

Maximization of (15) can be performed by using well established routines like `nllminb` or `optimize` in the R statistical package. The routines are able to locate the maximum in all cases if we take different starting values for the parameters. However, it is desirable to have reasonable starting values, which can be chosen using the estimates from the fitted BIII distribution.

The MLEs in θ , say $\hat{\theta}$, can also be determined numerically as simultaneous solutions of the equations $\partial \ell(\theta) / \partial \theta = \mathbf{0}$. For interval estimation of the components in θ , we require the observed information matrix for θ , say $-\ddot{\mathbf{L}}(\theta) = \{-L_{rv}\}$ ($r, v = c, \alpha, \beta, s$), whose elements can be obtained from the authors upon request.

The log-gamma Burr III regression model

Different forms of regression models have been studied in survival analysis. Among them, the location-scale regression model (Lawless 2003) is distinguished since it is frequently used in clinical trials. We propose a new location-scale regression model based on the logarithm of the GBIII random variable named the *log-gamma Burr III* (LGBIII) regression model as a feasible alternative for modeling the four existing types of hazard rates.

If X is a random variable having the GBIII density function (5), we define $Y = \log(X)$. The pdf of Y obtained by replacing $s = e^\mu$ and $\alpha = 1/\sigma$ is given by

$$f(y) = \frac{\beta \exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]}{\sigma \Gamma(c)} \left[\frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{1 + \exp\left(\frac{y-\mu}{\sigma}\right)} \right]^{\beta+1} \left[-\log \left\{ 1 - \left[\frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{1 + \exp\left(\frac{y-\mu}{\sigma}\right)} \right]^\beta \right\} \right]^{c-1}, \tag{16}$$

where $y \in \mathbb{R}$, $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $c > 0$ and $\beta > 0$ are shape parameters. We refer to Eq. (16) as the LGBIII distribution, say $Y \sim \text{LGBIII}(c, \beta, \sigma, \mu)$. The density of the random variable $Z = (Y - \mu)/\sigma$ follows from (16).

The survival function of Y is

$$S(y) = 1 - \frac{1}{\Gamma(c)} \gamma \left(c, -\log \left\{ 1 - \left[\frac{\exp\left(\frac{y-\mu}{\sigma}\right)}{1 + \exp\left(\frac{y-\mu}{\sigma}\right)} \right]^\beta \right\} \right). \tag{17}$$

In many practical applications, parametric models with explanatory variables are used to estimate univariate survival functions for censored data. A parametric model that provides a good fit to lifetime data tends to yield more precise estimates for the quantities of interest. Based on the LGBIII density, we propose a linear location-scale regression model linking the response variable y_i and the explanatory variable vector $\mathbf{v}_i^T = (v_{i1}, \dots, v_{ip})$ as

$$y_i = \mathbf{v}_i^T \boldsymbol{\tau} + \sigma z_i, \quad i = 1, \dots, n, \tag{18}$$

where the random error z_i has density function (16) with $\mu = 0$ and $\sigma = 1$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)^T$, $\sigma > 0$, $c > 0$ and $\beta > 0$ are unknown parameters. The parameter $\mu_i = \mathbf{v}_i^T \boldsymbol{\tau}$ is the location of y_i . The location parameter vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ is represented by a linear model $\boldsymbol{\mu} = \mathbf{V}\boldsymbol{\tau}$, where $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)^T$ is a known model matrix. The LGBIII regression model (18) opens new possibilities for fitting many different types of data. For example, it contains as sub-model the log-Burr III (LBIII) regression model when $c = 1$.

Consider a sample $(y_1, \mathbf{v}_1), \dots, (y_n, \mathbf{v}_n)$ of n independent observations, where each random response is defined by $y_i = \min\{\log(x_i), \log(c_i)\}$. We assume non-informative censoring such that the observed lifetimes and censoring times are independent. Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. Conventional likelihood estimation techniques can be applied here. The log-likelihood function for $\boldsymbol{\theta} = (c, \beta, \sigma, \boldsymbol{\tau}^T)^T$ from model (18) is given by $l(\boldsymbol{\theta}) = \sum_{i \in F} l_i(\boldsymbol{\theta}) +$

$\sum_{i \in C} l_i^{(c)}(\theta)$, where $l_i(\theta) = \log[f(y_i)]$, $l_i^{(c)}(\theta) = \log[S(y_i)]$, $f(y_i)$ is the density (16) and $S(y_i)$ is the survival function (17) of Y_i . The total log-likelihood function for θ reduces to

$$\begin{aligned}
 l(\theta) = & r \log \left(\frac{\beta}{\sigma \Gamma(c)} \right) - \sum_{i \in F} z_i + (\beta + 1) \sum_{i \in F} z_i - (\beta + 1) \sum_{i \in F} \log[1 + \exp(z_i)] \\
 & + (c - 1) \sum_{i \in F} \log \left\{ -\log \left[1 - \left(\frac{\exp(z_i)}{1 + \exp(z_i)} \right)^\beta \right] \right\} \\
 & + \sum_{i \in C} \log \left\{ \frac{1}{\Gamma(c)} \gamma \left(c, -\log \left\{ 1 - \left[\frac{\exp(z_i)}{1 + \exp(z_i)} \right]^\beta \right\} \right) \right\}, \tag{19}
 \end{aligned}$$

where $z_i = (y_i - \mathbf{v}_i^T \boldsymbol{\tau})/\sigma$ and r is the number of the uncensored observations (failures). The MLE $\hat{\theta}$ can be evaluated by maximizing (19). We use the NLMixed procedure in SAS to obtain $\hat{\theta}$. Initial values for $\boldsymbol{\tau}$ and σ are taken from the fitted LBIII regression model (with $c = 1$). We can fit the LBIII model to the uncensored observations only and then take the parameter estimates as initial values to fit the LGBIII regression model.

The estimated survival function for y_i follows from the fitted LGBIII model ($\hat{z}_i = (y_i - \mathbf{v}_i^T \hat{\boldsymbol{\tau}})/\hat{\sigma}$) as

$$S(y_i; \hat{c}, \hat{\beta}, \hat{\sigma}, \hat{\boldsymbol{\tau}}^T) = \frac{1}{\Gamma(\hat{c})} \gamma \left(\hat{c}, -\log \left\{ 1 - \left[\frac{\exp(\hat{z}_i)}{1 + \exp(\hat{z}_i)} \right]^{\hat{\beta}} \right\} \right). \tag{20}$$

Under first-order asymptotic theory, the $(p + 3) \times (p + 3)$ asymptotic covariance matrix $K(\theta)^{-1}$ of $\hat{\theta}$, where $K(\theta)$ is the expected information matrix for θ , can be approximated by the inverse of the observed information matrix $-\ddot{L}(\theta)$. The elements of this matrix can be computed numerically to construct approximate confidence intervals for the parameters in θ . We can use likelihood ratio (LR) statistics for comparing some sub-models with the LGBIII model in the classical way.

Applications and simulation

Application of GBIII to cigarettes data

In order to illustrate the estimation results in Section 3, we work with carbon monoxide measurements made in several brands of cigarettes in 1994. The data have been collected by the Federal Trade Commission (FTC), which is an independent agency of the United States Government, whose main mission is the promotion of consumer protection.

The data can be found at <http://www.econdataus.com/cigrs94.html> and contain $n = 384$ observations. We analyze the carbon monoxide (CO), measured in milligrams per cigarette, from several cigarette brands. We fit the GBIII, BIII and other sub-models to these data by the method of maximum likelihood. We also fit two more models to the current data: the beta Burr III (BBIII) (Gomes et al. 2013) and beta Weibull (BW) (Lee et al. 2007) distributions. The MLEs of the parameters, their standard errors (SEs) and the Akaike Information Criterion (AIC) for the fitted models are listed in Table 1. The required numerical evaluations are implemented using the `nlminb` function of the R language. When the GBIII model (5) is evaluated at $c = 1, \beta = 1$ it gives rise to the log-logistic (LL) model. For $\alpha = 1$ and $c = 1$, we have as a special case the exponentiated log-logistic (ELL) model. From the figures in Table 1, we note that the smallest AIC value corresponds to the GBIII model.

Table 1 MLEs of the model parameters for the cigarettes data, the corresponding SEs (given in parentheses) and the AIC measure

Model	α	β	s	c	b	γ	AIC
GBIII	17.0973 (0.1153)	0.0282 (0.0006)	1.4992 (0.0028)	2.6338 (0.0287)	-	-	336.827
BIII	18.4311 (0.1482)	0.1069 (0.0010)	1.6511 (0.0015)	1 (-)	-	-	938.113
LL	3.7222 (0.0092)	1 (-)	1.1094 (0.0014)	1 (-)	-	-	4554.549
ELL	1 (-)	3.6292 (0.0269)	0.2635 (0.0023)	1 (-)	-	-	1022.258
BBIII	28.7468 (0.5524)	0.5810 (0.0190)	1.5547 (0.0042)	0.1143 (0.0041)	0.4787 (0.0131)	-	348.247
BW	5.0892 (0.0321)	-	-	0.4410 (0.0037)	3.8626 (0.1545)	2.0235 (0.0166)	368.887

A comparison of the new distribution with three of its sub-models using LR statistics is described in Table 2. These statistics indicate that the new distribution is the most adequate model to these data.

The plots of the fitted densities of the GBIII distribution and its sub-models along with the fitted densities of the BBIII and BW models are displayed in Fig. 2. They reveal that the GBIII distribution provides a better fit to the current data than those of the other models.

The measures of skewness and kurtosis for the GBIII distribution are, respectively, -0.3001116 and 0.05107192.

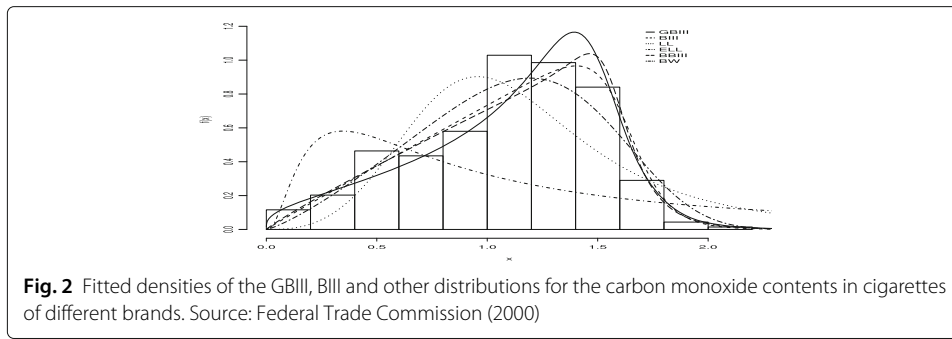
Further, we apply the Cramér–von Mises (W^*) and Anderson–Darling (A^*) goodness-of-fit statistics in order to verify which model yields the best fit to these data. These statistics are described by Chen and Balakrishnan (1995). In general, the smaller the values of these statistics indicates a better fit. The values of W^* and A^* for six fitted models to the current data are listed in Table 3. The figures in this table reveal that the GBIII model provides the best fit among the fitted models.

Some computing issues and a simulation study

As mentioned before, the optimization for estimating the parameters can be performed by minimizing the negative log-likelihood and, for that, we use the `nllminb` function of the R language. Optimization can also be tackled through simulated annealing (Kirkpatrick et al. 1983) using the `optim` function of the R. Reasonable starting values are chosen such that the estimated pdf of a sub-model fits well the histogram of the data. We now discuss some estimation issues related to the GBIII distribution. Mäkeläinen et al. (1981), in their Theorem 2.1, have established conditions for existence and uniqueness of the MLEs. However, proving that the likelihood function satisfies those conditions is a very hard task that could be addressed in a separate paper.

Table 2 LR tests for the cigarettes data

Model	Hypotheses	Statistic LR	p-value
GBIII vs BIII	$H_0 : a = 1$ vs $H_1 : H_0$ is false	10.79832	0.0010159
GBIII vs LL	$H_0 : c = \beta = 1$ vs $H_1 : H_0$ is false	189.9402	0.0000000
GBIII vs ELL	$H_0 : c = \alpha = 1$ vs $H_1 : H_0$ is false	587.5724	0.0000000



We conduct a Monte Carlo simulation study to assess the finite sample behaviour of the MLEs of the GBIII parameters. Random samples from the GBIII model are obtained using the qf given by (8). We consider as the true parameter values the average between two parameter vector estimates obtained for the cigarettes data when two different starting points were chosen. Even though those starting points differ substantially, the estimates do not.

Let $\theta = (\alpha, \beta, s, c) = (17.11192, 0.02785, 1.49817, 2.64952)$ be the *true parameter vector*. For each simulated sample, we estimate the true parameter vector using as starting points the so-called *initial point 1* and *initial point 2*, which correspond to the estimated vectors of θ of the sub-models 1 and 2 given in Table 1. The results are obtained from $k = 500$ replicates for each of the sample sizes $n = 50, 100, 500, 1,000$ and $5,000$. For a specified sample size and $k = 500$ estimated values obtained using one of the two referred starting values, we evaluate the average of those estimated vectors and the square root mean squared errors (SRMSEs). Afterwards, we take the averages of the results obtained using each of the starting points to produce Table 4. As we can observe from the figures in this table, the estimated expected vector does approach the true vector, but the SRMSEs decrease slowly.

In Fig. 3, we compare the GBIII density for the *true vector* and the corresponding estimated expected vectors obtained for the cases when the sample sizes are 50, 100, 500, 1,000 and 5,000, when we use the initial points 1 and 2. Regarding the consistency of the GBIII parameter estimators, Fig. 3 shows that when the sample size increases, the estimated vector tends to the true parameter vector, regardless of the initial values used in the estimation procedure. However, as mentioned before, the convergence seems rather slow.

In order to provide some empirical evidence about the asymptotic distribution of the MLEs of the GBIII density, we simulate 5,000 samples of size 10,000 from the GBIII

Table 3 Goodness-of-fit statistics

Distribution	W^*	A^*
GBIII	0.23988	1.46001
BIII	0.30316	1.83086
LL	2.47901	13.86572
ELL	5.53160	30.68406
BBIII	0.29278	1.75430
BW	0.66758	3.89611

Table 4 Monte Carlo results: means and SRMSEs (in parentheses) of $\hat{\alpha}$, $\hat{\beta}$, \hat{s} and \hat{c}

Parameter	α	β	s	c
True values	17.11192	0.02785	1.49817	2.64952
$n = 50$	20.60594 (10.48504)	0.06942 (0.09636)	1.51430 (0.12745)	2.56537 (1.55388)
$n = 100$	19.09881 (8.89715)	0.04589 (0.05139)	1.50674 (0.09234)	2.61276 (1.04975)
$n = 500$	17.38320 (1.77489)	0.03260 (0.01378)	1.50329 (0.04358)	2.59845 (0.48186)
$n = 1,000$	17.22701 (1.27852)	0.03018 (0.00780)	1.49953 (0.03029)	2.63054 (0.31852)
$n = 5,000$	17.16906 (0.49828)	0.02872 (0.00329)	1.50008 (0.01300)	2.63229 (0.14757)

model evaluated at the *true vector*. Figure 4 displays the histograms of the corresponding 5,000 estimates of the parameters. Each curve represents the normal density with the mean and standard deviation parameters fixed, respectively, at the average and sample standard deviation of the 5,000 parameter estimates. The vertical dotted lines represent the true parameter values. We can note, for each of the parameters, that the normal model provides a good approximation of the distributions of the parameter estimates. In summary, the simulations confirm that the GBIII estimators are consistent and their distributions can be approximated by the normal distribution in accordance with the first-order likelihood theory.

An application of the LGBIII regression model

In this section, we fit the LGBIII regression model defined in Section 3 to a data set obtained from a study carried out at the Department of Entomology of the Luiz de Queiroz School of Agriculture, University of São Paulo, which aims to assess the longevity of the mediterranean fruit fly (*ceratitis capitata*). The need for this fly to seek food just after emerging from the larval stage has permitted the use of toxic baits for its management in Brazilian orchards for at least fifty years. This pest control technique consists of using small portions of food laced with an insecticide, generally an organophosphate, that quickly kills the flies, instead of using an insecticide alone. Recently, there have been reports of the insecticidal effect of extracts of the neem tree leading to proposals to adopt various extracts (aqueous extract of the seeds, methanol extract of the leaves and dichloromethane extract of the branches) to control pests such as the mediterranean

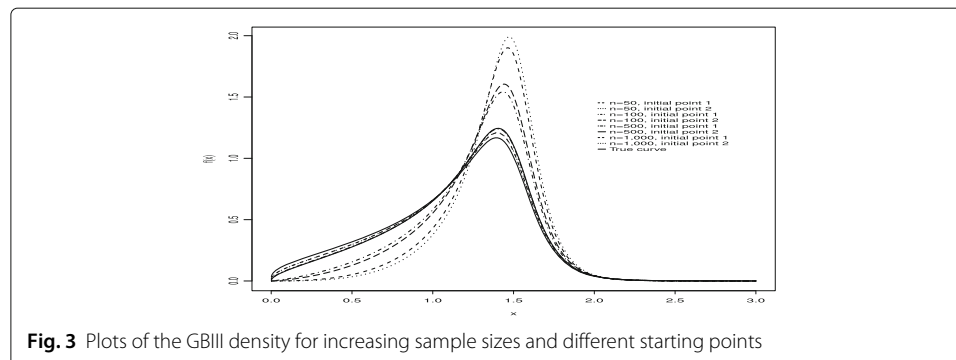
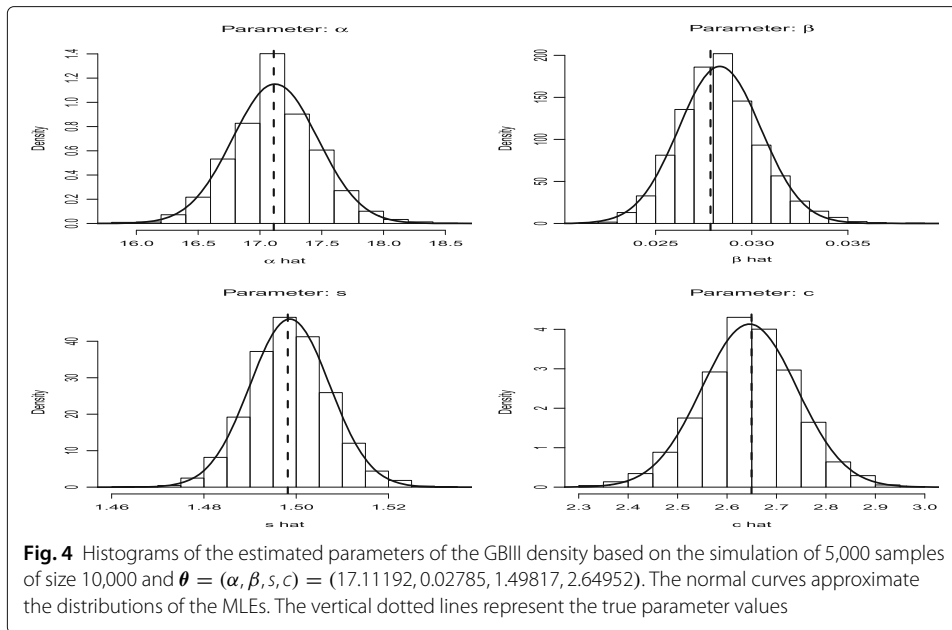


Fig. 3 Plots of the GBIII density for increasing sample sizes and different starting points



fruit fly. The experiment was completely randomized with eleven treatments, consisting of different extracts of the neem tree, at concentrations of 39, 225 and 888 ppm. After preliminary statistical analysis, these eleven treatments were allocated into two groups, namely:

- Group 1: Control 1 (deionized water); Control 2 (acetone - 5%); aqueous extract of seeds (AES) (39 ppm); AES (225 ppm); AES (888 ppm); methanol extract of leaves (MEL) (225 ppm); MEL (888 ppm); and dichloromethane extract of branches (DMB) (39 ppm).
- Group 2: MEL (39 ppm); DMB (225ppm) and DMB (888 ppm).

For more details, see Silva et al. (2013). The response variable in the experiment is the lifetime of the adult flies in days after exposure to the treatments. The experimental period is set at 51 days, so that the numbers of larvae that survived beyond this period are considered as censored data. The total sample size is $n = 72$, because four observations are lost. Therefore, we use the following variables: t_i -lifetime of ceratitis capitata adults in days, δ_i -censoring indicator, v_{i1} -sex of the larvae and v_{i2} -group (0=group 1, 1=group 2). Next, we present results by fitting the regression model

$$y_i = \tau_0 + \tau_1 v_{i1} + \tau_2 v_{i2} + \sigma z_i,$$

Table 5 MLEs of the parameters from the LGBIII and LBIII regression models fitted to the entomology data, the corresponding SEs (given in parentheses), p -values in [.] and the AIC, CAIC and BIC statistics

Model	c	β	σ	τ_0	τ_1	τ_2	AIC	CAIC	BIC
LGBIII	0.1722 (0.0161)	1.7304 (0.2708)	0.1685 (0.0244)	3.4935 (0.0624)	-0.0567 (0.0534)	-0.2830 (0.0574)	273.7	274.2	292.6
				[< 0.001]	[0.2890]	[< 0.001]			
LBIII	1	0.4016 (0.0807)	0.2199 (0.0286)	3.4048 (0.0835)	0.0402 (0.0843)	-0.3410 (0.0888)	338.8	339.2	354.6
	-			[< 0.001]	[0.6339]	[0.0002]			

Table 6 LR statistic w for the entomology data

Model	Hypotheses	w	p -value
LGBIII vs LBIII	$H_0 : c = 1$ vs $H_1 : H_0$ is false	67.1	< 0.00001

where the random variable z_i has the LGBIII density (16) (with $\mu = 0$ and $\sigma = 1$) for $i = 1, \dots, 72$. The MLEs of the model parameters are evaluated using the NLMixed procedure in SAS. Iterative maximization of the log-likelihood function (19) starts with initial values for β and σ taken from the fit of the LBIII regression model with $c = 1$, i.e. $\beta = 0.4, \sigma = 0.2, \tau_0 = 3.4, \tau_1 = 0.04$ and $\tau_2 = -0.3$. Table 5 lists the MLEs of the model parameters. The values of the AIC, CAIC and BIC statistics are smaller for the LGBIII regression model when compared to those values for the LBIII regression model.

A comparison of the new distribution with its LBIII sub-model using the LR statistic w is presented in Table 6. The value of w indicates that the LGBIII distribution provides a better fit to these data than the null model.

We note from the fitted LGBIII regression model that x_1 is not significant at 5%. So, the final model is given by

$$y_i = \tau_0 + \tau_2 v_{i2} + \sigma z_i.$$

The MLEs of the parameters in the final model are listed in Table 7.

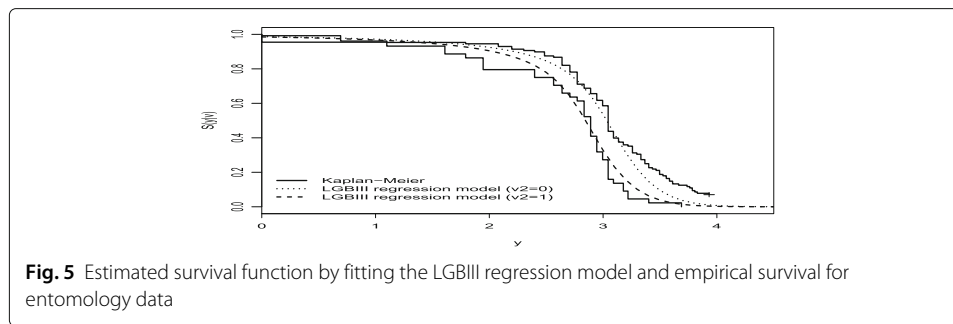
Note again that the values of the AIC, CAIC and BIC statistics are smaller for the LGBIII regression model when compared to those values for the LBIII regression model. In order to assess if the model is appropriate, the plots comparing the empirical survival function and the estimated survival function by the LGBIII regression model, see Eq. (20), are displayed in Fig. 5. Based on these plots, it is evident that the LGBIII model provides a superior fit.

Conclusions

Providing a new class of distributions is always precious for statisticians. There has been an increased interest in developing generalized classes of distributions by adding a single shape parameter to a baseline distribution. There is no doubt that some of these classes have attracted several applied researchers. Following this idea, Zografos and Balakrishnan (2009) introduced a gamma-generated family of distributions by adding an extra positive shape parameter to a baseline model. In this paper, we study some mathematical properties of the new four-parameter *gamma Burr III* distribution based on the gamma-generated family. We prove empirically that the proposed distribution can provide a better fit than important generated models such as the beta Burr III (Gomes et al. 2013) and

Table 7 MLEs of the parameters from the LGBIII and LBIII regression models fitted to the entomology data, considering only the significant variables, the corresponding SEs (given in parentheses), p -values in [.] and the AIC, CAIC and BIC statistics

Model	c	β	σ	τ_0	τ_2	AIC	CAIC	BIC
LGBIII	6.3762 (1.6458)	0.005080 (0.0062)	0.1157 (0.0136) [< 0.001]	2.9722 (0.0843) [0.0056]	- 0.1864 (0.0664)	256.3	256.6	272.0
LBIII	1 -	0.4084 (0.0791)	0.2218 (0.0279)	3.4173 (0.0791) [< 0.001]	- 0.3402 (0.0886) [0.0002]	337.1	337.3	349.6



beta Weibull (Lee et al. 2007) distributions. Finally, we propose a new log-gamma Burr III regression model and illustrate its importance by means of one application to a real data set.

Authors' contributions

GMC proposed the gamma Burr III model, wrote some parts of Sections 1 to 2, and also drafted the manuscript. AEG wrote some parts of Section 2. CQd-S wrote some parts of Section 2, subsections 4.1 and 4.2 and prepared Figs. 1, 2, 3 and 4. EMMO proposed the log-gamma Burr III model described in Section 3, performed the application in subsection 4.3, and also prepared Fig. 5. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Departamento de Estatística, Universidade Federal de Pernambuco, 52171-900 Recife, PE, Brazil. ²Departamento de Estatística, Universidade de Brasília, 70910-900 Brasília, DF, Brazil. ³Departamento de Ciências Exatas, ESALQ-Universidade de São Paulo, Piracicaba, SP, Brazil.

Received: 18 November 2016 Accepted: 13 October 2017

Published online: 01 November 2017

References

- Alexander, C, Cordeiro, GM, Ortega, EMM, Sarabia, JM: Generalized beta-generated distributions. *Computational Statistics and Data Analysis*. **56**, 1880–1897 (2012)
- Alzaatreh, A, Lee, C, Famoye, F: A new method for generating families of continuous distributions. *Metron*. **71**, 63–79 (2013)
- Chen, G, Balakrishnan, N: A general purpose approximate goodness-of-fit test. *J. Qual. Technol.* **27**, 154–161 (1995)
- Cordeiro, GM, de Castro, M: A new family of generalized distributions. *J. Stat. Comput. Simul.* **81**, 883–898 (2011)
- Cordeiro, GM, Ortega, EMM, Popovic, BV, Pescim, RR: The Lomax generator of distributions: Properties, minification process and regression model. *Appl. Math. Comput.* **247**, 465–486 (2014)
- Cordeiro, GM, Ortega, EMM, Ramires, TG: A new generalized Weibull family of distributions: mathematical properties and applications. *J. Stat. Distrib. Appl.* **2**, 131–145 (2015)
- Cordeiro, GM, Alizadeh, M, Ozel, G, Hossein, B, Ortega, EMM, Altun, E: The generalized odd log-logistic family of distributions: properties, regression models and applications. *J. Stat. Comput. Simul.* **87**, 908–932 (2017)
- Dagum, C: A new model of personal income distribution: specification and estimation. *Econ. Appl.* **30**, 413–437 (1977)
- Eugene, N, Lee, C, Famoye, F: Beta-Normal distribution and its applications. *Commun. Stat. Theory Methods*. **31**, 497–512 (2002)
- Gomes, AE, da-Silva, CQ, Cordeiro, GM, Ortega, EMM: The beta Burr III model for lifetime data. *Braz. J. Probab. Stat.* **27**, 502–543 (2013)
- Gove, JH, Ducey, MJ, Leak, WB, Zhang, L: Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr type III distribution. *Forestry* (2008). doi:10.1093/forestry/cpm025
- Gradshteyn, IS, Ryzhik, IM: *Table of Integrals, Series, and Products*, seventh edition. Academic Press, San Diego (2000)
- Kirkpatrick, S, Gelatt Jr, CD, Vecchi, MP: Optimization by Simulated Annealing. *Science*. **220**, 671–680 (1983)
- Kleiber, C, Kotz, S: *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley, New York (2003)
- Kleiber, C: *A Guide to the Dagum Distributions*. Springer, New York (2008)
- Klugman, SA, Panjer, HH, Willmot, GE: *Loss Models*. Wiley, New York (1998)
- Lawless, JF: *Statistical Models and Methods for Lifetime Data*. John Wiley, New York (2003)
- Lee, C, Famoye, F, Olumolade, O: Beta-Weibull Distribution: Some Properties and Applications to Censored Data. *J. Mod. Appl. Stat. Methods*. **6**, 173–186 (2007)
- Lindsay, SR, Wood, GR, Woollons, RC: Modelling the diameter distribution of forest stands using the Burr distribution. *J. Appl. Stat.* **23**, 609–619 (1996)

- Mäkeläinen, T, Schmidt, K, Styan, GPH: On the Existence and Uniqueness of the Maximum Likelihood Estimate of a Vector-Valued Parameter in Fixed-Size. *Ann. Stat.* **9**, 758–767 (1981)
- Mielke, PW: Another family of distributions for describing and analyzing precipitation data. *J. Appl. Meteorology.* **12**, 275–280 (1973)
- Nofal, ZM, Afify, AZ, Yousof, HM, Cordeiro, M: The generalized transmuted-G family of distributions. *Commun. Stat. Theory Methods.* **46**, 4119–4136 (2016)
- Silva, MA, Bezerra-Silva, GCD, Vendramim, JD, Mastrangelo, T: Sublethaleffect of neem extract on Mediterranean fruit fly adults. *Rev. Bras. Frutic.* **35**, 93–101 (2013)
- Tahir, MH, Cordeiro, GM, Alizadeh, M, Mansoor, M, Zubair, M, Hamedani, GG: The odd generalized exponential family of distributions with applications. *J. Stat. Distrib. Appl.* **2**, 1–28 (2015)
- Zografos, K, Balakrishnan, N: On families of beta- and generalized gamma-generated distributions and associated inference. *Stat. Methodol.* **6**, 344–362 (2009)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
