



HAL
open science

The odd generalized gamma-G family of distributions: Properties, regressions and applications

Arslan Nasir, Farrukh Jamal, Christophe Chesneau

► To cite this version:

Arslan Nasir, Farrukh Jamal, Christophe Chesneau. The odd generalized gamma-G family of distributions: Properties, regressions and applications. 2019. hal-01916828v4

HAL Id: hal-01916828

<https://hal.science/hal-01916828v4>

Preprint submitted on 5 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The odd generalized gamma-G family of distributions: Properties, regressions and applications

Arslan Nasir¹, Farrukh Jamal² and Christophe Chesneau³

^{1,2}Department of Statistics, The Islamia University Bahawalpur, Pakistan.

³LMNO, University of Caen, France.

¹arslannasir147@gmail.com ²drfarrukh1982@gmail.com

³christophe.chesneau@gmail.com

March 5, 2019

Abstract

In this article, a new “odd generalized gamma-G” family of distributions, called the GG-G family of distributions, is introduced. We propose a complete mathematical and statistical studies of this family, with a special focus on the Fréchet distribution as baseline distribution. In particular, we provide infinite mixture representations of its probability density function and its cumulative distribution function, the expressions for the Rényi entropy, the reliability parameter and the probability density function of i th order statistic. Then the statistical properties of the family are explored. Model parameters are estimated by the maximum likelihood method. A regression model is also investigated. A simulation study is performed to check the validity of the obtained estimators. Applications on real data sets are also included, with favorable comparisons to existing distributions in terms of goodness-of-fit.

Keywords— Gamma distribution; Moments; Order statistics; Rényi entropy; Maximum likelihood method; Regression model; Data analysis.

2000 Mathematics Subject Classification:— 60E05, 62E15, 62F10.

1 Introduction

The deep analysis of data from sophisticated phenomenons can be limited by the use of classical (probability) distributions. This motivates the statisticians to develop new distributions/models having the ability to capture special features of the phenomenons from the data. One of the most common approach is the use of the so called “generators of distributions” which consists in injecting more flexibility into an existing distribution. The construction of such generators often uses existing distributions with one or several parameters and with well-known structural properties. Among the most popular generators, there are the beta generator (see Eugene *et al.* (2002) and Jones (2004)), the Transmuted generator (see Shaw and Buckley (2007)), the Kumaraswamy generator (see Cordeiro and de-Castro (2011)), the McDonald generator (see Alexander *et al.* (2012)), the

Kummer beta generator (see Peskim *et al.* (2012)), the gamma generator (see Zografos and Balakrishnan (2009), Ristić and Balakrishnan (2012) and Tobarı and Montazari (2012)), the log-gamma generator (see Amini *al.* (2014)), the logistic generator (see Torabi and Montazeri (2014)), the beta extended Weibull generator (see Cordeiro *al.* (2012)), the Transformed-Transformer (T-X) generator (see Alzaatreh *et al.* (2013)), the exponentiated TX generator (see Alzaghal *et al.* (2013)), the Weibull generator (see Alzaatreh *et al.* (2013) and Bourguignon *et al.* (2014)), the exponentiated halflogistic generator (see Cordeiro *al.* (2014)), the sine generator (see Kumar *et al.* (2015)), the odd Burr III generator (see Jamal *et al.* (2017)), the cosine-sine generator (see Chesneau *et al.* (2018)) and the generalized odd gamma generator (see Hosseini *et al.* (2018)).

In particular, Hosseini *et al.* (2018) demonstrate that the combination of gamma generator with the generalized odd transformation generates very flexible distributions, with great advantages in data analysis (more flexible kurtosis in comparison to the baseline distribution, produce a skewness for symmetrical distributions, generate distributions with symmetric, left-skewed, right-skewed and reversed-J shaped ...). The aim of this paper is to propose an alternative to the generator introduced by Hosseini *et al.* (2018). It is constructed from the generalized gamma distribution and the (standard) odd transformation. It allows us to define a new family of distributions described below. Let us recall that the probability density function (pdf) of generalized gamma distribution is given by

$$f(x; \alpha, \beta, \delta) = \frac{\alpha \delta^{\alpha\beta}}{\Gamma(\beta)} x^{\alpha\beta-1} \exp[-(\delta x)^\alpha], \quad x > 0,$$

where $\alpha > 0$ and $\beta > 0$ are the shape parameters and $\delta > 0$ is the scale parameter. If we take $\delta = 1$, then $f(x; \alpha, \beta, \delta)$ is reduced to the pdf $r(x; \alpha, \beta)$ given by

$$r(x; \alpha, \beta) = \frac{\alpha}{\Gamma(\beta)} x^{\alpha\beta-1} \exp(-x^\alpha),$$

where $\Gamma(\beta) = \int_0^{+\infty} w^{\beta-1} \exp(-w) dw$ is the gamma function. Let us now consider a cumulative distribution function (cdf) of a baseline distribution denoted by $G(x; \xi)$, where ξ denotes the related parameter vector. We then propose to use a generalized gamma generator with the standard odd transformation defined by $W[G(x; \xi)] = G(x; \xi)/[1 - G(x; \xi)]$. This yields the following cdf:

$$F(x; \alpha, \beta, \xi) = \int_{-\infty}^{W[G(x; \xi)]} r(t; \alpha, \beta) dt = \frac{\gamma \left[\beta, \left(\frac{G(x; \xi)}{1 - G(x; \xi)} \right)^\alpha \right]}{\Gamma(\beta)}, \quad (1.1)$$

where $\gamma(\beta, x) = \int_0^x w^{\beta-1} \exp(-w) dw$ is the incomplete gamma function. By differentiation, the corresponding pdf is given by

$$f(x; \alpha, \beta, \xi) = \frac{\alpha}{\Gamma(\beta)} \frac{g(x; \xi) G^{\alpha\beta-1}(x; \xi)}{(1 - G(x; \xi))^{\alpha\beta+1}} \exp \left\{ - \left[\frac{G(x; \xi)}{1 - G(x; \xi)} \right]^\alpha \right\}. \quad (1.2)$$

The corresponding hazard rate function (hrf) is given by

$$\begin{aligned} h(x; \alpha, \beta, \xi) &= \frac{f(x; \alpha, \beta, \xi)}{1 - F(x; \alpha, \beta, \xi)} \\ &= \frac{\alpha g(x; \xi) G^{\alpha\beta-1}(x; \xi)}{(1 - G(x; \xi))^{\alpha\beta+1} \left\{ \Gamma(\beta) - \gamma \left[\beta, \left(\frac{G(x; \xi)}{1 - G(x; \xi)} \right)^\alpha \right] \right\}} \exp \left\{ - \left[\frac{G(x; \xi)}{1 - G(x; \xi)} \right]^\alpha \right\}. \end{aligned} \quad (1.3)$$

For the purpose of the paper, the generator characterized by the cdf (1.1) is called the odd generalized gamma generator and the corresponding family of distributions is called odd generalized gamma G (GG-G) family of distributions. To the best of our knowledge, it is new in the literature, even if connections exist with the Generalized Odd Gamma-G family introduced by Hosseini *et al.* (2018) (the families coincide by taking $\alpha = 1$ and, in the definition of the Generalized Odd Gamma-G family, $\beta = 1$). We show in this study how our family is complementary on several aspects, and can be superior in term of goodness-of-fit to the Generalized Odd Gamma-G family.

The rest of this paper is organized as follows. In Section 2, three special models are given, with plots of their pdfs and hrfs to illustrate their flexibility. The one using the Fréchet distribution as baseline will be at the heart of our applied study. The main mathematical properties of the GG-G family are studied in Section 3, including shapes and asymptotes of the pdf and the hrf, mixture representations of the pdf and cdf in terms of baseline distribution, explicit expressions for the r th moment, r th incomplete moment, moment generating function, mean deviations, Rényi entropy, reliability parameter and the pdf of the i th order statistic. The estimations of the related parameters by the maximum likelihood method are discussed in Section 4. In Section 5, a regression model is given, with a simulation study. Section 6 is devoted to the residual analysis. Applications to real life data sets are performed in Section 7.

2 Special sub models

In this section, we study three special sub models, namely odd generalized gamma Fréchet (GGFr), odd generalized gamma Weibull (GGW) and odd generalized gamma Lomax (GGLx) distributions.

2.1 Odd generalized gamma Fréchet distribution

Let us consider the Fréchet distribution as baseline distribution, i.e. with pdf cdf, $g(x; a, b) = (a^b b/x^{b+1}) \exp[-(a/x)^b]$ and cdf $G(x; a, b) = \exp[-(a/x)^b]$, where $a, b, x > 0$. Then the cdf of the GGFr distribution is given by

$$F(x; \alpha, \beta, a, b) = \frac{\gamma\left(\beta, \left[\exp\left[\left(\frac{a}{x}\right)^b\right] - 1\right]^{-\alpha}\right)}{\Gamma(\beta)}.$$

The pdf is given by

$$f(x; \alpha, \beta, a, b) = \frac{\alpha a^b b \exp\left[-\alpha \beta \left(\frac{a}{x}\right)^b\right]}{\Gamma(\beta) x^{b+1} \left\{1 - \exp\left[-\left(\frac{a}{x}\right)^b\right]\right\}^{\alpha\beta+1}} \exp\left\{-\left[\exp\left[\left(\frac{a}{x}\right)^b\right] - 1\right]^{-\alpha}\right\}. \quad (2.1)$$

The hrf is given by

$$h(x; \alpha, \beta, a, b) = \frac{\alpha a^b b \exp \left[-\alpha \beta (a/x)^b \right]}{x^{b+1} \left\{ 1 - \exp \left[-(a/x)^b \right] \right\}^{\alpha\beta+1} \left[\Gamma(\beta) - \gamma \left(\beta, \left[\exp \left[(a/x)^b \right] - 1 \right]^{-\alpha} \right) \right]} \times \exp \left\{ - \left[\exp \left[\left(\frac{a}{x} \right)^b \right] - 1 \right]^{-\alpha} \right\}.$$

2.2 Odd generalized gamma Weibull distribution

Let us consider the gamma Weibull distribution as baseline distribution, i.e. with pdf $g(x; a, b) = a b x^{b-1} \exp(-a x^b)$ and cdf $G(x; a, b) = 1 - \exp(-a x^b)$, where $a, b, x > 0$. Then the cdf of the GGW distribution is given by

$$F(x; \alpha, \beta, a, b) = \frac{\gamma(\beta, [\exp(a x^b) - 1]^\alpha)}{\Gamma(\beta)}.$$

The pdf is given by

$$f(x; \alpha, \beta, a, b) = \frac{\alpha}{\Gamma(\beta)} a b x^{b-1} \left[1 - \exp(-a x^b) \right]^{\alpha\beta-1} \exp \left\{ a \alpha \beta x^b - \left[\exp(a x^b) - 1 \right]^\alpha \right\}.$$

The hrf is given by

$$h(x; \alpha, \beta, a, b) = \frac{\alpha a b x^{b-1} \left[1 - \exp(-a x^b) \right]^{\alpha\beta-1}}{\Gamma(\beta) - \gamma(\beta, [\exp(a x^b) - 1]^\alpha)} \exp \left\{ a \alpha \beta x^b - \left[\exp(a x^b) - 1 \right]^\alpha \right\}.$$

2.3 Odd generalized gamma Lomax distribution

Let us consider the Lomax distribution as baseline distribution, i.e. with pdf $g(x; a, b) = (a/b) (1 + x/b)^{-a-1}$ and cdf $G(x; a, b) = 1 - (1 + x/b)^{-a}$, where $a, b, x > 0$. Then the cdf of the GGLx distribution is given by

$$F(x; \alpha, \beta, a, b) = \frac{\gamma(\beta, [(1 + x/b)^a - 1]^\alpha)}{\Gamma(\beta)}.$$

The pdf is given by

$$f(x; \alpha, \beta, a, b) = \frac{\alpha}{\Gamma(\beta)} \frac{a}{b} \left(1 + \frac{x}{b} \right)^{a\alpha\beta-1} \left[1 - \left(1 + \frac{x}{b} \right)^{-a} \right]^{\alpha\beta-1} \exp \left\{ - \left[\left(1 + \frac{x}{b} \right)^a - 1 \right]^\alpha \right\}.$$

The hrf is given by

$$h(x; \alpha, \beta, a, b) = \frac{\alpha a (1 + x/b)^{a\alpha\beta-1} \left[1 - (1 + x/b)^{-a} \right]^{\alpha\beta-1}}{b [\Gamma(\beta) - \gamma(\beta, [(1 + x/b)^a - 1]^\alpha)]} \exp \left\{ - \left[\left(1 + \frac{x}{b} \right)^a - 1 \right]^\alpha \right\}.$$

Figure 1 illustrates the pdfs of the GGFr, GGW and GGLx distributions for selected values of the parameters. Figure 2 illustrates the hrfs of the GGFr, GGW and GGLx distributions for selected values of the parameters. Various forms of shapes are observed, showing the great flexibility of these special sub models, and a fortiori, the GG-G family.

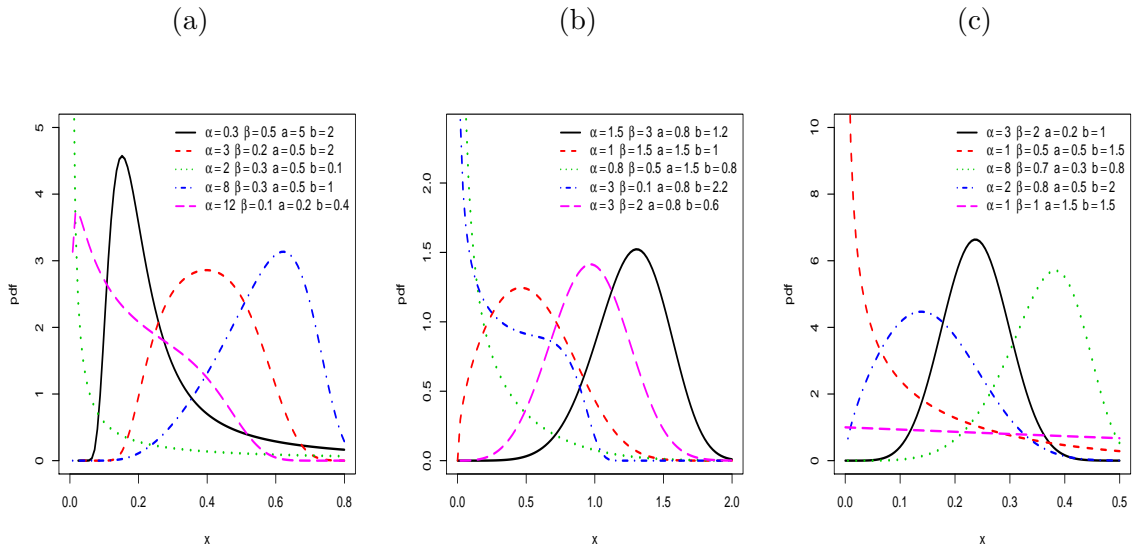


Figure 1: Plots of pdfs of the GGFr (a), GGW (b) and GGLx (c) distributions.

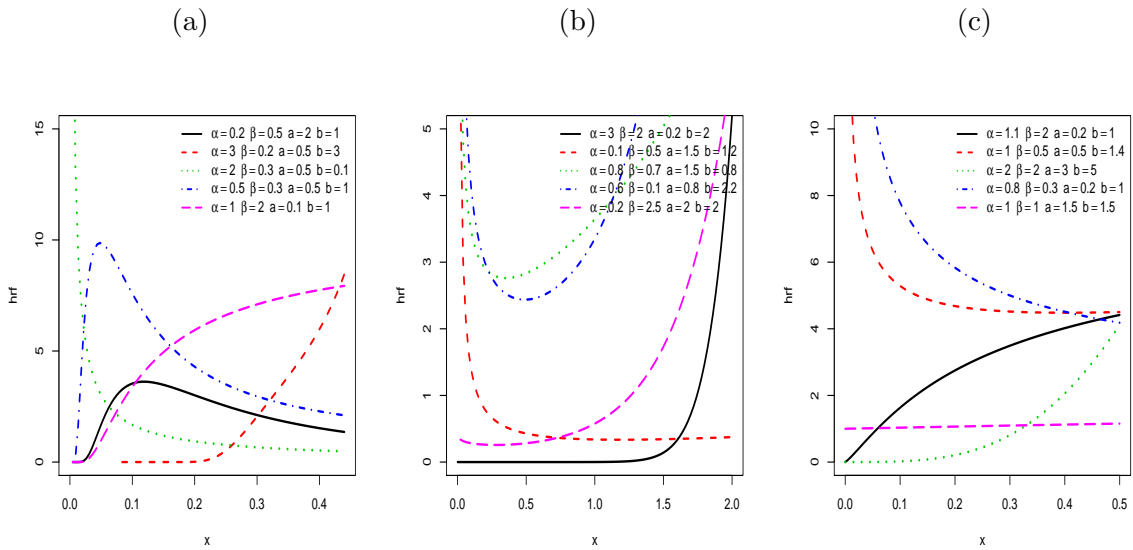


Figure 2: Plots of hrfs of the GGFr (a), GGW (b) and GGLx (c) distributions.

3 Mathematical properties

This section is devoted to the mathematical properties of the GG-G family of distributions.

3.1 Characterization

Let $G^{-1}(x, \xi)$ be the inverse function of $G(x, \xi)$ and V a random variable having the gamma distribution with parameters 1 and β . Then the random variable $X = G^{-1} \{V^{1/\alpha}/(1 + V^{1/\alpha}); \xi\}$ follows the GG-G family.

3.2 Quantile function

Let $\gamma^{-1}(\beta, x)$ be the inverse function of $\gamma(\beta, x)$. The quantile function of the GG-G family is given by

$$Q(y, \xi) = G^{-1} \left(\frac{[\gamma^{-1}(\beta, y\Gamma(\beta))]^{1/\alpha}}{1 + [\gamma^{-1}(\beta, y\Gamma(\beta))]^{1/\alpha}}, \xi \right), \quad y \in (0, 1).$$

The median of the GG-G family is given by $M_{ed} = Q(0.5, \xi)$.

3.3 Shape of the crucial functions

The shapes of the pdf and the hrf of the GG-G family can be described analytically. The critical points of the pdf $f(x; \xi)$ given by (1.2) are the roots of the following equation:

$$\frac{g^x(x; \xi)}{g(x; \xi)} + (\alpha\beta - 1) \frac{g(x; \xi)}{G(x; \xi)} + (\alpha\beta + 1) \frac{g(x; \xi)}{1 - G(x; \xi)} - \alpha \left\{ \frac{G(x; \xi)}{1 - G(x; \xi)} \right\}^{\alpha-1} \frac{g(x; \xi)}{\{1 - G(x; \xi)\}^2} = 0, \quad (3.1)$$

where $g^x(x; \xi) = \partial g(x; \xi)/\partial x$. The critical points of $h(x)$ given by (1.3) are obtained from the following equation:

$$\begin{aligned} & \frac{g^x(x; \xi)}{g(x; \xi)} + (\alpha\beta - 1) \frac{g(x; \xi)}{G(x; \xi)} + (\alpha\beta + 1) \frac{g(x; \xi)}{1 - G(x; \xi)} - \alpha \left\{ \frac{G(x; \xi)}{1 - G(x; \xi)} \right\}^{\alpha-1} \frac{g(x; \xi)}{\{1 - G(x; \xi)\}^2} \\ & + \frac{\alpha g(x; \xi) G^{\alpha\beta-1}(x; \xi)}{(1 - G(x; \xi))^{\alpha\beta+1} \left\{ \Gamma(\beta) - \gamma \left[\beta, \left(\frac{G(x; \xi)}{1 - G(x; \xi)} \right)^\alpha \right] \right\}} \exp \left\{ - \left[\frac{G(x; \xi)}{1 - G(x; \xi)} \right]^\alpha \right\} = 0. \end{aligned} \quad (3.2)$$

By using most of the symbolic computation software platforms, we can examine the equations (3.1) and (3.2) to determine the local maximums and minimums and inflexion points.

3.4 Mixture representation

In this subsection, we will give the mixture representations of the GG-G pdf and GG-G cdf in terms of exp-G distribution of the baseline distribution, which will be useful for the derivation of further properties.

First of all, let us recall that the exponential series expansion is given by, for any $x \in \mathbb{R}$,

$$\exp(-ax) = \sum_{i=0}^{+\infty} \frac{(-1)^i a^i}{i!} x^i. \quad (3.3)$$

On the other side, the generalized binomial series expansion is given by, for any x such that $|x| < 1$,

$$(1 - x)^b = \sum_{j=0}^{+\infty} \binom{b}{j} (-1)^j x^j, \quad (3.4)$$

where $\binom{b}{j} = \frac{b(b-1)\dots(b-j+1)}{j!}$. These two formulas will be useful in our mathematical developments. Using the series (3.3) and (3.4), the pdf $f(x; \alpha, \beta, \xi)$ given by (1.2) becomes

$$f(x; \alpha, \beta, \xi) = \sum_{i,j=0}^{+\infty} u_{i,j} g(x; \xi) G^{\alpha(\beta+i)+j-1}(x; \xi),$$

with $u_{i,j} = \frac{\alpha}{\Gamma(\beta)} \binom{-\alpha(\beta+i)-1}{j} \frac{(-1)^{i+j}}{i!}$. Now, by writing $G^{\alpha(\beta+i)+j-1}(x; \xi) = [1-(1-G(x; \xi))]^{\alpha(\beta+i)+j-1}$, we get

$$f(x; \alpha, \beta, \xi) = \sum_{i,j,m=0}^{+\infty} \sum_{\ell=m}^{+\infty} w_{i,j,\ell,m} g(x; \xi) G^m(x; \xi),$$

where $w_{i,j,\ell,m} = u_{i,j} \binom{\alpha(\beta+i)+j-1}{\ell} \binom{\ell}{m} (-1)^{\ell+m}$. Rewriting the above equation we have

$$f(x; \alpha, \beta, \xi) = \sum_{m=0}^{+\infty} a_m h_{m+1}(x; \xi), \quad (3.5)$$

where

$$a_m = \frac{\alpha}{\Gamma(\beta)} \frac{(-1)^m}{m+1} \sum_{i,j=0}^{+\infty} \sum_{\ell=m}^{+\infty} \binom{-\alpha(\beta+i)-1}{j} \binom{\alpha(\beta+i)+j-1}{\ell} \binom{\ell}{m} \frac{(-1)^{i+j+\ell}}{i!}$$

and $h_{m+1}(x; \xi) = (m+1) g(x; \xi) G^m(x; \xi)$.

Moreover, by integrating this equation with respect to x , the cdf of GG-G family is given by

$$F(x) = \sum_{m=0}^{+\infty} a_m H_{m+1}(x; \xi), \quad (3.6)$$

where $H_{m+1}(x; \xi) = G^{m+1}(x; \xi)$. The expression in (3.5) and (3.6) are the infinite mixtures representation of the GG-G pdf in terms of functions (pdf and cdf) of the exp-G distribution.

In the context of the GGFr distribution: We can express $f(x; \alpha, \beta, \xi)$ as (3.5) with $h_{m+1}(x; \xi) = (m+1) (a^b b/x^{b+1}) \exp[-(m+1)(a/x)^b]$, $\xi = (a, b)$, $x, a, b > 0$.

3.5 Moments and moment generating function

In this subsection, we give the explicit expression for the r th moment, r th incomplete moment, moment generating function and the mean deviation about the mean.

The r th moment of GG-G family can be obtained by using following formula

$$\mu'_r = \int_{-\infty}^{+\infty} x^r f(x; \alpha, \beta, \xi) dx,$$

where $f(x; \alpha, \beta, \xi)$ is given by (1.2). It follows from the infinite mixture representation given by (3.5) that

$$\mu'_r = \sum_{m=0}^{+\infty} a_m \tau_m^r,$$

where

$$\tau_m^r = \int_{-\infty}^{+\infty} x^r h_{m+1}(x; \xi) dx \quad (3.7)$$

and $h_{m+1}(x; \xi) = (m + 1) g(x; \xi) G^m(x; \xi)$.

In the context of the GGF_r distribution: One can show that, for $r < b$, we have $\tau_m^r = (m + 1)^{r/b} a^r \Gamma(1 - r/b)$.

Similarly, the r th incomplete moment of GG-G family is given by

$$T^r(x; \alpha, \beta, \xi) = \sum_{m=0}^{+\infty} a_m \Delta_m^r(x; \xi),$$

where

$$\Delta_m^r(x; \xi) = \int_{-\infty}^x t^r h_{m+1}(t; \xi) dt. \quad (3.8)$$

In the context of the GGF_r distribution: One can show that, for $r < b$, we have $\Delta_m^r(x; \xi) = (m + 1)^{r/b} a^r \Gamma(1 - r/b, (m + 1)a^b x^{-b})$, $\xi = (a, b)$, $a, b > 0$, where $\Gamma(s, x)$ denotes the upper incomplete gamma function defined by $\Gamma(s, x) = \int_x^{+\infty} t^{s-1} e^{-t} dt$, $s, x > 0$.

The mean deviation about the mean of the GG-G family is given by

$$D_1 = 2 \mu'_1 F(\mu'_1; \alpha, \beta, \xi) - 2 T^1(\mu'_1; \alpha, \beta, \xi)$$

and the mean deviation about the median of the GG-G family is given by

$$D_{82} = \mu'_1 F(\mu'_1; \alpha, \beta, \xi) - 2 T^1(M_{ed}; \alpha, \beta, \xi).$$

The moment generating function of GG-G family is given by

$$M_0(t; \alpha, \beta, \xi) = \sum_{m=0}^{+\infty} a_m M_x(t; \xi),$$

where

$$M_x(t; \xi) = \int_{-\infty}^{+\infty} \exp(tx) h_{m+1}(x; \xi) dx. \quad (3.9)$$

Note that the integrals in (3.7), (3.8) and (3.9) only depends for any choice of base line distribution.

3.6 Reliability parameter

In the context of reliability, the stress-strength model defines the life of a element which has a random strength X_1 that is subjected to an accidental stress X_2 . The component fails at the instant that the stress applied to it exceeds the strength, and the component will function suitably whenever $X_1 > X_2$. Hence, by modeling X_1 and X_2 as random variables, the probability $R = P(X_2 < X_1)$ is a measure of components reliability. It has many applications especially in the area of reliability and engineering. In what follows, we derive the reliability R when X_1 and X_2 are independent and

identically distributed and belongs to the GG-G family with same parameters, with pdf $f_1(x; \alpha, \beta, \xi)$ for X_1 and with cdf $F_2(x; \alpha, \beta, \xi)$ for X_2 . From (1.2) and (1.1), we can write

$$R = \int_{-\infty}^{+\infty} f_1(x; \alpha, \beta, \xi) F_2(x; \alpha, \beta, \xi) dx.$$

Using the mixture representations of the pdf and the cdf given by (3.5) and (3.6), we have

$$R = \sum_{m=0}^{+\infty} \sum_{k=0}^{+\infty} a_m a_k \int_{-\infty}^{+\infty} h_{m+1}(x; \xi) H_{k+1}(x; \xi) dx.$$

Since $h_{m+1}(x; \xi) H_{k+1}(x; \xi) = (m+1) g(x; \xi) G^{m+k+1}(x; \xi)$, we have

$$R = \sum_{m=0}^{+\infty} \sum_{k=0}^{+\infty} a_m a_k (m+1) \int_{-\infty}^{+\infty} g(x; \xi) G^{m+k+1}(x; \xi) dx.$$

Let us observe that the above integral depends only on the choice of baseline distribution.

In the context of the GGFr distribution: One can show that $\int_{-\infty}^{+\infty} g(x; \xi) G^{m+k+1}(x; \xi) dx = 1/(m+k+2)$.

3.7 Entropies

Let $\delta > 0$ with $\delta \neq 1$. The Rényi entropy of the GG-G family is given by

$$I_\delta = \frac{1}{1-\delta} \log \left[\int_{-\infty}^{+\infty} f^\delta(x; \alpha, \beta, \xi) dx \right]. \quad (3.10)$$

Using (3.3) and (3.4), we have

$$f^\delta(x; \alpha, \beta, \xi) = \left(\frac{\alpha}{\Gamma(\beta)} \right)^\delta \sum_{i,j=0}^{+\infty} u_{i,j,\delta} g^\delta(x; \xi) G^{j+\delta(\alpha\beta-1)+\alpha i}(x; \xi),$$

where $u_{i,j,\delta} = \binom{-\alpha i - \delta(\alpha\beta+1)}{j} \frac{(-1)^{i+1} \delta^i}{i!}$. Now equation (3.10) becomes

$$I_\delta = \frac{1}{1-\delta} \left[\delta \log \alpha - \delta \log \Gamma(\beta) + \log \left[\sum_{i,j=0}^{+\infty} u_{i,j,\delta} \int_{-\infty}^{+\infty} g^\delta(x; \xi) G^{j+\delta(\alpha\beta-1)+\alpha i}(x; \xi) dx \right] \right].$$

We observe that I_δ depends only for any choice of base line distribution.

The δ -entropy is defined by

$$H_\delta = \frac{1}{\delta-1} \log \left[1 - \int_{-\infty}^{+\infty} [f(x; \alpha, \beta, \xi)]^\delta dx \right].$$

So we have

$$H_\delta = \frac{1}{\delta-1} \log \left[1 - \left(\frac{\alpha}{\Gamma(\beta)} \right)^\delta \sum_{i,j=0}^{+\infty} u_{i,j,\delta} \int_{-\infty}^{+\infty} g^\delta(x; \xi) G^{j+\delta(\alpha\beta-1)+\alpha i}(x; \xi) dx \right].$$

Finally, the Shannon entropy of GG-G family is defined by

$S = - \int_{-\infty}^{+\infty} \log[f(x; \alpha, \beta, \xi)] f(x; \alpha, \beta, \xi) dx$. It is in fact a particular case of the Rényi entropy when δ tends to 1^+ .

3.8 Order statistics

Let X_1, \dots, X_n be the random sample population with pdf given by (1.2), then the pdf of the i th order statistic is given by

$$f_{i:n}(x; \alpha, \beta, \xi) = \frac{n!}{(i-1)!(n-i)!} f(x; \alpha, \beta, \xi) F^{i-1}(x; \alpha, \beta, \xi) [1 - F(x; \alpha, \beta, \xi)]^{n-i}.$$

Using series expansion in equation (3.4) we have

$$f_{i:n}(x; \alpha, \beta, \xi) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} \binom{n-i}{j} (-1)^j f(x; \alpha, \beta, \xi) F^{j+i-1}(x; \alpha, \beta, \xi). \quad (3.11)$$

By the infinite mixture representation in (3.5) and (3.6), we obtain

$$f_{i:n}(x; \alpha, \beta, \xi) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} \binom{n-i}{j} (-1)^j \sum_{m=0}^{+\infty} a_m h_{m+1}(x; \xi) \left[\sum_{k=0}^{+\infty} a_k H_{k+1}(x; \xi) \right]^{j+i-1}. \quad (3.12)$$

Using power series raised to a positive power (see Gradshteyn and Ryzhik (2000)), we get

$$\left(\sum_{k=0}^{+\infty} c_k x^k \right)^n = \sum_{k=0}^{+\infty} d_k x^k,$$

where $d_0 = c_0^n$ and $d_m = \frac{1}{m c_0} \sum_{k=1}^m (k(n+1) - m) c_k d_{m-k}$ for $m \geq 1$. Therefore

$$\left[\sum_{k=0}^{+\infty} a_k H_{k+1}(x; \xi) \right]^{j+i-1} = G^{j+i-1}(x; \xi) \sum_{k=0}^{+\infty} d_k G^k(x; \xi),$$

with $d_0 = a_0^{j+i-1}$ and $d_m = \frac{1}{m a_0} \sum_{k=1}^m (k(j+i) - m) a_k d_{m-k}$ for $m \geq 1$. Now equation (3.12) becomes

$$f_{i:n}(x; \alpha, \beta, \xi) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} \binom{n-i}{j} (-1)^j \sum_{m=0}^{+\infty} \sum_{k=0}^{+\infty} d_k a_m (m+1) g(x; \xi) G^{m+j+i+k-1}(x; \xi).$$

From this expression, several mathematical properties can be obtained, as moments, moment generating function...

In the context of the GGFr distribution: One can observe that $g(x; \xi) G^{m+j+i+k-1}(x; \xi) = (a^b b/x^{b+1}) \exp \left[-(m+j+i+k) \left(\frac{a}{x} \right)^b \right] = (1/(m+j+i+k)) u_{m+j+i+k}(x)$, where $u_{m+j+i+k}(x)$ denotes the pdf of the Fréchet distribution with parameters $(m+j+i+k)^{1/b} a$ and b . So the pdf of i th order statistics of the GGFr distribution can be expressed as a linear combination of Fréchet pdfs.

4 Estimation of GG-G family parameter

Let x_1, \dots, x_n be a random sample of size n from a population with pdf given by (1.2) with vector of parameters $\Theta = [\alpha, \beta, \xi]^T$. The log-likelihood function for the vector of parameters is given by

$$\begin{aligned} \ell(\Theta) = & n \log \alpha - n \log \Gamma(\beta) + \sum_{i=1}^n \log g(x_i, \xi) + (\alpha\beta - 1) \sum_{i=1}^n \log G(x_i, \xi) \\ & - (\alpha\beta + 1) \sum_{i=1}^n \log(1 - G(x_i; \xi)) - \sum_{i=1}^n \left(\frac{G(x_i, \xi)}{1 - G(x_i; \xi)} \right)^\alpha. \end{aligned}$$

The components of score vector $\Theta = [\alpha, \beta, \xi]^T$ are given by

$$\begin{aligned} U_\alpha = & \frac{n}{\alpha} + \beta \sum_{i=1}^n \log G(x_i, \xi) - \beta \sum_{i=1}^n \log(1 - G(x_i; \xi)) - \sum_{i=1}^n \left(\frac{G(x_i, \xi)}{1 - G(x_i; \xi)} \right)^\alpha \log \left(\frac{G(x_i, \xi)}{1 - G(x_i; \xi)} \right), \\ U_\beta = & -n \frac{d}{d\beta} \log \Gamma(\beta) + \alpha \sum_{i=1}^n \log G(x_i, \xi) - \alpha \sum_{i=1}^n \log(1 - G(x_i; \xi)), \\ U_\xi = & \sum_{i=1}^n \frac{g^\xi(x_i, \xi)}{g(x_i, \xi)} + (\alpha\beta - 1) \sum_{i=1}^n \frac{G^\xi(x_i, \xi)}{G(x_i, \xi)} + (\alpha\beta + 1) \sum_{i=1}^n \frac{G^\xi(x_i, \xi)}{1 - G(x_i, \xi)} \\ & - \alpha \sum_{i=1}^n \left[\frac{G(x_i, \xi)}{1 - G(x_i, \xi)} \right]^{\alpha-1} \frac{G^\xi(x_i, \xi)}{\{1 - G(x_i, \xi)\}^2}, \end{aligned}$$

where $G^\xi(x_i, \xi) = \partial G(x_i, \xi) / \partial \xi$ and $g^\xi(x_i, \xi) = \partial g(x_i, \xi) / \partial \xi$. Since these equations are nonlinear equations according to the parameters, they can not be solved analytically but can be solved numerically by any software like R-language or mathematica.

5 Log-generalized gamma Fréchet

In many applied area, the lifetimes are affected by explanatory variables such as the cholesterol level, blood sugar, gender and many other explanatory variables. Parametric survival models to estimate the survival functions for censored data are widely used. For example, recently Lanjoni *et al.* (2016) defined the extended Burr XII regression model and Pratavera *et al.* (2018) proposed the heteroscedastic odd log-logistic generalized gamma regression model for censored data. Thus, using the some approach adopted in these papers, a distribution obtained from the Log-generalized gamma Fréchet (LGGFr) distribution will be expressed in the form of the class of location-scale models with two additional parameters to the shape. In this way, we propose a model of regression location, scale and shape.

Let X be a random variable having the pdf (2.1). A class of regression models for location and scale is characterized by the fact that the random variable $Y = \log(X)$. The cdf and pdf of Y can be expressed in terms of the basic function as

$$G(y; \mu, \sigma) = \exp \left\{ - \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) \right\} \right\}, \quad g(y; \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) - \exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) \right\} \right\}.$$

Then the pdf (2.1) becomes:

$$f(y; \alpha, \beta, \mu, \sigma) = \frac{\alpha g(y) G(y)^{(\beta\alpha-1)} (1 - G(y))^{-(\beta\alpha+1)} \exp\{-[G(y)/(1 - G(y))]^\alpha\}}{\Gamma(\beta)}, \quad (5.1)$$

where $y, \mu \in \mathbb{R}$ and $\sigma > 0$, μ is the location parameter, σ is the scale parameter and α and β are shape parameters. Thus, if $X \sim \text{GGFr}(a, b, \alpha, \beta)$ then $Y = \log(X) \sim \text{LGGFr}(\mu, \sigma, \alpha, \beta)$. However, if $Y = \log(X) \sim \text{LGGFr}(\mu, \sigma, \alpha, \beta)$, the cdf and pdf of the standardized random variable $Z = (Y - \mu)/\sigma$ are given by

$$G(z) = \exp\{-\exp\{-z\}\}, \quad g(z) = \exp\{-z - \exp\{-z\}\}.$$

Then the pdf (2.1) becomes :

$$f(z; \alpha, \beta) = \frac{\alpha g(z) G(z)^{(\beta\alpha-1)} (1 - G(z))^{-(\beta\alpha+1)} \exp\{-[G(z)/(1 - G(z))]^\alpha\}}{\Gamma(\beta)}. \quad (5.2)$$

We write $Z \sim \text{LGGFr}(0, 1, \alpha, \beta)$.

In order to introduce a regression structure in the class of models (5.2), we assume that both parameters $\mu_i, \sigma_i, \alpha_i$ and β_i vary across observations through regression structures

$$y_i = \mu_i + \sigma_i z_i, \quad i = 1, \dots, n, \quad (5.3)$$

where the random error z_i has pdf (5.2), μ_i and σ_i are parameterized as

$$\mu_i = \mu_i(\boldsymbol{\theta}_1), \quad \sigma_i = \sigma_i(\boldsymbol{\theta}_2), \quad \alpha_i = \alpha_i(\boldsymbol{\theta}_3), \quad \beta_i = \beta_i(\boldsymbol{\theta}_4),$$

where $\boldsymbol{\theta}_1 = (\theta_{11}, \dots, \theta_{1p_1})^T$, $\boldsymbol{\theta}_2 = (\theta_{21}, \dots, \theta_{2p_2})^T$, $\boldsymbol{\theta}_3 = (\theta_{31}, \dots, \theta_{3p_3})^T$ and $\boldsymbol{\theta}_4 = (\theta_{41}, \dots, \theta_{4p_4})^T$. The usual systematic component for the location parameter is $\mu_i = \mathbf{x}_i^T \boldsymbol{\theta}_1$, where $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip_1})$ is a vector of known explanatory variables, i.e. $\boldsymbol{\mu} = \mathbf{X} \boldsymbol{\theta}_1$, and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ is a specified $n \times p_1$ matrix of full rank with $p_1 < n$. Analogously, we consider the systematic component $g(\sigma_i) = \eta_i = \mathbf{v}_i^T \boldsymbol{\theta}_2$ for the dispersion parameter, where $g(\cdot)$ is the dispersion link function, and $\mathbf{v}_i^T = (v_{i1}, \dots, v_{ip_2})$ is a vector of known explanatory variables. We have $g(\boldsymbol{\sigma}) = \boldsymbol{\eta} = \mathbf{V} \boldsymbol{\theta}_2$, where $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n)^T$, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ and $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_n)^T$ is a specified $n \times p_2$ matrix of full rank with $p_2 < n$. For α_i and β_i , we consider the systematic component analogous. We have $g(\boldsymbol{\alpha}) = \boldsymbol{\delta} = \mathbf{W} \boldsymbol{\theta}_3$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T$ and $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ is a specified $n \times p_3$ matrix of full rank with $p_3 < n$ and $g(\boldsymbol{\beta}) = \boldsymbol{\lambda} = \mathbf{S} \boldsymbol{\theta}_4$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)^T$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$ and $\mathbf{S} = (\mathbf{s}_1, \dots, \mathbf{s}_n)^T$ is a specified $n \times p_4$ matrix of full rank with $p_4 < n$. It is assumed that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$ and $\boldsymbol{\theta}_4$ are functionally independent and that $g(\cdot)$ is a known one-to-one continuously twice differentiable function.

5.1 Maximum Likelihood Estimation

Consider a sample $(y_1, \mathbf{x}_1, \mathbf{v}_1, \mathbf{w}_1, \mathbf{s}_1), \dots, (y_n, \mathbf{x}_n, \mathbf{v}_n, \mathbf{w}_n, \mathbf{s}_n)$ of n independent observations, where each random response is defined by $y_i = \min\{\log(x_i), \log(c_i)\}$. Here, the x_i 's are the failure times and the c_i 's are the censored times. We assume non-informative censoring such that the observed lifetimes and censoring times are independent. Let F and C be the sets of individuals for which

y_i is the log-lifetime or log-censoring, respectively. We can apply conventional likelihood estimation techniques to estimate the model parameters. The log-likelihood function for the vector of parameters. The log-likelihood function for the vector of parameters $\phi = (\beta_1^T, \beta_2^T, \beta_3^T, \beta_4^T)^T$ from model (5.3) has the form $l(\phi) = \sum_{i \in F} l_i(\phi) + \sum_{i \in C} l_i^{(c)}(\phi)$, where $l_i(\phi) = \log[f(y_i)]$, $l_i^{(c)}(\phi) = \log[S(y_i)]$, $f(y_i)$ is the pdf (5.1) and $S(y_i)$ is the survival function of Y_i , respectively. The maximum likelihood estimate (MLE) $\hat{\phi}$ of the vector of model parameters can be computed by maximizing the log-likelihood $l(\phi)$.

The asymptotic distribution of $(\hat{\phi} - \phi)$ is multivariate normal $N_{p_1+p_2+p_3+p_4}(0, K(\phi)^{-1})$ under standard regularity conditions, where $K(\phi)$ is the total expected information matrix. The asymptotic covariance matrix $K(\phi)^{-1}$ of $\hat{\phi}$ can be approximated by the inverse of the $(p_1 + p_2 + p_3 + p_4) \times (p_1 + p_2 + p_3 + p_4)$ observed information matrix $-\ddot{\mathbf{L}}(\phi)$. The elements of $-\ddot{\mathbf{L}}(\phi)$ can be evaluated numerically. The approximate multivariate normal distribution $N_{p_1+p_2+p_3+p_4}(0, -\ddot{\mathbf{L}}(\phi)^{-1})$ for $\hat{\phi}$ can be used in the classical way to construct approximate confidence intervals for the components of ϕ .

5.2 Simulation study

We now perform a Monte Carlo simulation study to assess the finite sample behavior of the MLEs. The results are obtained from 1,000 Monte Carlo simulations performed using the R software. In each replication, a random sample of size n is drawn from the LGGFr($\theta_1, \theta_2, \theta_3, \theta_4$) model and the parameters are estimated by maximum likelihood. The log-lifetimes denoted by $\log(x_1), \dots, \log(x_n)$ are generated from the LGGFr regression model (5.3), where $\mu_i = \theta_{10} + \theta_{11} x_i$, $\sigma_i = \exp(\theta_{20} + \theta_{21} x_i)$, $\alpha_i = \exp(\theta_{30} + \theta_{31} x_i)$ and $\beta_i = \exp(\theta_{40} + \theta_{41} x_i)$ and x_i is generated from a normal distribution $N(0, 0.65)$. Thus, we consider for the simulations with sample sizes $n = 100$, $n = 350$ and $n = 850$, and censoring percentages approximately equal to 0%, 10% and 20%. The values considered for the parameters are, $\theta_{10} = 5.6350$, $\theta_{11} = 0.4948$, $\theta_{20} = 0.1688$, $\theta_{21} = 0.5219$, $\theta_{30} = -0.5805$, $\theta_{31} = 0.0313$, $\theta_{40} = -2.6156$ and $\theta_{41} = 3.2250$. The survival times are generated considering the random censoring mechanism as follows:

- Generate $x_i \sim \text{normal}(1, 0.65)$.
- Generate $c \sim \text{uniform}(0, \tau)$, where τ denotes the proportion of censored observations.
- Generate $z \sim \text{LGGFr}(0, 1, \alpha_i, \alpha_i)$, the values from the pdf (5.2).
- Write $y^* = \mu_i x_i + \sigma_i z$.
- Set $y = \min(y^*, c)$.
- Create a vector κ of dimension n which receives 1's if $(y^* \leq c)$ and zero otherwise.

We fit the LGGFr regression model (5.3) to each generated data set, where $\mu_i = \theta_{10} + \theta_{11} x_i$, $\sigma_i = \exp(\theta_{20} + \theta_{21} x_i)$, $\alpha_i = \exp(\theta_{30} + \theta_{31} x_i)$ and $\beta_i = \exp(\theta_{40} + \theta_{41} x_i)$. From the simulations reported in Table 1, we can verify that the mean squared errors (MSEs) and bias of the MLEs of θ_{10} , θ_{11} , θ_{20} , θ_{21} , θ_{30} , θ_{31} , θ_{40} and θ_{41} decay toward zero when the sample size increases, as usually expected under first-order asymptotic theory. The mean estimates of the parameters tend

to be closer to the true parameter values when the sample size n increases. This fact supports that the asymptotic normal distribution provides an adequate approximation to the finite sample distribution of the estimates.

Table 1: Simulations for the LGGFr regression model for the parameters values $\theta_{10} = 5.6350$, $\theta_{11} = 0.4948$, $\theta_{20} = 0.1688$, $\theta_{21} = 0.5219$, $\theta_{30} = -0.5805$, $\theta_{31} = 0.0313$, $\theta_{40} = -2.6156$ and $\theta_{41} = 3.2250$.

n	θ	0% censored			10% censored			20% censored		
		AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
100	θ_{10}	5.6488	0.0138	0.1191	5.6612	0.0262	0.1423	5.7045	0.0695	0.1265
	θ_{11}	0.4844	-0.0104	0.0648	0.4682	-0.0266	0.0680	0.4439	-0.0509	0.0665
	θ_{20}	0.1212	-0.0476	0.0150	0.1203	-0.0485	0.0166	0.1252	-0.0436	0.0150
	θ_{21}	0.5294	0.0075	0.0079	0.5360	0.0141	0.0086	0.5382	0.0163	0.0098
	θ_{30}	-0.6982	-0.1177	0.1017	-0.6816	-0.1011	0.1039	-0.6601	-0.0796	0.1110
	θ_{31}	0.1097	0.0784	0.1426	0.0862	0.0549	0.1780	0.0297	-0.0016	0.1873
	θ_{40}	-2.6302	-0.0146	0.0908	-2.6663	-0.0507	0.1153	-2.7000	-0.0844	0.1554
	θ_{41}	3.2491	0.0241	0.1119	3.2953	0.0703	0.1487	3.3558	-0.0844	0.1817
350	θ_{10}	5.6583	0.0233	0.0155	5.6628	0.0278	0.0171	5.6746	0.0396	0.0222
	θ_{11}	0.4792	-0.0156	0.0091	0.4776	-0.0172	0.0106	0.4700	-0.0248	0.0129
	θ_{20}	0.1558	-0.0130	0.0023	0.1551	-0.0137	0.0024	0.1550	-0.0138	0.0026
	θ_{21}	0.5286	0.0067	0.0007	0.5284	0.0065	0.0009	0.5302	0.0083	0.0011
	θ_{30}	-0.6318	-0.0513	0.0190	-0.6122	-0.0317	0.0189	-0.6078	-0.0273	0.0209
	θ_{31}	0.0724	0.0411	0.0241	0.0493	0.0180	0.0238	0.0365	0.0052	0.0271
	θ_{40}	-2.6105	0.0051	0.0170	-2.6323	-0.0167	0.0203	-2.6478	-0.0322	0.0255
	θ_{41}	3.2236	-0.0014	0.0169	3.2433	0.0183	0.0199	3.2648	0.0398	0.0267
850	θ_{10}	5.6479	0.0129	0.0038	5.6522	0.0172	0.0043	5.6574	0.0224	0.0057
	θ_{11}	0.4861	-0.0087	0.0022	0.4843	-0.0105	0.0025	0.4814	-0.0134	0.0033
	θ_{20}	0.1620	-0.0068	0.0006	0.1605	-0.0083	0.0008	0.1599	-0.0089	0.0010
	θ_{21}	0.5260	0.0041	0.0001	0.5266	0.0047	0.0002	0.5272	0.0053	0.0002
	θ_{30}	-0.6118	-0.0313	0.0056	-0.6018	-0.0213	0.0050	-0.5963	-0.0158	0.0058
	θ_{31}	0.0591	0.0278	0.0067	0.0479	0.0166	0.0063	0.0377	0.0064	0.0080
	θ_{40}	-2.6081	0.0075	0.0046	-2.6258	-0.0102	0.0056	-2.6344	-0.0188	0.0085
	θ_{41}	3.2163	-0.0087	0.0048	3.2310	0.0060	0.0054	3.2421	0.0171	0.0080

6 Residual analysis

The objective of the analysis of the residuals is to verify the adequacy of the model for a given data set, which includes the investigation of intrinsic characteristics in the data. In order to check some of these characteristics, for example, outliers, several approaches have been proposed by Cox and Snell (1968), Cook and Weisberg (1982), Ortega *et al.* (2008) and Silva *et al.* (2011). In the context of survival analysis, the deviance residuals have been more widely used because they take into account the information of censored times (Silva *et al.* (2011)). Thus, the plot of the deviance

residuals versus the observed times provides a way to test the adequacy of the fitted model and to detect atypical observations.

If the model is appropriate, the martingale and modified deviance residuals must present a random behavior around zero. The plots of the residuals, martingale or modified deviance residuals versus the adjusted values provide a simple way to verify the adequacy of the model and to detect outliers. Atkinson (1985) suggested the construction of envelopes to enable better interpretation of the normal probability plot of the residuals. These envelopes are simulated confidence bands that contain the residuals, such that if the model is well-fitted, the majority of points will be within these bands and randomly distributed.

We perform a simulation study to assess the accuracy of the MLEs of the parameters in the LGGFr regression model with censored data, and also to investigate the behavior of the empirical distribution of the martingale and deviance residuals. For the simulation study, we generate the variables z_1, \dots, z_n from the LGGFr distribution given by (5.2).

So, 1000 samples are generated for each scenario presented in Subsection 5.2 as well as the algorithm for generating the survival times considering censored.

In Figure 3 we display the plots of the residuals versus the expected values of the order statistics of the standard normal distribution. These plots are known as the normal probability plots and serve to assess the departure from the normality assumption of the residuals Weisberg (2005). Therefore, the following interpretation is obtained from these plots: the empirical distribution of the deviance residuals agrees with the standard normal distribution when the sample size increases.

7 Data Analysis

7.1 Analysis of three data sets

Here, we present three applications to real data to illustrate the potentiality of the GG-G family. To compare its performance, we consider its sub model Generalize Gamma Fréchet (GGFr) to other competitive models given in Table 2.

Table 2: The competitive models of the GGFr distributions.

Distribution	Author(s)
Generalized Odd Gamma-Fréchet (GOFr)	Hosseini <i>et al.</i> (2018)
Log-gamma generated Fréchet (LGFr)	Amini <i>al.</i> (2014)
Kumaraswamy Fréchet (KwFr)	Mead and Abd-Eltawab (2014)
Beta Fréchet (BFR)	Nadarajah and Gupta (2004)
Exponentiated Fréchet (EFr)	Nadarajah and Kotz (2003)

We consider the $-\hat{\ell}$ (where $\hat{\ell}$ denotes the maximized log-likelihood), AIC (Akaike information criterion), BIC (Bayesian information criterion), CVM (Cramér-Von Mises), AD (Anderson-Darling) and KS (Kolmogorov Smirnov with its p-value (PV)) statistics to compare the fitted distributions. The results in this section are obtained using the R PROGRAM.

The first data set (Data Set 1) is from Xu *at al.* (2003) and it represents the time to failure (103 h) of turbocharger of one type of engine. Recently used by Alzaatreh *et al.* (2015).

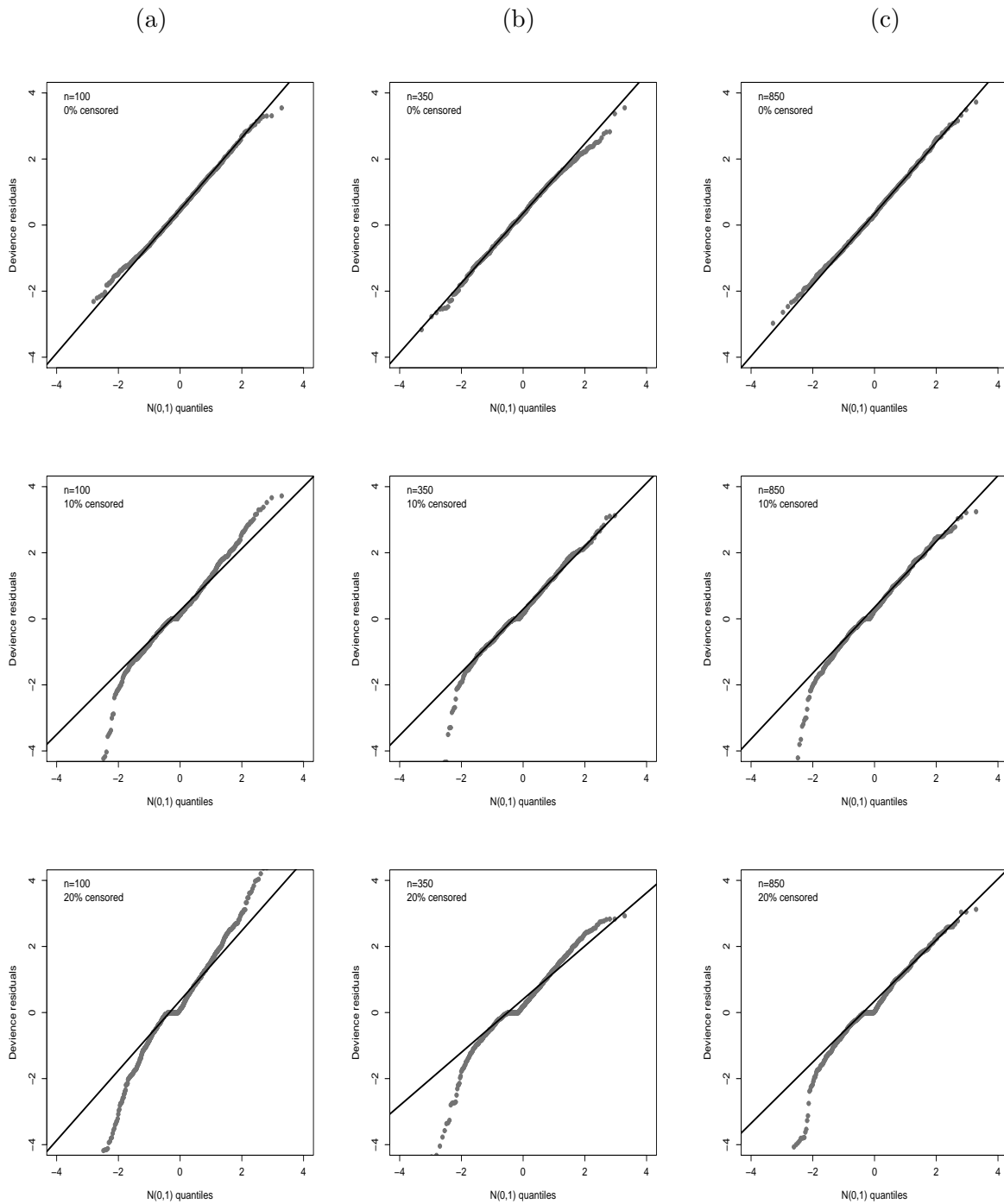


Figure 3: Normal probability plots for the deviance residuals. (a) Sample size $n = 100$ for censored percentages of (0%, 10% and 20%). (b) Sample size $n = 350$ for censored percentages of (0%, 10% and 20%). (c) Sample size $n = 850$ for censored percentages of (0%, 10% and 20%).

The time to failure of turbocharger data ($n = 40$)

1.6 3.5 4.8 5.4 6.0 6.5 7.0 7.3 7.7 8.0 8.4 2.0 3.9 5.0 5.6 6.1 6.5 7.1 7.3 7.8 8.1 8.4 2.6 4.5 5.1 5.8 6.3
6.7 7.3 7.7 7.9 8.3 8.5 3.0 4.6 5.3 6.0 8.7 8.8 9.0.

The second data set (Data Set 2) was obtained in Proschan (2000) 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), Gupta and Kundu (2001), Kus (2007) and De Andrade *al.* (2017)), 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.

The third data set (Data Set 3) represents the annual maximum temperatures at Oxford and Worthing in England for the period of 1901-1980. Chandler and Bate (2007) used the generalized extreme value distribution to model the annual maximum temperatures. Recently study this data by Alzaatreh *et al.* (2015).

The annual maximum temperatures data ($n = 80$)

75, 92, 87, 86, 85, 95, 84, 87, 86, 82, 77, 89, 79, 83, 79, 85, 89, 84, 84, 82, 86, 81, 84, 84, 87, 89, 80, 86, 85, 84, 89, 80, 87, 84, 85, 82, 86, 87, 86, 89, 90, 90, 91, 81, 85, 79, 83, 93, 87, 83, 88, 90, 83, 82, 80, 81, 95, 89, 86, 89, 87, 92, 89, 87, 87, 83, 89, 88, 84, 84, 77, 85, 77, 91, 94, 80, 80, 85, 83, 88.

Tables 3, 5 and 7 provide the values of goodness-of-fit measures for the GGFr model and other fitted models, whereas the MLEs and their corresponding standard errors (SEs) (in parentheses) are listed in Tables 4, 6 and 8 respectively. Some plots of the fitted GGFr pdf and cdf and other fitted pdfs and cdfs are displayed in Figure 4. From Tables 3, 5 and 7 we can see that GGFr perform best performance than other competitive models and plots in Figure 4 also reveal that the GGFr distribution yields the best fit and it can be considered a very competitive model to other fitted distributions.

Table 3: Goodness-of-fit measures for Data Set 1.

Model	$-\widehat{\ell}$	AIC	BIC	CVM	AD	KS	PV
GGFr	77.6861	163.3722	170.1278	0.0147	0.1128	0.0627	0.9975
GOFr	87.8787	183.7575	190.5130	0.2201	1.4468	0.1353	0.4562
LGFr	85.8029	179.6059	186.3614	0.1659	1.1365	0.1434	0.3831
KwFr	86.1506	180.3014	187.0569	0.1710	1.1615	0.1218	0.5928
BFr	91.2658	190.5317	197.2872	0.3161	1.9815	0.1523	0.3109
EFr	101.5918	209.1836	214.2502	0.6067	3.4798	0.2438	0.0172

Table 4: MLEs and SEs (in parentheses) for Data Set 1.

Model	Estimates			
GGFr	39.9826	0.0554	5.4318	0.7286
(α, β, a, b)	(35.4823)	(0.0508)	(1.8028)	(0.4728)
GOFr	13.0778	4.6917	0.1739	1.1748
(α, β, a, b)	(14.14513)	(3.4168)	(0.8502)	(0.3630)
LGFr	0.2839	135.0816	41.7514	1.0413
(α, β, a, b)	(0.1061)	(78.3133)	(0.0755)	(0.0919)
KwFr	5.2239	554.6118	9.8825	0.5019
(a, b, λ, α)	(36.8518)	(222.0773)	(138.8463)	(0.0927)
BFr	71.1502	259.1749	74.9748	0.1685
(a, b, λ, α)	(2.1131)	(111.2123)	(38.1707)	(0.0323)
EFr	1.6108	1.9440	7.9296	
(a, b, λ)	(0.1674)	(0.2033)	(3.3140)	

Table 5: Goodness-of-fit measures for Data Set 2.

Model	$-\widehat{\ell}$	AIC	BIC	CVM	AD	KS	PV
GGFr	1174.049	2352.098	2339.543	0.0299	0.2265	0.0348	0.9579
GOFr	1175.01	2358.021	2371.466	0.0330	0.2710	0.0359	0.9460
LGFr	1180.145	2368.289	2381.735	0.1203	0.8585	0.0583	0.4623
KwFr	1174.79	2357.581	2371.026	0.0306	0.2466	0.0380	0.9176
BFr	1179.905	2367.809	2381.254	0.1158	0.8285	0.0572	0.4875
EFr	1210.316	2426.632	2436.716	0.7114	4.5756	0.1024	0.0228

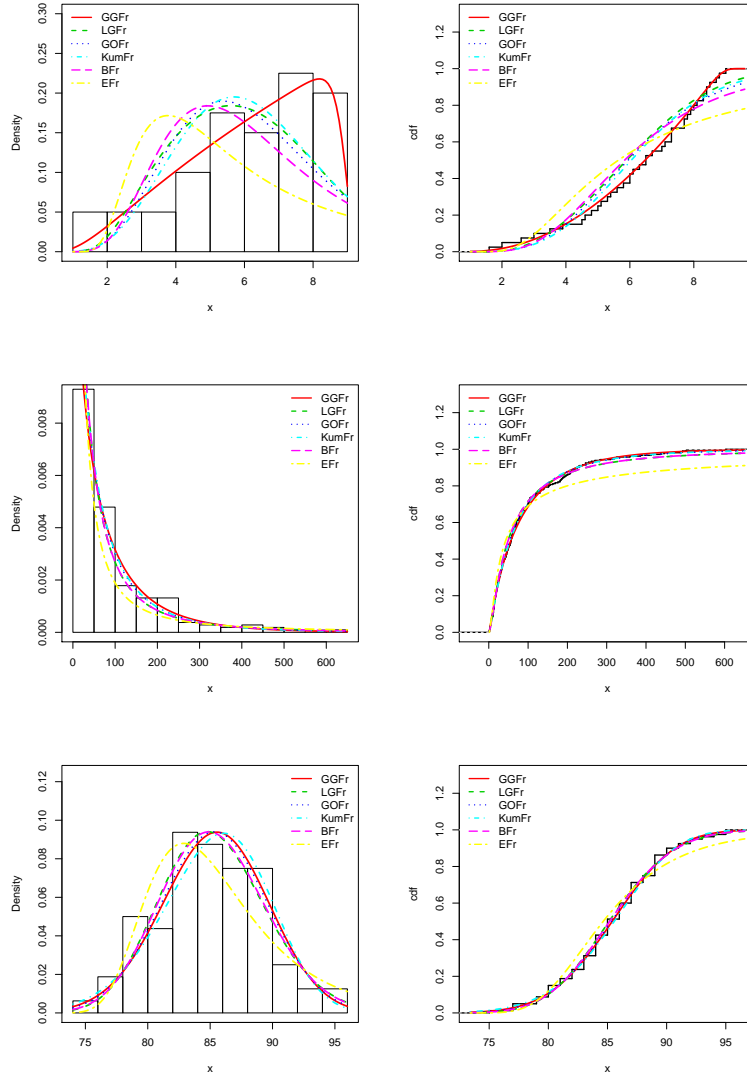


Figure 4: Plots of estimated pdfs and cdfs for Data Sets 1, 2 and 3 respectively.

Table 6: MLEs and SEs (in parentheses) for Data Set 2.

Model	Estimates			
GGFr	1.9151	0.6879	50.9094	0.3914
(α, β, a, b)	(0.4837)	(0.5710)	(46.1448)	(0.1184)
GOFr	10.2114	2.8055	0.0193	0.4308
(α, β, a, b)	(0.4837)	(0.5710)	(46.1448)	(0.1184)
LGFr	74.8549	165.0835	57.9521	0.0724
(α, β, a, b)	(86.4516)	(162.3931)	(29.0281)	(0.0352)
KwFr	7.9721	152.5672	5.8233	0.1756
(a, b, λ, α)	(2.5710)	(170.4326)	(9.5828)	(0.0365)
BFr	123.6822	219.1236	69.6975	0.0565
(a, b, λ, α)	(127.1976)	(204.1889)	(11.2324)	(0.0312)
EFr	3.9932	0.7360	3.9688	
(a, b, λ)	(21.5108)	(0.0345)	(4.4654)	

Table 7: Goodness-of-fit measures for Data Set 3.

Model	$-\hat{\ell}$	AIC	BIC	CVM	AD	KS	PV
GGFr	229.1030	465.2059	474.7340	0.0489	0.3041	0.0628	0.8993
GOFr	229.0553	466.1105	475.6386	0.0507	0.3053	0.0640	0.8978
LGFr	229.4772	466.9545	476.4826	0.0662	0.3923	0.0799	0.6855
KwFr	229.0371	466.0743	475.6024	0.0499	0.3057	0.0688	0.8198
BFr	229.3539	466.7079	476.2360	0.0623	0.3694	0.0772	0.7269
EFr	236.7190	479.4380	486.5841	0.2524	1.5095	0.1342	0.1120

Table 8: MLEs and SEs (in parentheses) for Data Set 3.

Model	Estimates			
GGFr	18.2831	14.3326	9.7718	0.2178
(α, β, a, b)	(11.7818)	(13.2666)	(5.1681)	(0.03312)
GOFr	40.2115	34.2970	10.1598	3.3992
(α, β, a, b)	(16.4392)	(20.8552)	(11.5038)	(0.1216)
LGFr	85.6583	111.0738	69.3561	2.3037
(α, β, a, b)	(3.1215)	(119.7371)	(12.2132)	(2.2619)
KwFr	0.6140	35.5674	117.4816	5.8279
(a, b, λ, α)	(6.9305)	(16.5385)	(28.7539)	(2.7076)
BFr	123.4322	239.6551	90.1387	1.3572
(a, b, λ, α)	(1.2342)	(146.6493)	(21.3231)	(1.6669)
EFr	69.0908	19.8504	38.9689	
(a, b, λ)	(38.8814)	(1.5900)	(8.5356)	

7.2 Regression model for turbine data

In this application, we use a real data set available in the book of (Lawless, 2003, page 262, Table 5.9) to study the LGGFr regression model. The model parameters are estimated by package GAMLSS in R. This data set presents an experiment designed to compare the performances of high-speed turbine engine bearings made out of five different compounds (McCool (1979)). The experiment tested 10 bearings of each type; the times to fatigue failure are given in units of millions of cycles. The analysis considering the LGGFr regression model is performed with the definition of dammy variables as follows: levels type I ($d_{i1} = 0, d_{i2} = 0, d_{i3} = 0$ and $d_{i4} = 0$), levels type II ($d_{i1} = 1, d_{i2} = 0, d_{i3} = 0$ and $d_{i4} = 0$), levels type III ($d_{i1} = 0, d_{i2} = 1, d_{i3} = 0$ and $d_{i4} = 0$), type IV ($d_{i1} = 0, d_{i2} = 0, d_{i3} = 1$ and $d_{i4} = 0$) and levels type V ($d_{i1} = 0, d_{i2} = 0, d_{i3} = 0$ and $d_{i4} = 1$).

The LGGFr regression model for the turbine data can be expressed as follows

$$y_i = \mu_i + \sigma_i z_i,$$

where z_1, \dots, z_{50} are independent random variables with pdf (5.2). We consider the three following configurations. The first configuration corresponds to the homoscedastic case: we consider the model parameters $\alpha, \beta, \mu_i, \sigma$, where $\mu_i = \theta_{10} + \theta_{11} d_{i1} + \theta_{12} d_{i2} + \theta_{13} d_{i3} + \theta_{14} d_{i4}$ and σ is a common variance: $\sigma = \sigma_1 = \dots = \sigma_{50}$. The second configuration corresponds to the heteroscedastic case: we consider the parameters $\alpha, \beta, \mu_i, \sigma_i$, where $\mu_i = \theta_{10} + \theta_{11} d_{i1} + \theta_{12} d_{i2} + \theta_{13} d_{i3} + \theta_{14} d_{i4}$ and $\sigma_i = \exp(\theta_{20} + \theta_{21} d_{i1} + \theta_{22} d_{i2} + \theta_{23} d_{i3} + \theta_{24} d_{i4})$. The third configuration is the more general; we consider the parameters:

$$\begin{aligned} \mu_i &= \theta_{10} + \theta_{11} d_{i1} + \theta_{12} d_{i2} + \theta_{13} d_{i3} + \theta_{14} d_{i4}, \quad \sigma_i = \exp(\theta_{20} + \theta_{21} d_{i1} + \theta_{22} d_{i2} + \theta_{23} d_{i3} + \theta_{24} d_{i4}), \\ \alpha_i &= \exp(\theta_{30} + \theta_{31} d_{i1} + \theta_{32} d_{i2} + \theta_{33} d_{i3} + \theta_{34} d_{i4}), \quad \beta_i = \exp(\theta_{40} + \theta_{41} d_{i1} + \theta_{42} d_{i2} + \theta_{43} d_{i3} + \theta_{44} d_{i4}). \end{aligned}$$

The MLEs for the LGGFr model are presented in Tables 9 and 10. Thus, when establishing a significance level of 5%, we note that the compounds type level is significant and should be used to model the location, scale and shape.

In order to see if the considered regression model is appropriate, the plot comparing the empirical survival function and estimated survival function for the LGGFr regression model is displayed in Figure 5 under the three presented configurations. We observe that LGGFr regression model shows a suitable fit. Figure 6 (a) presents the fitted hazard functions, (b) the index plot of the deviance residual for the turbine data and (c) the normal probability plot for the deviance component residual with envelopes from the fitted of LGGFr regression model to the turbine data. All these figures shows that the considered LGGFr regression model is appropriate.

Table 9: MLEs, SEs and p -values for the LGGFr regression model fitted to the turbine data.

Homoscedastic LGGFr regression model				Heteroscedastic LGGFr regression model			
Parameter	Estimate	SE	p -Value	Parameter	Estimate	SE	p -Value
$\log(\sigma)$	-0.4432	0.0780	-	$\log(\alpha)$	-0.297	0.119	-
$\log(\alpha)$	0.2504	0.1033	-	$\log(\beta)$	-0.0314	0.1149	-
$\log(\beta)$	0.0027	0.0919	-	θ_{10}	2.3104	0.1665	< 0.0001
θ_{10}	2.0162	0.1068	< 0.0001	θ_{11}	-0.4664	0.2065	0.0289
θ_{11}	-0.4858	0.1757	0.0082	θ_{12}	-0.1732	0.2101	0.4140
θ_{12}	-0.1621	0.1910	0.4004	θ_{13}	0.0422	0.2088	0.8408
θ_{13}	0.0126	0.1532	0.9344	θ_{14}	0.3663	0.1998	0.0734
θ_{14}	0.3804	0.2167	0.0860	θ_{20}	-0.3842	0.1983	0.0590
				θ_{21}	-0.3091	0.3122	0.3270
				θ_{22}	-0.2708	0.2913	0.3570
				θ_{23}	-0.2141	0.2654	0.4240
				θ_{24}	-0.3968	0.2740	0.1550

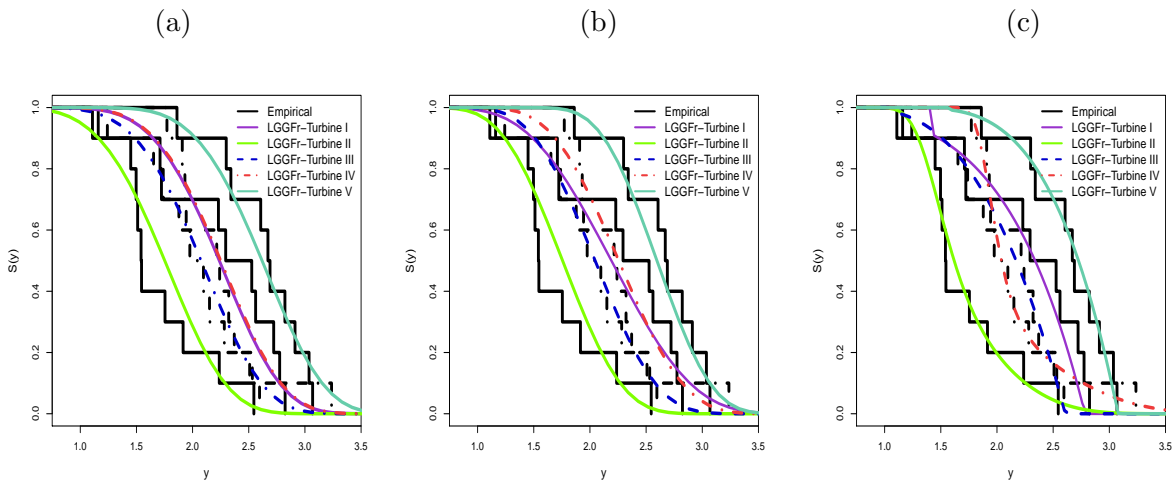


Figure 5: Estimated survival considering the LGGFr regression model for the three considered configurations in Table 9 and Table 10.

Table 10: MLEs, SEs and p -values for the LGGFr regression model fitted to the turbine data.

Location, scale and shape LGGFr regression model							
Parameter	Estimated	SE	p -Value	Parameter	Estimated	SE	p -Value
θ_{10}	2.2989	0.0883	< 0.0001	θ_{30}	-3.0051	0.3193	< 0.0001
θ_{11}	-0.2361	0.1228	0.0609	θ_{31}	2.10526	0.4235	< 0.0001
θ_{12}	-0.0443	0.1264	0.7272	θ_{32}	0.9625	0.4655	0.0444
θ_{13}	0.3523	0.1141	0.0034	θ_{33}	1.6414	0.4568	0.0008
θ_{14}	0.4168	0.1224	0.0014	θ_{34}	1.4204	0.4564	0.0032
θ_{20}	0.4083	0.1874	0.0347	θ_{40}	3.5051	0.2891	< 0.0001
θ_{21}	-1.6849	0.2164	< 0.0001	θ_{41}	-4.4531	0.4086	< 0.0001
θ_{22}	-0.4542	0.331	0.1769	θ_{42}	-1.3238	0.4216	0.0029
θ_{23}	-1.9999	0.2119	< 0.0001	θ_{43}	-5.2839	0.4363	< 0.0001
θ_{24}	-0.4597	0.3349	0.1766	θ_{44}	-1.5485	0.4164	0.0005

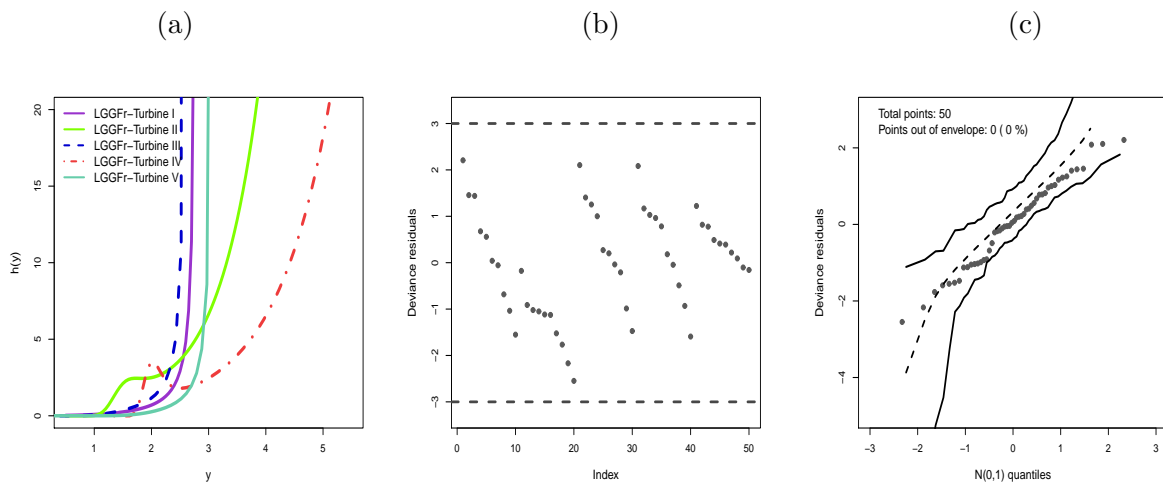


Figure 6: (a) Fitted hazard functions, (b) index plot and (c) normal probability plot for the LGGFr regression model.

References

- Alexander, C., Cordeiro, G.M., Ortega, E.M.M. and Sarabia, J.M. (2012). Generalized beta generated distributions. *Computational Statistics and Data Analysis*, 56, 1880-1897.
- Alzaatreh, A., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63-79.
- Alzaatreh, A., Lee, C. and Famoye, F. (2015). Family of generalized gamma distributions: Properties and applications. *Hacet. J. Math. Stat* , 45, 869-886.
- Alzaatreh, A., Lee, C. and Famoye, F. (2016). Family of generalized gamma distributions: Properties and applications. *Hacetepce Journal of Mathematics and Statistics*, 45, 869-886.
- Alzaghal, A., Lee, C. and Famoye, F. (2013). Exponentiated T-X family of distributions with some applications. *International Journal of Statistics and Probability*, 2, 31-49.
- Amini, M., MirMostafaei, S.M.T.K. and Ahmadi, J. (2014). Log-gamma-generated families of distributions, *Statistics*, 48, 4, 913-932.
- Atkinson, A.C. (1985). *Plots, Transformations, and Regression*. Oxford: Oxford University Press.
- Bourguignon, M., Silva, R.B. and Cordeiro, G.M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12, 1253-1268.
- Chandler, R.E. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood, *Biometrika*, 94, 1, 167-183.
- Chesneau, C., Bakouch, H. and Hussain, T. (2018). A new class of probability distributions via cosine and sine functions with applications, *Communications in Statistics: Simulation and Computation*, (to appear).
- Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, New York: Chapman and Hall.
- Cordeiro, G.M., Alizadeh, M., and Ortega, E.M.M. (2014). The exponentiated half-logistic family of distributions: Properties and applications. *Journal of Probability and Statistics*, Article ID 864396, 21 pages.
- 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 Silva, G. (2012). The beta extended Weibull family. *Journal of Probability and Statistical Science*, 10, 15-40.
- Cox, D.R. and Snell, E.J. (1968). A general definition of residuals, *J. R. Stat. Soc., Ser. B* 30, 248-265, Discussion 265-275.

- 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., Zea, L.M., Gomes-Silva, S. and Cordeiro, G.M. (2017). The gamma generalized pareto distribution with applications in survival analysis. *International Journal of Statistics and Probability*, 6, 141-156.
- Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Comm. Statist. Theory Methods*, 31, 497-512.
- Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Table of Integrals, Series and Products*, Academic Press, New York.
- Gupta, R.D. and Kundu, D. (2001). Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, 43, 117-130.
- Hosseini, B., Afshari, M. and Alizadeh, M. (2018). The Generalized Odd Gamma-G Family of Distributions: Properties and Applications, *Austrian Journal of Statistics*, 47, 69-89.
- Jamal, F., Nasir, M.A., Tahir, M.H. and Montazeri, N.H. (2017). The odd Burr-III family of distributions. *Journal of Statistics Applications and Probability*, 6, 105-122.
- Jones, M.C. (2004). Families of distributions arising from distributions of order statistics. *Test*, 13, 1-43.
- Kumar, D., Singh, U. and Singh, S. K. (2015). A New Distribution Using Sine Function- Its Application to Bladder Cancer Patients Data. *J. Stat. Appl. Pro.*, 4 (3), 417- 427.
- Kus, C. (2007). A New Lifetime Distribution. *Computational Statistics and Data Analysis*, 51, 4497-4509.
- Lanjoni, B.R., Ortega, E.M.M. and Cordeiro, G. (2016). Extended burr xii regression models: Theory and applications. *Journal of Agricultural Biological and Environmental Statistics*, 21, 203-224.
- Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.
- McCool, C.J. (1979). Distribution of *Cysticercus bovis* in lightly infected young cattle. *Aust. Vet. J.*, 55, 214-216.
- Mead, M.E. and Abd-Eltawab A.R. (2014). A note on Kumaraswamy Fréchet distribution. *Aust. J. Basic and Appl. Sci.*, 8, 294-300.
- Nadarajah, S. and Gupta, A.K. (2004). The beta Fréchet distribution. *Far East Journal of Theoretical Statistics*, 14, 15-24.
- Nadarajah, S. and Kotz, S. (2003). The exponentiated Fréchet distribution, *Statisticson the Internet*. <http://interstat.statjournals.net/YEAR/2003/articles/0312002.pdf>

- Ortega, E.M.M., Paula G.A. and Bolfarine, H. (2008). Deviance residuals in generalized loggamma regression models with censored observations. *Journal of Statistical Computation and Simulation*, 78, 747-764.
- Pescim, R.R., Cordeiro, G.M., Demetrio, C.G.B., Ortega, E.M.M. and Nadarajah, S. (2012). The new class of Kummer beta generalized distributions. *Statistics and Operations Research Transactions (SORT)*, 36, 153-180.
- Pratavia, F., Ortega, E.M., Cordeiro, G.M., Pescim, R.R. and Verssani, B.A. (2018). A new generalized odd log-logistic flexible Weibull regression model with applications in repairable systems. *Reliab. Eng. Syst. Saf.*, 176, 13-26.
- Proschan, F. (2000). Theoretical explanation of observed decreasing failure rate. *American Statistical Society* 1963, reprinted in *Technometrics*, 42, 1, 7-11.
- Ristić, M.M. and Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82, 1191-1206.
- Shaw. W. and Buckley. I. (2007). The alchemy of probability distributions: beyond Gram-Charlier expansions and a skewkurtotic-normal distribution from a rank transmutation map. Research report, King's College, London, U.K.
- Silva, G.O., Ortega, E.M.M. and Paula, G.A. (2011). Residuals for log-Burr XII regression models in survival analysis. *Journal of Applied and Statistics*, 38, 1435-1445.
- Torabi, H. and Montazari, N.H. (2012). The gamma-uniform distribution and its application. *Kybernetika*, 48, 16-30.
- Torabi, H. and Montazeri, N.H. (2014). The logistic-uniform distribution and its applications. *Communications in Statistics- Simulation and Computation*, 43, 2551-2569.
- Weisberg, S. (2005). *Applied linear regression* (3rd ed.). Hoboken, NJ: John Wiley and Sons, Inc.
- Xu, K., Xie, M., Tang, L.C. and Ho, S.L. (2003). Application of Neural Networks in forecasting Engine Systems Reliability. *Applied Soft Computing*, 2, 4, 255-268.
- Zografos, K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, 6, 344-362.