Spanish Journal of Statistics
Vol. 2 No. 1 2020, Pages 23–40
doi:https://doi.org/10.37830/SJS.2020.1.03

Regular Article

The Gamma-Chen distribution: a new family of distributions with applications

Lucas David R. Reis[1], Gauss M. Cordeiro[2], Maria do Carmo S. Lima[3]
[1]Federal University of Pernambuco, econ.lucasdavid@gmail.com
[2]Federal University of Pernambuco, gausscordeiro@gmail.com
[3]Federal University of Pernambuco, maria@de.ufpe.br

Received: July 21, 2020. Accepted: February 15, 2021.

Abstract: The generalized gamma-generated family adds one shape parameter to a baseline distribution. We define the gamma-Chen and derive some of its mathematical properties. Its hazard rate may have increasing, decreasing, bathtub and unimodal shapes due to the extra parameter, which portrays a positive point of the proposed model. We perform Monte Carlo simulations to prove that the asymptotic properties of the maximum likelihood estimators hold. We show empirically that the new distribution is better than ten others known distributions using engineering-related data sets.

Keywords: Bathtub, Chen distribution, gamma-Chen distribution, Maximum likelihood, Simulation study

MSC: 33B15, 33C15, 62E20, 62P30, 65C05

1 Introduction

In the area of survival analysis and new distributions, much is said about proposing families and, consequently, distributions, which model fatigue data sets, failure time of electronic components, etc., which constitute engineering data.

Several types of data sets have been used for new distributions from medical and different branches of engineering and industry. However, Brazilian data sets are seldom used in international statistical papers. In this context, this work focuses on two applications in engineering area from Brazil. In addition, given that many workers used data collected over 20 years, we adopt more recent data sets.

The Weibull and Birnbaum-Saunders models are among the most widely distributions taken for baseline for several generators in different areas of engineering. The main goal here is to propose a new distribution that is as flexible as, or more than, the aforementioned, and that fits recent real engineering data. We believe that this purpose is valid and innovative.

One of the most used methods in the construction of new lifetime distributions is based on well-established generators by adding shape(s) parameters to parent models. This method is adopted in this paper. The proposed distribution is interesting for lifetime data analysis as a further option, where some known distributions do not fit well.

The probability density function (pdf) and cumulative distribution function (cdf) of the gamma-G family (Zografos and Balakrishnan, 2009) for a baseline G are

$$f_{\mathrm{GG}}(x;a,\boldsymbol{\eta}) = \frac{1}{\Gamma(a)}\{-\log[1-G(x;\boldsymbol{\eta})]\}^{a-1}\, g(x;\boldsymbol{\eta}) \tag{1}$$

and

$$F_{\mathrm{GG}}(x;a,\boldsymbol{\eta}) = \frac{\gamma\left(a,-\log[1-G(x;\boldsymbol{\eta})]\right)}{\Gamma(a)} = \frac{1}{\Gamma(a)}\int_0^{-\log[1-G(x;\boldsymbol{\eta})]} t^{a-1}\mathrm{e}^{-t}\,dt,$$

respectively, where $a > 0$, $\boldsymbol{\eta}$ is the q-parameter vector of the baseline distribution, $g(x;\boldsymbol{\eta}) = dG(x;\boldsymbol{\eta})/dx$, $\Gamma(a) = \int_0^\infty t^{a-1}\mathrm{e}^{-t}dt$ is the gamma function and $\gamma(a,z) = \int_0^z t^{a-1}\mathrm{e}^{-t}dt$ denotes the lower incomplete gamma function. This family is flexibilized by the shape parameter a and the support of $f_{\mathrm{GG}}(x)$ is the same of $g(x)$.

Many papers adopt the gamma-G family in order to fit several types of data sets. The great advantage in choosing this family over that one proposed by Torabi and Hedesh (2012) is the reduction in the problems of parameter estimation, since the proposal by Zografos and Balakrishnan (2009) has one parameter less than the other one. There are many works involving this family and its structural properties (Nadarajah et al., 2015). Table 1 lists some of its special models and associated data sets. In all cases, the gamma-G provides better fits when compared with another well-know distributions including beta-G models.

Table 1: Some gamma-G models

Model	Authors (year)	Application
Gamma-Birnbaum-Saunders	Cordeiro et al. (2016)	Failure and fatigue
Gamma-Normal	Lima (2015)	Agronomy and levels nicotine
Gamma-Lindley	Lima (2015)	Reliability and SAR images
Gamma-Nadarajah-Haghighi	Lima (2015)	Failure and fatigue
Gamma-Extended Weibull	Lima (2015)	Failure and fatigue
Gamma-Pareto	Alzaatreh et al. (2012)	River flood rates, fatigue and frequencies for Tribolium Confusum Strain
Gamma-Exponentiated Weibull	Pinho et al. (2012)	Daily minimum wind speed
Gamma exponentiated exponential-Weibull distribution	Pogány and Saboor (2016)	Remission times and fatigue

Major topics studied in the sections are as follows. In Section 2, we define the gamma-Chen (GC) model. In Section 3, we obtain some of its properties. In Section 4, we examine the accuracy of the maximum likelihood estimators (MLEs). The superiority of the GC model in relation to ten known distributions (including the well-known exponentiated Weibull model) is proved by means of two engineering data sets in Section 5. These competitors were chosen based on previous works in engineering data management (focus of the applications of this work). In Section 6, we conclude the paper.

2 Proposed model

In survival analysis it is very common to look for new distributions that have great versatility in the hazard rate function (hrf). The most common forms of hrfs are bathtub and unimodal. Chen (2000) proposed a two-parametric distribution that accommodates increasing and bathtub hrf forms, thus showing the great flexibility of this distribution.

Recently, some extensions of the Chen distribution (Chen, 2000) have appeared in the literature. Dey et al. (2017) proposed the exponentiated-Chen (exp-Chen) and showed that this distribution also has unimodal hrf. Among others extensions, we can mention Kumaraswamy-exponentiated-Chen (Khan et al., 2018) distribution, Weibull-Chen (Tarvirdizade and Ahmadpour, 2019) distribution, modified Weibull extension (Xie et al., 2002) and odd Chen-G family (Anzagra et al., 2020).

The cdf and pdf of the random variable $Y \sim \mathrm{Chen}(\lambda, \beta)$ are

$$G(y;\lambda,\beta) = 1 - \mathrm{e}^{\lambda(1-\mathrm{e}^{y^{\beta}})}, \quad y > 0$$

and

$$g(y;\lambda,\beta) = \lambda\beta y^{\beta-1}\mathrm{e}^{y^{\beta}+\lambda(1-\mathrm{e}^{y^{\beta}})}, \tag{2}$$

respectively, where $\lambda > 0$ is a scale parameter and $\beta > 0$ is a shape parameter.

The GC density is determined from (1) and the last two equations

$$f_{\mathrm{GC}}(x;a,\lambda,\beta) = \frac{\lambda^{a}\beta}{\Gamma(a)}\, x^{\beta-1}\,(\mathrm{e}^{x^{\beta}}-1)^{a-1}\,\mathrm{e}^{x^{\beta}+\lambda(1-\mathrm{e}^{x^{\beta}})}, \quad x > 0. \tag{3}$$

For $a = 1$, we have the Chen density. Henceforth, $X \sim \mathrm{GC}(a, \lambda, \beta)$ denotes a random variable with pdf (3). The three-parameter GC distribution has no problem of identifiability. The Chen distribution is clearly identifiable since different parameter vectors imply different cumulative distributions. So, the GC is also identifiable.

A simple motivation for the GC density follows from Zografos and Balakrishnan (2009). If $Y_{(1)} < \cdots < Y_{(p)} < \cdots$ are upper record values arising from a sequence of Chen independent and identically random variables $Y_1, Y_2, \cdots$, then the order statistic $Y_{(p)}$ has the GC density with $a = p$. So, the density of X can approximate the density of the pth order statistic of the Chen(λ, β) distribution by taking p as the greatest integer less than or equal to a. So, the GC distribution is generated by Chen record value densities. This explicitly means that the GC distribution is a direct record-Chen analog.

A practical relevance and applicability of the GC distribution is for the lifetime system with n independent components which function if and only if at least k of the components function is a "k out of n" system. For such a system, k is less than n, and it includes some parallel, fail-safe and series systems all as special cases for $k = 1, k = n-1$ and $k = n$, respectively. Suppose $Y_1, \cdots, Y_n$ denote the lifetimes of n components having the Chen distribution of a system, where k is assumed unknown and n is very large. Then, the lifetime of a k-out-of-n system consisting of these components can be represented by the order statistic $Y_{(n-k+1)}$, which can be modeled by the GC distribution to estimate a and then k.

Figure 1(a) displays plots of the density of X for some parameter values, which show that it accommodates several forms. By combining different values of β and a provide great flexibility for the GC density. In fact, this density can be symmetric, left-skewed or right-skewed, and the parameter a has significant effects on both skewness and kurtosis.

The cdf and hrf of X are

$$F_{\mathrm{GC}}(x;a,\lambda,\beta) = \frac{\gamma(a, -\lambda(1-\mathrm{e}^{x^{\beta}}))}{\Gamma(a)},$$

and

$$\tau_{\rm GC}(x;a,\lambda,\beta) = \frac{\lambda^a \beta\, x^{\beta-1}\,(\mathrm{e}^{x^\beta}-1)^{a-1}\,\mathrm{e}^{x^\beta+\lambda(1-\mathrm{e}^{x^\beta})}}{\Gamma(a)-\gamma(a,-\lambda(1-\mathrm{e}^{x^\beta}))},$$

respectively. Note that for $a=\beta=1$ the shape of the hrf is independent of λ. Chen (2000) showed that the Chen hrf can only be increasing ($\beta\geq 1$) and bathtub ($\beta<1$). However, the hrf of X can be increasing, decreasing, unimodal and bathtub-shaped as shown in Figure 1(b). Further, the bathtub shape can be obtained even when $\beta>1$. This fact reveals that the hrf of X gains more flexibility with the extra parameter a since it can take the most four common forms for applications to real data: increasing for any positive value of β, bathtub-shaped, unimodal, and also decreasing, which shows that it has great flexibility due to the parameter a (see Figure 1(b)).

Following the idea in Qian (2012), we can to determine the parameter ranges for the density shapes. Setting $z=\exp(x^\beta)$, we obtain from (3)

$$r(z) = f([\log z]^{1/\beta}) = \frac{\lambda^a \beta \log z}{\Gamma(a)}(\log z)^{-1/\beta}(z-1)^{a-1}\exp[\log z+\lambda(1-z)].$$

Applying logarithms of both sides of the previous equation,

$$\begin{aligned}\log r(z) =& a\log\lambda+\log\beta+\log(\log z)-\log\Gamma(a)-\frac{1}{\beta}\log(\log z)\\ &+(a-1)\log(z-1)+\log z+\lambda(1-z).\end{aligned}$$

By taking derivatives of both sides of the last equation, we have

$$\begin{aligned}\frac{r'(z)}{r(z)} =& \frac{1}{z\log z}-\frac{1}{\beta z\log z}+\frac{a-1}{z-1}+\frac{1}{z}-\lambda\\ =& \frac{\beta(z-1)-(z-1)+\beta z(a-1)\log z+\beta(z-1)\log z-\lambda\beta z(z-1)\log z}{\beta z(z-1)\log z}.\end{aligned}$$

If $s(z)$ is the numerator of the right side of this equation, we can write

$$r'(z) = \frac{r(z)s(z)}{\beta z(z-1)\log z}.$$

Hence, $r'(z)$ and $s(z)$ have the same signs, since $r(z)>0$ and $\beta z(z-1)\log z>0$ for $z>1$. The condition $z>1$ holds since $x>0$. In this case, $x=\log(z)^{1/\beta} \iff z>1$.

Note that in Region I (Figure 2(a)), $s(z)$ takes positive values first and then negative values, which indicates the unimodal property of the density. In Region II (Figure 2(b)), $s(z)$ has only negative values, thus indicating decreasing shape. So, Figures 2(a) and 2(b) reveal that the pdf is unimodal for $a\geq 1$ and that it is decreasing for $a\in(0,1)$, respectively, as noted in Figure 1(a).

3 Properties

It is not possible to obtain some mathematical properties of the GC distribution in closed form, that is, according to known mathematical functions. Then, we determine these quantities from the weighted linear combination for its density function given in Theorem 2 below.

For a given cdf $G(z;\boldsymbol{\eta})$ with q-parameter vector $\boldsymbol{\eta}$, the cdf and pdf of the exponentiated-G (exp-G) random variable Z_a with power parameter $a>0$, say $Z_a\sim$ exp-G$(a,\boldsymbol{\eta})$, are

$$H(z;a,\boldsymbol{\eta}) = G(z;\boldsymbol{\eta})^a \quad\text{and}\quad h(z;a,\boldsymbol{\eta}) = a\,g(z;\boldsymbol{\eta})G(z;\boldsymbol{\eta})^{a-1},$$

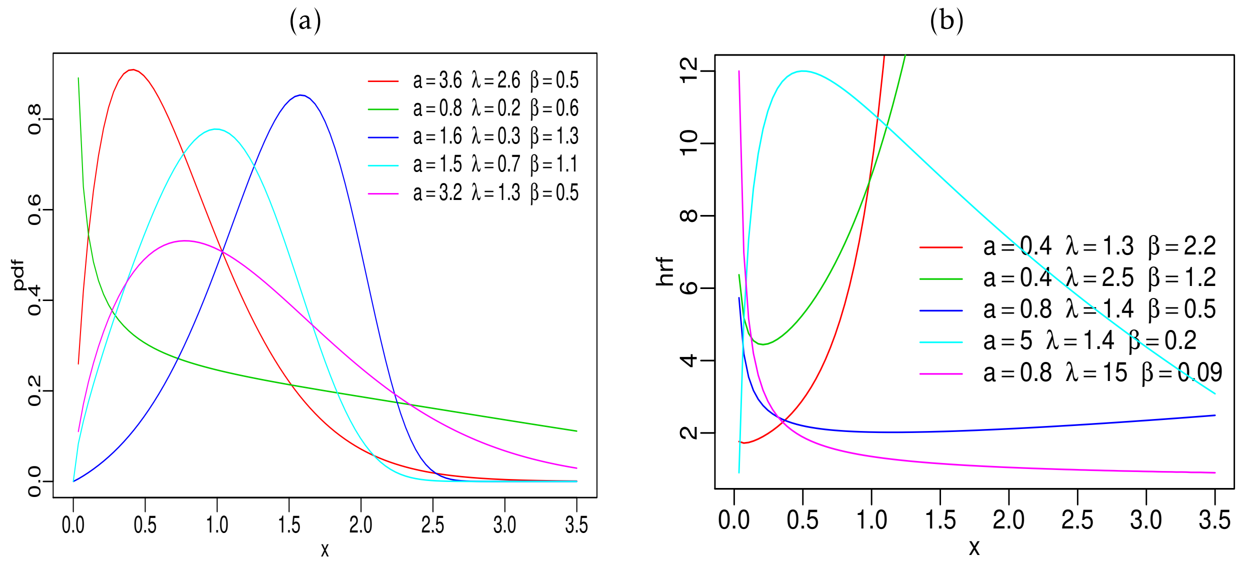

Figure 1: Plots of the pdf (a) and hrf (b) of the GC distribution.

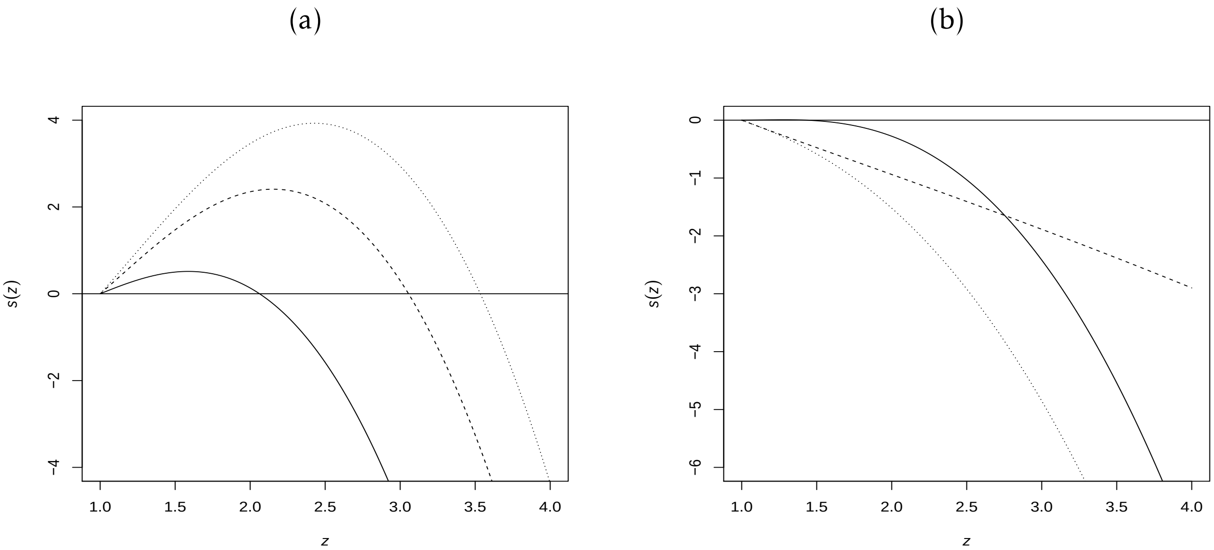

Figure 2: Regions for the density shapes. (a) Region I: $a \geq 1$ and (b) Region II: $a \in (0,1)$.

respectively, where $g(z;\boldsymbol{\eta}) = dG(z;\boldsymbol{\eta})/dz$.

The gamma-G cdf can be expressed as (Castellares and Lemonte, 2015)

$$F_{\mathrm{GG}}(x;a,\boldsymbol{\eta}) = \sum_{k=0}^{\infty} \frac{\varphi_k(a)}{(a+k)} H(x;(a+k),\boldsymbol{\eta}), \tag{4}$$

where $\varphi_0(a) = \frac{1}{\Gamma(a)}$, $\varphi_k(a) = \frac{(a-1)}{\Gamma(a)}\,\psi_{k-1}(k+a-2)$ $(k \geq 1)$, $\psi_{k-1}(\cdot)$ are the Stirling polynomials

$$\psi_{k-1}(w) = \frac{(-1)^{k-1}}{(k+1)!}\Bigg[T_k^{k-1} - \frac{(w+2)}{(k+2)}T_k^{k-2} + \frac{(w+2)(w+3)}{(k+2)(k+3)}T_k^{k-3} - \cdots + (-1)^{k-1}\frac{(w+2)(w+3)\cdots(w+k)}{(k+2)(k+3)\cdots(2k)}T_k^0\Bigg],$$

$T_0^0 = 1$, $T_{k+1}^0 = 1 \times 3 \times \ldots \times (2k+1)$, $T_{k+1}^k = 1$ and T_k^m are positive integers determined from

$$T_{k+1}^m = (2k+1-m)T_k^m + (k-m+1)T_k^{m-1}.$$

The function $H(x;(a+k),\boldsymbol{\eta})$ denotes the cdf of Z_{a+k}. Thus, we can obtain the properties of the gamma-G model from those of the exp-G class.

Theorem 1. *Let Y be a random variable having density (2). Then, the exp-Chen(a,λ,β) density can be expressed as*

$$h(y;a,\lambda,\beta) = \sum_{m=1}^{\infty} p_m\, g(y;m\lambda,\beta),$$

where $p_m = p_m(a) = (-1)^{m+1}\binom{a}{m}$ and $g(y;m\lambda,\beta)$ is the Chen density with scale $m\lambda$ and shape β.

Proof. For $|x| < 1$ and any real $a \neq 0$, the power series

$$(1-x)^a = \sum_{m=0}^{\infty}(-1)^m\binom{a}{m}x^m$$

converges. Thus, the exp-Chen cdf can be expanded as

$$H(y;a,\lambda,\beta) = \left[1 - \mathrm{e}^{\lambda(1-\mathrm{e}^{y^\beta})}\right]^a = 1 + \sum_{m=1}^{\infty}(-1)^m\binom{a}{m}[1 - G(y;m\lambda,\beta)].$$

By differentiating the last equation,

$$h(y;a,\lambda,\beta) = \sum_{m=1}^{\infty}(-1)^{m+1}\binom{a}{m}g(y;m\lambda,\beta),$$

and then the exp-Chen density is a linear combination of Chen densities. □

Theorem 2. *The pdf of X in Equation (3) can be expressed as*

$$f_{\mathrm{GC}}(x;a,\lambda,\beta) = \sum_{m=1}^{\infty} w_m\, g(x;m\lambda,\beta),$$

where $g(x;m\lambda,\beta)$ is the Chen density with scale $m\lambda$ and shape β and the weights are

$$w_m = w_m(a) = (-1)^{m+1}\sum_{k=0}^{\infty}\frac{\varphi_k(a)}{(a+k)}\binom{a+k}{m}.$$

Proof. The proof comes directly from Equation (4) and Theorem 1. □

By using Theorem 2, the rth moment of X has the form

$$\mathbb{E}[X^r] = \sum_{m=1}^{\infty} w_m \, \mathbb{E}[Y_m^r],$$

where $Y_m \sim \text{Chen}(m\lambda, \beta)$.

If $Y \sim \text{Chen}(\lambda, \beta)$ has pdf (2), we can write (Pogány et al., 2017)

$$\mathbb{E}[Y^r] = \lambda \, \mathrm{e}^{\lambda} \, \mathbb{D}_t^{r\beta^{-1}} \left[\frac{\Gamma(t+1, \lambda)}{\lambda^{t+1}} \right]_{t=0}. \tag{5}$$

Here,

$$\mathbb{D}_t^p \left[\frac{\Gamma(t+1, \lambda)}{\lambda^{t+1}} \right]_{t=0} = \Gamma(p+1) \sum_{k \geq 0} \frac{(2)_k}{k!} \, \Phi_{\mu,1}^{(0,1)}(-k, p+1, 1) \; {}_1F_1(k+2; 2; -\lambda),$$

where $\Phi_{\mu,1}^{(0,1)}(-a, p+1, 1) = \sum_{n \geq 0} \frac{(-a)^n}{n!(n+1)^{p+1}}$ for $\mu \in \mathbb{C}$, ${}_1F_1(a; b; x) = \sum_{n \geq 0} \frac{(a)_n}{(b)_n} \frac{x^n}{n!}$, for $x, a \in \mathbb{C}$ and $b \in \mathbb{C} \setminus Z_0^-$, is the confluent hypergeometric function (Kilbas et al., 2006, page 29, Eq. 1.6.14) and $(\lambda)_\eta = \frac{\Gamma(\lambda+\eta)}{\Gamma(\lambda)}$, for $\lambda \in \mathbb{C} \setminus \{0\}$, is the generalized Pochhammer symbol with $(0)_0 = 1$.

Thus, using (5), the rth moment of X can be reduced to

$$\mathbb{E}[X^r] = \lambda \sum_{m=1}^{\infty} m \, w_m \, \mathrm{e}^{m\lambda} \, \mathbb{D}_t^{r\beta^{-1}} \left[\frac{\Gamma(t+1, m\lambda)}{(m\lambda)^{t+1}} \right]_{t=0}.$$

Figures 3, 4 and 5 provide the plots of the mean and variance of X as functions of a, λ and β, respectively, the other parameters being fixed. The mean and variance of X increase when a increases. In turn, these measures decrease when λ increases. Further, the mean of X increases when β increases and the variance of X increases to a maximum point and starts to decrease.

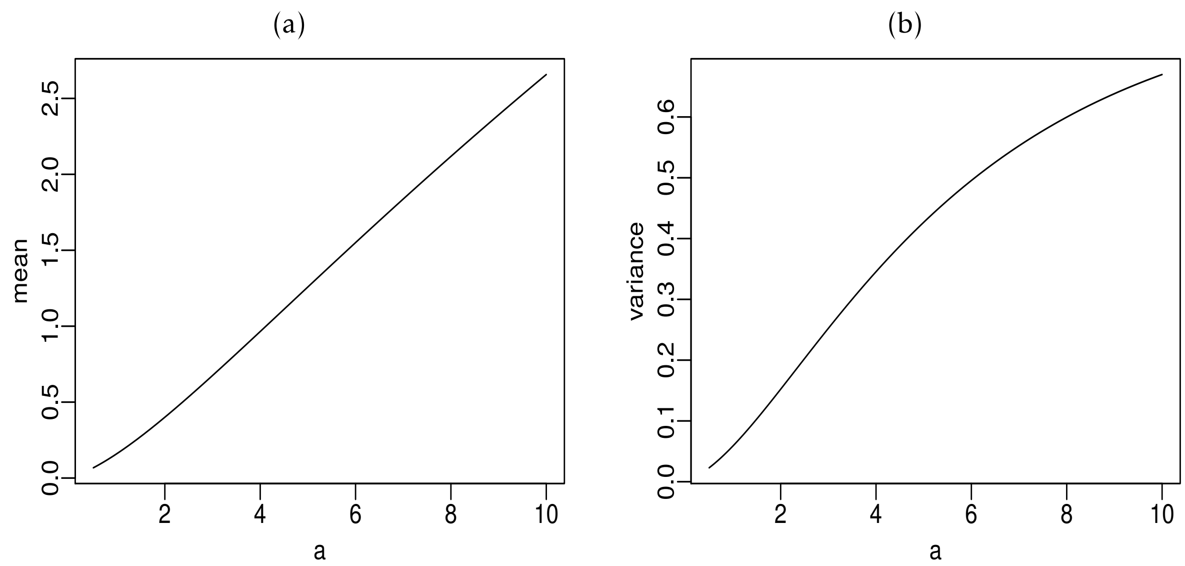

Figure 3: Mean (a) and variance (b) plots of X as functions of a ($\lambda = 2.4, \beta = 0.5$).

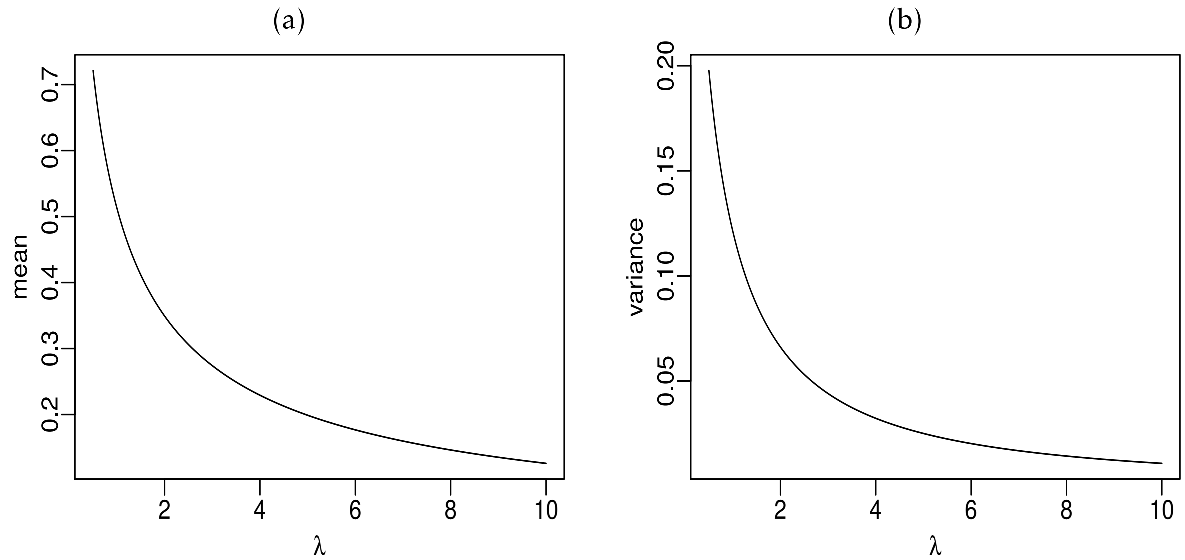

Figure 4: Mean (a) and variance (b) plots of X as functions of λ ($a = 0.7, \beta = 1.4$).

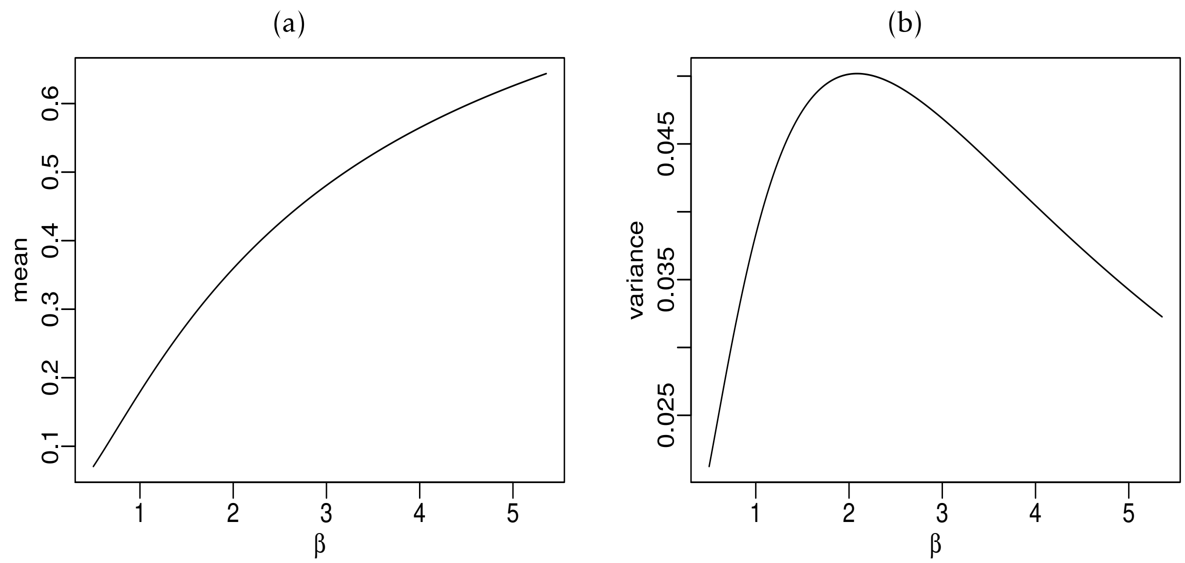

Figure 5: Mean (a) and variance (b) plots of X as functions of β ($a = 0.6, \lambda = 2.7$).

Another type of measure that has great applicability is the incomplete moment. For $z > 0$, the rth incomplete moment of the random variable Y with Chen distribution, say $q_r(z;\lambda,\beta) = \int_0^z y^r g(y;\lambda,\beta)dy$, follows from Pogány et al. (2017) as

$$q_r(z;\lambda,\beta) = \lambda e^{\lambda} \sum_{n,k\geq 0} \sum_{j=1}^{k} \frac{(2)_{n+k}}{(2)_n} \frac{(-1)^{n+j}\,\lambda^n}{n!k!(j+1)^{r\beta^{-1}+1}} \binom{k}{j} \gamma\left(r\beta^{-1}, (j+1)(1-z^{-1})\right). \tag{6}$$

Thus, using Theorem 2 and Equation (6), the rth incomplete moment of X is

$$m_r(z) = \lambda \sum_{m=1}^{\infty} m\,\mathrm{e}^{m\lambda}\, w_m \sum_{n,k\geq 0} \sum_{j=1}^{k} \frac{(2)_{n+k}}{(2)_n} \frac{(-1)^{n+j}\,(m\lambda)^n}{n!k!(j+1)^{r\beta^{-1}+1}} \binom{k}{j} \gamma\left(r\beta^{-1}, (1-z^{-1})(j+1)\right).$$

The first incomplete moment is used to obtain Lorenz and Bonferroni curves and mean deviations.

The generating function (gf) of $Y \sim \text{Chen}(\lambda,\beta)$, $M_Y(-t) = \mathbb{E}[\mathrm{e}^{-tY}]$, $t > 0$, can be written, according to Pogány et al. (2017), by

$$M_Y(-t) = \lambda\beta\,\mathrm{e}^{\lambda}\, t^{-\beta} \sum_{n\geq 0} \frac{(-\lambda)^n}{n!}\, {}_1\Psi_0\left[(\beta,\beta);-;\frac{n+1}{t^{\beta}}\right], \tag{7}$$

where

$${}_1\Psi_0\left[(a,b);-;z\right] = \sum_{n\geq 0} \frac{\Gamma(a+bn)\,z^n}{n!}, \quad z, a \in \mathbb{C}, b > 0,$$

is the generalized Fox-Wright function.

Thus, from Theorem 2 and Equation (7), the gf of X follows as (for $t > 0$)

$$M_X(-t) = \lambda\beta\, t^{-\beta} \sum_{m=1}^{\infty} \sum_{n\geq 0} \frac{m\,\mathrm{e}^{m\lambda}\,(-m\lambda)^n\, w_m}{n!}\, {}_1\Psi_0\left[(\beta,\beta);-;\frac{n+1}{t^{\beta}}\right].$$

The quantile function (qf) of X, say $Q_{\mathrm{GG}}(u;a,\boldsymbol{\eta}) = F_{\mathrm{GG}}^{-1}(u;a,\boldsymbol{\eta})$, can be expressed as (Nadarajah et al., 2015)

$$Q_{\mathrm{GG}}(u;a,\boldsymbol{\eta}) = Q_{\mathrm{G}}\left(1-\mathrm{e}^{-Q_1(a,u)};\boldsymbol{\eta}\right), \quad 0 < u < 1,$$

where Q_{G} is the qf of the baseline $G(x;\boldsymbol{\eta})$ and $Q_1(a,u)$ is the inverse function of $\gamma_1(a,w) = \gamma(a,w)/\Gamma(a)$. Further, we can write

$$Q_{\mathrm{GC}}(u;a,\lambda,\beta) = \left\{\log\left[1+\lambda^{-1}Q_1(a,u)\right]\right\}^{1/\beta}. \tag{8}$$

We can obtain skewness and kurtosis measures of X from Equation (8). The Bowley skewness and Moors kurtosis are based on quartiles and octiles, respectively. Letting $Q_{\mathrm{GC}}(u) = Q_{\mathrm{GC}}(u;a,\lambda,\beta)$, the skewness and kurtosis of X are

$$\mathcal{B}(a,\lambda,\beta) = \frac{Q_{\mathrm{GC}}(3/4) + Q_{\mathrm{GC}}(1/4) - 2Q_{\mathrm{GC}}(2/4)}{Q_{\mathrm{GC}}(3/4) - Q_{\mathrm{GC}}(1/4)}$$

and

$$\mathcal{M}(a,\lambda,\beta) = \frac{Q_{\mathrm{GC}}(7/8) - Q_{\mathrm{GC}}(5/8) - Q_{\mathrm{GC}}(3/8) + Q_{\mathrm{GC}}(1/8)}{Q_{\mathrm{GC}}(6/8) - Q_{\mathrm{GC}}(2/8)},$$

respectively. Plots of these measures as functions of a are displayed in Figure 6, which show that both of them decrease when a increases. Both measures grow when a decreases from one, and they can take negative values and higher positive values when a increases from one.

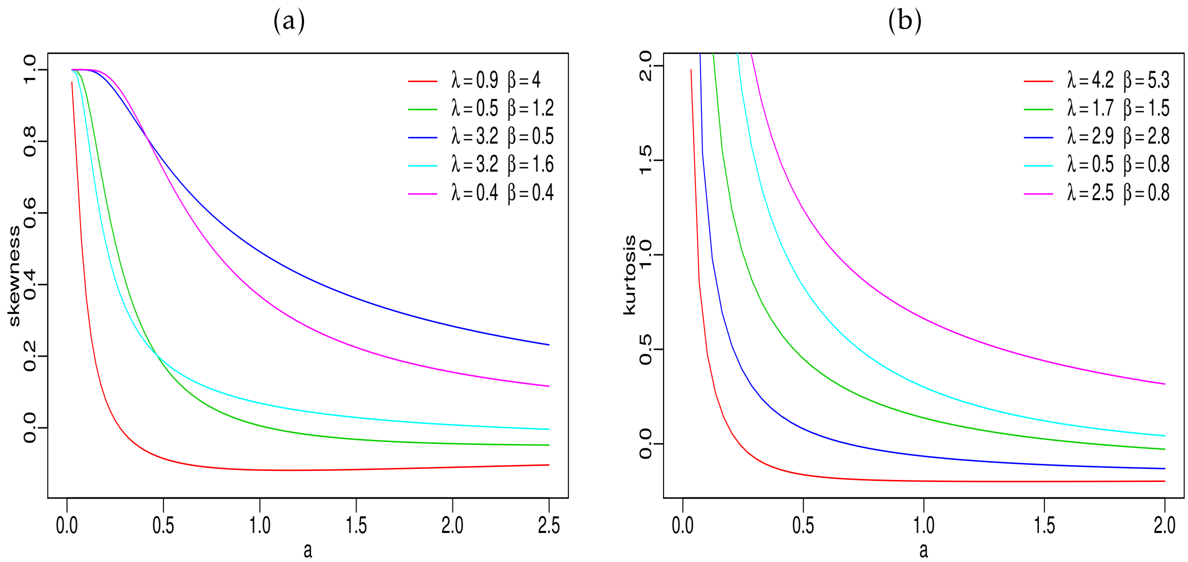

Figure 6: Skewness (a) and kurtosis (b) plots of X as functions of a.

4 Estimation

Let $\boldsymbol{\theta} = (a, \lambda, \beta)^\top$ be the parameter vector of the GC model. Consider the random variables $X_1, \ldots, X_n \sim \mathrm{GC}(a, \lambda, \beta)$ with observed values $x_1, \ldots, x_n$. The log-likelihood function for $\boldsymbol{\theta}$ is

$$\begin{aligned}\ell(\boldsymbol{\theta}) =& n[a \log \lambda + \log \beta - \log \Gamma(a) + \lambda] + \sum_{i=1}^{n} x_i^\beta + (\beta - 1) \sum_{i=1}^{n} \log x_i \\ &+ (a-1) \sum_{i=1}^{n} \log(\mathrm{e}^{x_i^\beta} - 1) - \lambda \sum_{i=1}^{n} \mathrm{e}^{x_i^\beta}.\end{aligned}$$

The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}$, can be found by maximizing $\ell(\boldsymbol{\theta})$ numerically with respect to its components. Some routines such as SAS (`PROC NLMIXED`), R (`optim` function) and Ox (sub-routine `MaxBFGS`) can be used for the maximization.

We now study the behavior of MLEs in the GC model from 1,000 Monte Carlo replications. All simulations are performed using R Project (R Core Team, 2019). The sample sizes chosen are $n = 25$, 50, 100, 200, 300 and 400 and the true parameter vectors are: $(a, \lambda, \beta) = (1.4, 0.7, 1.9)$ for scenario 1, and $(a, \lambda, \beta) = (2.5, 1.5, 0.8)$ for scenario 2. There were no special reasons for choosing these parameters.

Table 2 reports the average estimates (AEs), biases and mean squared errors (MSEs) for both scenarios. The MLEs converge to the true parameters and the biases and MSEs decrease to zero when the sample size n increases, that makes us conclude that the consistency criterion holds.

5 Engineering data

In order to show a superior performance of the new distribution when compared to others already published in the literature, we provide two applications in recent real data sets in the engineering

Table 2: Findings under scenarios 1 to 2.

Par	$n=25$			$n=50$			$n=100$		
	scenario 1								
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
a	1.965	0.565	5.503	1.656	0.256	1.634	1.496	0.096	0.485
λ	1.065	0.365	2.138	0.868	0.168	0.641	0.762	0.062	0.206
β	2.216	0.316	1.074	2.049	0.149	0.415	1.985	0.085	0.167
Par	$n=200$			$n=300$			$n=400$		
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
a	1.442	0.042	0.155	1.436	0.036	0.108	1.419	0.019	0.081
λ	0.727	0.027	0.068	0.727	0.027	0.045	0.715	0.015	0.029
β	1.942	0.042	0.077	1.917	0.017	0.048	1.916	0.016	0.040
	scenario 2								
Par	$n=25$			$n=50$			$n=100$		
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
a	1.965	0.465	7.624	1.808	0.308	4.813	1.621	0.121	1.326
λ	1.869	0.369	4.307	1.787	0.287	2.159	1.608	0.108	0.806
β	1.393	0.593	2.348	1.043	0.243	0.778	0.891	0.091	0.104
Par	$n=200$			$n=300$			$n=400$		
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
a	1.539	0.039	0.400	1.547	0.047	0.225	1.477	-0.023	0.345
λ	1.542	0.042	0.246	1.546	0.046	0.155	1.506	0.006	0.134
β	0.851	0.051	0.077	0.819	0.019	0.021	0.839	0.039	0.053

area. In this work, we choose two types of engineering data sets to show the flexibility of the proposed model. One related to ore wagon fleets and the other to natural gas; both sets belong to the production engineering branch. Table 3 presents the descriptive statistics of the two data sets.

The first data set is obtained in a work by (Sivini, 2006), which involves the execution of a pilot project of a reliability data applied in natural gas pressure reducing stations (ERPGN) of a company that operates in Pernambuco (Brazil). The data in question refer to the time until the maintenance time (Tm) in one of the Pressure Reduction and Measurement Stations (ERPM - A) between 10/14/2002 to 05/16/2005.

The second application has a data set taken from the same work (Sivini, 2006). Here, we consider Tm in ERPM B and C, collected between 11/14/2002 and 6/16/2005.

Table 3 gives the descriptive statistics of the two data sets. Note that the two data sets differ widely. The first with a mean of 2.9222 and the second with a mean of 6.2062. Their maximum values and standard deviations (SDs) are also very different.

5.1 Competitive distributions

We compare the GC model with other ten distributions: Chen, exponentiated Weibull (Mudholkar and Hutson, 1993), Kumaraswamy-log-logistic (de Santana et al., 2012), gamma-extended Frèchet (da Silva et al., 2013), beta-log-logistic (Lemonte, 2014), Birnbaum-Saunders (Birnbaum and Saunders, 1969), gamma-Birnbaum-Saunders (Cordeiro et al., 2016), beta Birnbaum-Saunders (Cordeiro and Lemonte, 2011), odd-log-logistic Birnbaum-Saunders (Ortega et al., 2016) and odd-log-logistic Birnbaum-Saunders Poisson (Cordeiro et al., 2018).

Table 3: Descriptive statistics.

Description	data set 1	data set 2
Min.	1.1700	1.0000
1st Qu.	1.1900	3.0625
Median	1.5850	4.8350
Mean	2.9222	6.2062
3rd Qu.	4.9175	6.1250
Max.	8.0000	30.3300
SD	2.2553	6.7021

The choice of the previous distributions is based on suitable ones with good fits to engineering data. We emphasize that other distributions could also be used.

The densities of the exponentiated Weibull (EW), Kumaraswamy-log-logistic (KLL), gamma-extended Frèchet (GEF) and beta-log-logistic (BLL) are (for $x > 0$)

$$f_{\text{EW}}(x;a,\lambda,\beta) = \frac{a\lambda}{\beta}\left(\frac{x}{\beta}\right)^{\lambda-1}\left\{1-\exp\left[-\left(\frac{x}{\beta}\right)^{\lambda}\right]\right\}^{(a-1)}\exp\left[-\left(\frac{x}{\beta}\right)^{\lambda}\right],$$

$$f_{\text{KLL}}(x;a,b,\alpha,\delta) = \frac{ab\delta}{\alpha^{a\delta}}\,x^{a\delta-1}\left[1-\left(\frac{x}{\alpha}\right)^{\delta}\right]^{-a-1}\left\{1-\left[1-\frac{1}{1+\left(\frac{x}{\alpha}\right)^{\delta}}\right]^{a}\right\}^{b-1},$$

$$f_{\text{GEF}}(x;a,\lambda,\sigma,\alpha) = \frac{\alpha\lambda\sigma^{\lambda}}{\Gamma(a)}\,x^{-\lambda-1}\exp\left[-\left(\frac{\sigma}{x}\right)^{\lambda}\right]\left\{1-\exp\left[-\left(\frac{\sigma}{x}\right)^{\lambda}\right]\right\}^{\alpha-1}$$
$$\times\left\{-\log\left\{1-\exp\left[-\left(\frac{\sigma}{x}\right)^{\lambda}\right]\right\}^{\alpha}\right\}^{a-1},$$

and

$$f_{\text{BLL}}(x;a,b,\alpha,\beta) = \frac{\beta\,\Gamma(a)\Gamma(b)}{\alpha\,\Gamma(a+b)}\,\frac{(x/\alpha)^{a\beta-1}}{[1+(x/\alpha)^{\beta}]^{a+b}},$$

respectively, where all parameters are positive.

The cdf and pdf of the Birnbaum-Saunders (BS) are

$$F_{\text{BS}}(x;\alpha,\beta) = \Phi\left(\frac{1}{\alpha}\left[\left(\frac{x}{\beta}\right)^{\frac{1}{2}}-\left(\frac{\beta}{x}\right)^{\frac{1}{2}}\right]\right), \quad x>0 \tag{9}$$

and

$$f_{\text{BS}}(x;\alpha,\beta) = \frac{\exp(\alpha^{-2})}{2\alpha\sqrt{2\pi\beta}}\,x^{-\frac{3}{2}}(x+\beta)\exp\left[-\frac{1}{2\alpha^2}\left(\frac{x}{\beta}+\frac{\beta}{x}\right)\right], \tag{10}$$

respectively, where $\alpha, \beta > 0$ and $\Phi(\cdot)$ is the standard normal cdf.

The densities of the gamma-Birnbaum-Saunders (GBS), beta-Birnbaum-Saunders (BBS), odd-log-logistic Birnbaum-Saunders (OLLBS) and odd-log-logistic Birnbaum-Saunders Poisson (OLLBSP) distributions are given by

$$f_{\text{BG}}(x;a,b,\boldsymbol{\eta}) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,g(x;\boldsymbol{\eta})G(x;\boldsymbol{\eta})^{a-1}\,\bar{G}(x;\boldsymbol{\eta})^{b-1},$$

$$f_{\text{OLLG}}(x;a,\boldsymbol{\eta}) = \frac{a\,g(x;\boldsymbol{\eta})\left[G(x;\boldsymbol{\eta})\bar{G}(x;\boldsymbol{\eta})\right]^{a-1}}{\left[G(x;\boldsymbol{\eta})^a + \bar{G}(x;\boldsymbol{\eta})^a\right]^2}$$

and

$$f_{\text{OLLG-P}}(x;a,b,\boldsymbol{\eta}) = \frac{ab\,g(x;\boldsymbol{\eta})\left[G(x;\boldsymbol{\eta})\bar{G}(x;\boldsymbol{\eta})\right]^{a-1}}{(\mathrm{e}^b-1)\left[G(x;\boldsymbol{\eta})^a + \bar{G}(x;\boldsymbol{\eta})^a\right]^2}\exp\left[\frac{b\,G(x;\boldsymbol{\eta})^a}{G(x;\boldsymbol{\eta})^a + \bar{G}(x;\boldsymbol{\eta})^a}\right],$$

respectively, where $a,b > 0$ and $\bar{G}(x;\boldsymbol{\eta}) = 1 - G(x;\boldsymbol{\eta})$.

As for the simulations, we adopt the *open source* computing platform: R Project (R Core Team, 2019). The MLEs of the parameters of the fitted densities are calculated using the `goodness.fit` function of the script `AdequacyModel` (Marinho et al., 2019) available in programming environment R Project (R Core Team, 2019) with the BFGS method. The best models fitted to the data sets are chosen based on the statistics: Cramèr-von Mises (W^*), Anderson-Darling (A^*), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC) and Kolmogorov-Smirnov (KS) and its p-value.

Initial parameter values are chosen based on a function created using the GenSA package from R Project (Xiang et al., 2013). Such a package allows implementing a function that seeks the global minimum of a given function with a large number of optimum points. Therefore, we insert functions in R that take as arguments the data set to be used and the desired density. The functions in question return the initial shots of the parameters in question.

5.2 Findings

Tables 4 and 5 give the MLEs and their standard errors (SEs) in parentheses and the information criteria, respectively. The values for all statistics (except KS) in Table 5 indicate that the GC distribution is the best model to these data. Further, the p-values of the KS statistic also reveal that all distributions (except BS, OLLBS and EW models) can be used to fit the current data. So, the information criteria support that the CG distribution provides the best fit to these data. The plots of the estimated pdfs and cdfs and the Kaplan-Meier (KM) estimate, for the two best models, displayed in Figure 7 reveal that the GC distribution is the most adequate model to these data.

For the data set 2, the MLEs, SEs and information criteria are reported in Tables 6 and 7, respectively. All information criteria also indicate to the CG distribution is the best model when compared to the others. The p-values of the KS statistic show that the BS, OLLBS and OLLBSP models can not be used for the current data. Based on the histogram, the estimated pdfs and cdfs and the KM estimate (Figure 8), we can conclude that the GC distribution provides a better fit to these data.

The likelihood ratio (LR) statistics that compare the GC and Chen models, for the two data sets, are reported in Table 8. For both data sets, the null hypothesis is rejected, and the GC distribution is a more appropriate model for both data sets.

6 Conclusions

We introduce the gamma-Chen (GC) distribution which extends the Chen model. The new distribution adds an extra shape parameter thus giving greater flexibility. We obtain some of its mathematical properties. The hazard rate function of the GC distribution may have increasing, decreasing and bathtub shapes. We show the consistency of the maximum likelihood estimators via Monte Carlo

Table 4: Fitted models to data set 1.

Model	Estimates			
BS(α,β)	29.9850	0.0034		
	(9.6966)	(0.0017)		
Chen(λ,β)	0.1410	0.5849		
	(0.0546)	(0.0685)		
OLLBS(α,β,a)	67.3595	2.1694	95.2768	
	(0.0294)	(0.0294)	(0.1584)	
GBS(α,β,a)	0.6451	3.5546	0.5637	
	(0.1265)	(0.3059)	(0.1775)	
GC(a,λ,β)	159.6704	84.6636	0.0685	
	(0.8199)	(0.1208)	(0.0088)	
EW(a,λ,β)	0.0438	7.9268	40.2123	
	(0.0103)	(0.0890)	(0.0762)	
BLL(a,b,α,β)	7.6014	493.3696	6999.9700	0.5272
	(0.0003)	(<0.0001)	(1.8495)	(0.0109)
GEF(a,λ,σ,α)	1.1000	0.3263	1110.5080	999.9747
	(<0.0001)	(<0.0001)	(<0.0001)	0.0027
KLL(a,b,α,δ)	20.5704	0.1668	0.7889	8.4532
	(<0.0001)	(0.0002)	(0.0004)	(0.0001)
OLLBSP(α,β,a,b)	479.9736	0.0513	182.0844	6.1178
	(0.0004)	(0.0004)	(0.0019)	(0.1940)
BBS(α,β,a,b)	86.2317	0.0551	0.0905	0.0800
	(0.0010)	(0.0120)	(0.0010)	(9.9186)

Table 5: Information criteria for data set 1.

Model	W^*	A^*	AIC	CAIC	BIC	HQIC	KS	p-value (KS)
BS	0.3054	1.7138	113.8926	114.6926	115.6733	114.1381	0.7303	<0.0001
Chen	0.3229	1.7662	78.7075	79.5075	80.4882	78.9530	0.2450	0.2300
OLLBS	0.2973	1.6868	117.663	119.3773	120.3341	118.0313	0.7325	<0.0001
GBS	0.2999	1.6924	77.6396	79.3538	80.3107	78.0079	0.2339	0.2782
GC	**0.2613**	**1.5455**	**73.9505**	**75.6648**	**76.6216**	**74.3188**	0.2254	0.3198
EW	0.2775	1.5989	109.3646	111.0789	112.0357	109.7329	0.4288	0.0027
BLL	0.2716	1.5835	76.7815	79.8584	80.3429	77.2725	0.2368	0.2652
GEF	0.2900	1.6348	77.9778	81.0547	81.5392	78.4689	0.2553	0.1911
KLL	0.2673	1.5726	76.6320	79.7089	80.1934	77.1231	0.2255	0.3194
OLLBSP	0.2731	1.6114	78.3789	81.4559	81.9404	78.8700	0.2189	0.3544
BBS	0.2947	1.6678	78.8362	81.9131	82.3976	79.3273	0.2535	0.1976

simulations. We prove empirically that the new distribution is better than ten known distributions by means of two real engineering data sets.

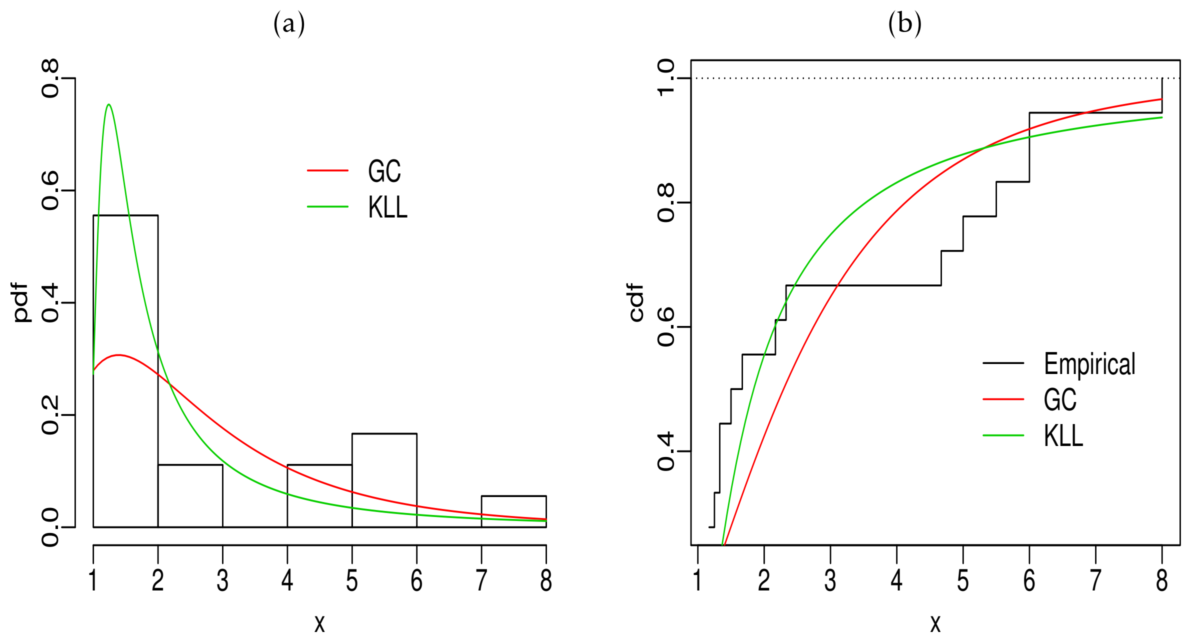

Figure 7: Estimated (a) pdfs and (b) cdfs and empirical cdf for data set 1.

Table 6: Fitted models to data set 2.

Model	Estimates			
BS(α,β)	0.7743	4.7946		
	(0.1678)	(0.0029)		
Chen(λ,β)	0.1247	0.3990		
	(0.0473)	(0.0385)		
OLLBS(α,β,a)	0.9426	1252.8320	0.0075	
	(0.0004)	(0.0319)	(0.0014)	
GBS(α,β,a)	0.6457	9.0578	0.4192	
	(0.0104)	(0.0001)	(0.1064)	
GC(a,λ,β)	18.1587	6.5901	0.1705	
	(<0.0001)	(0.2862)	(0.0178)	
EW(a,λ,β)	0.0685	5.5721	64.0574	
	(0.0171)	(0.0018)	(0.0018)	
BLL(a,b,α,β)	5.4816	465.9178	7001.0060	0.6199
	(0.0001)	(0.0142)	(0.0471)	(0.0153)
GEF$(a,\lambda,\sigma,\alpha)$	0.2547	0.6594	102.7024	78.3443
	(0.0634)	(0.0013)	(0.0051)	(0.0015)
KLL(a,b,α,δ)	20.3658	37.5799	0.1393	0.4325
	(0.0025)	(<0.0001)	(0.0763)	(0.0633)
OLLBSP(α,β,a,b)	96.3598	9.0409	150.5198	0.1610
	(0.0005)	(0.0457)	(1.7901)	(0.0760)
BBS(α,β,a,b)	27.7816	12.3196	827.4882	879.9439
	(7.9993)	(4.5306)	(0.0011)	(0.0016)

Table 7: Information criteria for data set 2.

Model	W^*	A^*	AIC	CAIC	BIC	HQIC	KS	p-value (KS)
BS	0.2187	1.3995	125.2946	126.2177	126.8398	125.3737	0.6726	<0.0001
Chen	0.2981	1.8112	100.0162	100.9393	101.5614	100.0953	0.2800	0.1626
OLLBS	0.1231	0.7648	127.8833	129.8833	130.2011	128.0020	0.6861	<0.0001
GBS	0.1217	0.7848	90.1436	92.1436	92.4614	90.2623	0.2116	0.4709
GC	**0.1109**	**0.7403**	**89.9507**	**91.9507**	**92.2685**	**90.0694**	0.2030	0.5245
EW	0.1154	0.7746	90.3318	92.3318	92.6496	90.4505	0.2156	0.4468
BLL	0.1295	0.8653	93.2454	96.8818	96.3358	93.4037	0.2162	0.4433
GEF	0.2212	1.3587	98.9108	102.5472	102.0012	99.0691	0.2226	0.4059
KLL	0.1179	0.7944	92.7261	96.3625	95.8165	92.8844	0.1992	0.5494
OLLBSP	0.1779	1.0750	193.8526	197.4889	196.9429	194.0108	0.5065	0.0005
BBS	0.1515	0.9833	94.2090	97.8454	97.2994	94.3673	0.2599	0.2300

Table 8: LR test (GC vs Chen).

Description	hypothesis	LR	p-value
data set 1	$\mathcal{H}_0 : a = 1$ vs $\mathcal{H}_1 : a \neq 1$	6.7570	0.0093
data set 2	$\mathcal{H}_0 : a = 1$ vs $\mathcal{H}_1 : a \neq 1$	12.0655	0.0005

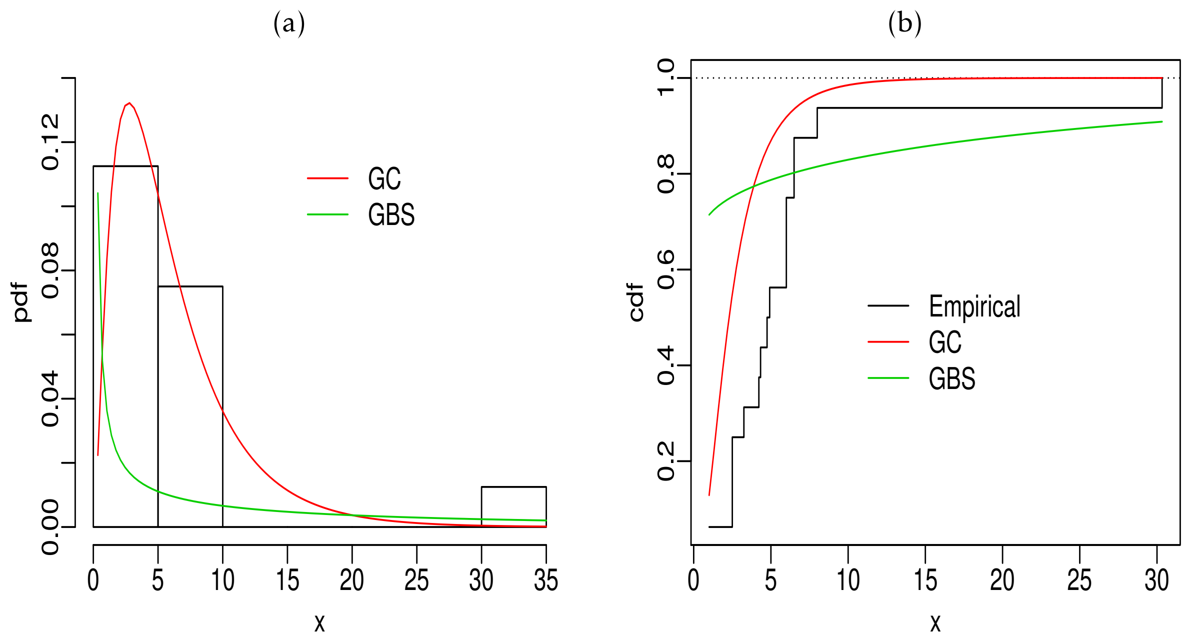

Figure 8: Estimated (a) pdfs and (b) cdfs and empirical cdf for data set 2.

Acknowledgments

The authors would like to thank the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) for funding the doctoral scholarship. Thanks to Associate Editor and reviewers.

References

Alzaatreh, A., F. Famoye, and C. Lee (2012). Gamma-Pareto Distribution and its Applications. *Journal of Modern Applied Statistical Methods 11*(1), 78–94.

Anzagra, L., S. Sarpong, and S. Nasiru (2020). Odd Chen-G Family of Distributions. *Annals of Data Science 16*(3), 1–23.

Birnbaum, Z.W. and S.C. Saunders (1969). A new family of life distributions. *Journal of Applied Probability 6*(1), 319–327.

Castellares, F. and A.J. Lemonte (2015). A new generalized Weibull distribution generated by gamma random variables. *Journal of the Egyptian Mathematical Society 23*(2), 382–390.

Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters 49*(2), 155–161.

Cordeiro, G.M. and A.J. Lemonte (2011). The β-Birnbaum-Saunders distribution: an improved distribution for fatigue life modeling. *Computational Statistics & Data Analysis 55*(3), 1445–1461.

Cordeiro, G.M., M.C.S. Lima, A.H.M.A. Cysneiros, M.A.R. Pascoa, R.R. Pescim, and E.M.M. Ortega (2016). An extended Birnbaum-Saunders distribution: Theory, estimation, and applications. *Communications in Statistics-Theory and Methods 45*(8), 2268–2297.

Cordeiro, G.M., M.C.S. Lima, E.M.M. Ortega, and A.K. Suzuki (2018). A new extended Birnbaum-Saunders model: properties, regression and applications. *Stats 1*(1), 32–47.

da Silva, R.V., T.A.N. de Andrade, D.B.M. Maciel, R.P.S. Campos, and G.M. Cordeiro (2013). A new lifetime model: The gamma extended Fréchet distribution. *Journal of Statistical Theory and Applications 12*(1), 39–54.

de Santana, T.V.F., E.M.M. Ortega, G.M. Cordeiro, and G.O. Silva (2012). The Kumaraswamy-log-logistic distribution. *Journal of Statistical Theory and Applications 11*(3), 265–291.

Dey, S., D. Kumar, P.L. Ramos, and F. Louzada (2017). Exponentiated Chen distribution: properties and estimation. *Communications in Statistics-Simulation and Computation 46*(10), 8118–8139.

Khan, M.S., R. King, and I.L. Hudson (2018). Kumaraswamy exponentiated Chen distribution for modelling lifetime data. *Applied Mathematics and Information Sciences 12*(3), 617–623.

Kilbas, A.A., H.M. Srivastava, and J.J. Trujillo (2006). *Theory and applications of fractional differential equations*, Volume 204. Amsterdam: Elsevier.

Lemonte, A.J. (2014). The beta log-logistic distribution. *Brazilian Journal of Probability and Statistics 28*(3), 313–332.

Lima, M.C.S. (2015). *Mathematical properties of some generalized gamma models*. Ph. D. thesis, Universidade Federal de Pernambuco.

Marinho, P.R.D., R.B. Silva, M. Bourguignon, G.M. Cordeiro, and S. Nadarajah (2019). AdequacyModel: An R package for probability distributions and general purpose optimization. *PLOS ONE 14*(8), 1–30.

Mudholkar, G.S. and A.D. Hutson (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability 42*(2), 299–302.

Nadarajah, S., G.M. Cordeiro, and E.M.M. Ortega (2015). The ZografosâĂŞBalakrishnan-G Family of Distributions: Mathematical Properties and Applications. *Communications in Statistics - Theory and Methods 44*(1), 186–215.

Ortega, E.M.M., A.J. Lemonte, G.M. Cordeiro, and J.N. da Cruz (2016). The odd Birnbaum-Saunders regression model with applications to lifetime data. *Journal of Statistical Theory and Practice 10*(4), 780–804.

Pinho, L.G.B., G.M. Cordeiro, and J.S. Nobre (2012). The gamma-exponentiated Weibull distribution. *Journal of Statistical Theory and Applications 11*(4), 379–395.

Pogány, T.K., G.M. Cordeiro, M.H. Tahir, and H.M. Srivastava (2017). Extension of generalized integro-exponential function and its application in study of Chen distribution. *Applicable Analysis and Discrete Mathematics 11*(2), 434–450.

Pogány, T.K. and A. Saboor (2016). The gamma exponentiated exponential-Weibull distribution. *Filomat 30*(12), 3159–3170.

Qian, L (2012). The Fisher information matrix for a three-parameter exponentiated Weibull distribution under type II censoring. *Statistical Methodology 9*(3), 320–329.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sivini, P.G.L. (2006). *Desenvolvimento de banco de dados de confiabilidade: uma aplicação em estações redutoras de pressão de gás natural*. Master thesis, Universidade Federal de Pernambuco.

Tarvirdizade, B. and M. Ahmadpour (2019). A new extension of Chen distribution with applications to lifetime data. *Communications in Mathematics and Statistics 14*(5), 1–16.

Torabi, H. and N.M. Hedesh (2012). The gamma-uniform distribution and its applications. *Kybernetika 48*(1), 16–30.

Xiang, Y., S. Gubian, B. Suomela, and J. Hoeng (2013). Generalized Simulated Annealing for Global Optimization: The GenSA Package. *R Journal 5*(1), 13–28.

Xie, M, Y Tang, and T N Goh (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety 76*(3), 279–285.

Zografos, K. and N. Balakrishnan (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology 6*(4), 344–362.