Chilean Journal of Statistics
Volume 14, Number 2 (December 2023), Pages 99–122
DOI: 10.32372/ChJS.14-02-03

Distributions Theory
Research Article

New exponentiated-Weibull-G family: Properties, simulations, regression and applications to COVID-19 data

Gauss M. Cordeiro[1], Elisângela C. Biazatti[1,2,*], Maria do Carmo S. de Lima[1], and Luís H. de Santana[3]

[1]Department of Statistics, Federal University of Pernambuco, Recife, Brazil
[2]Department of Mathematics and Statistics, Federal University of Rondônia, Ji-Paraná, Brazil
[3]Technology Department, State University of Maringá, Umuarama, Brazil

(Received: 22 June 2022 · Accepted in final form: 29 May 2023)

Abstract

In this article, a new exponentiated-Weibull-G family with two extra shape parameters is defined, which incorporates certain special classes of distributions. Some of its mathematical properties are presented. Maximum likelihood estimation and simulations are addressed. Two applications to COVID-19 data show the flexibility of the new family. A simulation experiment is developed to show the validity of the asymptotic properties of the maximum likelihood estimators. A regression model, called the log-exponentiated Weibull-Weibull, is constructed and applied to COVID-19 data, which has better performance when compared to two other regression models.

Keywords: Exponentiated-Weibull distribution · Generalized distributions · Maximum likelihood estimation · Moment · T-X family.

Mathematics Subject Classification: Primary 62E10 · Secondary 62F10.

1. Introduction

The process of developing new distributions has been improved over the last few years. We can mention two recent classes of distributions: the arctan (Gómez-Déniz et al, 2022) and the erf-G (Fernández and de Andrade, 2020) families. One of the most recent techniques to generate new families of distributions has revolutionized the area, bringing a way to generate more flexible models. This technique, called the transformed-transformer (T-X) family of distributions, was developed by Alzaatreh et al (2013). However, despite the great effort to find distributions or families that adequately model a large number of types of data sets, many models proposed in the literature are still not very competitive when compared to the beta-G (B) and Kumaraswamy-G (Kw-G) classes proposed by Eugene et al (2002) and Cordeiro and de Castro (2011), respectively.

*Corresponding author. Email: elisangela.biazatti@unir.br

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

http://www.soche.cl/chjs

There is a structural difference between the two aforementioned classes. Despite the fact that the B-G distribution provides in general an excellent performance in the adjustment to real data sets, the structure of its probability density and cumulative distribution functions (PDF and CDF, respectively) include complicated mathematical functions, which makes the parameter estimation process more difficult. In addition, the Kw-G class does not have any complicated functions in its PDF.

The main idea of the present work is to present, based on the T-X family of distributions, a competitive class to the B-G and Kw-G families. The sections of the article are organized as follows. The new family is introduced in Section 2 and this same section provides four special models of the EW-G family. Section 3 gives the skewness and kurtosis of X based on quantiles, a linear representation for the EW-G PDF given in Equation (2.5), and some of its properties, say moments and generating function. The maximum likelihood (ML) method applied to the PDF given in Equation (2.5) is addressed in Section 4 together with a regression model and a simulation study. In Section 5, two applications to real data sets prove the potentiality of the EW-G family. Finally, Section 6 presents the main conclusions of this work and future research motivated by this work.

2. The new family and related-models

2.1 The new family

Let $G(x)$ be the CDF of a random variable X with parameter vector $\boldsymbol{\theta}$ and $r(t)$ be the PDF of a random variable T with parameter vector $\boldsymbol{\eta}$ defined on $[a, b]$, for $-\infty \leq a < b \leq \infty$. Henceforth, the parameters are omitted in all functions.

The CDF of the T-X family was defined by Alzaatreh et al (2013) as

$$F(x) = \int_a^{W(G(x))} r(t)\,\mathrm{d}t, \tag{2.1}$$

where W is a differentiable and monotonically non-decreasing function such that

$$\lim_{x\to-\infty} W(G(x)) = a, \quad \lim_{x\to\infty} W(G(x)) = b.$$

If a random variable T is defined on $[0, \infty)$ and $W(G(x)) = -\ln(1 - G(x))$, the CDF of the T-X family given in Equation (2.1) reduces to

$$F(x) = \int_0^{-\ln(1-G(x))} r(t)\,\mathrm{d}t. \tag{2.2}$$

By considering the three-parameter exponentiated-Weibull (EW) PDF (Mudholkar and Srivastava, 1993) for T, namely

$$r(t) = \frac{\alpha\, p}{\lambda} \left(\frac{t}{\lambda}\right)^{p-1} \exp\left(-\left(\frac{t}{\lambda}\right)^{p}\right) \left(1 - \exp\left(-\left(\frac{t}{\lambda}\right)^{p}\right)\right)^{\alpha-1},$$

in Equation (2.2), we obtain

$$F(x) = \left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x)\right)\right)^{p}\right)\right)^{\alpha}. \tag{2.3}$$

Equation (2.3) defines the exponentiated-Weibull-G (EW-G) family (generated by G) with scale parameter $\lambda > 0$, shape parameters $\alpha > 0$ and $p > 0$. Henceforth, let $X \sim$

EW-G$(\alpha, \lambda, p, \boldsymbol{\theta})$ be a random variable with CDF given in Equation (2.3).

By inverting Equation (2.3), the quantile function (QF) of X has the form

$$Q(u) = G^{-1}\Big(1 - \exp\Big(- \sqrt(p) - \lambda^p \ln\Big(1 - \sqrt(\alpha) u\Big)\Big)\Big), \tag{2.4}$$

where $G^{-1}(u)$ is the baseline QF.

As long as we have a closed form for $G^{-1}(u)$ or a good approximation for it, the QF of X is easily obtained from the parent QF. Thus, if U has a uniform $U(0,1)$ distribution, then $X = Q(U)$.

By differentiating Equation (2.3), the PDF of X follows as

$$\begin{aligned} f(x) = {} & \frac{\alpha\, p\, g(x)}{\lambda^p (1 - G(x))} \left(-\ln\left(1 - G(x)\right)\right)^{p-1} \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x)\right)\right)^p\right) \\ & \times \left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x)\right)\right)^p\right)\right)^{\alpha - 1}, \end{aligned} \tag{2.5}$$

where $g(x) = \mathrm{d}G(x)/\mathrm{d}x$. Two examples presented in Section 5 with $\lambda = 1$ (that is, EW-G$(\alpha, 1, p, \boldsymbol{\theta})$) prove that the EW-G family can be a competitive alternative for both B-G and Kw-G classes mentioned before. The hazard rate function (HRT) of X can be expressed as $h(x) = f(x)/(1 - F(x))$.

2.2 Special models

Equation (2.5) is most tractable whenever $G(x)$ and $g(x)$ have simple analytic expressions. The baseline $G(x)$ is clearly a special case of Equation (2.3) when $\alpha = \lambda = p = 1$. Setting $\lambda = p = 1$ leads to the exponentiated type of distributions (Gupta et al, 1998). If $p = 1$ and $\beta = \lambda^{-1}$, we have that

$$F(x) = \left(1 - (1 - G(x))^{\beta}\right)^{\alpha}$$

is identical to the exponentiated generalized class (Cordeiro et al, 2013). Moreover, the exponentiated-exponential logistic defined by Ghosh and Alzaatreh (2018) is another special model when the baseline is the logistic and $p = 1$. In the following, some special models are discussed.

2.3 Exponentiated-Weibull gamma (EW-Ga) model

The gamma CDF (for $x > 0$) with shape parameter $a > 0$ and scale parameter $b > 0$ is $G(x) = \gamma(a, bx)/\Gamma(a)$, where $\Gamma(a) = \int_0^\infty w^{a-1} \mathrm{e}^{-w}\, \mathrm{d}w$ and $\gamma(a, x) = \int_0^x w^{a-1} \mathrm{e}^{-w}\, \mathrm{d}w$ are the gamma and incomplete gamma functions, respectively.

The EW-Ga PDF can be expressed as

$$\begin{aligned} f(x) = {} & \frac{\alpha\, p\, b^a\, x^{a-1} \exp(-\mathrm{bx})}{\lambda^p\left(\Gamma(a) - \gamma(a, bx)\right)} \left(-\ln\left(1 - \frac{\gamma(a, bx)}{\Gamma(a)}\right)\right)^{p-1} \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \frac{\gamma(a, bx)}{\Gamma(a)}\right)\right)^p\right) \\ & \times \left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \frac{\gamma(a, bx)}{\Gamma(a)}\right)\right)^p\right)\right)^{\alpha - 1}. \end{aligned}$$

2.4 Exponentiated-Weibull Weibull (EW-W) model

The Weibull CDF (for $x > 0$) with shape parameter $k > 0$ and scale parameter $b > 0$ is $G(x) = 1 - \exp(-(x/b)^k)$. The EW-W PDF has the form given by

$$f(x) = \frac{\alpha\, p\, k}{b\lambda^p}\Big(\frac{x}{b}\Big)^{kp-1} \exp\Big(-\lambda^{-p}\Big(\frac{x}{b}\Big)^{kp}\Big)\Big(1-\exp\Big(-\lambda^{-p}\Big(\Big(\frac{x}{b}\Big)^{k}\Big)^{p}\Big)\Big)^{\alpha-1}. \tag{2.6}$$

For $k = 1$, it gives the exponentiated-Weibull exponential distribution.

2.5 Exponentiated-Weibull log-logistic (EW-LL) model

The CDF of the log-logistic (LL) distribution is (for $x, a, b > 0$) $G(x) = 1 - (1 + (x/a)^b)^{-1}$. The EW-LL PDF can be expressed as

$$f(x) = \frac{\alpha\, p\, b}{\lambda^p a^b}\left(1+\Big(\frac{x}{a}\Big)^b\right)^{-3}\left(-\ln\left(1+\Big(\frac{x}{a}\Big)^b\right)^{-1}\right)^{p-1}\exp\left(-\lambda^{-p}\left(-\ln\left(1+\Big(\frac{x}{a}\Big)^b\right)^{-1}\right)^{p}\right)$$
$$\times\left(1-\exp\left(-\lambda^{-p}\left(-\ln\left(1+\Big(\frac{x}{a}\Big)^b\right)^{-1}\right)^{p}\right)\right)^{\alpha-1}.$$

2.6 Exponentiated-Weibull Birnbaum-Saunders (EW-BS) model

The CDF of the Birnbaum-Saunders (BS) distribution is (for $x, a, b > 0$) $G(x) = \Phi(1/a(\sqrt{x/b} - \sqrt{b/x}))$. The EW-BS PDF has the form stated as

$$f(x) = \frac{\alpha\, p\, x^{-3/2}(x+b)e^{a^{-2}}\exp\left(-\frac{1}{2a^2}\left(\frac{x}{b}+\frac{b}{x}\right)\right)}{2a\lambda^p\sqrt{2\pi b}\left(1-\Phi\left(\frac{1}{a}\left(\sqrt{\frac{x}{b}}-\sqrt{\frac{b}{x}}\right)\right)\right)}\left(-\ln\left(1-\Phi\left(\frac{1}{a}\left(\sqrt{\frac{x}{b}}-\sqrt{\frac{b}{x}}\right)\right)\right)\right)^{p-1}$$
$$\times\exp\left(-\lambda^{-p}\left(-\ln\left(1-\Phi\left(\frac{1}{a}\left(\sqrt{\frac{x}{b}}-\sqrt{\frac{b}{x}}\right)\right)\right)\right)^{p}\right)$$
$$\times\left(1-\exp\left(-\lambda^{-p}\left(-\ln\left(1-\Phi\left(\frac{1}{a}\left(\sqrt{\frac{x}{b}}-\sqrt{\frac{b}{x}}\right)\right)\right)\right)^{p}\right)\right)^{\alpha-1}.$$

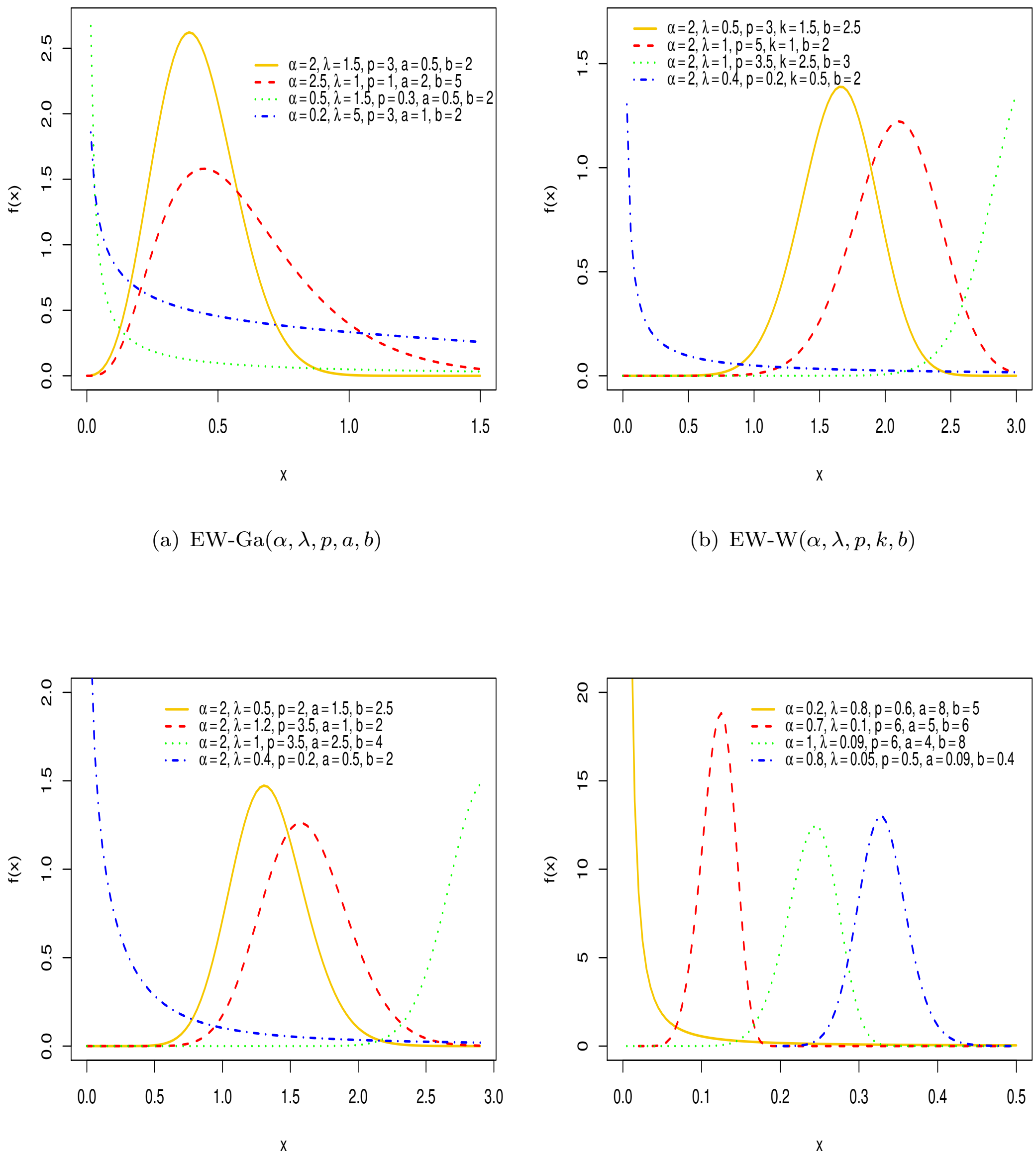
Figures 1 and 2 report some shapes of four generated PDFs and HRTs, respectively. We note the flexibility of the new family for selected baselines.

3. Other functions and properties

3.1 Quantile function

The skewness and kurtosis of X can be based on quantile measures easily calculated from Equation (2.4). In fact, the Bowley skewness S (Kenney and Keeping, 1961) and the Moors kurtosis K (Moors, 1988) are given by

$$S = \frac{Q(3/4) + Q(1/4) - 2Q(1/2)}{Q(3/4) - Q(1/4)}$$



(a) EW-Ga$(\alpha, \lambda, p, a, b)$ (b) EW-W$(\alpha, \lambda, p, k, b)$

(c) EW-LL$(\alpha, \lambda, p, a, b)$ (d) EW-BS$(\alpha, \lambda, p, a, b)$

Figure 1. EW-G PDFs for indicated models.

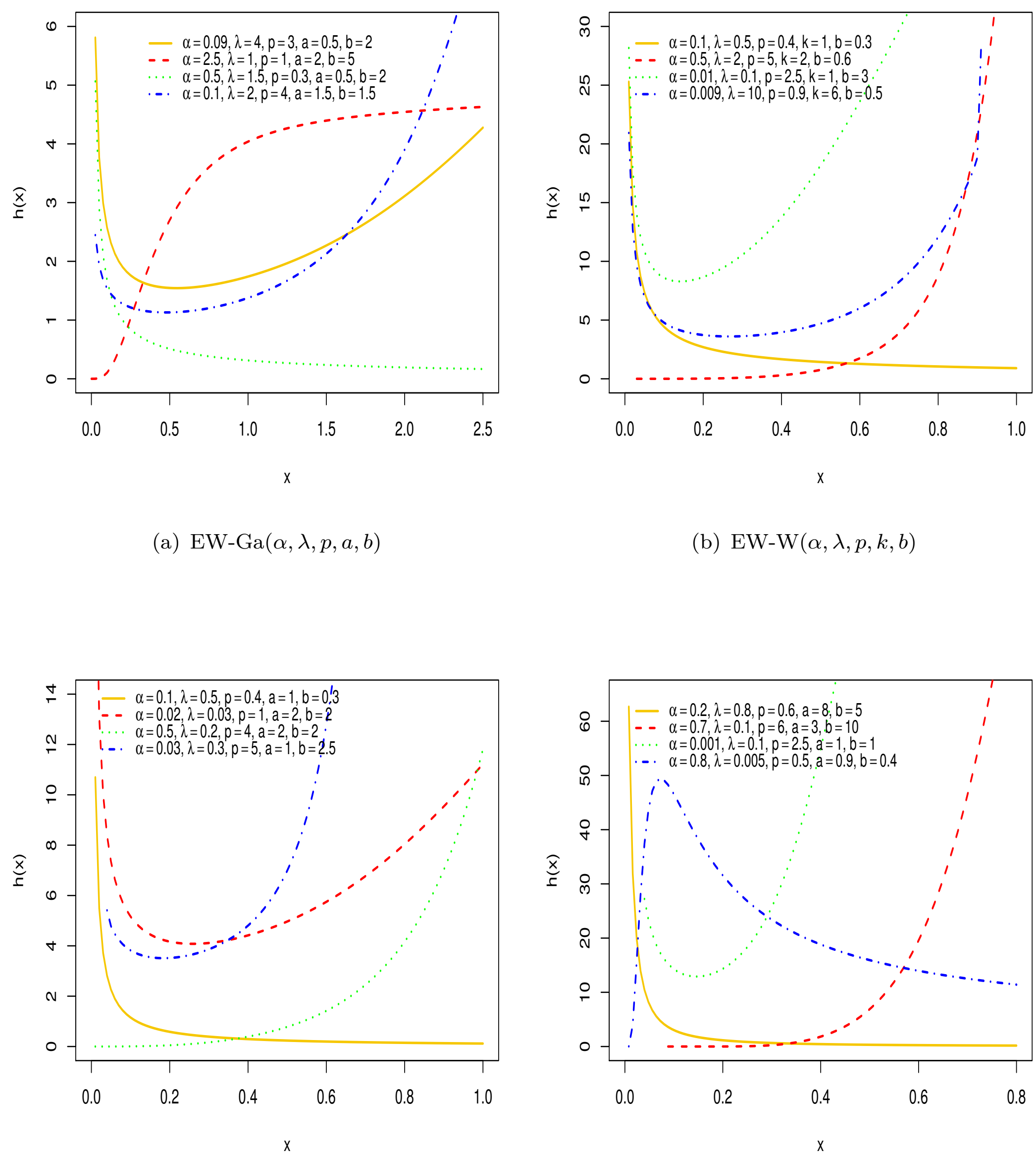
and

$$K = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)},$$

respectively.

Figures 3(a), 3(b), 4(a) and 4(b) display the measures S and K for the EW-W$(\alpha, \lambda, p, k, b)$ distribution for some values of k as functions of both α and p, respectively.

If α increases in Figure 3(a) when $k = 3$, a very sudden drop occurs, initially to 0.2 and then the asymmetry remains practically constant. The other configurations maintain practically the same behavior, but when k increases, the asymmetry decreases.



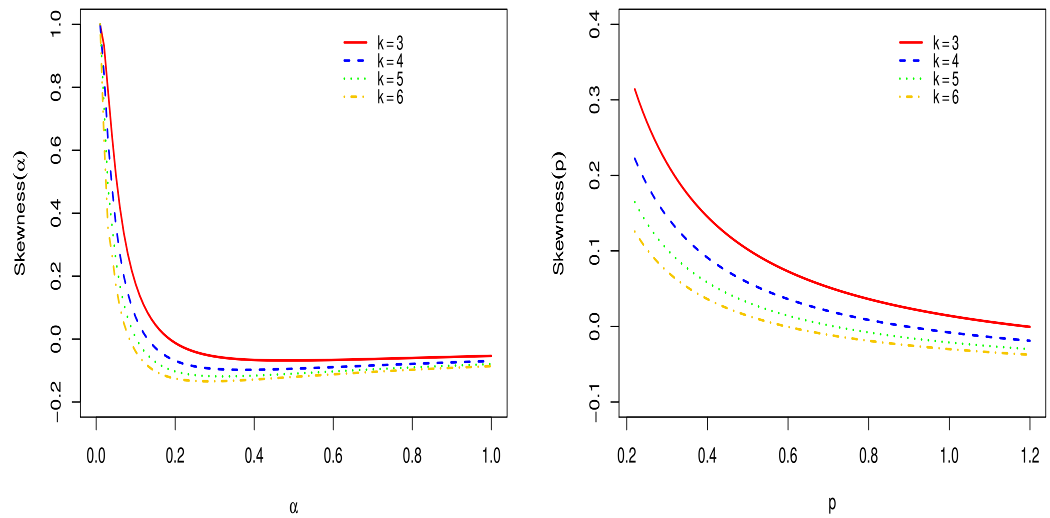
(a) EW-Ga$(\alpha, \lambda, p, a, b)$ (b) EW-W$(\alpha, \lambda, p, k, b)$

(c) EW-LL$(\alpha, \lambda, p, a, b)$ (d) EW-BS$(\alpha, \lambda, p, a, b)$

Figure 2. EW-G HRTs for indicated models.

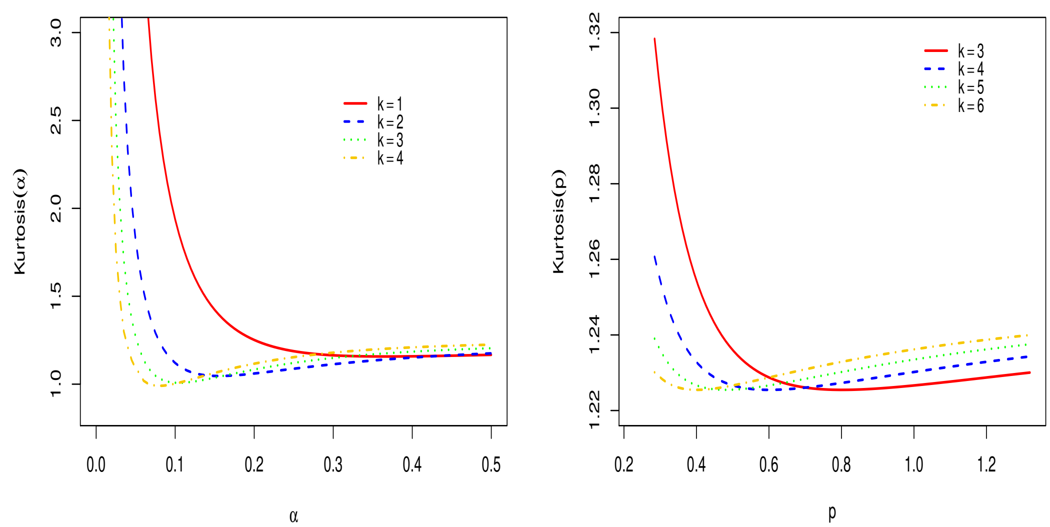
Figure 3(b) shows that when p increases, there is a decrease in the asymmetry, but it does not occur abruptly (for all scenarios). If p is fixed, the values for the asymmetry are smaller when k increases.

The behavior in Figures 4(a) and 4(b) are similar. In Figure 4(a) for $k = 1$, the kurtosis decreases, and for $k = 2$, $k = 3$ and $k = 4$, there is an inversion in the behavior of the kurtosis, that is, fixing α, the kurtosis increases when k increases. In Figure 4(b), for values of p between 0 and approximately 0.3, there is no kurtosis. For some value of p, except when $k = 3$, there is an inversion of the behavior of the kurtosis for the cases $k = 4$, $k = 5$ and $k = 6$, that is, fixing p, the kurtosis values increase when k increases.



(a) (b)

Figure 3. Bowley skewness of the EW-W distribution. (a) for some values of k ($p = b = 2$ and $\lambda = 1$) and (b) for some values of k ($\alpha = b = 2$ and $\lambda = 1$).



(a) (b)

Figure 4. Moors' kurtosis of the EW-W distribution. (a) for some values of k ($p = b = 2$ and $\lambda = 1$) and (b) for some values of k ($\alpha = b = 2$ and $\lambda = 1$)

3.2 Linear representation

Here, we present a linear representation for the EW-G PDF given in Equation (2.5).

First, the CDF and PDF of the exponentiated-G (Exp-G) distribution for an arbitrary parent CDF $G(x)$ with power parameter $c > 0$, are $\Pi_c(x) = G(x)^c$ and PDF $\pi_c(x) = c\, g(x)\, G(x)^{c-1}$, respectively. Some Exp-G properties have been studied by many authors in recent years; see, for instance, Mudholkar and Srivastava (1993) for exponentiated Weibull, Nadarajah and Kotz (2003) for exponentiated Fréchet, Nadarajah (2005) for exponentiated Gumbel, and Nadarajah and Gupta (2007) for exponentiated gamma distributions.

In Appendix A, it is shown that a linear representation of the EW-G is given by

$$f(x) = \sum_{r=0}^{\infty} c_{r+1}\, \pi_{r+1}(x), \tag{3.7}$$

where $\pi_{r+1}(x) = (r+1)\; g(x)\, G(x)^r$ is the PDF of the Exp-G random variable V_{r+1} with power parameter $r+1$. Thus, some structural properties of the EW-G family can be determined from those of the Exp-G class.

3.3 Properties

First, the nth ordinary moment of X, say $\mu'_n = \mathrm{E}(X^n)$, follows from Equation (3.7) and it is given by

$$\mu'_n = \sum_{r=0}^{\infty} c_{r+1}\, \mathrm{E}(V_{r+1}^n) = \sum_{r=0}^{\infty} (r+1)\, c_{r+1}\, \delta_{n,r},$$

where $\delta_{n,r} = \int_0^1 \left(Q_G(u)\right)^n\; u^r\, \mathrm{d}u$ is easily calculated numerically from the parent QF.

Some moments of the EW-E distribution are obtained in this way using the `R` software. Let $\alpha = 1.7$, $\lambda = 1.5$, $p = 3.5$ and $\beta = 1.5$ be the reference setup, and we study the behavior of the variance, skewness and kurtosis of X increasing α, λ and p. Table 1 reports some numerical findings for these quantities.

For all setups, the asymmetry and kurtosis are positive. Further, we note that when α increases, keeping the other parameters fixed, there is an increase in the first four moments and in the variance, but the skewness and kurtosis values decrease.

Table 1. Some moments of the EW-E model.

Parameter	μ'_1	μ'_2	μ'_3	μ'_4	Variance	Skewness	Kurtosis
Reference	0.06463	0.08201	0.11836	0.19267	0.07783	4.74316	27.08344
α=2.7	0.07614	0.10780	0.16755	0.28627	0.10200	4.41467	22.96156
α=3.7	0.08830	0.13207	0.21366	0.37540	0.12427	4.10998	19.80985
α=4.7	0.10028	0.15524	0.25777	0.46159	0.14518	3.85219	17.42437
λ=2.5	0.03047	0.05452	0.11010	0.24845	0.05359	8.47762	81.94044
λ=3.5	0.00758	0.01527	0.03462	0.08721	0.01521	18.26429	372.27010
λ=4.5	0.00215	0.00460	0.01110	0.02976	0.00460	35.51967	1404.38641
p=4.5	0.13653	0.20569	0.34827	0.66034	0.18705	3.32676	14.06580
p=5.5	0.35587	0.62986	1.24980	2.76824	0.50321	1.86989	5.60635
p=6.5	1.09053	2.26383	5.25519	13.55143	1.07458	0.39743	2.19823

The first four moments and the variance decrease when λ increases (with the other parameters fixed). The skewness and kurtosis decrease slowly when α increases. And if λ increases, there is an increase in the skewness and kurtosis values, being much faster for the kurtosis. Thus, α exerts an influence on the skewness and kurtosis in the sense of decreasing (slow), and λ exerts influence on increasing the skewness and kurtosis, which occurs more quickly.

By analyzing results varying p, we note that the effect of increasing p is similar to the effect caused when α increases, that is, when p increases, fixing the other parameters, the first four moments and the variance (higher for p=6.5) increase and the skewness and kurtosis decrease. The observed behavior may be related to the fact that α and p are shape parameters, and λ is a scale parameter. The scale parameter λ exerts profound effect over the skewness and kurtosis.

Second, the nth incomplete moment of X, say $m_n(t)$, can be expressed from Equation (3.7) as

$$m_n(t) = \int_{-\infty}^{t} x^n\, f(x)\, \mathrm{d}x = \sum_{r=0}^{\infty} c_{r+1} \int_{-\infty}^{t} x^n\, \pi_{r+1}(x)\, \mathrm{d}x.$$

The last integral is just the nth incomplete moment of V_{r+1}.

Third, the moment generating function $M_X(s) = \mathrm{E}(\exp(\mathrm{s\,X}))$ of X follows from Equation (3.7)

$$M_X(s) = \sum_{r=0}^{\infty} c_{r+1}\, M_{r+1}(s) = \sum_{r=0}^{\infty} (r+1)\, c_{r+1}\, \tau_{s,r},$$

where $M_{r+1}(s)$ is the moment generating function of V_{r+1} (for $r \geq 0$), and $\tau_{s,r} = \int_0^1 \exp\left(s\, Q_G(u)\right)\, u^r\, \mathrm{d}u$ can be evaluated numerically from the baseline QF.

4. Estimation, regression, and simulations

4.1 Estimation

The ML method is used to estimate the unknown parameters of the EW-G family. We adopt the adequacy measures to compare fitted models: Akaike information criterion (AIC), corrected Akaike information criterion (CAIC), Bayesian information criterion (BIC), and the Kolmogorov-Smirnov (K-S), Anderson-Darling (A^*) and Cramér-von Mises (W^*) statistics (Chen and Balakrishnan, 1995).

Consider a sample $x_1, \ldots, x_n$ from Equation (2.5), where $\boldsymbol{\theta}$ is a parameter vector in $G(x)$. The log-likelihood function $l = \log L(\alpha, \lambda, p, \boldsymbol{\theta})$ for the parameters is

$$\begin{aligned} l = {} & n \ln\left(\frac{\alpha p}{\lambda^p}\right) + \sum_{i=1}^{n} \ln g(x_i) - \sum_{i=1}^{n} \ln(1 - G(x_i)) \\ & + (p-1) \sum_{i=1}^{n} \ln\left(-\ln(1 - G(x_i))\right) - \frac{1}{\lambda^p} \sum_{i=1}^{n} \left(-\ln(1 - G(x_i))\right)^p \\ & + (\alpha - 1) \sum_{i=1}^{n} \ln\left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)\right). \end{aligned} \tag{4.8}$$

A good way to obtain the ML estimates and their standard errors (SE) is the `AdequacyModel` library of the `R` software (Marinho et al, 2019). In addition, this library also provides some statistics to evaluate the adequacy of a fitted distribution.

The `AdequacyModel` package can also maximize Equation (4.8) using the particle swarm optimization (PSO) approach from the quasi-Newton BFGS, Nelder-Mead, and simulated-annealing methods and it does not require initial values. In fact, depending on the complexity of the log-likelihood function, problems can be faced, depending on the initial values when using Newton or quasi-Newton methods to optimize the function. These are situations where the numerical derivative may not give us good faster ascent directions. For these problems, algorithms such as BFGS, Nelder-Mead, among others, can be used. However, in the present article, and based on Marinho et al (2019), to which one of the authors of this article is a member, good initial guesses were taken, through meta-heuristic algorithms that do not use derivatives and that are numerically quite consistent to obtain a global optimum, as is the Simulated Annealing (SANN) and PSO algorithms. They are not influenced by approximately flat regions (derivatives close to zero). The link https://www.youtube.com/watch?v=qfRbX54RHf4 provides a video on YouTube of the PSO algorithm from the `AdequacyModel` package of an extreme situation, in which the initial guess was given in a region whose derivative is zero and where the object is to minimize the objective function. Further, In a convex function, as is the case of this likelihood function, taking as initial guesses the responses of the SANN and/or PSO algorithms in a quasi-Newton method such as the BSFS is a great strategy for obtaining a consistent iterative ML estimate. Other details are available at https://rdrr.io/cran/AdequacyModel/.

The `optim` function of `R` software is a good alternative for maximizing l, since it provides numerical checks of the behavior of the Hessian matrix, and then we can select an optimization method, for example, Nelder-Mead, BFGS, CG, L-BFGS-B, SANN and Brent. The elements of the score vector $U(\alpha,\lambda,p,\boldsymbol{\theta})$ are given in Appendix B.

The ML estimates are obtained by solving $U(\alpha,\lambda,p,\boldsymbol{\theta})=(0,0,0,0)^\top$. Note that from U_α, it is possible to obtain a semi closed-form ML estimate for α, namely

$$\widehat{\alpha}=-\frac{n}{\sum_{i=1}^{n}\ln\Big(1-\exp\Big(-\widehat{\lambda}^{-\widehat{p}}\,(-\ln(1-G(x_i)))^{\widehat{p}}\Big)\Big)}.$$

Thus, the ML estimate of α can be written as a function of the ML estimates of λ, p and $\boldsymbol{\theta}$. Then, we move from a (q + 3) × (q + 3) system to a (q + 2) × (q + 2) system of nonlinear equations.

4.2 Log-exponentiated Weibull-Weibull (LEW-W) regression

If X is a random variable having the EW-W PDF given in Equation (2.6), then $Y=\log(X)$ follows the LEW-W PDF, which reparameterized in terms of $k=\sigma^{-1}$ and $b=\exp(\mu)$, has the form (for $y\in\mathrm{R}$)

$$\begin{aligned}
f_Y(y)=&\frac{\alpha\, p\,\exp\left(-\frac{y-\mu}{\sigma}-\exp\left(-\frac{y-\mu}{\sigma}\right)\right)}{\sigma\,\lambda^p\left(1-\exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right)\right)}\left(-\ln\left(1-\exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right)\right)\right)^{p-1}\\
&\times\exp\left(-\lambda^{-p}\left(-\ln\left(1-\exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right)\right)\right)^{p}\right)\\
&\times\left(1-\exp\left(-\lambda^{-p}\left(-\ln\left(1-\exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right)\right)\right)^{p}\right)\right)^{\alpha-1},
\end{aligned}\qquad(4.9)$$

where $\mu\in\mathrm{R}$, $\sigma>0$, $\alpha>0$, $\lambda>0$ and $p>0$.

The survival function corresponding to Equation (4.9) is stated as

$$S(y) = 1 - \left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right)\right)\right)^{p}\right)\right)^{\alpha}.$$

The standardized random variable $Z = (Y-\mu)/\sigma$ has PDF (for $z \in \mathrm{R}$) defined as

$$\begin{aligned}\pi(z) = {} & \frac{\alpha\, p \exp\left(-z - \exp\left(-z\right)\right)}{\lambda^{p}\left(1 - \exp\left(-\exp\left(-z\right)\right)\right)} \left(-\ln\left(1 - \exp\left(-\exp\left(-z\right)\right)\right)\right)^{p-1} \\ & \times \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \exp\left(-\exp\left(-z\right)\right)\right)\right)^{p}\right) \\ & \times \left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \exp\left(-\exp\left(-z\right)\right)\right)\right)^{p}\right)\right)^{\alpha-1}. \end{aligned} \tag{4.10}$$

Many studies involve the presence of explanatory variables that can influence the lifetimes x_i. Therefore, consider the location and scale model defined by

$$y_i = \boldsymbol{v}_i^{\top}\boldsymbol{\gamma} + \sigma\, z_i, \; i = 1, \ldots, n, \tag{4.11}$$

where $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{ip})^{\top}$ is the explanatory variable vector modeling the location parameter $\mu_i = \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}$ for the ith response variable y_i (for $i = 1, \ldots, n$), $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^{\top}$ is the vector of regression coefficients, the random error z_i has the PDF given in Equation (4.10), and $\sigma > 0$ is a scale parameter.

Let F be the group of individuals who failed and C the group of censored individuals. The log-likelihood function for $\boldsymbol{\theta} = (\alpha, \lambda, p, \sigma, \boldsymbol{\gamma}^{\top})^{\top}$ can be determined from Equations (4.10) and (4.11) as

$$\begin{aligned} l(\boldsymbol{\theta}) = {} & q \ln\left(\frac{\alpha p}{\lambda^{p}}\right) + \sum_{i \in F}\left(\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma}\right) - \exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma}\right)\right) \\ & - \sum_{i \in F}\ln\left(1 - \exp\left(-\exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma}\right)\right)\right) \\ & + (p-1)\sum_{i \in F}\ln\left(-\ln\left(1 - \exp\left(-\exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma}\right)\right)\right)\right) \\ & - \frac{1}{\lambda^{p}}\sum_{i \in F}\left(-\ln\left(1 - \exp\left(-\exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma}\right)\right)\right)\right)^{p} \\ & + (\alpha-1)\sum_{i \in F}\ln\left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \exp\left(-\exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma} v\right)\right)\right)\right)^{p}\right)\right) \quad (4.12) \\ & + (n-q)\sum_{i \in C}\ln\left(1 - \left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - \exp\left(-\exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\boldsymbol{\gamma}}{\sigma}\right)\right)\right)\right)^{p}\right)\right)^{\alpha}\right), \end{aligned}$$

where q is the number of failures. The ML estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be found by maximizing Equation (4.12). From the fitted regression given in Equation (4.11), the survival function of y_i can be estimated by

$$\widehat{S}(y_i) = \left(1 - \left(1 - \exp\left(-\widehat{\lambda}^{-\widehat{p}}\left(-\ln\left(1 - \exp\left(-\exp\left(-\frac{y_i - \boldsymbol{v}_i^{\top}\widehat{\boldsymbol{\gamma}}}{\widehat{\sigma}}\right)\right)\right)\right)^{\widehat{p}}\right)\right)^{\widehat{\alpha}}\right).$$

4.3 Simulation study

A Monte Carlo simulation study is carried out by taking the exponential (E) for baseline (with mean $1/\beta$) to investigate the accuracy of the ML estimators of the parameters. The simulation procedure consists of generating observations from the EW-E$(\alpha, \lambda, p, \beta)$ distribution using the inversion method for different parameter combinations. The number of replications is 10,000, the sample size is $n \in [50, 100, 150]$, and we adopt the BFGS algorithm in the `R` software to maximize Equation (4.8). The simulation process follows the steps:

1. Generate $U \sim \mathrm{U}(n, 0, 1)$
2. Return $X = F^{-1}(U)$, where $X \sim$ EW-E$(\alpha, \lambda, p, \beta)$
3. Simulate EW-E observations for fixed $n \in [50, 100, 150]$ by means of the previous scheme.
4. Consider three scenarios: $\alpha = 1.5$, $\lambda = 1$, $p = 2$ and $\beta = 1.5$ (Setup 1); $\alpha = 2$, $\lambda = 1$, $p = 1.5$ and $\beta = 2.5$ (Setup 2); and $\alpha = 1.7$, $\lambda = 1$, $p = 1.5$ and $\beta = 2$ (Setup 3).
5. Calculate the ML estimates from each generated data set, and then obtain the averages, biases and mean square errors (MSEs).

Table 2 reports the findings on step 5. The averages of the estimates approach to the true parameters and the biases decrease (for all scenarios) when n increases.

5. Numerical applications

5.1 Context

Here, the importance of the EW-G family is illustrated in two applications to real data sets. The gamma, Birnbaum-Saunders, Weibull and log-logistic distributions are taken as baselines in the proposed family to prove their flexibility. The data sets are obtained from database of the Severe Acute Respiratory Syndrome, available from the platform of the Ministry of Health linked to the Brazilian Open Data Portal[1], which comprise events from 2020-2021 and passed for filter process to obtain the COVID-19 patients.

The `MASS` package in the `R` language is used with a heuristic method to obtain the initial parameter values to maximize the log-likelihood. Further, the `GenSA`, `MASS` and `AdequacyModel` libraries and `goodness.fit()` function with the SANN method of the `R` software are used in the others computations. The data sets and application codes can be accessed at https://github.com/elisangelacbiazatti/EW-G.

The competing distributions for comparison are listed below:

- Beta-Birnbaum-Saunders (B-BS) (Cordeiro and Lemonte, 2011);
- Kumaraswamy-Birnbaum-Saunders (Kw-BS) (Saulo et al, 2012);
- Beta-gamma (B-Ga) (Kong et al, 2007);
- Kumaraswamy-gamma (Kw-Ga) (Cordeiro and de Castro, 2011);
- Kumaraswamy-Weibull (Kw-W) (Cordeiro et al, 2010);
- Kumaraswamy-log-logistic (Kw-LL) (Santana et al, 2012);
- Beta Weibull distribution (B-W) (Lee et al, 2007);
- Beta log-logistic (B-LL) (Lemonte, 2014).

Some descriptive statistics for the data sets are provided, the ML estimators and their SEs of the fitted distributions, and the adequacy statistics to compare them. The data sets used, and fitted models are described below.

[1]BD-SRAG. Available in: https://opendatasus.saude.gov.br/dataset/srag-2020. Accessed on: August 23, 2021.

Table 2. Simulation Results

Setup	Sample size	Parameter	Average	Bias	MSE
Setup 1	n=50	α	1.53766	0.03766	0.05809
		λ	1.00843	0.00843	0.00079
		p	2.03546	0.03546	0.05231
		β	1.50792	0.00792	0.02785
	$n = 100$	α	1.51719	0.01719	0.03487
		λ	1.00673	0.00673	0.00007
		p	2.01971	0.01971	0.02636
		β	1.50566	0.00566	0.02762
	$n = 150$	α	1.51414	0.01414	0.02996
		λ	1.00667	0.00667	0.00004
		p	2.01489	0.01489	0.01596
		β	1.50267	0.00267	0.02759
Setup 2	n=50	α	2.04752	0.04752	0.09619
		λ	1.01065	0.01065	0.00211
		p	1.51895	0.01895	0.03352
		β	2.51535	0.01535	0.04084
	$n = 100$	α	2.02632	0.02632	0.05089
		λ	1.00699	0.00699	0.00016
		p	1.50812	0.00812	0.02816
		β	2.50888	0.00888	0.02976
	$n = 150$	α	2.01862	0.01862	0.03554
		λ	1.00670	0.00669	0.00006
		p	1.50454	0.00453	0.02768
		β	2.50459	0.00459	0.02796
Setup 3	$n = 50$	α	1.74158	0.04158	0.07357
		λ	1.01172	0.01172	0.00300
		p	1.51982	0.01982	0.03445
		β	2.01958	0.01958	0.03028
	$n = 100$	α	1.72463	0.02463	0.03744
		λ	1.00733	0.00733	0.00032
		p	1.50988	0.00988	0.02839
		β	2.01207	0.01207	0.01339
	$n = 150$	α	1.71643	0.01643	0.02589
		λ	1.00677	0.00677	0.00007
		p	1.50527	0.00527	0.02771
		β	2.01064	0.01064	0.00648

5.2 COVID-19 data in Goiânia, Brazil

The first application refers to the times (in days) of 1105 COVID-19 patients from the date of entry in the Intensive Care Unit (ICU) until death in Goiânia, considering the municipal code 520870, in the Goiás state. By COVID-19 patient, we mean someone who has a positive RT-PCR test. The descriptive statistics for the time until death for COVID-19 data in Goiânia include: mean = 13.18, standard deviation (SD) = 12.31, skewness = 2.58, kurtosis = 14.50, and minimum and maximum values 1 and 126, respectively.

To investigate which distribution fits better to these data, formal goodness-of-fit tests are used. The values of the W^* and A^* statistics (from `AdequacyModel` package) are reported in Table 3. The EW-G (for $\lambda = 1$) is better than the Kw-G and B-G classes for both Birnbaum-Saunders and gamma baselines. By verifying the measures of these formal goodness-of-fit tests in Table 3, we conclude that the EW-BS model outperforms all other fitted distributions. Thus, the two special models in the new generator provide better fits than the distributions belonging to the B-G and Kw-G classes and then can be an efficient alternative to these distributions for modeling data sets.

Table 3. Estimation results for COVID-19 data in Goiânia.

Distribution	ML estimates and SEs				W^*	A^*
EW-Ga	α 1.18685 (0.26250)	p 0.73965 (0.04488)	a 1.99341 (0.42817)	b 0.18321 (0.02364)	0.20301	1.93795
EW-BS	α 4.92674 (0.77151)	p 0.84445 (0.10890)	a 1.72221 (0.19646)	b 1.40769 (0.28523)	0.23278	2.19773
Kw-Ga	α 15.98273 (0.00013)	β 0.49066 (0.01993)	a 0.12328 (0.00284)	b 0.14244 (0.00338)	0.39119	2.96695
Kw-BS	α 3.60128 (0.47410)	β 0.93352 (0.52569)	a 1.74557 (0.25979)	b 2.06262 (0.77221)	0.44681	3.66681
B-Ga	α 25.19377 (3.235e-07)	β 41.16548 (3.944e-07)	a 0.16931 (1.712e-06)	b 0.00022 (6.040e-06)	0.31399	2.51569
B-BS	α 41.01784 (1.06967)	β 32.73182 (1.18535)	a 7.75890 (0.38043)	b 3.37354 (0.90847)	0.33102	2.86836

The Vuong test (Vuong, 1989) in Table 4 reveals that the EW-BS distribution is better than the EW-Ga, B-Ga, B-BS, Kw-BS and Kw-Ga distributions for a level of significance of 5%.

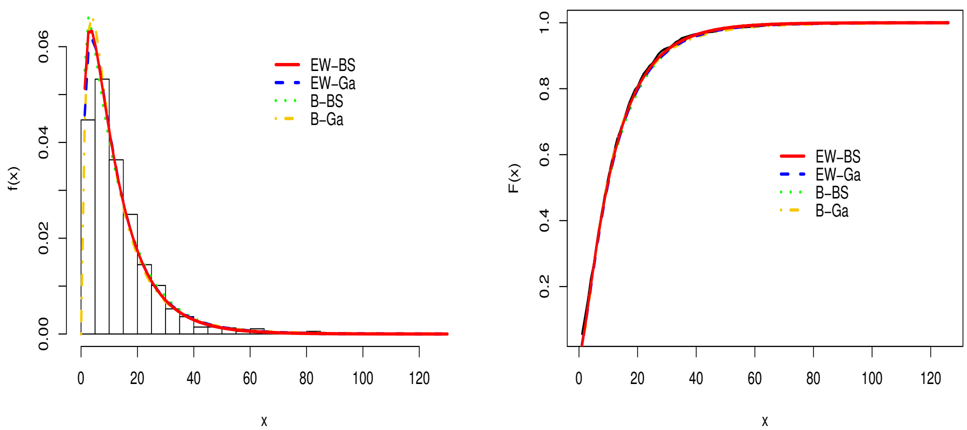
Table 4. The Vuong test for some fitted models to COVID-19 data in Goiânia.

Distribution	Vuong statistic	Decision
EW-Ga × EW-BS	-110.7868	EW-BS is chosen
EW-BS × B-Ga	26.03541	EW-BS is chosen
EW-BS × B-BS	17.24543	EW-BS is chosen
EW-BS × Kw-BS	7.22972	EW-BS is chosen
EW-BS × Kw-Ga	93.83827	EW-BS is chosen

Figure 5 displays the histogram of the data, where x represents the time to death and the fitted EW-BS PDF, some other PDFs, and empirical and estimated CDFs. We note that the EW-BS distribution provides a better fit to the time until death for COVID-19 in Goiânia than those of the other models.

5.3 COVID-19 data in Natal, Brazil

The second application refers to COVID-19 data from the city of Natal, considering the municipal code 240810, in Rio Grande do Norte state. The 35 survival times (in days) of COVID-19 patients from the date of hospitalization until death are: 4, 23, 12, 3, 2, 22, 1, 15, 4, 10, 9, 1, 34, 1, 16, 10, 11, 14, 5, 6, 13, 9, 9, 1, 5, 3, 27, 17, 10, 6, 8, 1, 10, 8, 3. By COVID-19 patient, we mean someone who has a positive RT-PCR test.



(a) (b)

Figure 5. Estimated EW-BS, EW-Ga, B-BS and B-Ga PDFs (a); and empirical and estimated CDFs (b).

The descriptive statistics for these data include: mean = 9.51, SD = 7.84, skewness = 1.27, kurtosis = 4.44, and minimum and maximum values 1 and 34, respectively.

Table 5 reports the ML estimates of the parameters (SEs in parentheses) and the values of the statistics W^* and A^* (from `AdequacyModel` package). In this application, we show that the EW-G (for $\lambda = 1$) is better than the Kw-G and B-G classes for both the log-logistic and Weibull baselines.

Table 5. Findings from the fitted models to COVID-19 data in Natal.

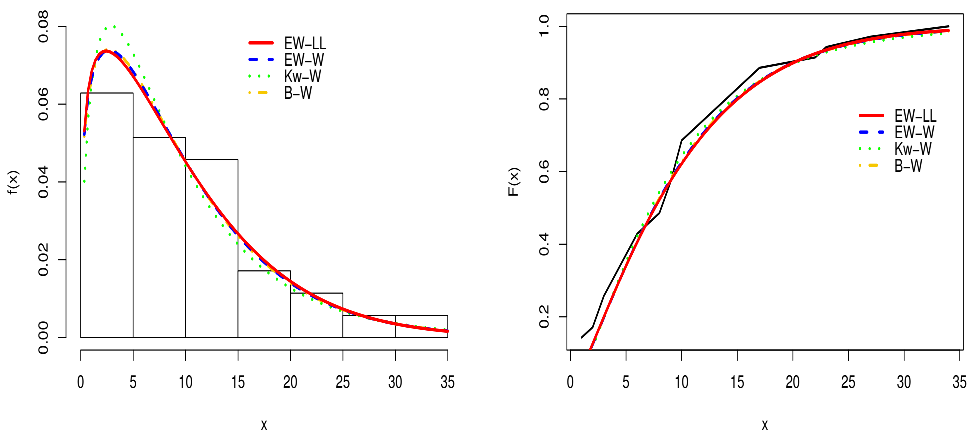
Distribution	ML estimates and SEs				W^*	A^*
	α	p	a	b		
EW-LL	0.59274	5.09809	4.63721	0.49906	0.04283	0.33661
	(0.34082)	(3.45002)	(8.796e-06)	(0.16256)		
	α	p	k	b		
EW-W	1.06493	2.23626	0.53249	9.73429	0.04393	0.34133
	(0.20430)	(0.31642)	(0.00001)	(5.656e-07)		
	α	β	a	b		
Kw-LL	4.94667	50.04956	14.92093	0.43042	0.07575	0.53421
	(7.834e-06)	(0.00004)	(2.50239)	(0.00002)		
	α	β	k	b		
Kw-W	4.85694	19.98867	0.36364	20.03681	0.05570	0.41001
	(0.01187)	(0.00005)	(0.03636)	(0.00004)		
	α	β	a	b		
B-LL	0.61195	3.61587	26.47914	1.86317	0.11272	0.77509
	(0.30367)	(0.17226)	(5.178e-06)	(0.69671)		
	α	β	k	b		
B-W	1.09556	2.71432	1.16828	22.03596	0.06493	0.46725
	(0.23859)	(0.66934)	(0.00514)	(0.00002)		

The Vuong test (Vuong, 1989) in Table 6 reveals that the EW-LL distribution is better than the EW-W, Kw-W, and B-W distributions at a level of significance of 5%.

Figure 6 reports the histogram of the data, where x represents the time to death and the fitted EW-LL PDF and some other PDFs, and empirical and estimated CDFs. We note that the EW-LL distribution gives a better fit to the time until death for COVID-19 in Natal than those other models.

Table 6. The Vuong test for some fitted models to COVID-19 data in Natal.

Distribution	Vuong statistic	Decision
EW-LL × EW-W	55.68025	EW-LL is chosen
EW-LL × Kw-W	7.261374	EW-LL is chosen
EW-LL × B-W	45.67538	EW-LL is chosen



(a) (b)

Figure 6. Estimated EW-LL, EW-W, Kw-W and B-W PDFs (a); and empirical and estimated CDFs (b).

5.4 Regression applied to COVID-19 data in Porto Alegre, Brazil

We present an application of the LEW-W regression to COVID-19 data. The study comprises the time (in days) elapsed from the date of hospitalization until death by the coronavirus of 599 patients in Porto Alegre-RS, considering the municipal code 431490, with all observations failing, that is, censored times are not considered in the study.

The explanatory variables are (for $i = 1, \ldots, 599$):

- v_{i1}: age (in years);
- v_{i2}: icu (1=Admitted to the ICU, 2=Not admitted to the ICU).

The `survival` library, with `optim()` function and the SANN method of the `R` language are used to develop the computational part.

We compare the fits of the Log Exponentiated Weibull-Weibull (LEW-W) given in Equation (4.9) (for $\lambda = 1$) with the log-beta Weibull (LBW) (Ortega et al, 2013) and the log-Kumaraswamy-Weibull (or Kumaraswamy Gumbel) (Kw-Gu) (Cordeiro et al, 2012).

It is important to use a methodology to identify the most appropriate model for the time to death. In this context, the total time on test plot (not shown here) for the data under study shows an increasing appearance for the most part risk function. The descriptive statistics for the time until death for the current data include mean = 22.20, SD = 18.61, skewness = 2.01, kurtosis = 8.85, and minimum and maximum values 1 and 135, respectively. The results from the fitted regression are presented as

$$y_i = \gamma_0 + \gamma_1\, v_{i1} + \gamma_2\, v_{i2} + \sigma\, z_i,$$

where the errors $z_1, \ldots, z_{599}$ are independent random variables with PDF given in Equation (4.10).

Table 7 gives the ML estimates of the parameters for the LEW-W, LBW and Kw-Gu regressions fitted to these data and an asymptotic 95% confidence interval for $\boldsymbol{\theta}$. These results indicate that the LEW-W regression has the lowest AIC, CAIC and BIC values among those of the fitted models, which indicates that LEW-W model provides the best fit to the data. This is also confirmed by the generalized likelihood ratio (GLR) test (Vuong, 1989), for a significance level of 5%. Further, we note from the fitted LEW-W regression that all covariates (v_{i1}: age and v_{i2}: icu) are significant at 1%, and that there is a significant difference between the categories of the explanatory variables.

Table 7. Estimates from the listed regression to COVID-19 data in Porto Alegre, and adequacy measures.

Parameter	Estimate	SE	p-value	95% Lower	95% Upper	AIC	CAIC	BIC
LEW-W								
γ_0	2.66507	0.34182	<0.00001	1.99509	3.33504	1554.873	1555.015	1581.244
γ_1	-0.00817	0.00256	0.00142	-0.01319	-0.00315			
γ_2	-0.44590	0.08424	<0.00001	-0.61102	-0.28079			
σ	1.97677	0.24216		1.50213	2.45142			
α	0.95458	0.16907		0.62321	1.28596			
p	2.95376	0.27963		2.40568	3.50184			
LBW								
γ_0	4.73123	0.39204	<0.00001	3.96283	5.49964	1634.352	1634.494	1660.724
γ_1	-0.00586	0.00261	0.02478	-0.01097	-0.00074			
γ_2	-0.37765	0.08939	0.00002	-0.55286	-0.20243			
σ	1.65511	0.10378		1.45171	1.85851			
a	0.86137	0.19776		0.47376	1.24898			
b	4.22907	0.85459		2.55406	5.90408			
Kw-Gu								
γ_0	3.37041	0.39806	<0.00001	2.59021	4.15061	1659.315	1659.457	1685.687
γ_1	-0.00891	0.00259	0.00059	-0.01399	-0.00383			
γ_2	-0.54654	0.09185	<0.00001	-0.72657	-0.36651			
σ	1.53454	0.07806		1.38154	1.68753			
a	2.40749	0.55253		1.32454	3.49046			
b	3.12223	0.41468		2.30945	3.93501			

GLR = 32.639 (LEW-W × LBW)
GLR = 26.599 (LEW-W × Kw-Gu)

The plots of the empirical survival function (estimated by Kaplan-Meier) and the estimated survival function by fitting the LEW-W model (reparameterized) are displayed in Figure 7 which reveal that this regression provides a good fit for the current data. Verify if there are observations influencing the model adjustment is an important aspect. The Cook distance is presented to investigate this aspect.

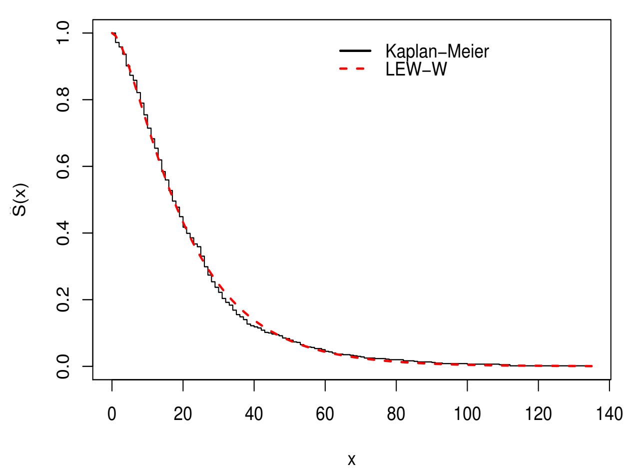


Figure 7. Empirical and estimated survival functions of the LEW-W model.

Next, we conduct a sensitivity analysis and influential observations. The observations #133 and #396 are the ones that stand out the most, as shown in Figure 8, under the generalized Cook distance (Cook, 1977), so indicating that these can be taken as possible influential observations.

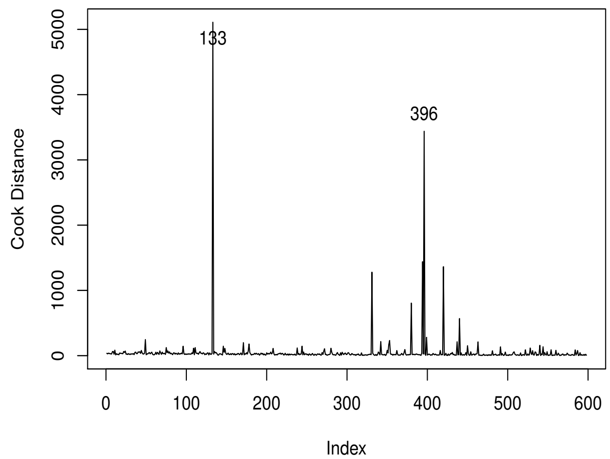


Figure 8. Cook distance for COVID-19 data in Porto Alegre.

The observation #133 refers to an individual aged 57 years, who was admitted to the ICU and had a length of stay until death of 10 days. And the observation #396 refers to the individual aged 67 years, who was admitted to the ICU and whose length of stay until death was 28 days. We can note that the two individuals were admitted to the ICU and are not very elderly patients.

To better investigate whether these individuals are influencing the model estimated, the parameters are estimated from a sub-sample of the original sample, which was selected withdrawing individually when a group the observations is considered influential. Relative change (RC), parameter estimates, and their significance are adopted to verify the impact of withdrawing possible observations from the analysis. Table 8 reports these findings.

By analyzing Table 8, it is noted that the parameter estimates do not suffer major changes and that the significance of the coefficients is maintained at 5%, which indicates that the LEW-W regression model is robust in this application, that is, there are no influential observations in this study.

Table 8. RC, parameter estimates and corresponding (p-value) for COVID-19 data in Porto Alegre.

Sub-sample	α	p	σ	γ_0	γ_1	γ_2
I - [Full model]	[-] 0.95458 (-)	[-] 2.95376 (-)	[-] 1.97677 (-)	[-] 2.66507 (<0.00001)	[-] -0.00817 (0.00142)	[-] -0.44590 (<0.00001)
I - [133]	[0.43177] 0.54242 (-)	[-0.6876] 3.15686 (-)	[0.14160] 1.69685 (-)	[-0.28043] 3.41243 (<0.00001)	[-0.01102] -0.00826 (0.00119)	[-0.27964] -0.57059 (<0.00001)
I - [396]	[0.11234] 0.84734 (-)	[0.01120] 2.92068 (-)	[0.07591] 1.82672 (-)	[-0.01228] 2.69779 (<0.00001)	[0.25214] -0.00611 (0.01415)	[0.10357] -0.39972 (<0.00001)
I - [133, 396]	[0.49280] 0.48416 (-)	[0.04378] 2.82443 (-)	[0.22078] 1.54033 (-)	[-0.30761] 3.48486 (<0.00001)	[-0.36475] -0.01115 (0.00003)	[0.35885] -0.28589 (0.00091)

6. Conclusions and future research

We proposed the new exponentiated-Weibull-G family which can generalize all classical continuous distributions. The maximum likelihood method was used to estimate the parameters, and the consistency of the estimators was accessed from a simulation study. The flexibility of the new family was illustrated by means of two real COVID-19 data sets. We showed that the new log-exponentiated Weibull-Weibull (for $\lambda = 1$) regression outperformed regressions based on well-known Kumaraswamy-G and beta-G generators. After verifying the good fit of the new regression, a sensitivity analysis was done to show that the proposed Log Exponentiated Weibull-Weibull regression model is robust in this application. In future works, for example, we can investigate other methods of sensitivity analysis, such as the local influence, and carry out a residual analysis for the new regression model.

Appendix A - Linear representation

Expansion (3.7) derived in the Appendix converges, since it was obtained by taking the term by term derivative of the power series expansion for the CDF $F(x)$ in Equation (5) which converges when $G(x) < 1$. In turn, the CDF is convergent because it is obtained by applying the following power series that guarantees convergence when $G(x) < 1$: the generalized Binomial Theorem, Proposition 2 of Castellares and Lemonte and Theorem 4.1 of Munir (2013) given below.

From the Proposition 2 of Castellares and Lemonte (2015), we can write

$$(-\ln(1-z))^{\delta} = z^{\delta} \sum_{m=0}^{\infty} \rho_m(\delta)\, z^m, \tag{1}$$

where $\delta \in \mathrm{R}$, $|z| < 1$, $\rho_0(\delta) = 1$, $\rho_m(\delta) = \delta\, \psi_{m-1}(m + \delta - 1)$ (for $m \geq 1$), and $\psi_0(m) = 1/2$, $\psi_1(m) = (2 + 3m)/24$, $\psi_2(m) = (m + m^2)/48$ and so on, are the Stirling polynomials.

Theorem 4.1 (Munir, 2013) If g is an analytic function on the open ball $B(0, R)$ with power series representation about the origin given as $g(z) = \sum_{n=0}^{\infty} b_n z^n$, then for all $z \in B(0, R)$, we have $\exp(g(z)) = \sum_{n=0}^{\infty} a_n\, z^n$, where

$$a_n = \begin{cases} \exp(b_0), & n = 0 \\ n^{-1} \sum_{k=1}^{n} k\ b_k\, a_{n-k}, & n \geq 1. \end{cases}$$

If c is a real number,and $\binom{c}{i}$ is the generalized binomial coefficient, the power series holds

$$(1 - z)^c = \sum_{i=0}^{\infty} (-1)^i \binom{c}{i} z^i, \quad |z| < 1. \tag{2}$$

Then, we can write from Equations (2.3) and (2) that

$$F(x) = 1 + \sum_{i=1}^{\infty} (-1)^i \binom{\alpha}{i} \exp\left(-i\, \lambda^{-p} \left(-\ln(1 - G(x))\right)^p\right). \tag{3}$$

Further, after simple algebra from Equations (1) and (2), we have $(-\ln(1-G(x)))^p = \sum_{k,m=0}^{\infty} A_{k,m}(p)\, G(x)^{k+m}$, where (for $k,m = 0,1,\ldots$)

$$A_{k,m}(p) = \sum_{j=k}^{\infty} (-1)^{j+k} \binom{p}{j} \binom{j}{k} \rho_m(p).$$

This double power series can be converted into a simple power series

$$(-\ln(1-G(x)))^p = \sum_{n=0}^{\infty} b_n(p)\, G(x)^n,$$

where (for $n \geq 0$) $b_n(p) = \sum_{l=0}^{n} A_{l,n-l}(p)$. Hence, from Theorem 4.1, we can write (for $i \geq 1$)

$$\exp\Big(-i\lambda^{-p}(-\ln(1-z))^p\Big) = \exp\left(\sum_{n=0}^{\infty} \frac{i\, b_n(p)}{\lambda^p}\, z^n\right) = \sum_{n=0}^{\infty} a_{n,i}(p,\lambda)\, z^n\,, \qquad (4)$$

where

$$a_{0,i} = \exp\left(\frac{i b_0(p)}{\lambda^p}\right) = \exp\left(\frac{i}{\lambda^p} \sum_{j=0}^{\infty} (-1)^j \binom{p}{j}\right),$$

$$a_{n,i} = \frac{i}{n\lambda^p} \sum_{k=1}^{n} k\, b_k(p)\, a_{n-k,i}, \quad n \geq 1.$$

Theferore, Equations (3) and (4) imply that

$$F(x) = 1 + \sum_{n=1}^{\infty} c_n\; G(x)^n, \qquad (5)$$

where

$$c_n = c_n(p,\lambda,\alpha) = \sum_{i=0}^{\infty} (-1)^i \binom{\alpha}{i} a_{n,i}.$$

By differentiating Equation (5) and changing indices, we obtain

$$f(x) = \sum_{r=0}^{\infty} c_{r+1}\, \pi_{r+1}(x),$$

where $\pi_{r+1}(x) = (r+1)\; g(x)\, G(x)^r$.

Appendix B - Elements of the score vector $U(\alpha, \lambda, p, \boldsymbol{\theta})$

The components of the score vector $U(\alpha, \lambda, p, \boldsymbol{\theta})$ are given by

$$\frac{\mathrm{d}l}{\mathrm{d}\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \ln\left(1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)\right)$$

$$\frac{\mathrm{d}l}{\mathrm{d}\lambda} = -\frac{np}{\lambda} + \sum_{i=1}^{n} \ln\left(-\ln(1 - G(x_i))\right) + \frac{p}{\lambda^{p+1}} \sum_{i=1}^{n} \left(-\ln(1 - G(x_i))\right)^p$$

$$-(\alpha - 1)\frac{p}{\lambda^{p+1}} \sum_{i=1}^{n} \frac{\exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)\left(-\ln(1 - G(x_i))\right)^p}{1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)} \frac{\mathrm{d}l}{\mathrm{dp}}$$

$$= \frac{n}{p} + n\ln\lambda + \sum_{i=1}^{n} \ln\left(-\ln(1 - G(x_i))\right) - \sum_{i=1}^{n} \left(\frac{-\ln(1 - G(x_i))}{\lambda}\right)^p \ln\left(\frac{-\ln(1 - G(x_i))}{\lambda}\right)$$

$$+(\alpha - 1)\sum_{i=1}^{n} \frac{\exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)}{1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)}$$

$$\times \left(\frac{-\ln(1 - G(x_i))}{\lambda}\right)^p \ln\left(\frac{-\ln(1 - G(x_i))}{\lambda}\right)$$

$$\frac{\mathrm{d}l}{\mathrm{d}\boldsymbol{\theta}_{\mathrm{j}}} = \sum_{i=1}^{n} \frac{1}{g(x_i)} \partial_{\boldsymbol{\theta}_j} g(x_i) - \sum_{i=1}^{n} \frac{1}{1 - G(x_i)} \partial_{\boldsymbol{\theta}_j} G(x_i) + (p - 1)\sum_{i=1}^{n} \frac{1}{\ln(1 - G(x_i))} \frac{1}{1 - G(x_i)} \partial_{\boldsymbol{\theta}_j} G(x_i)$$

$$-\frac{p}{\lambda^p} \sum_{i=1}^{n} \left(-\ln(1 - G(x_i))\right)^{p-1} \frac{1}{1 - G(x_i)} \partial_{\boldsymbol{\theta}_j} G(x_i)$$

$$+\frac{p\,(\alpha - 1)}{\lambda^p} \sum_{i=1}^{n} \frac{\exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)}{1 - \exp\left(-\lambda^{-p}\left(-\ln\left(1 - G(x_i)\right)\right)^p\right)}$$

$$\times \left(-\ln(1 - G(x_i))\right)^{p-1} \frac{1}{1 - G(x_i)} \partial_{\boldsymbol{\theta}_j} G(x_i).$$

Authors' contributions Conceptualization, G.M.C., E.C.B., M.C.S.L., L.H.S.; data curation, G.M.C., E.C.B., M.C.S.L.; formal analysis, G.M.C., E.C.B., M.C.S.L.; investigation, G.M.C., E.C.B.; methodology, G.M.C., E.C.B., L.H.S.; software, E.C.B., M.C.S.L.; supervision, G.M.C. validation, G.M.C., E.C.B., M.C.S.L., L.H.S.; visualization, G.M.C., E.C.B., M.C.S.L., L.H.S.; writing—original draft preparation, G.M.C., E.C.B., M.C.S.L., L.H.S.; writing—review and editing, G.M.C., E.C.B., M.C.S.L., L.H.S. All authors have read and agreed to the published version of the article.

Acknowledgements The authors thank to the editors, associate editor, and the anonymous referees for their valuable comments to improve this article.

Funding This research received no external funding.

Conflicts of interest The authors declare no conflict of interest.

References

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

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

Chen, G, and Balakrishnan, N., 1995. A general purpose approximate goodness-of-fit test. Journal of Quality Technology, 27, 154–161.

Cook, R.D., 1977. Detection of influential observation in linear regression. Technometrics, 19, 15–18.

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

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

Cordeiro, G.M., Nadarajah, S., and Ortega, E.M.M., 2012. The Kumaraswamy Gumbel distribution. Statistical Methods and Applications, 21, 139–168.

Cordeiro, G.M., Ortega, E.M.M., and Nadarajah, S., 2010. The Kumaraswamy Weibull distribution with application to failure data. Journal of the Franklin Institute, 347, 1399–1429.

Cordeiro, G.M., Ortega, E.M.M., and da Cunha, D.C.C., 2013. The exponentiated generalized class of distributions. Journal of Data Science, 11, 1–27.

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

Fernández, L.M.Z. and de Andrade, T.A.N., 2020. The erf-G family: New unconditioned and log-linear regression models. Chilean Journal of Statistics, 11, 3–23.

Gómez-Déniz, E., Calderón-Ojeda, E., and Sarabia, J.M., 2022. The arctan family of distributions: New results with applications. Chilean Journal of Statistics, 13, 113–132.

Ghosh, I. and Alzaatreh, A., 2018. A new class of generalized logistic distribution. Communications in Statistics: Theory and Methods, 47, 2043–2055.

Gupta, R.C., Gupta, P.L., and Gupta, R.D., 1998. Modeling failure time data by Lehman alternatives. Communications in Statistics: Theory and Methods, 27, 887–904.

Kenney, J.F. and Keeping, E.S., 1961. Mathematics of Statistics. D. Van Nostrand Company, 1, 429.

Kong, L., Lee, C., and Sepanski, J., 2007. On the properties of beta-gamma distribution. Journal of Modern Applied Statistical Methods, 6, 187–211.

Lee, C., Famoye, F., and Olumolade, O., 2007. Beta-Weibull distribution: Some properties and applications to censored data. Journal of Modern Applied Statistical Methods, 6, 173–186.

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

Marinho, P.R.D., Silva, R.B., Bourguignon, M., Cordeiro, G. M., and Nadarajah, S., 2019. AdequacyModel: An R package for probability distributions and general purpose optimization. PLoS ONE, 14, e0221487.

Moors, J., 1988. A quantile alternative for kurtosis. Journal of the Royal Statistical Society D, 37, 25–32.

Mudholkar, G. and Srivastava, D., 1993. Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Transactions on Reliability, 42, 299–302.

Munir, A.U.K., 2013. Series Representations and Approximation of some Quantile Functions appearing in Finance. University College London, 1, 164.

Nadarajah, S. and Kotz, S., 2003. The exponentiated Fréchet distribution. Interstat Electronic Journal, 14, 1–7.

Nadarajah, S., 2005. The exponentiated Gumbel distribution with climate application. Environmetrics, 17, 13–23.

Nadarajah, S. and Gupta, A.K., 2007. The exponentiated gamma distribution with application to drought data. Calcutta Statistical Association Bulletin, 59, 29–54.

Ortega, E.M., Cordeiro, G.M., and Kattan, M.W., 2013. The log-beta Weibull regression model with application to predict recurrence of prostate cancer. Statistical Papers, 54, 113–132.

Saulo, H., Leão, J., and Bourguignon, M., 2012. The Kumaraswamy Birnbaum-Saunders distribution. Journal of Statistical Theory and Practice, 6, 745–759.

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

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57, 307–333.