UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PÓS-GRADUAÇÃO EM ESTATÍSTICA

THIAGO ALEXANDRO NASCIMENTO DE ANDRADE

THE EXPONENTIATED GENERALIZED FAMILY OF CONTINUOUS DISTRIBUTIONS

Recife
2017

Thiago Alexandro Nascimento de Andrade

THE EXPONENTIATED GENERALIZED FAMILY OF CONTINUOUS DISTRIBUTIONS

Doctoral Thesis submitted to the Graduate Program in Statistics, Department of Statistics, Federal University of Pernambuco, as a partial requirement for obtaining a Ph.D. in Statistics.

Advisor: Ph.D. Gauss Moutinho Cordeiro

Recife

2017

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

A553e Andrade, Thiago Alexandro Nascimento de
The exponentiated generalized family of continuous distributions / Thiago Alexandro Nascimento de Andrade. – 2017.
122 f.: il., fig., tab.

Orientador: Gauss Moutinho Cordeiro.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências.

1. Estatística matemática. 2. Distribuições exponencializadas. I. Cordeiro, Gauss Moutinho (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- MEI 2017-49

THIAGO ALEXANDRO NASCIMENTO DE ANDRADE

THE EXPONENTIATED GENERALIZED FAMILY OF CONTINUOUS DISTRIBUTIONS

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 16 de janeiro de 2017.

BANCA EXAMINADORA

Prof. Gauss Moutinho Cordeiro
UFPE

Prof. Getúlio José Amorim do Amaral
UFPE

Prof. Alex Dias Ramos
UFPE

Prof. Marcelo Bourguignon Pereira
UFRN

Prof. Rodrigo Bernardo da Silva
UFPB

I dedicate this work to my grandmother Quiteria Ferreira do Nascimento, my mother Célia Maria Nascimento de Andrade, my wife Renata Cristina de Andrade and my mother-in-law Maria do Carmo. My eternal gratitude.

Acknowledgment

At this moment, it is difficult to think of gratitude and not start by thanking God. There is my God. To the God that I serve, a living God who always held my right hand, walked with me and blessed me saying day by day: “Son, do not fear, I am your God, who protects you and guides you”. To my God, to Abraham, to Isaac, and to Jacob, my thank you.

Well, now I begin thanking my fellow men. The list will be extensive and the jokes about "Thanks have gotten bigger than the thesis" will surely come. But to the joker: It’s my thesis!

I thank all my family, who have always been and will be my safe haven. Thank you for being with me along this walk. It was difficult, but we did it ... We have a doctor in the family! In particular, I thank my grandmother for her prayers and words of comfort.

I thank my wife, companion for this whole life, love as immortal as this thesis. Thank you so much for existing in my life. I love you and I will love you for a thousand years if I can.

I thank my great friend Marcelo Mesquita Monte. I love you my friend. You’re like a father to me.

I thank my advisor, Ph.D. Gauss Cordeiro, for the safe guidance and at the same time for the human welcome in the necessary moments. I am deeply grateful and proud to have been mentored by one of the world’s greatest researchers in his field. Professor Gauss, you are to me a model of person and researcher. Thank you so much for everything.

I thank my collaborators Marcelo Bourguignon, Heloisa Rodrigues, Gustavo Pinho, Morad Alizadeh and Frank Gomes-Silva. Thank you so much for the many scientific partnerships and shared experiences.

I thank all the professors from the statistics department at the Federal University of Pernambuco, in particular the teacher Audrey Cysneiros, with whom I had the opportunity to work in the master's degree. I also thank the teachers Francisco Cribari, Leandro Rêgo and Klaus Vasconcellos for their fantastic courses in which I have learned a lot.

I thank my friends Marcelo Bourguignon, Heloisa Rodrigues, Gustavo Pinho, Luz Milena, Ana Hermìnia, Sadraque Lucena, Alberto Lopes, Diêgo Maciel and Flàvio J. Marques.

I thank to Valéria Bittencourt, for the competence, dedication and affection that have become trademarks of her work with the graduate students.

Thanks to the CAPES for the financial support.

If any of you lacks wisdom, you should ask God, who gives generously to all without finding fault, and it will be given to you. But when you ask, you must believe and not doubt, because the one who doubts is like a wave of the sea, blown and tossed by the wind. That person should not expect to receive anything from the Lord.
(James 1; 5-7).

Commit your way to the Lord; trust in him and he will do this. He will make your righteous reward shine like the dawn, your vindication like the noonday sun.
(Psalms 37; 5-6).

Abstract

Statistical and applied researchers have shown great interest in building new extended probability models that generalize well-known distributions and are more flexible for data modeling in many fields of applications. Probably, one of the most popular ways to extend well-known models is to consider distribution generators. A class of univariate distributions called the *exponentiated generalized* (EG for short) class was recently proposed in the literature. We believe that the EG class of distributions can be widely used to generalize continuous distributions. For this reason, the present doctoral thesis presents some extended models using the EG class. For each model presented in the chapters that follow, we provide a complete mathematical treatment, simulation studies and applications to real data that illustrate the usefulness of the models under study.

Keywords: Exponentiated generalized class. Generalized distribution. Probability distribution.

Resumo

Estatísticos e pesquisadores aplicados têm mostrado grande interesse em propor novos modelos de probabilidade estendidos que generalizam distribuições bem estabelecidas na literatura. Provavelmente, uma das formas mais populares de estender modelos bem conhecidos é considerando os chamados geradores de distribuições. Uma classe de distribuições univariadas chamada de classe *exponencializada generalizada* (EG) foi proposta recentemente na literatura. Acreditamos que a classe EG de distribuições pode ser amplamente utilizada para generalizar distribuições contínuas. Por esta razão, a presente tese de doutorado apresenta alguns modelos estendidos usando a classe EG. Para cada modelo apresentado nos capítulos que se seguem, fornecemos um tratamento matemático completo, estudos de simulação e aplicações a dados reais que ilustram a utilidade dos modelos em estudo.

Palavras-chave: Classe exponencializada generalizada. Distribuição generalizada. Distribuição de probabilidade.

List of Figures

List of Tables

Contents

CHAPTER 1

Introduction

The recent statistical literature offers a broad arsenal of univariate probability distributions that can be (and indeed are) widely used in data modeling in several fields of application. Notwithstanding, this large number of available distributions does not seem to deal with the huge variety of data arising from several fields such as medicine, engineering, demography, biology, actuarial, economics, finance, reliability, among others ([1]). Indeed, statistical and applied researchers have shown great interest in building new extended probability distributions, which are more flexible for data modeling ([2]). There are several ways described in the literature to extend well-known distributions ([3]). Probably, one of the most popular ways is to consider distribution generators.

In the generator approach, we refer to the following papers: [4] for the *Marshall-Olkin* class, [5] for the *beta* class, [6], [7] and [8] for the *gamma* class and [9] for the *Kumaraswamy* class of distributions. More recently, for any baseline cdf $G(x)$, and $x \in \mathbb{R}$, [10] defined the *exponentiated generalized* (EG) class of distributions with two extra shape parameters.

We believe that the addition of parameters to the well-known models may generate new distributions with great adjustment capability. We also believe that the EG class of distributions proposed by [10] can be widely used to generalize continuous distributions. For this reason, the present doctoral thesis introduces some extended models using the methodology proposed in [10]. For each model presented in the chapters that follow, we provide a complete mathematical treatment, simulation studies and applications to real data.

Besides this introduction, the present doctoral thesis is organized as follows. In Chapter 2, we study the *exponentiated generalized Gumbel* distribution, henceforth referred as EGGu distribution, that generalizes the *Gumbel* model. We discuss several properties for this model, including shapes of its density function, explicit expressions for the ordinary moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves, Rényi entropy and the density function of the order statistics. The method of maximum likelihood is considered to estimate the model parameters and we determine the observed information matrix. We provide a Monte Carlo simulation study to evaluate the maximum likelihood estimates (MLEs) of the model parameters and two applications to real data to illustrate the importance of the EGGu model. In Chapter 3, we study the *exponentiated generalized extended exponential* (EGEE for short) distribution that generalizes the *extended exponential* (EE) model introduced by [11]. We have shown that the hazard function of the EGEE model can take the classic shapes: bathtub, inverted bathtub, increasing, decreasing and constant, among others. We also investigate many of its mathematical properties such as a representation for the density function as a double linear combination of Erlang densities, explicit expressions for the quantile function, ordinary and incomplete moments, mean deviations, Bonferroni and Lorenz curves, generating function, Rényi entropy, density of order statistics and reliability. We use the maximum likelihood method to estimate the model parameters and provide the elements of the score vector. Two applications to real data illustrates the flexibility of the proposed model. In Chapter 4, we introduce a new two-parameter lifetime model, called the *exponentiated generalized standard half-logistic* (EGSHL) distribution, and study some of its general structural properties. This distribution extends the *half-logistic* distribution proposed by Balakrishnan in the eighties. We provide explicit expressions for the density function, ordinary and incomplete moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves, and order statistics. Our formulas are manageable using modern computer resources with analytical and numerical capabilities and they may turn into adequate tools for applied statisticians. For most of the functions associated with the proposed model, we provide numerical and graphical studies to illustrate their practical use. The model parameters are estimated by maximum likelihood and the observed information matrix is derived. An extensive Monte Carlo simulation study reveals that these estimators have good properties such as low biases and variances, even in small or moderate sample sizes. We also show that the proposed model can be superior to some other lifetime models by means of a real data set. Finally, Chapter 5 is dedicated to general considerations and also presents some paths for future research on the EG class of distributions.

It should be mentioned that, although having an unifying theme, this doctoral thesis consists of a collection of independent essays published in different journals. Nevertheless, each chapter is arranged according to its period of elaboration. This is reflected in the characteristics in each of them. For example, in Chapter 4, because it is more recent, an effort has been made to provide several practical answers, which illustrates the increasing computational demand in statistics. We hope that these works will be useful for researchers applied in various fields of activity.

CHAPTER 2

The Exponentiated Generalized Gumbel Distribution

Paper published in the *Colombian Journal of Statistics*, 38, 123-143, 2015.

Abstract

A family of univariate distributions called the exponentiated generalized class was recently proposed in the literature. A four-parameter model within this class, named the exponentiated generalized Gumbel distribution, is defined. We discuss the shapes of its density function and obtain explicit expressions for the ordinary moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves and Rényi entropy. The density function of the order statistic is derived. The method of maximum likelihood is used to estimate the model parameters. We determine the observed information matrix. We provide a Monte Carlo simulation study to evaluate the maximum likelihood estimates of the model parameters and two applications to real data to illustrate the importance of the model.

Keywords: Exponentiated generalized family. Gumbel distribution. Maximum likelihood, Moments.

Resumo

Recentemente foi proposta na literatura uma nova família de distribuições contínuas univariadas denominada de classe de distribuições exponencializadas generalizadas. Dentro desta família, denfine-se uma distribuição com quatro parâmetros denominada de Gumbel exponencializada generalizada. Nós discutimos as formas da densidade desta distrubuição e obtemos fórmulas explícitas para seus momentos ordinários, funções geratriz de momentos e quantílica, desvios médios ao redor da média e mediana, curvas de Bonferroni e Lorenz, além de entropia de Rényi. Nós derivamos a função densidade da estatística de ordem. Utilizamos o método de máxima verossimilhança para estimar os parâmetros do modelo e derivamos a matriz de informação observada. Fornecemos um estudo de simulação de Monte Carlo para avaliar os estimadores de máxima verossimilhança e duas aplicações a dados reais, que ilustram a importância do modelo.

Palavras-chave: Distribuição Gumbel. Família exponencializada generalizada. Máxima verossimilhança. Momentos.

2.1 Introduction

The Gumbel distribution is a very popular statistical model due to its wide applicability. An extensive list of the Gumbel model applications can be obtained in [12]. In the area of climate modeling, for example, some applications of the Gumbel model include: global warming problems, offshore modeling, rainfall and wind speed modeling ([13]). We can find applications of this model in various areas of engineering such as flood frequency analysis, network, nuclear, risk-based, space, software reliability, structural and wind engineering ([14]). Due to its wide applicability, several works aimed to extend the Gumbel model became important. Here, we refer to the papers: [15], [13] and [14].

The cumulative distribution function (cdf) $G(x)$ and probability density function (pdf) $g(x)$ of the Gumbel (Gu) distribution are given by

$$G(x) = G(x;\mu,\sigma) = \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\} \tag{2.1}$$

and

$$g(x) = g(x;\mu,\sigma) = \frac{1}{\sigma}\exp\left\{-\left[\frac{x-\mu}{\sigma}+\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}, \tag{2.2}$$

respectively, for $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$.

In recent years, some different generalizations of continuous distributions have received great attention in the literature. An excellent review of these generalizations is provided by [16]. Here, we refer to the papers: [4] for the *Marshall-Olkin* class, [5] for the *beta* class, [6] and [7] for the *gamma* class and [9] for the *Kumaraswamy* class of distributions. In a similar manner, for any baseline cdf $G(x)$, and $x \in \mathbb{R}$, [10] defined the *exponentiated generalized* (EG) class of distributions with two extra parameters $a > 0$ and $b > 0$ and cdf $F(x)$ and pdf $f(x)$ given by

$$F(x) = F(x;a,b) = \{1-[1-G(x)]^a\}^b \tag{2.3}$$

and

$$f(x) = f(x;a,b) = ab[1-G(x)]^{a-1}\{1-[1-G(x)]^a\}^{b-1}g(x), \tag{2.4}$$

respectively, in which the dependence of the $G(x)$ parameters are implicit.

The generator proposed by [10] presents some important characteristics that, in our opinion, makes it more attractive to generalize distributions. The first important point to highlight is the simplicity of the (2.3) and (2.4) equations. They do not involve any complicated functions and will be as simple as $G(x)$ is simple. Despite the simplicity of the model, the two extra parameters a and b in the density (2.4) can control both tail weights, allowing generate flexible distributions, with heavier or lighter tails, as appropriate. Another important feature is that the EG model contains as especial cases the two classes of Lehmann's alternatives. In fact, for $a = 1$, (2.3) reduces to $F(x) = G(x)^b$ and for $b = 1$ we obtain $F(x) = 1-[1-G(x)]^a$, which correspond to the cdf's of the Lehmann type I and II families ([17]), respectively. For this reason, the EG model *encompasses both Lehmann type I and Lehmann type II classes*. So, the EG family can be derived from a double transformation using these classes. There is also an attractive physical interpretation of the model (2.3) when a and b are positive integers.

This interpretation is described in [3] and reproduced below. Suppose initially that a certain device is composed of b components in a parallel system. Consider also that, for each component b, there exists a series of subcomponents a independent and distributed according to $G(x)$. Suppose also that each component b fails if some a subcomponent fails. Let $X_{j1}, \ldots, X_{ja}$ denote the lifetimes of the subcomponents within the jth component, $j = 1, \ldots, b$, with common cdf $G(x)$. Let X_j denote the lifetime of the jth component and let X denote the lifetime of the device. Thus, the cdf of X is

$$\begin{aligned}
P(X \leq x) &= P(X_1 \leq x, \ldots, X_b \leq x) = P(X_1 \leq x)^b = [1 - P(X_1 > x)]^b \\
&= [1 - P(X_{11} > x, \ldots, X_{1a} > x)]^b = [1 - P(X_{11} > x)^a]^b \\
&= [1 - \{1 - P(X_{11} \leq x)\}^a]^b.
\end{aligned}$$

So, the lifetime of the device obeys the EG family of distributions. The above properties and many others have been discussed and explored in recent works for the EG class. In this chapter, we study the so-called *exponentiated generalized Gumbel* ("EGGu" for short) distribution by inserting the formula (2.1) in equation (2.3). As we will see later, our model in study is very flexible and with great potential of adjustment to real data.

The rest of the chapter is organized as follows. In Section 2.2, we define the EGGu distribution. Shapes of the density function are discussed in Section 2.3. Explicit expressions for the cumulative and density functions, quantile function, ordinary moments, mean deviations, Bonferroni and Lorenz curves, generating function, Rényi entropy and order statistics are derived in Section 2.4. We discuss maximum likelihood estimation and present a Monte Carlo simulation experiment to evaluate the maximum likelihood estimates (MLEs) of the model parameters in Section 2.5. Two applications in Section 2.6 illustrate the usefulness of the EGGu distribution for data modeling. Lastly, concluding remarks are given in Section 2.7.

2.2 The EGGu distribution

The EGGu distribution was proposed by [10], but they did not provide a complete mathematical treatment for this model. The cdf and pdf of the EGGu distribution are given by

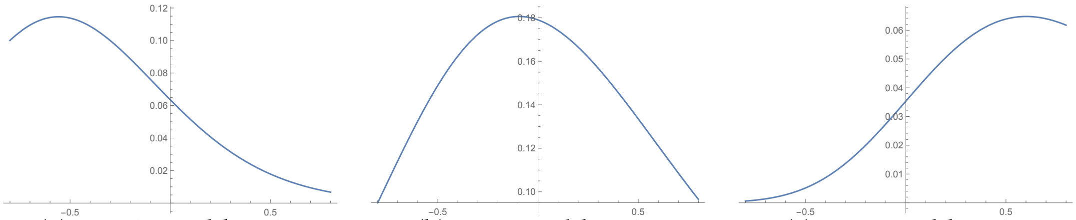
$$F(x) = F(x; a, b, \mu, \sigma) = \left\{1 - \left\{1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}^{a}\right\}^{b} \tag{2.5}$$

and

$$\begin{aligned} f(x) = f(x; a, b, \mu, \sigma) =& \frac{ab}{\sigma} \exp\left\{-\left[\frac{x-\mu}{\sigma} + \exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\} \\ & \left\{1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}^{a-1} \\ & \left\{1 - \left[1 - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right]^{a}\right\}^{b-1}, \end{aligned} \tag{2.6}$$

respectively, with $\mu \in \mathbb{R}$ and $\sigma > 0$.

Hence, a continuous random variable X with support on the real line and having density function (2.6) is denoted by X $\sim$ EGGu(a, b, μ, σ). We write $F(x) = F(x; a, b, \mu, \sigma)$ in order to eliminate the dependence on the model parameters. In this model, $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters, respectively, whereas $a > 0$ and $b > 0$ are the shape parameters. The Gumbel distribution is clearly an especial case of (2.5) when $a = b = 1$. Setting $b = 1$ we obtain the exponentiated Gumbel distribution defined by [13]. Plots of the EGGu density function for selected parameter values are displayed in Figure 2.1. As we will see later, our model in study is very flexible and with great potential of adjustment to real data.



(a) $a = 4.7$ and $b = 1.5$ (b) $a = 2$ and $b = 1.5$ (c) $a = 2$ and $b = 4.7$

Figure 2.1: Plots of the EGGu density function for $\mu = 0$, $\sigma = 1$ and some a and b parameter values.

2.3 Shapes

The main features of the density shape can be perceived through the study of its first and second derivative. Regarding the EGGu distribution, the first derivative of $\log\{f(x)\}$ is

$$\frac{d\log\{f(x)\}}{dx} = \frac{1}{\sigma}\left\{-1-\ln(z)\left[1-\frac{(a-1)\,z}{(1-z)}+\frac{a\,(b-1)\,z\,(1-z)^{a-1}}{[1-(1-z)^{a}]}\right]\right\},$$

where $z = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]$. Here, $0 < z < 1$.

The critical values of $f(x)$ are the roots of the equation:

$$\frac{(a-1)\,z}{(1-z)} - \frac{a\,(b-1)\,z\,(1-z)^{a-1}}{[1-(1-z)^{a}]} = \frac{\ln(z)+1}{\ln(z)}. \tag{2.7}$$

If the point $x = x_0$ is a root of (2.7), then we can classify it as local maximum, local minimum or inflection point when we have, respectively, $\lambda(x_0) < 0$, $\lambda(x_0) > 0$ and $\lambda(x_0) = 0$, where $\lambda(x) = d^2\log\{f(x)\}/dx^2$.

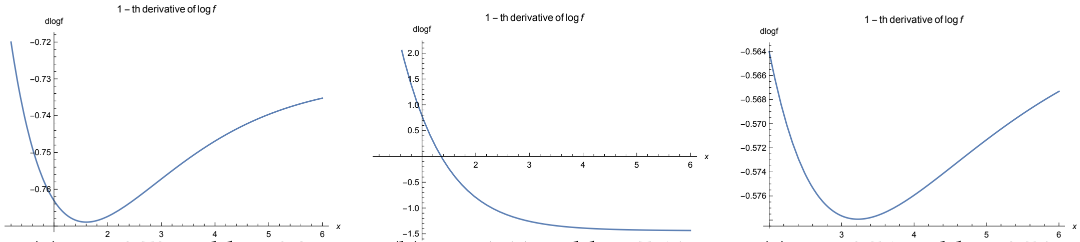
Particularly noteworthy are the following special cases for equation (2.7):

- For $a = 1$, the critical values of $f(x)$ are the roots of the equation $b = 1/\ln(z)$;
- For $b = 1$, the critical values of $f(x)$ are the roots of the equation $a = [1-z+\ln(z)]/[z\ln(z)]$;
- For $a = b = 1$, the critical values of $f(x)$ are the roots of the equation $[\ln(z)+1]/[\ln(z)] = 0$.

It is often difficult to obtain an analytical solution for the critical value of this function. Thus, it is very common to obtain numerical solutions from optimization routines in most mathematical and statistical platforms. Some plots of the first derivative of $\log\{f(x)\}$ for some a and b parameter values ($\mu = 0$, $\sigma = 1$) are displayed in Figure 2.2. These plots are constructed using the *Wolfram Mathematica* software.

2.4 Properties

In this section, we study some structural properties of the EGGu distribution.



(a) $a = 0.73$ and $b = 0.2$ (b) $a = 1.44$ and $b = 7.64$ (c) $a = 0.56$ and $b = 0.54$

Figure 2.2: Plots of the first derivative of $\log\{f(x)\}$.

2.4.1 A useful representation

Several properties of the EGGu distribution can be derived using the concept of exponentiated distributions. The class of exponentiated distributions have been studied by many authors in recent years. To give an idea of the importance of this class, a recent paper by [16] lists over seventy works related to exponentiated distributions. Here, we refer to a few of these papers: [18] for exponentiated exponential, [19] for exponentiated Fréchet, [13] for exponentiated Gumbel, [20] for exponentiated lognormal, [21] for exponentiated Gamma, [22] for exponentiated modified Weibull and [23] for exponentiated generalized Gamma.

For an arbitrary baseline cdf $G(x)$, a random variable is said to have the exponentiated-G ("exp-G" for short) distribution with power parameter $c > 0$, say $Y \sim$exp-$G(c)$, if its cdf and pdf are $H_c(x) = G(x)^c$ and $h_c(x) = c\, g(x) G(x)^{c-1}$, respectively. We consider the generalized binomial expansion

$$(1-z)^b = \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} z^k, \tag{2.8}$$

which holds for any real non-integer b and $|z| < 1$. Using expansion (2.8) twice in equation (2.3), [3] expressed the EG cdf as

$$F(x) = \sum_{j=0}^{\infty} w_{j+1}\, H_{j+1}(x), \tag{2.9}$$

where $w_{j+1} = \sum_{m=1}^{\infty} (-1)^{j+m+1} \binom{b}{m} \binom{m\,a}{j+1}$ and $H_{j+1}(x) = G(x)^{j+1}$ is the exponentiated-G (exp-G) cdf with power parameter $j+1$ (for $j \geq 0$). By differentiating (2.9), we obtain

$$f(x) = \sum_{j=0}^{\infty} w_{j+1}\, h_{j+1}(x), \tag{2.10}$$

where $h_{j+1}(x)$ is the exp-G pdf.

By using (2.9) and (2.10) for the Gumbel distribution (2.1), $h_{j+1}(x)$ becomes the exp-Gu pdf with power parameter $j+1$ (for $j \geq 0$) given by

$$h_{j+1}(x) = \frac{(j+1)}{\sigma} \exp\left\{ - \left[\frac{x-\mu}{\sigma} + (j+1) \exp\left(- \frac{x-\mu}{\sigma} \right) \right] \right\}. \tag{2.11}$$

Equations (2.10) and (2.11) reveal that the EGGu density function is a linear combination of exp-Gu densities. This result is important to derive some structural properties of the EGGu distribution like the ordinary and incomplete moments, generating function and mean deviations from those of the exp-Gu distribution.

2.4.2 Quantile function

In applied work, we are interested in the quantile function (qf) of a continuous distribution. Based on the qf, we can generate occurrences of the distribution and obtain measures of skewness and kurtosis. The EGGu qf, say $x_u = Q(u)$, follows by inverting the EGGu cdf (2.5) as

$$x_u = Q(u) = \mu - \sigma \log\{- \log[1 - (1 - u^{1/b})^{1/a}]\}. \tag{2.12}$$

The median of X is simply $x_{1/2} = Q(1/2)$. Furthermore, it is possible to generate EGGu variates by $X = Q(U)$, where U is an uniform variate on the unit interval $(0, 1)$.

The effect of the additional shape parameters a and b on the skewness and kurtosis of the EGGu distribution can be based on quantile measures. In this sense, two important measures are the Bowley's skewness (B) and the Moors's kurtosis (M). Recent papers used these measures to determine the skewness and kurtosis, for example, [24], [25] and [26] derived the B and M measures for the Beta normal, Beta exponentiated Pareto and exponentiated Lomax Poisson distributions, respectively.

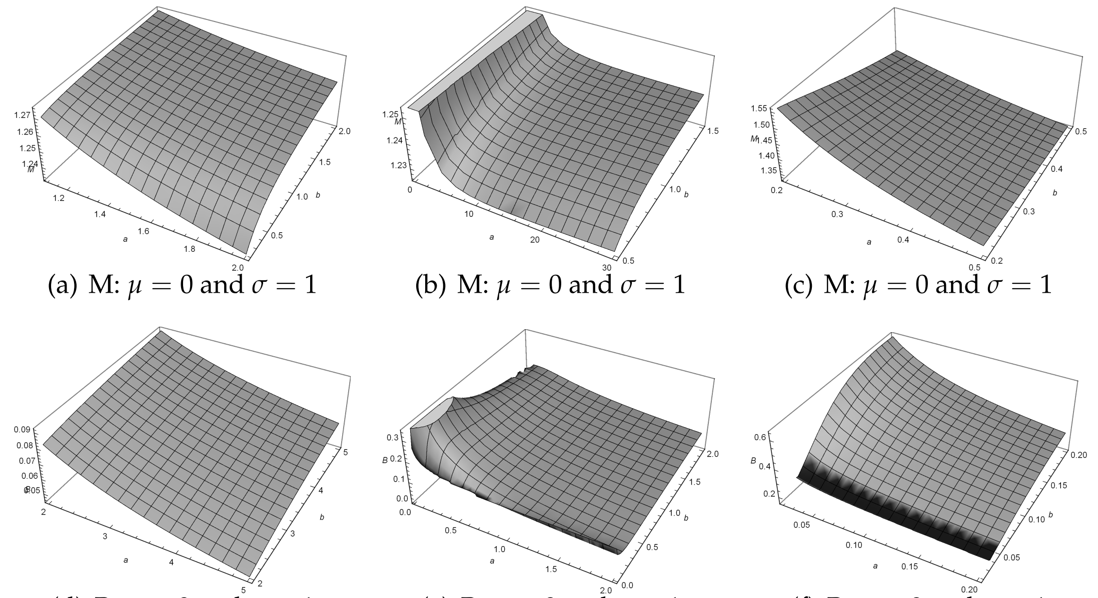
The Bowley's skewness ([27]) based on quartiles is given by

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

On the other hand, the Moors's kurtosis ([28]) based on octiles is given by

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

These measures are less sensitive to outliers and they exist even for distributions without moments. For the normal distribution, $B = M = 0$. Plots of these skewness and kurtosis measures for some choices of the parameter b as functions of a, and for some choices of a as functions of b, for $\mu = 0, \sigma = 1$, are displayed in Figure 2.3. These plots indicate that the skewness and kurtosis decrease when b increases for fixed a and when a increases for fixed b.

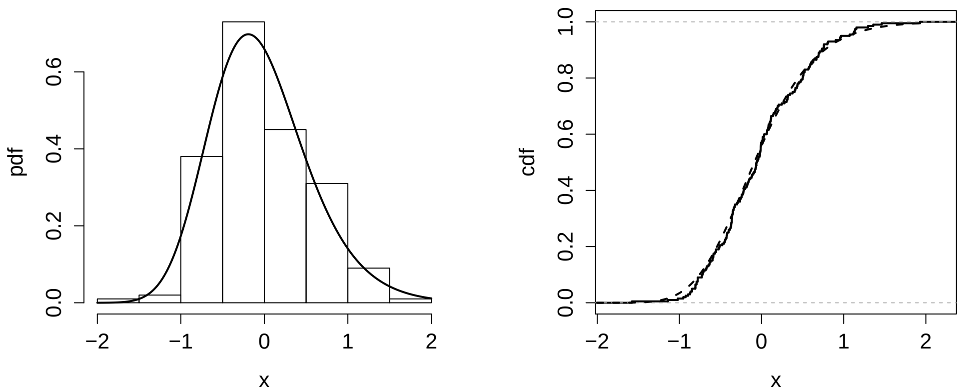


(a) M: $\mu = 0$ and $\sigma = 1$ (b) M: $\mu = 0$ and $\sigma = 1$ (c) M: $\mu = 0$ and $\sigma = 1$

(d) B: $\mu = 0$ and $\sigma = 1$ (e) B: $\mu = 0$ and $\sigma = 1$ (f) B: $\mu = 0$ and $\sigma = 1$

Figure 2.3: Plots of the Bowley's skewness and Moors's kurtosis for the EGGu distribution.

The EGSGu (we refer EGSGu to denote EGGu with $\mu = 0$ and $\sigma = 1$) distribution is easily simulated from a uniform random variable U by $X = Q(U)$. Next, we use (2.12) to generate 200 EGSGu(3, 2) occurrences. Figure 2.4 displays the histogram and empirical cdf for the simulated data and also the exact pdf and cdf of the EGSGu model. As we can see, the setting is quite adequate and reinforces that the model has good potential for simulation studies. For similar studies, we refer [29] and [30], among others.



(a) Histogram and exact pdf

(b) Empirical and exact cdfs

Figure 2.4: Plots of the EGSGu$(3,\ 2)$ pdf, histogram, exact and empirical cdfs for simulated data with $n = 200$.

2.4.3 Moments

It is hardly necessary to emphasize the importance of calculating the moments of a random variable in statistical analysis, especially in applied work. Some key features of a distribution such as skewness and kurtosis can be studied through its moments. The nth moment of X can be determined from (2.6) as $E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$. This integral has no closed form, but is easily manageable using symbolic computing softwares that have numerical integration routines. To illustrate this, we present below a small numerical example, in which we compute the first six moments for the EGSGu distribution, considering some a and b parameters values. The results are present in Table 2.1. All computations are obtained using *Wolfram Mathematica* software, which have numerical integration routines with great precision.

As we see, it is very simple to get the moments of X using numerical integration. But, we go further and we present below a closed expression for $E(X^n)$. Using equation (2.10), we obtain

$$
\begin{aligned}
E(X^n) &= \int_{-\infty}^{\infty} x^n \sum_{j=0}^{\infty} w_{j+1}\, h_{j+1}(x)dx \\
&= \frac{1}{\sigma}\sum_{j=0}^{\infty}(j+1)\, w_{j+1} \int_{-\infty}^{\infty} x^n \exp\left\{-\left[\frac{x-\mu}{\sigma} + (j+1)\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\} dx,
\end{aligned}
$$

Table 2.1: First sixth moments of X for several a and b values (with $\mu = 0$ and $\sigma = 1$).

a	b	$\mathrm{E}(X)$	$\mathrm{E}(X^2)$	$\mathrm{E}(X^3)$	$\mathrm{E}(X^4)$	$\mathrm{E}(X^5)$	$\mathrm{E}(X^6)$
	1	0.57722	1.97811	5.44487	23.5615	117.839	715.067
	2	1.27036	3.25876	10.7232	45.3600	233.071	1418.73
1	3	1.67583	4.45333	15.3804	66.3318	345.015	2113.31
	4	1.96351	5.50031	19.6637	86.4083	454.276	2798.98
	5	2.18665	6.42639	23.6502	105.693	561.110	3476.31
	1	−0.11593	0.69747	0.166507	1.76299	2.60769	11.4045
	2	0.34165	0.72200	1.03513	2.52093	6.50153	20.6720
2	3	0.59545	0.92435	1.59047	3.63454	9.74663	30.7523
	4	0.77030	1.14191	2.07825	4.75215	12.8901	40.7867
	5	0.90328	1.34995	2.53233	5.84379	15.9738	50.7334
	1	−0.40361	0.61140	−0.45472	0.93635	−0.68052	2.31902
	2	−0.02993	0.37204	0.10596	0.51195	0.57386	1.68848
3	3	0.17170	0.36710	0.31147	0.60373	1.02392	2.32992
	4	0.30846	0.41307	0.45178	0.75980	1.38992	3.06458
	5	0.41138	0.47383	0.57072	0.62827	1.73627	3.81059
	1	−0.57351	0.67294	−0.70212	1.00506	−1.28616	2.13690
	2	−0.24490	0.33249	−0.14312	0.29771	−0.09027	0.46581
4	3	−0.07069	0.24739	0.01646	0.20673	0.12820	0.38704
	4	0.04630	0.22709	0.09714	0.20835	0.23144	0.45403
	5	0.13374	0.23112	0.15195	0.23451	0.30806	0.54894

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

Setting $u = \exp\{-(x-\mu)/\sigma\}$, its reduces to

$$E(X^n) = \sum_{j=0}^{\infty}(j+1)w_{j+1}\int_0^{\infty}[\mu - \sigma\log(u)]^n \exp\{-u(j+1)\}du.$$

Using the binomial expansion for $[\mu - \sigma \log(u)]^n$, $E(X^n)$ can be expressed as

$$E(X^n) = \sum_{j=0}^{\infty}\sum_{i=0}^{n}(j+1)\binom{n}{i}(-\sigma)^i \mu^{n-i} w_{j+1} \int_0^{\infty} [\log(u)]^i \exp\{-u(j+1)\} du. \quad (2.13)$$

Using a result by [13], $I(i,j) = \int_0^{\infty}[\log(u)]^i \exp\{-u(j+1)\}du$ reduces to

$$I(i,j) = \left(\frac{\partial}{\partial c}\right)^i [(j+1)^{-c}\Gamma(c)] \mid_{c=1} . \quad (2.14)$$

By combining (2.13) and (2.14), the nth moment of X becomes

$$E(X^n) = \sum_{j=0}^{\infty}\sum_{i=0}^{n}(j+1)\binom{n}{i}(-\sigma)^i \mu^{n-i} w_{j+1} \left(\frac{\partial}{\partial c}\right)^i [(j+1)^{-c}\Gamma(c)] \mid_{c=1} .$$

2.4.4 Generating function

The moment generating function (mgf) of X can be obtained using the fact that the EGGu density function is a linear combination of exp-Gu densities. Thus,

$$\begin{aligned} M(t) &= \sum_{j=0}^{\infty} w_{j+1} \int_{-\infty}^{\infty} \mathrm{e}^{tx}\, h_{j+1}(x) dx \\ &= \frac{1}{\sigma}\sum_{j=0}^{\infty}(j+1) w_{j+1} \int_{-\infty}^{\infty} \mathrm{e}^{tx} \exp\left\{-\left[\frac{x-\mu}{\sigma} + (j+1)\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\} dx. \end{aligned}$$

Setting $u = \exp\{-(x-\mu)/\sigma\}$, $M(t)$ reduces to

$$M(t) = \mathrm{e}^{t\mu} \sum_{j=0}^{\infty}(j+1)\, w_{j+1} \int_0^{\infty} u^{-t\sigma} \exp[-(j+1)u]\, du.$$

Using a result by [14], we have

$$I(j) = \int_0^{\infty} u^{-t\sigma} \exp[-(j+1)u]\, du = \Gamma(1-t\sigma)(j+1)^{t\sigma-1},$$

and then

$$M(t) = \mathrm{e}^{t\mu}\, \Gamma(1-t\sigma) \sum_{j=0}^{\infty}(j+1)^{t\sigma}\, w_{j+1}.$$

2.4.5 Mean deviations

Generally, there has been a great interest in obtaining the first incomplete moment of a distribution. Based on this quantity, we can obtain, for example, the mean deviations that provide important information about characteristics of a population. Indeed, the amount of dispersion in a population may be measured to some extent by all the deviations from the mean and median.

For calculating the mean deviations about the mean and the median, we require the first incomplete moment of X given by $T(z) = \int_{-\infty}^{z} x\, f(x)\, dx$. Using equation (2.10) and setting $u = \exp\{-(x-\mu)/\sigma\}$, $T(z)$ reduces to

$$\begin{aligned} T(z) &= \sum_{j=0}^{\infty}(j+1)w_{j+1}\int_{t}^{\infty}[\mu - \sigma\log(u)]\exp[-u(j+1)]du \\ &= \sum_{j=0}^{\infty} w_{j+1}\{\exp[-t(j+1)][\mu-\sigma\log(t)] - \sigma\Gamma[0,(j+1)t]\}, \qquad (2.15) \end{aligned}$$

where $t = \exp\{-(z-\mu)/\sigma\}$ and $\Gamma(k,x) = \int_{x}^{\infty} v^{k-1}\, \mathrm{e}^{-v} dv$ is the complementary incomplete gamma function.

The mean deviations about the mean and the median are defined by

$$\delta_1 = 2\mu_1'\, F(\mu_1') - 2T(\mu_1') \qquad \text{and} \qquad \delta_2 = \mu_1' - 2T(M),$$

respectively, where $\mu_1' = E(X)$, the median M of X is determined from the qf by $M = Q(1/2)$, $F(M)$ and $F(\mu_1')$ are easily obtained from (2.5) and $T(z)$ is given by (2.15).

Another important application of the first incomplete moment is to determine the Bonferroni and Lorenz curves, which are commonly used in applied works in areas such as economy, reliability, demography, insurance, medicine and others. For a given probability π, these curves are defined by $B(\pi) = T(q)/(\pi\mu_1')$ and $L(\pi) = T(q)/\mu_1'$, where $\mu_1' = E(X)$ and $q = Q(\pi)$ is given by (2.12). Here, we make reference to the following papers: [31], [32], [33] and [34].

2.4.6 Rényi entropy

Given a certain random phenomenon under study, it is important to quantify the uncertainty associated with the random variable of interest. In this context, several statistical methods are available in the literature. One of the most popular measures used to quantify the variability of X is the Rényi entropy. Here, we make reference to the following papers: [25], [35], [36] and [37], among others.

The entropy of a random variable X with density function $f(x)$ for any real parameter $\lambda > 0$ and $\lambda \neq 1$, the Rényi entropy is given by

$$I_R(\lambda) = \frac{1}{(1-\lambda)} \log\left(\int_{-\infty}^{\infty} f(x)^{\lambda} dx\right).$$

The equation above can be easily implemented computationally with $I_R(\lambda)$ values being obtained in a few seconds. Table 2.2 shows some values of $I_R(\lambda)$ for the EGGu model, considering different parameter values. Naturally, the higher the value of $I_R(\lambda)$, indicates the greater uncertainty about the phenomenon under study. All computations are obtained using *Wolfram Mathematica* software, which have numerical integration routines with great precision.

Table 2.2: Rényi entropy of X for some λ, a and b values (with $\mu = 0$ and $\sigma = 1$).

a	b	$\lambda = 2$	$\lambda = 4$	$\lambda = 6$	$\lambda = 8$	$\lambda = 10$
2	3	0.90814	0.77951	0.72323	0.69037	0.66843
2	2	0.94522	0.81770	0.76180	0.72913	0.70731
3	2	0.72436	0.59999	0.54513	0.51299	0.49148

Using the binomial expansion (2.8) twice in equation (2.4), we can write

$$f(x)^{\lambda} = (ab)^{\lambda} \sum_{j=0}^{\infty} \delta_j \, G(x)^j \, g(x)^{\lambda}, \tag{2.16}$$

where δ_j is given by

$$\delta_j = \sum_{i=0}^{\infty} (-1)^{i+j} \binom{\lambda(b-1)}{i} \binom{ai + \lambda(a-1)}{j}.$$

Inserting (2.1) and (2.2) in equation (2.16) and, after some algebra, we obtain

$$f(x)^{\lambda} = \left(\frac{ab}{\sigma}\right)^{\lambda} \sum_{j=0}^{\infty} \delta_j \exp\left\{-\left[\frac{\lambda(x-\mu)}{\sigma} + (j+\lambda)\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}.$$

Finally,

$$I_R(\lambda) = \frac{1}{(1-\lambda)} \log\Bigg\{ \left(\frac{ab}{\sigma}\right)^{\lambda} \sum_{j=0}^{\infty} \delta_j \times$$
$$\int_{-\infty}^{\infty} \exp\left\{-\left[\frac{\lambda(x-\mu)}{\sigma} + (j+\lambda)\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\} dx \Bigg\}.$$

2.4.7 Order statistics

We derive an explicit expression for the density of the ith order statistic $X_{i:n}$, say $f_{i:n}(x)$, in a random sample of size n from the EGGu distribution. It is well-known that

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x)\, F(x)^{i+j-1}. \qquad (2.17)$$

Replacing (2.3) and (2.4) in equation (2.17) and applying the binomial expansion (2.8) twice, we can write

$$f_{i:n}(x) = \frac{ab}{B(i, n-i+1)} \sum_{\ell=0}^{\infty} \vartheta_\ell \; g(x)\, G(x)^{\ell},$$

where ϑ_ℓ is given by

$$\vartheta_\ell = \sum_{j=0}^{n-i} \sum_{k=0}^{\infty} (-1)^{j+k+\ell} \binom{n-i}{j} \binom{b(i+j)-1}{k} \binom{a(k+1)-1}{\ell}.$$

Thus, replacing $G(x)$ and $g(x)$ by the cdf and pdf of the Gumbel distribution given by (2.1) and (2.2), respectively, we can write $f_{i:n}(x)$ as

$$f_{i:n}(x) = \frac{ab}{\sigma B(i, n-i+1)} \sum_{\ell=0}^{\infty} \vartheta_\ell \; \exp\left\{-\left[\frac{x-\mu}{\sigma} + (\ell+1)\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}.$$

After simple algebraic manipulation, we can rewrite the last equation as

$$f_{i:n}(x) = \frac{ab}{B(i, n-i+1)} \sum_{\ell=0}^{\infty} \vartheta_\ell^* \, h_{\ell+1}(x), \qquad (2.18)$$

where $\vartheta_\ell^* = \vartheta_\ell/(\ell+1)$ and $h_{\ell+1}(x)$ is given by (2.11).

Equation (2.18) reveals that the density function of the EGGu order statistic is a linear combination of exp-Gu densities. A direct application of (2.18) is to calculate the moments and the mgf of the EGGu order statistics.

The rth moment of $X_{i:n}$ is given by

$$E(X_{i:n}^r) = \frac{ab}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} \vartheta_\ell^* \int_{-\infty}^{\infty} x^r \, h_{\ell+1}(x) \, dx.$$

From the results presented in Section (2.4.3), the last equation reduces to

$$E(X_{i:n}^r) = \frac{ab}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} \sum_{k=0}^{r} (\ell+1) \, (-\sigma)^k \, \mu^{r-k} \binom{r}{k} \vartheta_\ell^* \left(\frac{\partial}{\partial c}\right)^k \left[(\ell+1)^{-c}\Gamma(c)\right] \big|_{c=1} \, .$$

The mgf of $X_{i:n}$ is given by

$$M(t) = \frac{ab}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} \vartheta_\ell^* \int_{-\infty}^{\infty} \mathrm{e}^{tx} \, h_{\ell+1}(x) \, dx.$$

Finally, based on the results in Section (2.4.4), the last equation can be rewritten as

$$M(t) = \frac{a\, b\, \mathrm{e}^{t\mu} \, \Gamma(1-t\sigma)}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} (\ell+1)^{t\sigma} \, \vartheta_\ell^*.$$

2.5 Estimation and inference

Several approaches for parameter point estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The MLEs enjoy desirable properties and can be used when constructing confidence intervals and regions and also in test statistics. Large sample theory for these estimates delivers simple approximations that work well in finite samples. The resulting approximation for the estimates in distribution theory is easily handled either analytically or numerically. So, we consider the estimation of the unknown parameters a, b, μ and σ of the EGGu distribution from complete samples only by the method of maximum likelihood. Let $x_1, \ldots, x_n$ be a sample of size n from X. The log-likelihood function for the vector of parameters $\boldsymbol{\theta}^\top = (a, b, \mu, \sigma)^\top$ can be expressed as

$$\begin{aligned}
\ell(\boldsymbol{\theta}) = n \log\left(\frac{ab}{\sigma}\right) &- \sum_{i=1}^{n} \left[\frac{x_i-\mu}{\sigma} + \exp\left(-\frac{x_i-\mu}{\sigma}\right)\right] \\
&+ (a-1) \sum_{i=1}^{n} \log\left\{1 - \exp\left[-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]\right\} \\
&+ (b-1) \sum_{i=1}^{n} \log\left\{1 - \left[1 - \exp\left[-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]\right]^a\right\}.
\end{aligned}$$

The elements of the score vector are given by

$$
\begin{aligned}
U_a(\theta) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \log[H(x_i)] + (1-b)\sum_{i=1}^{n} \frac{H(x_i)^a \log H(x_i)}{1-H(x_i)^a},\\
U_b(\theta) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} \log[1-H(x_i)^a],\\
U_\mu(\theta) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} = \frac{n}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n} \exp\left(-\frac{x_i-\mu}{\sigma}\right) + (a-1)\sum_{i=1}^{n} \frac{g(x_i)}{H(x_i)}\\
&\quad + a(1-b)\sum_{i=1}^{n} \frac{g(x_i)H(x_i)^{a-1}}{1-H(x_i)^a},\\
U_\sigma(\theta) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{n\mu}{\sigma^2} - \frac{n}{\sigma} - \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\exp\left(-\frac{x_i-\mu}{\sigma}\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i\\
&\quad + (a-1)\sum_{i=1}^{n} \frac{(x_i-\mu)g(x_i)}{\sigma H(x_i)}\\
&\quad + (b-1)\sum_{i=1}^{n} -\frac{a(x_i-\mu)g(x_i)H(x)^{a-1}}{\sigma[1-H(x)^a]},
\end{aligned}
$$

$H(x_i) = 1-\exp\left[-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]$ and $g(x_i) = \frac{1}{\sigma}\exp\left\{-\left[\frac{x_i-\mu}{\sigma} + \exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]\right\}$.

The MLEs $\widehat{\theta}$ of θ is obtained by solving simultaneously the nonlinear equations $U_a(\theta) = 0$, $U_b(\theta) = 0$, $U_\mu(\theta) = 0$ and $U_\sigma(\theta) = 0$. They cannot be solved analytically and require statistical software with iterative numerical techniques. There exists many maximization methods in R scripts like NR (Newton-Raphson), BFGS (Broyden-Fletcher-Goldfarb-Shanno), BHHH (Berndt-Hall-Hall-Hausman), SANN (Simulated-Annealing), NM (Nelder-Mead) and L-BFGS-B. For interval estimation and hypothesis tests on the parameters a, b, μ and σ, we determine the 4×4 observed information matrix $J(\theta) = \{-U_{rs}\}$, where $U_{rs} = \partial^2\ell(\boldsymbol{\theta})/(\partial\boldsymbol{\theta}_r\partial\boldsymbol{\theta}_s)$ for $r, s \in \{a, b, \mu, \sigma\}$. The elements of $J(\theta)$ are given in the Appendix.

Next, a small Monte Carlo simulation experiment based on $10,000$ replications will be conducted to evaluate the MLEs of the parameters of the EGGu distribution. We set the sample size at $n = 100, 200, 400$ and 800, the parameter a at $a = 1.5$ and 3.0, and the parameter b at $b = 1.5$ and 3.0. The location and scale parameters are fixed at $\mu = 0$ and $\sigma = 1$, respectively, without loss of generality. The Monte Carlo simulation experiments are performed using the R programming language; see http://www.r-project.org. Table 2.3 reports the empirical means and mean squared errors (in parentheses) of the corresponding estimators. From the figures in this table, we note that, as the sample size increases, the empirical biases and mean squared errors decrease in all cases, as expected.

Table 2.3: MLEs and mean squared errors (in parentheses); $\mu = 0$ and $\sigma = 1$.

a	b	$\widehat{a}$	$\widehat{b}$	$\widehat{\mu}$	$\widehat{\sigma}$
		$n = 100$			
1.5	1.5	1.8709 (2.4748)	1.8402 (4.2093)	0.4719 (2.5600)	0.9405 (0.2061)
1.5	3.0	1.7451 (2.1280)	2.7433 (4.5480)	0.5222 (2.3406)	0.8962 (0.2240)
3.0	1.5	3.8777 (11.916)	1.4662 (3.1887)	0.4057 (11.311)	0.8999 (0.1775)
3.0	3.0	3.3444 (8.8191)	2.3874 (4.3211)	0.3190 (8.8023)	0.8200 (0.1654)
		$n = 200$			
1.5	1.5	1.6795 (0.9176)	1.7512 (4.1874)	0.4188 (1.0608)	0.9684 (0.1151)
1.5	3.0	1.5910 (0.8141)	2.8413 (4.1928)	0.3743 (0.9459)	0.9364 (0.1237)
3.0	1.5	3.6424 (7.1481)	1.5338 (3.1388)	0.3643 (6.8681)	0.9632 (0.1720)
3.0	3.0	3.1394 (3.8597)	2.5946 (4.2360)	0.2676 (3.9119)	0.8860 (0.1213)
		$n = 400$			
1.5	1.5	1.5886 (0.2920)	1.6240 (2.9118)	0.3542 (0.4096)	0.9839 (0.0496)
1.5	3.0	1.5559 (0.3442)	2.9562 (3.6910)	0.2807 (0.4199)	0.9713 (0.0674)
3.0	1.5	3.4946 (5.0507)	1.6924 (3.6551)	0.2981 (4.8949)	1.0017 (0.1697)
3.0	3.0	3.1413 (3.3052)	2.7812 (4.4744)	0.2236 (3.3352)	0.9406 (0.1107)
		$n = 800$			
1.5	1.5	1.5495 (0.1265)	1.5625 (1.9864)	0.2688 (0.1963)	0.9914 (0.0274)
1.5	3.0	1.5328 (0.1513)	2.9838 (2.5010)	0.1724 (0.1799)	0.9889 (0.0332)
3.0	1.5	3.2975 (3.6003)	1.5810 (2.8298)	0.2321 (3.5656)	0.9932 (0.1335)
3.0	3.0	2.9814 (1.4341)	2.7585 (2.5881)	0.1302 (1.4507)	0.9467 (0.0599)

2.6 Applications to real data

In this section, we provide two applications to real data sets to illustrate the importance of the EGGu distribution. The MLEs of the parameters are computed (as discussed in Section 2.5) and the goodness-of-fit statistics for this model are compared with other competing models. All computations are performed using the SAS subroutine NLMixed. The four-parameter Beta Gumbel (BGu) [15] and Kumaraswamy Gumbel (KumGu) [14] distributions are used to make a comparison with the EGGu model. Their pdfs are given by

$$\pi_{\mathrm{BGu}}(x;a,b,\mu,\sigma)=\frac{\exp[-(x-\mu)/\sigma]\exp\{-a\exp[-(x-\mu)/\sigma]\}\{1-\exp[-(x-\mu)/\sigma]\}^{b-1}}{\sigma\,B(a,b)},$$

and

$$\begin{aligned}\pi_{\mathrm{KumGu}}(x;a,b,\mu,\sigma)&=a\,b\,\exp[-(x-\mu)/\sigma]\exp\{-a\exp[-(x-\mu)/\sigma]\}\\&\quad\{1-\exp\{-a\exp[-(x-\mu)/\sigma]\}\}^{b-1},\end{aligned}$$

respectively.

The first data set is obtained from [38]. They consist of thirty successive values of March precipitation (in inches) in Minneapolis/St Paul. The data are: 0.77, 1.74, 0.81, 1.20, 1.95, 1.20, 0.47, 1.43, 3.37, 2.20, 3.00, 3.09, 1.51, 2.10, 0.52, 1.62, 1.31, 0.32, 0.59, 0.81, 2.81, 1.87, 1.18, 1.35, 4.75, 2.48, 0.96, 1.89, 0.90, 2.05. Table 2.4 gives some descriptive statistics for these data, which includes central tendency statistics, variance, among others. Table 2.5 lists the MLEs of the model parameters (standard errors in parentheses) for all fitted models. It is also given the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC).

Table 2.4: Descriptives statistics for Hinkley's data set.

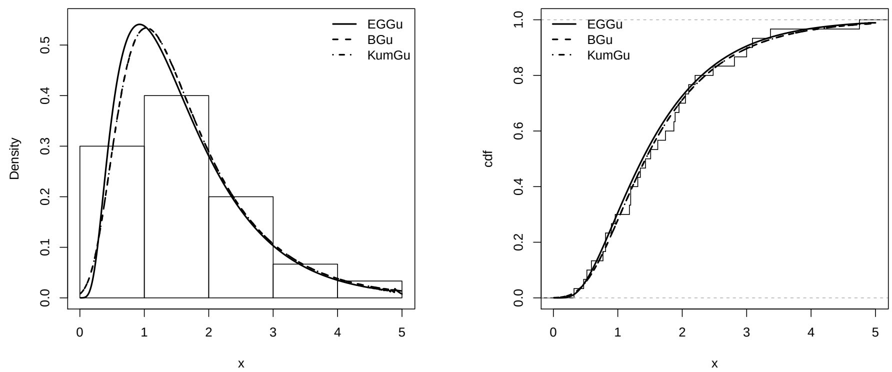
Statistic	
Mean	1.675
Median	1.470
Variance	1.001
Minimum	0.320
Maximum	4.750

Table 2.5: MLEs (standard errors in parentheses), AIC, BIC and CAIC statistics for Hinkley's data.

Distribution	$\widehat{a}$	$\widehat{b}$	$\widehat{\mu}$	$\widehat{\sigma}$	AIC	BIC	CAIC
EGGu	0.1440	2.1935	0.2190	0.1285	**84.1**	**85.7**	**89.7**
	(0.0358)[a]	(1.3392)	(0.2508)	(0.0078)			
BGu	0.8294	0.4449	0.9133	0.4301	84.7	86.3	90.3
	(2.0008)[a]	(0.4855)	(1.5573)	(0.3628)			
KumGu	0.7632	0.4349	0.9225	0.4206	84.7	86.3	90.3
	(0.2094)[a]	(0.4428)	(0.3800)	(0.3151)			

[a] Denotes the standard deviations of the MLEs of a, b, μ and σ.

Plots of the estimated pdf and cdf of the fitted EGGu, BGu and KumGu models to these data are displayed in Figure 2.5. They indicate that the EGGu distribution is superior to the other distributions in terms of model fitting.



(a) (b)

Figure 2.5: (a) Plots of the fitted EGGu, BGu and KumGu densities; (b) Plots of the estimated cdfs of the EGGu, BGu and KumGu models.

Next, we shall apply formal goodness-of-fit tests in order to verify which distribution fits better to the current data. We consider the Cramér-von Mises (CM) and Anderson-Darling (AD) statistics, which are described in [39]. Table 2.6 gives the values of the CM and AD statistics (and the p-values of the tests in parentheses) for the fitted models. Thus, according to these formal tests, the EGGu model fits the current data better than the other models, i.e., these values indicate that the null hypothesis is strongly not rejected for the EGGu distribution. Based on the plots of Figure 2.5, we conclude that the EGGu distribution provides a better fit to these data than the BGu and KumGu models.

Table 2.6: Goodness-of-fit tests for Hinkley's data set.

Model	Statistics	
	CM	AD
EGGu	**0.0151 (0.9932)**[a]	**0.1169 (0.9891)**[a]
BGu	0.0205 (0.9611)	0.1606 (0.9415)
KumGu	0.0193 (0.9718)	0.1520 (0.9548)

[a] Denotes the p-value of the test.

The second data set is given by [40]. The data referrers to the time between failures for repairable item: 1.43, 0.11, 0.71, 0.77, 2.63, 1.49, 3.46, 2.46, 0.59, 0.74, 1.23, 0.94, 4.36, 0.40, 1.74, 4.73, 2.23, 0.45, 0.70, 1.06, 1.46, 0.30, 1.82, 2.37, 0.63, 1.23, 1.24, 1.97, 1.86, 1.17. Table 2.7 gives some descriptive statistics for these data. Table 2.8 gives the MLEs of the model parameters (standard errors in parentheses) for all fitted models and the values of the AIC, BIC and CAIC statistics.

Table 2.7: Descriptives statistics for the times between failures.

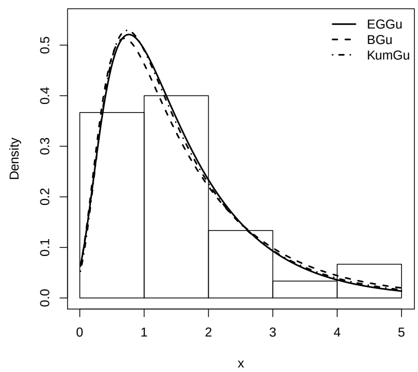
Statistic	
Mean	1.543
Median	1.235
Variance	1.272
Minimum	0.110
Maximum	4.730

Table 2.8: MLEs (standard errors in parentheses) and the AIC, BIC and CAIC statistics for the times between failures.

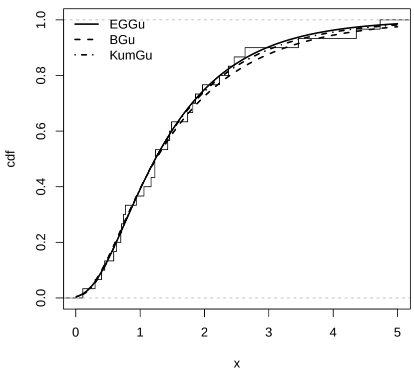
Distribution	$\widehat{a}$	$\widehat{b}$	$\widehat{\mu}$	$\widehat{\sigma}$	AIC	BIC	CAIC
EGGu	0.2914	1.3294	0.3146	0.3004	87.55	93.16	89.15
	(0.3659)[a]	(1.0088)	(0.5421)	(0.3222)			
BGu	7.7144	0.2089	-0.2351	0.2600	87.82	93.43	89.42
	(10.521)[a]	(0.1715)	(0.5534)	(0.1884)			
KumGu	2.4766	0.2749	0.1804	0.3115	87.55	93.16	89.15
	(4.4419)[a]	(0.2469)	(0.5544)	(0.2508)			

[a] Denotes the standard deviation of the MLEs of a, b, μ and σ.

Plots of the estimated pdfs and cdfs of the EGGu, BGu and KumGu models to the current data are displayed in Figure 2.6.



(a)



(b)

Figure 2.6: (a) Plots of the fitted EGGu, BGu and KumGu densities; (b) Plots of the estimated cdfs of the EGGu, BGu and KumGu models.

Table 2.9 gives the values of the CM and AD statistics (p-values between parentheses). Thus, according to these formal tests, the EGGu model fits the current data better than the other models, i.e., these values indicate that the null hypotheses are strongly not rejected for the EGGu distribution.

Table 2.9: Goodness-of-fit tests for the times between failures.

Model	Statistics	
	CM	AD
EGGu	**0.0168 (0.9885)**[a]	**0.1198 (0.9892)**[a]
BGu	0.0181 (0.9821)	0.1231 (0.9874)
KumGu	0.0176 (0.9848)	0.1204 (0.9889)

[a] Denotes the p-value of the test.

2.7 Conclusions

In this chapter, we study a four-parameter model named the exponentiated generalized Gumbel (EGGu) distribution. This model generalizes the Gumbel distribution, which is one of the most important models for fitting data with support in $\mathbb{R}$. We provide some mathematical properties of the EGGu distribution including the ordinary moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves and Rényi entropy. The density function of the order statistics is obtained as a combination of exponentiated Gumbel densities. We discuss the parameter estimation by maximum likelihood and provide the observed information matrix. We provide a Monte Carlo simulation study to evaluate the maximum likelihood estimation of the model parameters. Two applications to real data indicate that the EGGu distribution provides a good fit and can be used as a competitive model to fit real data.

Appendix A: Observed information matrix.

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a^2} &= -\frac{n}{a^2} + (b-1)\sum_{i=1}^{n}\left\{-\frac{H(x_i)^{2a}\log[H(x_i)]^2}{[1-H(x_i)^a]^2} - \frac{H(x_i)^a \log[H(x_i)]^2}{1-H(x_i)^a}\right\},\\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial b^2} &= -\frac{n}{b^2},\\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu^2} &= (a-1)\sum_{i=1}^{n}\left\{-\frac{g(x_i)^2}{H(x_i)^2} + \frac{g(x_i)\left[\frac{1-\exp(-\frac{x_i-\mu}{\sigma})}{\sigma}\right]}{H(x_i)}\right\}\\
&\quad + (b-1)\sum_{i=1}^{n}\left\{-\frac{a(a-1)g(x_i)^2H(x_i)^{a-2}}{1-H(x_i)^a} - \frac{a^2g(x_i)^2H(x_i)^{2(a-1)}}{[1-H(x_i)^a]^2}\right.\\
&\qquad\left. - \frac{a[H(x_i)]^{a-1}g(x_i)\left[\frac{1-\exp(-\frac{x_i-\mu}{\sigma})}{\sigma}\right]}{1-H(x_i)^a}\right\} - \sum_{i=1}^{n}\frac{\exp(\frac{x_i-\mu}{\sigma})}{\sigma^2},\\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma^2} &= \frac{2n\mu}{\sigma^3} + \frac{n}{\sigma^2} - \frac{2}{\sigma^3}\sum_{i=1}^{n}x_i\\
&\quad - \sum_{i=1}^{n}\left[\frac{-2(x_i-\mu)\exp\left(-\frac{x_i-\mu}{\sigma}\right)}{\sigma^3} + \frac{(x_i-\mu)^2\exp\left(-\frac{x_i-\mu}{\sigma}\right)}{\sigma^4}\right]\\
&\quad + (a-1)\sum_{i=1}^{n}\left\{-\frac{2(x_i-\mu)g(x_i)}{\sigma^2H(x_i)} - \frac{(x_i-\mu)^2g(x_i)^2}{\sigma^2H(x_i)}\right.\\
&\qquad\left. + \frac{(x_i-\mu)g(x_i)\left\{\frac{(x_i-\mu)\left[1-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]}{\sigma^2}\right\}}{\sigma H(x_i)}\right\}\\
&\quad + (b-1)\sum_{i=1}^{n}\left\{\frac{2a(x_i-\mu)g(x_i)H(x_i)^{a-1}}{\sigma^2[1-H(x_i)^a]}\right.\\
&\qquad - \frac{a(a-1)(x_i-\mu)^2g(x_i)^2H(x_i)^{a-2}}{\sigma^2[1-H(x_i)^a]}\\
&\qquad - \frac{a^2(x_i-\mu)^2g(x_i)^2H(x_i)^{2(a-1)}}{\sigma^2[1-H(x_i)^a]^2}\\
&\qquad\left. - \frac{a(x_i-\mu)g(x_i)H(x_i)^{a-1}\left\{\frac{(x_i-\mu)\left[1-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]}{\sigma^2}\right\}}{\sigma[1-H(x_i)^a]}\right\},\\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a\partial b} &= \sum_{i=1}^{n}-\frac{H(x_i)^a\log[H(x_i)]}{1-H(x_i)^a},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a \partial \mu} &= \sum_{i=1}^{n} \frac{g(x_i)}{H(x_i)} + (b-1)\sum_{i=1}^{n}\Bigg\{ -\frac{g(x_i)H(x_i)^{a-1}}{1-H(x_i)^a} - \frac{a g(x_i) H(x_i)^{2a-1}\log[H(x_i)]}{[1-H(x_i)^a]^2} \\
&\qquad - \frac{a g(x_i) H(x_i)^{a-1}\log[H(x_i)]}{1-H(x_i)^a}\Bigg\}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a \partial \sigma} &= \sum_{i=1}^{n} \frac{(x_i-\mu)g(x_i)}{\sigma H(x_i)} + (b-1)\sum_{i=1}^{n}\Bigg\{ -\frac{(x_i-\mu)g(x_i)H(x_i)^{a-1}}{\sigma[1-H(x_i)^a]} \\
&\qquad - \frac{a(x_i-\mu)g(x_i)H(x_i)^{2a-1}\log[H(x_i)]}{\sigma[1-H(x_i)^a]^2} \\
&\qquad - \frac{a(x_i-\mu)g(x_i)H(x_i)^{a-1}\log[H(x_i)]}{\sigma[1-H(x_i)^a]}\Bigg\}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial b \partial \mu} &= \sum_{i=1}^{n} -\frac{a g(x_i)H(x_i)^{a-1}}{1-H(x_i)^a}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial b \partial \sigma} &= \sum_{i=1}^{n} -\frac{a(x_i-\mu)g(x_i)H(x_i)^{a-1}}{\sigma[1-H(x_i)^a]}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu \partial \sigma} &= -\frac{n}{\sigma^2} - \sum_{i=1}^{n}\Bigg\{ -\frac{\exp\left(-\frac{x_i-\mu}{\sigma}\right)}{\sigma^2} - \frac{(x_i-\mu)\exp\left(-\frac{x_i-\mu}{\sigma}\right)}{\sigma^3}\Bigg\} \\
&\quad + (a-1)\sum_{i=1}^{n}\Bigg\{ -\frac{g(x_i)}{\sigma H(x_i)} - \frac{(x_i-\mu)g(x_i)^2}{\sigma H(x_i)^2} + \frac{g(x_i)[(x_i-\mu)(1-\exp\left(-\frac{x_i-\mu}{\sigma}\right)]}{\sigma^2 H(x_i)}\Bigg\} \\
&\quad + (b-1)\sum_{i=1}^{n}\Bigg\{ \frac{a g(x_i)H(x_i)^{a-1}}{\sigma[1-H(x_i)^a]} - \frac{a(a-1)(x_i-\mu)g(x_i)^2 H(x_i)^{a-2}}{\sigma[1-H(x_i)^a]} \\
&\qquad - \frac{a^2(x_i-\mu)g(x_i)^2 H(x_i)^{2(a-1)}}{\sigma[1-H(x_i)^a]^2} \\
&\qquad - \frac{a g(x_i)H(x_i)^{a-1}\left[\frac{(x_i-\mu)(1-\exp\left(-\frac{x_i-\mu}{\sigma}\right)}{\sigma^2}\right]}{1-H(x_i)^a}\Bigg\},
\end{aligned}
$$

where $H(x_i) = 1-\exp\left[-\exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]$ and $g(x_i) = \frac{1}{\sigma}\exp\left\{-\left[\frac{x_i-\mu}{\sigma} + \exp\left(-\frac{x_i-\mu}{\sigma}\right)\right]\right\}$.

Appendix B: Additional application study to real data

In this appendix we provide an additional application to real data, in order to reinforce the importance of the EGGu model. Here, we consider the EG generator applied to the standardized Gumbel distribution. The cumulative distribution function (cdf) $G(x)$ and probability density function (pdf) $g(x)$ of the standardized Gumbel (SGu) distribution are given by

$$G(x;\mu,\sigma) = \exp\left[-\exp\left(-x\right)\right]$$

and

$$g(x;\mu,\sigma) = \exp[-x - \exp(-x)]$$

respectively, for $x \in I\!R$, $\mu = 0$ and $\sigma = 1$.

So, the cdf $F(x)$ and pdf $f(x)$ of the exponentiated generalized standardized Gumbel (EGSGu) are given by

$$F(x) = \{1 - [1 - \exp\left[-\exp\left(-x\right)\right]]^{a}\}^{b}$$

and

$$f(x) = a\,b\,\exp[-x - \exp(-x)]\ \{1 - \exp\left[-\exp\left(-x\right)\right]\}^{a-1}\ \{1 - [1 - \exp\left[-\exp\left(-x\right)\right]]^{a}\}^{b-1},$$

respectively, where $a > 0$, $b > 0$.

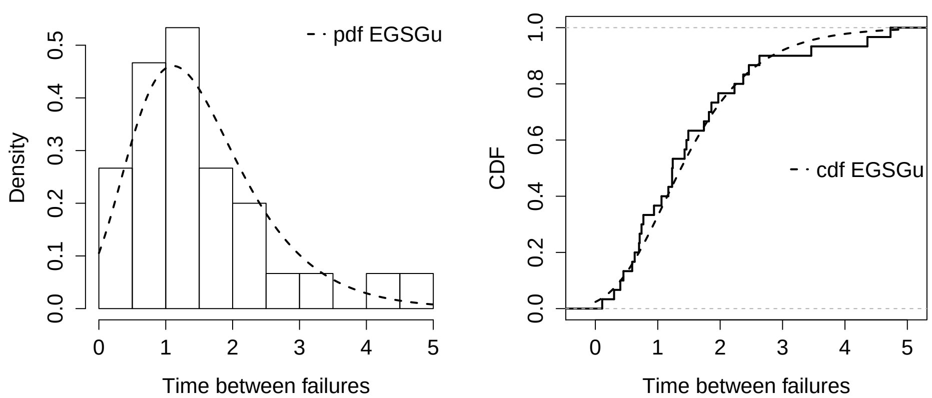
Clearly, the model EGSGu is a special case of the equation (2.5) when $\mu = 0$ and $\sigma = 1$. We adjusted the EGSGu model, that contain just two parameters, and compare the results with other important models in the literature. We consider the EGSGu distribution and two Gumbel Lehmann's alternatives sub-models, denoted by EISGu and EIISGu, respectively. In addition, we have adjusted the models proposed by [41], [42] and [43]. The data set was obtained from [40], and consists of the times between failures for repairable items: 1.43, 0.11, 0.71, 0.77, 2.63, 1.49, 3.46, 2.46, 0.59, 0.74, 1.23, 0.94, 4.36, 0.40, 1.74, 4.73, 2.23, 0.45, 0.70, 1.06, 1.46, 0.30, 1.82, 2.37, 0.63, 1.23, 1.24, 1.97, 1.86, 1.17. Table 2.10 gives some descriptive statistics for these data. In Table 2.11 we provide the MLEs (and their standard errors in parentheses) for all fitted models.

Table 2.10: Descriptives statistics for the times between failures.

Statistic	
Mean	1.543
Median	1.235
Variance	1.272
Minimum	0.110
Maximum	4.730

Table 2.12 lists the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) statistics. In general, it is considered that the lower values of these criteria indicates the better fit to the data. The figures in Table 2.12 revels that the EGSGu model has the lowest AIC, BIC and CAIC values among all fitted models. Thus, the proposed EGSGu distribution is the best model to explain these data.

Plots of the estimated pdf and cdf of the EGSGu distribution and the histogram of the data are displayed in Figure 2.7. These plots clearly reveal that the EGSGu model fits the data adequately and then it can be chosen for modeling these data.



(a) Estimated pdf of the EGSGu model (b) Estimated cdf of the EGSGu model

Figure 2.7: Estimated pdf and cdf of the EGSGu model for the times between failures for repairable items.

Table 2.11: The MLEs (and their standard errors in parentheses) for the times between failures.

Distribution					
EGSGu	$\widehat{a}$	$\widehat{b}$			
	1.340	4.800			
	(0.247)	(1.670)			
EISGu	1	3.118			
	(−)	(0.569)			
EIISGu	0.589	1			
	(0.108)[a]	(−)			
KwGP	$\widehat{a}$	$\widehat{b}$	$\widehat{\xi}$	$\widehat{\sigma}$	
	1.917	14.432	1.119	4.972	
	(0.837)[a]	(12.687)	(2.371)	(4.817)	
BGP	$\widehat{a}$	$\widehat{b}$	$\widehat{\xi}$	$\widehat{\sigma}$	
	1.979	13.850	0.156	11.061	
	(0.460)[a]	(1.983)	(1.960)	(3.852)	
BMW	$\widehat{a}$	$\widehat{b}$	$\widehat{\alpha}$	$\widehat{\gamma}$	$\widehat{\lambda}$
	2.901	0.344	2.374	0.854	0.093
	(1.903)[a]	(0.168)	(0.164)	(0.210)	(0.109)

[a] Denotes the standard deviation of the MLE's.

Appendix C: Additional simulation study

We understand that the inferential process depends heavily on the quality of the estimates obtained for the model under study. For this reason, we have extended the simulation study presented in Section 2.5. We investigate the behavior of the MLEs

Table 2.12: AIC, BIC and CAIC statistics for the times between failures for repairable items.

Distribution	AIC	BIC	CAIC
EGSGu	86.181	88.983	86.826
EISGu	86.334	87.735	86.477
EIISGu	103.73	105.13	103.87
KwGP	87.283	92.888	88.883
BGP	87.257	92.862	88.857
BMW	89.239	96.245	91.739

for the parameters of the EGGu model by generating from (2.12) samples sizes $n =$ 50, 100, 150, 200 with $\mu = 0$, $\sigma = 1$ and selected values for a and b.

The simulation process is based in $10,000$ Monte Carlo replications, performed in the `R` software using the *simulated-annealing*(SANN) maximization method in the *maxLik* script. To ensure the reproducibility of the experiment, we use the seed for the random number generator: set.seed (103). Initial kicks are taken as equal to half of the true values of the parameters in each scenario.

The results of these new simulations are presented in Tables 2.13 and 2.14, which contain the estimates and their estimated asymptotic variances in parentheses. These results reveal that the for all estimates, in general, the biases and variances decrease as the sample size increases.

Table 2.13: MLEs for several a and b parameter values (variances in parentheses).

		$n=50$		$n=100$		$n=150$		$n=200$	
a	b	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
1	1	1.0521	1.0601	1.0262	1.0287	1.0170	1.0192	1.0131	1.0145
		(0.0400)	(0.0447)	(0.0185)	(0.0189)	(0.0121)	(0.0122)	(0.0089)	(0.0091)
1	2	1.0416	2.1546	1.0210	2.0733	1.0135	2.0487	1.0105	2.0368
		(0.0280)	(0.2526)	(0.0131)	(0.1029)	(0.0087)	(0.0661)	(0.0064)	(0.0484)
1	3	1.0380	3.2710	1.0191	3.1277	1.0123	3.0845	1.0096	3.0640
		(0.0240)	(0.7163)	(0.0113)	(0.2837)	(0.0075)	(0.1811)	(0.0055)	(0.1314)
1	4	1.0361	4.4045	1.0182	4.1898	1.0117	4.1254	1.0091	4.0947
		(0.0219)	(1.5130)	(0.0103)	(0.5873)	(0.0068)	(0.3728)	(0.0050)	(0.2687)
1	5	1.0350	5.5516	1.0176	5.2585	1.0114	5.1705	1.0088	5.1284
		(0.0206)	(2.6931)	(0.0097)	(1.0349)	(0.0064)	(0.6539)	(0.0047)	(0.4695)
1	6	1.0339	6.7013	1.0172	6.3317	1.0111	6.2184	1.0086	6.1645
		(0.0194)	(4.1665)	(0.0093)	(1.6460)	(0.0061)	(1.0363)	(0.0045)	(0.7410)
2	1	2.1041	1.0601	2.0524	1.0287	2.0340	1.0193	2.0262	1.0146
		(0.1599)	(0.0446)	(0.0739)	(0.0189)	(0.0485)	(0.0122)	(0.0355)	(0.0091)
2	2	2.0832	2.1545	2.0419	2.0732	2.0271	2.0488	2.0210	2.0368
		(0.1122)	(0.2526)	(0.0526)	(0.1028)	(0.0348)	(0.0661)	(0.0255)	(0.0484)
2	3	2.0760	3.2711	2.0382	3.1277	2.0247	3.0847	2.0192	3.0640
		(0.0960)	(0.7158)	(0.0451)	(0.2837)	(0.0299)	(0.1812)	(0.0219)	(0.1314)
2	4	2.0723	4.4044	2.0363	4.1900	2.0234	4.1254	2.0182	4.0947
		(0.0875)	(1.5131)	(0.0412)	(0.5872)	(0.0273)	(0.3726)	(0.0200)	(0.2685)
2	5	2.0699	5.5510	2.0352	5.2581	2.0227	5.1702	2.0176	5.1282
		(0.0821)	(2.6869)	(0.0387)	(1.0351)	(0.0257)	(0.6539)	(0.0189)	(0.4691)
2	6	2.0678	6.7009	2.0343	6.3319	2.0222	6.2184	2.0173	6.1646
		(0.0778)	(4.1584)	(0.0370)	(1.6465)	(0.0246)	(1.0359)	(0.0180)	(0.7399)
3	1	3.1562	1.0601	3.0786	1.0287	3.0509	1.0192	3.0392	1.0145
		(0.3598)	(0.0447)	(0.1663)	(0.0189)	(0.1093)	(0.0122)	(0.0799)	(0.0091)
3	2	3.1248	2.1546	3.0628	2.0732	3.0405	2.0487	3.0315	2.0369
		(0.2525)	(0.2527)	(0.1182)	(0.1028)	(0.0782)	(0.0661)	(0.0573)	(0.0484)
3	3	3.1141	3.2710	3.0573	3.1277	3.0371	3.0847	3.0287	3.0639
		(0.2159)	(0.7162)	(0.1014)	(0.2837)	(0.0673)	(0.1811)	(0.0493)	(0.1314)
3	4	3.1085	4.4044	3.0545	4.1898	3.0352	4.1255	3.0273	4.0947
		(0.1968)	(1.5079)	(0.0927)	(0.5872)	(0.0615)	(0.3726)	(0.0450)	(0.2686)
3	5	3.1047	5.5492	3.0527	5.2580	3.0341	5.1702	3.0265	5.1285
		(0.1846)	(2.6673)	(0.0871)	(1.0340)	(0.0579)	(0.6537)	(0.0424)	(0.4689)
3	6	3.1018	6.7026	3.0515	6.3317	3.0333	6.2185	3.0259	6.1646
		(0.1750)	(4.1982)	(0.0833)	(1.6463)	(0.0553)	(1.0364)	(0.0406)	(0.7403)
4	1	4.2082	1.0601	4.1048	1.0287	4.0679	1.0192	4.0523	1.0146
		(0.6396)	(0.0446)	(0.2956)	(0.0189)	(0.1945)	(0.0122)	(0.1421)	(0.0091)
4	2	4.1663	2.1544	4.0838	2.0732	4.0541	2.0487	4.0419	2.0368
		(0.0487)	(0.2527)	(0.2101)	(0.1028)	(0.1390)	(0.0661)	(0.1019)	(0.0484)
4	3	4.1521	3.2711	4.0765	3.1277	4.0495	3.0848	4.0383	3.0640
		(0.3839)	(0.7160)	(0.1804)	(0.2838)	(0.1195)	(0.1811)	(0.0876)	(0.1314)
4	4	4.1447	4.4046	4.0727	4.1899	4.0470	4.1255	4.0365	4.0946
		(0.3502)	(1.5122)	(0.1647)	(0.5872)	(0.1092)	(0.3724)	(0.0801)	(0.2686)
4	5	4.1398	5.5499	4.0703	5.2581	4.0455	5.1704	4.0353	5.1284
		(0.3283)	(2.6737)	(0.1549)	(1.0347)	(0.1028)	(0.6539)	(0.0754)	(0.4691)
4	6	4.1356	6.7007	4.0687	6.3318	4.0444	6.2185	4.0345	6.1646
		(0.3107)	(4.1586)	(0.1481)	(1.6447)	(0.0983)	(1.0354)	(0.0721)	(0.7403)

Table 2.14: MLEs for several a and b parameter values (variances in parentheses).

		$n=50$		$n=100$		$n=150$		$n=120$	
a	b	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
5	1	5.2603	1.0601	5.1309	1.0287	5.0847	1.0192	5.0651	1.0145
		(0.9996)	(0.0447)	(0.4618)	(0.0189)	(0.3037)	(0.0122)	(0.2220)	(0.0091)
5	2	5.2077	2.1544	5.1047	2.0732	5.0677	2.0487	5.0523	2.0368
		(0.7010)	(0.2525)	(0.3284)	(0.1029)	(0.2172)	(0.0660)	(0.1591)	(0.0484)
5	3	5.1902	3.2711	5.0956	3.1277	5.0617	3.0847	5.0478	3.0638
		(0.6001)	(0.7159)	(0.2817)	(0.2836)	(0.1869)	(0.1812)	(0.1369)	(0.1314)
5	4	5.1809	4.4045	5.0908	4.1898	5.0587	4.1254	5.0456	4.0947
		(0.5471)	(1.5100)	(0.2573)	(0.5867)	(0.1707)	(0.3726)	(0.1252)	(0.2687)
5	5	5.1748	5.5504	5.0879	5.2579	5.0568	5.1702	5.0442	5.1284
		(0.5130)	(2.6726)	(0.2420)	(1.0348)	(0.1606)	(0.6536)	(0.1177)	(0.4690)
5	6	5.1691	6.6984	5.0858	6.3316	5.0555	6.2184	5.0431	6.1647
		(0.4844)	(4.1254)	(0.2315)	(1.6452)	(0.1536)	(1.0357)	(0.1126)	(0.7403)
6	1	6.3122	1.0601	6.1572	1.0287	6.1017	1.0192	6.0783	1.0145
		(1.4391)	(0.0447)	(0.6648)	(0.0189)	(0.4373)	(0.0122)	(0.3197)	(0.0091)
6	2	6.2496	2.1545	6.1256	2.0732	6.0811	2.0487	6.0628	2.0368
		(1.0098)	(0.2526)	(0.4729)	(0.1028)	(0.3129)	(0.0661)	(0.2292)	(0.0484)
6	3	6.2280	3.2709	6.1146	3.1277	6.0741	3.0846	6.0575	3.0639
		(0.8640)	(0.7163)	(0.4057)	(0.2838)	(0.2690)	(0.1811)	(0.1972)	(0.1315)
6	4	6.2171	4.4047	6.1090	4.1897	6.0705	4.1255	6.0547	4.0947
		(0.7875)	(1.5108)	(0.3705)	(0.5867)	(0.2459)	(0.3727)	(0.1802)	(0.2687)
6	5	6.2096	5.5503	6.1055	5.2581	6.0681	5.1701	6.0529	5.1283
		(0.7389)	(2.6818)	(0.3486)	(1.0346)	(0.2313)	(0.6535)	(0.1695)	(0.4688)
6	6	6.2033	6.6998	6.1031	6.3317	6.0666	6.2186	6.0518	6.1646
		(0.6989)	(4.1357)	(0.3334)	(1.6468)	(0.2212)	(1.0354)	(0.1622)	(0.7403)
7	1	7.3644	1.0601	7.1832	1.0287	7.1187	1.0192	7.0915	1.0146
		(1.9582)	(0.0446)	(0.9051)	(0.0189)	(0.5953)	(0.0122)	(0.4351)	(0.0091)
7	2	7.2913	2.1545	7.1467	2.0733	7.0947	2.0487	7.0734	2.0368
		(1.3737)	(0.2526)	(0.6438)	(0.1029)	(0.4259)	(0.0661)	(0.3118)	(0.0483)
7	3	7.2661	3.2710	7.1338	3.1277	7.0865	3.0846	7.0670	3.0639
		(1.1757)	(0.7158)	(0.5522)	(0.2837)	(0.3661)	(0.1811)	(0.2683)	(0.1314)
7	4	8.2893	4.4043	7.1272	4.1899	7.0823	4.1255	7.0639	4.0947
		(1.4000)	(1.5082)	(0.5044)	(0.5872)	(0.3346)	(0.3726)	(0.2453)	(0.2686)
7	5	7.2447	5.5505	7.1230	5.2581	7.0796	5.1703	7.0618	5.1285
		(1.0061)	(2.6839)	(0.4743)	(1.0349)	(0.3149)	(0.6535)	(0.2308)	(0.4690)
7	6	7.2380	6.7034	7.1202	6.3318	7.0778	6.2187	7.0604	6.1646
		(0.9544)	(4.1961)	(0.4535)	(1.6468)	(0.3011)	(1.0358)	(0.2208)	(0.7402)
8	1	8.4162	1.0600	8.2095	1.0286	8.1358	1.0193	8.1044	1.0145
		(2.5586)	(0.0447)	(1.1824)	(0.0189)	(0.7775)	(0.0122)	(0.5680)	(0.0091)
8	2	8.3329	2.1545	8.1677	2.0733	8.1084	2.0488	8.0838	2.0368
		(1.7944)	(0.2525)	(0.8407)	(0.1029)	(0.5563)	(0.0661)	(0.4073)	(0.0484)
8	3	8.3042	3.2711	8.1528	3.1277	8.0989	3.0846	8.0765	3.0639
		(1.5356)	(0.7159)	(0.7213)	(0.2837)	(0.4782)	(0.1811)	(0.3505)	(0.1314)
8	4	8.2893	4.4043	8.1452	4.1898	8.0939	4.1255	8.0729	4.0946
		(1.4000)	(1.5082)	(0.6589)	(0.5871)	(0.4371)	(0.3725)	(0.3205)	(0.2686)
8	5	8.2796	5.5503	8.1406	5.2580	8.0909	5.1702	8.0706	5.1284
		(1.3132)	(2.6715)	(0.6196)	(1.0344)	(0.4112)	(0.6534)	(3.3015)	(0.4691)
8	6	8.2715	6.7023	8.1373	6.3317	8.0890	6.2186	8.0689	6.1645
		(1.2452)	(4.1991)	(0.5922)	(1.6454)	(0.3933)	(1.0358)	(0.2885)	(0.7402)

CHAPTER 3

The Exponentiated Generalized Extended Exponential Distribution

Paper published in the *Journal of Data Science*, 14, 393-414, 2016.

Abstract

We introduce and study a new four-parameter lifetime model named the *exponentiated generalized extended exponential* distribution. The proposed model has the advantage of including as special cases the exponential and exponentiated exponential distributions, among others, and its hazard function can take the classic shapes: bathtub, inverted bathtub, increasing, decreasing and constant, among others. We derive some mathematical properties of the new model such as a representation for the density function as a double linear combination of Erlang densities, explicit expressions for the quantile function, ordinary and incomplete moments, mean deviations, Bonferroni and Lorenz curves, generating function, Rényi entropy, density of order statistics and reliability. We use the maximum likelihood method to estimate the model parameters and provide the elements of the score vector. Two applications to real data illustrates the flexibility of the proposed model.

Keywords: Erlang distribution. Extended exponential distribution. Hazard rate function. Lifetime distribution. Moments.

Resumo

Nós introduzimos e estudamos um novo modelo com quatro parâmetros que pode ser usado para ajustar dados de sobrevivência, chamado de distribuição exponencial estendida exponencializada generalizada. O modelo proposto tem a vantagem de incluir como casos especiais as distribuições exponencial e exponencial exponencializada, dentre outras, além de possuir função de taxa de falha que assume as formas clássicas descritas na literatura: banheira, banheira invertida, crescente, decrescente e constante. Nós derivamos algumas propriedades do novo modelo, tais como uma representação para a função densidade como uma combinação linear dupla de densidades Erlang, expressões explícitas para a função quantílica, momentos ordinários e incompletos, desvios médios, curvas de Bonferroni e Lorenz, função geratriz de momentos, entropia de Rényi, densidade da estatística de ordem e confiabilidade. Nós usamos o método de máxima verossimilhança para estimar os parâmetros do modelo e fornecemos o vetor escore. Duas aplicações à dados reais ilustram a flexibilidade do modelo proposto.

Palavras-chave: Distribuição de Erlang. Distribuição exponencial estendida. Função de risco. Modelo de sobrevivência. Momentos.

3.1 Introduction

The exponential distribution is a very popular statistical model and, probably, is one of the parametric models most extensively applied in several fields ([44]). The popularity of this distribution can be explained, perhaps, by the simplicity of their cumulative function, which involves only one unknown parameter $\lambda > 0$ and takes a simple form $G(x) = 1 - \mathrm{e}^{-\lambda x}$, for $x > 0$, in addition to having constant failure rate function. Due to its importance, several studies introducing and/or studying extensions of the exponential distribution are available in the literature. Here, we refer to the following papers: [45], [18], [46] and [44], to mention a few. Recently, [11] introduced the *extended exponential* (EE for short) distribution. Its cumulative distribution function (cdf) and probability density function (pdf) (for $x > 0$) are given by

$$G(x;\alpha,\beta) = \frac{\alpha+\beta-(\beta+\alpha+\alpha\beta x)\,\mathrm{e}^{-\alpha x}}{\alpha+\beta} \tag{3.1}$$

and

$$g(x;\alpha,\beta) = \frac{\alpha^2\,(1+\beta x)\,\mathrm{e}^{-\alpha x}}{\alpha+\beta}, \tag{3.2}$$

respectively, where $\alpha > 0$ and $\beta \geq 0$.

Here, we point out that the density (3.2) given above can be also obtained as a special case of the generalized Lindley distribution, proposed by [47]. However, these latter authors do not address this particular case in their research.

Several mathematical properties of the EE distribution, including expectation, variance, moment generating function (mgf), asymmetry and kurtosis coefficients, among others, were studied by [11]. In particular, these authors proved that the density of the EE model is a linear combination of the exponential and gamma densities. We believe that the addition of parameters to the EE model may generate new distributions with great adjustment capability and, for this reason, we propose a generalization of it.

In a recent paper, [10] proposed a new way of adding two parameters to a continuous distribution. For a given continuous baseline cdf $G(x)$, and $x \in \mathbb{R}$, they defined the *exponentiated generalized* (EG) class of distributions with two extra shape parameters $a > 0$ and $b > 0$ and cdf $F(x)$ and pdf $f(x)$ given by

$$F(x) = \{1-[1-G(x)]^a\}^b \tag{3.3}$$

and

$$f(x) = a\,b\,[1-G(x)]^{a-1}\,\{1-[1-G(x)]^a\}^{b-1}\,g(x), \tag{3.4}$$

respectively, in which are implicit the dependence on the parameters of $G(x)$.

To illustrate the flexibility of the EG model, [10] applied (3.3) to extend some well-known distributions such as the Fréchet, normal, gamma and Gumbel distributions. Moreover, these authors presented several properties for the EG class, which constitute motivations to adopt this generator. Next, we discuss some of these motivations. The first important point to note is the simplicity of equations (3.3) and (3.4). They have no complicated functions and will be always tractable when the cdf and pdf of the baseline distribution have simple analytic expressions. It is very easy, for example, to obtain the inverse of the cdf (3.3). Another important feature is that the EG model contains as especial cases the two classes of Lehmann's alternatives. In fact, for $a = 1$, (3.3) reduces to $F(x) = G(x)^b$ and for $b = 1$ we obtain $F(x) = 1 - [1 - G(x)]^a$, which correspond to the cdf's of the Lehmann type I and II families ([17]), respectively. For this reason, the EG model *encompasses both Lehmann type I and Lehmann type II classes*. So, the EG family can be derived from a double transformation using these classes. The two extra parameters a and b in the density (3.4) can control both tail weights, allowing generate flexible distributions, with heavier or lighter tails, as appropriate. There is also an attractive physical interpretation of the model (3.3) when a and b are positive integers. This interpretation is described in [3].

The above properties and many others have been discussed and explored in recent works for the EG class. Here, we refer to the papers: [48], [3], [49], [50] and [51], which used the EG class to extend the Burr III, Birnbaum-Saunders, inverse Weibull, inverted exponential and generalized gamma distributions, respectively.

In this chapter, we propose the so-called *exponentiated generalized extended exponential* (EGEE) distribution determined by inserting (3.1) in equation (3.3). The EGEE model includes as special cases the exponential, Lindley and exponentiated exponential distributions, among others, which are very important statistical models, especially for applied works. The density function of the new distribution is a double linear combination of Erlang densities, and thus the derivation of several properties of the EGEE model can be simplified from this relationship. Moreover, the proposed model has monotonic and non-monotonic hazard rate functions. We hope that this new distribution can be widely used for data modelling in areas such as economics, finance, reliability, biology and medicine, among others.

The rest of the chapter is organized as follows. In Section 3.2, we provide the density and hazard functions of the EGEE model with corresponding plots for selected parameter values. We present explicit expressions for the cumulative and reversed hazard functions. In Section 3.3, we discuss the shapes of the EGEE density function. In Section 3.4, we derive several mathematical properties of the proposed model, including linear combination representations for the density and culmulative functions, explicit expressions for the quantile and generating functions, ordinary and incomplete moments, among others. Estimation and inference by maximum likelihood are discussed in Section 3.5. Two applications to real data are presented in Section 3.6. Section 3.7 provides concluding remarks.

3.2 The EGEE distribution

The cdf and pdf of the EGEE distribution (by omitting the dependence on the parameters $a > 0$, $b > 0$, $\alpha > 0$ and $\beta \geq 0$) are given by

$$F(x) = \frac{[(\alpha+\beta)^a - (\beta+\alpha+\alpha\beta x)^a \, \mathrm{e}^{-a\alpha x}]^b}{(\alpha+\beta)^{ab}} \tag{3.5}$$

and

$$f(x) = \frac{ab\alpha^2}{(\alpha+\beta)^{ab}} \, (1+\beta x) \, \mathrm{e}^{-a\alpha x} \, (\beta+\alpha+\alpha\beta x)^{a-1} \, [(\alpha+\beta)^a - (\beta+\alpha+\alpha\beta x)^a \, \mathrm{e}^{-a\alpha x}]^{b-1}, \tag{3.6}$$

respectively, for $x > 0$. Henceforth, a continuous random variable X with positive support and having pdf (3.6) is denoted by $X \sim EGEE\big(a, b, \alpha, \beta\big)$.

Several distributions are special cases of the EGEE model. Here, we mention some of them. Clearly, the EE distribution is a basic exemplar when $a = b = 1$. The exponential and Lindley distributions are obtained from (3.6) by setting $\beta = 0$ and $\beta = 1$, respectively, in addition to $a = b = 1$. The exponentiated generalized exponential (EGE) model is obtained from (3.6) by setting $\beta = 0$ and the exponentiated generalized Lindley (EGL) distribution follows when $\beta = 1$. The EGEE model also includes the Lehmann type I and type II transformations of the EE, exponential and Lindley distributions. For example, the widely known exponentiated exponential distribution ([45]), also referred in the literature as the generalized exponential distribution, follows when $\beta = 0$ and $a = 1$. For a brief discussion and some properties of the exponentiated exponential distribution, see a recent paper by [44]. The exponentiated Lindley (EL) model by [52] (they called the generalized Lindley distribution, but here we adopt the EL terminology) comes when $a = \beta = 1$.

The hazard rate function (hrf) and reversed hazard rate function (rhrf) of X are given by

$$h(x) = \frac{ab\alpha^2\,(1+\beta x)\,\mathrm{e}^{-a\alpha x}\,(\beta+\alpha+\alpha\beta x)^{a-1}\,[(\alpha+\beta)^a - (\beta+\alpha+\alpha\beta x)^a\,\mathrm{e}^{-a\alpha x}]^{b-1}}{(\alpha+\beta)^{ab} - [(\alpha+\beta)^a - (\beta+\alpha+\alpha\beta x)^a\,\mathrm{e}^{-a\alpha x}]^b}$$

and

$$\tau(x) = \frac{ab\alpha^2\,(1+\beta x)\,\mathrm{e}^{-a\alpha x}\,(\beta+\alpha+\alpha\beta x)^{a-1}}{[(\alpha+\beta)^a - (\beta+\alpha+\alpha\beta x)^a\,\mathrm{e}^{-a\alpha x}]},$$

respectively.

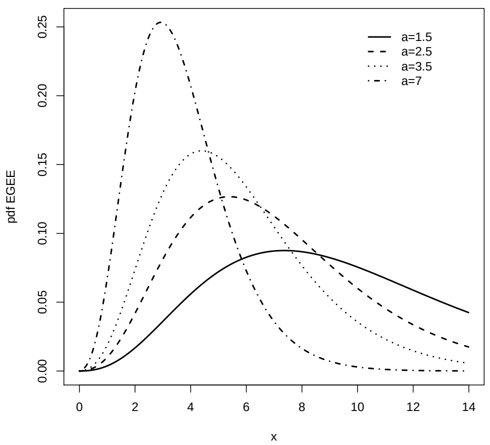
Plots of the EGEE density for selected parameter values are displayed in Figure 3.1. Figure 3.2 provides some possible shapes of the EGEE hazard function for appropriate choice of the parameter values, including bathtub, inverted bathtub, increasing, decreasing, constant and decreasing-increasing-decreasing shapes. These plots indicate that the EGEE model is fairly flexible and can be used to fit several types of positive data.

3.3 Shapes

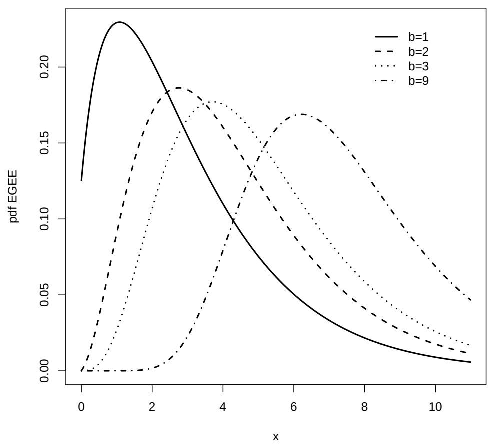
The first derivative of $\log\{f(x)\}$ for the EGEE model is given by

$$\frac{\mathrm{d}\log\{f(x)\}}{\mathrm{d}x} = -a\,\alpha + \frac{\beta}{1+\beta x} + \frac{(a-1)\,\alpha\,\beta}{z(x)} + \frac{a\,(b-1)\,\alpha\,z(x)^a\,[1-\beta/z(x)]\,\mathrm{e}^{-a\alpha x}}{(\alpha+\beta)^a - z(x)^a\,\mathrm{e}^{-a\alpha x}},$$

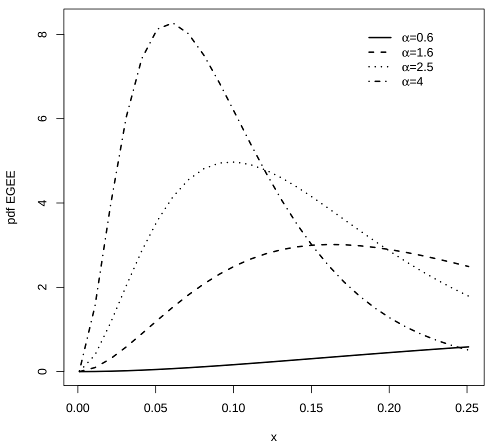
where $z(x) = \beta + \alpha + \alpha\beta x$.



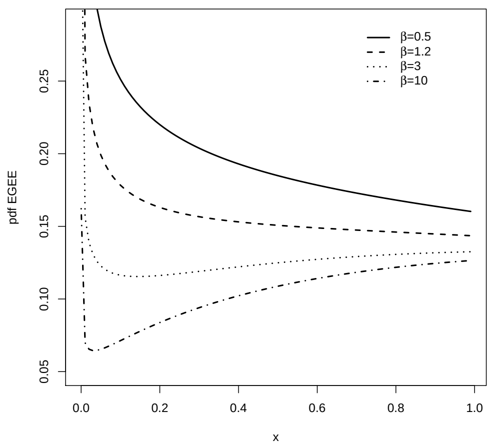
(a) $b = 2$, $\alpha = 0.2$ and $\beta = 6$



(b) $a = 0.5$, $\alpha = 1$ and $\beta = 3$

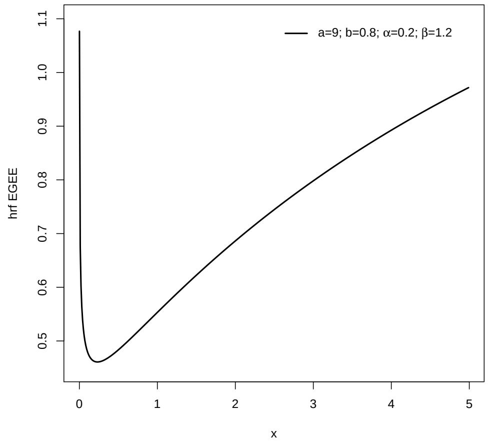


(c) $a = 5$, $b = 3$ and $\beta = 0.3$

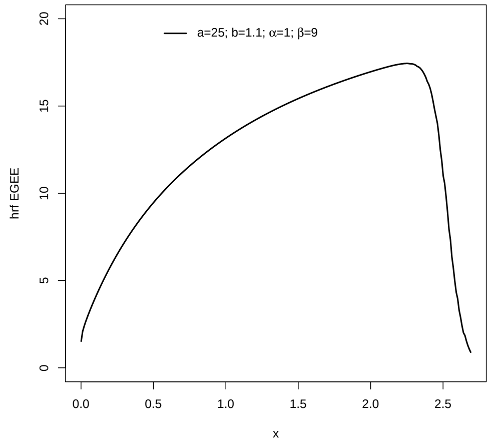


(d) $a = 0.4$, $b = 0.8$ and $\alpha = 0.6$

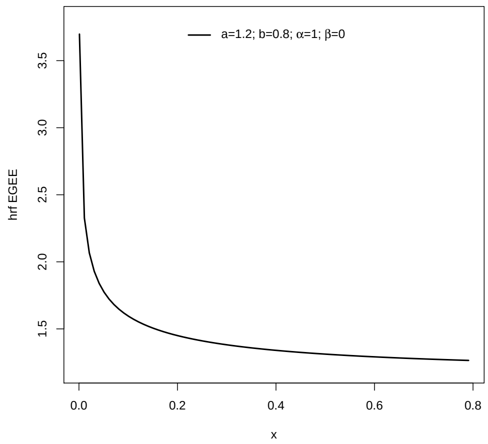
Figure 3.1: Plots of the EGEE density function for some parameter values.



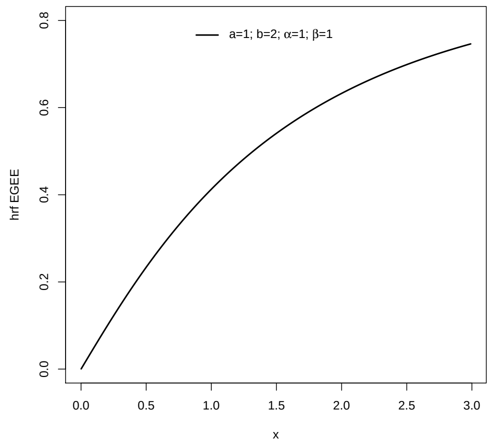
(a) bathtub



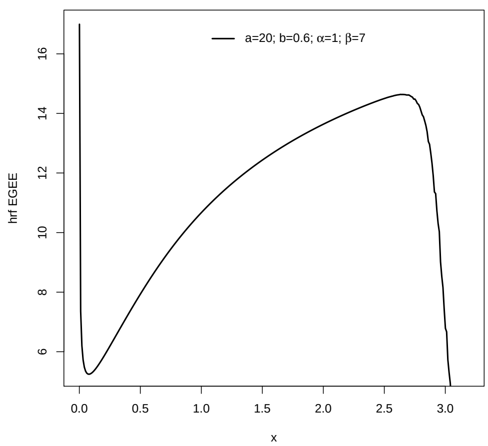
(b) inverted bathtub



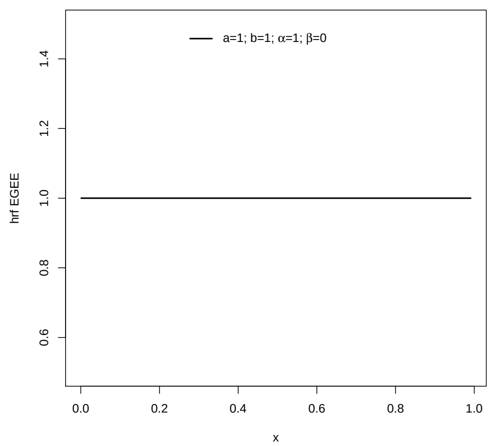
(c) decreasing



(d) increasing



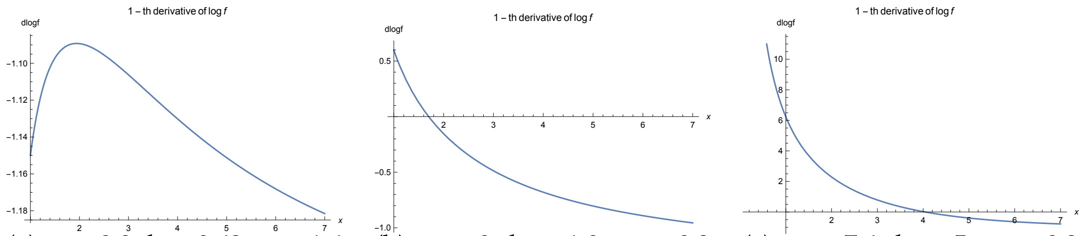
(e) decreasing-increasing-decreasing



(f) constant

Figure 3.2: Plots of the EGEE hazard function for some parameter values.

Some plots of the first derivative of $\log\{f(x)\}$ for selected parameter values are displayed in Figure 3.3. These plots are constructed using the *Wolfram Mathematica* software.



(a) $a = 0.8$, $b = 0.62$, $\alpha = 1.6$ and $\beta = 0.8$

(b) $a = 8$, $b = 1.9$, $\alpha = 0.2$ and $\beta = 0.7$

(c) $a = 7.6$, $b = 5$, $\alpha = 0.2$ and $\beta = 5$

Figure 3.3: Plots of the first derivative of $\log\{f(x)\}$.

Thus, the critical values of $f(x)$ are the roots of the equation:

$$\frac{a\,(b-1)\,\alpha\, z(x)^a\,[1-\beta/z(x)]\,\mathrm{e}^{-a\alpha x}}{(\alpha+\beta)^a - z(x)^a\,\mathrm{e}^{-a\alpha x}} = a\,\alpha - \frac{\beta}{1+\beta x} - \frac{(a-1)\,\alpha\,\beta}{z(x)}. \tag{3.7}$$

If the point $x = x_0$ is a root of (3.7), then we can classify it as local maximum, local minimum or inflection point when $\lambda(x_0) < 0$, $\lambda(x_0) > 0$ and $\lambda(x_0) = 0$, respectively, where $\lambda(x) = \mathrm{d}^2 \log\{f(x)\}/\mathrm{d}x^2$ is given by

$$\begin{aligned}\lambda(x) = &-\frac{\beta^2}{(1+\beta x)^2} - \frac{(a-1)\,\alpha^2\,\beta^2}{z(x)^2} - \frac{(b-1)\{a\,\alpha\, z(x)^a\,[1-\beta/z(x)]\,\mathrm{e}^{-a\alpha x}\}^2}{[(\alpha+\beta)^a - z(x)^a\,\mathrm{e}^{-a\alpha x}]^2}\\ &+ \frac{a\,(b-1)\,\alpha^2\, z(x)^a\,[a\,\beta\, z(x)^{-1} - a - (a-1)\,\beta^2\, z(x)^{-2} + a\,\beta\, z(x)^{-1}]\,\mathrm{e}^{-a\alpha x}}{(\alpha+\beta)^a - z(x)^a\,\mathrm{e}^{-a\alpha x}}.\end{aligned}$$

3.4 Properties

In this section, we study some structural properties of the EGEE distribution.

3.4.1 A useful representation

First, we derive simple representations for the density and cumulative functions of the EGEE distribution. The starting point of our approach is the class of exponentiated distributions, which has been widely explored in recent works. For an arbitrary continuous baseline cdf $G(x)$, a random variable Y is said to have the exponentiated-G ("exp-G" for short) distribution with power parameter $c > 0$, say $Y \sim$ exp-$G(c)$, if its cdf and pdf are $H_c(x) = G(x)^c$ and $h_c(x) = c\, g(x) G(x)^{c-1}$, respectively. Thus, "exp-G" denotes the Lehmann type I transformation of $G(x)$. Using some results in [3], we can express the EG cdf (3.3) as

$$F(x) = \sum_{j=0}^{\infty} w_{j+1}\, H_{j+1}(x), \tag{3.8}$$

where $w_{j+1} = \sum_{m=1}^{\infty} (-1)^{j+m+1} \binom{b}{m} \binom{m\,a}{j+1}$ and $H_{j+1}(x) = G(x)^{j+1}$ is the exp-G cdf with power parameter $j+1$. By differentiating (3.8), we obtain a similar linear combination representation for $f(x)$ as

$$f(x) = \sum_{j=0}^{\infty} w_{j+1}\, h_{j+1}(x), \tag{3.9}$$

where $h_{j+1}(x) = \mathrm{d}H_{j+1}(x)/\mathrm{d}x$.

By using (3.8) and (3.9) for the EE distribution (3.1), $h_{j+1}(x)$ becomes the exp-EE pdf with power parameter $j+1$ (for $j \geq 0$) given by

$$h_{j+1}(x) = \frac{(j+1)\,\alpha^2}{(\alpha+\beta)^{j+1}}\,(1+\beta x)\,\mathrm{e}^{-\alpha x}\;[\alpha+\beta-(\beta+\alpha+\alpha\beta x)\,\mathrm{e}^{-\alpha x}]^j. \tag{3.10}$$

Combining equations (3.9) and (3.10) we have an important result: the EGEE density function is a linear combination of exp-EE densities. This result can be used to derive some EGEE mathematical properties.

Next, we apply the binomial expansion in equation (3.10) to obtain a simple representation for the exp-EE density. We have

$$h_{j+1}(x) = \sum_{k=0}^{j}\sum_{\ell=0}^{k}\sum_{i=0}^{\ell+1} \frac{(j+1)\,(-1)^k\,\alpha^{\ell+2}\,\beta^{i+k-\ell}}{(\alpha+\beta)^{k+1}} \binom{j}{k}\binom{k}{\ell}\binom{\ell+1}{i} x^i\,\mathrm{e}^{-\alpha(k+1)x}.$$

By interchanging $\sum_{\ell=0}^{k}\sum_{i=0}^{\ell+1}$ by $\sum_{i=0}^{k+1}\sum_{\ell=\delta_i}^{k}$ in the last equation, where

$$\delta_i = \begin{cases} 0, & \text{if } i = 0, 1 \\ i-1, & \text{if } i \geq 2, \end{cases}$$

and, after a simple algebraic manipulation, we obtain

$$h_{j+1}(x) = \sum_{k=0}^{j} \sum_{i=0}^{k+1} p_{k,i}^{(j+1)} \, \pi(x; i+1, (k+1)\alpha), \tag{3.11}$$

where

$$p_{k,i}^{(j+1)} = \sum_{\ell=\delta_i}^{k} \frac{(j+1)\,(-1)^k\, \alpha^{\ell-i+1}\, \beta^{i+k-\ell}\, i!}{(k+1)^{i+1}(\alpha+\beta)^{k+1}} \binom{j}{k} \binom{k}{\ell} \binom{\ell+1}{i}. \tag{3.12}$$

Here, $\pi(x; i+1, (k+1)\alpha)$ denotes the Erlang density with shape parameter $i+1$ (for $i \geq 0$) and scale parameter $(k+1)\alpha$. If Z is a Erlang random variable with shape parameter $s\,(= 1, 2, 3, \ldots)$ and scale parameter $\lambda > 0$, its pdf is given by $\pi(z; s, \lambda) = \lambda^s\, z^{s-1}\, \mathrm{e}^{-\lambda z} / (s-1)!$.

Second, combining equations (3.9) and (3.11) and changing $\sum_{j=0}^{\infty} \sum_{k=0}^{j}$ by $\sum_{k=0}^{\infty} \sum_{j=k}^{\infty}$, the EGEE density function reduces to

$$f(x) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} v_{k,i}\, \pi(x; i+1, (k+1)\alpha), \tag{3.13}$$

where $v_{k,i} = \sum_{j=k}^{\infty} w_{j+1}\, p_{k,i}^{(j+1)}$.

Equation (3.13) is the main result of this section. Based on this equation, we conclude that the density function of X can be expressed as a double linear combination of Erlang densities. This result is important to derive some mathematical properties of X such as the ordinary and incomplete moments, generating function and mean deviations from those of the Erlang distribution. We can take the upper limit of k to be equal 20 in equation (3.13) for most practical purposes.

3.4.2 Quantile function

For many applications it is important to determine the quantile function (qf) of X. Based on this function, we can, for example, generate variates and obtain the median of the EGEE distribution. By inverting (3.5), the qf of X can be expressed as

$$Q(u) = -\frac{1}{\alpha} - \frac{1}{\beta} - \frac{1}{\alpha}\,\mathrm{W}\left\{-\frac{1}{\beta}\,(\alpha+\beta)\,(1-u^{1/b})^{1/a}\,\exp[-(\alpha+\beta)/\beta]\right\}, \qquad (3.14)$$

where $0 < u < 1$ and $\mathrm{W}(\cdot)$ denotes the Lambert W-function. The median of X, say M, is obtained by $M = Q(1/2)$.

In a recent paper, [52] use the Lambert W-function to obtain the qf of the EL distribution. For any complex t, the Lambert W-function is defined as the inverse of the function $g(t) = t\,\mathrm{e}^t$. For more details, see http://mathworld.wolfram.com/LambertW-Function.html. An implementation in `R` software is available through the *LambertW* package. See http://cran.r-project.org/web/packages/LambertW/LambertW.pdf.

Using the Lagrange inversion theorem, the power series for the W-function holds:

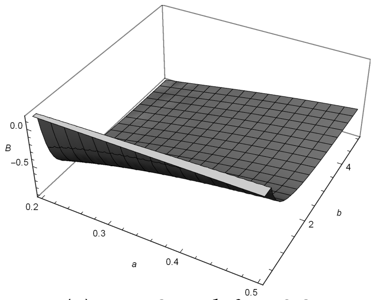
$$\mathrm{W}(z) = \sum_{k=1}^{\infty} \frac{(-k)^{k-1}}{k!}\,z^k. \qquad (3.15)$$

By applying (3.15) in equation (3.14), we have

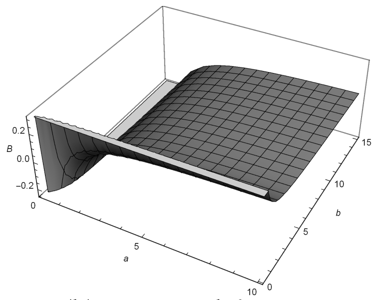
$$Q(u) = -\frac{1}{\alpha} - \frac{1}{\beta} - \frac{1}{\alpha}\sum_{k=1}^{\infty} \frac{(-1)^k\,(-k)^{k-1}}{\beta^k\,k!}\,(\alpha+\beta)^k\,(1-u^{1/b})^{k/a}\,\exp[-k(\alpha+\beta)/\beta].$$

The effects of the additional shape parameters a and b on the skewness and kurtosis of the EGEE model can be based on quantile measures. The Bowley skewness is based on quartiles $\mathrm{B} = [Q(3/4) - 2Q(1/2) + Q(1/4)]/[Q(3/4) - Q(1/4)]$, whereas the Moors kurtosis is based on octiles $\mathrm{M} = [Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)]/[Q(6/8) - Q(2/8)]$. These measures are fairly considered in the literature. Here, we refer to the following works: [53], [54], [55] and [56], among others.

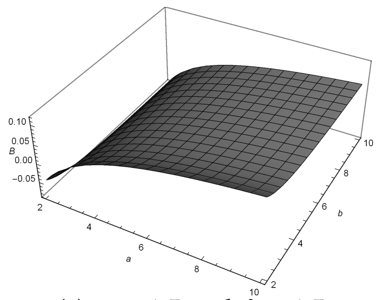
In Figures 3.4 and 3.5, we present 3D plots of B and M measures for some parameters values. These plots are obtained using the *Wolfram Mathematica* software. Based on these plots, it is possible to conclude that changes in the additional parameters a and b have a considerable impact on the skewness and kurtosis of the EGEE model, thus showing its greater flexibility. So, theses plots reinforce the importance of the additional parameters.



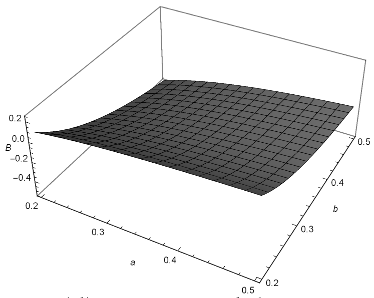
(a) $\alpha = 3$ and $\beta = 0.3$



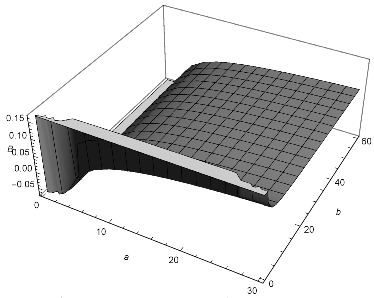
(b) $\alpha = 3$ and $\beta = 0.7$



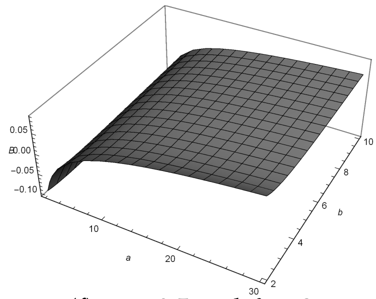
(c) $\alpha = 1.5$ and $\beta = 1.5$



(d) $\alpha = 0.2$ and $\beta = 5$

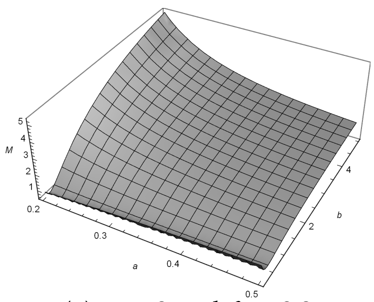


(e) $\alpha = 0.5$ and $\beta = 0.9$

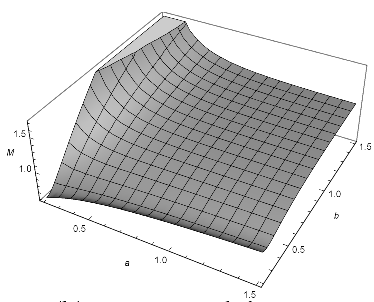


(f) $\alpha = 0.5$ and $\beta = 2$

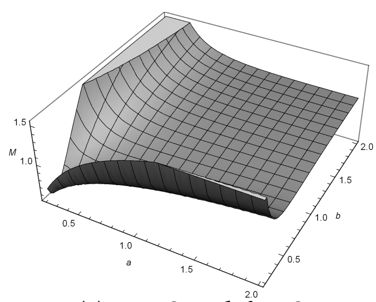
Figure 3.4: Plots of the Bowley skewness for the EGEE distribution.



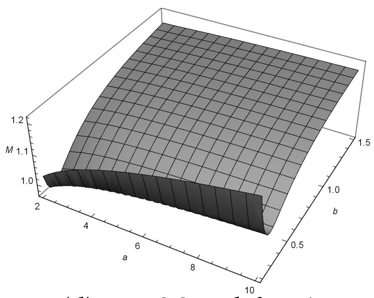
(a) $\alpha = 3$ and $\beta = 0.3$



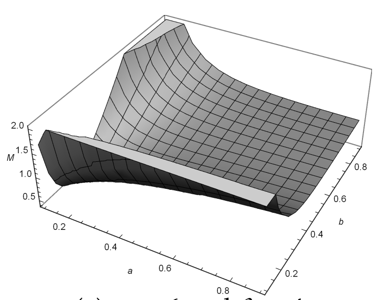
(b) $\alpha = 0.3$ and $\beta = 0.3$



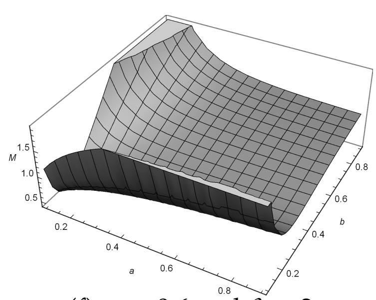
(c) $\alpha = 2$ and $\beta = 2$



(d) $\alpha = 0.2$ and $\beta = 4$



(e) $\alpha = 6$ and $\beta = 4$



(f) $\alpha = 0.6$ and $\beta = 2$

Figure 3.5: Plots of the Moors kurtosis for the EGEE distribution.

3.4.3 Moments

It is hardly necessary to emphasize the importance of the moments of a random variable. Here, for illustrative purposes, we provide the first six moments of X obtained by numerical integration. The results from (3.6) as $E(X^n) = \int_0^\infty x^n f(x)dx$ are present in Table 3.1. All computations are obtained using *Wolfram Mathematica* software, which have numerical integration routines with great precision. We consider some a and b parameter values with $\alpha = 2$ and $\beta = 1$ fixed.

Based on the values in Table 3.1, we conclude that the additional parameters a and b have large impact on the moments of X. Theses values reveals that, in general, for fixed a parameter value, the moments increases when b increase. The inverse happens when we set values for b and the parameter a increases.

The nth moment of X can be also determined using the linear combination (3.13) of Erlang densities. Thus, the nth moment of X is given by

$$E(X^n) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} v_{k,i} \int_0^\infty x^n \, \pi(x; i+1, (k+1)\alpha) \, \mathrm{d}x$$

and we can be also write as

$$E(X^n) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \frac{(n+i)! \, v_{k,i}}{i! \, [(k+1)\alpha]^n}.$$

3.4.4 Incomplete moments

The incomplete moments of a distribution play an important role in applications. The nth incomplete moment of X is given by $T_n(z) = \int_0^z x^n f(x) \, \mathrm{d}x$, and using equation (3.13), we can write

$$T_n(z) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} v_{k,i} \int_0^z x^n \, \pi(x; i+1, (k+1)\alpha) \, \mathrm{d}x.$$

Thus, we have

$$T_n(z) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \frac{\{\Gamma(n+i+1) - \Gamma[n+i+1, z\,(k+1)\alpha]\} \, v_{k,i}}{i! \, [(k+1)\alpha]^n}, \tag{3.16}$$

where $\Gamma(a) = \int_0^\infty t^{a-1} \, \mathrm{e}^{-t} \mathrm{d}t$ is the gamma function and $\Gamma(a,z) = \int_z^\infty t^{a-1} \, \mathrm{e}^{-t} \mathrm{d}t$ denotes the upper incomplete gamma function.

Table 3.1: First sixth moments of X for several a and b values (with $\alpha = 2$ and $\beta = 1$).

a	b	$\mathrm{E}(X)$	$\mathrm{E}(X^2)$	$\mathrm{E}(X^3)$	$\mathrm{E}(X^4)$	$\mathrm{E}(X^5)$	$\mathrm{E}(X^6)$
	1	0.66667	0.83333	1.50000	3.50000	10.0000	33.7500
	2	0.98611	1.43750	2.78125	6.72917	19.5898	66.7676
1	3	1.19427	1.91927	3.91394	9.74748	28.8323	99.1301
	4	1.34813	2.32359	4.93712	12.5945	37.7736	130.900
	5	1.46994	2.67395	5.87523	15.2981	46.4496	162.126
	1	0.34722	0.22917	0.21875	0.27083	0.41016	0.73242
	2	0.51548	0.39652	0.40637	0.52125	0.80389	1.44926
2	3	0.62564	0.53048	0.57266	0.75566	1.18365	2.15212
	4	0.70724	0.64317	0.72314	0.97700	1.55126	2.84227
	5	0.77194	0.74097	0.86127	1.18737	1.90811	3.52080
	1	0.23594	0.10677	0.07019	0.05998	0.06274	0.07741
	2	0.35102	0.18511	0.13058	0.11554	0.12304	0.15323
3	3	0.42660	0.24801	0.18422	0.16762	0.18126	0.22760
	4	0.48271	0.30102	0.23284	0.21686	0.23765	0.30067
	5	0.52725	0.34710	0.27752	0.26370	0.29243	0.37253
	1	0.17896	0.06181	0.03113	0.02042	0.01642	0.01558
	2	0.26662	0.10732	0.05797	0.03936	0.03222	0.03085
4	3	0.32433	0.14394	0.08186	0.05715	0.04748	0.04584
	4	0.36723	0.17484	0.10353	0.07397	0.06227	0.06057
	5	0.40132	0.20173	0.12347	0.08999	0.07664	0.07506
	1	0.14425	0.04034	0.01650	0.00881	0.00577	0.00446
	2	0.21512	0.07012	0.03076	0.01699	0.01132	0.00884
5	3	0.26185	0.09412	0.04345	0.02467	0.01669	0.01314
	4	0.29663	0.11439	0.05499	0.03195	0.02190	0.01736
	5	0.32429	0.13205	0.06561	0.03888	0.02696	0.02152

The first incomplete moment of X is important to determine the mean deviations, which can be used to measure the amount of scatter in a population, and the Bonferroni and Lorenz curves, which are useful for applications in areas such as economics, reliability, demography and many others. Based on equation (3.16), for $n = 1$, we have

$$T_1(z) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \frac{\{\Gamma(i+2) - \Gamma[i+2, z\,(k+1)\alpha]\}\, v_{k,i}}{i!\,(k+1)\alpha}. \tag{3.17}$$

The mean deviations of X about the mean $\mu = E(X)$ and about the median M are given by $\delta_1 = \int_0^\infty |x-\mu|\, f(x)\,\mathrm{d}x$ and $\delta_2 = \int_0^\infty |x-M|\, f(x)\,\mathrm{d}x$, respectively, where $f(x)$ is the pdf (3.6). Using equation (3.17), these measures follow as

$$\delta_1(X) = 2[\mu\, F(\mu) - T_1(\mu)] \qquad \text{and} \qquad \delta_2(X) = \mu - 2\,T_1(M),$$

where $F(\mu)$ is the cdf (3.5) evaluated at μ and $T_1(z)$ is given by (3.17).

Equation (3.17) can also be adopted to obtain the Bonferroni and Lorenz curves of X given by $B(p) = T_1(q)/(p\mu)$ and $L(p) = T_1(q)/\mu$, respectively, where $q = Q(p)$ is given by (3.14) and p is a specified probability.

3.4.5 Generating function

The mgf of X can be determined from (3.13) as

$$M(t) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} v_{k,i} \int_0^\infty \mathrm{e}^{tx}\, \pi(x; i+1, (k+1)\alpha)\,\mathrm{d}x.$$

Then, for all $t < (k+1)\alpha$, we have

$$M(t) = \sum_{k=0}^{\infty} \sum_{i=0}^{k+1} \left[1 - \frac{t}{(k+1)\alpha}\right]^{-(i+1)} v_{k,i}.$$

3.4.6 Rényi entropy

The entropy of X is a measure of variation of the uncertainty. There are many entropy measures studied and discussed in the literature, but the Rényi entropy is perhaps one of the most popular. The Rényi entropy of X with density (3.6) is given by

$$I_R(\rho) \quad = \quad \frac{1}{(1-\rho)} \log\left(\int_0^\infty f(x)^\rho \mathrm{d}x\right). \tag{3.18}$$

where $\rho > 0$ and $\rho \neq 1$.

Now, we consider the generalized binomial expansion

$$(1-z)^b = \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} z^k, \tag{3.19}$$

which holds for any real non-integer b and $|z| < 1$. Using (3.19) twice in equation (3.4), we can write

$$f(x)^\rho = (ab)^\rho \sum_{k,\ell=0}^{\infty} (-1)^{k+\ell} \binom{\rho(b-1)}{k} \binom{ak+\rho(a-1)}{\ell} g(x)^\rho G(x)^\ell. \tag{3.20}$$

Inserting (3.1) and (3.2) in equation (3.20) and applying the binomial expansion twice from $[\alpha+\beta-(\beta+\alpha+\alpha\beta x)\,\mathrm{e}^{-\alpha x}]^\ell$, we obtain

$$\begin{aligned} f(x)^\rho = (ab)^\rho \sum_{k,\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{n=0}^{m} \frac{(-1)^{k+\ell+m}\, \alpha^{2\rho+n}\, \beta^{m-n}}{(\alpha+\beta)^{\rho+m}} \binom{\rho(b-1)}{k} \\ \times \binom{ak+\rho(a-1)}{\ell} \binom{\ell}{m} \binom{m}{n} (1+\beta x)^{\rho+n}\, \mathrm{e}^{-\alpha(\rho+m)x}. \end{aligned} \tag{3.21}$$

Then, by inserting (3.21) in equation (3.18), the Rényi entropy reduces to

$$\begin{aligned} I_R(\rho) = \frac{1}{(1-\rho)} \log\Bigg\{ (ab)^\rho \sum_{k,\ell=0}^{\infty} \sum_{m=0}^{\ell} \sum_{n=0}^{m} \frac{(-1)^{k+\ell+m}\, \alpha^{2\rho+n}\, \mathrm{e}^{\alpha(\rho+m)/\beta}}{\beta^{n+1-m}\, (\alpha+\beta)^{\rho+m}} \\ \times \binom{\rho(b-1)}{k} \binom{ak+\rho(a-1)}{\ell} \binom{\ell}{m} \binom{m}{n} \mathbb{E}[-(\rho+n), \alpha(\rho+m)/\beta] \Bigg\}, \end{aligned}$$

where $\mathbb{E}[\rho, z] = \int_1^\infty t^{-\rho}\, \mathrm{e}^{-zt} \mathrm{d}t$ is the *exponential integral function*.

3.4.7 Order statistics

The density function $f_{i:n}(x)$ of the ith order statistic, say $X_{i:n}$, for $i = 1, \ldots, n$, from a random sample $X_1, \ldots, X_n$ having the EG distribution can be expressed as

$$f_{i:n}(x) = \frac{f(x)}{B(i, n-i+1)}\, F(x)^{i-1}\, [1 - F(x)]^{n-i},$$

where $f(x)$ is the pdf (3.4) and $F(x)$ is the cdf (3.3).

Applying the binomial expansion in the last equation, we have

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x)\, F(x)^{i+j-1}. \tag{3.22}$$

Substituting (3.3) and (3.4) in equation (3.22) and applying the generalized binomial expansion (3.19), we can write

$$f_{i:n}(x) = \frac{ab}{B(i, n-i+1)} \sum_{\ell=0}^{\infty} \sum_{j=0}^{n-i} \sum_{k=0}^{\infty} (-1)^{j+k+\ell} \binom{n-i}{j} \binom{b(i+j)-1}{k} \binom{a(k+1)-1}{\ell} \times g(x)\, G(x)^{\ell}.$$

Then, after a simple algebraic manipulation, we have

$$f_{i:n}(x) = \sum_{\ell=0}^{\infty} q_\ell\, h_{\ell+1}(x), \tag{3.23}$$

where q_ℓ is given by

$$q_\ell = \frac{ab}{B(i, n-i+1)} \sum_{j=0}^{n-i} \sum_{k=0}^{\infty} \frac{(-1)^{j+k+\ell}}{\ell+1} \binom{n-i}{j} \binom{b(i+j)-1}{k} \binom{a(k+1)-1}{\ell},$$

and $h_{\ell+1}(x)$ denotes the exp-G density function with power parameter $\ell+1$ (for $\ell \geq 0$).

Equation (3.23) reveals that the density function of the EG order statistic is a linear combination of exp-G densities. We emphasize that this result is not new and it is given by [10]. However, we now give an alternative formula for the weights that compose this linear combination.

By combining equations (3.11) and (3.23) and after some algebra, we obtain

$$f_{i:n}(x) = \sum_{m=0}^{\infty} \sum_{s=0}^{m+1} d_{m,s}\, \pi(x; s+1, (m+1)\alpha), \tag{3.24}$$

where $d_{m,s} = \sum_{\ell=m}^{\infty} q_\ell \, p_{m,s}^{(\ell+1)}$, the quantity $p_{m,s}^{(\ell+1)}$ is given by (3.12) and $\pi(x; s+1, (m+1)\alpha)$ denotes the Erlang density with shape parameter $s+1$ and scale parameter $(m+1)\alpha$.

Thus, based on (3.24), we obtain an important result that gives the density of $X_{i:n}$ as a double linear combination of Erlang densities. Undoubtedly, there are many applications for equation (3.24), but the most important is to obtain the moments and the mgf of the ith order statistic. The rth moment of $X_{i:n}$ is given by

$$E(X_{i:n}^r) = \sum_{m=0}^{\infty} \sum_{s=0}^{m+1} d_{m,s} \int_0^{\infty} x^r \, \pi(x; s+1, (m+1)\alpha) \, \mathrm{d}x.$$

Based on the results presented in Section 3.4.3, the last equation reduces to

$$E(X_{i:n}^r) = \sum_{m=0}^{\infty} \sum_{s=0}^{m+1} \frac{(r+s)! \, d_{m,s}}{s! \, [(m+1)\alpha]^r}.$$

Next, the mgf of $X_{i:n}$ is given by

$$\varphi(t) = \sum_{m=0}^{\infty} \sum_{s=0}^{m+1} d_{m,s} \int_0^{\infty} \mathrm{e}^{tx} \, \pi(x; s+1, (m+1)\alpha) \, \mathrm{d}x.$$

Based on the results in Section 3.4.5, the last equation can be rewritten as

$$\varphi(t) = \sum_{m=0}^{\infty} \sum_{s=0}^{m+1} \left[1 - \frac{t}{(m+1)\alpha}\right]^{-(s+1)} d_{m,s},$$

for all $t < (m+1)\alpha$.

3.4.8 Reliability

In this section, we derive the reliability, say R, for the EGEE model when $X_1 \sim EGEE(a_1, b_1, \alpha, \beta)$ and $X_2 \sim EGEE(a_2, b_2, \alpha, \beta)$ are two independent random variables with the same baseline parameters α and β. Let $f_1(x)$ denote the pdf of X_1 and $F_2(x)$ denote the cdf of X_2. The reliability can be expressed as $R = \mathrm{P}(X_1 > X_2) = \int_0^{\infty} f_1(x) \, F_2(x) \, \mathrm{d}x$ and using equations (3.8) and (3.9) gives

$$R = \sum_{j,k=0}^{\infty} \mathcal{I}_{j,k} \int_0^{\infty} h_{j+1}(x)\, H_{k+1}(x)\, \mathrm{d}x,$$

where $\mathcal{I}_{j,k} = \sum_{m,n=1}^{\infty} (-1)^{j+k+m+n+2} \binom{b_1}{m} \binom{m\, a_1}{j+1} \binom{b_2}{n} \binom{n a_2}{k+1}$.

Thus, the reliability of X reduces to

$$R = \sum_{j,k=0}^{\infty} \sum_{\ell=0}^{j+k+1} \sum_{r=0}^{\ell} \sum_{i=0}^{r+1} \frac{(-1)^{\ell}\, (j+1)\, \alpha^{r-i+1}\, \beta^{\ell+i-r}\, i!}{(\ell+1)^{i+1}\, (\alpha+\beta)^{\ell+1}} \binom{j+k+1}{\ell} \binom{\ell}{r} \binom{r+1}{i} \mathcal{I}_{j,k}.$$

3.5 Estimation and inference

Several approaches for parameter estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used when constructing confidence intervals for the parameters and also in test statistics. Under conditions of regularity already known, the normal approximation for these estimators in large sample distribution theory is easily handled either analytically or numerically. So, we consider the estimation of the unknown parameters a, b, α and β of the EGEE distribution from complete samples only by maximum likelihood. Let $x_1, \ldots, x_n$ be a random sample of size n from the EGEE distribution. The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (a, b, \alpha, \beta)^{\top}$, say $\ell(\boldsymbol{\theta})$, can be expressed as

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= n \log(ab\alpha^2) - nab \log(\alpha+\beta) - a\alpha \sum_{i=1}^{n} x_i + (a-1) \sum_{i=1}^{n} \log[z(x_i)] \\ &\quad + \sum_{i=1}^{n} \log(1+\beta x_i) + (b-1) \sum_{i=1}^{n} \log[(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}], \qquad (3.25)\end{aligned}$$

where $z(x_i) = \beta + \alpha + \alpha\beta x_i$.

Equation (3.25) can be maximized either directly by using the `Ox` program (subroutine `MaxBFGS`), `R` (`optim` function) and `SAS` (`PROC NLMIXED`), or by solving the nonlinear likelihood equations obtained by differentiating $\ell(\boldsymbol{\theta})$. The elements of the score vector are given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial a} = \frac{n}{a} + (b-1) \sum_{i=1}^{n} \frac{\{(\alpha+\beta)^a\, \log(\alpha+\beta) - z(x_i)^a\, \log[z(x_i)]\, \mathrm{e}^{-a\alpha x_i} + \alpha\, x_i\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}\}}{[(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]}$$

$$
\begin{aligned}
&-n\,b\log(\alpha+\beta)+\sum_{i=1}^{n}\log[z(x_i)]-\alpha\sum_{i=1}^{n}x_i,\\
\frac{\partial\ell(\boldsymbol{\theta})}{\partial b} &= \frac{n}{b}-n\,a\,\log(\alpha+\beta)+\sum_{i=1}^{n}\log[(\alpha+\beta)^a-z(x_i)^a\,\mathrm{e}^{-a\alpha x_i}],\\
\frac{\partial\ell(\boldsymbol{\theta})}{\partial \alpha} &= \frac{2n}{\alpha}+(b-1)\sum_{i=1}^{n}\frac{a\,(\alpha+\beta)^{a-1}-a\,z(x_i)^{a-1}\,(1+\beta x_i)\,\mathrm{e}^{-a\alpha x_i}+a\,x_i\,z(x_i)^a\,\mathrm{e}^{-a\alpha x_i}}{[(\alpha+\beta)^a-z(x_i)^a\,\mathrm{e}^{-a\,\alpha x_i}]}\\
&-\frac{n\,a\,b}{\alpha+\beta}-a\sum_{i=1}^{n}x_i+(a-1)\sum_{i=1}^{n}\frac{(1+\beta x_i)}{z(x_i)},\\
\frac{\partial\ell(\boldsymbol{\theta})}{\partial \beta} &= -\frac{n\,a\,b}{\alpha+\beta}+(b-1)\sum_{i=1}^{n}\frac{a\,(\alpha+\beta)^{a-1}-a\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\alpha x_i}}{[(\alpha+\beta)^a-z(x_i)^a\,\mathrm{e}^{-a\,\alpha x_i}]}\\
&+\sum_{i=1}^{n}\frac{x_i}{(1+\beta x_i)}+(a-1)\sum_{i=1}^{n}\frac{(1+\alpha x_i)}{z(x_i)}.
\end{aligned}
$$

The MLE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained numerically. For interval estimation and hypothesis tests on the parameters a, b, α and β, we determine the 4×4 observed information matrix given by $J(\boldsymbol{\theta}) = \{-U_{rs}\}$, whose elements $U_{rs} = \partial^2\ell(\boldsymbol{\theta})/(\partial r\partial s)$ for $r,s\in\{a,b,\alpha,\beta\}$ can be obtained in the Appendix.

3.6 Applications to real data

Here, we present two applications to real data to illustrate the potentiality of the new distribution. First, in addition to the EGEE model, we consider the three-parameter EGEE$(a,b,\alpha,0)$ and EGEE$(a,b,\alpha,1)$ sub-models. Also, the three-parameter beta-Lindley (BL) distribution, proposed by [57], is compared with the EGEE distribution and its sub-models. All computations are performed using the SAS subroutine NLMixed.

The BL density is given by

$$
\eta(x;a,b,\alpha)=\frac{\alpha^2\,(1+x)\,\mathrm{e}^{-b\alpha x}}{(\alpha+1)^{a+b-1}\,B(a,b)}\,(1+\alpha+\alpha x)^{b-1}\,[\alpha+1-(1+\alpha+\alpha x)\,\mathrm{e}^{-\alpha x}]^{a-1},
$$

where $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function.

First, we consider the number of failures for the air conditioning system of jet airplanes. These data were reported by [58] and [59]: 194, 413, 90, 74, 55, 23, 97, 50, 359, 50, 130, 487, 57, 102, 15, 14, 10, 57, 320, 261, 51, 44, 9, 254, 493, 33, 18, 209, 41, 58, 60, 48, 56, 87, 11, 102, 12, 5, 14, 14, 29, 37, 186, 29, 104, 7, 4, 72, 270, 283, 7, 61, 100, 61, 502, 220, 120, 141, 22, 603, 35, 98, 54, 100, 11, 181, 65, 49, 12, 239, 14, 18, 39, 3, 12, 5, 32, 9, 438, 43, 134, 184, 20, 386, 182, 71, 80, 188, 230, 152, 5, 36, 79, 59, 33, 246, 1, 79, 3, 27, 201, 84, 27, 156, 21, 16, 88, 130, 14, 118, 44, 15, 42, 106, 46, 230, 26, 59, 153, 104, 20, 206, 5, 66, 34, 29, 26, 35, 5, 82, 31, 118, 326, 12, 54, 36, 34, 18, 25, 120, 31, 22, 18, 216, 139, 67, 310, 3, 46, 210, 57, 76, 14, 111, 97, 62, 39, 30, 7, 44, 11, 63, 23, 22, 23, 14, 18, 13, 34, 16, 18, 130, 90, 163, 208, 1, 24, 70, 16, 101, 52, 208, 95, 62, 11, 191, 14, 71. Some descriptive statistics for these data are given below. The smallest and the largest values are 1 and 603, respectively. Further, the mean, median and variance are 92.07, 54.00 and 11645.93, respectively.

Table 3.2 lists the MLEs of the model parameters (with the corresponding standard errors in parentheses) for all fitted models and also the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) statistics. In general, it is considered that the lower values of these criteria indicates the better fit to the data. The figures in Table 3.2 revels that the EGEE model has the lowest AIC and BIC values among all fitted models. Thus, the proposed EGEE distribution is the best model to explain these data.

Table 3.2: MLEs (and their standard errors in parentheses), AIC, BIC and CAIC statistics for the number of successive failures for the air conditioning system.

Distribution	$\widehat{a}$	$\widehat{b}$	$\widehat{\alpha}$	$\widehat{\beta}$	AIC	BIC	CAIC
EGEE	0.0639	0.6470	0.1497	183.90	**2077.4**	**2077.6**	2090.3
	(0.0721)[a]	(0.1254)	(0.1667)	(1248.4)			
EGE	0.1010	0.9100	0.1010	0	2081.5	2081.6	2091.2
	(2.3212)[a]	(0.0866)	(0.9112)	(−)			
EGL	0.05653	0.6951	0.1734	1	2078.0	2078.1	**2087.7**
	(0.0577)[a]	(0.1224)	(0.1758)	(−)			
BL	0.4842	0.5379	0.0234	−	2086.6	2086.7	2096.3
	(0.0577)[a]	(0.3188)	(0.0132)	(−)			

[a] Denotes the standard deviation of the MLEs.

Further, we consider the formal goodness-of-fit tests based on the Cramér-von Mises (W^*) and Anderson-Darling (A^*) test statistics in order to verify which distribution fits better the current data. The W^* and A^* statistics are described in [39]. In general, the lower values of these statistics indicate the better fit to the data. Table 3.3 gives the values of the W^* and A^* statistics for all fitted models. Based on the figures in this table, we conclude that the EGEE distribution provides a better fit to these data than its sub-models and the BL distribution.

Table 3.3: Goodness-of-fit tests for the number of successive failures for the air conditioning system.

Model	Statistics	
	W^*	A^*
EGEE	**0.1137**	**0.7136**
EGE	0.1940	1.2030
EGL	0.1415	0.8886
BL	0.2491	1.5472

Plots of the estimated pdf and cdf of the EGEE distribution and the histogram of the data are displayed in Figure 3.6. These plots clearly reveal that the EGEE model fits the data adequately and then it can be chosen for modeling these data.

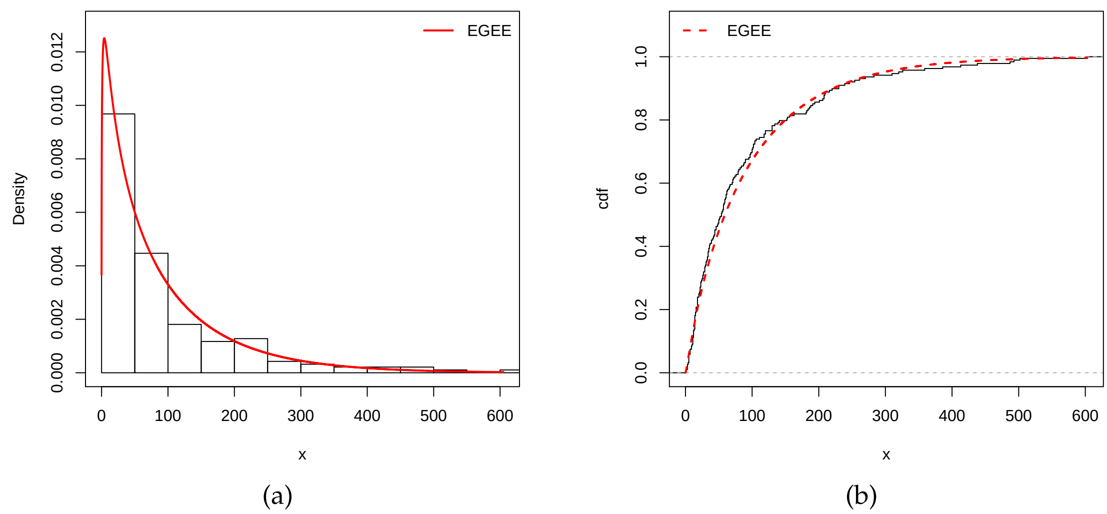


Figure 3.6: (a) Estimated pdf of the EGEE model; (b) Estimated cdf of the EGEE model.

Second, we consider the data presented by [40] on the failure times (in weeks) of 50 components. The data are: 0.013, 0.065, 0.111, 0.111, 0.163, 0.309, 0.426, 0.535, 0.684, 0.747, 0.997, 1.284, 1.304, 1.647, 1.829, 2.336, 2.838, 3.269, 3.977, 3.981, 4.520, 4.789, 4.849, 5.202, 5.291, 5.349, 5.911, 6.018, 6.427, 6.456, 6.572, 7.023, 7.087, 7.291, 7.787, 8.596, 9.388, 10.261, 10.713, 11.658, 13.006, 13.388, 13.842, 17.152, 17.283, 19.418, 23.471, 24.777, 32.795, 48.105. Some descriptive statistics of these data are presented below. The minimum observed value is 0.013, while the maximum value is 48.105. The mean, median and variance are 7.821, 5.320 and 84.76, respectively.

For [40]'s data, we compare the EGEE model with the EE ([11]) and Lindley sub-models and other commonly used models in survival analysis, namely the log-logistic, Fréchet and Birnbaum-Saunders (BS) distributions. The densities of these models are given in the Wolfram alpha website (https://www.wolframalpha.com). Table 3.4 gives the MLEs of the fitted models to the current data with their corresponding standard errors, in addition to the AIC, BIC and CAIC statistics. Table 3.5 lists the values of the A^* and W^* statistics.

The figures in Tables 3.4 and 3.5 suggest at least two important conclusions. The first one is that the proposed model EGEE has the lowest values for the AIC, CAIC, A^* and W^* statistics, and therefore, may be chosen as the best model to analyze the current data. Moreover, these results confirm what has already been demonstrated in the recent statistical literature: generalized models, as the proposed in this paper, usually have superior performance in terms of adjustment when compared to non-generalized models. These conclusions emphasize the importance of the proposed model.

Finally, Figure 3.7 displays the estimated pdf and cdf of the EGEE model and the histogram of the data. These plots reveal that the proposed model is quite suitable for these data.

3.7 Conclusions

Recently, [10] introduced the *exponentiated generalized* (EG) class of continuous distributions with two extra shape parameters. In this chapter, we consider the EG

Table 3.4: MLEs (and their standard errors in parentheses), AIC, BIC and CAIC statistics for the failure times of 50 components.

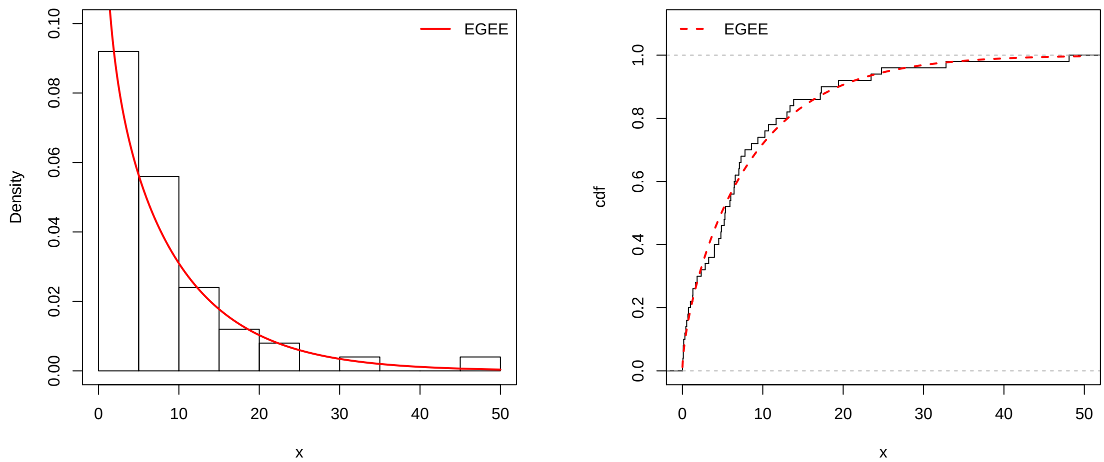
Distribution	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{a}$	$\widehat{b}$	AIC	BIC	CAIC
EGEE	0.3659	0.3103	0.3239	0.6041	**308.3**	316.0	**309.2**
	(0.9972)[a]	(0.7977)	(1.0123)	(0.1946)			
EE	0.1279	2.338E-7	1	1	309.7	**313.5**	309.9
	(0.0352)[a]	(0.0302)	(-)	(-)			
Lindley	0.2317	1	1	1	324.6	326.5	324.6
	(0.0234)[a]	(−)	(−)	(−)			
	$\widehat{\alpha}$	$\widehat{\beta}$					
Log-logistic	4.0938	1.0834			316.0	319.8	316.3
	(0.9218)[a]	(0.1304)					
	$\widehat{\sigma}$	$\widehat{\lambda}$					
Fréchet	1.2802	0.4791			341.3	345.1	341.5
	(0.4028)[a]	(0.0454)					
	$\widehat{\alpha}$	$\widehat{\beta}$					
BS	2.7621	1.2576			327.4	331.2	327.7
	(0.2973)[a]	(0.2721)					

[a] Denotes the standard deviation of the MLE's.

class to generalize the *extended exponential* (EE) distribution. So, we define a new four-parameter lifetime model named the *exponentiated generalized extended exponential* (EGEE) distribution, witch include as special cases the exponential, Lindley and exponentiated exponential distributions, among others. The hazard function of the new model can take the classic bathtub, inverted bathtub, increasing, decreasing and constant shapes. We demonstrate that the EGEE density can be expressed as a double

Table 3.5: Goodness-of-fit tests for the failure times of 50 components.

Model	Statistics	
	W^*	A^*
EGEE	**0.0512**	**0.2670**
EE	0.0658	0.3295
Lindley	0.0657	0.3284
Log-logistic	0.2572	1.3816
Fréchet	0.6097	3.3138
BS	0.2794	1.5364



(a) (b)

Figure 3.7: (a) Estimated pdf of the EGEE model; (b) Estimated cdf of the EGEE model.

linear combination of Erlang densities. Further, we derive several basic mathematical properties of the EGEE model, including explicit expressions for the quantile function, ordinary and incomplete moments, mean deviations, Bonferroni and Lorenz curves, generating function, Rényi entropy, density of the order statistics and reliability. We discuss the estimation of the model parameters by maximum likelihood. We provide the elements of the score vector. We conduct two applications to real data to illustrate the flexibility of the new model.

Appendix A: Observed information matrix.

Here, we provide the elements of the 4×4 observed information matrix $J(\boldsymbol{\theta})$:

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a^2} &= -\frac{n}{a^2} - (b-1) \sum_{i=1}^{n} \frac{\{(\alpha+\beta)^a \log(\alpha+\beta) - z(x_i)^a \log[z(x_i)]\, \mathrm{e}^{-a\alpha x_i}\}^2}{[(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]^2} \\
&- (b-1) \sum_{i=1}^{n} \frac{[\alpha\, x_i\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]^2}{[(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]^2} \\
&+ (b-1) \sum_{i=1}^{n} \frac{(\alpha+\beta)^a \log^2(\alpha+\beta) - z(x_i)^a \log^2[z(x_i)]\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}} \\
&+ (b-1) \sum_{i=1}^{n} \frac{2\,\alpha\, x_i\, z(x_i)^a \log[z(x_i)]\, \mathrm{e}^{-a\alpha x_i} - \alpha^2\, x_i^2\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}},
\end{aligned}
$$

$$
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial b^2} = -\frac{n}{b^2},
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha^2} &= -\frac{2n}{\alpha^2} + \frac{nab}{(\alpha+\beta)^2} - (a-1) \sum_{i=1}^{n} \frac{(1+\beta x_i)^2}{z(x_i)^2} \\
&- (b-1) \sum_{i=1}^{n} \frac{[a\,(\alpha+\beta)^{a-1} - a\, z(x_i)^{a-1}\,(1+\beta x_i)\, \mathrm{e}^{-a\alpha x_i} + a\, x_i\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]^2}{[(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]^2} \\
&+ (b-1) \sum_{i=1}^{n} \frac{a\,(a-1)\,(\alpha+\beta)^{a-2} - a\,(a-1)\, z(x_i)^{a-2}\,(1+\beta x_i)^2\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}} \\
&+ (b-1) \sum_{i=1}^{n} \frac{2\, a^2\, x_i\, z(x_i)^{a-1}\,(1+\beta x_i)\, \mathrm{e}^{-a\alpha x_i} - a^2\, x_i^2\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta^2} &= \frac{nab}{(\alpha+\beta)^2} - \sum_{i=1}^{n} \frac{x_i^2}{(1+\beta x_i)^2} - (a-1) \sum_{i=1}^{n} \frac{(1+\alpha x_i)^2}{z(x_i)^2} \\
&- (b-1) \sum_{i=1}^{n} \frac{[a\,(\alpha+\beta)^{a-1} - a\, z(x_i)^{a-1}\,(1+\alpha x_i)\, \mathrm{e}^{-a\alpha x_i}]^2}{[(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}]^2} \\
&+ (b-1) \sum_{i=1}^{n} \frac{a\,(a-1)\,(\alpha+\beta)^{a-2} - a\,(a-1)\, z(x_i)^{a-2}\,(1+\alpha x_i)^2\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a b} &= -n \log(\alpha+\beta) + \sum_{i=1}^{n} \frac{(\alpha+\beta)^a \log(\alpha+\beta) - z(x_i)^a \log[z(x_i)]\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}} \\
&+ \sum_{i=1}^{n} \frac{\alpha\, x_i\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a \alpha} &= -\frac{n\, b}{\alpha+\beta} - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{(1+\beta\, x_i)}{z(x_i)} \\
&+ (b-1) \sum_{i=1}^{n} \frac{(\alpha+\beta)^{a-1} + a\,(\alpha+\beta)^{a-1} \log(\alpha+\beta) - z(x_i)^{a-1}\,(1+\beta x_i)\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}} \\
&+ (b-1) \sum_{i=1}^{n} \frac{a\, z(x_i)^{a-1}\,(1+\beta x_i) \log[z(x_i)]\, \mathrm{e}^{-a\alpha x_i} + a\,\alpha\, x_i\, z(x_i)^{a-1}\,(1+\beta x_i)\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}} \\
&+ (b-1) \sum_{i=1}^{n} \frac{x_i\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i} + a\, x_i\, z(x_i)^a \log[z(x_i)]\, \mathrm{e}^{-a\alpha x_i} - a\,\alpha\, x_i^2\, z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^a - z(x_i)^a\, \mathrm{e}^{-a\alpha x_i}}
\end{aligned}
$$

$$
\begin{aligned}
&- (b-1)\sum_{i=1}^{n}\left\{\frac{[a\,(\alpha+\beta)^{a-1} - a\,z(x_i)^{a-1}\,(1+\beta x_i)\,\mathrm{e}^{-a\,\alpha\,x_i} + a\,x_i\,z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}{[(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}\right.\\
&\qquad \left.\times\,\frac{[(\alpha+\beta)^{a}\,\log(\alpha+\beta) - z(x_i)^{a}\,\log[z(x_i)]\,\mathrm{e}^{-a\alpha x_i} + \alpha\,x_i\,z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}{[(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial a\beta} &= -\frac{nb}{\alpha+\beta} + \sum_{i=1}^{n}\frac{1+\alpha x_i}{z(x_i)}\\
&+ (b-1)\sum_{i=1}^{n}\frac{(\alpha+\beta)^{a-1} + a\,(\alpha+\beta)^{a-1}\,\log(\alpha+\beta) - z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}}\\
&+ (b-1)\sum_{i=1}^{n}\frac{a\,\alpha\,x_i\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\alpha x_i} - a\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\log[z(x_i)]\,\mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}}\\
&- (b-1)\sum_{i=1}^{n}\left\{\frac{[a\,(\alpha+\beta)^{a-1} + a\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\,\alpha\,x_i}]}{[(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}\right.\\
&\qquad \left.\times\,\frac{[(\alpha+\beta)^{a}\,\log(\alpha+\beta) - z(x_i)^{a}\,\log[z(x_i)]\,\mathrm{e}^{-a\alpha x_i} + \alpha\,x_i\,z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}{[(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial b\alpha} &= -\frac{na}{\alpha+\beta} + \sum_{i=1}^{n}\frac{a\,(\alpha+\beta)^{a-1} - a\,z(x_i)^{a-1}\,(1+\beta x_i)\,\mathrm{e}^{-a\alpha x_i} + a\,x_i\,z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial b\beta} &= -\frac{na}{\alpha+\beta} + \sum_{i=1}^{n}\frac{a\,(\alpha+\beta)^{a-1} - a\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial \alpha\beta} &= \frac{nab}{(\alpha+\beta)^2} + (a-1)\sum_{i=1}^{n}\left[\frac{x_i}{z(x_i)} - \frac{(1+\alpha x_i)\,(1+\beta x_i)}{z(x_i)^2}\right]\\
&+ (b-1)\sum_{i=1}^{n}\frac{a\,(a-1)\,(\alpha+\beta)^{a-2} - a\,(a-1)\,z(x_i)^{a-2}\,(1+\alpha x_i)\,(1+\beta x_i)\,\mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}}\\
&+ (b-1)\sum_{i=1}^{n}\frac{a^2\,x_i\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\alpha x_i} - a\,x_i\,z(x_i)^{a-1}\,\mathrm{e}^{-a\alpha x_i}}{(\alpha+\beta)^{a} - z(x_i)^{a}\mathrm{e}^{-a\alpha x_i}}\\
&- (b-1)\sum_{i=1}^{n}\left\{\frac{[a\,(\alpha+\beta)^{a-1} - a\,z(x_i)^{a-1}\,(1+\alpha x_i)\,\mathrm{e}^{-a\,\alpha\,x_i}]}{[(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}\right.\\
&\qquad \left.\times\,\frac{[a\,(\alpha+\beta)^{a-1} - a\,z(x_i)^{a-1}\,(1+\beta x_i)\,\mathrm{e}^{-a\alpha x_i} + a\,x_i\,z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}{[(\alpha+\beta)^{a} - z(x_i)^{a}\,\mathrm{e}^{-a\alpha x_i}]}\right\},
\end{aligned}
$$

where $z(x_i) = \beta + \alpha + \alpha\beta x_i$.

CHAPTER 4

The Exponentiated Generalized Standard Half-Logistic Distribution

Paper accepted for publication in the *International Journal of Statistics and Probability*.

Abstract

We introduce a new two-parameter lifetime model, called the exponentiated generalized standard half-logistic distribution, and study some of its general structural properties. This distribution extends the half-logistic distribution proposed by Balakrishnan in the eighties. We provide explicit expressions for the density function, ordinary and incomplete moments, generating and quantile functions, mean deviations, Bonferroni and Lorenz curves, and order statistics. Our formulas are manageable using modern computer resources with analytic and numerical capabilities and they may turn into adequate tools for applied statisticians. For most of the functions associated with the proposed model, we provide numerical and graphical studies to illustrate their practical use. The model parameters are estimated by maximum likelihood and the observed information matrix is derived. An extensive Monte Carlo simulation study reveals that these estimators have good properties such as low biases and variances, even in small or moderate sample sizes. We also show that the proposed model can be superior to some other lifetime models by means of a real data set.

Keywords: Exponentiated generalized distribution. Half-logistic distribution. Hazard rate function. Lifetime distribution. Moments.

Resumo

Nós introduzimos um novo modelo com dois parâmetros que pode ser usado para ajustar dados de sobrevivência, chamado de distribuição semi-logística padronizada exponencializada generalizada. Esta distribuição estende a distribuição semi-logística proposta por Balakrishnan nos anos oitenta. Nós fornecemos expressões explícitas para a função de densidade, momentos ordinários e incompletos, funções geratriz de momentos e quantílica, desvios médios, curvas de Bonferroni e Lorenz e estatísticas de ordem. Nossas fórmulas são manejáveis usando recursos computacionais modernos com capacidades analíticas e numéricas e podem se transformar em ferramentas adequadas para estatísticos aplicados. Para a maioria das funções associadas ao modelo proposto, fornecemos estudos numéricos e gráficos para ilustrar seu uso prático. Os parâmetros do modelo são estimados por máxima verossimilhança e a matriz de informação observada é derivada. Um extenso estudo de simulação de Monte Carlo revela que estes estimadores têm boas propriedades, tais como baixos viéses e variâncias, mesmo em amostras pequenas ou moderadas. Mostramos também que o modelo proposto pode ser superior a alguns outros modelos de vida por meio de um conjunto de dados reais.

Palavras-chave: Distribuição exponencializada generalizada. Distribuição semi-logística padronizada. Função de risco. Modelo de sobrevivência. Momentos.

4.1 Introduction

The statistical literature points out that Balakrishnan (1985) pioneered the half-logistic HL distribution as a lifetime model, which is the distribution of the absolute value of a random variable following the logistic distribution. It has a monotonically increasing hazard rate function (hrf) for all parameter values, which is a property shared by relatively few distributions with support on the positive real line. Recently, the HL distribution has been discussed by several authors. We shall refer to the following works: [60] obtained approximate maximum likelihood estimates (MLEs) for the location and scale parameters with type-II right-censoring; [61] presented the estimation for the scaled HL distribution under type II censoring; [62] investigated bootstrap confidence intervals for the process capability index under the HL distribution. More recently, [63] introduced a new extension for the HL model by considering a standardized version, say the standardized half-logistic (SHL) distribution, described below.

The fact that it does not have parameters makes the SHL distribution an attractive model for statisticians and applied researchers. Its mathematical simplicity allows its properties to be obtained in a very simple way. In particular, this distribution allows the addition of parameters without difficulties. The cumulative distribution function (cdf) and probability density function (pdf) (for $t > 0$) of the SHL distribution has simple closed-forms. They are given by

$$G(t) = \frac{1 - \mathrm{e}^{-t}}{1 + \mathrm{e}^{-t}} \tag{4.1}$$

and

$$g(t) = \frac{2\mathrm{e}^{-t}}{(1 + \mathrm{e}^{-t})^2}, \tag{4.2}$$

respectively.

Let T be a random variable having density (4.2). The HL distribution is defined by a linear transformation $W = \mu + \sigma T$, with $\mu \in \Re^+$ and $\sigma > 0$. Without loss of generality, we can work with the SHL model. The nth moment of T can be obtained by direct integration

$$E(T^n) = 2\int_0^\infty \frac{t^n\,\mathrm{e}^{-\mathrm{t}}}{(1+\mathrm{e}^{-\mathrm{t}})^2}dt = 2n!(1-2^{1-n})\zeta(n),$$

where $\zeta(\cdot)$ is the Riemann zeta function. For details on the Riemann zeta function, see the Wolfram website http: //mathworld.wolfram.com/RiemannZetaFunction.html. In particular, the first two moments of T are $E(T) = \log(4)$ and $E(T^2) = \pi^2/3$. In addition, the hazard rate function (hrf) of T is given by $\lambda(t) = 1/(1+\mathrm{e}^{-t})$. The moment generating function (mgf) of T, say $M_T(s) = E(\mathrm{e}^{-sT})$, is

$$M_T(s) = 2\int_0^\infty \mathrm{e}^{-st}\frac{\mathrm{e}^{-t}}{(1+\mathrm{e}^{t})^2}dt = 2J_1(1+s,1-s),$$

where $J_p(a,b) = \int_0^p \frac{u^{a-1}}{(1+u)^{a+b}}$ $(a,\, b > 0)$ is the type II incomplete beta function. For more properties of the HL distribution we recommend to the readers the following papers: [60], [61], [62], [63] and [64].

We believe that the addition of parameters to the SHL model may generate new distributions with great adjustment capability and, for this reason, we propose a generalization of it. The recent literature has suggested several ways of extending well-known distributions, among them, the generator approach. The generators allow us to extend well-known distributions and at the same time develop more realistic statistical models in a great variety of applications. Some of the most important generators were recently discussed by [65].

For a baseline continuous cdf $G(x)$, [10] defined the *exponentiated generalized* (EG for short) class of distributions by

$$F(x) = \{1 - [1 - G(x)]^a\}^b, \tag{4.3}$$

where $a > 0$ and $b > 0$ are two extra parameters whose role is to govern skewness and generate distributions with heavier/ligther tails. They are sought as a manner to furnish a more flexible distribution. Because of its tractable distribution function (4.3), this class can be used quite effectively even if the data are censored. The EG class is suitable for modeling continuous univariate data that can be in any interval of the real line. The pdf corresponding to (4.3) is given by

$$f(x) = a\,b\,[1 - G(x)]^{a-1}\,\{1 - [1 - G(x)]^{a}\}^{b-1}\,g(x), \tag{4.4}$$

where $g(x) = dG(x)/dx$ is the baseline pdf. The two parameters in (4.4) can control both tail weights and possibly adding entropy to the center of the EG density. The baseline pdf $g(x)$ is a special case of (4.4) when $a = b = 1$. Setting $a = 1$ it gives the exponentiated-G ("exp-G") distribution. If $b = 1$, we obtain the Lehmann type II distribution. So, the distribution (4.4) generalizes both Lehmann types I and II distributions; that is, this method can be interpreted as a double construction of Lehmann alternatives. Note that even if $g(x)$ is a symmetric density, the density $f(x)$ will not be symmetric.

The above properties and many others have been discussed and explored in recent works for the EG class. We refer to the papers: [37], [53], [66], [36] [67], [55], [68], [56], [69] and [54], which used the EG class to extend the Burr III, Birnbaum-Saunders, inverse Weibull, inverted exponential, generalized gamma, Gumbel, extended exponential, Fréchet, modified Weibull and Dagum distributions, respectively.

In this Chapter, we define the *exponentiated generalized standard half-logistic* (EGSHL) distribution by inserting (4.1) in equation (4.3). As we shall see later, our model has only two parameters, does not involve any complicated functions and is quite flexible. Its hrf can take non-monotonous forms, such as bathtub and inverted bathtub, which explain many real phenomenons. In addition, we prove that the density of the proposed model can be expressed as a linear combination of exp-SHL densities ([63]). Thus, many properties of the new model can be obtained using this representation. Additionally, for each important equation associated with the new model, we provide plots and numerical studies in order to illustrate its usefulness. We hope that the EGSHL distribution can be useful for modelling real data and, for this reason, we present a practical study that illustrates the power of adjustment of the proposed model.

The rest of the Chapter is organized as follows. In Section 4.2, we present the EGSHL distribution. In Section 4.3, we provide a small study for the shapes of the EGSHL pdf and hrf. A detailed study of the quantile function (qf) and its applications is presented in Section 4.4. In Section 4.5, we obtain a useful linear representation for the density function as a linear combination of exp-SHL densities. So, many properties of the exp-SHL model are presented in Section 4.6. Explicit expressions for the ordinary and incomplete moments, cumulants, mean deviations, Bonferroni and Lorenz curves and generating function of the EGSHL distribution are determined in Section 4.7. Sections 4.8 and 4.9 are related to the probability weighted moments (PWMs) and Rényi entropy, respectively. The order statistics and their moments are investigated in Section 4.10. We discuss maximum likelihood estimation of the model parameters in Section 4.11. In Section 4.12, we present a simulation study. An application to real data in Section 4.13 shows the usefulness of the proposed distribution. Finally, concluding remarks are addressed in Section 4.14.

4.2 The new distribution

Let X be a random variable with support on the positive real line having the EGSHL(a, b) distribution, say $X \sim$ EGSHL(a, b). The cdf of X is defined by inserting (4.1) in equation (4.3)

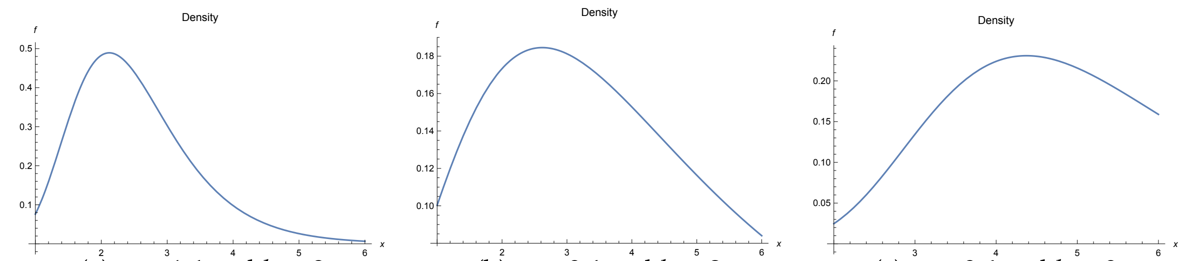
$$F(x) = F(x; a, b) = \frac{[(1+\mathrm{e}^{-x})^{a} - 2^{a}\,\mathrm{e}^{-ax}]^{b}}{(1+\mathrm{e}^{-x})^{ab}}, \tag{4.5}$$

where $a > 0$ and $b > 0$. Note that (4.5) has a simple closed-form, which is an important aspect to generate EGSHL variables in a very simple manner by using the method of inversion. The density of X, for $x > 0$, can be reduced to

$$f(x) = f(x; a, b) = \frac{a\,b\,2^{a}\,\mathrm{e}^{-ax}\,[(1+\mathrm{e}^{-x})^{a} - 2^{a}\,\mathrm{e}^{-ax}]^{b-1}}{(1+\mathrm{e}^{-x})^{ab+1}}. \tag{4.6}$$

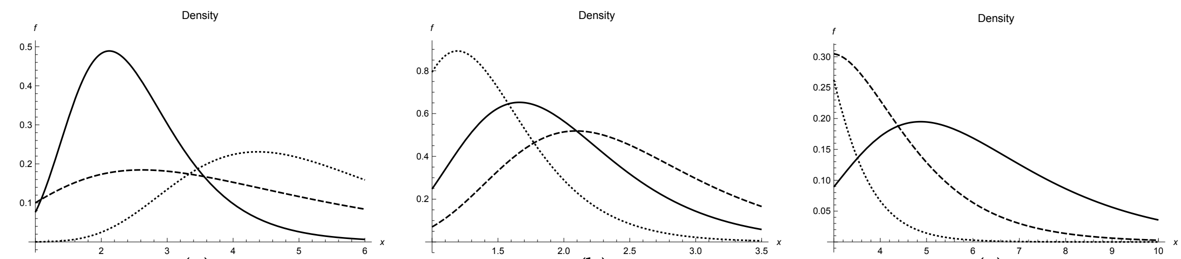
For brevity of notation, we shall drop the explicit reference to the parameters a and b unless otherwise stated.

For $a = b = 1$, equation (4.6) reduces to the SHL density. The EGSHL model also includes the Lehmann type I and type II transformations of the SHL distribution, from now on denoted by ESHLI and ESHLII. For example, the exponentiated SHL distribution, say ESHLI, follows when $a = 1$. Some plots of the pdf (4.6) are displayed in Figure 4.1. Theses plots reveal that the EGSHL pdf is quite flexible and can take symmetric and asymmetric forms, among others. In summary, these plots reinforce the importance of the proposed model.



(a) $a = 1.4$ and $b = 8$ (b) $a = 0.4$ and $b = 2$ (c) $a = 0.6$ and $b = 9$

Figure 4.1: Plots of the EGSHL density function for some parameter values.



(a) (b) (c)

Figure 4.2: Plots of the EGSHL density function for some parameter values.

Besides the cdf (4.5) and pdf (4.6), other functions can be used to characterize the EGSHL model such as the survival function (sf) and hrf. These are particularly important to analyze survival data that involve the time associated to an event of interest such as the time that a certain component fails, the death of a patient or a disease relapse. Here, it is worth quoting [70] Chapter 2, page 8:

> *The distribution of survival times is usually described or characterized by three functions:* (1) *the survivorship function,* (2) *the probability density function, and* (3) *the hazard function. These three functions are mathematically equivalent - if one of them is given, the other two can be derived.*

The sf and hrf of X are given by

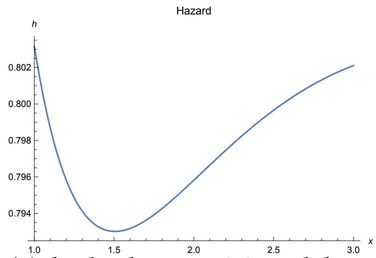
$$S(x) = \frac{(1+\mathrm{e}^{-x})^{ab} - [(1+\mathrm{e}^{-x})^{a} - 2^{a}\,\mathrm{e}^{-ax}]^{b}}{(1+\mathrm{e}^{-x})^{ab}} \tag{4.7}$$

and

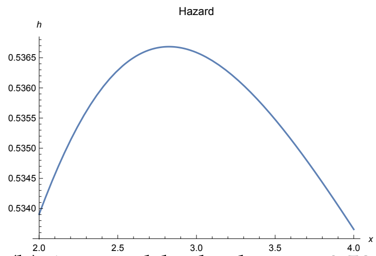
$$\tau(x)=\frac{a\,b\,2^{a}\,\mathrm{e}^{-ax}\,[(1+\mathrm{e}^{-x})^{a}-2^{a}\,\mathrm{e}^{-ax}]^{b-1}}{(1+\mathrm{e}^{-x})\{(1+\mathrm{e}^{-x})^{ab}-[(1+\mathrm{e}^{-x})^{a}-2^{a}\,\mathrm{e}^{-ax}]^{b}\}},\tag{4.8}$$

respectively.

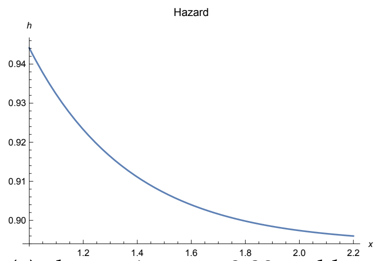
Some plots of the hrf (4.8) are displayed in Figure 4.3. Besides monotone forms, the hrf of X can take bathtub and inverted bathtub shapes. This non-monotone form is particularly important because of its great practical applicability. The time of human life is just one of many phenomena that the bathtub shape hrf is applicable [70].



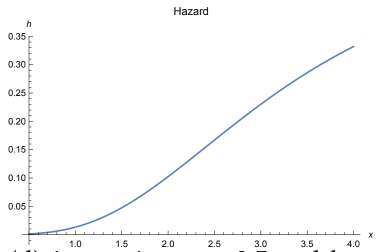
(a) bathtub $a = 0.8$ and $b = 0.39$



(b) inverted bathtub $a = 0.52$ and $b = 0.56$



(c) decreasing $a = 0.89$ and $b = 0.2$



(d) increasing $a = 0.5$ and $b = 4.4$

Figure 4.3: Plots of the EGSHL hazard function for some parameter values.

As a further characterization of the EGSHL distribution, we provide the cumulative hazard rate (chrf) $H(x)$ and reversed hazard rate (rhrf) $r(x)$ functions:

$$H(x)=-\log\left[\frac{(1+\mathrm{e}^{-x})^{ab}-[(1+\mathrm{e}^{-x})^{a}-2^{a}\,\mathrm{e}^{-ax}]^{b}}{(1+\mathrm{e}^{-x})^{ab}}\right]$$

and

$$r(x)=\frac{a\,b\,2^{a}\,\mathrm{e}^{-ax}}{(1+\mathrm{e}^{-x})[(1+\mathrm{e}^{-x})^{a}-2^{a}\,\mathrm{e}^{-ax}]},$$

respectively. These last two functions are less used in practical situations. However, they have outstanding theoretical importance. For example, we can expressed $f(x) = r(x)\exp\{-H(x)\}$. For more details, see [70].

4.3 Shapes

For a detailed mathematical approach for the EGSHL model, we investigate the shapes of its pdf and hrf using their first and second derivatives. The first derivative of $\log\{f(x)\}$ is given by

$$\frac{\mathrm{d}\log\{f(x)\}}{\mathrm{d}x} = -\,a + (1+ab)\,\mathrm{e}^{-x}\,\eta^{-1}(x) - a\,(1-b)\,\mathrm{e}^{-x}\,v_2(x)\,v_1^{-1}(x),$$

where $\eta(x) = 1 + \mathrm{e}^{-x}$, $v_1 = -\,2^a\,\mathrm{e}^{-a\,x} + \eta^a(x)$ and $v_2 = 2^a\,\mathrm{e}^{x(1-a)} - \eta^{a-1}(x)$.

Thus, the critical values of $f(x)$ are the roots of the equation:

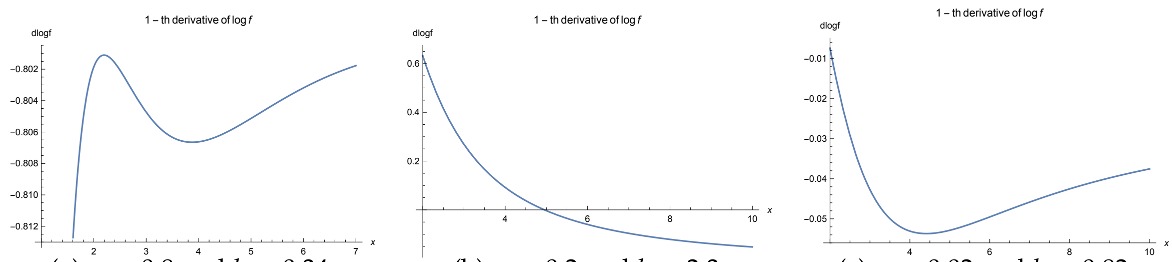
$$-\,a + (1+ab)\,\mathrm{e}^{-x}\,\eta^{-1}(x) = a\,(1-b)\,\mathrm{e}^{-x}\,v_2(x)\,v_1^{-1}(x).$$

The value x_0, which solves the equation above can be a maximum, minimum or inflection point. To check this, we evaluate the sign of the second derivative of $\log\{f(x)\}$ at $x = x_0$. We have

$$\begin{aligned}\frac{\mathrm{d}^2\log\{f(x)\}}{\mathrm{d}x^2} \quad = \quad & \frac{(1+ab)\,\mathrm{e}^{-x}\,[\mathrm{e}^{-x} - \eta(x)]}{\eta^2(x)} \\ & + a\,(1-b)\,\mathrm{e}^{-2x}\,v_1^{-2}(x)\left\{a\,2^{2a}\,\mathrm{e}^{2x(1-a)} + \eta^{-2}(x)\big[1 - a + a\,\eta^{2a}(x) - \mathrm{e}^{x}\,\eta(x)\big]\right. \\ & \left. -\,2^a\,\mathrm{e}^{-a\,x}\,\eta^{a-2}(x)\big[1 - a + \mathrm{e}^{x}\,\eta(x)(1 + a - a\,\mathrm{e}^{x}\,\eta(x))\big]\right\}.\end{aligned}$$

It is often difficult to obtain an analytical solution for the critical value of this function. Therefore, it is common to obtain numerical solutions with high accuracy through optimization routines in most mathematical and statistical platforms. Some plots of the first derivative of $\log\{f(x)\}$ for selected parameter values are displayed in Figure 4.4. These plots are constructed using the *Wolfram Mathematica* software.

Similarly, we provide the first and second derivatives of $\log\{h(x)\}$ for the EGSHL model. The critical values of $\log\{h(x)\}$ are the roots of the equation:



(a) $a = 0.8$ and $b = 0.34$ (b) $a = 0.2$ and $b = 2.3$ (c) $a = 0.02$ and $b = 0.82$

Figure 4.4: Plots of the first derivative of $\log\{f(x)\}$.

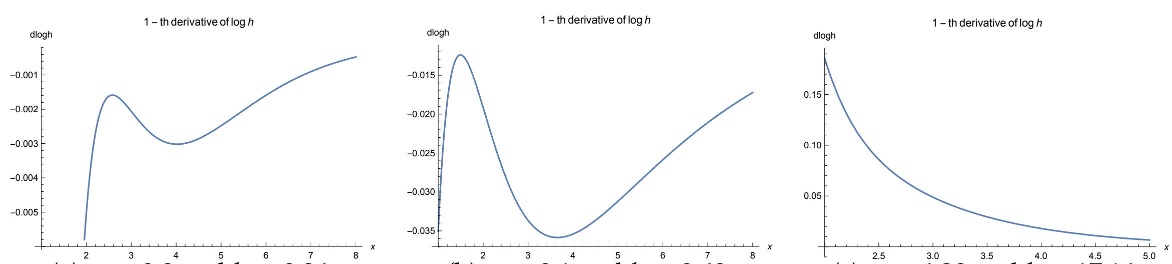
$$
\begin{aligned}
\frac{\mathrm{d}\log\{h(x)\}}{\mathrm{d}x} &= -\,a\,(1-b)\,\mathrm{e}^{-x}\,v_2(x) + \frac{\mathrm{e}^{-x} + (4-a)\,\eta(x)}{\eta(x)} \\
&\quad + \frac{a\,b\,\mathrm{e}^{-x}\,\eta^{a\,b-1}(x) + a\,b\,\mathrm{e}^{-x}\,v_2(x)\,v_1^{b-1}(x)}{\eta^{a\,b}(x) - v_1^{b}(x)}.
\end{aligned}
$$

The second derivative of $\log\{h(x)\}$ is given by

$$
\begin{aligned}
\frac{\mathrm{d}^2\log\{h(x)\}}{\mathrm{d}x^2} &= \frac{\mathrm{e}^{-2\,x} - \mathrm{e}^{-\,x}\,\eta(x)}{\eta^2(x)} + a\,\mathrm{e}^{-\,x}\,v_1^{-2}(x)\,\Big\{(1-b)\,\big[a\,\mathrm{e}^{-\,x}\,v_2^2(x) + v_1(x)\,v_3(x)\big] \\
&\quad -\,b\,\big[\eta^{a\,b-1}(x) + v_2(x)\,v_1^{b-1}(x)\big]^2 - b\,\big[\eta^{a\,b}(x) - v_1^{b}(x)\big] \\
&\quad \times\,\big[\eta^{a\,b-1}(x) - (1-a\,b)\,\mathrm{e}^{-\,x}\,\eta^{a\,b-2}(x) + a\,(1-b)\,\mathrm{e}^{-\,x}\,v_2^2(x)\,v_1^{b-2}(x) \\
&\quad + v_1^{b-1}(x)\,v_3(x)\big]\Big\},
\end{aligned}
$$

where $v_3(x) = 2^a\,a\,\mathrm{e}^{x(1-a)} + (1-a)\,\mathrm{e}^{-x}\,\eta^{a-2}(x) - \eta^{a-1}(x)$.

Some plots of the first derivative of $\log\{h(x)\}$ for the EGSHL distribution for selected parameter values are displayed in Figure 4.5.



(a) $a = 0.8$ and $b = 0.34$ (b) $a = 0.1$ and $b = 0.68$ (c) $a = 4.28$ and $b = 17.14$

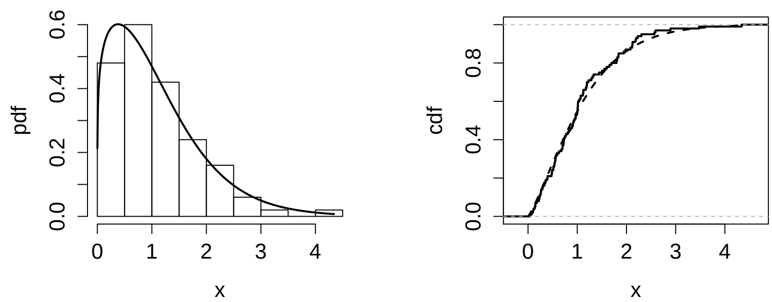
Figure 4.5: Plots of the first derivative of $\log\{h(x)\}$.

4.4 Quantile function

In previous sections, we provide several important functions that characterize the random variable $X \sim \text{EGSHL}(a,\, b)$. Here, we present the qf for the EGSHL model. By inverting (4.5), we obtain

$$Q(u) = -\log\left[\frac{(1-u^{1/b})^{1/a}}{2-(1-u^{1/b})^{1/a}}\right], \tag{4.9}$$

where $u \in (0,1)$. The EGSHL distribution is easily simulated from a uniform random variable U by $X = Q(U)$. Next, we use (4.9) to generate 100 EGSHL $(1.5,\, 1.2)$ occurrences. Figure 4.6 displays the histogram and empirical cdf for the simulated data and also the exact pdf and cdf of the EGSHL model. As we can see, the setting is quite adequate and reinforces that the model has good potential for simulation studies. For similar studies, we refer [29] and [30], among others.



(a) Histogram and exact pdf (b) Empirical and exact cdfs

Figure 4.6: Plots of the EGSHL$(1.5,\, 1.2)$ pdf, histogram, exact and empirical cdfs for simulated data with $n = 100$.

As mentioned earlier, the qf practical uses are numerous. For example, $Q(1/2)$ determines the median of the model. Table 4.1 shows a small simulation study using the `R` software. The goal is to compare the empirical medians (EMed), calculated from random samples of size n = 10, 20, 40, 100, generated for different parameter values, with their corresponding theoretical medians (Med) obtained by $Q(1/2)$. As expected, the difference between EMed and Med decreases when n increases.

Finally, we use the qf of X to determine the Bowley skewness [27] (B) and Moors

Table 4.1: Theoretical and empirical medians (for $n = 10, 20, 40, 100$) of X for some parameter values.

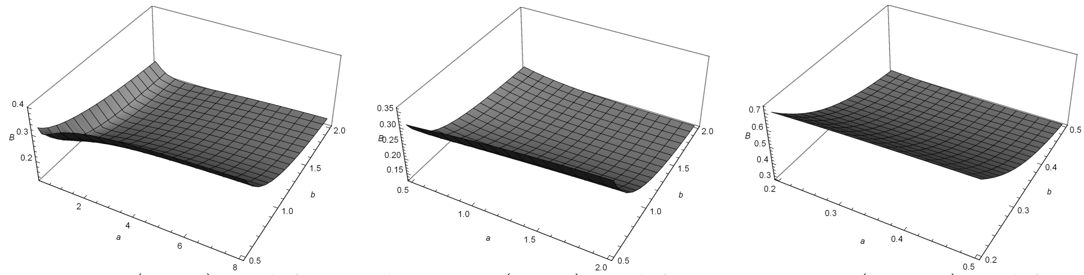
a	b	Med	EMed ($n = 10$)	EMed ($n = 20$)	EMed ($n = 40$)	EMed ($n = 100$)
1.5	3.3	1.6220	1.0295	1.2260	1.5061	1.6586
1.5	1.5	1.0579	0.5052	0.6796	0.9446	1.0940
3.3	1.5	0.5325	0.2436	0.3328	0.4720	0.5518

kurtosis [28] (M). The shortcomings of the classical kurtosis measure are well-known. The effects of the additional shape parameters a and b on the skewness and kurtosis of the EGSHL model can be based on quantile measures. The Bowley skewness is based on quartiles $\mathrm{B} = [Q(3/4) - 2Q(1/2) + Q(1/4)]/[Q(3/4) - Q(1/4)]$, whereas the Moors kurtosis is based on octiles $\mathrm{M} = [Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)]/[Q(6/8) - Q(2/8)]$. These two measures are less sensitive to outliers and they exist even for distributions without moments. Since M is based on octiles, it is not sensitive to variations of the values in the tails or to variations of the values around the median. These measures are fairly considered in the literature. Here, we refer to the following works: [53], [54], [55] and [56], among others.

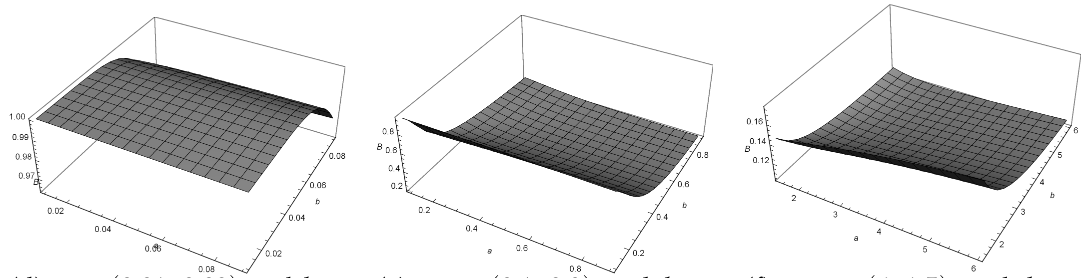
In Figures 4.7 and 4.8, we present 3D plots of B and M measures for several parameters values. These plots are obtained using the *Wolfram Mathematica* software. Based on these plots, it is possible to conclude that changes in the additional parameters a and b have a considerable impact on the skewness and kurtosis of the EGSHL model, thus showing its greater flexibility. So, theses plots reinforce the importance of the additional parameters.

4.5 Linear representation

We provide useful linear representations for equations (4.3) and (4.4) based on the exponentiated class of distributions. Mathematical properties of the exponentiated distributions have been published by many authors in the 90s and more recently. See, for example, [18] for exponentiated exponential, [35] for exponentiated Lindley, [71] for exponentiated linear failure rate and, more recently, [33] for the exponentiated

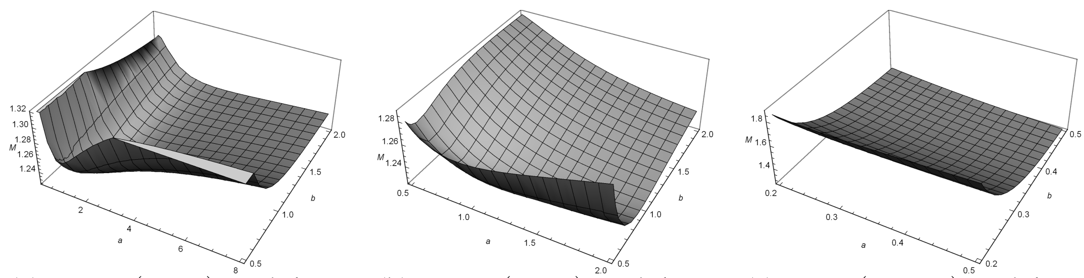


(a) $a \in (0.2, 8)$ and $b \in (0.5, 2)$ (b) $a \in (0.5, 2)$ and $b \in (0.5, 2)$ (c) $a \in (0.2, 0.5)$ and $b \in (0.2, 0.5)$

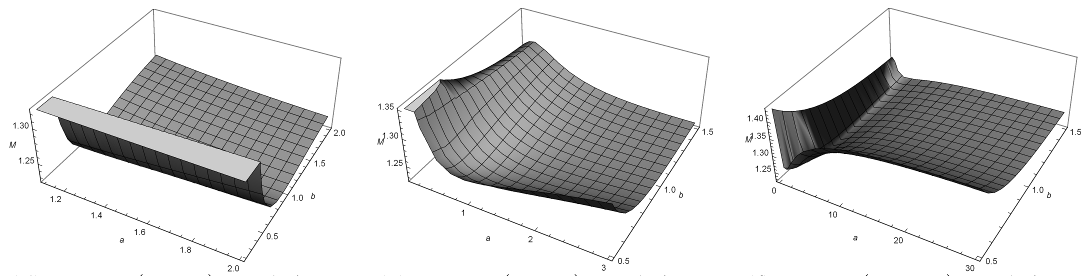


(d) $a \in (0.01, 0.09)$ and $b \in (0.01, 0.09)$ (e) $a \in (0.1, 0.9)$ and $b \in (0.1, 0.9)$ (f) $a \in (6, 1.5)$ and $b \in (1.5, 6)$

Figure 4.7: Plots of the Bowley skewness for the EGSHL distribution.



(a) $a \in (0.2, 8)$ and $b \in (0.5, 2)$ (b) $a \in (0.5, 2)$ and $b \in (0.5, 2)$ (c) $a \in (0.2, 0.5)$ and $b \in (0.2, 0.5)$



(d) $a \in (1.1, 2)$ and $b \in (0.2, 2)$ (e) $a \in (0.1, 3)$ and $b \in (0.5, 1.5)$ (f) $a \in (0.1, 30)$ and $b \in (0.5, 1.5)$

Figure 4.8: Plots of the Moors kurtosis for the EGSHL distribution.

Nadarajah-Haghighi and [63] for the SHL distribution distribution.

For an arbitrary baseline cdf $G(x)$, a random variable Y_c has the exponentiated-G (exp-G) class with power parameter $c > 0$, say Y_c $\sim$exp-G(c), if its cdf and pdf are given by $H_c(x) = G(x)^c$ and $h_c(x) = c\, g(x)\, G(x)^{c-1}$, respectively. For a comprehensive discussion about the exponentiated class, see a recent paper by [65].

Here, we consider the generalized binomial expansion

$$(1-z)^b = \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} z^k, \tag{4.10}$$

which holds for any real non-integer b and $|z| < 1$.

Using (4.10) twice in equation (4.4), the EG density function can be expressed as

$$f(x) = \sum_{j=0}^{\infty} w_{j+1}\, h_{j+1}(x), \tag{4.11}$$

where $w_{j+1} = \sum_{m=1}^{\infty} (-1)^{j+m+1} \binom{b}{m} \binom{m\,a}{j+1}$ and $h_{j+1}(x) = (j+1)\, g(x)\, G(x)^j$ is the exp-G pdf with power parameter $j+1$. Equation (4.11) reveals that the EG density is a linear combination of exp-G densities. It is the main result of this section, which aims to derive some structural properties of the EG class from those exp-G properties. The cdf $F(x)$ comes from (4.11) by simple integration, namely

$$F(x) = \sum_{j=0}^{\infty} w_{j+1}\, H_{j+1}(x), \tag{4.12}$$

where $H_{j+1}(x) = G(x)^{j+1}$ is the exp-G cdf with power parameter $j+1$.

Here, it is worth mentioning that the results presented in equations (4.11) and (4.12) are general and, therefore, are valid for any baseline distribution $G(x)$. These expressions agree with those provided by [53]. It is not difficult to show numerically that $\sum_{j=0}^{\infty} w_{j+1} = 1$. Moreover, for most practical purposes, we can set the upper limits equal to 50.

We can adopt (4.11) for the EGSHL distribution and obtain its mathematical properties from those properties of the exp-SHL distribution. Let Y_{j+1} be a random variable having the exp-SHL density with power parameter $j+1$ ($j \geq 0$) given by

$$h_{j+1}(x) = \frac{2\,(j+1)\,\mathrm{e}^{-x}\,(1-\mathrm{e}^{-x})^{j}}{(1+\mathrm{e}^{-x})^{j+2}}. \tag{4.13}$$

Clearly, several mathematical properties of X (such as the ordinary and incomplete moments, mean deviations and generating function) can be determined from those of the exp-SHL distribution using the linear representation (4.11). Some mathematical properties of the exp-SHL distribution are obtained by [63], which are reported in the next section.

4.6 Properties of the exp-SHL distribution

Henceforth, let Y_{j+1} $\sim$exp-SHL$(j+1)$ have the density function (4.13). We use throughout an equation for a power series raised to an integer $j = 1, 2, \ldots$

$$\left(\sum_{i=0}^{\infty} a_i\, x^i\right)^j = \sum_{i=0}^{\infty} c_{j,i}\, x^i, \tag{4.14}$$

where $a_0 \neq 0$, $c_{j,0} = a_0^j$ and the coefficients $c_{j,i}$ (for $i \geq 1$) are determined recursively by

$$c_{j,i} = \frac{1}{i a_0} \sum_{m=0}^{i} [m(j+1) - i]\, a_m\, c_{j,i-m}. \tag{4.15}$$

The nth moment of Y_{j+1} derived by expanding the binomial terms is given by

$$E(Y_{j+1}^n) = 2(j+1) \int_0^{\infty} x^n\, \mathrm{e}^{-x}\, \frac{(1-\mathrm{e}^{-x})^j}{(1+\mathrm{e}^{-x})^{j+2}} dx = (j+1) \sum_{i=0}^{\infty} \frac{c_{n,i}}{(j+1+i+n)}, \tag{4.16}$$

where the quantities $c_{n,i}$'s are obtained from equation (4.15) by taking $a_i = [1+(-1)^i]/(i+1)$ (for $i \geq 0$).

For empirical purposes, the shape of many distributions can be usefully described by the incomplete moments. They form natural building blocks for measuring inequality: for example, the Lorenz and Bonferroni curves depend upon the first incomplete moment of an income distribution. The nth incomplete moment of Y_{j+1} is given by

$$m_{j+1}(n; z) = \int_0^{z} x^n\, h_{j+1}(x)\, dx = (j+1) \sum_{i=0}^{\infty} \frac{c_{n,i}\, \tanh(z/2)^{j+1+i+n}}{(j+1+i+n)}, \tag{4.17}$$

where $\tanh(\cdot)$ is the hyperbolic tangent function.

The mgf of Y_{j+1}, say $M_{j+1}(s) = E(\mathrm{e}^{sY_{j+1}})$, can be expressed as

$$M_{j+1}(s) = (j+1)!\,\Gamma(1-s)\,{}_2\tilde{F}_1[j+1, -s; j+2-s; -1], \tag{4.18}$$

where ${}_2\tilde{F}_1$ is the regularized hypergeometric function defined by

$${}_2\tilde{F}_1[a, b; c; z] = \sum_{k=0}^{\infty} \frac{(a)_k\,(b)_k}{\Gamma(c+k)}\,\frac{z^k}{k!}, \quad |z| < 1,$$

$(a)_k = a(a-1)\ldots(a-k+1)$ (for $k > 1$) is the falling factorial, $(a)_0 = 1$, and $\Gamma(a) = \int_0^\infty x^{a-1}\,\mathrm{e}^{-\mathrm{x}}\mathrm{dx}$ is the gamma function. For $|z| < 1$ and arbitrary parameters a, b and c, the above infinite sum is convergent. See [63].

4.7 Properties of the EGSHL distribution

In this section, we obtain explicit expressions for moments, cumulants, mean deviations, Bonferroni and Lorenz curves and generating function of the EGSHL distribution. The formulae derived can be handled in most symbolic computation platforms such as *Mathematica* and *Maple* and they can be more efficient than computing the mathematical quantities directly by numerical integration of the density function (4.6). The infinity limit in these formulae can be substituted by a large positive integer such as 20 or 30 for most practical purposes.

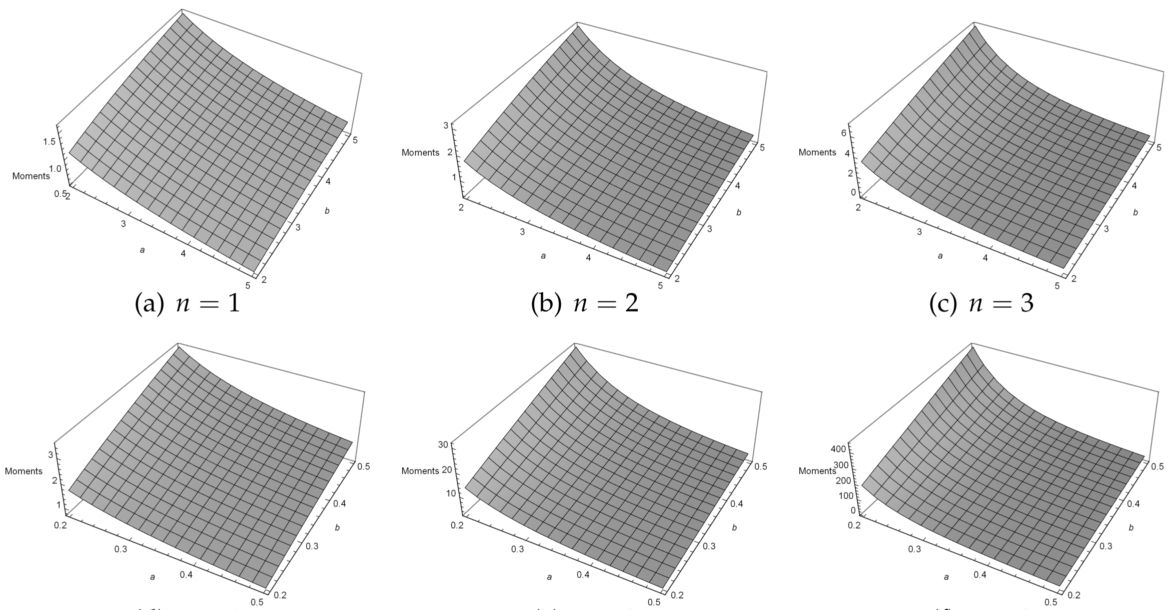
4.7.1 Moments

The statistical relevance for calculating moments, especially in applied research, is widely know in the literature. Next, we provide two ways to compute the nth moment of X with density (4.6). The first formula follows as

$$\mu_n' = E(X^n) = \int_0^\infty x^n\,\frac{a\,b\,2^a\,\mathrm{e}^{-ax}\,[(1+\mathrm{e}^{-x})^a - 2^a\,\mathrm{e}^{-ax}]^{b-1}}{(1+\mathrm{e}^{-x})^{ab+1}}\mathrm{d}x. \tag{4.19}$$

Although we do not have a closed-solution to the integral above, it is very simple to calculate the nth moment of X computationally, from equation (4.19). For illustrative purposes, we provide a small numerical study by computing $E(X^n)$ and the variance of X from (4.19) numerically. We consider several parameters values and $n = 1, 2, 3, 4, 5$. The results are given in Table 4.2 with five decimal digits of precision. All computations are performed using *Wolfram Mathematica* platform. Some plots of the EGSHL moments for some parameter values are display in Figure 4.9.

Based on the values in Table 4.2 and the plots in Figure 4.9, we conclude that the additional parameters a and b have large impact on the moments of X. Theses values and plots reveals that, in general, for fixed a parameter value, the moments and the variance increases when b increase. The inverse happens when we set values for b and the parameter a increases.



(a) $n = 1$ (b) $n = 2$ (c) $n = 3$

(d) $n = 1$ (e) $n = 2$ (f) $n = 3$

Figure 4.9: Plots of the EGSHL moments for some parameter values.

Alternatively, the nth moment of X can be obtained from equations (4.11) and (4.16) as

$$\mu'_n = \sum_{j,i=0}^{\infty} \frac{(j+1)\, w_{j+1}\, c_{n,i}}{(j+1+i+n)}, \tag{4.20}$$

where the quantities $c_{n,i}$ are defined in (4.16).

The central moments (μ_n) and cumulants (κ_n) of X can be determined from the

Table 4.2: First five moments and variance of X for several a and b values.

a	b	$\mathrm{E}(X)$	$\mathrm{E}(X^2)$	$\mathrm{E}(X^3)$	$\mathrm{E}(X^4)$	$\mathrm{E}(X^5)$	$\mathrm{E}(X^6)$	$\mathrm{Var}(X)$
	1	1.38629	3.28987	10.8185	45.4576	233.309	1419.19	1.36807
	2	2.00000	5.54518	19.7392	86.5481	454.576	2799.70	1.54518
1	3	2.38629	7.28987	27.4540	124.414	666.049	4146.65	1.59549
	4	2.66667	8.72690	34.3189	159.759	869.290	5463.90	1.61577
	5	2.88629	9.95653	40.5444	1.93052	1065.45	6754.52	1.62586
	1	0.77259	1.03456	1.89782	4.36705	12.0417	38.6822	0.43766
	2	1.12149	1.74814	3.45960	8.29363	23.3966	76.1291	0.49040
2	3	1.34151	2.29938	4.80462	11.8926	34.1919	112.506	0.49989
	4	1.50061	2.75176	5.99586	15.2347	44.5172	147.941	0.49993
	5	1.62474	3.13723	7.07113	18.3678	54.4388	182.532	0.49745
	1	0.54518	0.52394	0.69195	1.14285	2.24819	5.11402	0.22672
	2	0.79560	0.88925	1.26465	2.17282	4.36906	10.0624	0.25627
3	3	0.95453	1.17288	1.75929	3.11793	6.38539	14.8667	0.26175
	4	1.06985	1.40625	2.19804	3.99600	8.31344	19.5434	0.26167
	5	1.15990	1.60536	2.59437	4.81925	10.1654	24.1058	0.25999
	1	0.42369	0.32098	0.33603	0.44048	0.68689	1.23528	0.14147
	2	0.62075	0.54682	0.61577	0.83882	1.33604	2.43152	0.16149
4	3	0.74656	0.72307	0.85829	1.20520	1.95400	3.59362	0.16572
	4	0.83814	0.86856	1.07395	1.54616	2.54548	4.72537	0.16608
	5	0.90980	0.99296	1.26911	1.86623	3.11402	5.82982	0.16522
	1	0.34738	0.21827	0.19059	0.20891	0.27257	0.40988	0.09760
	2	0.51045	0.37296	0.35006	0.39844	0.53067	0.80725	0.11240
5	3	0.61507	0.49422	0.48878	0.57318	0.77674	1.19363	0.11591
	4	0.69146	0.59461	0.61244	0.73610	1.01256	1.57020	0.11649
	5	0.75134	0.68063	0.72456	0.88925	1.23945	1.93788	0.11612

ordinary moments μ'_n as

$$\mu_n = \sum_{k=0}^{r} (-1)^k \binom{n}{k} \mu_1'^n \, \mu'_{n-k} \qquad \text{and} \qquad \kappa_n = \mu'_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \kappa_k \, \mu'_{n-k},$$

respectively, where $\kappa_1 = \mu'_1$. The skewness $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and kurtosis $\gamma_2 = \kappa_4/\kappa_2^2$ of X follow from the third and fourth standardized cumulants.

The pth descending factorial moment of X is

$$\mu'_{(p)} = E[X^{(p)}] = E[X(X-1) \times \cdots \times (X-p+1)] = \sum_{k=0}^{p} s(p,k)\, \mu'_k,$$

where $s(r,k) = (k!)^{-1}[d^k x^{(r)}/dx^k]_{x=0}$ is the Stirling number of the first kind. So, we can obtain the factorial moments from the ordinary moments given before.

4.7.2 Incomplete moments and their applications

The nth incomplete moment of X, say $m(n;y) = \int_0^y x^n\, f(x)\, dx$, can be determined from (4.11) and (4.17) as

$$m(n;y) = \sum_{j,i=0}^{\infty} \frac{(j+1)\, w_{j+1}\, c_{n,i} \tanh(z/2)^{j+1+i+n}}{(j+1+i+n)}. \tag{4.21}$$

Generally, there has been a great interest in obtaining the first incomplete moment of a distribution. The mean residual function follows from (4.21) with $n = 1$ as $\mu'_1 - m(1;y) - y$. Based on the first incomplete moment, we can obtain mean deviations from the mean and the median defined by $\delta_1 = E(|X - \mu'_1|) = 2\mu'_1\, F(\mu'_1) - 2m(1;\mu'_1)$ and $\delta_2 = E(|X - M|) = \mu'_1 - 2m(1;M)$, where the mean μ'_1 and the median M follow from (4.20) and (4.9), respectively.

Equation (4.21) with $n = 1$ is also useful to derive the Bonferroni and Lorenz curves defined (for a given probability π) by $B(\pi) = m(1;q)/(\pi\, \mu'_1)$ and $L(\pi) = m(1;q)/\mu'_1$, respectively, where $q = Q(\pi)$ follows from (4.9).

As further applications of $m_n(y)$ one has the mean residual life and mean waiting time given by $v(t) = \{[1 - m_1(t)]/S(t)\} - t$ and $\mu(t) = t - [m_1(t)/F(t)]$, respectively, where $S(t)$ and $F(t)$ are obtained from (4.7) and (4.5).

4.7.3 Generating function

The mgf of X, say $M(s) = E(\mathrm{e}^{sX})$, can be obtained from (4.11) and (4.18) (for $s \neq 0, 1, 2, \ldots$) as

$$M(s) = \sum_{j=0}^{\infty} (j+1)\, w_{j+1}\, \Gamma(1-s)\, {}_2\tilde{F}_1[j+1, -s; j+2-s; -1].$$

4.7.4 Reliability

Here, we derive the reliability, say R, for the EGSHL model when $X_1 \sim \mathrm{EGSHL}(a_1, b_1)$ and $X_2 \sim \mathrm{EGSHL}(a_2, b_2)$ are two independent random variables. Let $f_1(x)$ denote the pdf of X_1 and $F_2(x)$ denote the cdf of X_2. The reliability can be expressed as $R = \mathrm{P}(X_1 > X_2) = \int_0^\infty f_1(x)\, F_2(x)\, \mathrm{d}x$ and using equations (4.11) and (4.12) gives

$$R = \sum_{j,k=0}^{\infty} \mathcal{I}_{j,k} \int_0^\infty h_{j+1}(x)\, H_{k+1}(x)\, \mathrm{d}x,$$

where $\mathcal{I}_{j,k} = \sum_{m,n=1}^{\infty} (-1)^{j+k+m+n+2} \binom{b_1}{m} \binom{m\, a_1}{j+1} \binom{b_2}{n} \binom{n a_2}{k+1}$.

Thus, the reliability of X reduces to

$$R = \sum_{j,k=0}^{\infty} \mathcal{I}_{j,k} \int_0^\infty \frac{2\,(j+1)\, \mathrm{e}^{-x}\, (1-\mathrm{e}^{-x})^{j+k+1}}{(1+\mathrm{e}^{-x})^{j+k+3}}\, \mathrm{d}x$$

and then

$$R = \sum_{j,k=0}^{\infty} \frac{(j+1)\, \mathcal{I}_{j,k}}{(j+k+2)}. \tag{4.22}$$

Table 4.3 gives some values of R for the EGSHL model considering different parameter values. Naturally, for $a_1 = a_2$ and $b_1 = b_2$ we obtain $R = \mathrm{P}(X_1 > X_2) = 1/2$. All computations are obtained using *Wolfram Mathematica* software and we consider de upper limits equal 30 in (4.22).

4.8 Probability weighted moments

The PWMs are used to derive estimators of the parameters and quantiles of generalized distributions. The moment method of estimation is formulated by equating

Table 4.3: The reliability of $X \sim$ EGSHL for $(a_1 = 2, a_2 = 2)$ and some of b_1 and b_2 values.

b_2	2	3	4	5	6
b_1					
2	0.50000	0.40000	0.33333	0.28571	0.25000
3	0.60000	0.50000	0.42857	0.37500	0.33333
4	0.66667	0.57143	0.50000	0.44444	0.40000
5	0.71429	0.62500	0.55556	0.50000	0.45455
6	0.75000	0.66667	0.60000	0.54545	0.50000

the population and sample PWMs. These moments have low variances and no severe biases, and they compare favorably with estimators obtained by maximum likelihood. The (s, r)th PWM of X is defined by $\delta_{s,r} = E[X^s\, F(x)^r]$. Clearly, the ordinary moments follow as $\delta_{s,0} = E(X^s)$. Next, we derive simple expressions for the PWMs of X defined as

$$\delta_{s,r} = \int_0^\infty x^s\, F(x)^r\, f(x)\, \mathrm{d}x. \tag{4.23}$$

Inserting (4.5) and (4.6) in equation (4.23), the PWMs of X can be expressed in a simple form:

$$\delta_{s,r} = \int_0^\infty x^s\, \frac{a\, b\, 2^a\, \mathrm{e}^{-ax}\, [(1+\mathrm{e}^{-x})^a - 2^a\, \mathrm{e}^{-ax}]^{b(r+1)-1}}{(1+\mathrm{e}^{-x})^{ab(r+1)+1}}\, \mathrm{d}x. \tag{4.24}$$

Table 4.4 gives the values of $\delta_{s,r}$ for $X \sim \text{EGSHL}(2,\, 3)$ and several values of s and r by using (4.24). All computations are obtained using *Wolfram Mathematica* software.

We can go further and present a simpler expression to the PWM of X. Under simple algebraic manipulation, we can write $\delta_{s,r}$ as

$$\delta_{s,r} = \frac{1}{(r+1)} \int_0^\infty x^s\, f[x;\, a,\, (r+1)b]\, \mathrm{d}x. \tag{4.25}$$

Table 4.4: The PWM of $X \sim$ EGSHL(2, 3) and several values of s and r.

r	1	2	3	4	5	6	7
s							
1	0.86311	0.65040	0.52742	0.44636	0.38851	0.34493	0.31083
2	1.73709	1.43084	1.23227	1.09076	0.98370	0.89926	0.83057
3	4.02758	3.53658	3.18774	2.92247	2.71156	2.53844	2.39258
4	10.6632	9.79486	9.13100	8.59828	8.15643	7.78109	7.45605

where $f[x;\, a,\, (r+1)b]$ is the EGSHL density with parameters a and $(r+1)b$. This is the most important result of this section. Equation (4.25) revels that the PWM of X can be expressed in terms of the ordinary moments of $X \sim \text{EGSHL}[a,\, (r+1)b]$.

4.9 Rényi Entropy

Given a certain random phenomenon under study, it is important to quantify the uncertainty associated with the random variable of interest. In this context, several statistical methods are available in the literature. One of the most popular measures used to quantify the variability of X is the Rényi entropy. Here, we make reference to the following papers: [72], for the gamma extended Fréchet model; [73] for the Marshall-Olkin extended modified Weibull distribution and [74] for an extension of the logistic distribution, among others.

The Rényi entropy of X with density (4.6), say $I_R(\rho)$, is given by

$$I_R(\rho) \quad = \quad \frac{1}{(1-\rho)} \log\left(\int_0^\infty f(x)^\rho \mathrm{d}x\right), \tag{4.26}$$

where $\rho > 0$ and $\rho \neq 1$.

By inserting (4.6) in equation (4.26), we obtain

$$I_R(\rho) = \frac{1}{(1-\rho)} \log\left(\int_0^\infty \left[\frac{a\, b\, 2^a\, \mathrm{e}^{-ax}\, [(1+\mathrm{e}^{-x})^a - 2^a\, \mathrm{e}^{-ax}]^{b-1}}{(1+\mathrm{e}^{-x})^{ab+1}}\right]^\rho \mathrm{d}x\right). \tag{4.27}$$

Equation (4.27) can be easily implemented computationally and values of $I_R(\rho)$ are obtained in a few seconds. Table 4.5 shows some values of $I_R(\rho)$ for the EGSHL model, considering different parameter values. Naturally, the higher the value of $I_R(\rho)$, indicates the greater uncertainty about the phenomenon under study. All computations are obtained using *Wolfram Mathematica* software, which have numerical integration routines with great precision. Based on the figures in Table 4.5, we note that, independently of a and b, $I_R(\rho)$ decreases when ρ increases. For fixed ρ, the Rényi entropy is larger for $a < b$.

Table 4.5: Rényi entropy of X for several ρ, a and b values.

a	b	$\rho = 2$	$\rho = 4$	$\rho = 6$	$\rho = 8$	$\rho = 10$
2	3	0.80620	0.68292	0.62861	0.59677	0.57546
2	2	0.76897	0.64952	0.59687	0.56595	0.54522
3	2	0.44012	0.32050	0.26770	0.23670	0.21590

Although, as we cited before, equation (4.27) is easily manageable computationally, we provide an expression in closed-form to compute $I_R(\rho)$. Using (4.10) twice in equation (4.4), we can write

$$f(x)^{\rho} = (ab)^{\rho} \sum_{k,\ell=0}^{\infty} (-1)^{k+\ell} \binom{\rho(b-1)}{k} \binom{ak+\rho(a-1)}{\ell} g(x)^{\rho} G(x)^{\ell}. \tag{4.28}$$

Substituting (4.1) and (4.2) in the equation (4.28), we have

$$\begin{aligned} I_R(\rho) \quad = \quad & \frac{1}{(1-\rho)} \log \Big\{ \frac{1}{2} \sqrt{\pi}\, (a\,b)^{\rho}\, \Gamma(1+\rho) \sum_{k,\ell=0}^{\infty} (-1)^{k+\ell}\, 2^{\ell} \binom{\rho(b-1)}{k} \\ & \times \binom{ak+\rho(a-1)}{\ell} \mathrm{K}(\ell,\rho) \Big\}, \end{aligned}$$

where

$$\begin{aligned} \mathrm{K}(\ell,\rho) \quad = & \; \ell\, \Gamma\left(\tfrac{1+\ell}{2}\right) {}_3\tilde{F}_2\left[\tfrac{1}{2}-\ell,\, 1-\ell,\, \tfrac{1+\ell}{2};\, \tfrac{3}{2},\, \tfrac{3+\ell}{2}+\rho;\, 1\right] \\ & + \Gamma\left(\tfrac{1}{2}\right) {}_3\tilde{F}_2\left[\tfrac{1}{2}-\ell,\, -\ell,\, \tfrac{1}{2};\, \tfrac{1}{2},\, 1+\rho+\tfrac{\ell}{2};\, 1\right] \end{aligned}$$

is the regularized hypergeometric function defined by

$$ {}_3\tilde{F}_2[a,b,c;d,e;z] = \sum_{k=0}^{\infty} \frac{(a)_k\,(b)_k\,(c)_k}{\Gamma(d+k)\,\Gamma(e+k)}\,\frac{z^k}{k!}, \quad |z|<1. $$

4.10 Order statistics

The importance of order statistics and their applications is widely disseminated in the literature. As define by [75], the main objective of the order statistics is the investigation of properties and applications of ordered random variables, as well as functions of these variables. The density function $f_{i:n}(x)$ of the ith order statistic, say $X_{i:n}$, based on a random sample $X_1,\ldots,X_n$, can be expressed as (for $i=1,\ldots,n$)

$$ f_{i:n}(x) = \frac{1}{B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x)\,F(x)^{i+j-1}. $$

By inserting (4.5) and (4.6) in the above expression, the density function of the EGSHL order statistics follow as

$$ f_{i:n}(x) = \frac{1}{B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \frac{a\,b\,2^a\,\mathrm{e}^{-ax}\left[(1+\mathrm{e}^{-x})^a - 2^a\,\mathrm{e}^{-ax}\right]^{b(i+j)-1}}{(1+\mathrm{e}^{-x})^{ab(i+j)+1}}. \tag{4.29} $$

There are many practical applications in which we can employ the above equation. Perhaps, the most important of these refers to the moments of $X_{i:n}$. The r-th moment of $X_{i:n}$ comes from (4.29) as

$$ E(X_{i:n}^r) = \frac{1}{B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \int_0^{\infty} x^r \frac{a\,b\,2^a\mathrm{e}^{-ax}\left[(1+\mathrm{e}^{-x})^a - 2^a\,\mathrm{e}^{-ax}\right]^{b(i+j)-1}}{(1+\mathrm{e}^{-x})^{ab(i+j)+1}} \mathrm{d}x. \tag{4.30} $$

The r-th moment of $X_{i:n}$ can be easily obtained numerically using (4.30) through any symbolic computing platform. In Table 4.6, we present a small illustration, in which we calculate the first five moments of $X_{i:10}$ for $a=b=2$ and some values of r and i. All computations are performed using the *Wolfram Mathematica* platform. For a similar study, readers may see a paper by [76], who evaluated $E(X_{i:n}^r)$ numerically for the Weibull-geometric distribution.

Table 4.6: The first five moments of $X_{i:10}$ for $a = b = 2$ and some values of r and i.

$r \rightarrow$	1	2	3	4	5
$i \downarrow$					
1	0.30388	0.12086	0.05784	0.03183	0.01991
5	0.93164	0.92874	0.98474	1.10008	1.33621
9	1.81708	3.49876	7.13339	15.3918	35.1363
10	2.38815	6.16922	17.2845	52.6660	174.968

Finally, we provide a linear representation for $f_{i:n}(x)$. After a simple algebraic manipulation, we can write

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} \sum_{j=0}^{n-i} \xi_{i,j}\, f[x;\, a,\, (i+j)b], \tag{4.31}$$

where $\xi_{i,j} = [(-1)^j/(i+j)]\binom{n-i}{j}$ and $f[x;\, a,\, (i+j)b]$ is the EGSHL density with parameters a and $(i+j)b$. Equation (4.31) revels that the pdf of $X_{i:n}$ is a linear combination of EGSHL densities. So, the moments, incomplete moments and other quantities for the EGSHL order statistics can be determined from the above expression.

4.11 Estimation and Inference

Several approaches for parameter estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used when constructing confidence intervals and regions and also in test statistics. The normal approximation for these estimators in large sample distribution theory is easily handled either analytically or numerically. So, we consider the estimation of the unknown parameters for this family from complete samples only by maximum likelihood. Let $x_1, \ldots, x_n$ be observed values from the EGSHL distribution with parameters a and b.

The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (a, b)^\top$, say $\ell(\boldsymbol{\theta})$, can be expressed as

$$\ell(\boldsymbol{\theta}) = n\log(ab2^a) - a\sum_{i=1}^{n} x_i + (b-1)\sum_{i=1}^{n}\log[(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}] - (ab+1)\sum_{i=1}^{n}\log(1+\mathrm{e}^{-x_i}). \tag{4.32}$$

Equation (4.32) can be maximized either directly by using the `R` (`optim` function), `SAS` (`PROC NLMIXED`) or `Ox` program (sub-routine `MaxBFGS`) or by solving the nonlinear likelihood equations obtained by differentiating (4.32). The components of the score function are:

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\theta})}{\partial a} &= -\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n}\log(1+\mathrm{e}^{-x_i}) + n\,a^{-1}\,[1 + a\,\log(2)] \\
&\quad + (b-1)\sum_{i=1}^{n}\frac{2^a\,x_i\,\mathrm{e}^{-ax_i} - 2^a\,\log(2)\,\mathrm{e}^{-ax_i} + (1+\mathrm{e}^{-x_i})^a\,\log(1+\mathrm{e}^{-x_i})}{(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}}, \\
\frac{\partial\ell(\boldsymbol{\theta})}{\partial b} &= \frac{n}{b} - a\sum_{i=1}^{n}\log(1+\mathrm{e}^{-x_i}) + \sum_{i=1}^{n}\log[(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}].
\end{aligned}$$

The elements of the observation matrix $J(\boldsymbol{\theta})$ are given by

$$\begin{aligned}
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial a^2} &= -(b-1)\sum_{i=1}^{n}\frac{2^{a+1}\,\log(2)\,x_i\,\mathrm{e}^{-ax_i} - 2^a\,x_i^2\,\mathrm{e}^{-ax_i} - 2^a\,[\log(2)]^2\,\mathrm{e}^{-ax_i}}{(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}} \\
&\quad + (b-1)\sum_{i=1}^{n}\frac{(1+\mathrm{e}^{-x_i})^a[\log(1+\mathrm{e}^{-x_i})]^2}{(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}} \\
&\quad - (b-1)\sum_{i=1}^{n}\frac{[2^a\,x_i\,\mathrm{e}^{-ax_i} - -2^a\,[\log(2)]^2\,\mathrm{e}^{-ax_i} + (1+\mathrm{e}^{-x_i})^a\log(1+\mathrm{e}^{-x_i})]^2}{[(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}]^2} \\
&\quad + n\,a^{-1}\,\{2\,\log(2) + a\,[log(2)]^2\} - n\,a^{-2}\,[1 + a\,\log(2)] - n\,a^{-1}\,\log(2)\,[1 + a\,\log(2)], \\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial b^2} &= \frac{n}{b^2}, \\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial ab} &= \sum_{i=1}^{n}\log(1+\mathrm{e}^{-x_i}) + \sum_{i=1}^{n}\frac{2^a\,x_i\,\mathrm{e}^{-ax_i} - 2^a\,\log(2)\,\mathrm{e}^{-ax_i} + (1+\mathrm{e}^{-x_i})^a\,\log(1+\mathrm{e}^{-x_i})}{(1+\mathrm{e}^{-x_i})^a - 2^a\,\mathrm{e}^{-ax_i}}.
\end{aligned}$$

For large n and under conditions of regularity already known, the distribution of $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ can be approximated to a bivariate normal distribution with zero means and variance-covariance matrix $J(\boldsymbol{\theta})^{-1}$. Some structural properties of $\hat{\boldsymbol{\theta}}$ can be derived based on this normal approximation.

4.12 Simulation Study

Among the many estimation methods presented in the literature, the maximum likelihood method is the most commonly used by applied researchers. But, once the estimation method is chosen, it is necessary to observe if the parameters of the model are obtained with precision, because the resulting inferences and the decision processes will depend directly on the quality of the estimates. In this section we present a Monte Carlo simulation study, with the aim of investigating the behavior of the MLEs for the parameters of the EGSHL model. For doing this, we generate samples sizes $n = 20, 40, 80, 120$ using (4.9) and $10,000$ replications. We select some values for a and b. The entire simulation process is performed in the `R` software, using the *simulated-annealing* (SANN) maximization method in *maxLik* package. To ensure the reproducibility of the experiment we use the seed for the random number generator: set.seed (103). Finally, it should be mentioned that our methodology used, as an ad hoc rule, initial kicks are equal to half of the true value of the parameters in each scenario considered.

The results of the simulations are presented in Tables 4.7 and 4.8, which contains all the estimates for each case considered and also the variance between parentheses. These results show that EGSHL estimates have good properties even for small to moderate sample sizes . In general, the biases and variances decrease as the sample size increases, as expected.

4.13 Application to real data

In this section, we present a small application study to a real data set. The objective is to demonstrate that the EGSHL distribution can be used in practical situations for real data modeling. We consider the set of data presented by [63] referring to the soil fertility influence and the characterization of the biologic fixation of N2 for the Dimorphandra wilsonii rizz growth. The phosphorus concentration, in the leaves, for 128 plants are:

0.22, 0.17, 0.11, 0.10, 0.15, 0.06, 0.05, 0.07, 0.12, 0.09, 0.23, 0.25, 0.23, 0.24, 0.20, 0.08, 0.11, 0.12, 0.10, 0.06, 0.20, 0.17, 0.20, 0.11, 0.16, 0.09, 0.10, 0.12, 0.12, 0.10, 0.09, 0.17, 0.19, 0.21, 0.18, 0.26, 0.19, 0.17, 0.18, 0.20, 0.24, 0.19, 0.21, 0.22, 0.17, 0.08, 0.08, 0.06, 0.09, 0.22, 0.23, 0.22, 0.19, 0.27, 0.16, 0.28, 0.11, 0.10, 0.20, 0.12, 0.15, 0.08, 0.12, 0.09, 0.14, 0.07, 0.09, 0.05, 0.06, 0.11, 0.16, 0.20, 0.25, 0.16, 0.13, 0.11, 0.11, 0.11, 0.08, 0.22, 0.11, 0.13, 0.12, 0.15, 0.12, 0.11, 0.11, 0.15, 0.10, 0.15, 0.17, 0.14, 0.12, 0.18, 0.14, 0.18, 0.13, 0.12, 0.14, 0.09, 0.10, 0.13, 0.09, 0.11, 0.11, 0.14, 0.07, 0.07, 0.19, 0.17, 0.18, 0.16, 0.19, 0.15, 0.07, 0.09, 0.17, 0.10, 0.08, 0.15, 0.21, 0.16, 0.08, 0.10, 0.06, 0.08, 0.12, 0.13.

Figure 4.10 shows the dispersion of the data, while Table 4.9 provides some descriptive statistics. It is possible to observe, for example, that the mean and the variance are 0.1408 and 0.0030, respectively. We adopt the maximum likelihood method to estimate the model parameters and all computations are performed using the SAS subroutine NLMixed.

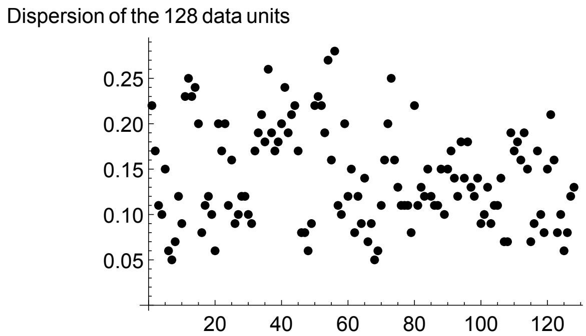
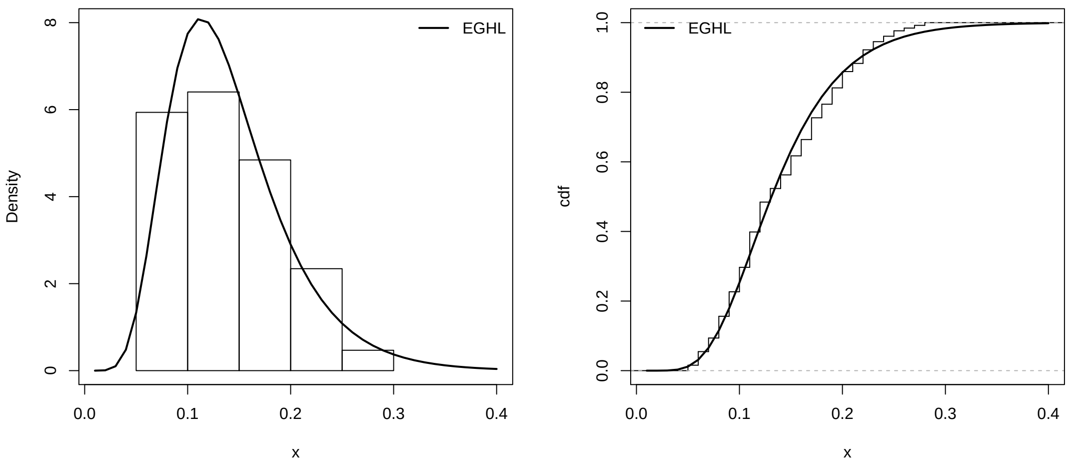


Figure 4.10: Dispersion of the 128 data units of the phosphorus concentration in the leaves

In addition to the EGSHL model and its sub-models ESHLI and ESHLII, we consider the McDonald half-logistic (MCSHL) distribution and Kumaraswamy half-logistic (KWSHL) models, introduced by [63]. Table 4.10 lists the MLEs of the model parameters (with the corresponding standard errors in parentheses) for all fitted models and also the values of the Akaike information criterion (AIC), Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) statistics. In general, it is considered that the lower values of these criteria indicates the better fit to the data. The figures in Table 4.10 revels that the EGSHL model has the lowest AIC, BIC and CAIC values among all fitted models. Thus, the proposed EGSHL distribution is the best model to explain these data.

Finally, Figure 4.11 displays the estimated pdf and cdf of the EGSHL model and the histogram of the data. These plots reveal that the proposed model is quite suitable for these data.



(a) Estimated pdf of the EGSHL model (b) Estimated cdf of the EGSHL model

Figure 4.11: Estimated pdf and cdf of the EGSHL model for phosphorus concentration in leaves data

4.14 Conclusions

In this paper, we introduce a univariate continuous distribution with two parameters that govern the asymmetry and kurtosis, named *the exponentiated generalized standard half-logistic* model, say EGSHL. We provide a comprehensive mathematical treatment and show by numerical studies that the formulas related to the new model are computationally manageable. In particular, the maximum likelihood estimators are easily estimated. These estimators have good properties, such as low biases and variances, even in small or moderate sample sizes. A study using real data shows that the new distribution can be used in practical situations due to its great power of adjustment when compared to other competitive models. We hope that this EGSHL model can be useful for applied statisticians and other researchers who refer to a model with few parameters but flexible to accommodate supported data in real positives. For future research, we will study bias correction via bootstrap for estimators in small samples.

Table 4.7: MLEs for several a and b parameter values (variances in parentheses).

		$n = 20$		$n = 40$		$n = 80$		$n = 120$	
a	b	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
1	1	1.1364	1.1665	1.0679	1.0780	1.0322	1.0362	1.0219	1.0244
		(0.1291)	(0.1816)	(0.0533)	(0.0608)	(0.0234)	(0.0244)	(0.0152)	(0.0156)
1	2	1.1079	2.4438	1.0542	2.2022	1.0257	2.0922	1.0176	2.0620
		(0.0861)	(1.1979)	(0.0369)	(0.3524)	(0.0166)	(0.1336)	(0.0108)	(0.0843)
1	3	1.0975	3.7774	1.0495	3.3564	1.0235	3.1610	1.0160	3.1078
		(0.0714)	(3.2781)	(0.0315)	(1.0175)	(0.0142)	(0.3703)	(0.0093)	(0.2316)
1	4	1.0901	5.1139	1.0471	4.5328	1.0224	4.2394	1.0152	4.1599
		(0.0617)	(6.0177)	(0.0286)	(2.1581)	(0.0130)	(0.7694)	(0.0085)	(0.4777)
1	5	1.0825	6.3945	1.0453	5.7225	1.0216	5.3249	1.0147	5.2169
		(0.0535)	(8.6183)	(0.0267)	(3.7875)	(0.0122)	(1.3586)	(0.0080)	(0.8401)
1	6	1.0739	7.5949	1.0435	6.9057	1.0211	6.4185	1.0144	6.2786
		(0.0465)	(10.8839)	(0.0250)	(5.6409)	(0.0117)	(2.1707)	(0.0076)	(1.3332)
2	1	2.2728	1.1665	2.1358	1.0780	2.0644	1.0362	2.0439	1.0244
		(0.5161)	(0.1815)	(0.2130)	(0.0608)	(0.0938)	(0.0244)	(0.0606)	(0.0156)
2	2	2.2158	2.4432	2.1083	2.2021	2.0515	2.0923	2.0351	2.0620
		(0.3442)	(1.1851)	(0.1476)	(0.3523)	(0.0663)	(0.1336)	(0.0433)	(0.0844)
2	3	2.1946	3.7732	2.0989	3.3562	2.0471	3.1610	2.0320	3.1078
		(0.2844)	(3.2076)	(0.1259)	(1.0163)	(0.0569)	(0.3700)	(0.0372)	(0.2317)
2	4	2.1784	5.0913	2.0940	4.5316	2.0447	4.2392	00000	00000
		(0.2433)	(5.6459)	(0.1143)	(2.1459)	(0.0519)	(0.7688)	(000000)	(000000)
2	5	2.1610	6.3515	2.0905	5.7206	2.0433	5.3254	2.0295	5.2170
		(0.2085)	(8.1286)	(0.1066)	(3.7591)	(0.0488)	(1.3598)	(0.0319)	(0.8398)
2	6	2.1413	7.5129	2.0873	6.9111	2.0423	6.4184	2.0289	6.2790
		(0.1784)	(10.0107)	(0.1002)	(5.7568)	(0.0467)	(2.1702)	(0.0306)	(1.3345)
3	1	3.4091	1.1665	3.2037	1.0780	3.0967	1.0362	3.0658	1.0244
		(1.1614)	(0.1816)	(0.4793)	(0.0608)	(0.2110)	(0.0244)	(0.1364)	(0.0156)
3	2	3.3236	2.4429	3.1626	2.2023	3.0773	2.0923	3.0527	2.0620
		(0.7739)	(1.1783)	(0.3322)	(0.3523)	(0.1493)	(0.1336)	(0.0973)	(0.0843)
3	3	3.2920	3.7749	3.1484	3.3562	3.0706	3.1609	3.0481	3.1078
		(0.6405)	(3.2347)	(0.2832)	(1.0168)	(0.1280)	(0.3701)	(0.0836)	(0.2318)
3	4	3.2681	5.0951	3.1411	4.5327	3.0670	4.2391	3.0457	4.1599
		(0.5486)	(5.6934)	(0.2575)	(2.1610)	(0.1168)	(0.7687)	(0.0764)	(0.4779)
3	5	3.2416	6.3449	3.1355	5.7184	3.0649	5.3253	3.0442	5.2170
		(0.4708)	(7.9364)	(0.2393)	(3.7228)	(0.1099)	(1.3602)	(0.0719)	(0.8397)
3	6	3.2104	7.4949	3.1293	6.8942	3.0634	6.4181	3.0432	6.2787
		(0.4016)	(9.7370)	(0.2218)	(5.4756)	(0.1050)	(2.1678)	(0.0687)	(1.3336)
4	1	4.5455	1.1665	4.2715	1.0780	4.1289	1.0362	4.0878	1.0244
		(2.0640)	(0.1815)	(0.8520)	(0.0607)	(0.3753)	(0.0244)	(0.2424)	(0.0156)
4	2	4.4313	2.4427	4.2168	2.2024	4.1031	2.0923	4.0702	2.0619
		(1.3750)	(1.1782)	(0.5907)	(0.3527)	(0.2654)	(0.1336)	(0.1729)	(0.0843)
4	3	4.3891	3.7720	4.1979	3.3560	4.0941	3.1609	4.0641	3.1077
		(1.1352)	(3.1640)	(0.5032)	(1.0154)	(0.2275)	(0.3701)	(0.1486)	(0.2316)
4	4	4.3588	5.1031	4.1880	4.5314	4.0895	4.2391	4.0609	4.1599
		(0.9799)	(5.7925)	(0.4573)	(2.1404)	(0.2077)	(0.7686)	(0.1359)	(0.4780)
4	5	4.3254	6.3667	4.1805	5.7169	4.0865	5.3253	4.0590	5.2171
		(0.8448)	(8.2116)	(0.4245)	(3.7004)	(0.1953)	(1.3602)	(0.1278)	(0.8398)
4	6	4.2843	7.5198	4.1724	6.8951	4.0846	6.4182	4.0577	6.2787
		(0.7193)	(9.9206)	(0.3944)	(5.4947)	(0.1866)	(2.1674)	(0.1221)	(1.3317)

Table 4.8: MLEs for several a and b parameter values (variances in parentheses).

		$n = 20$		$n = 40$		$n = 80$		$n = 120$	
a	b	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
5	1	5.6817	1.1665	5.3393	1.0780	5.1612	1.0362	5.1097	1.0244
		(3.2234)	(0.1815)	(1.3307)	(0.0607)	(0.5864)	(0.0244)	(0.3787)	(0.0156)
5	2	5.5387	2.4418	5.2707	2.2021	5.1287	2.0923	5.0877	2.0620
		(2.1444)	(1.1620)	(0.9228)	(0.3524)	(0.4147)	(0.1336)	(0.2702)	(0.0844)
5	3	5.4866	3.7734	5.2473	3.3561	5.1176	3.1609	5.0802	3.1078
		(1.7733)	(3.1906)	(0.7866)	(1.0163)	(0.3555)	(0.3702)	(0.2323)	(0.2317)
5	4	5.4483	5.1013	5.2350	4.5321	5.1117	4.2390	5.0762	4.1598
		(1.5275)	(5.7708)	(0.7149)	(2.1531)	(0.3245)	(0.7685)	(0.2122)	(0.4777)
5	5	5.4088	6.3755	5.2258	5.7183	5.1083	5.3253	5.0738	5.2172
		(1.3238)	(8.3123)	(0.6646)	(3.7192)	(0.3051)	(1.3594)	(0.1996)	(0.8400)
5	6	5.3610	7.5440	5.2161	6.8971	5.1057	6.4179	5.0721	6.2787
		(1.1379)	(10.1558)	(0.6190)	(5.5159)	(0.2915)	(2.1674)	(0.1909)	(1.3327)
6	1	6.8182	1.1665	6.4072	1.0780	6.1935	1.0362	6.1315	1.0244
		(4.6354)	(0.1815)	(1.9166)	(0.0607)	(0.8443)	(0.0244)	(0.5458)	(0.0156)
6	2	6.6464	2.4416	6.3250	2.2021	6.1544	2.0922	6.1052	2.0620
		(3.0857)	(1.1553)	(1.3287)	(0.3524)	(0.5973)	(0.1336)	(0.3893)	(0.0844)
6	3	6.5814	3.7674	6.2971	3.3564	6.1410	3.1609	6.0961	3.1077
		(2.5331)	(3.1065)	(1.1329)	(1.0158)	(0.5120)	(0.3702)	(0.3342)	(0.2316)
6	4	6.5345	5.0899	6.2817	4.5315	6.1340	4.2391	6.0914	4.1598
		(2.1763)	(5.5942)	(1.0288)	(2.1487)	(0.4672)	(0.7686)	(0.3055)	(0.4777)
6	5	6.4870	6.3639	6.2714	5.7206	6.1298	5.3254	6.0884	5.2170
		(1.8878)	(8.1698)	(0.9581)	(3.7620)	(0.4393)	(1.3597)	(0.2875)	(0.8395)
6	6	6.4294	7.5291	6.2615	6.9092	6.1268	6.4178	6.0865	6.2788
		(1.6218)	(9.9311)	(0.8991)	(5.7015)	(0.4197)	(2.1642)	(0.2749)	(1.3334)
7	1	7.9522	1.1663	7.4750	1.0780	7.2256	1.0362	7.1535	1.0244
		(6.2724)	(0.1811)	(2.6087)	(0.0608)	(1.1489)	(0.0244)	(0.7424)	(0.0156)
7	2	7.7540	2.4417	7.3791	2.2022	7.1803	2.0923	7.1228	2.0620
		(4.1999)	(1.1631)	(1.8074)	(0.3523)	(0.8131)	(0.1336)	(0.5296)	(0.0844)
7	3	7.6759	3.7623	7.3464	3.3563	7.1645	3.1610	7.1121	3.1078
		(3.4208)	(3.0301)	(1.5413)	(1.0165)	(0.6967)	(0.3701)	(0.4551)	(0.2318)
7	4	7.6208	5.0854	7.3288	4.5315	7.1564	4.2391	7.1066	4.1598
		(2.9346)	(5.5863)	(1.3999)	(2.1457)	(0.6361)	(0.7687)	(0.4159)	(0.4778)
7	5	7.5606	6.3387	7.3166	5.7200	7.1514	5.3252	7.1033	5.2171
		(2.5330)	(7.8692)	(1.3043)	(3.7473)	(0.5979)	(1.3601)	(0.3912)	(0.8395)
7	6	7.4924	7.4992	7.3040	6.9033	7.1480	6.4181	7.1008	6.2785
		(2.1695)	(9.6662)	(1.2178)	(5.6022)	(0.5715)	(2.1686)	(0.3742)	(1.3331)
8	1	9.0832	1.1657	8.5427	1.0780	8.2579	1.0362	8.1753	1.0244
		(8.0944)	(0.1800)	(3.4079)	(0.0608)	(1.5003)	(0.0244)	(0.9699)	(0.0156)
8	2	8.8614	2.4406	8.4332	2.2021	8.2060	2.0923	8.1402	2.0619
		(5.4792)	(1.1425)	(2.3624)	(0.3523)	(1.0618)	(0.1336)	(0.6919)	(0.0844)
8	3	8.7708	3.7586	8.3956	3.3560	8.1882	3.1610	8.1282	3.1077
		(4.4672)	(2.9786)	(2.0137)	(1.0162)	(0.9102)	(0.3703)	(0.5945)	(0.2316)
8	4	8.6995	5.0610	8.3757	4.5316	8.1788	4.2391	8.1219	4.1598
		(3.7735)	(5.2886)	(1.8292)	(2.1457)	(0.8304)	(0.7687)	(0.5433)	(o.4778)
8	5	8.6215	6.2896	8.3616	5.7194	8.1730	5.3253	8.1180	5.2170
		(3.1863)	(7.3155)	(1.7022)	(3.7422)	(0.7809)	(1.3598)	(0.5109)	(0.8396)
8	6	8.5478	7.4592	8.3467	6.9028	8.1691	6.4183	8.1151	6.2784
		(2.7600)	(9.3019)	(1.5899)	(5.6189)	(0.7459)	(2.1677)	(0.4886)	(1.3321)

Table 4.9: Descriptives statistics for phosphorus concentration in leaves data.

Statistic	
n	128
Mean	0.1408
Median	0.1300
Variance	0.0030
Minimum	0.0500
Maximum	0.2800

Table 4.10: MLEs (and the corresponding standard errors in parentheses), AIC, BIC and CAIC statistics for phosphorus concentration in leaves data.

Distribution	$\widehat{a}$	$\widehat{b}$		AIC	BIC	CAIC
EGSHL	39.5048	9.7074		**−388.6**	**−382.9**	**-388.5**
	(3.1128)	(1.8281)				
ESHLI	1	0.3659		−5.4	−2.5	−5.3
	(−)	(0.03234)				
ESHLII	13.6554	1		−251.1	−248.2	−251.0
	(1.2070)	(−)				
	$\widehat{a}$	$\widehat{b}$	$\widehat{c}$			
MCSHL	13.915	58.358	0.614	−388.1	−379.5	−387.9
	(2.781)	(0.682)	(125.480)			
KWSHL	1314.13	1	2.8298	−385.7	−380.0	−385.6
	(18.9075)	(−)	(0.03691)			

CHAPTER 5

General conclusions and future works

Based on the present doctoral thesis, follows the general conclusions:

- The *exponentiated generalized* (EG) class of distributions with two extra shape parameters $a > 0$ and $b > 0$ constitutes a simple way of adding two parameters to a continuous distribution.

- The EG family of distributions has desirable properties such as: they have no complicated functions and will be always tractable when the cdf and pdf of the baseline distribution have simple analytical expressions; contain as special cases the two classes of Lehmann's alternatives and then, the EG family can be derived from a double transformation using these classes; the two extra parameters a and b in the EG density can control both tail weights, allowing generating flexible distributions, with heavier or lighter tails, as appropriate and among others.

- The EG hazard function can take the classic shapes: bathtub, inverted bathtub, increasing, decreasing and constant, among others. These shapes are important for reliability studies, survival analysis, among others.

- The sub-models proposed and studied in this thesis, say the *exponentiated generalized Gumbel* ("EGGu" for short), the *exponentiated generalized extended expo-*

nential (EGEE) and *exponentiated generalized standard half-logistic* (EGSHL) distributions, are mathematically and computationally manageable. Moreover, they have shown to be competitive in terms of adjustment to real data, when compared to several classical models well-established in the literature. In summary, the models studied in this thesis have proved to be important and can be widely used by applied researchers.

Due to the recognized importance of the EG family of distributions, we believe that more needs to be done in the study of their properties and sub-models. Thus, an additional contribution of the present thesis is the proposition of new researches related to this class. Specifically, we propose an exploration of the sub-models:

- The *exponentiated generalized generalized Gompertz* (EGEG) distribution:

$$F(x) = \left\{1 - \left[1 - \left(1 - \mathrm{e}^{-\frac{\beta}{\gamma}(\mathrm{e}^{\gamma x}-1)}\right)^{\theta}\right]^{a}\right\}^{b},$$

  where $a > 0$, $b > 0$, $\beta > 0$, $\gamma \geq 0$, $\theta > 0$ and $x \geq 0$.

- The *exponentiated generalized generalized power Weibull* (EGGPW) distribution:

$$F(x) = \{1 - \exp[\{1 - [1 + (x/\beta)^{\alpha}]^{\theta}\}^{a}]\}^{b},$$

  where $a > 0$, $b > 0$, $\beta > 0$, $\alpha > 0$, $\theta > 0$ and $x \geq 0$.

- The *Odd log-logistic Generalized Exponentiated Gumbel* (OLLEGGu) distribution:

$$F(x;\mu,\sigma,a,b,\alpha) = \frac{\left\{1 - [1 - \mathrm{e}^{-\mathrm{u}}]^{a}\right\}^{b\alpha}}{\left\{1 - [1 - \mathrm{e}^{-\mathrm{u}}]^{a}\right\}^{b\alpha} + \left\{1 - \left\{1 - [1 - \mathrm{e}^{-\mathrm{u}}]^{a}\right\}^{b}\right\}^{\alpha}},$$

  where $u = e^{-(x-\mu)/\sigma}$, $a, b, \alpha, \sigma > 0$ and $\mu, x \in I\!R$.

- The *Exponentiated Generalized Marshall–Olkin* (EGMO-G) family of distributions:

$$F(x) = \left\{1 - \left[\frac{\alpha \bar{G}(x)}{1 - \bar{\alpha}\bar{G}(x)}\right]^{a}\right\}^{b},$$

  in which the dependence on the parameters of $G(x)$ are implicit.

Although it has not been detailed here, all the above models are being studied and the future submission of related works will refer to the present thesis.

References

[1] Gauss M. Cordeiro, Artur J. Lemonte, and Edwin M.M. Ortega. An extended fatigue life distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 47:626–653, 2013.

[2] Gauss M. Cordeiro, Elizabeth M. Hashimoto, and Edwin M.M. Ortega. The McDonald Weibull model. *Statistics: A Journal of Theoretical and Applied Statistics*, 48:256–278, 2014.

[3] Gauss M. Cordeiro and Artur J. Lemonte. The exponentiated generalized Birnbaum-Saunders distribution. *Applied Mathematics and Computation*, 247:762–779, 2014.

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

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

[6] K. Zografos and N. Balakrishnan. On families of Beta- and generalized Gamma-generated distributions and associated inference. *Statistical Methodology*, 6:344–362, 2009.

[7] M. M. Ristic and N. Balakrishnan. The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation, doi: 10.1080/00949655.2011.574633*, 2011.

[8] Saralees Nadarajah, Gauss M. Cordeiro, and Edwin M. M. Ortega. The Zografos-Balakrishnan-G Family of Distributions: Mathematical Properties and Applications. *Communications in Statistics - Theory and Methods*, 44:186–215, 2015.

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

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

[11] Y. M. Gómez, H. Bolfarine, and H. W. Gómez. A new extension of the exponential distribution. *Revista Colombiana de Estadística*, 37:25–34, 2014.

[12] S. Kotz and S. Nadarajah. *Extreme value distributions: theory and applications*. Imperial College Press, London, 2000.

[13] S. Nadarajah. The exponentiated Gumbel distribution with climate application. *Environmetrics*, 17:13–23, 2006.

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

[15] S. Nadarajah and S. Kotz. The beta Gumbel distribution. *Mathematical Problems in Engineering*, 4:323–332, 2004.

[16] Muhammad H. Tahir and Saralees Nadarajah. Parameter induction in continuous univariate distributions: Well-established G families. *Anais da Academia Brasileira de Ciências*, 87:539–568, 2015.

[17] E. L. Lehmann. The power of rank tests. *Annals of Mathematical Statistics*, 24:23–43, 1953.

[18] R. D. Gupta and D. Kundu. Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal*, 43:117–130, 2001.

[19] S. Nadarajah and S. Kotz. The exponentiated Fréchet distribution. *Interstat. Available online from interstat.statjournals.net/YEAR/2003/articles/0312001.pdf*, 2003.

[20] D. T. Shirke and C. S. Kakde. On exponentiated lognormal distribution. *International Journal of Agricultural and Statistical Sciences*, 2:319–326, 2006.

[21] S. Nadarajah and A. K. Gupta. The exponentiated Gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin*, 59:29–54, 2007.

[22] J. M. F. Carrasco, E. M. M. Ortega, and G. M. Cordeiro. A generalized modified Weibull distribution for lifetime modeling. *Computational Statistics and Data Analysis*, 53:450–462, 2008.

[23] G. M. Cordeiro, E. M. M. Ortega, and G. O. Silva. The exponentiated generalized Gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, 81:827–842, 2011.

[24] L. C. Rêgo, R. J. Cintra, and G. M. Cordeiro. On Some Properties of the Beta Normal Distribution. *Communications in Statistics - Theory and Methods*, 41:3722–3738, 2012.

[25] L. M. Zea, R. B. Silva, M. Bourguignon, A. M. Santos, and G. M. Cordeiro. The Beta Exponentiated Pareto Distribution with Application to Bladder Cancer Susceptibility. *International Journal of Statistics and Probability*, 1:8–19, 2012.

[26] M. W. Ramos, P. R. Marinho, R. V. Silva, and G. M. Cordeiro. The Exponentiated Lomax Poisson Distribution with an Application to lifetime data. *Advances and Applications in Statistics*, 34:107–135, 2013.

[27] J. F. Kenney and E. S. Keeping. *Mathematics of Statistics*. 3rd ed., Part 1, New Jersey, 1962.

[28] J. J. Moors. A quantile alternative for kurtosis. *J. Royal Statist. Soc. D*, 37:25–32, 1988.

[29] A. A. Jafari, S. Tahmaseb, and M. Alizadeh. The beta–gompertz distribution. *Revista Colombiana de Estadística*, 37:139–156, 2014.

[30] A. A. Jafari and E. Mahmoudi. Beta-linear failure rate distribution and its applications. *Journal of The Iranian Statistical Society*, 14:89–105, 2015.

[31] Gauss M. Cordeiro and Artur J. Lemonte. The β-Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, 55:1445–1461, 2011.

[32] Shujiao Huang and Broderick O Oluyede. Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data. *Journal of Statistical Distributions and Applications*, 1:1–20, 2014.

[33] Artur J. Lemonte. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, 62:149–170, 2013.

[34] Saralees Nadarajah and Firoozeh Haghighi. An extension of the exponential distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 45:543–558, 2011.

[35] Saralees Nadarajah, Hassan S. Bakouch, and Rasool Tahmasbi. A generalized Lindley distribution. *Sankhya B*, 73:331–359, 2011.

[36] P. E. Oguntunde, A. O. Adejumo, and O. S. Balogun. Statistical properties of the exponentiated generalized inverted exponential distribution. *Applied Mathematics*, 4:47–55, 2014.

[37] Gauss M. Cordeiro, Antonio E. Gomes, and Cibele Q. da-Silva. Another extended Burr III model: some properties and applications. *Journal of Statistical Computation and Simulation*, 84:2524–2544, 2014.

[38] D. Hinkley. On quick choice of power transformations. *The American Statistician*, 26:67–69, 1977.

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

[40] D.N.P. Murthy, M. Xie, and R. Jiang. *Weibull models.Wiley series in probability and statistics*. Hoboken (NJ): JohnWiley and Sons, 2004.

[41] Eisa Mahmoudi. The beta generalized Pareto distribution with application to lifetime data. *Mathematics and Computers in Simulation*, 81:2414–2430, 2011.

[42] S. Nadarajah and S. Eljabri. The Kumaraswamy GP Distribution. *Journal of Data Science*, 11:739–766, 2013.

[43] O. G. Silva, E. M. M. Ortega, and G. Cordeiro. The beta modified weibull distribution. *Lifetime Data Anal*, 16:409–430, 2010.

[44] Artur J. Lemonte. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, 62:149–170, 2013.

[45] Ramesh C. Gupta, Pushpa L. Gupta, and Rameshwar D. Gupta. Modeling failure time data by Lehmann alternatives. *Communications in Statistics—Theory and Methods*, 27:887–904, 1998.

[46] Saralees Nadarajah and Firoozeh Haghighi. An extension of the exponential distribution. *Statistics: A Journal of Theoretical and Applied Statistics*, 45:543–558, 2011.

[47] H. Zakerzadeh and A. Dolati. Generalized Lindley distribution. *Journal of Mathematical Extension*, 3:13–25, 2009.

[48] Gauss M. Cordeiro, Antonio E. Gomes, and Cibele Q. da-Silva. Another extended Burr III model: some properties and applications. *Journal of Statistical Computation and Simulation*, 84:2524–2544, 2014.

[49] I. Elbatal and Hiba Z. Muhammed. Exponentiated generalized inverse Weibull distribution. *Applied Mathematical Sciences*, 8:3997–4012, 2014.

[50] P. E. Oguntunde, A. O. Adejumo, and O. S. Balogun. Statistical properties of the exponentiated generalized inverted exponential distribution. *Applied Mathematics*, 4:47–55, 2014.

[51] Ronaldo V. da Silva, Frank Gomes-Silva, Manoel Wallace A. Ramos, and Gauss M. Cordeiro. A new extended gamma generalized model. *International Journal of Pure and Applied Mathematics*, 100:309–335, 2015.

[52] Saralees Nadarajah, Hassan S. Bakouch, and Rasool Tahmasbi. A generalized Lindley distribution. *Sankhya B*, 73:331–359, 2011.

[53] Gauss M. Cordeiro and Artur J. Lemonte. The exponentiated generalized BirnbaumŰ-Saunders distribution. *Applied Mathematics and Computation*, 247:762–779, 2014.

[54] A. O. Silva, L. C. M. da Silva, and G. M. Cordeiro. The extended dagum distribution: Properties and application. *Journal of Data Science*, 13:53–72, 2015.

[55] T. A. N. De Andrade, H. Rodrigues, M. Bourguignon, and G. M. Cordeiro. The exponentiated generalized gumbel distribution. *Revista Colombiana de Estadística*, 38:123–143, 2015.

[56] M. Mansoor, M. H. Tahir, A. Alzaatreh, G. M. Cordeiro, M. Zubair, and S. S. A. Ghazali. An extended frechet distribution: Properties and applications. *Journal of Data Science*, 14:167–188, 2016.

[57] Faton Merovci and Vikas K. Sharma. The beta-Lindley distribution: properties and applications. *Journal of Applied Mathematics*, 2014:1–10, 2014.

[58] Gauss M. Cordeiro and Artur J. Lemonte. The β-Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, 55:1445–1461, 2011.

[59] Shujiao Huang and Broderick O Oluyede. Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data. *Journal of Statistical Distributions and Applications*, 1:1–20, 2014.

[60] N. Balakrishnan and K. H. T. Wong. Approximate mles for the location and scale parameters of the half–logistic distribution with type–II right–censoring. *IEEE Transactions on Reliability*, 40:140–145, 1991.

[61] N. Balakrishnan and P.S. Chan. Estimation for the scaled half logistic distribution under type II censoring. *Computional Statistics & Data Analysis*, 13:123–141, 1992.

[62] W. Panichkitkosolkul and Saothayanun. Bootstrap confidence intervals for the process capability index under half–logistic distribution. *Maejo International Journal of Science and Technology*, 6:272–281, 2012.

[63] J. Oliveira, J. Santos, C. Xavier, D Trindade, and G. M. Cordeiro. The McDonald half–logistic distribution: theory and practice. *Communications in Statistics – Theory and Methods*, 45(7):2005–2022, 2016.

[64] G. M. Cordeiro, M. Alizadeh, and P. R. D. Marinho. The type I half–logistic family of distributions. *Journal of Statistical Computation and Simulation*, (10.1080/00949655.2015.1031233), 2015.

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

[66] I. Elbatal and Hiba Z. Muhammed. Exponentiated generalized inverse Weibull distribution. *Applied Mathematical Sciences*, 8:3997–4012, 2014.

[67] Ronaldo V. da Silva, Frank Gomes-Silva, Manoel Wallace A. Ramos, and Gauss M. Cordeiro. A new extended gamma generalized model. *International Journal of Pure and Applied Mathematics*, 100:309–335, 2015.

[68] T. A. N. De Andrade, M. Bourguignon, and Gauss M. Cordeiro. The exponentiated generalized extended exponential distribution. *Journal of Data Science*, 14:393–414, 2016.

[69] G. Arya and I. Elbata. On the exponentiated generalized modified weibull distribution. *Communications for Statistical Applications and Methods*, 22:333–348, 2015.

[70] Elisa T. Lee. *STATISTICAL METHODS FOR SURVIVAL DATA ANALYSIS*. Wiley & Sons, Inc, 1992.

[71] Ammar M. Sarhan and Debasis Kundu. Generalized linear failure rate distribution. *Communications in Statistics - Theory and Methods*, 38:642–660, 2009.

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

[73] A. M. Alshangiti, M. Kayid, and B. Alarfaj. A new family of marshall–olkin extended distributions. *Journal of Computational and Applied Mathematics*, 271:369–379, 2014.

[74] M. A. C. Castellares, F. and. Santos, L. C. Montenegro, and G. M. Cordeiro. A gamma–generated logistic distribution: properties and inference. *American Journal of Mathematical and Management Sciences*, 34:14–39, 2015.

[75] N. Balakrishnan and A. C. Cohen. *Order Statistics and Inference: Estimation Methods*, volume 2014. ACADEMIC PRESS, 2014.

[76] W. Barreto-Souza, A. L. de Morais, and G. M. Cordeiro. The weibull–geometric distribution. *Journal of Statistical Computation and Simulation*, (10.1080/00949650903436554), 2010.