CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

Volume 11 Number 1
April 2020
ISSN: 0718-7912 (print)
ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society
SOCHE
SOCIEDAD CHILENA DE ESTADÍSTICA

Aims

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (`www.soche.cl`). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS is an international scientific forum strongly committed to gender equality, open access of publications and data, and the new era of information. The ChJS covers a broad range of topics in statistics, data science, data mining, artificial intelligence, and big data, including research, survey and teaching articles, reviews, and material for statistical discussion. In particular, the ChJS considers timely articles organized into the following sections: Theory and methods, computation, simulation, applications and case studies, education and teaching, development, evaluation, review, and validation of statistical software and algorithms, review articles, letters to the editors.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

Contents

Chilean Journal of Statistics
Vol. 11, No. 1, April 2020, 3–23

Statistical Modeling
Research Paper

The erf-G family: new unconditioned and log-linear regression models

Luz Milena Zea Fernández[1] and Thiago A. N. de Andrade[2,*]

[1]Department of Statistics, Federal University of Rio Grande do Norte, Natal, Brazil,
[2]Department of Statistics, Federal University of Pernambuco, Recife, Brazil,

(Received: 08 March 2019 · Accepted in final form: 04 June 2019)

Abstract

In this paper, we propose a new generator of distributions called the erf-G family. Our proposal provides special distributions without adding complexity to parametric spaces of resulting models. We also furnish empirical evidence that the proposed family may solve issues of flat or quasi-red likelihoods in some baselines. In particular, we detail six special models from the erf-G family. We also derive a new log-linear regression model considering a kind of censoring. We discuss censored and uncensored maximum likelihood estimation methods for the proposed models. In order to study asymptotic properties of considered estimators, we carry out a Monte Carlo simulation study. Finally, using applications to real data we illustrate that proposed models may outperform classic lifetime models.

Keywords: Error function · Flat likelihood · Generalized distributions · Log-linear regression models.

Mathematics Subject Classification: Primary 60E10 · Secondary 60E05.

1. Introduction

From both theoretical and applied perspectives, the proposal of new probability distributions is crucial to describe natural phenomena. There are several ways to extend well-known distributions. One of the most popular ways is to consider distribution generators. Some of them are: Marshall-Olkin (Marshall and Olkin, 1997), beta (Eugene et al., 2002), gamma (Zografos and Balakrishnan, 2009), (Ristic and Balakrishnan, 2012) and (Nadarajah et al., 2015), Kumaraswamy (Cordeiro and de Castro, 2011), exponentiated generalized (Cordeiro et al., 2013), red odd exponentiated half-logistic (Afify et al., 2017) classes of models, among others.

Several generators (beyond of these referred above) have provided models more flexible than classic ones, used widely in applications into the lifetime context. However, from a literature review, such generators have the disadvantage of adding complexity to

*Corresponding author. Email: thiagoan.andrade@gmail.com

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

http://www.soche.cl/chjs

the parametric space of resulting models. In this paper, we use the error function (erf) as a way to outperform this issue. The erf (also known as Gauss error function) is an important special function, that appear often as solutions from several mathematical and physical problems. Its applications include probability theory, statistics, mass and momentum transfer, branches of mathematical physics, partial differential equations describing diffusion process, among others. For more details, we refer to Chevillard (2012).

We propose and study the erf-G family in details. Some of erf-G special cases are introduced and discussed. We derive explicit expressions for some of its mathematical properties and also propose a log-linear regression (llr) model with log-erf$G$ response variables. A discussion about estimation and hypothesis inference is furnished for both proposed unconditioned and llr models. Simulations results and two applications to real data indicates that our proposals may outperform well-defined lifetime models. We also highlight that our study of the erf-G model has very clear and forceful motivations: (i) it does not impose more complex parametric spaces to resulting models; (ii) it may provide concavity to distributions with flat or quasi-flat likelihoods (details are explored in Section 3); and (iii) it can generate bathtub failure rate functions. For the reasons listed above, we strongly believe it is important to study in detail the erf-G distribution. We hope that this new distribution is part of the arsenal of applied researchers and will be used in many practical situations.

This paper is organized as follows. In Section 2, we define some erf-G special models. Inferential tools, including: (i) linear representations for the erf-G probability density function (PDF) and cumulative distribution function (CDF), (ii) estimation and hypotheses inference procedures and (iii) regression models, are provided in Section 3. Mathematical properties of the new family are presented in Section 4. Simulations and applications to real data are provided in Section 5. In Section 6, main conclusions are listed.

2. Genesis of the new model and some of its special models

In this Section, we present the design of the new model and some of its many special models.

2.1 General context

First we consider the traditional error function given by

$$\mathrm{erf}(z) = \frac{1}{\sqrt{\pi}} \int_{-z}^{z} \exp(-t^2)\,\mathrm{d}t = \frac{2}{\sqrt{\pi}} \int_{0}^{z} \exp(-t^2)\,\mathrm{d}t, \quad z \in \mathrm{R}. \tag{1}$$

From now on, we advocate that replacing z in (1) by $G(x)/[1-G(x)]$ for $x \in \mathcal{X} \subset \mathrm{R}$ collapses a new and efficient generator of distributions. Let $G(x)$ be a cumulative distribution function (CDF). The following operator may be considered as the CDF of a potential family of models:

$$F(x) = \mathrm{erf}\left[\frac{G(x)}{1-G(x)}\right],\ x \in \mathcal{X}. \tag{2}$$

We denote this case as the erf-G family. A stochastic conception of this class which may furnish insight about the relation between new erf-G models and their respective baselines (with CDF G) is given by the following theorem.

Theorem *Let $Z > 0$ be a random variable with CDF given by $F_Z(z) = \mathrm{erf}(z)\,\mathrm{I}_{(0,\infty)}(z)$. Thus, $X = G^{-1}\big[Z\,(1+Z)^{-1}\big]$ is a stochastic transformation having CDF*

$$F_X(x) \;=\; \mathrm{erf}\left[\frac{G(x)}{1\,-\,G(x)}\right],$$

where $G(x)$ represents the CDF of a baseline distribution.

The proof of this theorem holds from the basic probability manipulations. It reveals that distributions into the new family can understood as a quantile of $Y \sim G$ associated with a mapping $\mathcal{X} \to (0,1)$.

Now, let $X \sim$ erf-$G(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathrm{R}^p$, where $\boldsymbol{\Theta}$ represents the parametric space. The PDF of X and hazard rate function (HRF) are given respectively by

$$f(x) = \frac{2g(x;\boldsymbol{\theta})\exp\left[-\left(\dfrac{G(x;\boldsymbol{\theta})}{1-G(x;\boldsymbol{\theta})}\right)^2\right]}{\sqrt{\pi}(1-G(x;\boldsymbol{\theta}))^2}, \quad x \in \mathrm{R}. \tag{3}$$

and

$$h(x) = \frac{2g(x;\boldsymbol{\theta})\exp\left[-\left(\dfrac{G(x;\boldsymbol{\theta})}{1-G(x;\boldsymbol{\theta})}\right)^2\right]}{\sqrt{\pi}(1-G(x;\boldsymbol{\theta}))^2\left\{1-\mathrm{erf}\left[\frac{G(x;\boldsymbol{\theta})}{1-G(x;\boldsymbol{\theta})}\right]\right\}}, \quad x \in \mathrm{R}.$$

2.2 Some special models

The erf-G model is completely new. There is, therefore, a great variety of new distributions, based on (2), that can be explored by statisticians and applied researchers. In what follows, we discuss some special models.

2.2.1 The erf-Gumbel model

The Gumbel distribution is a statistical model defined in real support widely used in engineering problems (de Andrade et al., 2015). Its CDF is given by $G(x;\mu,\sigma) = \exp\left\{-\exp\left[-(x-\mu)/\sigma\right]\right\}$, where $-\infty < \mu < \infty$ and $\sigma > 0$ are the location and scale parameters, respectively. Applying its CDF and PDF in (2) and (3), we obtain the erf-Gumbel (erfGum) model, having CDF and PDF given by

$$F(x) = \mathrm{erf}\left\{\frac{1}{\exp[z_1(x)]-1}\right\}, \quad x \in \mathrm{R},$$

and

$$f(x) = \frac{2z_1(x)\exp\left\{-z_1(x) - \left[\dfrac{1}{\exp[z_1(x)]-1}\right]^2\right\}}{\sqrt{\pi}\sigma\{1-\exp[z_1(x)]\}^2}, \quad x \in \mathrm{R},$$

respectively, where $z_1(x) = \exp\left[-(x-\mu)/\sigma\right]$. Figure 1 presents erfGum PDF curves for some selected parameters. The Gumbel distribution is asymmetric. As we can see in the Figure 1, the erfGum model can accommodate asymmetric shapes.

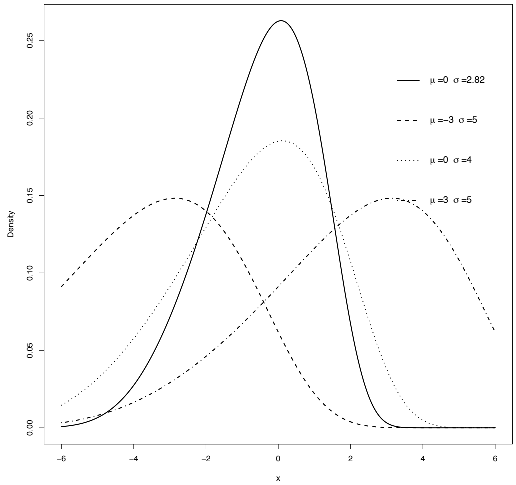


Figure 1. The PDF of the erfGumbel model for some σ and μ parameter values.

2.2.2 The erf-normal model

Let ϕ and Φ be the PDF and CDF of the standard normal model, respectively. Evaluating these equation in (2) and (3), we obtain the erf-normal (erfN) model, with CDF and PDF expressed by

$$F(x) = \operatorname{erf}\left[\frac{\Phi(z_2(x))}{\Phi(-z_2(x))}\right], \quad x \in \mathrm{R},$$

and

$$f(x) = \frac{\sqrt{2}\exp\big\{-z_2(x)^2/2 - [\Phi(z_2(x))/\Phi(-z_2(x))]^2\big\}}{\pi[\Phi(-z_2(x))]^2}, \quad x \in \mathrm{R},$$

where $z_2(x) = (x-\mu)/\sigma$. Plots for the erfN PDF at selected parameter values are displayed in Figure 2. Based on Figure 2, likewise that the erfGum, the erfN distribution may present asymmetrical behaviour in contrast with its baseline.

2.2.3 The erf-gamma model

As third special model, applying gamma model (having shape α and scale β) CDF and PDF in (2) and (3), we get the erf-gamma (erfΓ) model with CDF and PDF expressed as

$$F(x) = \operatorname{erf}\left[\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha,\beta x)}\right], \quad x > 0,$$

where $\Gamma(s,x) = \int_x^\infty \mathrm{t}^{\mathrm{s}-1}\exp(-\mathrm{t})\,\mathrm{dt}$ and $\gamma(s,x) = \int_0^x \mathrm{t}^{\mathrm{s}-1}\exp(-\mathrm{t})\,\mathrm{dt}$ are the upper and lower incomplete gamma functions, and

$$f(x) = \frac{2\beta^\alpha\Gamma(\alpha)x^{\alpha-1}\exp\left[-\left(\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha,\beta x)}\right)^2 - \beta x\right]}{\sqrt{\pi}\,[\Gamma(\alpha,\beta x)]^2}, \quad x > 0,$$

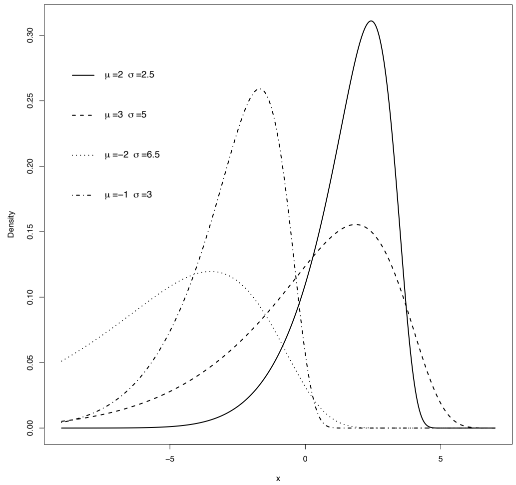
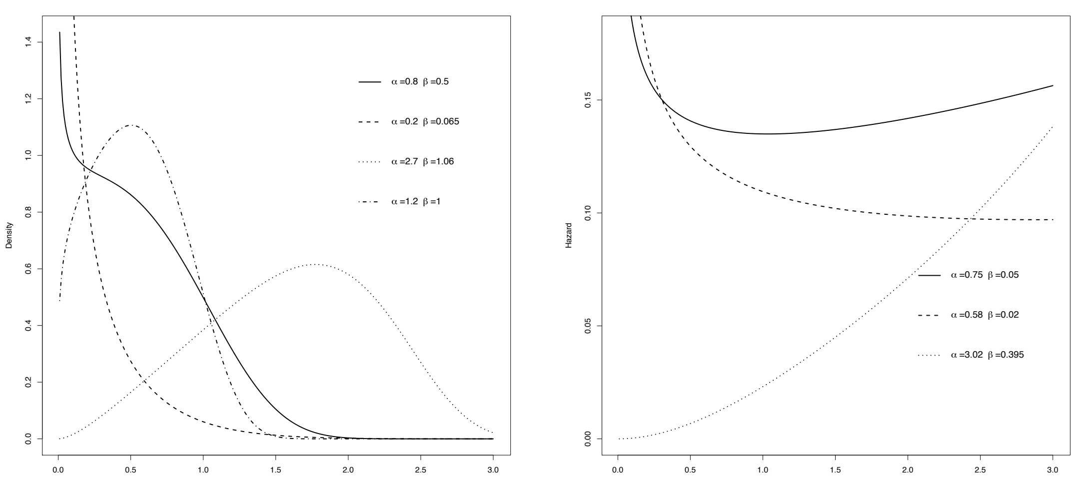


Figure 2. The PDF of the erfN model for some σ and μ parameter values.

where Γ represents the gamma function. The HRF of the erfΓ model is defined by

$$h(x) = \frac{2\beta^{\alpha}\Gamma(\alpha)x^{\alpha-1}\exp\left[-\left(\dfrac{\gamma(\alpha,\beta x)}{\Gamma(\alpha,\beta x)}\right)^2 - \beta x\right]}{\sqrt{\pi}\,[\Gamma(\alpha,\beta x)]^2\left\{1-\operatorname{erf}\left(\dfrac{\gamma(\alpha,\beta x)}{\Gamma(\alpha,\beta x)}\right)\right\}}, \quad x > 0.$$

Plots of the erfΓ PDF and HRF for selected parameter values are presented in Figure 3. At least, the associated HRF can assume bathtub, increasing and decreasing shapes. In contrast with the gamma model, which assumes only monotone HRF shapes.



(a) The PDF of the erfΓ model (b) The HRF of the erfΓ model

Figure 3. The PDF and HRF of the erfΓ model for some α and β parameter values.

2.2.4 The erf-Weibull model

The Weibull distribution can be considered as a standard model for lifetime data and, therefore, is interesting to study a special model generated from it. From evaluating Weibull

CDF and PDF in (2) and (3), we obtain the erf-Weibull (erfW) model, characterized by CDF and PDF given by

$$F(x) = \operatorname{erf}\left[\exp\left(\alpha x^{\beta}\right) - 1\right], \quad x > 0,$$

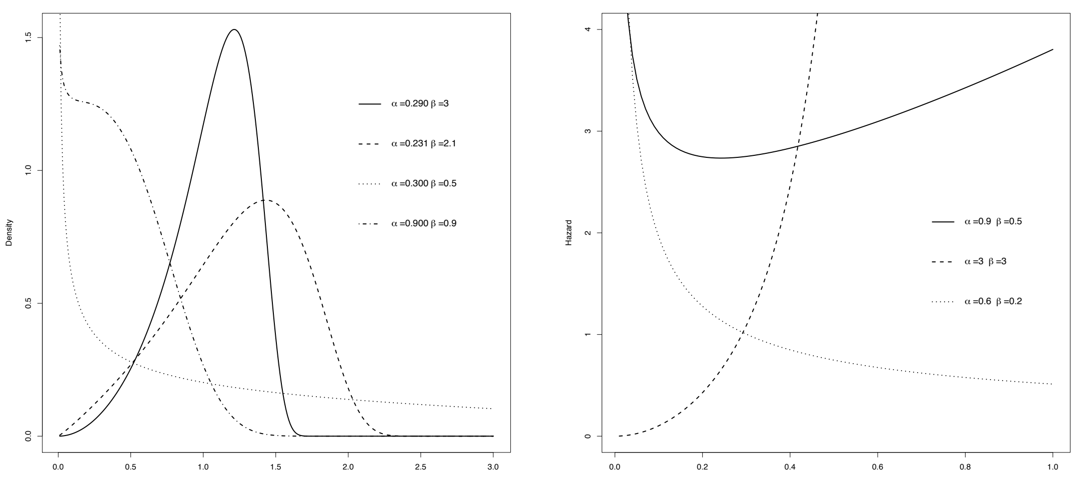
and

$$f(x) = 2\pi^{-1/2}\,\alpha\,\beta\,x^{\beta-1}\exp\left[\alpha x^{\beta} - \left(\exp(\alpha x^{\beta}) - 1\right)^{2}\right], \quad x > 0.$$

The erfW hazard can be expressed as

$$h(x) = \frac{2\,\alpha\,\beta\,x^{\beta-1}\exp\left[\alpha x^{\beta} - \left(\exp(\alpha x^{\beta}) - 1\right)^{2}\right]}{\sqrt{\pi}\,\{1 - \operatorname{erf}[\exp(\alpha x^{\beta}) - 1]\}}, \quad x > 0.$$

Plots of the erfW PDF for selected parameter values are displayed in Figure 4. This figure provides possible shapes of the erfW HRF, which includes the bathtub shape. It represents a gain on the Weibull model, which has constant and monotone shapes.



(a) The PDF of the erfW model (b) The HRF of the erfW model

Figure 4. The PDF and HRF of the erfW model for some α and β parameter values.

2.2.5 The erf-log-logistic distribution

In the survival analysis context, the log-logistic distribution is one of the possible choices when you want to model data with a unimodal failure rate. For $x > 0$, the CDF of the log-logistic model is given by $G(x; \alpha,\, \beta) = 1 - \left[1 + \left(\frac{x}{\alpha}\right)^{\beta}\right]^{-1}$, where $\alpha > 0$ and $\beta > 0$ are shape parameters. Thus, the CDF and PDF regard to the erf-log-logistic (erfLL) distribution are given by

$$F(x) = \operatorname{erf}\left[\left(\frac{x}{\alpha}\right)^{\beta}\right], \quad x > 0,$$

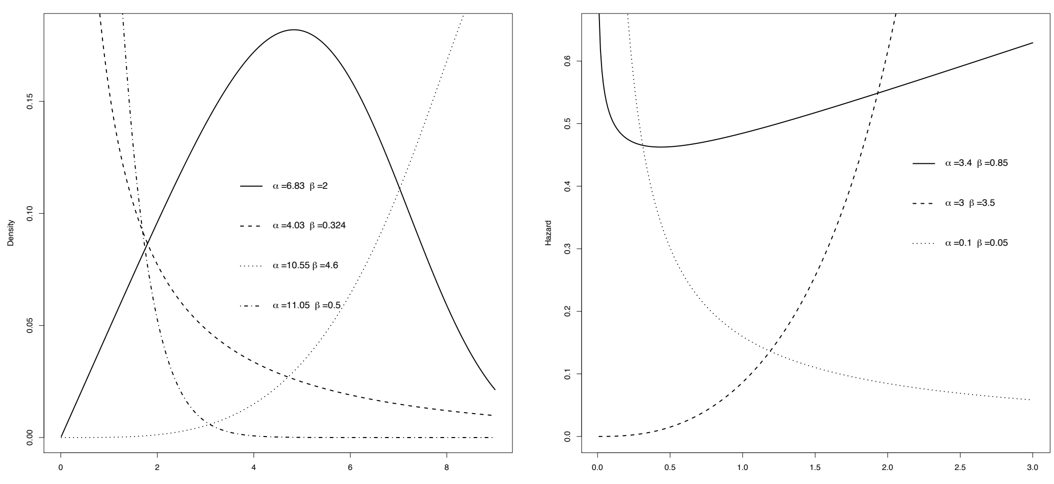
and

$$f(x) = \frac{2\,\beta\,x^{\beta-1}}{\sqrt{\pi}\,\alpha^{\beta}}\,\exp\left[-\left(\frac{x}{\alpha}\right)^{2\beta}\right], \quad x > 0.$$

The HRF of the erfLL distribution is easily defined as

$$h(x) = \frac{2\,\beta\,x^{\beta-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{2\beta}\right]}{\sqrt{\pi}\,\alpha^{\beta}\left\{1-\operatorname{erf}\left[\left(\dfrac{x}{\alpha}\right)^{\beta}\right]\right\}}, \quad x > 0.$$

Plots of the erfLL PDF for selected parameter values are displayed in Figure 5. Figure 5 also provides some possible shapes of the erfLL hazard function for appropriate parameter values, including bathtub, increasing and decreasing shapes. These plots indicate that the erfLL model is fairly flexible and can be used to fit several types of positive data.



(a) The PDF of the erfLL model (b) The HRF of the erfLL model

Figure 5. The PDF and HRF of the erfLL model for some α and β parameter values.

2.2.6 The erf-Frechet distribution

The CDF of the Frechet model is given by $G(x;\delta,\lambda) = \exp(-\delta^{\lambda}x^{-\lambda})$ for $x > 0$ and $\delta, \lambda > 0$. An important generalization based on this distribution was proposed by da Silva et al. (2013). Considering $G(x)$ as the Frechet CDF in equations (2) and (3), we get the erf-Frechet (erfF) model with CDF and PDF expressed as

$$F(x) = \operatorname{erf}\left[\left(\exp(\delta^{\lambda}x^{-\lambda}) - 1\right)^{-1}\right]$$

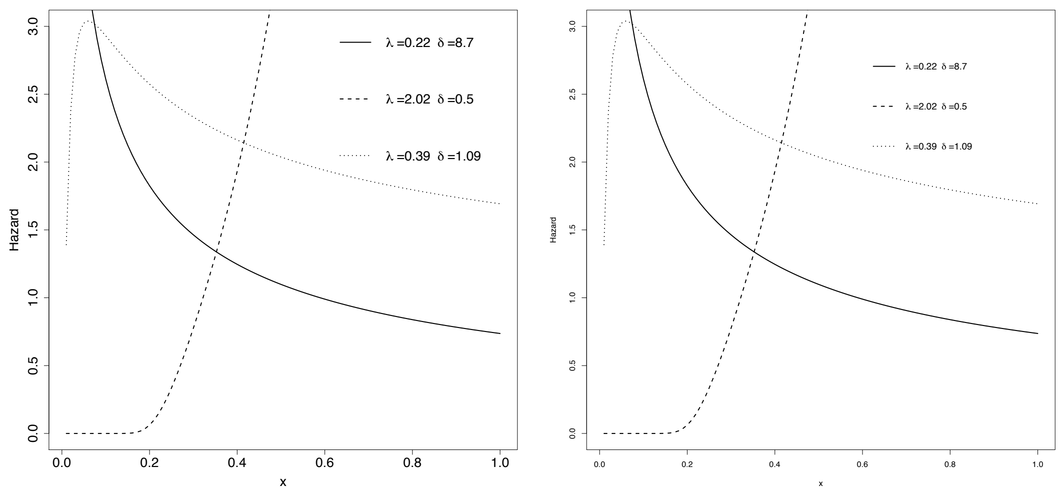
and

$$f(x) = \frac{2\lambda\,\delta^{\lambda}\,x^{-\lambda-1}\exp\left\{-\delta^{\lambda}x^{-\lambda} - \left[\exp(\delta^{\lambda}x^{-\lambda}) - 1\right]^{-2}\right\}}{\sqrt{\pi}\,\left[1-\exp(-\delta^{\lambda}x^{-\lambda})\right]^{2}}. \quad (4)$$

The risk function associated appears as

$$h(x) = \frac{2\lambda\,\delta^{\lambda}\,x^{-\lambda-1}\exp\left\{-\delta^{\lambda}x^{-\lambda} - \left[\exp(\delta^{\lambda}x^{-\lambda}) - 1\right]^{-2}\right\}}{\sqrt{\pi}\,[1-\exp(-\delta^{\lambda}x^{-\lambda})]^{2}\left\{1-\operatorname{erf}\left[(\exp(\delta^{\lambda}x^{-\lambda})-1)^{-1}\right]\right\}}.$$

Some plots for the erfF PDF and HRF are provide in Figure 6. The erfF HRF covers the inverted bathtub shape in contrast with the Frechet HRF, that assumes only monotone behavior.



(a) The PDF of the erfF model (b) The HRF of the erfF model

Figure 6. The PDF and HRF of the erfF model for some λ and δ parameter values.

3. Miscellaneous

In this Section, we provide a complete background for inferential processes.

3.1 A linear expansion

General expressions for the PDF and CDF functions are highly appreciated by applied researchers, as they allow approximate results to be obtained when analytical solutions are not available. Here, we refer to some works that consider these expansions: Cordeiro et al. (2015), Leao et al. (2013), de Andrade et al. (2016) and Afify et al. (2017). This section aims to provide expansions for (2) and (3) in order to determine representations for some erf-G mathematical properties, which do not present closed-forms. First, consider the Maclaurin expansion for the erf function given by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\sum_{k=0}^{\infty}\frac{(-1)^{k}x^{2k+1}}{k!(2k+1)}. \tag{5}$$

By applying (5) in (2), one has that

$$F(x) = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k \left[\frac{G(x)}{1-G(x)}\right]^{2k+1}}{k!(2k+1)}. \tag{6}$$

From the Taylor expansion, we have

$$\frac{x}{1-x} = \sum_{i=1}^{\infty} x^i \quad \text{for} \quad |x| < 1, \tag{7}$$

(7) applied in (6) collapses

$$F(x) = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(2k+1)} \left[\sum_{i=1}^{\infty} G(x)^i\right]^{2k+1}. \tag{8}$$

Setting ℓ as a positive integer number, we have

$$\left(\sum_{k=0}^{\infty} a_k x^k\right)^{\ell} = \sum_{m=0}^{\infty} c_{\ell,m} x^m, \tag{9}$$

where

$$c_{\ell,0} = a_0^{\ell}, \quad c_{\ell,m} = \frac{1}{m\, a_0} \sum_{j=1}^{m} (j\ell - m + j) a_j c_{\ell,m-j}, \quad m \geq 1.$$

From (9) in (8), we get

$$F(x) = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{(-1)^k d_{2k+1,m}}{k!(2k+1)} G(x)^{m+2k+1} = \sum_{k,m=0}^{\infty} b_{k,m}\, G(x)^{m+2k+1}, \tag{10}$$

where $d_{2k+1,0} = 1, d_{2k+1,m} = \dfrac{1}{m} \displaystyle\sum_{j=1}^{m} \big[2j(k+1) - m\big] d_{2k+1,m-j}, m \geq 1$ and

$$b_{k,m} = \frac{2\,(-1)^k\, d_{2k+1,m}}{\sqrt{\pi}\, k!\,(2k+1)}.$$

By applying the derivate with respect to x in (10), erf-G PDF can be express as

$$f(x) = \sum_{k,m=0}^{\infty} b_{k,m}\,(m+2k+1)\, g(x)\, G(x)^{m+2k} = \sum_{k,m=0}^{\infty} a_{k,m}\, g(x)\, G(x)^{m+2k}, \tag{11}$$

where $a_{k,m} = b_{k,m}\,(m + 2k + 1)$. Equations (10) and (11) indicate that erf-G random variables can be represented as a linear combination of exp-G distributions (discussed in detailed by Tahir and Nadarajah (2015)) having additional parameter $m + 1$.

3.2 Maximum likelihood estimation

Let $x_1, \ldots, x_n$ be a n-points observed sample obtained from $X \sim \text{erf}G(\boldsymbol{\theta})$. The log-likelihood function for the vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathrm{R}^p$ is expressed as

$$\ell(\boldsymbol{\theta}) = n\log\left(\frac{2}{\sqrt{\pi}}\right) + \sum_{i=1}^{n}\log\left[g(x_i|\boldsymbol{\theta})\right] - 2\sum_{i=1}^{n}\log\left[1 - G(x_i|\boldsymbol{\theta})\right] - \sum_{i=1}^{n}\frac{G(x_i|\boldsymbol{\theta})^2}{[1 - G(x_i|\boldsymbol{\theta})]^2}, \tag{12}$$

In this case, the jth element of the score vector, $\mathbf{U}(\boldsymbol{\theta}) = [U_1(\boldsymbol{\theta}), \ldots, U_p(\boldsymbol{\theta})]^\top = \left[\frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_1}, \ldots, \frac{\partial\ell(\boldsymbol{\theta})}{\partial\theta_p}\right]^\top$, is given by

$$\begin{aligned} U_j(\boldsymbol{\theta}) = &\sum_{i=1}^{n}\frac{\dot{g}(x_i|\boldsymbol{\theta})}{g(x_i|\boldsymbol{\theta})} + 2\sum_{i=1}^{n}\frac{\dot{G}(x_i|\boldsymbol{\theta})}{[1 - G(x_i|\boldsymbol{\theta})]} - 2\sum_{i=1}^{n}\frac{G(x_i|\boldsymbol{\theta})\dot{G}(x_i|\boldsymbol{\theta})[1 - G(x_i|\boldsymbol{\theta})]^2}{[1 - G(x_i|\boldsymbol{\theta})]^4} \\ &- 2\sum_{i=1}^{n}\frac{G(x_i|\boldsymbol{\theta})^2\dot{G}(x_i|\boldsymbol{\theta})[1 - G(x_i|\boldsymbol{\theta})]}{[1 - G(x_i|\boldsymbol{\theta})]^4}, \end{aligned}$$

where $\dot{g}(x_i|\boldsymbol{\theta}) = \partial g(x_i;\boldsymbol{\theta})/\partial\theta_j$ and $\dot{G}(x_i|\boldsymbol{\theta}) = \partial G(x_i;\boldsymbol{\theta})/\partial\theta_j$. Thus, the maximum likelihood estimator (ML estimator) are given by

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\theta\in\Theta}\{\ell(\boldsymbol{\theta})\}$$

or, equivalently, $\hat{\boldsymbol{\theta}}$ is a root of the non-linear equations system defined by $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$.

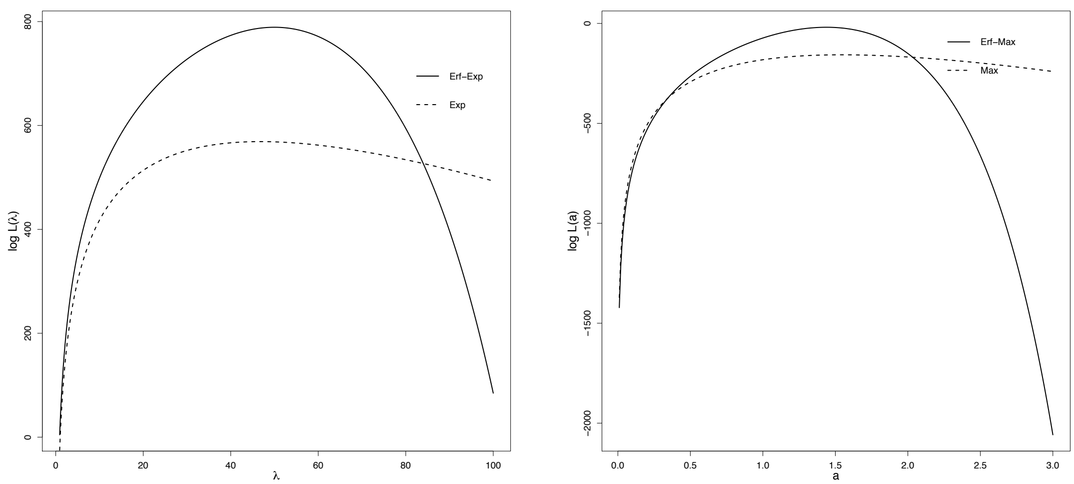
To illustrate as the erf-G model can modify geometrically a G distribution log-likelihood, we compare two pairs of distributions: (exponential (Exp), erf-exponential (erfExp)) and (Maxwell (Max), erf-Maxwell (erfMax)). The erfExp log-likelihood function is given by

$$\ell(\lambda) = n\log\left(\frac{2\lambda}{\sqrt{\pi}}\right) + \lambda\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}(1 - \mathrm{e}^{\lambda x_i}).$$

From Figure 7, it is noticeable that the erf-G structure may provide concavity to distributions with flat or quasi-flat likelihoods. It advocates in favor of the proposed family. Among other advantages, a greater concavity of likelihood provides better quality in the estimation process. In the next section, we illustrate that the maximum likelihood estimates (ML estimates) based on (12) may be more accurate than those obtained from the corresponding baseline.

3.3 The log-erf-Frechet regression model

In several applications, lifetimes are related to exatory variables. Regression models are sought for this end. Let T be a random variable with PDF (4), then $Y = \log(T)$ has the log-erf-Frechet (lerfF) distribution, denoted as $Y \sim \text{lerfF}$. Taking the parametrization $\delta = \exp(\mu)$ and $\lambda = 1/\sigma$, the PDF of Y can be written as



(a) The log-likelihood function for the Exp and erfExp distributions

(b) The log-likelihood function for the Max and erf-Max distribution

Figure 7. The log-likelihood function for the Exp, erfExp, Max and erfMax distributions.

$$
\begin{aligned}
f(y,\mu,\sigma) = \frac{2}{\sqrt{\pi}\sigma} \exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right] \exp\left\{\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\} \left(\exp\left\{\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}-1\right)^{-2} \\
\times \exp\left[-\left(\exp\left\{\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}-1\right)^{-2}\right],
\end{aligned}
\tag{13}
$$

for $-\infty < y < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$. Now, if $T \sim \mathrm{erfF}(\delta, \lambda)$, then $Y = \log(T) \sim \mathrm{lerfF}(\mu, \sigma)$ with CDF

$$
F_Y(y) = \mathrm{erf}\left[\left(\exp\left\{\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\right)^{-1}\right],
$$

and survival function (sf) given by

$$
S(y;\mu,\sigma) = 1 - \mathrm{erf}\left[\left(\exp\left\{\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\right)^{-1}\right]. \tag{14}
$$

Now, we are in position of defining the standardized random variable $Z = (Y-\mu)/\sigma$ with PDF

$$
\pi(z) = \frac{2}{\sqrt{\pi}} \exp(-z) \exp\big[\exp(-z)\big]\big\{\exp[\exp(-z)]-1\big\}^{-2} \exp\big(-\{\exp[\exp(-z)]-1\}^{-2}\big). \tag{15}
$$

Considering the substitution $u = \big\{\exp[\exp(-z)]-1\big\}^{-1}$, the r-th moment of Z is given by

$$
\mathrm{E}(Z^r) = \frac{2}{\sqrt{\pi}} \int_0^\infty \big\{-\log[\log(u^{-1}+1)]\big\}^r \exp(-u^2)\mathrm{d}u.
$$

Using the `Mathematica` software, it is possible to verify that the second ordinary moment of Z is finite:

$$\mathrm{E}(Z^2) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \left\{ -\log[\log(u^{-1}+1)] \right\}^2 \exp(-u^2) \mathrm{d}u = 0.321075 < \infty.$$

Let $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^\top$ be the exatory variable vector associated with the ith response variable Y_i for $i = 1, \ldots, n$.

Consider the sample $(Y_1, \boldsymbol{x_1}), \ldots, (Y_n, \boldsymbol{x_n})$ of n independent variables, where each random response is defined by $Y_i = \min\{\log(T_i), \log(c_i)\}$ and $\log(T_i)$ and $\log(c_i)$ are the $\log -$lifetime and $\log -$censoring, respectively. We consider non-informative censorship such that the lifetimes and censorship times are independent.

The linear regression model for the lerfF response variable, Y_i, is given by

$$Y_i = \boldsymbol{x_i}^\top \boldsymbol{\beta} + \sigma Z_i, \quad i = 1, 2, \ldots, n. \tag{16}$$

where Z_i is a random variable with PDF (15), $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ and $\sigma > 0$ are unknown parameters, and $\boldsymbol{x_i}$ is the ith explanatory random variables vector.

In this case, the location of $(Y_1, \ldots, Y_n)^\top$ is $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$ such that $\mu_i = \boldsymbol{x_i}^\top \boldsymbol{\beta}$ or, in matrix terms, $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ with model matrix $\boldsymbol{X} = (x_1, \ldots, x_n)^\top$.

Let F and C be the sets of individuals for which y_i is the $\log -$lifetime or $\log -$censoring, respectively.

The total log-likelihood function for the parameters $\boldsymbol{\theta} = (\sigma, \boldsymbol{\beta}^\top)^\top$ of model (16) has the form

$$\ell(\boldsymbol{\theta}) = \sum_{i \in F} \ell_i(\boldsymbol{\theta}) + \sum_{i \in C} \ell_i^{(c)}(\boldsymbol{\theta}),$$

where $\ell_i(\boldsymbol{\theta}) = \log[f(y_i)]$, $\ell_i^{(c)}(\boldsymbol{\theta}) = \log[S(y_i)]$, $f(y_i)$ and $S(y_i)$ are given in equations (13) and (14). Then, the $\log -$likelihood function reduces to

$$\begin{aligned}
\ell(\boldsymbol{\theta}) = q \left(\log(2) - \frac{\log(\pi)}{2} - \log(\sigma) \right) + \sum_{i \in F} \Bigg\{ -\left(\frac{y_i - \boldsymbol{x_i}^\top \boldsymbol{\beta}}{\sigma} \right) + \exp\left[-\left(\frac{y_i - \boldsymbol{x_i}^\top \boldsymbol{\beta}}{\sigma} \right) \right] \\
-2 \log \left(\exp\left\{ \exp\left[-\left(\frac{y_i - \boldsymbol{x_i}^\top \boldsymbol{\beta}}{\sigma} \right) \right] \right\} - 1 \right) - \left(\exp\left\{ \exp\left[-\left(\frac{y_i - \boldsymbol{x_i}^\top \boldsymbol{\beta}}{\sigma} \right) \right] \right\} - 1 \right)^{-2} \Bigg\} \\
+ \sum_{i \in C} \log \left\{ 1 - \operatorname{erf}\left[\left(\exp\left\{ \exp\left[-\left(\frac{y_i - \boldsymbol{x_i}^\top \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \right)^{-1} \right] \right\},
\end{aligned} \tag{17}$$

where q is the observed number of failures. The ML estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained by maximizing the Equation (17). Using the adjusted model (16), the sf of Y_i can be estimated by

$$\widehat{S}(y_i; \widehat{\sigma}, \widehat{\boldsymbol{\beta}}^\top) = 1 - \operatorname{erf}\left[\left(\exp\left\{ \exp\left[-\left(\frac{y_i - \boldsymbol{x_i}^\top \widehat{\boldsymbol{\beta}}}{\widehat{\sigma}} \right) \right] \right\} \right)^{-1} \right].$$

Under general regularity conditions, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ can be approximated by the multivariate normal $N_{p+1}(0, J(\boldsymbol{\theta})^{-1})$, where $J(\boldsymbol{\theta}) = \partial^2 \ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top \partial\boldsymbol{\theta}$ is the $(p+1)\times(p+1)$ observed information matrix. Statistical inference procedures for the parameter vector $\boldsymbol{\theta}$ can be made based on the asymptotic normality. In particular, an $100(1-\alpha)\%$ asymptotic confidence interval for each parameter θ_s is given by

$$\mathrm{ACI}_s = (\theta_s - z_{\alpha/2}\sqrt{\widehat{J}^{s,s}}, \theta_s + z_{\alpha/2}\sqrt{\widehat{J}^{s,s}}),$$

where $\widehat{J}^{s,s}$ denotes the sth diagonal element of the inverse of the estimated observed information matrix $J(\widehat{\boldsymbol{\theta}})^{-1}$ and $z_{\alpha/2}$ is the quantile $1-\alpha/2$ of the standard normal distribution.

4. Some mathematical properties

From now on, we present the process of obtaining the mathematical properties of the new model.

4.1 Quantile function

The quantile function (qf) of the erf-G distribution is obtained in an explicit form by inverting (2)

$$Q_F(u) = Q_G\left(\frac{\Phi^{-1}(\frac{u+1}{2})}{\sqrt{2}+\Phi^{-1}(\frac{u+1}{2})}\right), \tag{18}$$

where Q_G is the baseline quantile function and Φ^{-1} is the standard normal quantile function. Beyond to allow defining important quantiles (e.g., the median), (18) may also be used as a random variables generator, adopting uniform outcomes as inputs.

4.2 Ordinary and incomplete moments

Let X be a random variable following erf-G distribution. From Equation (11), the rth moment of X may be written as

$$\mathrm{E}(X^r) = \sum_{k,m=0}^{\infty} a_{m,k}\,\mathrm{E}(Y_{m+2k+1}^r),$$

where Y_{m+2k+1} follows the exponentiated distribution at the power parameter $m+2k+1$. Another way to represent the rth moment is through of the quantile function as follow:

$$\mathrm{E}(X^r) = \sum_{k,m=0}^{\infty} a_{m,k}\int_0^1 \left[Q_G\left(u^{\frac{1}{m+k+1}}\right)\right]^r \mathrm{d}u.$$

The rth incomplete moment of X can be given as follow

$$\mathrm{T}_r(z) = \int_{-\infty}^{z} x^r f(x)dx = \sum_{k,m=0}^{\infty} a_{m,k}\,\mathrm{T}_r^*(z),$$

where $\mathrm{T}_r^*(z)$ is the rth incomplete moment of the Y_{m+2k+1}. A second manner to obtain the rth incomplete moment of X is by using the quantile function, we have

$$\mathrm{T}_r(z) = \int_{-\infty}^{z} x^r f(x) dx = \sum_{m,k=0}^{\infty} a_{m,k} \int_0^{[G(z)]^{m+2k+1}} \left[Q_G\left(u^{\frac{1}{m+2k+1}}\right)\right]^r \mathrm{d}u.$$

4.3 Moment generating function

By using the Equation (11), the mgf of X can be expressed as

$$\mathrm{M}(t) = \sum_{m,k=0}^{\infty} b_{m,k}\, \mathrm{M}_{m+2k+1}(t),$$

where $\mathrm{M}_{m+2k+1}(t)$ is the mgf of Y_{m+2k+1} given by

$$\mathrm{M}_{m+2k+1}(t) = \int_{-\infty}^{\infty} \exp(tx)(m+2k+1) g(x) [G(x)]^{m+2k} \mathrm{d}x.$$

Another form to obtain an expansion of the mgf of X is by using the qf. We have

$$\mathrm{M}(t) = \sum_{m,k=0}^{\infty} (m+2k+1)\, b_{m,k} \int_0^1 \exp\left[t\, Q_G(u)\right] u^{m+2k} \mathrm{d}u.$$

4.4 Entropy

Two well-known variability measures are the Shannon and Rényi entropies. Determining their expressions consist an important task to quantify disorder in stochastic systems. In what follows, we derive these measures for the erf-G family. First consider the expansion: Assuming that $|z| < 1$ and $\rho > 0$,

$$(1-z)^{-\rho} = \sum_{j=1}^{\infty} w_j z_j, \quad w_j = \frac{\Gamma(\rho+j)}{j!\,\Gamma(\rho)}. \tag{19}$$

Considering the Taylor expansion and (19) an expression to the erf-G Rényi entropy is (for $\delta > 0$ and $\delta \neq 1$)

$$\begin{aligned}
\mathrm{I_R}(\delta) &= \frac{1}{1-\delta} \log\left(\int_0^{\infty} [f(x)]^{\delta} \mathrm{d}x\right) \\
&= \frac{1}{1-\delta} \log\left[\frac{2^{\delta}}{\pi^{\delta/2}} \sum_{k=0}^{\infty}\sum_{j=1}^{\infty} \frac{(-\delta)^k w_j}{k!} \int_0^{\infty} \left[g(x)\right]^{\delta} \left[G(x)\right]^{2k+j} \mathrm{d}x\right] \\
&= \frac{1}{1-\delta} \left\{\delta \log(2) - \frac{\delta}{2}\log(\pi) + \log\left(\sum_{k=0}^{\infty}\sum_{j=1}^{\infty} \frac{(-\delta)^k w_j}{k!} \int_0^{\infty} \left[g(x)\right]^{\delta} \left[G(x)\right]^{2k+j} \mathrm{d}x\right)\right\},
\end{aligned}$$

where $w_j = \dfrac{\Gamma[2(\delta+1)+j]}{j!\,\Gamma[2(\delta+1)]}$.

The Shannon entropy is defined as $\mathrm{E}\{-\log[f(X)]\}$ and it can be obtained from the Rnyi entropy doing $\delta \uparrow 1$. Note that

$$\mathrm{E}\big\{-\log[f(X)]\big\} = -2\log(2)+\frac{1}{2}\log(\pi)-\mathrm{E}\big[\log(X)\big]+\mathrm{E}\left\{\left[\frac{G(X)}{1-G(X)}\right]^2\right\}-2\mathrm{E}\big[\log(1-g(X))\big].$$

After some algebraic manipulations, we obtain

$$\mathrm{E}\big[\log(X)\big] = \sum_{m,k=0}^{\infty} b_{m,k}(m+2k+1)\int_0^1 u^{m+2k}\log\big[g(Q_G(u))\big]\mathrm{d}u,$$

$$\begin{aligned}\mathrm{E}\left\{\left[\frac{G(X)}{1-G(X)}\right]^2\right\} &= \int_{-\infty}^{\infty}\mathrm{E}\left\{\left[\frac{G(X)^2}{[1-G(X)]^2}\right]^2\right\} f(x)\mathrm{d}x \\ &= \sum_{m,k=0}^{\infty} b_{m,k}(m+2k+1)\int_0^1 \frac{u^{m+2k+2}}{(1-u)^2}\mathrm{d}u\end{aligned}$$

and

$$\mathrm{E}\big[\log(1-g(X))\big] = -\sum_{i=0}^{\infty}\sum_{m,k=0}^{\infty}\frac{b_{m,k}(m+2k+1)}{(i+1)(m+2k+i+2)}.$$

5. Numerical applications

In order to assess the performance of estimation procedures, we carry out a Monte Carlo study and two real data set applications.

5.1 A Monte Carlo study

This section aims to quantify the performance of ML estimators for erf-G parameters distribution. To that end, we consider the exponential (exp), Levy and Maxwell (Max) models, after we specify the following baseline models: erf{Exp, Levy, Max} using equation (3). The PDF's of the Exp, Levy and Max distributions are given, respectively, by

$$f(x,\lambda) = \lambda\exp(-\lambda x), \quad x>0, \quad \lambda>0,$$

$$f(x,\lambda) = \sqrt{\frac{\lambda}{2\pi}}\frac{\exp(-\frac{a}{2x})}{x^{\frac{3}{2}}}, \quad x>0, \quad \lambda>0$$

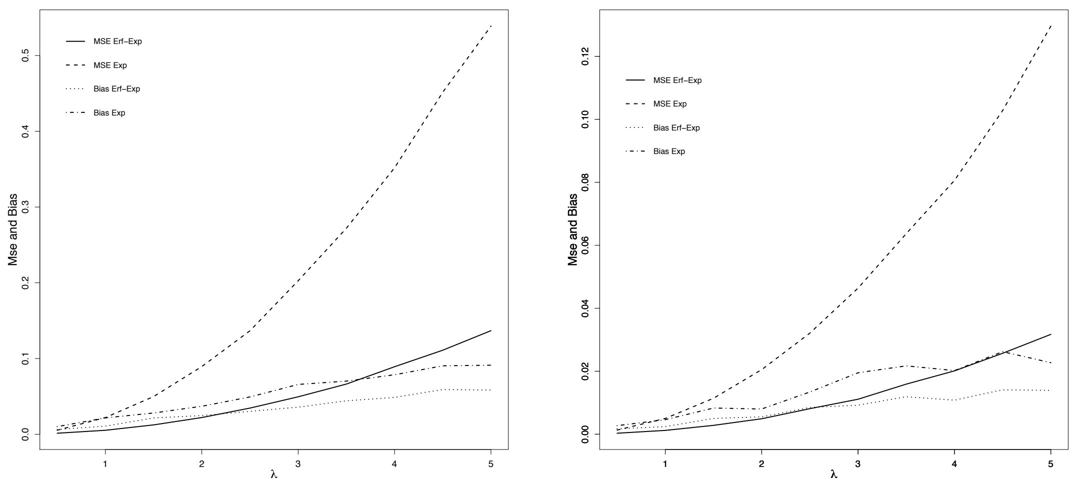
and

$$f(x;a) = \sqrt{\frac{2}{\pi}}a^{\frac{3}{2}}x^2\exp\left(\frac{1}{2}ay^2\right), \quad y>0, \quad a>0.$$

We make a Monte Carlo study with $10,000$ replications such that, for several baseline parameter values and sample sizes $n \in \{50, 200\}$, two comparison criteria are quantified:

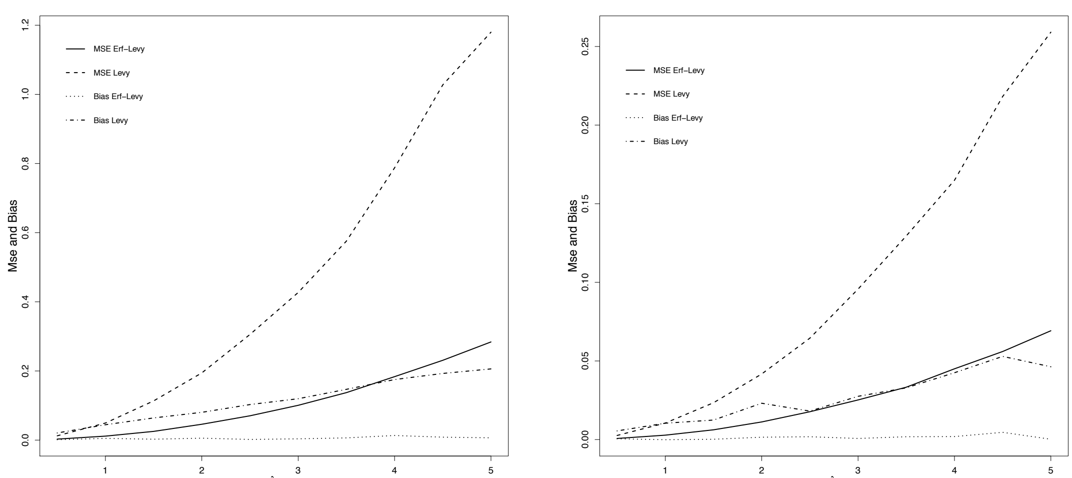
biases and root mean squared error (RMSE). All computations are implemented using the `R` programming language, which has numerous advantages, perhaps the main one being the fact that it is distributed free of charge through the so-called *GNU Public license*. For more information about `R`, visit the `https://www.r-project.org` website. To ensure the reproducibility of this experiment, the following comments are needed: It was utilized the `maxLik(.)` function of the `R` package `maxLik` . Specifically, the BFGS iterative method was used in the optimization process.

Simulation results are presented in Figures 8, 9 and 10. Based on these plots, we conclude that: (i) As expected, the biases and RMSE decreases as the sample size increases; (ii) The erf-G models has superior performance when compared to their respective baseline models.



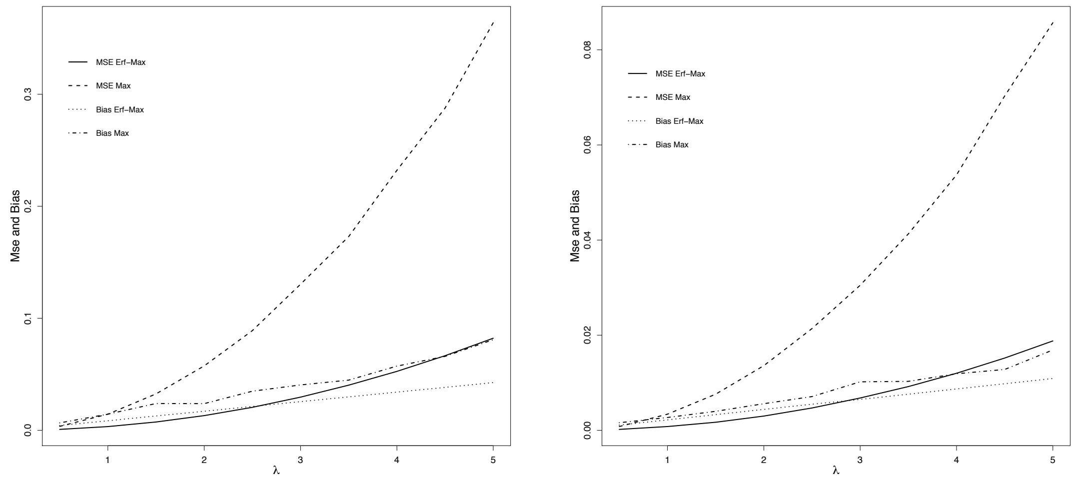
(a) $n = 50$ (b) $n = 200$

Figure 8. RMSEs and biases of $\widehat{\lambda}$ for the erfExp and Exp models.



(a) $n = 50$ (b) $n = 200$

Figure 9. RMSEs and biases of $\widehat{\lambda}$ for the erfLevy and Levy models.



(a) $n = 50$ (b) $n = 200$

Figure 10. RMSEs and biases of $\widehat{\lambda}$ for the erfMax and Max models.

5.2 Real data applications

Two applications to real data illustrate the performance of proposed models. First we describe a set of lifetime data by means of some erf-G models comparatively to the corresponding G distributions. Second the llr model performance is quantified and compared.

5.2.1 Unconditioned model

This section addresses an application to a real data set to illustrate the usefulness of the proposed family.

To that end, we consider three baseline distributions: exponential (Exp), Kumaraswamy (K) and Weibull (W). The main objective is to show that the distributions extended from the erf-G family perform better when compared with their baseline distributions.

We use a data set obtained in Proschan (1963) and corresponds to the time of successive failures of the air conditioning system of jet airplanes. These data were also studied by Dahiya and Gurland (1972), Gleser (1989) and Kuş (2007), among others. The data are

194, 413, 90, 74, 55, 23, 97, 50, 359, 50, 130, 487, 102, 15, 14, 10, 57, 320, 261, 51, 44, 9 , 254, 493, 18, 209, 41, 58, 60, 48, 56, 87, 11, 102, 12, 5, 100, 14, 29, 37, 186, 29, 104, 7, 4, 72, 270, 283, 7, 57, 33, 100, 61, 502, 220, 120, 141, 22, 603, 35, 98, 54, 181, 65, 49, 12, 239, 14, 18, 39, 3, 12, 5, 32, 9, 14, 70, 47, 62, 142, 3, 104, 85, 67, 169, 24, 21, 246, 47, 68, 15, 2, 91, 59, 447, 56, 29, 176, 225, 77, 197, 438, 43, 134, 184, 20, 386, 182, 71, 80, 188, 230, 152, 36, 79, 59, 33, 246, 1, 79, 3, 27, 201, 84, 27, 21, 16, 88, 130, 14, 118, 44, 15, 42, 106, 46, 230, 59, 153, 104, 20, 206, 5, 66, 34, 29, 26, 35, 5, 82, 5, 61, 31, 118, 326, 12, 54, 36, 34, 18, 25, 120, 31, 22, 18, 156, 11, 216, 139, 67, 310, 3, 46, 210, 57, 76, 14, 111, 97, 62, 26, 71, 39, 30, 7, 44, 11, 63, 23, 22, 23, 14, 18, 13, 34, 62, 11, 191, 14, 16, 18, 130, 90, 163, 208, 1, 24, 70, 16, 101, 52, 208, 95.

Some descriptive statistics for these data are given in Table 1. Note that the mean is greater than the median and the asymmetry coefficient is positive, i.e., the empirical distribution from data is positively asymmetric. There is a lot of variability in the data and they are overdispersed. Further, from the kurtosis coefficient, the distribution of the data is platykurtic.

Table 2 provides the ML estimates of considered model parameters (corresponding standard errors in parentheses) and the values of some goodness-of-fit measures: the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC). In general, it is considered that the lower values (AIC, BIC and CAIC) indicate better fits. In all the situations, the proposed models outperform the corresponding baselines.

Table 1. Descriptive statistics for the air conditioning system of airplanes data.

Statistic	
Mean	93.141
Median	57
Variance	11398.471
Minimum	1
Maximum	603
Skewness	2.322
Kurtosis	3.692

Table 2. The ML estimates (standard errors in parentheses) and the AIC, BIC and CAIC for the phosphorus concentration data.

Distribution	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\delta}$	$\widehat{\eta}$	Cramr	K-S	AD	AIC	BIC	CAIC
BGP	10.778	1.031	22.346	28.890	–	0.302	0.079	2.044	2386.988	2400.433	2387.180
	(0.791)	(0.356)	(1.549)	(0.872)							
KumaBXII	16.190	6.810	5.761	0.057	0.100	0.215	0.069	1.487	2381.423	2398.230	2381.713
	(3.375)	(1.831)	(2.008)	(0.019)	(0.059)						
Gama-Gama	10.997	0.001	22.975	–	–	0.851	0.122	5.085	2475.640	2485.724	2475.755
	(0.227)	(0.000)	(0.002)								
erf-We	0.043	0.524	–	–	–	0.475	0.109	2.857	2390.732	2397.455	2390.789
	(0.006)	(0.025)									

As qualitative comparison sources, plots ofnthe empirical and estimated PDF and CDF of the under discussion models are displayed in Figures 11. Results indicate the fitted erfW, erfExp and erfK models are better than the associated baselines for phosphorus concentration data. These are first practical evidences in favor of the use of the proposed family.

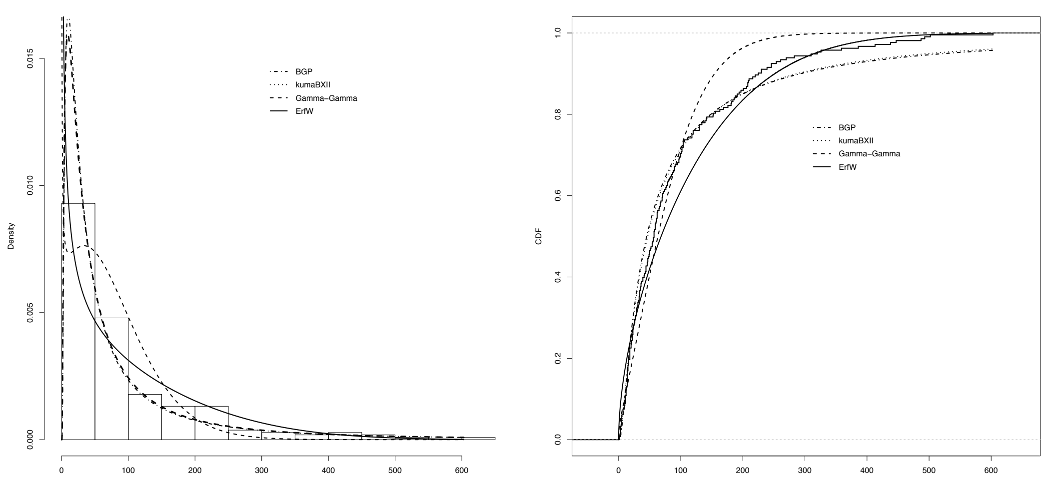


Figure 11. Plots of the fitted BGP, KumaBXII, Gamma-Gamma and erfW PDFs (left) and of the estimated CDFs of the BGP, KumaBXII, gamma-gamma and erfW models (right)s.

.

5.2.2 Regression model

Now, consider results obtained from a lifetime test experiment on 76 specimens of a type of electrical insulating fluid subjected to constant voltage stress, say x, at seven levels, $x = 26, 28, 30, 32, 34, 36$ and 38 kV. The time period until each sample has failed (or "broke"), say breaking time Y, was observed. Such study was firstly performed by Nelson (1972) and Vanegas et al. (2012). Now, we aim to investigate how the voltage level influences the failure time. Does the erf$G$ structure present advantage in the regression context likewise that for uncorrelated distributions?

To that end, we compare the lerfF and log-Frechet (L-F) regression models. Table 3 presents results for ML estimates of the adjusted models as well as their respective significance and standard error measures. We also provide values of the AIC, BIC and CAIC statistics as comparison means. From results of individual confidence intervals for β_i, one has that both considered slopes (and, as a consequence, used predictive variables) are meaningful at the level 5%, employing the asymptotic distribution of the t statistic for $\mathcal{H} : \beta_1 = 0$. From comparison point-of-view, lerfF regression model outperforms L-F, illustrating the importance of the erf-G family in the regression context.

Table 3. ML estimates of the parameters from some fitted regression models to the Minutes to breakdown data set, the corresponding standard errors (in parentheses), p-value (in brackets) and the AIC, BIC and CAIC measures.

Model	β_0	β_1	σ	AIC	BIC	CAIC
lerfF	13.5272	-0.3329	3.1597	297.5957	304.5879	297.9290
	(1.9256)	(0.0586)	(0.3053)			
	[<0.0001]	[<0.0001]	[<2e-16]			
L-F	7.1364	-0.1790	1.7176	320,9327	327,9249	321,2660
	(2.3629)	(0.0697)	(0.1601)			
	[0.0025]	[0.0102]	[<2e-16]			

6. Conclusions and future works

In this paper, we propose and study a new class of distributions called efr-G family. This family is based on the known error function and does not add parameters to its resulting models regard to the baseline distribution. As other advantage, the erf-G family seems to solve or at least to improve estimates based on flat likelihoods. We derive some of its mathematical properties, such as quantile function, ordinary and incomplete moments, moment generating function and Shannon and Rényi entropy measures. A log-linear regression model in the new family is also proposed. Simulation studies and real data applications illustrate the usefulness of the our proposals. For future works, new regression models and a complete study of residual analysis for the proposed models will be developed.

Acknowledgements

We thank two anonymous referees and the associate editor for their valuable suggestions, which certainly contributed to the improvement of this paper. Additionally, Thiago A. N. de Andrade is grateful the financial support from CAPES (Brazil), through its program to encourage post-doctoral researches. He also thanks to the statistics department of the Federal University of Pernambuco.

References

Afify, A., Altun, E., Alizadeh, M., Ozel Kadilar, G., and Hamedani, G., 2017. The odd exponentiated half-logistic-g family: Properties, characterizations and applications. Chilean Journal of Statistics, 8:65–91.

Chevillard, S., 2012. The functions erf and erfc computed with arbitrary precision and explicit error bounds. Information and Computation, 216:72 –95.

Cordeiro, G.M., Aristizabal, W.D., Suárez, D.M., and Lozano, S., 2015. The gamma modified Weibull distribution. Chilean Journal of Statistics, 6:37-48.

Cordeiro, G.M. and de Castro, M., 2011. A new family of generalized distributions. Journal of Statistical Computation and Simulation, 81:883–893.

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

da Silva, L.C.M., de Andrade, T.A.N., Maciel, D.B.M., Campos, R.P.S., and Cordeiro, G.M., 2013. A new lifetime model: the gamma extended Frechet distribution. Journal of Statistical Theory and Applications, 12:39–54.

Dahiya, R.C. and Gurland, J., 1972. Goodness of fit tests for the gamma and exponential distributions. Technometrics, 14:791–801.

de Andrade, T.A.N., Bourguignon, M., and Cordeiro, G.M., 2016. The exponentiated generalized extended exponential distribution. Journal of Data Science, 14:393–414.

de Andrade, T.A.N., Rodrigues, H., Bourguignon, M., and Cordeiro, G.M., 2015. The exponentiated generalized Gumbel distribution. Revista Colombiana de Estadística, 38:123–143.

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

Gleser, L.J., 1989. The gamma distribution as a mixture of exponential distributions. The American Statistician, 43:115–117.

Kuş, C., 2007. A new lifetime distribution. Computational Statistics and Data Analysis, 51:4497–4509.

Leao, J., Saulo, H., Bourguignon, M., Cintra, R., Rego, L., and Cordeiro, G.M., 2013. On some properties of the beta inverse Rayleigh distribution. Chilean Journal of Statistics, 4:111–131.

Marshall, A.N. and Olkin, I., 1997. A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. Biometrika, 84:641–652.

Nadarajah, S., Cordeiro, G.M., and Ortega, E.M.M., 2015. The Zografos-Balakrishnan-G family of distributions: Mathematical properties and applications. Communications in Statistics: Theory and Methods, 44:186–215.

Nelson, W., 1972. Graphical analysis of accelerated life test data with the inverse power law model. IEEE Transactions on Reliability, 21:1–10.

Proschan, F., 1963. Theoretical explanation of observed decreasing failure rate. Technometrics, 5:375–383.

Ristic, M.M. and Balakrishnan, N. (2012). The gamma exponentiated exponential distribution. Journal of Statistical Computation and Simulation, 82:1191–1206.

Tahir, M.H. and Nadarajah, S., 2015. Parameter induction in continuous univariate distributions: Well-established G families. Anais da Academia Brasileira de Ciencias, 87:539–568.

Vanegas, L., Rondon, L., and Cordeiro, G.M., 2012. Diagnostic tools in generalized Weibull linear regression models. Journal of Statistical Computation and Simulation, 83:1–24.

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

Information for authors

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF file of the manuscript in a free format to Editors of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a "cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

Preparation of accepted manuscripts

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at http://chjs.mat.utfsm.cl/files/ChJS.zip. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at http://www.ams.org/mathscinet/msc/. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

Copyright