Statistics

A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: https://www.tandfonline.com/loi/gsta20

A bimodal gamma distribution: properties, regression model and applications

Roberto Vila, Letícia Ferreira, Helton Saulo, Fábio Prataviera & Edwin Ortega

To cite this article: Roberto Vila, Letícia Ferreira, Helton Saulo, Fábio Prataviera & Edwin Ortega (2020) A bimodal gamma distribution: properties, regression model and applications, Statistics, 54:3, 469-493, DOI: 10.1080/02331888.2020.1764560

To link to this article: https://doi.org/10.1080/02331888.2020.1764560

Published online: 13 May 2020.

STATISTICS
2020, VOL. 54, NO. 3, 469–493
https://doi.org/10.1080/02331888.2020.1764560

Taylor & Francis
Taylor & Francis Group

A bimodal gamma distribution: properties, regression model and applications

Roberto Vila[a], Letícia Ferreira[a], Helton Saulo [a], Fábio Prataviera[b] and Edwin Ortega [b]

[a]Departamento de Estatística, Universidade de Brasília, Brasília, Brazil; [b]Departamento de Ciências Exatas, Universidade de São Paulo, São Paulo, Brazil

ABSTRACT
In this paper, we propose a bimodal gamma distribution using a quadratic transformation based on the alpha-skew-normal model. We discuss several properties of this distribution such as mean, variance, moments, hazard rate and entropy measures. Further, we propose a new regression model with censored data based on the bimodal gamma distribution. This regression model can be very useful to the analysis of real data and could give more realistic fits than other special regression models. Monte Carlo simulations were performed to check the bias in the maximum likelihood estimation. The proposed models are applied to two real data sets found in the literature.

ARTICLE HISTORY
Received 14 January 2020
Accepted 26 April 2020

KEYWORDS
Bimodal distribution; gamma distribution; Monte Carlo simulation; regression model

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
MSC 62E10; MSC 62F10; MSC 62E15

1. Introduction

The unimodal gamma distribution is well known due to its flexibility and good properties [1]. This model has been widely applied in several areas, such as physics [2], medicine [3,4], survival analysis [5], quality control [6,7], inventory [8,9], among others.

A general and effective way to introduce bimodality into a unimodal distribution is through a quadratic transformation, as it demands less computational effort in parameter estimation when compared to mixture-based bimodal models. In this sense, Elal-Olivero [10] introduced a prominent quadratic transformation in the normal distribution that produces asymmetry and bimodality. This transformation gave rise to the alpha-skew-normal (ASN) family of distributions. A random variable Z has an ASN distribution with parameter δ, if its probability density function (PDF) and cumulative distribution function (CDF) are given, respectively, by

$$g(z) = \frac{(1-\delta z)^2 + 1}{2+\delta^2}\phi(z) \quad \text{and} \quad G(z) = \Phi(z) + \delta\left(\frac{2-\delta z}{2+\delta^2}\right)\phi(z), \quad z, \delta \in \mathbb{R}, \quad (1)$$

where δ is an asymmetric parameter that controls the uni-bimodality effect and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal PDF and CDF, respectively; see [10]. We denote $Z \sim \text{ASN}(\delta)$.

CONTACT Edwin Ortega edwin@usp.br Departamento de Ciências Exatas, Universidade de São Paulo, São Paulo 05508-220, SP, Brazil

In this context, we introduce a bimodal gamma (BGamma) distribution through the multiplication of the gamma density by a quadratic function proposed and analysed by Elal-Olivero [10]. We present a statistical methodology based on the proposed BGamma model, including model formulation, mathematical properties and estimation based on the maximum likelihood (ML) method. Numerical evaluation is carried out by both Monte Carlo simulation and application to real data. In special, the proposed BGamma model is an alternative to the mixture generalized gamma model propose by Çankaya et al. [11].

Survival analysis is one of the areas of statistics that has grown steadily in recent decades. It is common for the response variable (time until the occurrence of the event of interest) to be related to the explanatory variables that explain its variability. We study the effects of these explanatory variables on the response variable using a regression model that is appropriate for censored data. In this paper, we also introduce a regression model using the BGamma distribution, for survival times analysis as a feasible alternative to the gamma regression model. The inferential part was carried out using the asymptotic distribution of the ML estimators.

The rest of the paper proceeds as follows. In Section 2, we introduce the bimodal gamma distribution. In Section 3, we discuss several mathematical properties of the proposed model. In Section 4, we consider likelihood-based methods to estimate the model parameters and we carry out a Monte Carlo simulation study to evaluate the performance of the ML estimators. In Section 5, we derive a regression model based on the proposed distribution. In Section 6, we illustrate the proposed methodologies with two real data sets. Finally, in Section 7, we make some concluding remarks.

2. The bimodal gamma distribution

We say that a random variable X has a BGamma distribution with parameter vector $\boldsymbol{\theta}_\delta := (\alpha, \beta, \delta)$, $\alpha > 0, \beta > 0$ and $\delta \in \mathbb{R}$, denoted by $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$, if its PDF is given by

$$f(x; \boldsymbol{\theta}_\delta) = \begin{cases} \dfrac{1 + (1 - \delta x)^2}{Z(\boldsymbol{\theta}_\delta)} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}, & x > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $Z(\boldsymbol{\theta}_\delta) := 2 + \frac{\alpha\delta}{\beta}[(1+\alpha)\frac{\delta}{\beta} - 2)]$ is the normalization constant, and $\Gamma(\alpha)$ is the gamma function. When $\delta = 0$, we obtain the classic gamma distribution with parameter vector $\boldsymbol{\theta}_0 = (\alpha, \beta, 0) := (\alpha, \beta)$. Figure 1 shows some different shapes of the BGamma PDF for different combinations of parameters. This figure reveals clearly the bimodality effect caused by the parameter δ.

If Y is a non-negative random variable following a gamma distribution with parameter vector $\boldsymbol{\theta}_0$, denoted by $Y \sim \text{BGamma}(\boldsymbol{\theta}_0)$, note that in fact the non-negative function $f(\cdot; \boldsymbol{\theta}_\delta)$ is a PDF since

$$\int_0^\infty f(x; \boldsymbol{\theta}_\delta)\,\mathrm{d}x = \frac{1 + \mathbb{E}(1 - \delta Y)^2}{Z(\boldsymbol{\theta}_\delta)} = \frac{1 + \frac{\alpha\delta^2}{\beta^2} + (1 - \frac{\alpha\delta}{\beta})^2}{Z(\boldsymbol{\theta}_\delta)} = 1.$$

Proposition 2.1 (Monotonicity of the PDF): *The PDF of the* BGamma *distribution* (2) *is decreasing as* $\alpha \leqslant 1$, $\delta > 0$ *and* $x < 1/\delta$.

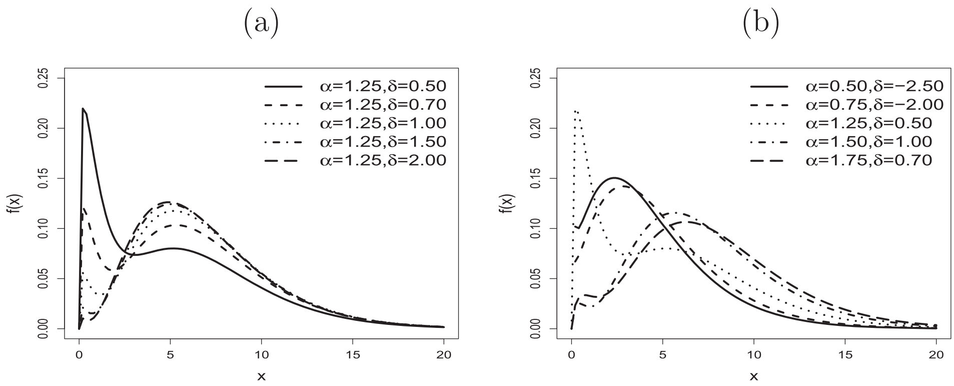

Figure 1. Bimodal gamma PDFs for some parameter values ($\beta = 0.50$).

Proof: Note that the function $g(x) := 1 + (1 - \delta x)^2$ is decreasing as $\alpha \leqslant 1$, $\delta > 0$ and $x < 1/\delta$. Furthermore, when $\alpha \leqslant 1$, the density $f(x; \boldsymbol{\theta}_\delta)$ is the product of the function $g(x)$ and a decreasing and non-negative function. Thus, the proof is complete. ■

3. Mathematical properties

3.1. Characterization of unimodality and bimodality

Proposition 3.1 (Modes): *The point x is a mode of the* BGamma *density* (2), *if and only if it is the solution of the following cubic polynomial equation:*

$$\left[\beta\delta^2 x^2 - 2\delta(\delta + \beta)x + 2(\delta + \beta)\right] x - \left[1 + (1 - \delta x)^2\right](\alpha - 1) = 0,$$

or equivalently

$$\beta\delta^2 x^3 - \delta\left[2(\delta + \beta) + \delta(\alpha - 1)\right] x^2 + 2\left[\delta + \beta + 2(\alpha - 1)\right] x - 2(\alpha - 1) = 0.$$

Proof: The proof is trivial and omitted. ■

Theorem 3.2 (Unimodality): *The PDF of the* BGamma *distribution* (2) *is unimodal in the following cases*:

(1) *for $\delta = 0$ and $\alpha > 1$;*
(2) *for $\delta \geqslant \beta$ and $\alpha = 1$.*

Proof: (1) When $\delta = 0$ and $\alpha > 1$ it is well-known that the density (2) increases and then decreases, with mode at $(\alpha - 1)/\beta$.

To prove Item (2), we suppose that x is a mode of the BGamma density and that $\alpha = 1$. In this case, the point x must be the solution of the quadratic polynomial equation $p_2(x) := \beta\delta^2 x^2 - 2\delta(\delta + \beta)x + 2(\delta + \beta) = 0$ (see Proposition 3.1). The discriminant of p_2 is given by $\Delta_2 = 4\delta^2(\delta + \beta)(\delta - \beta)$.

If $\delta = \beta$, $\alpha = 1$, $\Delta_2 = 0$ then, there is one real zero of multiplicity two for $p_2(x) = 0$, denoted by x_0. Note that $x_0 = 2/\beta$. Since $f(x; \boldsymbol{\theta}_\delta) \to 1$ as $x \to 0^+$ and $f(x; \boldsymbol{\theta}_\delta) \to 0$ as

$x \to \infty$, it follows that the density (2) increases on the interval $(0, x_0)$ and then decreases on (x_0, ∞). Then x_0 is the unique global maximum point.

On the other hand, if $\delta > \beta$ and $\alpha = 1$ note that $\Delta_2 > 0$ tn, the equation $p_2(x) = 0$ has two distinct rational zeros, denoted by x_1, x_2. Note that $x_1 = (\delta + \beta - \sqrt{\delta^2 - \beta^2})/(\beta\delta) > 0$ and $x_2 = (\delta + \beta + \sqrt{\delta^2 - \beta^2})/(\beta\delta) > 0$, and $x_1 < x_2$. Since $f(x; \boldsymbol{\theta}_\delta) \to \beta^3/[\beta^2 + \delta(\delta - \beta)]$ as $x \to 0^+$ and $f(x; \boldsymbol{\theta}_\delta) \to 0$ as $x \to \infty$, it follows that the BGamma density (2) decreases on the interval $(0, x_1)$, increases on (x_1, x_2) and then decreases on (x_2, ∞). That is, x_1 and x_2 are minimum and maximum points, respectively. ■

To state the following result, we define

$$a_{\delta,\beta} \coloneqq \delta(4 + \delta)(\delta + \beta) + \beta(3\delta - 4)(3\delta + 4); \tag{3}$$

$$b_{\delta,\beta} \coloneqq 16(1 + \delta)(\delta + \beta)^2 + \left[\delta^2 + 18\beta\delta(4 + \delta) - 96\beta\right](\delta + \beta) - 2\delta^2; \tag{4}$$

$$c_{\delta,\beta} \coloneqq 4(4 + \delta)(\delta + \beta)^3 + 12\beta(3\delta - 4)(\delta + \beta)^2 - 4\delta(\delta + \beta) - 27\beta. \tag{5}$$

Theorem 3.3 (Bimodality and unimodality): *The PDF of the* BGamma *distribution* (2), *as* $\alpha > 1$, *has the following shapes.*

(1) *It is bimodal as* $\delta > \beta$, $a_{\delta,\beta} > 0$, $b_{\delta,\beta} > 0$ *and* $c_{\delta,\beta} > 0$. *Just take, for example,* $\beta = 2$ *and* $\delta = 3$;
(2) *it is unimodal as* $0 < \delta < \beta$, $a_{\delta,\beta} < 0$, $b_{\delta,\beta} < 0$ *and* $c_{\delta,\beta} < 0$;
(3) *it is bimodal as* $\delta = \beta > \frac{8\sqrt{3}}{11} - \frac{4}{11}$;
(4) *it is unimodal as* $0 < \delta = \beta < \frac{\sqrt{1745}}{12} - \frac{35}{12}$;

where $a_{\delta,\beta}, b_{\delta,\beta}$ *and* $c_{\delta,\beta}$ *are as in* (3), (4) *and* (5), *respectively.*

Proof: If x is a mode of the BGamma density, by Proposition 3.1 the point x must be the solution of the cubic polynomial equation $p_3(x) \coloneqq \beta\delta^2 x^3 - \delta[2(\delta + \beta) + \delta(\alpha - 1)]x^2 + 2[\delta + \beta + 2(\alpha - 1)]x - 2(\alpha - 1) = 0$. By Descartes' rule of signs (see, e.g.,[12,13]), $p_3(x)$ has three or one positive roots. It is well-known that the discriminant of a cubic polynomial $ax^3 + bx^2 + cx + d$ is given by $\Delta_3 = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$. In our case, we have

$$\begin{aligned}\Delta_3 = \Delta_3(\alpha) &= 16\delta^4(\alpha - 1)^4 + 16\delta^2 a_{\beta,\delta}(\alpha - 1)^3 \\ &\quad + 4\delta^2 b_{\beta,\delta}(\alpha - 1)^2 + 4\delta^2 c_{\beta,\delta}(\alpha - 1) + 16\delta^2(\delta - \beta)(\delta + \beta)^3.\end{aligned}$$

(1) Since $\delta > \beta$, $a_{\delta,\beta}$, $b_{\delta,\beta}$ and $c_{\delta,\beta}$ are positive, we have $\Delta_3(\alpha) > 0$ for each $\alpha > 1$. Then the equation $p_3(x) = 0$ has three distinct positive roots, denoted by x_1, x_2, x_3. Let us assume that $x_1 < x_2 < x_3$. Since $f(x; \boldsymbol{\theta}_\delta) \to 0$ as $x \to 0^+$ and $f(x; \boldsymbol{\theta}_\delta) \to 0$ as $x \to \infty$, it follows that the BGamma density (2) increases on the intervals $(0, x_1)$ and (x_2, x_3), and decreases on (x_1, x_2) and (x_3, ∞). That is, x_1 and x_3 are two maximum points and x_1 is the unique minimum point.

(2) Since $0 < \delta < \beta$, $a_{\delta,\beta}$, $b_{\delta,\beta}$ and $c_{\delta,\beta}$ are negative, it follows that $\Delta_3(\alpha) < 0$ for each $\alpha > 1$. Hence, the polynomial equation $p_3(x) = 0$ has one positive root, denoted by x_0, and

two non-real complex conjugate roots. Since $f(x;\boldsymbol{\theta}_\delta) \to 0$ as $x \to 0^+$ and $f(x;\boldsymbol{\theta}_\delta) \to 0$ as $x \to \infty$, note that x_0 has to be a maximum point.

To prove Items (3) and (4), note that if $\delta = \beta$, then

$$
\begin{aligned}
a_{\delta,\beta} &= \beta(11\beta^2 + 8\beta - 16);\\
b_{\delta,\beta} &= \beta^2(18\beta^2 + 105\beta - 65);\\
c_{\delta,\beta} &= \beta(32\beta^3 + 272\beta^2 - 200\beta - 27).
\end{aligned}
$$

For each $\delta = \beta > \frac{8\sqrt{3}}{11} - \frac{4}{11}$, we obtain that $a_{\delta,\beta}$, $b_{\delta,\beta}$ and $c_{\delta,\beta}$ are positive quantities, then $\Delta_3(\alpha) > 0$ for each $\alpha > 1$, and the proof of Item (3) follows analogously to Item (1). On the other hand, for $0 < \delta = \beta < \frac{\sqrt{1745}}{12} - \frac{35}{12}$ note that $a_{\delta,\beta}$, $b_{\delta,\beta}$ and $c_{\delta,\beta}$ are negative. Hence, $\Delta_3(\alpha) < 0$ for each $\alpha > 1$, and the proof of Item (4) follows analogously to Item (2). ■

Remark 3.4: In the proof of Theorem 3.3, Item (1), another way to verify that the polynomial equation $p_3(x) = \beta\delta^2x^3 - \delta[2(\delta+\beta) + \delta(\alpha-1)]x^2 + 2[\delta+\beta+2(\alpha-1)]x - 2(\alpha-1) = 0$ has exactly three positive roots is to use Vieta's formula (see, e.g., [14]). Indeed, in our case Vieta's formula is expressed as

$$
\begin{aligned}
x_1 + x_2 + x_3 &= \frac{2(\delta+\beta) + \delta(\alpha-1)}{\beta\delta},\\
x_1x_2 + x_1x_3 + x_2x_3 &= \frac{2[\delta+\beta+2(\alpha-1)]}{\beta\delta^2},\\
x_1x_2x_3 &= \frac{2(\alpha-1)}{\beta\delta^2}.
\end{aligned}
$$

From the above equations, the claim follows.

3.2. Real moments, variance and moment generating function

The following result shows that the existence of the classic gamma moments is inherited for the BGamma distribution.

Proposition 3.5 (Moments): *If $X \sim \mathrm{BGamma}(\boldsymbol{\theta}_\delta)$, for each fixed real number ν such that $\nu > -\alpha$, we have*

$$
\mathbb{E}X^\nu = \frac{\left[2 - 2\frac{\delta}{\beta}(\nu+\alpha) + \frac{\delta^2}{\beta^2}(\nu+\alpha+1)\right]\Gamma(\nu+\alpha)}{Z(\boldsymbol{\theta}_\delta)\beta^\nu\Gamma(\alpha)}.
$$

Proof: A straightforward computation shows that

$$
\mathbb{E}X^\nu = \frac{2\mathbb{E}Y^\nu - 2\delta\mathbb{E}Y^{\nu+1} + \delta^2\mathbb{E}Y^{\nu+2}}{Z(\boldsymbol{\theta}_\delta)}, \qquad Y \sim \mathrm{BGamma}(\boldsymbol{\theta}_0).
$$

Since $\mathbb{E}Y^\nu = \frac{\Gamma(\nu+\alpha)}{\beta^\nu\Gamma(\alpha)}$, $\nu > -\alpha$, and $\Gamma(x+1) = x\Gamma(x)$, the proof follows. ■

Corollary 3.6 (Mean and variance): *Set $\mu := \mathbb{E}X$ and $\sigma := \sqrt{\mathrm{Var}(X)}$. By Proposition 3.5, it immediately follows that*

$$\mu = \frac{\alpha}{\beta Z(\boldsymbol{\theta}_\delta)}\kappa(\boldsymbol{\theta}_\delta),$$
$$\sigma^2 = \frac{\alpha}{\beta^2 Z(\boldsymbol{\theta}_\delta)}\left[\alpha\kappa^2(\boldsymbol{\theta}_\delta) + (1+\alpha)Z(\boldsymbol{\theta}_\delta)\kappa(\boldsymbol{\theta}_\delta) + \frac{\delta}{\beta}\Big(\frac{\delta}{\beta} - 2\Big)(1+\alpha)Z(\boldsymbol{\theta}_\delta)\right],$$

where $\kappa(\boldsymbol{\theta}_\delta) := 2 - 2\frac{\delta}{\beta}(1+\alpha) + \frac{\delta^2}{\beta^2}(2+\alpha)$.

Proposition 3.7 (Standardized moments): *If $X \sim \mathrm{BGamma}(\boldsymbol{\theta}_\delta)$, for each fixed natural number n, we have*

$$\mathbb{E}\Big(\frac{X-\mu}{\sigma}\Big)^n = \frac{1}{\sigma^n}\sum_{k=0}^{n}\binom{n}{k}(-\mu)^{n-k}\frac{2 - 2\frac{\delta}{\beta}(k+\alpha) + \frac{\delta^2}{\beta^2}(k+\alpha+1)}{Z(\boldsymbol{\theta}_\delta)\beta^k}\prod_{i=0}^{k-1}(\alpha+i),$$

where μ and σ is as in Corollary 3.6, and $\prod_{i=0}^{-1}(\alpha+i) := 1$. In particular, by taking $n = 3$ and $n = 4$ we have closed expressions for the skewness and kurtosis of X, respectively.

Proof: The proof of this proposition follows immediately by combining the Binomial expansion with the Proposition 3.5 and with the identity $\Gamma(x+1) = x\Gamma(x)$. ■

Proposition 3.8: *If $X \sim \mathrm{BGamma}(\boldsymbol{\theta}_\delta)$, for each fixed natural number n, we have*

(1)

$$\mathbb{E}\log X^n = \frac{\frac{n\delta}{\beta}[\frac{(2\alpha+1)\delta}{\beta} - 2] + n[2 - \frac{2n\alpha\delta}{\beta} + \frac{\alpha(\alpha+1)\delta^2}{\beta^2}][\Psi^{(0)}(\alpha) - \log\beta]}{Z(\boldsymbol{\theta}_\delta)};$$

(2)

$$\mathbb{E}(\log X)^n = \frac{\frac{1}{\Gamma(\alpha)}\{\frac{n(n-1)\delta^2}{\beta^2} - \frac{n\delta}{\beta}2 + [\frac{\alpha(\alpha+1)\delta^2}{\beta}]\} \sum_{k=0}^{n-2}\binom{n-2}{k}(-1)^{n-2-k}(\log\beta)^{n-2-k}\Psi^{(k)}(\alpha)}{Z(\boldsymbol{\theta}_\delta)}$$
$$+ \frac{\frac{1}{\Gamma(\alpha)}\{2 + [\frac{\alpha(\alpha+1)\delta^2}{\beta}]\}[\Psi^{(n)}(\alpha) - (n-1)\log\beta\,\Psi^{(n-1)}(\alpha)] - \frac{n\delta}{\beta\Gamma(\alpha)}[2 + \frac{(2\alpha+1)\delta}{\beta}]\Psi^{(n-1)}(\alpha)}{Z(\boldsymbol{\theta}_\delta)},$$

where $\Psi^{(m)}(z)$ is the polygamma function of order m defined by $\frac{\mathrm{d}^{m+1}}{\mathrm{d}z^{m+1}}\log\Gamma(z)$.

Proof: Let $Y \sim \text{BGamma}(\boldsymbol{\theta}_0)$. Integration by parts gives

$$\mathbb{E}Y \log Y^n = \frac{n}{\beta} + \frac{\alpha}{\beta}\mathbb{E}\log Y^n,$$
$$\mathbb{E}Y^2 \log Y^n = \frac{n(\alpha+1)}{\beta^2} + \frac{n}{\beta}\mathbb{E}Y + \frac{\alpha(\alpha+1)}{\beta^2}\mathbb{E}\log Y^n.$$

Since

$$\mathbb{E}\log X^n = \frac{2\mathbb{E}\log Y^n - 2\delta\mathbb{E}Y\log Y^n + \delta^2\mathbb{E}Y^2\log Y^n}{Z(\boldsymbol{\theta}_\delta)}$$

and $\mathbb{E}\log Y^n = n(\Psi^{(0)}(\alpha) - \log\beta)$, by combining the above identities with Proposition 3.5, the proof of first item follows.

On the other hand, to prove Item (2), note that integration by parts gives

$$\mathbb{E}Y(\log Y)^n = \frac{n}{\beta}\mathbb{E}(\log Y)^{n-1} + \frac{\alpha}{\beta}\mathbb{E}(\log Y)^n,$$
$$\mathbb{E}Y^2(\log Y)^n = \frac{n(n-1)}{\beta^2}\mathbb{E}(\log Y)^{n-2} + \frac{n(2\alpha+1)}{\beta^2}\mathbb{E}(\log Y)^{n-1} + \frac{\alpha(\alpha+1)}{\beta^2}\mathbb{E}(\log Y)^n.$$

Since

$$\mathbb{E}(\log X)^n = \frac{2\mathbb{E}(\log Y)^n - 2\delta\mathbb{E}Y(\log Y)^n + \delta^2\mathbb{E}Y^2(\log Y)^n}{Z(\boldsymbol{\theta}_\delta)},$$
$$\mathbb{E}(\log Y)^n = \frac{1}{\Gamma(\alpha)}\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}(\log\beta)^{n-k}\Psi^{(k)}(\alpha),$$

and

$$\mathbb{E}(\log Y)^{n-1} = \frac{\Psi^{(n-1)}(\alpha)}{\Gamma(\alpha)} + \mathbb{E}(\log Y)^{n-2},$$
$$\mathbb{E}(\log Y)^n = \frac{\Psi^{(n)}(\alpha)}{\Gamma(\alpha)} - (n-1)\frac{\log\beta\,\Psi^{(n-1)}(\alpha)}{\Gamma(\alpha)} + \mathbb{E}(\log Y)^{n-2}$$

by combining the above identities the proof follows. ■

Let $M_X(t) := \mathbb{E}e^{tX}$ be the moment generating function of X (if it exists). The known identity (see [15]) $M_X(t) = \sum_{r=0}^{\infty}\frac{t^r\mathbb{E}X^r}{r!}$, whenever it exists, simply provides an expression for the moment generating function of X since the moments of X exist (see Proposition 3.5). The following result gives us a closed expression for this function.

Proposition 3.9: *If* $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$, *then*

$$M_X(t) = \frac{2 + \delta^2\alpha(\alpha+1) - 2\delta\alpha(\beta - t)}{Z(\boldsymbol{\theta}_\delta)\beta^{-\alpha}}(\beta - t)^{-(\alpha+2)}, \quad \text{for } t < \beta.$$

Proof: Let $Y \sim \text{BGamma}(\boldsymbol{\theta}_0)$. For $t < \beta$, integration by parts gives

$$\mathbb{E}Y\mathrm{e}^{tY} = \frac{\alpha}{\beta - t} M_Y(t), \quad \mathbb{E}Y^2\mathrm{e}^{tY} = \frac{\alpha(\alpha+1)}{(\beta-t)^2} M_Y(t).$$

Since

$$M_X(t) = 2M_Y(t) - 2\delta\mathbb{E}Y\mathrm{e}^{tY} + \delta^2\mathbb{E}Y^2\mathrm{e}^{tY},$$

combining the above identities, we obtain

$$M_X(t) = \frac{2 + \delta^2\alpha(\alpha+1) - 2\delta\alpha(\beta - t)}{Z(\boldsymbol{\theta}_\delta)(\beta-t)^2} M_Y(t), \quad \text{for } t < \beta.$$

Since $M_Y(t) = (1 - \frac{t}{\beta})^{-\alpha}$ for $t < \beta$, the proof follows. ∎

Remark 3.10: The characteristic function of $X \sim \text{BGamma}(\boldsymbol{\theta})$, denoted by $\phi_X(t)$, can be obtained from the moment generating function by the relation $M_X(t) = \phi_X(-it)$.

The next result shows that the tail of the BGamma distribution (2) function decays to zero exponentially or faster.

Corollary 3.11 (Light-tailed distribution): *If* $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$, *then there exists* $t > 0$ *such that* $\mathbb{P}(X > x) \leqslant \mathrm{e}^{-tx}$ *for x large enough.*

Proof: Since, by Proposition 3.9, there exists $t < \beta$ such that $M_X(t) < \infty$, $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$, the proof follows. ∎

Remark 3.12: Let X an absolutely continuous random variable with density function $f_X(\cdot)$. Following [16], the rate of a random variable is

$$\tau_X := -\lim_{x\to\infty} \frac{\mathrm{d}\ln f_X(x)}{\mathrm{d}x}.$$

Note that

$$\begin{aligned}\tau_{\text{BGamma}(\boldsymbol{\theta}_\delta)} &= \lim_{x\to\infty}\left[\frac{2\delta(1-\delta x)}{1+(1-\delta x)^2} - (\alpha - 1)\frac{1}{x} + \beta\right] = \beta \\ &= \tau_{\text{BGamma}(\boldsymbol{\theta}_0)} = \tau_{\text{BGamma}(\alpha=1,\beta,\delta=0)} = \tau_{\exp(\beta)}.\end{aligned}$$

That is, the rate of a BGamma-distributed random variable depends only on its scale β. In other words, far enough out in the tail, every BGamma distribution looks like an exponential distribution. On the other hand, it is simple to verify that

$$\tau_{\text{InvGamma}(\boldsymbol{\theta}_0)} = \tau_{\text{LogNorm}(\mu,\sigma^2)} = \tau_{\text{GenPareto}(\boldsymbol{\theta}_0,\xi)} = 0 < \tau_{\text{BGamma}(\boldsymbol{\theta}_\delta)} < \tau_{\text{Normal}(\mu,\sigma^2)} = \infty.$$

Therefore, the tail of the normal distribution is lighter than the tail of the BGamma distribution, which is lighter than the tails of the generalized-Pareto, log-normal, and inverse-gamma distributions.

3.3. Reliability, hazard rate and the mean residual life

For each $t \geqslant 0$, the reliability, the hazard rate and the mean residual life functions are defined as

$$R(t;\boldsymbol{\theta}_\delta) := \int_t^\infty f(x;\boldsymbol{\theta}_\delta)\mathrm{d}x, \quad H(t;\boldsymbol{\theta}_\delta) := \frac{f(t;\boldsymbol{\theta}_\delta)}{R(t;\boldsymbol{\theta}_\delta)},$$
$$\mathrm{MRL}(t;\boldsymbol{\theta}_\delta) := \frac{1}{R(t;\boldsymbol{\theta}_\delta)} \int_t^\infty R(x;\boldsymbol{\theta}_\delta)\mathrm{d}x,$$

respectively.

Let $Y \sim \mathrm{BGamma}(\boldsymbol{\theta}_0)$. Integration by parts gives

$$\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y = \frac{\mathrm{e}^{-\beta t}}{\beta}t^\alpha + \frac{\alpha}{\beta}\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}, \tag{6}$$

$$\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y^2 = \frac{\mathrm{e}^{-\beta t}}{\beta}t^\alpha\left(t + \frac{\alpha+1}{\beta}\right) + \frac{\alpha(\alpha+1)}{\beta^2}\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}, \tag{7}$$

$$\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y^3 = \frac{\mathrm{e}^{-\beta t}}{\beta}t^\alpha\left[t^2 + \frac{\alpha+2}{\beta}t + \frac{(\alpha+1)(\alpha+2)}{\beta^2}\right] + \frac{\alpha(\alpha+1)(\alpha+2)}{\beta^3}\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}. \tag{8}$$

Proposition 3.13: *If* $X \sim \mathrm{BGamma}(\boldsymbol{\theta}_\delta)$ *then*

(1) *Reliability function*: $R(t;\boldsymbol{\theta}_\delta) = \frac{\frac{\delta t}{\beta}[\delta(t+\frac{\alpha+1}{\beta})-2]}{Z(\boldsymbol{\theta}_\delta)}f(t;\boldsymbol{\theta}_0) + R(t;\boldsymbol{\theta}_0)$.

(2) *Cumulative distribution function*: $F(t;\boldsymbol{\theta}_\delta) = -\frac{\frac{\delta t}{\beta}[\delta(t+\frac{\alpha+1}{\beta})-2]}{Z(\boldsymbol{\theta}_\delta)}f(t;\boldsymbol{\theta}_0) + F(t;\boldsymbol{\theta}_0)$.

(3) *Hazard rate*: $H(t;\boldsymbol{\theta}_\delta) = \frac{[1+(1-\delta t)^2]H(t;\boldsymbol{\theta}_0)}{\frac{\delta t}{\beta}[\delta(t+\frac{\alpha+1}{\beta})-2]H(t;\boldsymbol{\theta}_0)+Z(\boldsymbol{\theta}_\delta)}$,

where $R(t;\boldsymbol{\theta}_0) = 1 - F(t;\boldsymbol{\theta}_0) = \mathbb{E}\mathbb{1}_{\{Y\geqslant t\}} = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_t^\infty y^{\alpha-1}\mathrm{e}^{-\beta y}\,\mathrm{d}y$.

Proof: Since

$$R(t;\boldsymbol{\theta}_\delta) = 2\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}} - 2\delta\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y + \delta^2\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y^2, \quad Y \sim \mathrm{BGamma}(\boldsymbol{\theta}_0),$$

using the identities (6) and (7), the proof of Item (1) follows. The proof of items (2) and (3) follows directly by combining the definitions of $F(t;\boldsymbol{\theta}_\delta)$ and $H(t;\boldsymbol{\theta}_\delta)$ with Item (1), respectively. ∎

Figure 2 shows some different shapes of the BGamma hazard rate for different combinations of parameters.

Remark 3.14 (Monotonicity of the hazard function when $\delta = 0$): It is well-known that, when $\alpha > 1$, the hazard function $H(t;\boldsymbol{\theta}_0)$ is concave and increasing. When $\alpha < 1$, the hazard function is convex and decreasing. The case $\alpha = 1$ corresponds to the exponential distribution which has constant hazard function.

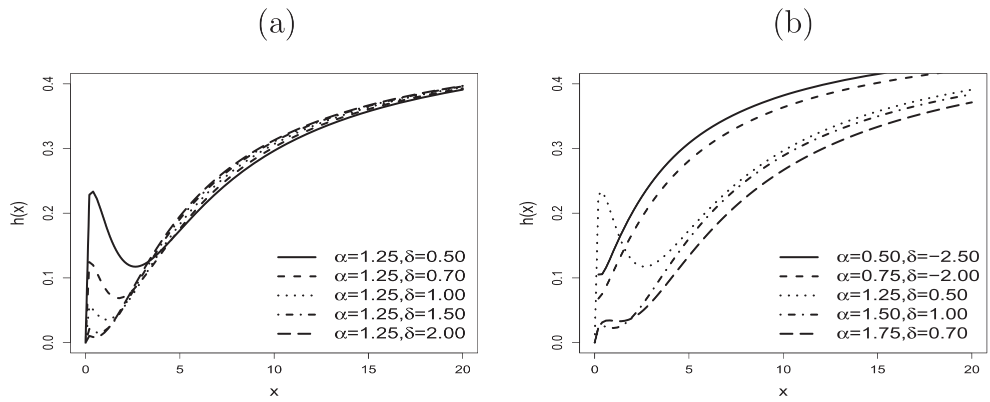

Figure 2. BGamma hazard rate for some parameter values ($\beta = 0.50$).

Proposition 3.15 (Decreasing monotonicity of the hazard rate): *The hazard rate $H(x;\boldsymbol{\theta}_\delta)$ of the* BGamma *distribution* (2) *is decreasing when $\alpha \leqslant 1$, $\delta > 0$ and $x \in (\frac{1}{\delta} - \frac{\alpha+1}{2\beta}, \frac{1}{\delta})$.*

Proof: By Proposition 3.13,

$$H(x;\boldsymbol{\theta}_\delta) = \frac{1 + (1-\delta x)^2}{\frac{\delta x}{\beta}\left[\delta(x + \frac{\alpha+1}{\beta}) - 2\right] + \frac{Z(\boldsymbol{\theta}_\delta)}{H(x;\boldsymbol{\theta}_0)}}. \tag{9}$$

A straightforward computation shows that the function $x \mapsto 1 + (1-\delta x)^2$ decreases when $x < 1/\delta$ and that the function $x \mapsto \frac{\delta x}{\beta}[\delta(x + \frac{\alpha+1}{\beta}) - 2]$ increases when $x > \frac{1}{\delta} - \frac{\alpha+1}{2\beta}$. Then, using Remark 3.14, $x \mapsto \frac{\delta x}{\beta}[\delta(x + \frac{\alpha+1}{\beta}) - 2] + \frac{Z(\boldsymbol{\theta}_\delta)}{H(x;\boldsymbol{\theta}_0)}$ is an increasing function when $\alpha \leqslant 1$ and $x > \frac{1}{\delta} - \frac{\alpha+1}{2\beta}$. Hence, by expression (9), the hazard rate is the product of two decreasing and non-negative functions when $\alpha \leqslant 1, \delta > 0$ and $x \in (\frac{1}{\delta} - \frac{\alpha+1}{2\beta}, \frac{1}{\delta})$. The proof follows. ■

Proposition 3.16 (Increasing monotonicity of the hazard rate): *The hazard rate $H(x;\boldsymbol{\theta}_\delta)$ of the* BGamma *distribution (2) is increasing in the following cases:*

(1) *for $\delta = \beta, \alpha = 1$ and $x \in (0, 2/\beta)$;*
(2) *for $\delta > \beta$, $\alpha = 1$ and $x \in (\delta + \beta - \sqrt{\delta^2 - \beta^2})/(\beta\delta), \delta + \beta + \sqrt{\delta^2 - \beta^2})/(\beta\delta))$;*
(3) *under the conditions $\alpha > 1$, $\delta > \beta$, $a_{\delta,\beta} > 0$, $b_{\delta,\beta} > 0$ and $c_{\delta,\beta} > 0$, for $x \in (0, x_1)$ or $x \in (x_2, x_3)$, where x_1, x_2, x_3 are the three distinct positive roots of the polynomial equation $p_3(x) = 0$;*
(4) *under the conditions $\alpha > 1$, $0 < \delta < \beta$, $a_{\delta,\beta} < 0$, $b_{\delta,\beta} < 0$ and $c_{\delta,\beta} < 0$ and $x \in (0, x_0)$, where x_0 is the unique positive root of $p_3(x) = 0$;*

where $p_3(x) = \beta\delta^2 x^3 - \delta[2(\delta+\beta) + \delta(\alpha-1)]x^2 + 2[\delta+\beta+2(\alpha-1)]x - 2(\alpha-1)$, *and $a_{\delta,\beta}, b_{\delta,\beta}$ and $c_{\delta,\beta}$ are as in* (3), (4) *and* (5), *respectively.*

Proof: As a sub-product of the proof of Theorems 3.2 and 3.3, note that the density $f(x;\boldsymbol{\theta}_\delta)$ is increasing on the above mentioned intervals. Since $H(x;\boldsymbol{\theta}_\delta)=\frac{f(x;\boldsymbol{\theta}_\delta)}{R(x;\boldsymbol{\theta}_\delta)}$ and $R(x;\boldsymbol{\theta}_\delta)$ is a decreasing function, in this case, we have that the hazard rate function is the product of the two increasing and non-negative functions, then the proof of Items (1)-(4) follows. ∎

Proposition 3.17: *If* $X\sim \mathrm{BGamma}(\boldsymbol{\theta}_\delta)$, *then*

$$\mathbb{E}\mathbb{1}_{\{X\geqslant t\}}X=\frac{\big\{2+\delta^2[t^2+\frac{\alpha+2}{\beta}t+\frac{(\alpha+1)(\alpha+2)}{\beta^2}]-2\delta(t+\frac{\alpha+1}{\beta})\big\}\beta^{\alpha-1}t^\alpha \mathrm{e}^{-\beta t}}{Z(\boldsymbol{\theta}_\delta)\Gamma(\alpha)}+\frac{\alpha}{\beta}\,\frac{\Big[2+\frac{\delta^2}{\beta^2}(\alpha+1)(\alpha+2)-2\frac{\delta}{\beta}(\alpha+1)\Big]R(t;\boldsymbol{\theta}_0)}{Z(\boldsymbol{\theta}_\delta)}.$$

Proof: Since

$$\mathbb{E}\mathbb{1}_{\{X\geqslant t\}}X=2\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y-2\delta\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y^2+\delta^2\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y^3,\quad Y\sim \mathrm{BGamma}(\boldsymbol{\theta}_0),$$

using the identities (6), (7) and (8), the proof follows. ∎

Remark 3.18 (Mean residual life function): Integration by parts gives

$$\mathbb{E}\mathbb{1}_{\{X\geqslant t\}}X=tR(t;\boldsymbol{\theta}_\delta)+\int_t^\infty R(x;\boldsymbol{\theta}_\delta)\,\mathrm{d}x,$$

since $xR(x;\boldsymbol{\theta}_\delta)\to 0$ as $x\to\infty$. Then

$$\mathrm{MRL}(t;\boldsymbol{\theta}_\delta)=\left[\frac{1}{R(t;\boldsymbol{\theta}_\delta)}\mathbb{E}\mathbb{1}_{\{X\geqslant t\}}X\right]-t,$$

where $R(t;\boldsymbol{\theta}_\delta)$ and $\mathbb{E}\mathbb{1}_{\{X\geqslant t\}}X$ are given in Propositions 3.13 and 3.17, respectively.

Remark 3.19: In the particular case $\delta=0$, note that

$$\mathbb{E}\mathbb{1}_{\{Y\geqslant t\}}Y=\frac{\beta^{\alpha-1}}{\Gamma(\alpha)}t^\alpha\mathrm{e}^{-\beta t}+\left(\frac{\alpha}{\beta}\right)R(t;\boldsymbol{\theta}_0),\quad Y\sim \mathrm{BGamma}(\boldsymbol{\theta}_0).$$

Then, by Remark 3.18,

$$\mathrm{MRL}(t;\boldsymbol{\theta}_0)=\frac{\beta^{\alpha-1}t^\alpha\mathrm{e}^{-\beta t}}{\Gamma(\alpha)R(t;\boldsymbol{\theta}_0)}+\left(\frac{\alpha}{\beta}\right)-t.$$

The above identity was also verified by Govil and Aggarwal [17], Equation (10).

3.4. Entropy measures

Entropy represents the amount of uncertainty of a probability distribution. Some of these measures are particular cases of Renyi's entropy, such as Shannon entropy and Quadratic entropy (Dukkipati [18]).

Let $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$. The Renyi's entropy measure is defined as

$$H_\gamma(X) := -\frac{1}{1-\gamma}\log\int_0^\infty f^\gamma(x;\boldsymbol{\theta}_\delta)\mathrm{d}x, \quad \gamma \geqslant 0 \quad \text{and} \quad \gamma \neq 1,$$

and for the quadratic entropy

$$H_2(X) := -\log\int_0^\infty f^2(x;\boldsymbol{\theta}_\delta)\,\mathrm{d}x.$$

We also define the Shannon entropy as

$$H_1(X) := -\int_0^\infty f(x;\boldsymbol{\theta}_\delta)\log f(x;\boldsymbol{\theta}_\delta)\,\mathrm{d}x.$$

Proposition 3.20 (Quadratic entropy): *If* $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$ *and* $\alpha > 1$, *then*

$$H_2(X) = \log 2 + \frac{1}{2}\log\pi + 2\log\Gamma(\alpha) + \log Z(\boldsymbol{\theta}_\delta) - \log\left[1+\delta^2\sigma^2+(1-\delta\mu)^2\right]$$
$$-\log\beta - \log(\alpha-1) - \log\Gamma(\alpha-1) - \log\Gamma\big(\alpha - \frac{1}{2}\big),$$

where μ *and* σ^2 *are as in Corollary* 3.6.

Proof: A straightforward computation shows that

$$\int_0^\infty f^2(x;\boldsymbol{\theta}_\delta)\,\mathrm{d}x = \frac{\beta\Gamma(2\alpha-1)}{2^{2\alpha-1}Z(\boldsymbol{\theta}_\delta)\Gamma^2(\alpha)}\left[1+\mathbb{E}(1-\delta X)^2\right],$$

where $X \sim \text{BGamma}(2\alpha-1, 2\beta, \delta)$ and $\mathbb{E}(1-\delta X)^2 = \delta^2\sigma^2 + (1-\delta\mu)^2$.

Combining the formulas $\Gamma(2\alpha+1) = 2\alpha\Gamma(2\alpha)$ and $\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}}\Gamma(z)\Gamma(z+\frac{1}{2})$, the expression of the right-hand side can be written as

$$= \frac{\beta(\alpha-1)\Gamma(\alpha-1)\Gamma\big(\alpha-\frac{1}{2}\big)}{2Z(\boldsymbol{\theta}_\delta)\sqrt{\pi}\Gamma^2(\alpha)}\left[1+\delta^2\sigma^2+(1-\delta\mu)^2\right].$$

Finally, taking logarithm and multiplying by -1 on both sides of the above identity, we complete the proof. ■

Proposition 3.21 (Shannon entropy): *Let* $X \sim \text{BGamma}(\boldsymbol{\theta}_\delta)$. *The Shannon entropy is given by*

$$H_1(X) = \log Z(\boldsymbol{\theta}_\delta) + \log\Gamma(\alpha) - \alpha\log\beta - \frac{1}{Z(\boldsymbol{\theta}_\delta)}\phi(s)\Big|_{s=1}$$
$$-(\alpha-1)\frac{\frac{\delta}{\beta}\big[\frac{(2\alpha+1)\delta}{\beta}-2\big]+\big[2-\frac{2\delta}{\beta}+\frac{\alpha(\alpha+1)\delta^2}{\beta^2}\big](\Psi^{(0)}(\alpha)-\log\beta)}{Z(\boldsymbol{\theta}_\delta)} + \beta\mu,$$

where $\phi(s) := \frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[1+(1-\delta Y)^2]^s$, $Y \sim \text{BGamma}(\boldsymbol{\theta}_0)$, *and* $\phi(s)|_{s=1}$ *exists. Here,* μ *is as in Corollary* 3.6 *and* $\Psi^{(m)}(z)$ *is the polygamma function of order* m *defined by* $\frac{\mathrm{d}^{m+1}}{\mathrm{d}z^{m+1}}\log\Gamma(z)$.

Proof: Note that the Shannon entropy can be rewritten as

$$H_1(X) = \log Z(\boldsymbol{\theta}_\delta) + \log \Gamma(\alpha) - \alpha \log \beta - \mathbb{E} \log g(X) - (\alpha - 1)\mathbb{E} \log X + \beta \mathbb{E}X,$$

where $Z(\boldsymbol{\theta}_\delta) = 2 + \frac{\alpha\delta}{\beta}[(1+\alpha)\frac{\delta}{\beta} - 2)]$ and $g(x) \coloneqq 1 + (1-\delta x)^2$. The expectation $\mathbb{E} \log X$ was obtained in Proposition 3.8 and $\mathbb{E}X = \mu$ is as in Corollary 3.6. By Teh et al. [19], we can approximate the function $\log g(x)$ using a second-order Taylor expansion about $\mathbb{E}g(X)$ and evaluate its expectation as follows:

$$\mathbb{E} \log g(X) \approx \log \mathbb{E}g(X) - \frac{\mathrm{Var}[g(X)]}{2\mathbb{E}g^2(X)}.$$

Since $\mathbb{E}X^\nu < \infty$, for each $\nu > -\alpha$ (see Proposition 3.5), we have that $\log \mathbb{E}g(X) < \infty$ and $\mathbb{E}g^2(X) < \infty$. Then, $\mathbb{E} \log g(X)$ exists. Finally, since

$$\begin{aligned}
\mathbb{E} \log g(X) &= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \mathbb{E}\left[g(Y)^s \log g(Y)\right]\Big|_{s=1}, \quad Y \sim \mathrm{BGamma}(\boldsymbol{\theta}_0) \\
&= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}s} g(Y)^s\right]\Big|_{s=1} = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \phi(s)\Big|_{s=1},
\end{aligned}$$

the proof follows. ■

4. Maximum likelihood estimation

Let X be a random variable with BGamma distribution $f(x; \boldsymbol{\theta}_\delta)$ that depends on a parameter vector $\boldsymbol{\theta}_\delta = (\alpha, \beta, \delta)^{\mathrm{T}}$ and let $(X_1, \ldots, X_n)$ be a random sample of X (i.e. the random variables $X_1, \ldots, X_n$ are independent and identically distributed with BGamma distribution) for $\boldsymbol{\theta}_\delta$ in an open subset (parameter space) Θ of $\mathbb{R}^3$, where distinct values of $\boldsymbol{\theta}_\delta$ yield distinct distributions for X_1. Denoting $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ as the corresponding observed values of the random sample $(X_1, \ldots, X_n)$, the log-likelihood function for $\boldsymbol{\theta}_\delta$ is given by

$$\begin{aligned}
l(\boldsymbol{\theta}_\delta; \mathbf{x}) = -\log Z(\boldsymbol{\theta}_\delta) &+ \sum_{i=1}^{n} \log\left[1 + (1-\delta x_i)^2\right] \\
&+ \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \log x_i - n\beta\overline{x},
\end{aligned} \tag{10}$$

where $Z(\boldsymbol{\theta}_\delta) = 2 + \frac{\alpha\delta}{\beta}[(1+\alpha)\frac{\delta}{\beta} - 2]$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. The first-order partial derivatives and the second-order (and mixed) partial derivatives of $Z(\boldsymbol{\theta}_\delta)$ are given by

$$\begin{aligned}
&\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial \alpha} = \frac{\delta}{\beta}\left[(1+2\alpha)\frac{\delta}{\beta} - 2\right], \quad \frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial \beta} = -\frac{2\alpha\delta}{\beta^2}\left[(1+\alpha)\frac{\delta}{\beta} - 1\right], \\
&\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial \delta} = \frac{2\alpha}{\beta}\left[(1+\alpha)\frac{\delta}{\beta} - 1\right], \quad \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial \alpha^2} = \frac{2\delta^2}{\beta^2}, \\
&\frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial \beta^2} = \frac{2\alpha\delta}{\beta^3}\left[(1+\alpha)\frac{3\delta}{\beta} - 2\right], \quad \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial \delta^2} = \frac{2\alpha(1+\alpha)}{\beta^2};
\end{aligned} \tag{11}$$

and

$$\frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial\alpha\partial\beta} = \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial\beta\partial\alpha} = -\frac{2\delta}{\beta^2}\left[(1+2\alpha)\frac{\delta}{\beta} - 1\right],$$
$$\frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial\alpha\partial\delta} = \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial\delta\partial\alpha} = \frac{2}{\beta}\left[(1+2\alpha)\frac{\delta}{\beta} - 1\right],$$
$$\frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial\beta\partial\delta} = \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial\delta\partial\beta} = -\frac{2\alpha}{\beta^2}\left[(1+\alpha)\frac{2\delta}{\beta} - 1\right].$$

Note that $f(x;\boldsymbol{\theta}_\delta)$ is a positive, differentiable function of $\boldsymbol{\theta}_\delta = (\alpha,\beta,\delta)^{\mathrm{T}}$. If a supremum $\widehat{\boldsymbol{\theta}}$ exists, it must satisfy the likelihood equations

$$\frac{\partial l(\widehat{\boldsymbol{\theta}};\mathbf{x})}{\partial\alpha} = 0, \quad \frac{\partial l(\widehat{\boldsymbol{\theta}};\mathbf{x})}{\partial\beta} = 0, \quad \frac{\partial l(\widehat{\boldsymbol{\theta}};\mathbf{x})}{\partial\delta} = 0. \tag{12}$$

Any (non-trivial) root of the likelihood equations (12) is called an ML estimator in the loose sense. In the case that the parameter value provides the absolute maximum of $l(\boldsymbol{\theta}_\delta;\mathbf{x})$, it is called an ML estimator in the strict sense.

Also notice that, using the polygamma function of order m, $\Psi^{(m)}(z) = \frac{\mathrm{d}^{m+1}}{\mathrm{d}z^{m+1}}\log\Gamma(z)$, the first-order partial derivatives of $l(\boldsymbol{\theta}_\delta;\mathbf{x})$ are

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\alpha} &= -\frac{1}{Z(\boldsymbol{\theta}_\delta)}\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial\alpha} + \log\beta - \Psi^{(0)}(\alpha) + \sum_{i=1}^{n}\log x_i \\
&= -\frac{\delta[(1+2\alpha)\delta - 2\beta]}{2\beta^2 + \alpha\delta[(1+\alpha)\delta - 2\beta]} + \log\beta - \Psi^{(0)}(\alpha) + \sum_{i=1}^{n}\log x_i, \\
\frac{\partial l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\beta} &= -\frac{1}{Z(\boldsymbol{\theta}_\delta)}\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial\beta} + \frac{\alpha}{\beta} - n\bar{x} \\
&= \frac{2\alpha\delta[(1+\alpha)\delta - \beta]}{2\beta^3 + \alpha\delta\beta[(1+\alpha) - 2\beta]} + \frac{\alpha}{\beta} - n\bar{x}, \\
\frac{\partial l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\delta} &= -\frac{1}{Z(\boldsymbol{\theta}_\delta)}\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial\delta} - 2\sum_{i=1}^{n}\frac{1-\delta x_i}{1+(1-\delta x_i)^2} \\
&= -\frac{2\alpha[(1+\alpha)\delta - \beta]}{2\beta^2 + \alpha\delta[(1+\alpha)\delta - 2\beta]} - 2\sum_{i=1}^{n}\frac{1-\delta x_i}{1+(1-\delta x_i)^2}.
\end{aligned} \tag{13}$$

Since the equations in (12) are not linear, numerical methods will be used to solve the problem. The solutions were found using Nelder–Mead method, since it is popular for unconstrained optimization and it is parsimonious in function evaluations per iteration [20].

The second-order partial derivatives of $l(\boldsymbol{\theta}_\delta;\mathbf{x})$ can be written as

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\alpha^2} &= D_{\boldsymbol{\theta}_\delta}(\alpha,\alpha) - \Psi^{(1)}(\alpha),\\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\beta^2} &= D_{\boldsymbol{\theta}_\delta}(\beta,\beta) - \frac{\alpha}{\beta^2},\\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\delta^2} &= D_{\boldsymbol{\theta}_\delta}(\delta,\delta) + 2\sum_{i=1}^{n}\frac{x_i[1-(1-\delta x_i)^2]}{[1+(1-\delta x_i)^2]^2};\end{aligned} \tag{14}$$

and the second-order mixed derivatives of $l(\boldsymbol{\theta}_\delta;\mathbf{x})$ are given by

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\alpha\partial\beta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\beta\partial\alpha} = D_{\boldsymbol{\theta}_\delta}(\alpha,\beta) + \frac{1}{\beta},\\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\alpha\partial\delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\delta\partial\alpha} = D_{\boldsymbol{\theta}_\delta}(\alpha,\delta),\\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\beta\partial\delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta;\mathbf{x})}{\partial\delta\partial\beta} = D_{\boldsymbol{\theta}_\delta}(\beta,\delta),\end{aligned}$$

where $D_{\boldsymbol{\theta}_\delta}(u,v) := \frac{1}{Z(\boldsymbol{\theta}_\delta)}\left[\frac{1}{Z(\boldsymbol{\theta}_\delta)}\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial u}\frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial v} - \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial u\partial v}\right]$, $u,v\in\{\alpha,\beta,\delta\}$. Here, by the well-known Schwarz's Theorem, the mixed partial differentiations are commutative at a given point $\boldsymbol{\theta}_\delta$ in $\mathbb{R}^3$ because the corresponding functions have continuous second partial derivatives at that point.

If $X\sim \text{BGamma}(\boldsymbol{\theta}_\delta)$, under mild regularity conditions the Fisher information matrix is given by

$$I_X(\boldsymbol{\theta}_\delta) = -\begin{bmatrix} D_{\boldsymbol{\theta}_\delta}(\alpha,\alpha)-\Psi^{(1)}(\alpha) & D_{\boldsymbol{\theta}_\delta}(\alpha,\beta)+\frac{1}{\beta} & D_{\boldsymbol{\theta}_\delta}(\alpha,\delta)\\ D_{\boldsymbol{\theta}_\delta}(\alpha,\beta)+\frac{1}{\beta} & D_{\boldsymbol{\theta}_\delta}(\beta,\beta)-\frac{\alpha}{\beta^2} & D_{\boldsymbol{\theta}_\delta}(\beta,\delta)\\ D_{\boldsymbol{\theta}_\delta}(\alpha,\delta) & D_{\boldsymbol{\theta}_\delta}(\beta,\delta) & D_{\boldsymbol{\theta}_\delta}(\delta,\delta)+2\mathbb{E}\frac{X[1-(1-\delta X)^2]}{[1+(1-\delta X)^2]^2}\end{bmatrix},$$

where $\mathbb{E}|\frac{X[1-(1-\delta X)^2]}{[1+(1-\delta X)^2]^2}| \leqslant 2\mathbb{E}X - 2\delta\mathbb{E}X^2 + \delta^2\mathbb{E}X^3 < \infty$, see Proposition 3.5.

Theorem 4.1: *Let* $\Theta=\{\alpha\in\mathbb{R}^+ : \varepsilon_0<\alpha<\alpha_0\}$ *be the parameter space, where* $\varepsilon_0=\varepsilon_0(\beta,\delta)\in(0,\alpha_0)$ *is fixed and* $\alpha_0=\alpha_0(\beta,\delta) := [(2\beta-\delta)+\sqrt{(2\beta-\delta)^2+2\delta(4\beta-\delta)}]/2\delta$ *with* β,δ *known such that* $0<\delta<2\beta$*. Then, with probability approaching 1, as* $n\to\infty$*, the likelihood equation* $\frac{\mathrm{d}l(\alpha;\mathbf{x})}{\mathrm{d}\alpha}=0$ *has a consistent solution, denoted by* $\widehat{\alpha}$*.*

Proof: Since β and δ are known, to simplify the notation, we will write BGamma(α), $f(x;\alpha)$, $Z(\alpha)$ and $D_\alpha(\alpha,\alpha)$ referring to BGamma$(\boldsymbol{\theta}_\delta)$, $f(x;\boldsymbol{\theta}_\delta)$, $Z(\boldsymbol{\theta}_\delta)$ and $D_{\boldsymbol{\theta}_\delta}(\alpha,\alpha)$, respectively.

Let $X\sim$ BGamma(α). By Cramér [21], it is sufficient to prove that

(1) $\mathbb{E}\frac{\mathrm{d}\log f(X;\alpha)}{\mathrm{d}\alpha}=0$ for all $\alpha\in\Theta$;

(2) $-\infty < \mathbb{E}\frac{\mathrm{d}^2 \log f(X;\alpha)}{\mathrm{d}\alpha^2} < 0$ for all $\alpha \in \Theta$;
(3) there exists a function $H(x)$ such that for all $\alpha \in \Theta$,

$$\left|\frac{\mathrm{d}^3 \log f(x;\alpha)}{\mathrm{d}\alpha^3}\right| < H(x) \quad \text{and} \quad \mathbb{E}H(X) = M(\alpha) < \infty.$$

Indeed, taking $n = 1$ in (13), we have

$$\frac{\mathrm{d} \log f(x;\alpha)}{\mathrm{d}\alpha} = -\frac{1}{Z(\alpha)}\frac{\mathrm{d}Z(\alpha)}{\mathrm{d}\alpha} + \log\beta - \Psi^{(0)}(\alpha) + \log x.$$

Then,

$$\mathbb{E}\frac{\mathrm{d} \log f(X;\alpha)}{\mathrm{d}\alpha} = \mathbb{E}\log X - \frac{1}{Z(\alpha)}\frac{\mathrm{d}Z(\alpha)}{\mathrm{d}\alpha} + \log\beta - \Psi^{(0)}(\alpha).$$

Using Proposition (3.8)-(1) and the identities in (11), a straightforward computation shows that

$$\mathbb{E}\log X - \frac{1}{Z(\alpha)}\frac{\mathrm{d}Z(\alpha)}{\mathrm{d}\alpha} = \Psi^{(0)}(\alpha) - \log\beta.$$

Therefore, $\mathbb{E}\frac{\mathrm{d} \log f(X;\alpha)}{\mathrm{d}\alpha} = 0$ for all $\alpha \in \Theta$, and the Item (1) is proved.

Taking $n = 1$ in (14), using the definition of $D_\alpha(\alpha,\alpha)$ and the identities in (11), it follows that

$$\begin{aligned}\frac{\mathrm{d}^2 \log f(x;\alpha)}{\mathrm{d}\alpha^2} &= D_\alpha(\alpha,\alpha) - \Psi^{(1)}(\alpha)\\ &= \frac{\delta^3/\beta^4}{Z^2(\alpha)}\left[2\delta\alpha^2 - 2(2\beta-\delta)\alpha - (4\beta-\delta)\right] - \Psi^{(1)}(\alpha). \qquad (15)\end{aligned}$$

Since $0 < \delta < 2\beta$, α_0 is well defined. For $0 < \alpha < \alpha_0$, note that $2\delta\alpha^2 - 2(2\beta-\delta)\alpha - (4\beta-\delta) < 0$. On the other hand, its known that $\Psi^{(1)}(\alpha) > \frac{\mathrm{e}^{1/\alpha}}{(\mathrm{e}^{1/\alpha}-1)\alpha^2} > 0$ (see [22], Corollary 1.2). Therefore, $\frac{\mathrm{d}^2 \log f(x;\alpha)}{\mathrm{d}\alpha^2} < 0$ for all $x > 0$. Hence, the Item (2) is satisfied.

To prove Item (3), deriving with respect to α in (15) we obtain

$$\begin{aligned}\frac{\mathrm{d}^3 \log f(x;\alpha)}{\mathrm{d}\alpha^3} = \frac{\delta^3/\beta^6}{Z^3(\alpha)}\Big\{&-2\delta\left[(1+2\alpha)\delta - 2\beta\right]\left[2\delta\alpha^2 - 2(2\beta-\delta)\alpha - (4\beta-\delta)\right]\\ &+ \left[4\alpha\delta - 2(2\beta-\delta)\right]\left\{2\beta^2 + \alpha\delta[(1+\alpha)\delta - 2\beta]\right\}\Big\} - \Psi^{(2)}(\alpha).\end{aligned}$$

Let $G(\alpha) := \frac{\delta^3/\beta^6}{Z^3(\alpha)}\{2\delta[(1+2\alpha)\delta+2\beta][2\delta\alpha^2+2(2\beta+\delta)\alpha+(4\beta+\delta)]+[4\alpha\delta+2(2\beta+\delta)]\{2\beta^2+\alpha\delta[(1+\alpha)\delta+2\beta]\}\}$. Then, for all $x > 0$ and $\alpha \in \Theta$,

$$\left|\frac{\mathrm{d}^3 \log f(x;\alpha)}{\mathrm{d}\alpha^3}\right| \leqslant G(\alpha) + |\Psi^{(2)}(\alpha)|. \qquad (16)$$

Since $G(\alpha)$ is an increasing function in α and $\alpha < \alpha_0$, we have

$$G(\alpha) \leqslant G(\alpha_0), \quad \text{forall } \alpha \in \Theta. \qquad (17)$$

Table 1. Simulated values of biases (MSEs within parentheses) of the estimators of the BGamma model.

		BG($\alpha=0.50,\beta=1.00,\delta$)		BG($\alpha=1.00,\beta=1.00,\delta$)		BG($\alpha=1.50,\beta=1.00,\delta$)	
n	δ	Bias($\widehat{\alpha}$)	Bias($\widehat{\beta}$)	Bias($\widehat{\alpha}$)	Bias($\widehat{\beta}$)	Bias($\widehat{\alpha}$)	Bias($\widehat{\beta}$)
20	−10	0.4053 (0.7019)	0.2028 (0.1850)	0.5157 (1.3893)	0.2018 (0.2165)	0.5574 (2.2031)	0.1770 (0.2427)
	−5	0.2785 (0.3735)	0.1617 (0.1328)	0.4580 (1.0668)	0.1895 (0.1878)	0.5453 (1.8772)	0.1795 (0.2199)
	1	0.0336 (0.0225)	0.1196 (0.1426)	0.1014 (0.0965)	0.0810 (0.0592)	0.2176 (0.3481)	0.0931 (0.0685)
	5	0.3599 (0.6604)	0.1661 (0.1435)	0.6729 (1.8993)	0.2345 (0.2417)	0.7525 (2.7966)	0.2253 (0.2561)
	10	0.5653 (1.2008)	0.2435 (0.2370)	0.4242 (2.0143)	0.1161 (0.3131)	0.6814 (2.7151)	0.2084 (0.2587)
60	−10	0.1294 (0.1080)	0.0632 (0.0302)	0.1567 (0.2546)	0.0601 (0.0391)	0.0321 (0.5791)	0.0021 (0.0697)
	−5	0.0774 (0.0503)	0.0449 (0.0215)	0.1392 (0.1843)	0.0565 (0.0330)	0.0835 (0.4332)	0.0207 (0.0561)
	1	−0.0033 (0.0059)	0.0134 (0.0191)	0.0287 (0.0241)	0.0225 (0.0142)	0.0623 (0.0699)	0.0268 (0.0152)
	5	0.0766 (0.0516)	0.0373 (0.0172)	0.2218 (0.3057)	0.0753 (0.0391)	0.2363 (0.5169)	0.0692 (0.0459)
	10	0.1912 (0.1766)	0.0798 (0.0358)	−0.2430 (0.5223)	−0.1394 (0.1098)	0.1682 (0.5757)	0.0482 (0.0551)
120	−10	0.0626 (0.0458)	0.0286 (0.0124)	0.0767 (0.1162)	0.0275 (0.0169)	−0.1228 (0.3672)	−0.0502 (0.0448)
	−5	0.0327 (0.0206)	0.0173 (0.0089)	0.0677 (0.0821)	0.0257 (0.0141)	−0.0376 (0.2493)	−0.0217 (0.0328)
	1	−0.0107 (0.0031)	−0.0054 (0.0085)	0.0133 (0.0118)	0.0090 (0.0065)	0.0298 (0.0321)	0.0114 (0.0068)
	5	0.0273 (0.0169)	0.0118 (0.0068)	0.1101 (0.1292)	0.0354 (0.0159)	0.1176 (0.2350)	0.0326 (0.0197)
	10	0.0944 (0.0706)	0.0373 (0.0140)	−0.4445 (0.3810)	−0.2162 (0.0848)	0.0386 (0.3181)	0.0073 (0.0300)

Combining the inequalities $\Psi^{(n)}(\alpha) > -(n-1)!\mathrm{e}^{-n\Psi^{(0)}(\alpha)}$, for n even (see the inequality just below Item (2.9) from [23]), and $\Psi^{(0)}(\alpha) > \log(\alpha+\frac{1}{2}) - \frac{1}{\alpha}$ (see [24]), we have that $-\Psi^{(2)}(\alpha) < \mathrm{e}^{2[\frac{1}{\alpha}-\log(\alpha+\frac{1}{2})]} < \mathrm{e}^{2[\frac{1}{\varepsilon_0}-\log(\varepsilon_0+\frac{1}{2})]}$, for all $\alpha\in\Theta$. On the other hand, by Corollary 1.2 from [22], $\Psi^{(2)}(\alpha) < \frac{\mathrm{e}^{1/\alpha}[1-2\alpha(\mathrm{e}^{1/\alpha}-1)]}{(\mathrm{e}^{1/\alpha}-1)^2\alpha^4} < 0$. Therefore,

$$|\Psi^{(2)}(\alpha)| = -\Psi^{(2)}(\alpha) < \mathrm{e}^{2[\frac{1}{\varepsilon_0}-\log(\varepsilon_0+\frac{1}{2})]}. \tag{18}$$

Combining (16), (17) and (18),

$$\left|\frac{\mathrm{d}^3\log f(x;\alpha)}{\mathrm{d}\alpha^3}\right| < G(\alpha_0) + \mathrm{e}^{2[\frac{1}{\varepsilon_0}-\log(\varepsilon_0+\frac{1}{2})]}, \quad \text{for all } \alpha\in\Theta.$$

Taking $H(x) = G(\alpha_0) + \mathrm{e}^{2[\frac{1}{\varepsilon_0}-\log(\varepsilon_0+\frac{1}{2})]}$ = constant, the proof of Item (3) follows. Thus, the proof of theorem is complete. ■

4.1. Monte Carlo simulation

We here carry out a Monte Carlo simulation study to evaluate the performance of the ML estimators of the BGamma model. All numerical evaluations were done in the `R` software [`www.r-project.org`]. The simulation study considers the following scenario: sample size $n\in\{10,60,120\}$, true shape parameter $\alpha\in\{0.50,1.00,1.50\}$, true scale parameter $\beta\in\{1.00\}$, true value of the asymmetric parameter as $\delta\in\{-10,-5,-1,1,5,10\}$, with 5000 Monte Carlo replications for each sample size.

For each value of the parameter δ and sample size, the empirical values for the bias and mean squared error (MSE) of the ML estimators are reported in Table 1. A look at the results in this table allows us to conclude that, as the sample size increases, the bias and MSE of all the estimators decrease, indicating that they are asymptotically unbiased, as expected.

5. The BGamma(θ_δ) regression model with censored data

In many practical applications, the lifetimes are affected by explanatory variables such as sex, age, grade of disease, tumour thickness and several others. So, it is important to explore the relationship between the response variable and the explanatory variables. Regression models can be proposed in different forms in statistical analysis.

In this section, we define a parametric regression model using the new distribution with censored data, called the BGamma($\boldsymbol{\theta}_\delta$) regression model, for reliability or survival analysis as a feasible alternative to the location-scale regression model. Considering that the BGamma($\boldsymbol{\theta}_\delta$) and BGamma($\boldsymbol{\theta}_0$) regression models are embedded models, the likelihood-ratio (LR) test can be used to discriminate between these models. We adopt a classic frequentist analysis for the BGamma($\boldsymbol{\theta}_\delta$) regression model.

Regression analysis of lifetimes involves specifications for the lifetime distribution of X given a vector of covariates denoted by $\mathbf{v} = (v_1, \dots, v_p)^{\mathrm{T}}$. Here, we relate the parameters α and β to covariates by the logarithm link functions $\alpha_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)$ and $\beta_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)$, $i = 1, \dots, n$, respectively, where $\boldsymbol{\tau}_1 = (\tau_{11}, \dots, \tau_{1p})^{\mathrm{T}}$ and $\boldsymbol{\tau}_2 = (\tau_{21}, \dots, \tau_{2p})^{\mathrm{T}}$ denote the vectors of regression coefficients and $\mathbf{v}_i^T = (v_{i1}, \dots, v_{ip})$.

The survival function of $X \mid \mathbf{v}$ follows from (Proposition 3.13) as

$$
\begin{aligned}
S(x \mid \mathbf{v}) = 1 + &\frac{\delta x^{\exp(\mathbf{v}^T \boldsymbol{\tau}_1)} \exp(\mathbf{v}^T \boldsymbol{\tau}_2)^{\exp(\mathbf{v}^T \boldsymbol{\tau}_1) - 1} \exp[-\exp(\mathbf{v}^T \boldsymbol{\tau}_2) x]}{\Gamma[\exp(\mathbf{v}^T \boldsymbol{\tau}_1)] Z(\boldsymbol{\theta}_\delta)} \\
&\times \left\{ \delta \left[x + \frac{\exp(\mathbf{v}^T \boldsymbol{\tau}_1) - 1}{\exp(\mathbf{v}^T \boldsymbol{\tau}_2)} \right] - 2 \right\} - I(\exp(\mathbf{v}^T \boldsymbol{\tau}_1), \exp(\mathbf{v}^T \boldsymbol{\tau}_2) x), \qquad (19)
\end{aligned}
$$

where $I(k, y) = \gamma(k, y)/\Gamma(k)$ is the incomplete gamma ratio function, $\gamma(k, y) = \int_0^y w^{k-1} e^{-w} dw$ is the incomplete gamma function and $\Gamma(\cdot)$ is the gamma function. Equation (19) is referred to as the survival function for the BGamma($\boldsymbol{\theta}_\delta$) regression model, which opens new possibilities for fitting many different types of reliability data.

Consider a sample $(x_1, \mathbf{v}_1), \dots, (x_n, \mathbf{v}_n)$ of n independent observations. We consider that each individual i has a lifetime X_i and a censoring time C_i, where X_i and C_i are independent random variables and the data consist of n independent observations and $x_i = \min(X_i, C_i)$, for $i = 1, \dots, n$. 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 x_i is the lifetime or censoring, respectively. Conventional likelihood estimation techniques can be applied here. The total log-likelihood function for the vector of parameters $\boldsymbol{\psi} = (\delta, \boldsymbol{\tau}_1^T, \boldsymbol{\tau}_2^T)^{\mathrm{T}}$ from model (19) has the form

$$
\begin{aligned}
l(\boldsymbol{\psi}) = &\sum_{i \in F} \log[1 + (1 - \delta x_i)^2] \sum_{i \in F} \log[Z_i(\boldsymbol{\theta}_\delta)] \sum_{i \in F} \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)(\mathbf{v}_i^T \boldsymbol{\tau}_2) \\
&- \sum_{i \in F} \log\{\Gamma[\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)]\} + \sum_{i \in F} (\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1) - 1) \log(x_i) - \sum_{i \in F} \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2) x_i \\
&+ \sum_{i \in C} l_i^{(c)}(\boldsymbol{\psi}), \qquad (20)
\end{aligned}
$$

where

$$l_i^{(c)}(\boldsymbol{\psi}) = \log\left\{1 + \frac{\delta\, x_i^{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)} \exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)^{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)-1} \exp[-\exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)\, x_i]}{\Gamma[\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)]Z_i(\boldsymbol{\theta}_\delta)}\right.$$
$$\left.\times \left\{\delta\left[x_i + \frac{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)-1}{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)}\right] - 2\right\} - I(\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1), \exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)x_i)\right\}$$

and

$$Z_i(\boldsymbol{\theta}_\delta) = 2 + \frac{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)\delta}{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)}\left\{\left[1+\exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)\right]\left[\frac{\delta}{\exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)}\right] - 2\right\}.$$

The MLE $\widehat{\boldsymbol{\psi}}$ of the vector of unknown parameters can be determined by maximizing the log-likelihood (20). We use the R software to compute $\widehat{\boldsymbol{\psi}}$. Initial values for $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are taken from the fit of the BGamma($\boldsymbol{\theta}_0$) regression model.

The multivariate normal $N_{2p+1}(0, J(\widehat{\boldsymbol{\psi}})^{-1})$ distribution under standard regularity conditions can be used to construct approximate confidence intervals for the model parameters. Further, we can compare the BGamma($\boldsymbol{\theta}_\delta$) model with its special models using LR statistics.

6. Applications

In this section, we provide two applications to real data to illustrate the flexibility of the BGamma($\boldsymbol{\theta}_\delta$) model. In the first application, we present a real situation in which the behaviour of the data is bimodal. In the second application, we consider a BGamma($\boldsymbol{\theta}_\delta$) regression model with censored data. In the applications, we determine the MLEs and the corresponding standard errors (SEs) (given in parentheses) of the model parameters and the values of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Cramer–von Mises (W^*) and Kolmogorov–Smirnov (KS) goodness-of-fit statistic for the fitted models. For all cases, the model parameters are estimated by the ML method using the R software.

6.1. Application 1: Wheaton River data

The data are the exceedances of flood peaks (in m^3/s) of the Wheaton River near Carcross in Yukon Territory, Canada. The data consist of 72 exceedances for the years 1958–1984, rounded to one decimal place. These data are presented and analysed by Choulakian and Stephens [25] and Akinsete et al. [26]. In [26], the authors present an analysis considering the following distributions: Pareto, three-parameter Weibull, generalized Pareto and Beta-Pareto. The authors use the KS measurement to select the most appropriate model. In Table 2, we present these values and the associated p-value.

We consider the Kumaraswamy generalized gamma (KumGG) distribution (for $x > 0$) defined by Pascoa et al. [27]. Note that the KumGG distribution contains as particular cases most of the classical distributions used in survival analysis. Hence, the associated density

Table 2. The KS measurements and associated p-value with the Wheaton River data.

Model	KS	p-Value
Pareto	2.7029	< 0.000
Three-parameter Weibull	1.6734	0.0074
Generalized Pareto	1.205	0.1094
Beta Pareto	1.2534	0.0864

Table 3. MLEs of the model parameters for the Wheaton River data.

Model	α	β	δ		
BGamma($\boldsymbol{\theta}_\delta$)	1.054	0.176	0.177		
	(0.145)	(0.111)	(0.032)		
	α	τ	k	λ	φ
Kw-GG	548.542	0.103	0.098	158.570	869.87
	(252.1)	(0.082)	(0.007)	(70.450)	(196.8)
gamma	14.558	1	0.838	1	1
	(2.816)	(–)	(0.121)	(–)	(–)
EW	11.278	1.380	0.591	1	1
	(1.506)	(0.284)	(0.149)	(–)	(–)
Weibull	11.632	0.901	1	1	1
	(1.601)	(0.085)	(–)	(–)	(–)
MW	0.124	0.775	0.010		
	(0.034)	(0.124)	(0.007)		

function with five positive parameters $\alpha,\ \tau,\ k,\ \lambda$ and φ has the form

$$f(x) = \frac{\lambda\varphi\tau}{\alpha\Gamma(k)}\left(\frac{x}{\alpha}\right)^{\tau k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\tau}\right]\left\{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}^{\lambda-1} \times\left(1-\left\{\gamma_1\left[k,\left(\frac{x}{\alpha}\right)^{\tau}\right]\right\}^{\lambda}\right)^{\varphi-1},$$

where $\gamma_1(k,y) = \gamma(k,y)/\Gamma(k)$ is the incomplete gamma ratio function, α is a scale parameter and the other positive parameters $\tau,\ k,\ \varphi$ and λ are shape parameters. This model has as particular cases, exponentiated Weibull (for $\lambda = 1$ and $\varphi = 1$), gamma for (for $\lambda = 1,\ \varphi = 1$ and $\tau = 1$) and Weibull for (for $\lambda = 1,\ \varphi = 1$ and $k = 1$).

The modified Weibull (MW) (for $x \geq 0$) was defined by Lai et al. [28], whose density function with three parameters $\alpha > 0,\ \tau \geq 0$ and $k \geq 0$ is given by

$$f(x) = \alpha x^{(\tau-1)}(\tau + kx)\exp[kx - \alpha x^{\tau}\exp(kx)].$$

The results are reported in Tables 3 and 4. The four statistics agree on the model's ranking. The lowest values of these criteria correspond to the BGamma($\boldsymbol{\theta}_\delta$) distribution, which could be chosen in this case. Also in relation to Table 2, we verified that the *KS* measurement of the proposed model presents smaller values and associated p-value is higher, indicating that the model is adequate to the data of Wheaton River data.

In Figure 3, we present the adjustment of the proposed model in relation to the PDF and CDF; see Figures 3(a,b). In Figure 3(c), we provide the QQ plot for the BGamma($\boldsymbol{\theta}_\delta$) distribution. We note that the quantile residuals follow more approximately a normal distribution for the BGamma($\boldsymbol{\theta}_\delta$) distribution. In fact, these plots reveal that the BGamma($\boldsymbol{\theta}_\delta$) distribution provides a good fit for Wheaton River data.

Table 4. Statistical measures.

Model	AIC	BIC	W^*	KS
BGamma($\boldsymbol{\theta}_\delta$)	501.51	508.34	0.038	0.065 (0.918)
	AIC	BIC	W^*	KS
Kw-GG	514.01	525.39	0.159	0.099 (0.473)
gamma	506.68	511.24	0.130	0.102 (0.433)
EW	505.85	512.68	0.074	0.096 (0.516)
Weibull	506.99	511.55	0.137	0.105 (0.402)
MW	507.34	514.17	0.097	0.100 (0.466)

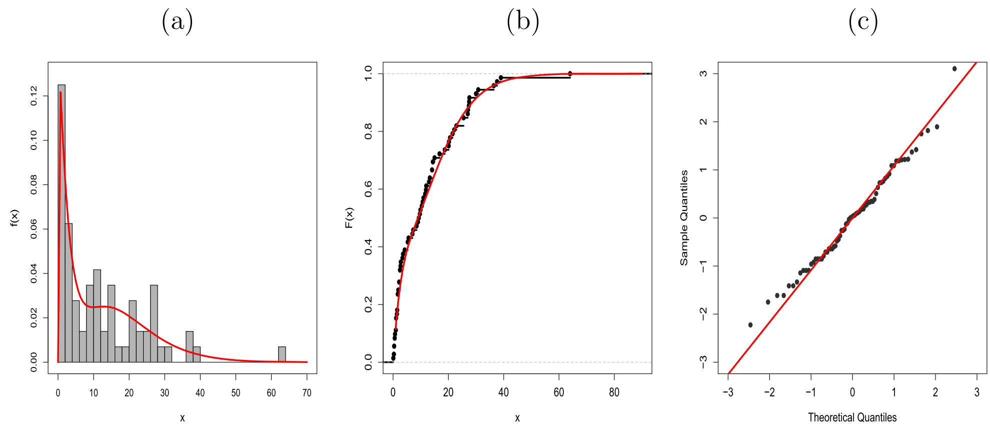

Figure 3. (a) Estimated PDF of the BGamma($\boldsymbol{\theta}_\delta$) model. (b) Empirical CDF and estimated CDF of the BGamma($\boldsymbol{\theta}_\delta$) model. (c) QQ plot for the quantile residual from the fitted BGamma($\boldsymbol{\theta}_\delta$) model to the Wheaton River data.

Table 5. The GD, AIC and BIC measurements for the BG, gamma and Weibull regression models for the gastric cancer.

Model	GD	AIC	BIC
BGamma($\boldsymbol{\theta}_\delta$)	866.53	876.53	893.04
BGamma($\boldsymbol{\theta}_0$)	871.28	879.28	892.49
Weibull	872.34	880.34	893.55

6.2. *Application 2: Gastric cancer data*

Stomach cancer is also known as gastric cancer. Stomach cancer develops slowly over many years. Prior to the appearance of cancer itself, precancerous changes occur in the inner lining of the stomach (mucosa). These early changes rarely cause symptoms and therefore often go unnoticed. Thus, new technologies to optimize medical decisions and the development of new therapies are of great importance to improve survival in gastric cancer. In this second application, in order to illustrate the use of BGamma($\boldsymbol{\theta}_\delta$) regression, we consider the data set analysed by Martinez et al. [29] and Ortega et al. [30]. These last two surveys use the healing fraction regression model to analyse this gastric cancer data. The sample size is $n = 201$ patients of different clinical stages, of which 76 patients who received

Table 6. MLEs, SE and p-value for the parameters from the BGamma($\boldsymbol{\theta}_\delta$) regression model on the gastric cancer.

Parameter	Estimate	SE	p-Value
τ_{10}	−0.221	0.073	0.003
τ_{11}	1.306	0.100	< 0.001
τ_{20}	−3.506	0.077	< 0.001
τ_{21}	1.077	0.102	< 0.001
δ	0.032	0.002	

adjuvant chemoradiotherapy and 125 who received resection alone. The response variable refers to times to death in months since surgery. We observed that we have 53.2% of the censored data. Thus the variables used were:

- x_i: time to death in months since surgery;
- v_{i1}: type of therapy (0 = adjuvante chemoradiotherapy; 1 = surgery alone) for $i = 1, \ldots, 271$.

We now present results by fitting the BGamma($\boldsymbol{\theta}_\delta$) regression model

$$\alpha_i = \exp(\tau_{10} + v_{i1}\tau_{11}) \quad \text{and} \quad \beta_i = \exp(\tau_{20} + v_{i1}\tau_{21}), \quad i = 1, \ldots, 271.$$

The results in Table 5 indicate that the BGamma($\boldsymbol{\theta}_\delta$) regression model has the lowest GD (Global Deviance) and AIC values among those of the fitted models, and so it could be chosen as the best regression model. If we consider the BIC statistic, then the BGamma($\boldsymbol{\theta}_\delta$) and gamma regressions models are more appropriate to model this data set.

Table 6 lists the MLEs, their SEs and p-values for the BGamma($\boldsymbol{\theta}_\delta$) regression model fitted to these data. We note from the fitted BGamma($\boldsymbol{\theta}_\delta$) regression model that v_1 is significant (at 5% level). Further, there is a significant difference between type of therapy (adjuvante chemoradiotherapy and surgery alone) for the time to death in months since surgery.

In order to detect possible outlying observations as well as departures from the assumptions of BGamma($\boldsymbol{\theta}_\delta$) regression model, we present, in Figure 4, the plots of the density, QQ-plot and worm plot for the quantile residuals. By analysing these plots, we conclude that the BGamma($\boldsymbol{\theta}_\delta$) regression model provides a good adjustment.

Finally, in order to assess if the model is appropriate, the empirical and estimated survival functions of the BGamma($\boldsymbol{\theta}_\delta$) regression model are plotted in Figure 5 for the different treatments. Figure 5(a) shows the fit of the BGamma($\boldsymbol{\theta}_\delta$) regression model considering regression structure only in the α parameter. Figure 5(b) shows the fit considering two regression structures in the α and β parameters. We may conclude from the plots that the BGamma($\boldsymbol{\theta}_\delta$) regression model considering two regression structures provides a suitable fit to the gastric cancer data.

7. Concluding remarks

In this work, we have introduced a bimodal generalization of the gamma distribution that can be an alternative to model bimodal data. It was obtained using a quadratic transformation based on the alpha-skew-normal model. Since this generalization has three

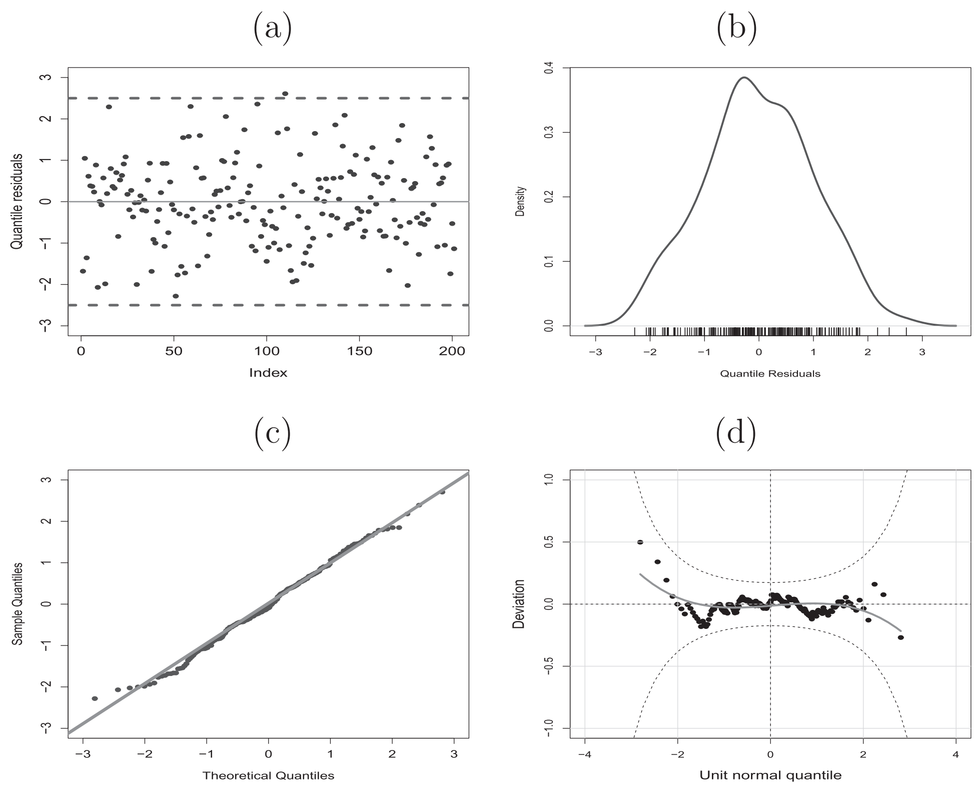

Figure 4. Plots of the diagnostic on fitting the BG regression model for gastric cancer. (a) Index plot for $\widehat{q}_i$. (b) Estimated density function for $\widehat{q}_i$. (c) QQ plot for $\widehat{q}_i$. (d) Worm plot for $\widehat{q}_i$.

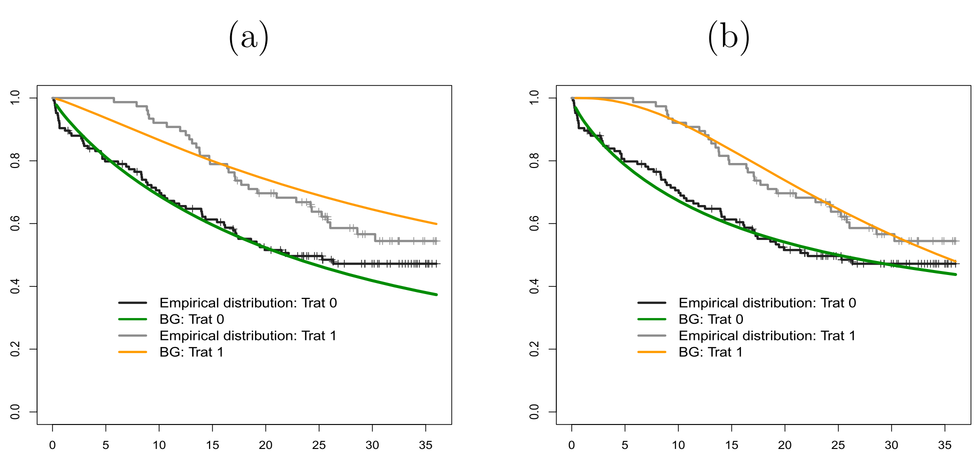

Figure 5. Estimated survival function for the BGamma$(\boldsymbol{\theta}_\delta)$ regression model and the empirical survival. (a) Adjustment considering only the $\alpha_i = \exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)$, parameter. (b) Adjustment considering both $\alpha_i = \exp(\mathbf{v}_i^T\boldsymbol{\tau}_1)$ and $\beta_i = \exp(\mathbf{v}_i^T\boldsymbol{\tau}_2)$, parameters for gastric cancer data.

parameters, the parameter estimation is simpler than in mixtures. We have discussed the properties of this density such as bimodality, moment generating function, hazard rate and entropy measures. In order to check the efficiency of the maximum likelihood estimators, we have carried out a Monte Carlo simulation study. We have also introduced a regression model based on the proposed bimodal gamma distribution. The fitting of the distribution along with its regression model was tested with two real data sets and it was shown that our model may outperform some distributions found in the literature. Thus, we have a flexible distribution that presented consistent results in data modelling.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

ORCID

Helton Saulo http://orcid.org/0000-0002-4467-8652
Edwin Ortega http://orcid.org/0000-0003-3999-7402

References

[1] Johnson NL, Kotz S, Balakrishnan N. Continuous univariate distributions. Vol. 1. New York (NY): Wiley; 1994. Volume 1.
[2] Ismadji S, Bhatia S. Characterization of activated carbons using liquid phase adsorption. Carbon. 2001;39:1237–1250.
[3] Balakrishnan N, Peng Y. Generalized gamma frailty model. Stat Med. 2006;25:2797–2816.
[4] Shankar FF, Lown L. Statistical modeling of atherosclerotic plaque in carotid b mode images-a feasibility study. Ultrasound Med Biol. 2003;29:1305–1309.
[5] Prataviera EMM, Ortega F, Cordeiro GM, et al. The heteroscedastic odd log-logistic generalized gamma regression model for censored data. Commun Stat Simul Comput. 2018;48:1815–1839.
[6] Derya K, Canan H. Control charts for skewed distributions: Weibull, gamma, and lognormal. Metodoloski Svezki. 2012;9:95–106.
[7] Hsu Y, Pearn W, Wu P. Capability adjustment for gamma processes with mean shift consideration in implementing six sigma program. Eur J Oper Res. 2008;119:517–529.
[8] Moors J, Strijbosch L. Exact fill rates for (R; s; S) inventory control with gamma distributed demand. J Oper Res Soc. 1988;53:1268–1274.
[9] Namit K, Chen J. Solutions to the inventory model for gamma lead-time demand. Int J Phys Distrib Logist Manage. 1999;29:138–154.
[10] Elal-Olivero D. Alpha-skew-normal distribution. Proyecciones J Math. 2010;29:224–240.
[11] Çankaya MN, Bulut YM, Doğru FZ, et al. A bimodal extension of the generalized gamma distribution. Rev Colombiana Estadíst. 2015;38:371–384.
[12] Griffiths L. Introduction to the theory of equations. New York: Wiley; 1947.
[13] Xue J. Loop tiling for parallelism. London: Springer; 2012. (The Springer international series in engineering and computer science).
[14] Vinberg ė. A course in algebra. London: American Mathematical Society; 2003. (Graduate studies in mathematics).
[15] Johnson NL, Kotz S, Kemp A. Univariate discrete distributions. New York (NY): Wiley; 1993.

[16] Klugman S, Panjer H, Willmot G. Loss models: from data to decisions. New York (NY): Wiley; 1998.
[17] Govil K, Aggarwal K. Mean residual life function for normal, gamma and lognormal densities. Reliab Eng. 1983;5:47–51.
[18] Dukkipati A, On generalized measures of information with maximum and minimum entropy prescriptions [PhD thesis]. Computer Science and Automation Indian Institute of Science Bangalore; 2006.
[19] Teh YW, Newman D, Welling M, A collapsed variational bayesian inference algorithm for latent Dirichlet allocation. Proceedings of the 19th International Conference on Neural Information Processing Systems, NIPS'06. Cambridge (MA): MIT Press; 2006. p. 1353–1360.
[20] Lagarias JC, Reeds JA, Wright MH, et al. Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J Optim. 1998;9:112–147.
[21] Cramér H. Mathematical methods of statistics. Princeton (NJ): Princeton University Press; 1946.
[22] Guo B-N, Qi F. Some properties of the psi and polygamma functions. Hacet Univ Bull Nat Sci Eng Ser B: Math Stat. 2010;39:219–231.
[23] Batir N. On some properties of digamma and polygamma functions. J Math Anal Appl. 2007;328:452–465.
[24] Elezovic CG, Pecaric J. The best bounds in Gautschi's inequality. Math Inequal Appl. 2000;3:239–252.
[25] Choulakian V, Stephens MA. Goodness-of-fit tests for the generalized Pareto distribution. Technometrics. 2001;43:478–484.
[26] Akinsete A, Famoye F, Lee C. The beta-Pareto distribution. Statistics. 2008;42:547–563.
[27] Pascoa MAR, Ortega EMM, Cordeiro GM. The Kumaraswamy generalized gamma distribution with application in survival analysis. Stat Methodol. 2011;8:411–433.
[28] Lai CD, Xie M, Murthy DNP. A modified Weibull distribution. IEEE Trans Reliab. 2003;52:33–37.
[29] Martinez EZ, Achcar JA, Jacome AA, et al. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. Comput Methods Programs Biomed. 2013;112:343–355.
[30] Ortega EMM, Cordeiro GM, Hashimoto EM, et al. Regression models generated by gamma random variables with long-term survivors. Commun Stat Appl Methods. 2017;24:43–65.