Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

HELOISA DE MELO RODRIGUES

SOME EXTENDED ALPHA MODELS: PROPERTIES AND APPLICATIONS

Recife
2017

Heloisa de Melo Rodrigues

SOME EXTENDED ALPHA MODELS: PROPERTIES AND APPLICATIONS

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

Orientadora: Professora Dra. Audrey Helen Mariz de A. Cysneiros

Recife
2017

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

R696s Rodrigues, Heloisa de Melo
Some extended Alpha models: properties and applications / Heloisa de Melo Rodrigues. – 2017.
117 f.: il., fig., tab.

Orientadora: Audrey Helen Mariz de A. Cysneiros.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências.

1. Análise de regressão. 2. Máxima verossimilhança. I. Cysneiros, Audrey Helen Mariz de A. (orientadora). II. Título.

519.536 CDD (23. ed.) UFPE- MEI 2017-98

HELOISA DE MELO RODRIGUES

SOME EXTENDED ALPHA MODELS: PROPERTIES AND APPLICATIONS

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

Aprovada em: 20 de fevereiro de 2017.

BANCA EXAMINADORA

Prof.ª Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof. Abraão David Costa do Nascimento
UFPE

Prof.ª Maria do Carmo Soares de Lima
UFPE

Prof. Rodrigo Bernardo da Silva
UFPB

Prof. Roberto Ferreira Manghi
UFC

Dedico esta tese aos meus pais Fátima e Paulo Cezar e ao meu marido Felipe.

Agradecimentos

Aos meus pais Fátima e Paulo Cezar, por sempre acreditarem em mim e por me dar muito amor e carinho, essencial para enfrentar qualquer desafio que a vida impõe.

Aos meus avós Eremita (*in memoriam*) e José Rodrigues (*in memoriam*).

Ao meu esposo Felipe, pelo amor, incentivo, companheirismo e compreensão nos momentos de tensão que passei no doutorado.

À professora Audrey Cysneiros, pela orientação e confiança depositada em mim.

À professora Maria do Carmo (Mel), pelo apoio imprescindível para a elaboração desta tese.

Ao professor Gauss Cordeiro pela colaboração neste trabalho e na minha formação acadêmica.

Aos membros da banca por terem aceitado o convite e pelas contribuições valiosas.

Aos meus tios Ana, Anália e Manoel e aos meus primos Luciana, Luan, Marília, Marquinhos, Letícia e Antônio por sempre estarem presente em minha vida, acompanhando de perto minha trajetória.

À minha sogra Socorro, contemporânea de doutorado, Rafael, Natália, tia Ró, Fernanda e Júlio, pelo convívio e torcida. Um agradecimento especial à vida dos meus sobrinhos Arthur e Bernardo, que estão sempre me proporcionando momentos de descontração e afeto.

À UFPE, por ter sido minha segunda casa ao longo de onze anos.

Aos amigos e colegas de graduação, especialmente Simara, Nathália, Samuel e Hélio.

Aos colegas de doutorado, em especial Thiago Andrade, Gustavo e Jéssica.

Aos professores do Departamento de Estatística da UFPE Alex, Andre Toom, Carla Monteiro, Claudia Lima, Cristina Raposo, Cristiano Ferraz, Francisco Cribari, Francisco Cysneiros, Getúlio Amaral, Isaac Xavier, Leandro Rêgo, Raydonal, Renato Cintra e Sylvio. Obrigada pela contribuição na minha formação acadêmica, tanto na graduação quanto no mestrado e doutorado.

À Valéria Bittencourt, pela simpatia, ombro amigo e disponibilidade de sempre. Meu agradecimento também aos demais funcionarios da graduação e pós-graduação com os quais convivi no meu tempo de UFPE.

"Falar sem aspas, amar sem interrogação, sonhar com reticências, viver sem ponto final."

Charles Chaplin

Abstract

Generalizing distributions provide some advantages, allowing us to define new families, to extend well-known distributions and provide great flexibility in modeling real data, which can be applied in several fields. The Alpha distribution was studied for the first time to analyze tool wear problems by Katsev (1968) and Wager and Barash (1971). Salvia (1985) provided its characterization. In this thesis, we discuss the Alpha distribution, we present a simulation study to verify the performance of its maximum likelihood estimators and four real data sets are used to evaluate the Alpha model when compared to some distributions well-known in literature. Furthermore, we developed new distributions considering this model as the baseline distribution applied to Exponentiated class (Gompertz, 1825; Verhulst, 1838, 1845, 1847) and Kumaraswamy class, proposed by Cordeiro and de Castro (2011). We also propose a new family of distributions, called Exponentiated Generalized Exponentiated-Generated (EG-Exp-G), which is an extension of the exponentiated generalized class proposed by Cordeiro et al. (2013). Some new distributions are proposed as submodels of this family, including the EG-Exp-Alpha distribution. We study some mathematical properties, such as quantile function, moments, moment generating function, mean deviations and order statistics. In addition, we use the maximum likelihood method to estimate the parameters of the proposed models. We perform Monte Carlo simulation studies to analyze the asymptotic properties of the maximum likelihood estimators and we illustrate the flexibility of the new models through applications to real data set in order to show their competitiveness compared to well-known distributions in the literature.

Keywords: Alpha Distribution. Exponentiated class. Exponentiated Generalized class. Kumaraswamy class. Maximum Likelihood Estimation.

Resumo

A generalização de distribuições oferece algumas vantagens, permitindo-nos definir novas famílias, estender distribuições conhecidas e proporcionar grande flexibilidade na modelagem de dados reais, que podem ser aplicados em vários campos. A distribuição Alpha foi estudada inicialmente para analisar problemas de desgaste de ferramentas por Katsev (1968) e Wager e Barash (1971). Salvia (1985) forneceu algumas características desta distribuição. Nesta tese, discutimos a distribuição Alpha, apresentamos um estudo de simulação para verificar a performance dos seus estimadores de máxima verssimilhança e quatro conjuntos de dados reais são utilizados para avaliar o modelo Alpha em relação à algumas distribuições de probabilidade já conhecidas na literatura. Além disso, desenvolvemos novas distribuições considerando tal modelo como distribuição de base aplicada aos geradores da Exponencializada (Gompertz, 1825; Verhulst, 1838, 1845, 1847) e da Kumaraswamy, proposta por Cordeiro e de Castro (2011). Propomos ainda uma nova família de distribuições, chamada exponencializada generalizada exponencializada (EG-Exp-G), que é uma extensão da classe exponencializada generalizada proposta por Cordeiro et al. (2013). Apresentamos alguns casos especiais deste novo gerador, entre eles a distribuição EG-Exp-Alpha. Desenvolvemos algumas propriedades matemáticas, a saber: desvios médios, estatísticas de ordem, função geratriz de momentos, função quantílica e momentos. Além disso, utilizamos o método de máxima verossimilhança para estimação dos parâmetros dos modelos propostos. Realizamos estudos de simulação de Monte Carlo visando analisar as propriedades assintóticas dos estimadores de máxima verossimilhança e ilustramos a flexibilidade dos novos modelos por meio de aplicações a dados reais a fim de mostrar a competitividade deles comparados às distribuições de probabilidade já conhecidas na literatura.

Palavras chave: Distribuição Alpha. Exponencializada-G. Exponencializada Generalizada-G. Kumaraswamy-G. Máxima verossimilhança.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

The Alpha distribution was studied for the first time to analyze tool wear problems by Katsev (1968) and Wager and Barash (1971). Sherif (1983) suggested its use in modeling lifetimes under accelerated test conditions and Salvia (1985) provided its characterization and a number of structural properties. This thesis has as its initial aims to discuss the Alpha distribution, to perform a simulation study and to compare this model with other already exist and, in this way, to show that the proposed distribution competes with other distributions very used for survival data, such as: Birnbaum-Saunders (BS), Burr XII, Gamma, Inverse Gaussian, Nadarajah-Haghighi (NH), Weibull, among others.

A random variable X has an Alpha distribution with shape parameter ($\alpha > 0$) and scale parameter ($\beta > 0$), if its cumulative distribution function (cdf), say $G(x)$, and probability density function (pdf), say $g(x)$, are given by

$$G(x;\alpha,\beta) = \frac{\Phi\left(\alpha - \frac{\beta}{x}\right)}{\Phi(\alpha)} \tag{1.1}$$

and

$$g(x;\alpha,\beta) = \frac{\beta}{\sqrt{2\pi}\, x^2\, \Phi(\alpha)} \exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\},$$

respectively, for $x > 0$, where $\Phi(\cdot)$ denote the standard normal cumulative function. The Alpha distribution is characterized by non existence of moments. So, some results derived in this thesis are valid for certain regions of the parametric space.

Furthermore, we extend the Alpha distribution, providing great flexibility in modeling real data, which can be applied in several fields. Generalized distributions provide some advantages, allowing us to define new families of distributions and to extend well-known distributions. For example, we can refer some papers: Eugene et al. (2002) for the *beta* class, Zografos and Balakrishnan (2009) for the *gamma* class and the more recent ones by Alexander et al. (2012) and Cordeiro et al. (2013), who defined the class generated by McDonald's (1984) generalized beta random variable and the class of *exponentiated generalized* distributions, respectively.

In this thesis, we propose to apply the Alpha distribution as baseline to the Exponentiated-G (Exp-G) and Kumaraswamy (Kw-G) families. In the first case, for an arbitrary baseline $G(x)$, a random variable has the Exp-G distribution with additional power parameter ($a > 0$) if its cdf and pdf are given by

$$F(x; a) = G(x)^a$$

and

$$f(x; a) = a\, g(x)\, G(x)^{a-1},$$

respectively, where $g(x) = dG(x)/dx$.

In the second one, for any continuous baseline G distribution, Cordeiro and de Castro (2011) proposed the Kw-G family of distributions with two additional shape parameters $a > 0$ and $b > 0$ and cdf and pdf given by

$$F(x; a, b) = 1 - [1 - G^a(x)]^b$$

and

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

respectively.

Besides, we propose a new family of distributions, called Exponentiated Generalized Exponentiated-Generated (EG-Exp-G, for short), which adds three shape parameters ($a, b, c > 0$) to the baseline distribution. This family is an extension of the exponentiated generalized class proposed by Cordeiro et al. (2013) and has cdf and pdf given by

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

and

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

respectively.

This thesis is organized as follows. In second chapter, we present the Alpha distribution and we provide a simulation study to evaluate the asymptotic properties of maximum likelihood estimates. In addition, we show the flexibility of the Alpha distribution in modeling data through of four applications to real data sets.

In third chapter, we propose a three-parameter model, called Exponentiated Alpha distribution (Exp-Alpha, for short). The new model is constructed using the exponentiated class (Gompertz, 1825; Verhulst, 1838, 1845, 1847), which is widely applied to baseline distributions in applications to real data set, because this class provide a greater flexibility. Furthermore, we provide plots of pdf and hazard rate function (hrf) and study some mathematical properties of the Exp-Alpha distribution, such as: moment generating function, mean deviations, moments, order statistics and quantile function. It is important to emphasize that the results presented are valid for certain regions of the parametric space. The estimation of the model parameters is discussed by maximum likelihood method. Besides, we provide a semiclosed estimator for the additional parameter. A Monte Carlo simulation study is performed to evaluate the maximum likelihood estimators of the proposed model. Furthermore, three applications to real data sets are provided in order to illustrate the flexibility of the Exp-Alpha distribution, when compared to well-known distributions, such as Exp-BS, Exp-Gompertz, Exp-Lomax, Exp-Weibull, among others.

In Chapter 4, we define a four-parameter model, called the Kumaraswamy Alpha (Kw-Alpha, for short) distribution, which is obtained combining results from Cordeiro and de Castro (2011) and Equation (1.1). We provide plots of pdf and hrf and obtain explicit expressions for the quantile function, ordinary moments, moment generating function and mean deviations. Further, we determine the density function of the order statistics. These results only are valid for some regions of the parametric space due to non existence of moments for the Alpha distribution. The maximum likelihood method is used to estimate the model parameters and we evaluate this procedure through a Monte Carlo simulation study. Besides, we found a semiclosed estimator for one of the extra parameters. We provide two applications to real data to illustrate the flexibility of the new model.

In the fifth chapter, we proposed the EG-Exp-G family and some new distributions are provided as submodels of this class, including the EG-Exp-Alpha. Some mathematical properties of this family are derived as mean deviations, moment generating function, order statistics, ordinary and incomplete moments, quantile function and Rényi and Shannon entropies. Estimation of model parameters are discussed by maximum likelihood method and a semiclosed estimator for one of the additional parameters is obtained. We provide a simulation study taking the EG-Exp-Alpha model as an example to evaluate the performance of the maximum likelihood estimates and four applications to real data is performed to illustrate the flexibility of the new family. Finally, the general conclusions are provided in Chapter 6.

Chapter 2

The Alpha distribution

Resumo

A distribuição Alpha foi primeiramente estudada por Katsev (1968), Wager e Barash (1971), Sherif (1983) e Salvia (1985). Porém, estudos de simulação de tal distribuição não foram estudados e o seu uso para modelar dados. Portanto, o objetivo deste capítulo é discutir a distribuição Alpha e propor seu uso à modelagem de conjuntos de dados em várias áreas. O método de máxima verossimilhança é utilizado para estimar os parâmetros do modelo e um estimador em forma semifechada para o parâmetro de escala é obtido. Nós avaliamos a performance dos estimadores de máxima verossimilhança por meio de um estudo de simulação de Monte Carlo e apresentamos quatro aplicações a conjuntos de dados reais, comparando o modelo proposto com outras distribuições já utilizadas em diferentes áreas.

Palavras-chave: Desvios médios. Distribuição Alpha. Estatísticas de ordem. Função quantílica. Máxima verossimilhança. Momentos.

Abstract

The Alpha distribution was studied previouly by Katsev (1968), Wager and Barash (1971), Sherif (1983) and Salvia (1985). But they do not perform simulation study of the Alpha distribution and its use for modeling data. Hence, the aim of this chapter is to discuss the Alpha distribution and to propose its use for modeling data set in various areas. Maximum likelihood method is used to estimate the model parameters and a semiclosed estimator for the scale parameter is obtained. We evaluate the performance of the maximum likelihood estimators by a Monte Carlo simulation study and we present four applications to real data sets, comparing the proposed model with other distributions already used in different areas.

Keywords: Alpha distribution. Maximum likelihood Estimation. Mean deviations. Moments. Order statistics. Quantile function.

2.1 Introduction

The Alpha distribution was studied for the first time to analyze tool wear problems by Katsev (1968) and Wager and Barash (1971). Sherif (1983) suggested its use in modeling lifetimes under accelerated test conditions and Salvia (1985) provided its characterization and a number of structural properties. The aim of this chapter is to discuss the Alpha distribution, to perform a simulation study and to compare this model with other already exist and, in this way, to show that the Alpha distribution competes with other distributions very used for survival data. In summary, the purpose of this chapter is to realize a numerical and applied study of the Alpha distribution discussed by Salvia (1985).

One of the main results of the Alpha distribution is that the mean does not exist, but this fact does not prevent its use as a model for accelerated life testing (as, for example, the Cauchy and certain Pareto models). As we will see in the applications, the Alpha model is skewed to the right, but its mode is finite. It is important to emphasize that the results presented here are valid for certain regions of the parametric space.

The rest of this chapter is organized as follows. The Alpha distribution is present in Section 2.2. Maximum likelihood estimation of the model parameters is investigated in Section 2.3. A simulation study and four applications to real data sets in Section 2.4 illustrate the flexibility of the Alpha distribution for data modeling. Concluding remarks are given in Section 2.5.

2.2 The Alpha distribution

A random variable X has an Alpha distribution with shape $\alpha > 0$ and scale $\beta > 0$ parameters, if its cumulative distribution function (cdf) and probability density function (pdf) are given by

$$G(x;\alpha,\beta) = \frac{\Phi\left(\alpha - \frac{\beta}{x}\right)}{\Phi(\alpha)} \tag{2.1}$$

and

$$g(x;\alpha,\beta) = \frac{\beta}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)}\,\exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\}, \tag{2.2}$$

respectively, for $x > 0$, where $\Phi(\cdot)$ denote the standard normal cumulative function.

Henceforth, a random variable X with density function (2.2) is denoted by $X \sim \text{Alpha}(\alpha,\beta)$. We write $G(x) = G(x;\alpha,\beta)$ and $g(x) = g(x;\alpha,\beta)$ in order to eliminate the dependence on the model parameters.

The hazard rate function (hrf) corresponding to (2.2) is given by

$$h(x;\alpha,\beta) = \frac{\beta}{\sqrt{2\pi}\,x^2[\Phi(\alpha) - \Phi(\alpha - \frac{\beta}{x})]}\,\exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\}.$$

Plots of the Alpha density function for selected parameter values are displayed in Figure 2.1. Figure 2.2 shows that the hrf of the Alpha distribution has increasing and decreasing shapes for different values of the parameters.

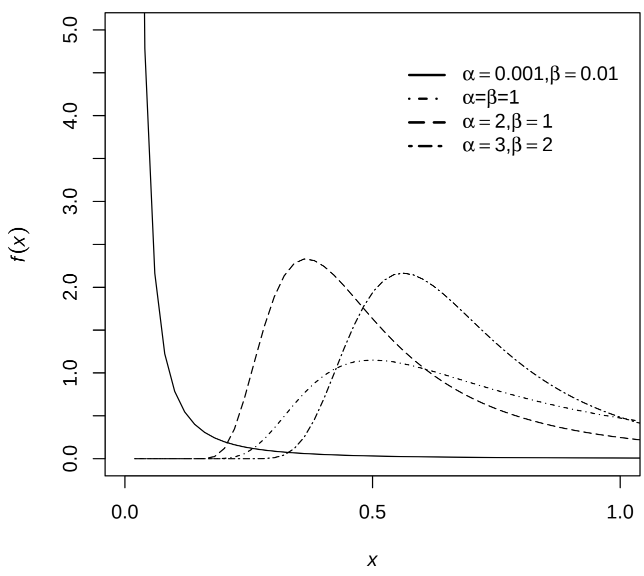


Figure 2.1: Plots of the Alpha density for some parameter values.

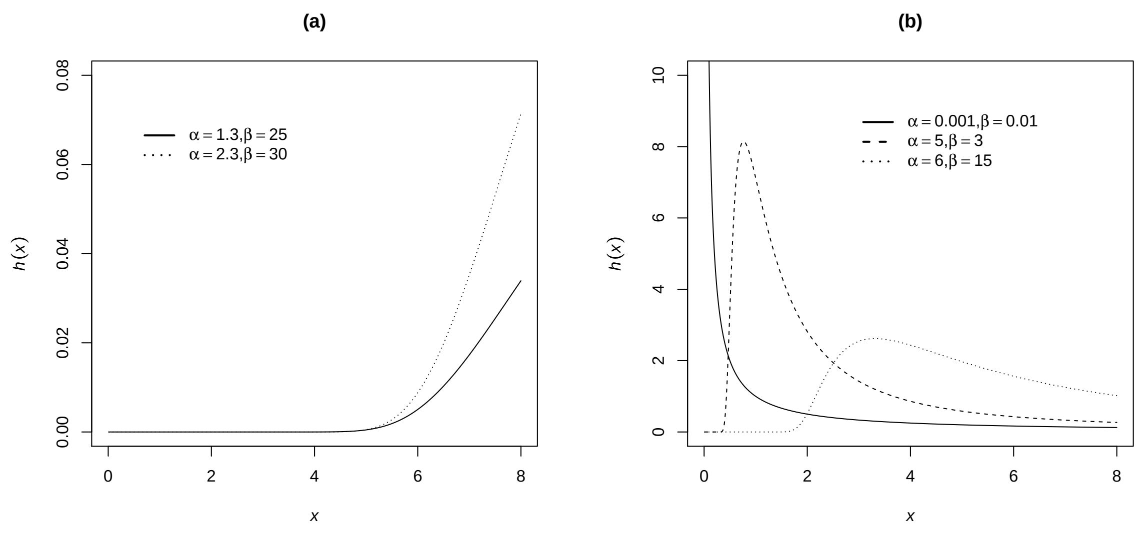


Figure 2.2: Plots of the Alpha hazard rate function for some parameter values. (a) Increasing hazard rate function. (b) Decreasing hazard rate function.

2.3 Estimation

In this section, we provide an analytic procedure to obtain the maximum likelihood estimators (MLEs) for the Alpha parameters using the profile log-likelihood function. The MLEs are employed due to their interesting asymptotic properties.

Let $(x_1,\dots,x_n)$ be a random sample of size n from the Alpha distribution with unknown parameter vector $\boldsymbol{\theta}=(\alpha,\beta)^\top$. The log-likelihood function for $\boldsymbol{\theta}$ based on a given sample can be expressed as

$$\ell(\boldsymbol{\theta}) = n\,\log\left(\frac{\beta}{\sqrt{2\pi}\Phi(\alpha)}\right) - 2\,v - \frac{n\alpha^2}{2} + \alpha\beta t - \frac{\beta^2\,w}{2}, \tag{2.3}$$

where $t=\sum_{i=1}^{n}\frac{1}{x_i}$, $w=\sum_{i=1}^{n}\frac{1}{x_i^2}$ and $v=\sum_{i=1}^{n}\log(x_i)$.

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

$$U_\alpha(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\alpha} = -n\,\alpha - n\,\frac{\phi(\alpha)}{\Phi(\alpha)} + \beta\,t,$$

$$U_\beta(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta} = \frac{n}{\beta} + \alpha\,t - \beta\,w. \tag{2.4}$$

The MLEs, $\hat{\boldsymbol{\theta}}=(\hat\alpha,\hat\beta)^\top$, can be obtained numerically by solving $U(\hat{\boldsymbol{\theta}})=0$.

From equation (2.4), we note that it is possible to obtain a semiclosed estimator for β. For $\frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta}=0$, the estimator for β is given by

$$\hat\beta_{\hat\alpha} = \frac{\hat\alpha\,t + \sqrt{(\hat\alpha\,t)^2 + 4\,w\,n}}{2\,w}. \tag{2.5}$$

The estimative of β can be obtained analytically by (2.5).

We can obtain the MLE of α using the profile log-likelihood. We suppose β fixed and rewrite the log-likelihood (2.3) as $\ell(\boldsymbol{\theta})=\ell_\beta(\alpha)$. By replacing β by $\hat\beta_\alpha$ in (2.3), we can obtain the profile log-likelihood function for α as

$$\ell_\beta(\alpha) = -\frac{n}{2}(\alpha^2-1) - 2\,v + n\log\left(\frac{\alpha\,t + \sqrt{(\alpha\,t)^2+4\,w\,n}}{2\,w\sqrt{2\pi}\Phi(\alpha)}\right) + \frac{(\alpha\,t)^2 + \alpha\,t\sqrt{(\alpha\,t)^2+4\,w\,n}}{4\,w}. \tag{2.6}$$

The MLE of α can be also determined by maximizing the profile log-likelihood function (2.6) with respect to α.

It should be noted that the maximization of (2.6) is more simpler than the maximization of (2.3) with respect to their respective parameters, since the profile log-likelihood function has one parameter, whereas the log-likelihood function has two parameters. But the profile log-likelihood is not a real likelihood function and hence some of the new properties that hold for a true likelihood do not hold for its profile version.

For interval estimation and hypothesis tests on the model parameters, we require the 2×2 observed information matrix $J = J(\boldsymbol{\theta})$, where its elements for the parameters (α, β) are given by

$$J_{\alpha,\alpha} = -\,n\left\{1 + \frac{[\dot{\phi}(\alpha)]_{\alpha}\,\Phi(\alpha) - \phi(\alpha)^2}{\Phi(\alpha)^2}\right\}, \quad J_{\alpha,\beta} = \sum_{i=1}^{n}\frac{1}{x_i}, \quad J_{\beta,\beta} = -\frac{n}{\beta^2} - \sum_{i=1}^{n}\frac{1}{x_i^2},$$

with $[\dot{\phi}(\alpha)]_{\alpha} = \frac{\partial\,\phi(\alpha)}{\partial\,\alpha}$ and $\phi(\alpha) = \frac{\partial\,\Phi(\alpha)}{\partial\,\alpha}$.

Under general regularity conditions, $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal $N_2(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. In practice, we can substitute $I(\boldsymbol{\theta})$ by the observed information matrix evaluated at $\hat{\boldsymbol{\theta}}$, say $J(\hat{\boldsymbol{\theta}})$. The asymptotic multivariate normal $N_2(0, J(\hat{\boldsymbol{\theta}})^{-1})$ distribution of $\hat{\boldsymbol{\theta}}$ can be used to construct approximate confidence intervals for the parameters.

2.4 Simulation study and applications

In this section, we perform a Monte Carlo simulation study and we provide four applications to real data to illustrate the flexibility of the Alpha distribution.

2.4.1 Simulation study

Here, a Monte Carlo simulation experiment based on $1,000$ replications will be conduced to evaluate the MLEs of the parameters of the Alpha distribution. Setting $\alpha = (1, 2)$ and $\beta = (1, 2)$, we evaluate the relative bias (RB) and the mean squared errors (MSE) for point estimators based on artificially generated samples of sizes $n = 20, 40, 60, 80$ and 100. The RB is defined by 100×(bias/real value of the parameter)%. The results are given in Table 2.1 and we note that, as the sample size increases, the empirical biases and mean squared errors decrease, as expected.

2.4.2 Applications to real data

In this section, we provide four applications to real data sets to illustrate the flexibility of the Alpha distribution, with pdf given by (2.2). Here, we compare the Alpha model with competitive distributions well-known to fit real data set. These distributions and their probability density functions are listed in Table 2.3. It is important to emphasize that all these distributions have support in the positive real set and their parameters are also real positives, except the location parameter ($\mu \in$ IR) of the Log-Normal distribution.

Table 2.1: RB (%) and MSE of the MLE $\hat{\boldsymbol{\theta}}$ for the Alpha distribution.

α	β	$\widehat{\alpha}$		$\widehat{\beta}$	
$n = 20$		RB (%)	MSE	RB (%)	MSE
1	1	-2.0100	0.0867	2.6000	0.0193
1	2	-1.9200	0.0854	3.1300	0.0783
2	1	0.0950	0.0539	1.3500	0.0086
2	2	-0.9400	0.0611	2.1900	0.0414
$n = 40$		RB (%)	MSE	RB (%)	MSE
1	1	-0.2100	0.0383	1.1100	0.0077
1	2	-1.8300	0.0397	1.8550	0.0352
2	1	0.1050	0.0292	0.5500	0.0043
2	2	-0.1700	0.0294	0.9150	0.0174
$n = 60$		RB (%)	MSE	RB (%)	MSE
1	1	-0.7100	0.0269	0.7600	0.0051
1	2	-0.7000	0.0279	0.7650	0.0217
2	1	-0.0500	0.0186	0.4800	0.0028
2	2	0.3400	0.0191	0.1900	0.0102
$n = 80$		RB (%)	MSE	RB (%)	MSE
1	1	0.1400	0.0196	0.4100	0.0039
1	2	-0.1900	0.0202	0.5660	0.0156
2	1	0.2200	0.0135	0.1500	0.0020
2	2	-0.0100	0.0149	0.3150	0.0085
$n = 100$		RB (%)	MSE	RB (%)	MSE
1	1	-0.5900	0.0162	0.5300	0.0033
1	2	-0.6100	0.0172	0.6050	0.0135
2	1	-0.1250	0.0115	0.3100	0.0017
2	2	0.2650	0.0111	0.0800	0.0061

The MLEs of the parameters are computed (as discussed in Section 2.3) and the goodness-of-fit statistics for Alpha model are compared with other distributions. We consider the Anderson-Darling (A*) and Cramér-von Misses (W*), which are described by Chen and Balakrishnan (1995), and Kolmogorov-Smirnov (KS) statistics. Are also calculated other measures of goodness of fit. These functions are: Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criteria (CAIC) and Hannan-Quinn Information Criterion (HQIC).

The W* and A* statistics are used to verify which distribution fits better to the data. Since the values of W* and A* are smaller for the Alpha model when compared with those values of the other models, the considered model seems to be a very competitive model for the data. We use these statistics, where we have a random sample $(x_1, \ldots, x_n)$ with empirical distribution function $F_n(x)$ and require to test if the sample comes from a special distribution. The W^* and A^* statistics are given by

$$W^* = \left\{ n \int_{-\infty}^{+\infty} \{F_n(x) - F(x;\hat{\theta}_n)\}^2 \, dF(x;\hat{\theta}_n) \right\} \left(1 + \frac{0.5}{n}\right) = W^2 \left(1 + \frac{0.5}{n}\right),$$

$$A^* = \left\{ n \int_{-\infty}^{+\infty} \frac{\{F_n(x) - F(x;\hat{\theta}_n)\}^2}{\{F(x;\hat{\theta})\,(1 - F(x;\hat{\theta}_n))\}} \, dF(x;\hat{\theta}_n) \right\} \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$$

$$= A^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right),$$

respectively, where $F_n(x)$ is the empirical distribution function and $F(x;\hat{\theta}_n)$ is the specified distribution function evaluated at the MLE $\hat{\theta}_n$ of θ. Note that the W^* and A^* statistics are given by difference distances of $F_n(x)$ and $F(x;\hat{\theta}_n)$. Thus, the lower are the W^* and A^* statistics more evidence we have that $F(x;\hat{\theta}_n)$ generates the sample.

The package `GenSA` available in the `R` programming language is used to obtain the initial values of the model parameters and the computations to the fit model are performed using the function `goodness.fit` of `AdequacyModel` package. The numerical BFGS (Broyden-Fletcher-Goldfarb-Shanno) procedure is used for minimization of the −log-likelihood function.

The data sets are:

(i) *Tumor-free time data*
This data set refers to the tumor-free time (days) of rats on low-fat diet. For details about these data, see Lee and Wang (2003).

(ii) *Cumulative proportion surviving data*
Cumulative proportion surviving from beginning of study to end of interval (1945-1954) for male patients with localized cancer of rectum diagnosed in Connecticut. For more details, see Lee and Wang (2003).

(iii) *Relief times data*
The data set represent the relief times of twenty patients receiving an analgesic and it is given by Gross and Clark (1975).

(iv) *Survival times of animals data*
These data refers to survival times (in 10 hour units) of animals in a experiment. More details from this data set can be obtained in Hand et al. (1993).

Table 2.2 gives a descriptive summary of each data set, which includes central tendency statistics, standard deviation, among others. The distribution of the all data have positive skewness (right skewed). With respect to kurtosis, we note that the distribution of tumor-free and proportion data have negative kurtosis whereas the distribution of relief and animals data present positive kurtosis.

Table 2.2: Descriptive statistics.

Statistic	Data			
	tumor-free	proportion	relief	animals
Mean	107.000	0.455	1.900	0.479
Median	86.000	0.418	1.700	0.400
Std. Dev.	48.918	0.151	0.704	0.252
Minimum	50.000	0.290	1.100	0.180
Maximum	191.000	0.751	4.100	1.240
Skewness	0.523	0.628	1.592	1.153
Kurtosis	−1.422	−1.050	2.346	0.626
n	15	10	20	48

Table 2.3: Distribution, probability density function and reference.

Distribution	Probability density function	Reference
Birnbaum-Saunders (BS) (α, β)	$(2\alpha e^{\alpha^2} \sqrt{2\pi\beta})^{-1}\, x^{-3/2}\, (x+\beta)\, \exp\left[-\frac{(x/\beta)-(x/\beta)^{-1}}{2\alpha^2}\right]$	Birnbaum and Saunders, 1969
Burr XII (α, β)	$\alpha\, \beta\, \frac{x^{\alpha-1}}{(1+x^{\alpha})^{\beta+1}}$	Burr, 1942
Exp-Half-Normal (σ, a)	$\frac{a\sqrt{2}}{\sigma\sqrt{\pi}}\, \exp\left\{-\frac{x^2}{2\,\sigma^2}\right\}\, \{2\,\Phi(x)-1\}^{a-1}$	Gupta et al., 1998; Cooray and Ananda, 2008
Exp-Lindley (a, θ)	$\frac{a\,\theta^2}{\theta+1}\,(1+x)\,e^{-\theta\, x}\left\{1-\frac{e^{-\theta\, x}(1+\theta+\theta\, x)}{1+\theta}\right\}^{a-1}$	Nadarajah et al., 2012
Gamma (k, θ)	$\frac{x^{k-1}\, e^{-x/\theta}}{\theta^k\, \Gamma(k)}$	Khodabin and Ahmadabadi, 2010 [1]
Gompertz (α, β)	$\alpha\beta \exp[\beta + \alpha\, x - \beta \exp(\alpha\, x)]$	Abu-Zinadah and Aloufi, 2014 [2]
Inverse Gaussian (λ, μ)	$\left[\frac{\lambda}{2\pi\, x^3}\right]^{1/2}\, \exp\left\{-\frac{\lambda\,(x-\mu)^2}{2\,\mu^2\, x}\right\}$	Chhikara and Folks, 1989
Log-Logistic II (α, β)	$\frac{\alpha\beta\,(\beta\, x)^{\alpha-1}}{[1+(\beta\, x)^{\alpha}]^2}$	Singh, 1994
Log-Normal (μ, σ)	$\frac{1}{\sqrt{2\pi}\sigma\, x}\, \exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}$	Johnson, Kotz and Balakrishnan, 1994
Nadarajah-Haghighi (NH) (α, λ)	$\alpha\lambda\,(1+\lambda\, x)^{\alpha-1}\, \exp\{1-[1+\lambda\, x]^{\alpha}\}$	Nadarajah and Haghighi, 2011
Rayleigh II (μ, λ)	$2\lambda\,(x-\mu)e^{-\lambda(x-\mu)^2}$	Johnson, Kotz and Balakrishnan, 1994
Weibull (λ, γ)	$\frac{\gamma}{\lambda}\, \left(\frac{x}{\lambda}\right)^{\gamma-1}\, \exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\}$	Weibull, 1951
Chi-Squared (μ)	$\frac{1}{2^{\mu/2}\Gamma(\mu/2)} x^{\mu/2-1}\, e^{-x/2}$	Johnson, Kotz and Balakrishnan, 1998
Exponential (λ)	$\lambda\, e^{-\lambda\, x}$	Gupta and Kundu, 1999, 2001
Half-Normal (σ^2)	$\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\, \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Cooray and Ananda, 2008
Log-Logistic I (α)	$\frac{\alpha\, x^{\alpha-1}}{(1+x^{\alpha})^2}$	Singh, 1998 [3]
Rayleigh I (σ)	$\frac{x}{\sigma^2}\, \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Hoffman and Karst, 1975

[1] with new parametrization and $\tau = 1$

[2] with $a = 1$

[3] with $\beta = 1$

One of the important device which can help selecting a particular model is the total time on test (TTT) plot (for more details see Aarset, 1987). This plot is constructed through the quantities

$$T(i/n) = \left[\sum_{j=1}^{i} X_{j:n} + (n-i)\, X_{i:n} \right] \Big/ \sum_{j=1}^{n} X_{j:n} \ versus \ i/n,$$

where $i = 1, \ldots, n$ and $X_{j:n}$ is the j-th order statistics of the sample (Mudholkar et al., 1995).

The Figure 2.3 shows the TTT plots for the tumor-free time, cumulative proportion surviving, relief times and survival times of animals data. These plots indicate an increasing hrf and then reveals the adequacy of the Alpha distribution to fit these data.

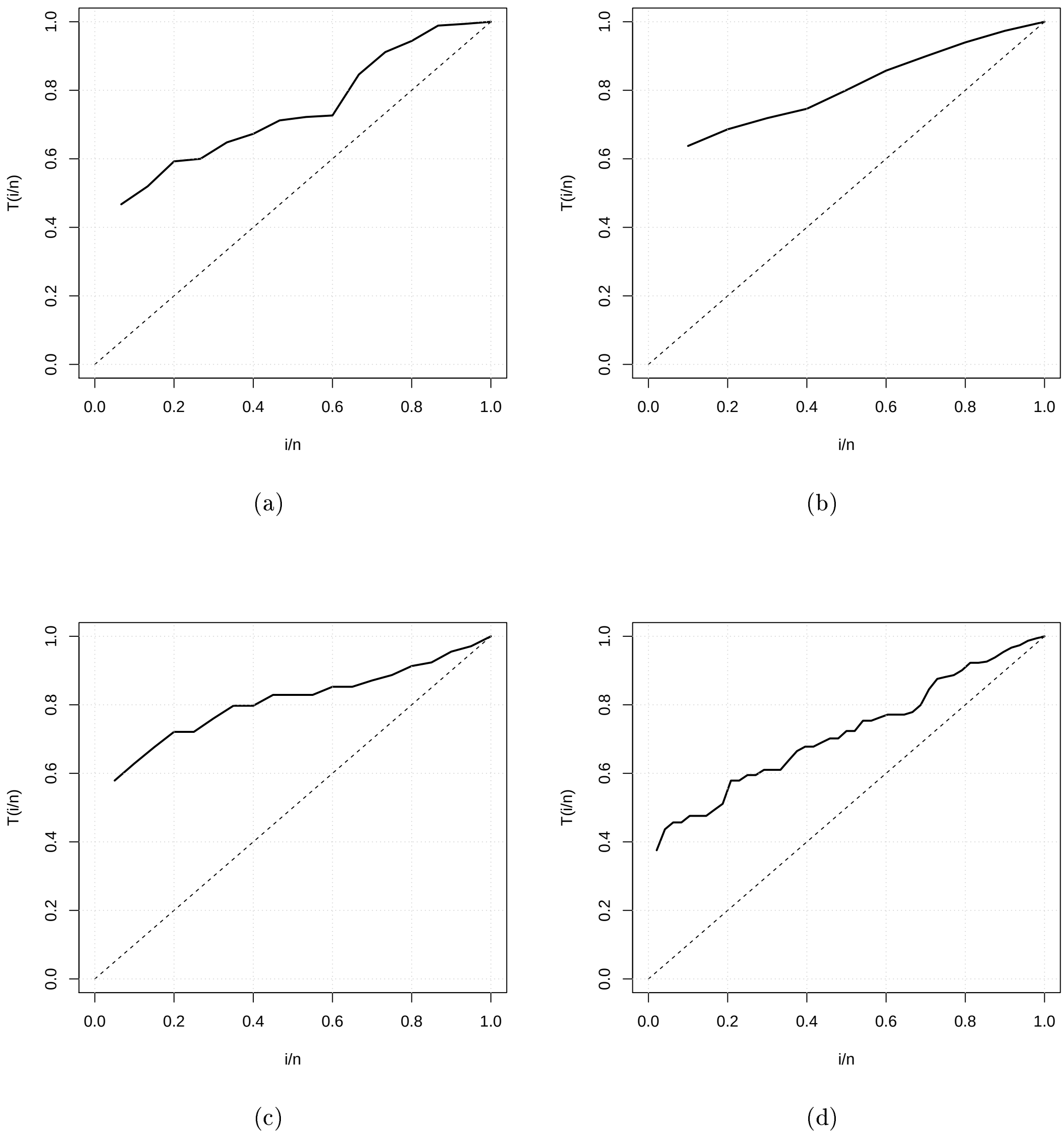


Figure 2.3: TTT plots (a) tumor-free time data; (b) cumulative proportion surviving data; (c) relief times data; (d) survival times of animals data.

Tables 2.4, 2.5, 2.6 and 2.7 give the MLEs (with standard errors in parentheses) for some fitted models for tumor-free time data, cumulative proportion surviving data, relief times data, survival times of animals, respectively. Table 2.8 lists the values of the W*, A*, KS, AIC, CAIC, BIC and HICQ for all fitted models for the considered four data sets. This table reveals that the Alpha model corresponds to the best fit for all data set, even when compared to distributions (with one and two parameters) well-known in the literature for fit to real data set. Note that we used differents scenarios to the data set in concerning the compared distributions. This fact proves that, in differents situations, the proposed model have a upper performance to the other distributions analyzed.

Table 2.4: Maximum likelihood estimates (standard errors in parentheses) for the tumor-free time data.

Distribution	MLEs (standard errors)	
Alpha (α, β)	2.378	213.689
	(0.585)	(43.716)
Exp-Half-Normal (σ, a)	77.782	3.352
	(12.605)	(1.304)
Gamma (k, θ)	5.413	0.050
	(1.913)	(0.018)
Log-Logistic II (α, β)	3.767	0.010
	(0.783)	(0.001)
Weibull (λ, γ)	121.310	2.460
	(13.504)	(0.491)
Chi-Squared (μ)	98.280	
	(3.601)	
Exponential (λ)	0.009	
	(0.002)	
Half-Normal (σ)	116.970	
	(21.355)	
Log-Logistic I (α)	0.336	
	(0.066)	

Table 2.5: Maximum likelihood estimates (standard errors in parentheses) for the cumulative proportion surviving data.

Distribution	MLEs (standard errors)	
Alpha (α, β)	3.541	1.475
	(0.861)	(0.333)
BS (α, β)	0.302	0.435
	(0.067)	(0.041)
Burr XII (α, β)	3.523	11.687
	(0.771)	(5.868)
Gamma (k, θ)	10.912	23.949
	(4.807)	(10.796)
Inverse Gaussian (λ, μ)	4.855	0.455
	(2.171)	(0.044)
Log-Logistic II (α, β)	5.510	2.338
	(1.409)	(0.239)
Rayleigh II (μ, λ)	0.218	13.049
	(0.049)	(5.767)
Chi-Squared (μ)	1.208	
	(0.333)	
Exponential (λ)	2.194	
	(0.694)	
Half-Normal (σ)	0.477	
	(0.106)	
Log-Logistic I (α)	1.795	
	(0.446)	
Rayleigh I (σ)	0.337	
	(0.053)	

Table 2.6: Maximum likelihood estimates (standard errors in parentheses) for the relief times data.

Distribution	MLEs (standard errors)	
Alpha (α, β)	3.586	6.187
	(0.615)	(0.986)
BS (α, β)	0.314	1.810
	(0.049)	(0.125)
Exp-Half-Normal (σ, a)	1.088	7.339
	(0.140)	(3.134)
Gamma (k, θ)	9.669	5.089
	(3.006)	(1.624)
Inverse Gaussian (λ, μ)	18.679	1.900
	(5.912)	(0.135)
Log-Logistic II (α, β)	5.889	0.570
	(1.115)	(0.037)
Log-Normal (μ, σ)	0.589	0.310
	(0.069)	(0.049)
Rayleigh II (μ, λ)	0.839	0.626
	(0.161)	(0.194)
Weibull (λ, γ)	2.129	2.787
	(0.182)	(0.427)
Chi-Squared (μ)	2.718	
	(0.432)	
Exponential (λ)	0.526	
	(0.117)	
Half-Normal (σ)	2.020	
	(0.319)	
Log-Logistic I (α)	2.491	
	(0.447)	
Rayleigh I (σ)	1.428	
	(0.159)	

Table 2.7: Maximum likelihood estimates (standard errors in parentheses) for the survival times of animals data.

Distribution	MLEs (standard errors)	
Alpha (α, β)	2.085	0.812
	(0.325)	(0.098)
Burr XII (α, β)	2.343	4.923
	(0.230)	(0.817)
Exp-Half-Normal (σ, a)	0.384	2.625
	(0.037)	(0.550)
Exp-Lindley (a, θ)	6.303	5.963
	(1.887)	(0.767)
Gamma (k, θ)	4.307	8.986
	(0.847)	(1.875)
Gompertz (α, β)	2.210	0.403
	(0.460)	(0.182)
Log-Logistic II (α, β)	3.527	2.409
	(0.418)	(0.174)
NH (α, λ)	645.929	0.002
	(125.455)	(0.0003)
Rayleigh II (μ, λ)	0.082	4.545
	(0.039)	(0.917)
Weibull (λ, γ)	0.544	2.060
	(0.040)	(0.218)
Chi-Squared (μ)	1.195	
	(0.150)	
Exponential (λ)	2.086	
	(0.301)	
Half-Normal (σ)	0.540	
	(0.055)	
Log-Logistic I (α)	1.656	
	(0.191)	
Rayleigh I (σ)	0.382	
	(0.027)	

Table 2.8: Goodness-of-fit statistics.

Distribution	W^*	A^*	KS	AIC	CAIC	BIC	HQIC
Tumor-free time data							
Alpha	**0.051**	**0.358**	**0.143**	**159.179**	**160.179**	**160.595**	**159.164**
Exp-Half-Normal	0.111	0.652	0.241	160.076	161.076	161.492	160.061
Gamma	0.100	0.587	0.225	159.481	160.481	160.898	159.466
Log-Logistic II	0.090	0.546	0.184	160.323	161.323	161.739	160.307
Weibull	0.123	0.716	0.243	160.612	161.612	162.028	160.597
Chi-Squared	0.100	0.589	0.385	254.309	254.617	255.017	254.302
Exponential	0.099	0.586	0.374	172.185	172.492	172.893	172.177
Half-Normal	0.115	0.672	0.331	166.631	166.939	167.339	166.624
Log-Logistic I	0.080	0.485	0.788	229.958	230.266	230.666	229.950
Cumulative proportion surviving data							
Alpha	**0.027**	**0.192**	**0.144**	**−8.736**	**−7.022**	**−8.131**	**−9.400**
BS	0.032	0.232	0.161	−8.385	−6.671	−7.780	−9.045
Burr XII	0.052	0.361	0.152	−6.631	−4.917	−6.026	−7.295
Gamma	0.038	0.270	0.164	−7.868	−6.154	−7.263	−8.532
Inverse Gaussian	0.032	0.231	0.161	−8.399	−6.685	−7.794	−9.063
Log-Logistic II	0.036	0.257	0.146	−7.482	−5.768	−6.877	−8.146
Rayleigh II	0.031	0.231	0.195	−8.616	−6.902	−8.011	−9.280
Chi-Squared	0.036	0.262	0.459	16.177	16.677	16.480	15.845
Exponential	0.037	0.270	0.471	6.278	6.778	6.581	5.947
Half-Normal	0.044	0.311	0.456	1.740	2.240	2.043	1.408
Log-Logistic I	0.036	0.258	0.625	12.545	13.045	12.848	12.213
Rayleigh I	0.044	0.313	0.309	−4.762	−4.262	−4.459	−5.094
Relief times data							
Alpha	**0.026**	**0.150**	**0.090**	**34.943**	**35.648**	**36.934**	**35.331**
BS	0.074	0.440	0.157	37.631	38.337	39.623	38.020
Exp-Half-Normal	0.110	0.654	0.182	40.087	40.793	42.078	40.475
Gamma	0.105	0.626	0.173	39.637	40.343	41.628	40.025
Inverse Gaussian	0.073	0.431	0.156	37.544	38.250	39.536	37.933
Log-Logistic II	0.051	0.309	0.110	36.953	37.659	38.944	37.342
Log-Normal	0.072	0.425	0.151	37.535	38.241	39.526	37.924
Rayleigh II	0.106	0.626	0.210	39.615	40.321	41.607	40.004
Weibull	0.185	1.092	0.184	45.172	45.878	47.164	45.561
Chi-Squared	0.103	0.610	0.309	64.568	64.791	65.564	64.763
Exponential	0.105	0.624	0.439	67.674	67.896	68.669	67.868
Half-Normal	0.146	0.864	0.413	59.158	59.380	60.154	59.352
Log-Logistic I	0.049	0.287	0.561	67.604	67.827	68.600	67.799
Rayleigh I	0.147	0.874	0.256	46.957	47.179	47.953	47.151
Survival times of animals data							
Alpha	**0.067**	**0.507**	**0.080**	**−10.614**	**−10.348**	**−6.872**	**−9.200**
Burr XII	0.180	1.042	0.144	−4.905	−4.638	−1.162	−3.490
Exp-Half-Normal	0.197	1.136	0.166	−4.402	−4.135	−0.659	−2.987
Exp-Lindley	0.103	0.620	0.113	−10.417	−10.150	−6.674	−9.003
Gamma	0.139	0.813	0.136	−8.340	−8.073	−4.597	−6.925
Gompertz	0.337	1.934	0.191	9.346	9.613	13.088	10.760
Log-Logistic II	0.085	0.561	0.087	−9.385	−9.118	−5.642	−7.970
NH	0.254	1.460	0.264	12.717	12.983	16.459	14.131
Rayleigh II	0.177	1.023	0.190	−5.518	−5.251	−1.776	−4.104
Weibull	0.219	1.264	0.160	−2.506	−2.239	1.236	−1.091
Chi-Squared	0.124	0.734	0.339	70.301	70.388	72.173	71.009
Exponential	0.138	0.808	0.334	27.413	27.500	29.285	28.121
Half-Normal	0.210	1.212	0.281	12.654	12.741	14.525	13.361
Log-Logistic I	0.110	0.662	0.486	59.220	59.307	61.091	59.927
Rayleigh I	0.215	1.237	0.151	−4.427	−4.340	−2.556	−3.720

2.5 Concluding remarks

We present the Alpha distribution discussed by Salvia (1985), which provided its characterization and a number of structural properties, but did not perform a simulation study and application to show the usefulness of this distribution. This model have one shape parameter and one scale parameter and its cdf and pdf involves the standard normal cumulative function. We study the shapes of the pdf and hrf of the Alpha model. We estimate the model parameters by maximum likelihood method, present a Monte Carlo simulation study to evaluate the assymptotic properties of the maximum likelihood estimators and provide four applications to real data, which indicate that the Alpha distribution is a competitive model to fit real data, if compared with well-known distributions in survival analysis field.

Chapter 3

The Exponentiated Alpha Distribution

Resumo

Um modelo com três parâmetros denominado Alpha Exponencializado (Exp-Alpha) é proposto como uma generalização da distribuição Alpha (Katsev, 1968; Wager e Barash, 1971; Sherif, 1983; Salvia, 1985). Discutimos as formas de sua função densidade de probabilidade e da função taxa de falha e obtemos três formas para a função taxa de falha do modelo proposto: banheira invertida, crescente e decrescente. Algumas propriedades matemáticas são estudadas, tais como: desvios médios, estatísticas de ordem, função geratriz de momentos, função quantílica e momentos. O método de máxima verossimilhança é utilizado para estimar os parâmetros do modelo e obtemos um estimador em forma semi fechada para o parâmetro adicional. Fornecemos um estudo de simulação de Monte Carlo para avaliar os estimadores de máxima verossimilhança e três aplicações a dados reais mostram a flexibilidade do novo modelo.

Palavras-chave: Desvios médios. Distribuição Alpha. Estatísticas de ordem. Exponencializada-G. Função quantílica. Máxima verossimilhança. Momentos.

Abstract

A three-parameter model called Exponentiated Alpha (Exp-Alpha, for short) distribution is proposed as a generalization of the Alpha distribution (Katsev, 1968; Wager and Barash, 1971; Sherif, 1983; Salvia, 1985). We discuss the shapes of its probability density function and hazard rate function and we obtain three forms to this function of the proposed model: upside-down bathtub, increasing and decreasing. Some mathematical properties are studied, such as: generating function, mean deviations, moments, order statistics and quantile function. The method of maximum likelihood is used to estimate the model parameters and we obtain a semiclosed estimator for the additional parameter. We provide a Monte Carlo simulation study to evaluate the maximum likelihood estimators and three applications to real data show the flexibility of the new model.

Keywords: Alpha Distribution. Exponentiated class. Maximum Likelihood Estimation. Mean deviations. Moments. Order statistics. Quantile function.

3.1 Introduction

The exponentiated class (Gompertz, 1825; Verhulst, 1838, 1845, 1847) is widely applied to baseline distributions in applications to real data set because this class provide a greater flexibility in this case. Over many years, many exponentiated type distributions have been proposed. Mudholkar and Srivastava (1993) studied the shapes of the exponentiated Weibull hazard rate function (hrf), provided application to real data set and used the maximum likelihood method to estimate the model parameters. Gupta et al. (1998) proposed a generalization of the standard exponential distribution, called the exponentiated exponential distribution with two parameters (shape and scale) similar to the Weibull and gamma families, for example. More details of the mathematical properties of the exponentiated exponential distribution can be found in Gupta and Kundu (2001), where they observed the similarity of the new model with the Weibull and gamma distributions. The exponentiated gamma distribution was proposed by Nadarajah and Gupta (2007) and they provided some mathematical properties of its distribution, the maximum likelihood method to estimate the model parameters and illustrated the flexibility of the new model through applications to drought data. For more examples of the usefulness of the exponentiated class, see Gupta et al. (1998) and Cordeiro et al. (2011) for the exponentiated Pareto and exponentiated generalized gamma distributions, respectively. In a general form, for mathematical properties and estimation of the parameters for some exponentiated distributions, see AL-Hussaini and Ahsanullah (2015).

The exponentiated class is the most known in survival analysis, more specifically in generation of the new class of distributions. For an arbitrary baseline $G(x)$, a random variable has the Exponentiated-G (Exp-G) distribution with additional power parameter $a > 0$ if its cumulative distribution function (cdf) and probability density function (pdf) are given by

$$F(x;a) = G(x)^a \tag{3.1}$$

and

$$f(x;a) = a\, g(x)\, G(x)^{a-1}, \tag{3.2}$$

respectively, where $g(x) = dG(x)/dx$. We note that there is no complicated function in equation (3.1).

In this chapter, we study a new model, so-called Exponentiated Alpha (Exp-Alpha, for short) distribution by combining the Exp-G class with the Alpha distribution, which is present in Chapter 2 and has cdf and pdf given by

$$G(x;\alpha,\beta) = \frac{\Phi\left(\alpha - \frac{\beta}{x}\right)}{\Phi(\alpha)} \tag{3.3}$$

and

$$g(x;\alpha,\beta) = \frac{\beta}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)}\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}, \tag{3.4}$$

respectively, for $x > 0$, where $\alpha > 0$ is the shape parameter, $\beta > 0$ is the scale parameter and $\Phi(\cdot)$ denote the standard normal cumulative function.

The Alpha distribution is generally used in tool wear problems (Katsev, 1968; Wager and Barash, 1971). Sherif (1983) suggested its use in modeling lifetimes under accelerated test conditions and Salvia (1985) described this distribution and its applicability as a model for accelerated life testing. In this sense, the main aim of this chapter is to provide evidence that the extension of the Alpha distribution (Exp-Alpha) is more flexible than the Alpha model and it can be used to fit several data sets.

The rest of this chapter is organized as follows. In Section 3.2, we define the Exp-Alpha distribution. Explicit expressions for quantile function, ordinary moments, moment generating function, mean deviations and order statistics are derived in Section 3.3. Maximum likelihood estimation of the model parameters is investigated in Section 3.4. A simulation study and applications to real data sets in Section 3.5 illustrate the flexibility of the new distribution for data modeling. Finally, concluding remarks are given in Section 3.6.

3.2 The Exp-Alpha distribution

In this section, we define the Exp-Alpha distribution. The cdf and pdf of the Exp-Alpha distribution are given by

$$F(x;a,\alpha,\beta) = \left[\frac{\Phi\left(\alpha-\frac{\beta}{x}\right)}{\Phi(\alpha)}\right]^a \tag{3.5}$$

and

$$f(x;a,\alpha,\beta) = \frac{a\beta\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)^a}\,\Phi\left(\alpha-\frac{\beta}{x}\right)^{a-1}, \tag{3.6}$$

respectively.

The hrf corresponding to (3.6) is given by

$$h(x;a,\alpha,\beta) = \frac{a\beta\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}\,\Phi\left(\alpha-\frac{\beta}{x}\right)^{a-1}}{\sqrt{2\pi}\,x^2\,\left[\Phi(\alpha)^a-\Phi\left(\alpha-\frac{\beta}{x}\right)^a\right]}.$$

A random variable X having density function (3.6) is denoted by $X \sim \text{Exp-Alpha}(a,\alpha,\beta)$. In this model, $a > 0$ and $\alpha > 0$ are shape parameters and $\beta > 0$ is a scale parameter. We write $F(x) = F(x;a,\alpha,\beta)$ and $f(x) = f(x;a,\alpha,\beta)$ in order to eliminate the dependence on the model parameters.We write $F(x) = F(x;a,\alpha,\beta)$ and $f(x) = f(x;a,\alpha,\beta)$ in order to eliminate the dependence on the model parameters. The Alpha distribution is clearly a special case of (3.5) when $a = 1$.

Plots of the Exp-Alpha pdf and hrf for selected parameter values are displayed in Figure 3.1, respectively. It is evident that the additional shape parameter a allows for a high degree of flexibility of the Exp-Alpha distribution, i.e., this distribution is more flexible than the Alpha distribution. So, the proposed model can be very useful for modeling positive real data sets.

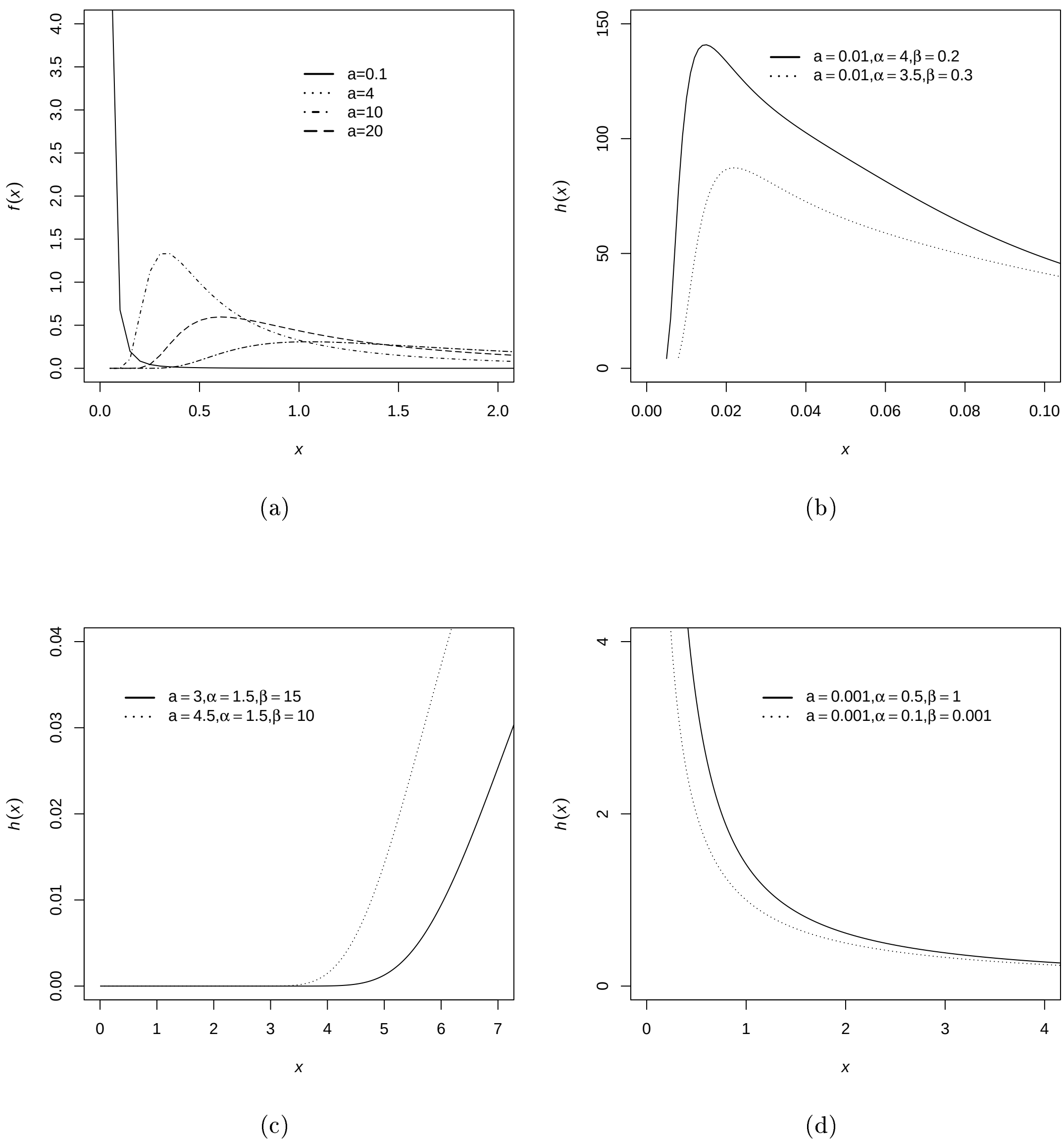


Figure 3.1: (a) Plots of the Exp-Alpha density for some parameter values. Plots of the Exp-Alpha hazard rate function for some parameter values: (b) Upside-down bathtub hazard rate function. (c) Increasing hazard rate function. (d) Decreasing hazard rate function.

3.3 Properties

In this section, we provide some mathematical properties of the Exp-Alpha distribution in order to show interesting characteristics of the proposed model. It is important to emphasize that the results presented here are only valid for certain regions of the parametric space due to non existence of moments for the Alpha distribution.

3.3.1 Quantile function

The quantile function (qf) of X can be obtained by inverting the Exp-Alpha cdf (3.5) as

$$x_u = Q(u) = \frac{\beta}{\alpha - \Phi^{-1}\{\Phi(\alpha)\, u^{1/a}\}}, \quad u \in (0,1).$$

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

The effect of the additional shape parameter a on the skewness and kurtosis of the new distribution can be based on quantile measures. In this sense, two important measures are the Bowley skewness (Kenney and Keeping, 1962) and the Moors kurtosis (Moors, 1988).

The Bowley skewness (B) is based on quartiles, while the Moors kurtosis (M) is based on octiles. These measures are given by

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

and

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

respectively.

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 parameters α and β as functions of a are displayed in Figure 3.2. These plots indicate that both skewness and kurtosis increase when a increases for some fixed values of α and β. Further, the skewness is negative for small values of a when $\alpha = \beta = 100$. This fact justify the adequacy of the Exp-Alpha distribution to real data set which has negative skewness, as we shall see in the Section 3.5.2.

3.3.2 Moments

The ordinary moments of X can be obtained in closed-form as an infinite weighted linear combination of the baseline probability weighted moments (PWMs). First, we review the PWMs of the Alpha distribution since they are required for the ordinary moments of the Exp-Alpha distribution. If $Z \sim \text{Alpha}(\alpha, \beta)$ has the pdf (3.4), the (s,r)th PWM of Z is formally defined by

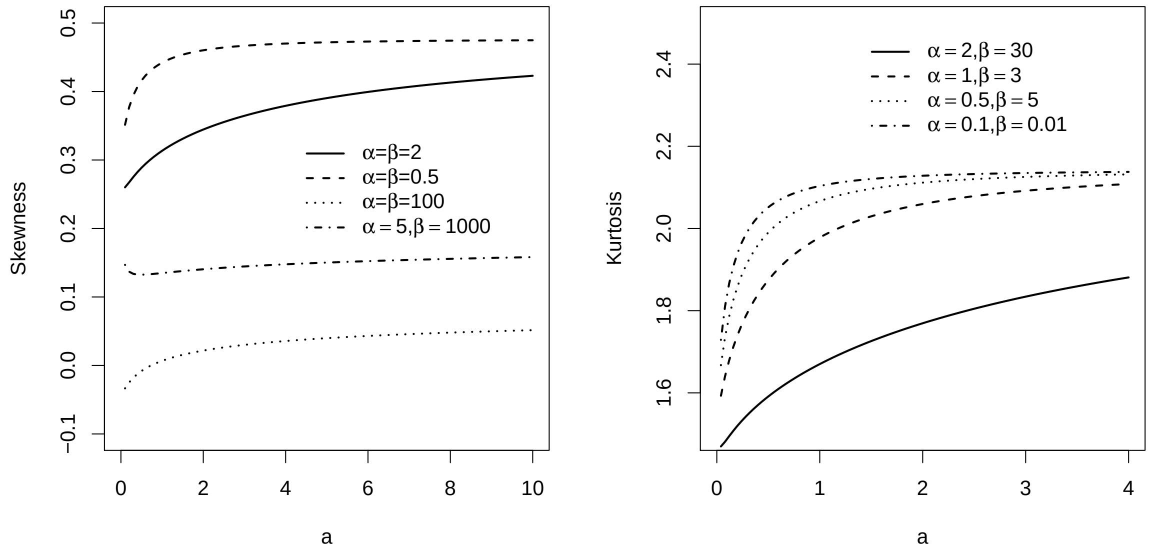


Figure 3.2: Plots of the Exp-Alpha skewness and kurtosis as functions of a for some values of α and β.

$$\tau_{s,r} = E(Z^s\, G(x)^r) = \int_0^\infty x^s\, G(x)^r\, g(x)\, dx. \tag{3.7}$$

Substituting equations (3.3) and (3.4) in (3.7) and after some algebraic manipulations we obtain $\tau_{s,r}$ as

$$\tau_{s,r} = \frac{c_1}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \sum_{m=0}^{2s_j+j} A(k_1,\ldots,k_j) \binom{2s_j+j}{m} \alpha^{2s_j+j-m} (-\beta)^m\, I(s-m-2,\alpha,\beta), \tag{3.8}$$

where $c_1 = \frac{\beta}{\sqrt{2\pi}\,\Phi(\alpha)^{r+1}}$, $a_k = \frac{(-1)^k\, 2^{(1-2k)/2}}{\sqrt{\pi}(2k+1)\,k!}$, $A(k_1,\ldots,k_j) = a_{k_1}\ldots a_{k_j}$ and $s_j = k_1+\ldots+k_j$. Here, $I(p,\alpha,\beta)$ is the integral given by

$$\begin{aligned} I(p,\alpha,\beta) &= \int_0^\infty x^p \exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\} dx \\ &= \frac{1}{\sqrt{\pi}}\left\{2^{(-3/2-p/2)}\,\beta^{(p+1)} \exp^{-\alpha^2/2}\left[-\frac{1}{2}\frac{\sqrt{2}\pi^2\alpha\beta\, L_{p/2}^{1/2}\left(\frac{\alpha^2}{2}\right)}{\sqrt{\beta^2}\sin\left(\frac{p}{2}\pi\right)\,\Gamma\left(\frac{3}{2}+\frac{p}{2}\right)}\right.\right. \\ &\quad \left.\left. -\frac{1}{2}\frac{\pi^2(\alpha^2+1)\, L_{p/2}^{1/2}\left(\frac{\alpha^2}{2}\right)}{\cos\left(\frac{p}{2}\pi\right)\,\Gamma\left(2+\frac{p}{2}\right)} + \frac{1}{2}\frac{\pi^2\alpha^2\, L_{p/2}^{3/2}\left(\frac{\alpha^2}{2}\right)}{\cos\left(\frac{1}{2}\pi\right)\,\Gamma\left(1+\frac{p}{2}\right)}\right]\right\}, \end{aligned}$$

where $L_{p/2}^{1/2}\left(\frac{\alpha^2}{2}\right)$, $\Gamma(\cdot)$, $\sin(\cdot)$ and $\cos(\cdot)$ are the Laguerre, gamma, sine and cosine functions, respectively.

Finally, the moments of X can be expressed from (3.2) as

$$E(X^s) = \int_0^\infty x^s\, f(x)\, dx$$
$$= a\, \tau_{s,a-1}. \tag{3.9}$$

Then, we obtain the moments of the Exp-Alpha distribution as an infinite linear combination of convenient PWMs of the Alpha distribution. Equation (3.9) is the main result of this section.

3.3.3 Generating function

Here, we derive the moment generating function (mgf) $M(t) = E(\mathrm{e}^{tX})$ of the Exp-Alpha distribution. Thus,

$$M(t) = \int_0^\infty \mathrm{e}^{tx}\, f(x)\, dx$$
$$= \int_0^\infty \mathrm{e}^{tx}\, a\, G(x)^{a-1}\, g(x)\, dx.$$

Using the power series for the exponential function, the Exp-Alpha mgf reduces to

$$M(t) = a \sum_{\ell=0}^{\infty} d_\ell\, \tau_{\ell,a-1},$$

where $d_\ell = \frac{t^\ell}{\ell!}$ and $\tau_{\ell,a-1} = \int_0^\infty x^\ell\, G(x)^{a-1}\, g(x)\, dx$ is the PWM of the Alpha distribution.

3.3.4 Mean Deviations

If X has the Exp-Alpha distribution with cdf $F(x)$ given by (3.5), we can derive the mean deviations about the mean $\mu = E(X)$ and about the median M from the relations

$$\delta_1 = \int_0^\infty |x-\mu|\, f(x)\, dx \qquad \text{and} \qquad \delta_2 = \int_0^\infty |x-M|\, f(x)\, dx,$$

respectively.

Defining the first incomplete moment of X by $I(s) = \int_0^s x\, f(x)\, dx$, these measures can be determined from

$$\delta_1 = 2\,\mu\, F(\mu) - 2\, I(\mu) \qquad \text{and} \qquad \delta_2 = E(X) + 2\, M\, F(M) - M - 2\, I(M), \tag{3.10}$$

where $F(\mu)$ and $F(M)$ are easily obtained from equation (3.5).

Setting

$$\rho(s,r) = \int_0^s x\, G(x)^r\, g(x)\, dx,$$

we obtain from equation (3.2)

$$I(s) = a \int_0^s x\, G(x)^{a-1}\, g(x)\, dx = a\, \rho(s, a-1).$$

Then, $I(\mu) = a\, \rho(\mu, a-1)$.

Combining (3.3) and (3.8) and defining

$$J(s, -m-1) = \int_0^s x^{-m-1} \exp\left\{ -\frac{1}{2} \left(\alpha - \frac{\beta}{x} \right)^2 \right\} dx,$$

we can write

$$\rho(s,r) = \frac{\beta}{\sqrt{2\pi}\, 2^r\, \Phi^{r+1}(\alpha)} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \sum_{m=0}^{2s_j+j} A(k_1,\ldots,k_j) \\ \times \binom{2s_j+j}{m} \alpha^{2s_j+j-m}\, (-\beta)^m\, J(s,-m-1).$$

Letting $t = \alpha - \beta/x$, the last integral reduces to

$$J(s,-m-1) = \int_{-\infty}^{\alpha-\beta/s} \frac{1}{\beta^m} (\alpha - t)^{m-1}\, \mathrm{e}^{-t^2/2}\, dt$$

and using the binomial expansion and interchanging terms, we have

$$J(s,-m-1) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell\, \alpha^{m+1-\ell}}{\beta^{m+1}} \binom{m+1}{\ell} \int_{-\infty}^{\alpha-\beta/s} t^\ell\, \mathrm{e}^{-t^2/2}\, dt.$$

We now define

$$G(\ell) = \int_0^\infty x^\ell\, \mathrm{e}^{-x^2/2} dx = 2^{(\ell-1)/2}\, \Gamma(\ell + 1/2).$$

For evaluating the integral in $J(s,-m-1)$, it is required to consider two cases. If $\alpha - \beta/s < 0$, we have

$$\int_{-\infty}^{\alpha-\beta/s} t^\ell\, \mathrm{e}^{-t^2/2}\, dt = (-1)^\ell\, G(\ell) + (-1)^{\ell+1} \int_0^{-(\alpha-\beta/s)} t^\ell\, \mathrm{e}^{-t^2/2}\, dt.$$

If $\alpha - \beta/s > 0$, we have

$$\int_{-\infty}^{\alpha-\beta/s} t^\ell\, \mathrm{e}^{-t^2/2}\, dt = (-1)^\ell\, G(\ell) + \int_0^{\alpha-\beta/s} t^\ell\, \mathrm{e}^{-t^2/2}\, dt.$$

Further, the integrals of the type $\int_0^q x^\ell \, \mathrm{e}^{-x^2/2}\, dx$ can be calculated easily as

$$\begin{aligned}\int_0^q x^\ell \, \mathrm{e}^{-x^2/2}\, dx = & \frac{2^{\ell/4+1/4}\, q^{\ell/2+1/2}\, \mathrm{e}^{-q^2/4}}{(\ell/2+1/2)(\ell+3)} M_{\ell/4+1/4,\ell/4+3/4}(q^2/4) \\ & + \frac{2^{\ell/4+1/4}\, q^{\ell/2-3/2}\, \mathrm{e}^{-q^2/4}}{\ell/2+1/2} M_{\ell/4+5/4,\ell/4+3/4}(q^2/4),\end{aligned}$$

where $M_{k,m}(x)$ is the Whittaker function, which can be expressed in terms of the confluent hypergeometric function ${}_1F_1$ by $M_{k,m}(x) = \mathrm{e}^{-x^2/2}\, x^{m+1/2}\, {}_1F_1(\frac{1}{2}+m-k; 1+2m; x)$, where ${}_1F_1(\alpha_1;\beta_1;x) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k}{(\beta_1)_k} \frac{x^k}{k!}$, for $\alpha_1 > 0$, $\beta_1 > 0$.

Hence, we have all quantities to obtain $J(s,-m-1)$, $\rho(s,r)$, $I(\mu)$ and then the mean deviations δ_1 and δ_2 in (3.10).

3.3.5 Order statistics

In this section, we derive an explicit expression for the probability density function of the ith order statistic $X_{i:n}$, say $f_{i:n}(x)$, in a random sample size n from the Exp-Alpha distribution with cdf and pdf given by (3.5) and (3.6), respectively. It is well-known that

$$f_{i:n}(x) = \frac{f(x)}{B(i,n-i+1)} F(x)^{i-1}[1-F(x)]^{n-i}, \tag{3.11}$$

where $B(\cdot,\cdot)$ is the Beta function.

Since $0 < F(x) < 1$ for $x > 0$, we can use the binomial expansion of $[1-F(x)]^{n-i}$ given as follows

$$[1-F(x)]^{n-i} = \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} F(x)^j. \tag{3.12}$$

Substituting from Equation (3.12) into Equation (3.11), we obtain

$$\begin{aligned} f_{i:n}(x) &= \frac{1}{B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x) F(x)^{i+j-1} \\ &= \sum_{j=0}^{n-i} w_j \, [F(x)]^{i+j-1}\, f(x), \end{aligned} \tag{3.13}$$

where $w_j = \frac{(-1)^j\, n!}{j!\,(i-1)!\,(n-i-j)!}$.

Equation (3.13) reveals that the density function of $X_{i:n}$ is a linear combination of Exp-Alpha densities with power parameter $(i+j)$. A direct application of (3.13) is to obtain the moments and the mgf of the Exp-Alpha order statistics.

The rth moment and the mgf of $X_{i:n}$ can be written as

$$E(X_{i:n}^r) = a \sum_{j=0}^{n-i} w_j \, \tau_{r,a(i+j)-1},$$

and

$$M(t) = a \sum_{j=0}^{n-i} \sum_{r=0}^{\infty} w_j \, d_r \, \tau_{r,a(i+j)-1},$$

respectively, where $\tau_{r,a(i+j)-1}$ is defined in (3.7) and $d_r = \frac{t^r}{r!}$.

3.4 Estimation

In this section, we provide an analytic procedure to obtain the maximum likelihood estimators (MLEs) for the Exp-Alpha parameters. The MLEs are employed due to their interesting asymptotic properties.

Let $(x_1, \ldots, x_n)$ be a random sample of size n from the Exp-Alpha distribution with unknown parameter vector $\boldsymbol{\theta} = (a, \alpha, \beta)^\top$. The log-likelihood function for $\boldsymbol{\theta}$ based on a given sample can be expressed as

$$\ell(\boldsymbol{\theta}) = n \, \log\left(\frac{a\beta}{\sqrt{2\pi}}\right) - n\, a \log[\Phi(\alpha)] - 2\sum_{i=1}^{n} \log(x_i) - \frac{n\alpha^2}{2} + \alpha\beta \sum_{i=1}^{n} \frac{1}{x_i} - \frac{\beta^2}{2} \sum_{i=1}^{n} \frac{1}{x_i^2} + (a-1) \sum_{i=1}^{n} \log[\Phi(t_i)], \tag{3.14}$$

where $t_i = \left(\alpha - \frac{\beta}{x_i}\right)$, $i = 1, \ldots, n$.

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

$$U_a(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \log[\Phi(t_i)] - n \, \log[\Phi(\alpha)], \tag{3.15}$$

$$U_\alpha(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = -n\,\alpha - n\,a \frac{\phi(\alpha)}{\Phi(\alpha)} + \beta \sum_{i=1}^{n} \frac{1}{x_i} + (a-1) \sum_{i=1}^{n} \frac{\phi(t_i)}{\Phi(t_i)},$$

$$U_\beta(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta} = \frac{n}{\beta} + \alpha \sum_{i=1}^{n} \frac{1}{x_i} - \beta \sum_{i=1}^{n} \frac{1}{x_i^2} - (a-1) \sum_{i=1}^{n} \frac{\phi(t_i)}{x_i \, \Phi(t_i)}.$$

The MLEs, $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{\alpha}, \hat{\beta})^\top$, can be obtained numerically by solving $U(\hat{\boldsymbol{\theta}}) = 0$.

From equation (3.15) we note that it is possible to obtain a semiclosed estimator of a. For $\frac{\partial \ell(\boldsymbol{\theta})}{\partial a} = 0$, the estimator for a is given by

$$\hat{a}_{\hat{\alpha},\hat{\beta}} = \frac{n}{n \, \log[\Phi(\hat{\alpha})] - \sum_{i=1}^{n} \, \log[\Phi(t_i^*)]}, \tag{3.16}$$

where $t_i^* = (\hat{\alpha} - \frac{\hat{\beta}}{x_i})$.

The estimative of a can be obtained analytically by (3.16).

Another way to obtain the MLEs is using the profile log-likelihood. We suppose a fixed and rewrite the log-likelihood (3.14) as $\ell(\boldsymbol{\theta}) = \ell_a(\alpha, \beta)$. By replacing a by $\hat{a}_{\alpha,\beta}$ in (3.14), we can obtain the profile log-likelihood function as

$$\begin{aligned}\ell_a(\alpha,\beta) = {} & n\log\left(\frac{n\,\beta}{\sqrt{2\pi}\,[n\,\log[\Phi(\alpha)] - \sum_{i=1}^{n}\log[\Phi(t_i)]]}\right) \\ & - 2\sum_{i=1}^{n}\log(x_i) - \frac{n\,\alpha^2}{2} + \alpha\,\beta\sum_{i=1}^{n}\frac{1}{x_i} - \beta^2\sum_{i=1}^{n}\frac{1}{2\,x_i^2} \\ & + \frac{\sum_{i=1}^{n}\log[\Phi(t_i)]\{n - n\log[\Phi(\alpha)] + \sum_{i=1}^{n}\log[\Phi(t_i)]\} - n^2\log[\Phi(\alpha)]}{n\log[\Phi(\alpha)] - \sum_{i=1}^{n}\log[\Phi(t_i)]}. \qquad (3.17)\end{aligned}$$

The MLEs of α and β can be also determined by maximizing the profile log-likelihood function (3.17) with respect to α and β.

It is possible note that the maximization of (3.17) is more simpler than the maximization of (3.14) with respect to their respective parameters, since the number of parameters of the profile log-likelihood is smaller, only two, when compared with the number of parameters of the log-likelihood function, which are three. However, some of the properties already known for the likelihood function are not valid for the case of the profile log-likelihood, since this is not a real likelihood function.

For interval estimation and hypothesis tests on the model parameters, we require the 3×3 observed information matrix $J = J(\boldsymbol{\theta})$ given in the Appendix. Under general regularity conditions, $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal $N_3(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. In practice, we can substitute $I(\boldsymbol{\theta})$ by the observed information matrix evaluated at $\hat{\boldsymbol{\theta}}$, say $J(\hat{\boldsymbol{\theta}})$. The asymptotic multivariate normal $N_3(0, J(\hat{\boldsymbol{\theta}})^{-1})$ distribution of $\hat{\boldsymbol{\theta}}$ can be used to construct approximate confidence intervals for the parameters.

3.5 Simulation study and applications

In this section, we perform a Monte Carlo simulation study and we provide three applications to real data to illustrate the flexibility of the Exp-Alpha distribution.

3.5.1 Simulation study

Here, we provide a Monte Carlo simulation study based on 1,000 replications to evaluate the MLEs of the parameters of the Exp-Alpha distribution. Setting $\alpha = 1$, $\beta = 2$ and three different values of the additional parameter a ($a = 2, 4, 6$), we evaluate the relative bias (RB) and the mean squared errors (MSE) for all point estimators based on artificially generated samples of sizes $n = 20, 40, 60, 80$ and 100. The RB is defined by $100\times$(bias/real value of the parameter)%.

The results are given in Table 3.1 and we note that, in general, the empirical biases and mean squared errors decrease when the sample size increases, as expected.

Table 3.1: RB (%) and MSE of the MLE $\hat{\boldsymbol{\theta}}$ for the Exp-Alpha distribution with $\alpha = 1$ and $\beta = 2$.

a	$\widehat{a}$		$\widehat{\alpha}$		$\widehat{\beta}$	
n=20	RB (%)	MSE	RB (%)	MSE	RB (%)	MSE
2	4.2850	0.2722	-1.0400	0.0640	2.5450	0.0958
4	4.8275	1.0004	-1.9200	0.0499	3.4800	0.1131
6	3.7467	1.9971	-1.2700	0.0448	3.4550	0.1278
n=40	RB (%)	MSE	RB (%)	MSE	RB (%)	MSE
2	2.8350	0.1220	-1.2600	0.0318	1.7100	0.0431
4	2.4750	0.4619	-0.9600	0.0250	1.8300	0.0525
6	2.4967	1.0150	-1.0500	0.0222	1.9750	0.0566
n=60	RB (%)	MSE	RB (%)	MSE	RB (%)	MSE
2	2.0450	0.0775	-0.9500	0.0207	1.2700	0.0280
4	1.9175	0.2956	-0.9100	0.0165	1.4000	0.0333
6	1.4450	0.5734	-0.6200	0.0129	1.1900	0.0331
n=80	RB (%)	MSE	RB (%)	MSE	RB (%)	MSE
2	1.1800	0.0553	-0.4600	0.0157	0.6850	0.0200
4	0.5675	0.2007	0.1900	0.0116	0.5200	0.0227
6	1.1367	0.5096	-0.3900	0.0114	0.9300	0.0290
n=100	RB (%)	MSE	RB (%)	MSE	RB (%)	MSE
2	0.7350	0.0408	-0.1400	0.0119	0.4400	0.0148
4	0.9000	0.1799	-0.3100	0.0101	0.6700	0.0203
6	1.5017	0.4187	-0.8500	0.0095	1.1450	0.0235

3.5.2 Applications to real data

In this section, we provide three applications to real data sets to illustrate the flexibility of the Exp-Alpha distribution, with pdf given by (3.6). Here, we compare the Exp-Alpha model with other well-known distributions to fit real data set. These distributions and their probability density functions are listed in Table 3.2. It is important to emphasize that all these distributions have support in the positive real set and their parameters are also real positives, except the location parameter $\mu \in$ IR of the Log-Normal distribution.

The MLEs of the parameters are computed (as discussed in Section 3.4) and the goodness-of-fit statistics for Exp-Alpha model are compared with other distributions. We consider the Anderson-Darling (A*) and Cramér-von Misses (W*), which are described by Chen and Balakrishnan (1995), and Kolmogorov-Smirnov (KS) statistics. Are also calculated other measures of goodness of fit, such as: Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criteria (CAIC) and Hannan-Quinn Information Criterion (HQIC).

The W* and A* statistics are used to verify which distribution fits better to the data. Since the values of W* and A* are smaller for the Exp-Alpha model when compared with those values of the other models, the new model seems to be a very competitive model for the data. We use these statistics, where we have a random sample $(x_1, \ldots, x_n)$ with empirical distribution function $F_n(x)$ and require to test if the sample comes from a special distribution. The W^* and A^* statistics are given by

Table 3.2: Distribution, probability density function and reference.

Distribution	Probability density function	Reference
Exp-Birnbaum-Saunders (Exp-BS) (a,α,β)	$a\,(2\alpha e^{\alpha^2}\sqrt{2\pi\beta})^{-1}\,x^{-3/2}\,(x+\beta)\,\exp\left[-\frac{(x/\beta)-(x/\beta)^{-1}}{2\alpha^2}\right]\times$ $\left\{\Phi\left\{\alpha^{-1}[(x/\beta)^{1/2}-(x/\beta)^{-1/2}]\right\}\right\}^{a-1}$	Cordeiro and Lemonte, 2014[1]
Exp-Gompertz (α,β,a)	$\alpha\,\beta\,a\,\exp[\beta+\alpha\,x-\beta\,\exp(\alpha\,x)]\,\{1-\exp[-\beta(\exp(\alpha\,x)-1)]\}^{a-1}$	Abu-Zinadah and Aloufi, 2014
Exp-Lomax (a,λ,α)	$a,\lambda\alpha\,[1+\lambda\,x]^{-(\alpha-1)}\,\{1-[1+\lambda\,x]^{-\alpha}\}^{a-1}$	Abdul-Moniem and Abdel-Hameed, 2012
Exp-Weibull (a,λ,γ)	$a\frac{\gamma}{\lambda}\left(\frac{x}{\lambda}\right)^{\gamma-1}\exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\}\left\{1-\exp\left[-\left(\frac{x}{\lambda}\right)^{\gamma}\right]\right\}^{a-1}$	Mudholkar and Srivastava, 1993
Alpha (α,β)	$\frac{\beta}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)}\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}$	Salvia, 1985
Birnbaum-Saunders (BS) (α,β)	$(2\alpha e^{\alpha^2}\sqrt{2\pi\beta})^{-1}\,x^{-3/2}\,(x+\beta)\,\exp\left[-\frac{(x/\beta)-(x/\beta)^{-1}}{2\alpha^2}\right]$	Birnbaum and Saunders, 1969
Burr XII (α,β)	$\alpha\,\beta\,\frac{x^{\alpha-1}}{(1+x^{\alpha})^{\beta+1}}$	Burr, 1942
Exp-Half-Normal (σ,a)	$\frac{a\,\sqrt{2}}{\sigma\,\sqrt{\pi}}\,\exp\left\{-\frac{x^2}{2\,\sigma^2}\right\}\,\{2\,\Phi(x)-1\}^{a-1}$	Gupta et al., 1998; Cooray and Ananda, 2008
Exp-Lindley (a,θ)	$\frac{a\,\theta^2}{\theta+1}\,(1+x)\,e^{-\theta\,x}\left\{1-\frac{e^{-\theta\,x}(1+\theta+\theta\,x)}{1+\theta}\right\}^{a-1}$	Nadarajah et al., 2011
Gamma (k,θ)	$\frac{x^{k-1}\,e^{-x/\theta}}{\theta^k\,\Gamma(k)}$	Khodabin and Ahmabadi, 2010[2]
Gompertz (α,β)	$\alpha\beta\exp[\beta+\alpha\,x-\beta\exp(\alpha\,x)]$	Abu-Zinadah and Aloufi, 2014[3]
Inverse Gaussian (λ,μ)	$\left[\frac{\lambda}{2\pi\,x^3}\right]^{1/2}\,\exp\left\{-\frac{\lambda\,(x-\mu)^2}{2\,\mu^2\,x}\right\}$	Chhikara and Folks, 1989
Log-Normal (μ,σ)	$\frac{1}{\sqrt{2\pi}\sigma\,x}\,\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}$	Johnson, Kotz and Balakrishnan, 1994
Nadarajah-Haghighi (NH) (α,λ)	$\alpha\lambda\,(1+\lambda\,x)^{\alpha-1}\,\exp\{1-[1+\lambda\,x]^{\alpha}\}$	Nadarajah and Haghighi, 2011
Rayleigh II (μ,λ)	$2\lambda\,(x-\mu)e^{-\lambda(x-\mu)^2}$	Johnson, Kotz and Balakrishnan, 1994
Weibull (λ,γ)	$\frac{\gamma}{\lambda}\,\left(\frac{x}{\lambda}\right)^{\gamma-1}\,\exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\}$	Weibull, 1951
Chi-Squared (μ)	$\frac{1}{2^{\mu/2}\Gamma(\mu/2)}x^{\mu/2-1}\,e^{-x/2}$	Johnson, Kotz and Balakrishnan, 1994
Exponential (λ)	$\lambda\,e^{-\lambda\,x}$	Gupta and Kundu, 1999, 2001
Half-Normal (σ^2)	$\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\,\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Cooray and Ananda, 2008
Log-Logistic I (α)	$\frac{\alpha\,x^{\alpha-1}}{(1+x^{\alpha})^2}$	Singh, 1998[4]
Rayleigh I (σ)	$\frac{x}{\sigma^2}\,\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Hoffman and Karst, 1975

[1]with $b=1$
[2]with new parametrization and $\tau=1$
[3]with $a=1$
[4]with $\beta=1$

$$W^* = \left\{ n \int_{-\infty}^{+\infty} \{F_n(x) - F(x;\hat{\theta}_n)\}^2 \, dF(x;\hat{\theta}_n) \right\} \left(1 + \frac{0.5}{n}\right) = W^2 \left(1 + \frac{0.5}{n}\right),$$

$$A^* = \left\{ n \int_{-\infty}^{+\infty} \frac{\{F_n(x) - F(x;\hat{\theta}_n)\}^2}{\{F(x;\hat{\theta})\,(1 - F(x;\hat{\theta}_n))\}} \, dF(x;\hat{\theta}_n) \right\} \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$$
$$= A^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right),$$

respectively, where $F_n(x)$ is the empirical distribution function and $F(x;\hat{\theta}_n)$ is the specified distribution function evaluated at the MLE $\hat{\theta}_n$ of θ. Note that the W^* and A^* statistics are given by difference distances of $F_n(x)$ and $F(x;\hat{\theta}_n)$. Thus, the lower are the W^* and A^* statistics more evidence we have that $F(x;\hat{\theta}_n)$ generates the sample.

The package `GenSA` available in the `R` programming language is used to obtain the initial values of the model parameters and the computations to the fit model are performed using the function `goodness.fit` of `AdequacyModel` package. The numerical BFGS (Broyden-Fletcher-Goldfarb-Shanno) procedure is used for minimization of the −log-likelihood function.

The data sets are:

(i) *Baboon data*

Wagner and Altmann (1973) report data from a study conducted in the Amboseli Reserve in Kenya on the time of the day at which members of a baboon troop descend from the trees in which they sleep. The time is defined as the time at which half of the troop has descended and begun that day's foraging. For more details about this data set, see Klein and Moeschberger (1997).

(ii) *Sitka size data*

This data set gives repeated measurements on the log-size (height times diameter squared) of 79 Sitka spruce trees, 54 of which were grown in ozone-enriched chambers and 25 of which were controls. The size was measured eight times in 1989, at roughly monthly intervals. These data, called 'Sitka89' in software `R`, are avaliable in package `MASS`.

(iii) *Survival times of animals data*

These data refers to survival times (in 10 hour units) of animals in a experiment. More details from this data set can be obtained in Hand et al. (1993).

Table 3.3 gives a descriptive summary of each data set, which includes central tendency statistics, standard deviation, among others. The distribution of the baboon and survival times of animals data have positive skewness (right skewed) and positive kurtosis. The sitka size data has negative skewness (left skewed) and positive kurtosis. We note that the considered data sets have both positive and negative skewness, whereas only positive values of the kurtosis were found. This fact can be confirmed in Section 3.3.1.

Table 3.3: Descriptive statistics.

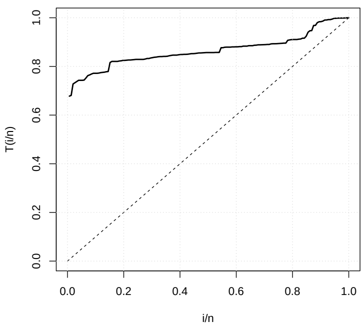
Statistic	Data		
	baboon	sitka size	animals
Mean	967.80	5.991	0.479
Median	858	6.060	0.400
Std. Dev.	281.51	0.704	0.252
Minimum	656	3.610	0.18
Maximum	1829	7.560	1.24
Skewness	1.81	−0.527	1.153
Kurtosis	2.08	0.061	0.626
n	152	632	48

One of the important device which can help selecting a particular model is the total time on test (TTT) plot (for more details see Aarset, 1987). This plot is constructed through the quantities

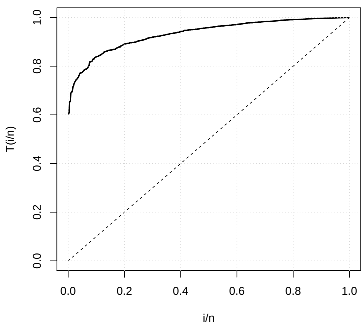
$$T(i/n) = \left[\sum_{j=1}^{i} X_{j:n} + (n-i)\, X_{i:n}\right] \Big/ \sum_{j=1}^{n} X_{j:n} \; versus \; i/n,$$

where $i = 1, \ldots, n$ and $X_{j:n}$ is the j-th order statistics of the sample (Mudholkar et al., 1995).

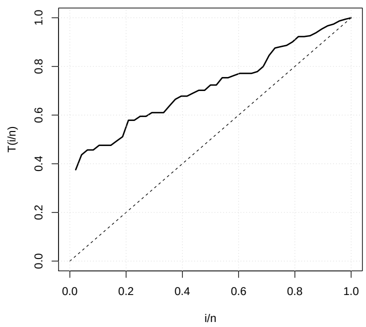
The Figure 3.3 shows the TTT plots for the baboon, sitka size and survival times of animals data. These plots indicate an increasing hrf and then reveals the adequacy of the Exp-Alpha distribution to fit these data.



(a)



(b)



(c)

Figure 3.3: TTT plots (a) baboon data; (b) sitka size data; (c) survival times of animals data.

Tables 3.4, 3.5 and 3.6 give the MLEs (with standard errors in parentheses) for some fitted models for baboon data, sitka size data and survival times of animals data, respectively. Table 3.7 lists the values of the W*, A*, KS, AIC, CAIC, BIC and HICQ for all fitted models for the considered three data sets. This table reveals that the Exp-Alpha model corresponds to the best fit for all data sets, even when compared to distributions (with one, two and three parameters) well-known in the literature for fit to real data set. Note that we used differents scenarios to the data set in concerning the compared distributions. This fact proves that, in differents situations, the proposed model have a upper performance to the other distributions analyzed.

Table 3.4: Maximum likelihood estimates (standard errors in parentheses) for the baboon data.

Distribution	MLEs (standard errors)		
Exp-Alpha (a, α, β)	382.227	4.576	1493.258
	(97.130)	(0.135)	(104.847)
Exp-BS (a, α, β)	97.198	227.070	0.003
	(24.852)	(12.232)	(0.0001)
Exp-Lomax (a, λ, σ)	987.383	2.765	0.012
	(359.422)	(0.164)	(0.001)
Exp-Weibull (a, λ, γ)	1262.253	0.517	0.322
	(298.775)	(0.160)	(0.015)
Alpha (α, β)	4.475	4081.848	
	(0.280)	(244.526)	
BS (α, β)	0.244	939.670	
	(0.014)	(18.515)	
Exp-Half-Normal (σ, a)	448.961	18.713	
	(20.231)	(3.903)	
Exp-Lindley (a, θ)	77.926	0.007	
	(29.977)	(0.0004)	
Gamma (k, θ)	15.314	0.015	
	(1.613)	(0.001)	
Inverse Gaussian (λ, μ)	15663.260	967.750	
	(1712.317)	(19.511)	
Log-Normal (mu, σ)	6.841	0.242	
	(0.019)	(0.013)	
Weibull (λ, γ)	1072.822	3.340	
	(27.804)	(0.185)	
Half-Normal (σ)	1000.002		
	(56.705)		
Log-Logistic I (α)	0.225		
	(0.014)		
Rayleigh I (σ)	712.48		
	(28.895)		

Table 3.5: Maximum likelihood estimates (standard errors in parentheses) for the sitka size data.

Distribution	MLEs (standard errors)		
Exp-Alpha (a, α, β)	0.021	29.789	221.865
	(0.005)	(3.363)	(25.806)
Exp-Lomax (a, λ, α)	719.902	125.468	0.009
	(107.046)	(3.459)	(0.0002)
Exp-Weibull (a, λ, γ)	1775.536	1.697	0.889
	(441.300)	(0.129)	(0.020)
Alpha (α, β)	7.620	44.958	
	(0.217)	(1.264)	
BS (α, β)	0.123	5.945	
	(0.003)	(0.029)	
Exp-Half-Normal (σ, a)	2.169	109.608	
	(0.034)	(12.594)	
Exp-Lindley (a, θ)	732.353	1.432	
	(135.584)	(0.036)	
Inverse Gaussian (λ, μ)	388.989	5.990	
	(21.882)	(0.029)	
Rayleigh II (μ, λ)	3.603	0.161	
	(0.005)	(0.006)	
Chi-Squared (μ)	6.919		
	(0.137)		
Exponential (λ)	0.166		
	(0.006)		
Log-Logistic I (α)	0.864		
	(0.026)		

Table 3.6: Maximum likelihood estimates (standard errors in parentheses) for the survival times of animals data.

Distribution	MLEs (standard errors)		
Exp-Alpha (a, α, β)	0.042	2.295	2.934
	(0.033)	(0.821)	(1.077)
Exp-Gompertz (α, β, a)	0.008	599.190	6.614
	(0.003)	(234.662)	(1.976)
Alpha (α, β)	2.085	0.812	
	(0.325)	(0.098)	
Burr XII (α, β)	2.343	4.923	
	(0.230)	(0.817)	
Exp-Half-Normal (σ, a)	0.384	2.625	
	(0.037)	(0.550)	
Gamma (k, θ)	4.307	8.986	
	(0.847)	(1.875)	
Gompertz (α, β)	2.210	0.403	
	(0.460)	(0.182)	
NH (α, λ)	645.929	0.002	
	(125.455)	(0.0003)	
Rayleigh II (μ, λ)	0.082	4.545	
	(0.039)	(0.917)	
Weibull (λ, γ)	0.544	2.060	
	(0.040)	(0.218)	
Chi-Squared (μ)	1.195		
	(0.150)		
Exponential (λ)	2.086		
	(0.301)		
Half-Normal (σ)	0.540		
	(0.055)		
Log-Logistic I (α)	1.656		
	(0.191)		
Rayleigh I (σ)	0.382		
	(0.027)		

Table 3.7: Goodness-of-fit statistics.

Distribution	W^*	A^*	KS	AIC	CAIC	BIC	HQIC
			Baboon data				
Exp-Alpha	**0.669**	**4.096**	**0.166**	**2022.472**	**2022.634**	**2031.544**	**2026.157**
Exp-BS	2.001	11.451	0.295	2110.355	2110.518	2119.427	2114.041
Exp-Lomax	1.480	8.676	0.326	2149.798	2149.960	2158.869	2153.483
Exp-Weibull	1.745	10.102	0.314	2170.152	2170.314	2179.223	2173.837
Alpha	1.242	7.391	0.196	2042.278	2042.358	2048.326	2044.735
BS	2.190	12.413	0.249	2085.255	2085.335	2091.303	2087.712
Exp-Half-Normal	2.248	12.709	0.246	2092.018	2092.098	2098.065	2094.475
Exp-Lindley	1.560	9.087	0.196	2057.138	2057.218	2063.185	2059.594
Gamma	2.533	14.164	0.262	2103.842	2103.923	2109.890	2106.299
Inverse Gaussian	2.179	12.358	0.248	2084.725	2084.806	2090.773	2087.182
Log-Normal	2.160	12.275	0.243	2084.313	2084.394	2090.361	2086.770
Weibull	3.450	18.595	0.283	2159.566	2159.646	2165.614	2162.023
Half-Normal	2.912	16.028	0.506	2324.918	2324.945	2327.942	2326.147
Log-Logistic I	2.130	12.119	0.812	3121.695	3121.722	3124.719	3122.923
Rayleigh I	2.925	16.090	0.373	2219.845	2219.872	2222.869	2221.073
			Sitka size data				
Exp-Alpha	**0.694**	**4.580**	0.086	**1378.434**	**1378.472**	**1391.780**	**1383.617**
Exp-Lomax	2.190	13.751	0.148	1572.675	1572.713	1586.021	1577.858
Exp-Weibull	2.375	14.832	0.146	1568.763	1568.801	1582.109	1573.946
Alpha	2.016	12.730	0.104	1494.123	1494.142	1503.020	1497.578
BS	1.041	6.806	0.080	1408.677	1408.696	1417.574	1412.132
Exp-Half-Normal	1.668	10.643	0.090	1469.040	1469.060	1477.938	1472.496
Exp-Lindley	2.381	14.869	0.110	1536.512	1536.531	1545.410	1539.967
Inverse Gaussian	1.043	6.818	0.080	1408.850	1408.869	1417.748	1412.305
Rayleigh II	1.350	8.651	0.222	1682.724	1682.743	1691.622	1686.180
Chi-Squared	0.803	5.310	0.383	2738.493	2738.500	2742.942	2740.221
Exponential	0.800	5.290	0.498	3528.822	3528.829	3533.271	3530.550
Log-Logistic I	1.119	7.294	0.765	4880.325	4880.331	4884.773	4882.052
			Survival times of animals data				
Exp-Alpha	**0.041**	**0.326**	0.094	**−11.425**	**−10.880**	−5.812	**−9.304**
Exp-Gompertz	0.099	0.600	0.112	−8.653	−8.107	−3.039	−6.531
Alpha	0.067	0.507	**0.080**	−10.614	−10.348	**−6.872**	−9.200
Burr XII	0.180	1.042	0.144	−4.905	−4.638	−1.162	−3.940
Exp-Half-Normal	0.197	1.136	0.166	−4.402	−4.135	−0.659	−2.987
Gamma	0.139	0.813	0.136	−8.340	−8.073	−4.597	−6.925
Gompertz	0.337	1.934	0.191	9.346	9.613	13.088	10.760
NH	0.254	1.460	0.264	12.717	12.983	16.459	14.131
Rayleigh II	0.177	1.023	0.190	−5.518	−5.251	−1.776	−4.104
Weibull	0.219	1.264	0.160	−2.506	−2.239	1.236	−1.091
Chi-Squared	0.124	0.734	0.339	70.301	70.388	72.173	71.009
Exponential	0.138	0.808	0.333	27.413	27.500	29.285	28.121
Half-Normal	0.210	1.212	0.281	12.654	12.741	14.525	13.361
Log-Logistic I	0.110	0.662	0.486	59.220	59.307	61.091	59.927
Rayleigh I	0.215	1.237	0.151	−4.427	−4.340	−2.556	−3.720

3.6 Concluding remarks

We introduce the Exponentiated Alpha (Exp-Alpha) distribution, which generalizes the Alpha distribution discussed by Salvia (1985). The new distribution is constructed using the exponentiated class (Gompertz, 1825; Verhulst, 1838, 1845, 1847), which is widely applied to baseline distributions in applications to real data set because this class provides a greater flexibility. The Exp-Alpha model has two shape parameters, one scale parameter and includes as special model the Alpha distribution. The Exp-Alpha hrf takes various forms depending on its shape parameters. We provide some mathematical properties of the new distribution including explicit expressions for the ordinary moments, generating function, mean deviations, density function, moments and generating function of the order statistics. These results are valid for certain regions of the parametric space. We estimate the model parameters by maximum likelihood method and provide a semiclosed estimator for the additional parameter. Besides that, a study simulation was performed to evaluated the asymptotic properties of the maximum likelihood estimates. Three applications to real data are provided to indicate that the new model best fits real data set when compared to well-known competitive distributions in the literature.

Appendix

The elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters (a, α, β) are

$$J_{a,a} = -\,\frac{n}{a^2}, \quad J_{a,\alpha} = -\frac{n\,\phi(\alpha)}{\Phi(\alpha)} + \sum_{i=1}^{n} \frac{\phi(t_i)}{\Phi(t_i)}, \quad J_{a,\beta} = -\sum_{i=1}^{n} \frac{\phi(t_i)}{x_i\,\Phi(t_i)},$$

$$J_{\alpha,\alpha} = -\,n\left\{1 + a\frac{[\dot{\phi}(\alpha)]_{\alpha}\,\Phi(\alpha) - \phi(\alpha)^2}{\Phi(\alpha)^2}\right\} + (a-1)\sum_{i=1}^{n} \frac{[\dot{\phi}(t_i)]_{\alpha}\,\Phi(t_i) - \phi(t_i)^2}{\Phi(t_i)^2},$$

$$J_{\alpha,\beta} = -\,(a-1)\sum_{i=1}^{n} \frac{[\dot{\phi}(t_i)]_{\beta}\,\Phi(t_i) - \phi(t_i)^2}{x_i\,\Phi(t_i)^2} + \sum_{i=1}^{n} \frac{1}{x_i},$$

$$J_{\beta,\beta} = -\,\frac{n}{\beta^2} - \sum_{i=1}^{n} \frac{1}{x_i^2} - (a-1)\sum_{i=1}^{n} \left\{\frac{[\dot{\phi}(t_i)]_{\beta}}{x_i\,\Phi(t_i)} + \frac{1}{x_i^2}\left[\frac{\phi(t_i)}{\Phi(t_i)}\right]^2\right\},$$

where $t_i = \left(\alpha - \frac{\beta}{x_i}\right)$, $[\dot{\phi}(\alpha)]_{\alpha} = \frac{\partial\,\phi(\alpha)}{\partial\,\alpha}$, $[\dot{\phi}(t_i)]_{\alpha} = \frac{\partial\,\phi(t_i)}{\partial\,\alpha}$ and $[\dot{\phi}(t_i)]_{\beta} = \frac{\partial\,\phi(t_i)}{\partial\,\beta}$.

Chapter 4

The Kumaraswamy Alpha Distribution

Resumo

Para qualquer distribuição de origem G contínua, Cordeiro e de Castro (2011) propuseram a família de distribuições Kumaraswamy (Kw-G) com dois parâmetros positivos adicionais. Dentro desta classe, definimos um modelo com quatro parâmetros denominado distribuição Kumaraswamy Alpha (Kw-Alpha), que é uma extensão da distribuição Alpha (Katsev, 1968; Wager e Barash, 1971; Sherif, 1983; Salvia, 1985). Fornecemos gráficos da função densidade de probabilidade e função taxa de falha. Obtemos expressões explícitas para os desvios médios, função geratriz de momentos e momentos ordinários. Determinamos a função densidade de probabilidade das estatísticas de ordem. O método de máxima verossimilhança é utilizado para estimar os parâmetros do modelo e obtemos um estimador em forma semifechada para um dos parâmetros adicionais. Este procedimento é avaliado por meio de um estudo de simulação de Monte Carlo. Fornecemos duas aplicações a dados reais para ilustrar a utilidade do novo modelo.

Palavras-chave: Desvios médios. Distribuição Alpha. Estatísticas de ordem. Função quantílica. Kumaraswamy-G. Máxima verossimilhança. Momentos.

Abstract

For any continuous baseline G distribution, Cordeiro and de Castro (2011) proposed the Kumaraswamy (Kw-G) family of distributions with two extra positive parameters. We define a four-parameter model within this class so-called the Kumaraswamy Alpha (Kw-Alpha) distribution, which is an extension of the Alpha distribution (Katsev, 1968; Wager and Barash, 1971; Sherif, 1983; Salvia, 1985). We provide plots of the probability density function and hazard rate function. Explicit expressions for the mean deviations, moment generating function, ordinary moments and quantile function are obtained. We determine the probability density function of the order statistics. The method of maximum likelihood is used to estimate the model parameters and we obtain a semiclosed estimator for one of the additional parameters. This procedure is assessed through a Monte Carlo simulation study. We provide two applications to real data to illustrate the usefulness of the new model.

Keywords: Alpha Distribution. Kumaraswamy class. Maximum Likelihood Estimation. Mean deviations. Moments. Order statistics. Quantile function.

4.1 Introduction

In recent years, some different generalizations of distributions have received increased attention in the literature. Generalizing new families of distributions to extend well-known distributions provide great flexibility in modeling real data, which can be applied in several fields. For example, we can refer some papers: Eugene et al. (2002) for the *beta* class, Zografos and Balakrishnan (2009) for the *gamma* class and the more recent ones by Alexander et al. (2012) and Cordeiro et al. (2013), who defined the generalized beta-generated and *exponentiated generalized* classes, respectively. In a similar manner, for any baseline cumulative distribution function (cdf) $G(x)$, Cordeiro and de Castro (2011) proposed the *Kumaraswamy* (Kw) class of distributions with two additional shape parameters $(a, b > 0)$ and cdf $F(x)$ and probability density function (pdf) $f(x)$ given by

$$F(x; a, b) = 1 - [1 - G^a(x)]^b \tag{4.1}$$

and

$$f(x; a, b) = ab\, g(x)\, G^{a-1}(x)\, [1 - G^a(x)]^{b-1}, \tag{4.2}$$

respectively, where $g(x) = dG(x)/dx$. We note that there is no complicated function in equation (4.1) in contrast with the beta generalized family (Eugene et al., 2002), which also includes two extra parameters but involves the incomplete beta function.

In this chapter, we study the so-called Kumaraswamy Alpha (Kw-Alpha, for short) distribution, which is provided combining the Kw-G family with the Alpha model presented in Chapter 2. A random variable X has an Alpha distribution with shape $\alpha > 0$ and scale $\beta > 0$ parameters, if its cdf $G(x)$ and pdf $g(x)$ are given by

$$G(x; \alpha, \beta) = \frac{\Phi\left(\alpha - \frac{\beta}{x}\right)}{\Phi(\alpha)} \tag{4.3}$$

and

$$g(x; \alpha, \beta) = \frac{\beta}{\sqrt{2\pi}\, x^2\, \Phi(\alpha)} \exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\}, \tag{4.4}$$

respectively, for $x > 0$ and $\Phi(\cdot)$ denotes the standard normal cumulative function.

In the last years, the Kw class has received wide attention of various authors and many papers are derived from this class proposed by Cordeiro and de Castro (2011). Table 4.1 lists some distributions that belong to this family.

The rest of this chapter is organized as follows. In Section 4.2, we define the Kw-Alpha distribution. Explicit expressions for cumulative and density functions, mean deviations, moment generating function, order statistics, ordinary moments and quantile function are derived in Section 4.3. Maximum likelihood estimation of the model parameters is investigated in Section 4.4. A simulation study and two applications to real data sets in Section 4.5 illustrate the

Table 4.1: Kumaraswamy distributions and respective references.

Distribution	Reference
Kumaraswamy exponential-Weibull	Cordeiro et al., 2016
Kumaraswamy Nadarajah-Haghighi	Lima, 2015
Kumaraswamy Gompertz	Silva et al., 2015
Kumaraswamy generalized Rayleigh	Gomes et al., 2014
Kumaraswamy Pareto	Bourguignon et al., 2013
Kumaraswamy Burr XII	Paranaíba et al., 2013
Kumaraswamy Generalized Half-Normal	Cordeiro et al., 2012a
Kumaraswamy Log-Logistic	Santana et al., 2012
Kumaraswamy Gumbel	Cordeiro et al., 2012b
Kumaraswamy modified Weibull	Cordeiro et al., 2012c
Kumaraswamy generalized gamma	de Pascoa et al., 2011
Kumaraswamy Weibull	Cordeiro et al., 2010

flexibility of the new distribution for data modeling. Finally, concluding remarks are given in Section 4.6.

4.2 The Kw-Alpha distribution

In this section, we define the Kw-Alpha distribution. The cdf and pdf of this model are given by

$$F(x;a,b,\alpha,\beta) = 1 - \left\{1 - \left[\frac{\Phi(\alpha - \frac{\beta}{x})}{\Phi(\alpha)}\right]^a\right\}^b \tag{4.5}$$

and

$$f(x;a,b,\alpha,\beta) = \frac{ab\beta\,\exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\}}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)^a}\,\Phi\left(\alpha - \frac{\beta}{x}\right)^{a-1}\left\{1 - \left[\frac{\Phi(\alpha - \frac{\beta}{x})}{\Phi(\alpha)}\right]^a\right\}^{b-1}, \tag{4.6}$$

respectively, where $a, b, \alpha, \beta > 0$.

The hazard rate function (hrf) corresponding to (4.6) is given by

$$h(x;a,b,\alpha,\beta) = \frac{ab\beta\,\exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\}\,\Phi\left(\alpha - \frac{\beta}{x}\right)^{a-1}}{\sqrt{2\pi}\,x^2\left[\Phi(\alpha)^a - \Phi\left(\alpha - \frac{\beta}{x}\right)^a\right]}.$$

A random variable X having density function (4.6) is denoted by $X \sim \text{Kw-Alpha}(a,b,\alpha,\beta)$. In this model, a, b, α are shape parameters and β is a scale parameter. We write $F(x) = F(x;a,b,\alpha,\beta)$ and $f(x) = f(x;a,b,\alpha,\beta)$ in order to eliminate the dependence on the model parameters. The Alpha distribution is clearly a special case of (4.5) when $a = b = 1$. Setting $b = 1$ gives the Exponentiated Alpha (Exp-Alpha) distribution. For $a = 1$, we have the Lehmann type II Alpha distribution corresponding to the cdf $F(x) = \{1 - [1 - G(x)]^b\}$.

Plots of the Kw-Alpha pdf and hrf for selected parameter values are displayed in Figures 4.1 and 4.2, respectively. It is evident that the additional shape parameters (a and b) allow for a high degree of flexibility of the Kw-Alpha distribution, i.e., this distribution is more flexible than the Alpha distribution, mainly with respect to the hrf. So, the proposed model can be very useful for modeling positive real data sets.

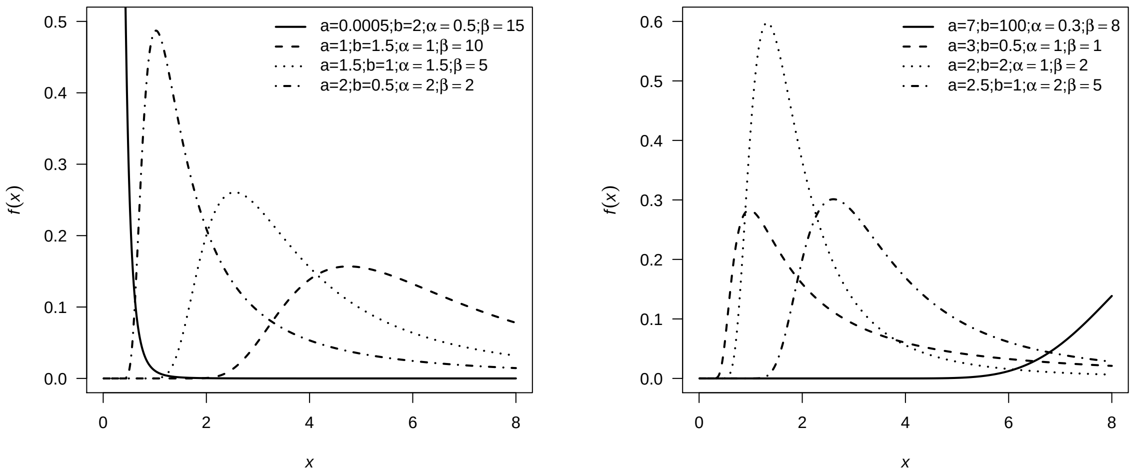


Figure 4.1: Plots of the Kw-Alpha density for some parameter values.

4.3 Properties

In this section, we provide some mathematical properties of the Kw-Alpha distribution in order to show interesting characteristics of the proposed model. It is important to emphasize that the results presented here are only valid for certain regions of the parametric space due to non existence of moments for the Alpha distribution.

4.3.1 A Useful Representation

For an arbitrary baseline cdf $G(x)$, a random variable has the Exponentiated-G (Exp-G) distribution with power parameter $a > 0$, say $Y \sim \text{Exp}^a(G)$, if its cdf and pdf are $H_a(x) = G(x)^a$ and $h_a(x) = a\,g(x)\,G(x)^{a-1}$, respectively.

We consider the generalized binomial expansion

$$(1-z)^b = \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} z^k, \tag{4.7}$$

which holds for any real non-integer b and $|z| < 1$. Using expansion (4.7) in equations (4.6) and (4.5), we can express the pdf and the cdf of the Kw-Alpha distribution as

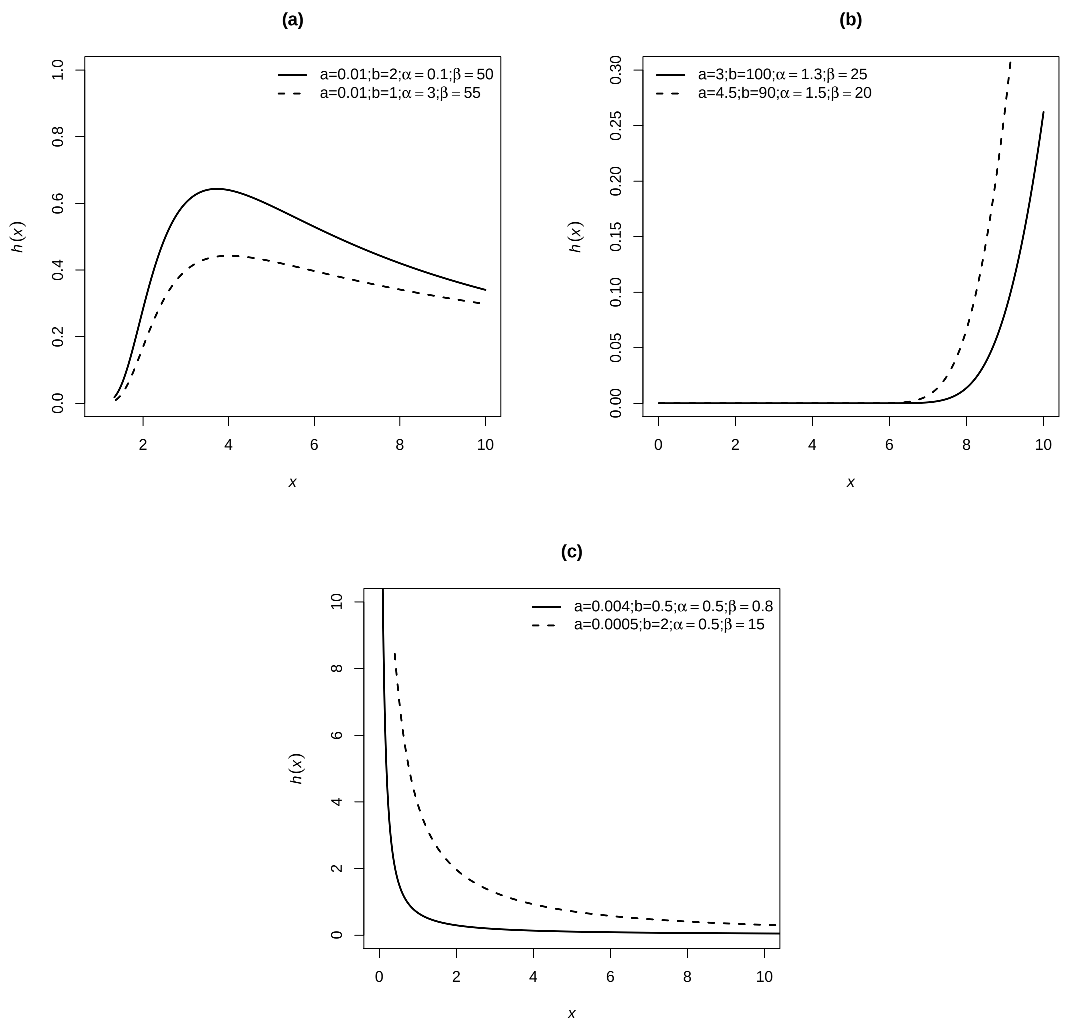


Figure 4.2: Plots of the Kw-Alpha hazard rate function for some parameter values. (a) Upside-down bathtub hazard rate function. (b) Increasing hazard rate function. (c) Decreasing hazard rate function.

$$f(x) = a^{-1} \sum_{k=0}^{\infty} \frac{w_k}{(k+1)}\, h_{(k+1)a}(x) \tag{4.8}$$

and

$$F(x) = 1 - \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} H_{ka}(x),$$

respectively, where $w_k = (-1)^k\, ab\, \binom{b-1}{k}$, $h_{(k+1)a}(x)$ is the Exp-Alpha pdf with power parameter $(k+1)a$ given by

$$h_{(k+1)a}(x) = \frac{\beta\,(k+1)\,a}{\sqrt{2\pi}\, x^2\, \Phi(\alpha)} \exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\} \left\{\frac{\Phi\left(\alpha - \frac{\beta}{x}\right)}{\Phi(\alpha)}\right\}^{(k+1)a-1} \tag{4.9}$$

and $H_{ka}(x) = G(x)^{ka}$ is the Exp-Alpha cdf with power parameter ka.

Equation (4.8) reveals that the Kw-Alpha density function is a linear combination of Exp-Alpha densities. This result is important to derive some structural properties of the new distribution.

4.3.2 Quantile function

The quantile function (qf) of X can be obtained by inverting the Kw-Alpha cdf (4.5) as

$$x_u = Q(u) = \frac{\beta}{\alpha - \Phi^{-1}\{\Phi(\alpha)\,[1-(1-u)^{1/b}]^{1/a}\}}, \quad u \in (0,1).$$

The median of X is simply $x_{1/2} = Q(1/2)$. Further, it is possible to generate Kw-Alpha variates by $X = Q(U)$, where U is a 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 new distribution can be based on quantile measures. In this sense, two important measures are the Bowley skewness (Kenney and Keeping, 1962) and the Moors kurtosis (Moors, 1988).

The Bowley skewness (B) is based on quartiles, while the Moors kurtosis (M) is based on octiles. These measures are given by

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

and

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

respectively.

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 (considering $\alpha = 20, \beta = 25$ for skewness and $\alpha = 1, \beta = 3$ for kurtosis) are displayed in Figure 4.3. These plots indicate that skewness increase when a increases for fixed b and decrease when b increases for fixed a. Further, the kurtosis increases and becomes constant when a increases for fixed b and decrease when b increases for fixed a.

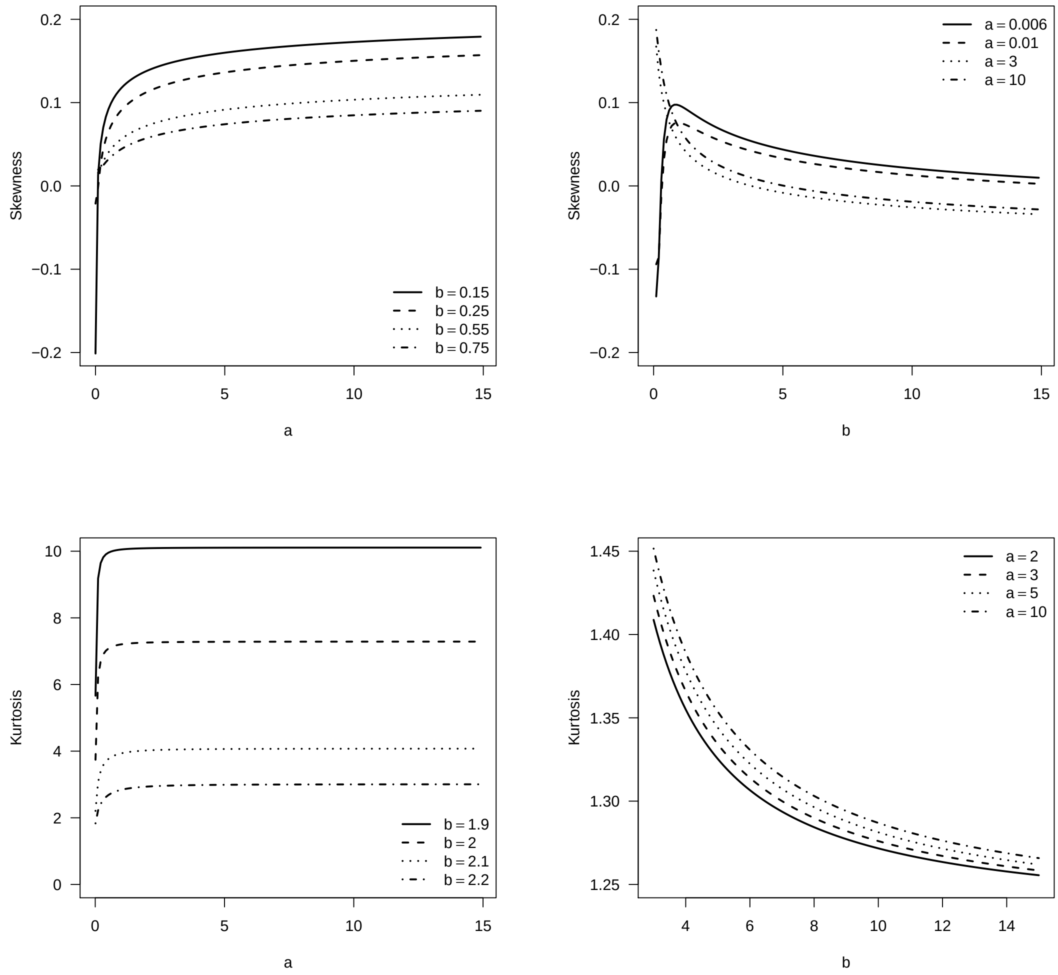


Figure 4.3: Plots of the Kw-Alpha skewness and kurtosis as functions of a for some values of b and as functions of b for some values of a.

4.3.3 Moments

The ordinary moments of X can be obtained in closed-form as an infinite weighted linear combination of the baseline probability weighted moments (PWMs). First, we review the PWMs of the Alpha distribution since they are required for the ordinary moments of the Kw-Alpha distribution. If $Z \sim \text{Alpha}(\alpha, \beta)$ has the pdf (4.4), the (s, r)th PWM of Z is formally defined by

$$\tau_{s,r} = E(Z^s\, G(x)^r) = \int_0^\infty x^s\, G(x)^r\, g(x)\, dx. \tag{4.10}$$

Substituting equations (4.3) and (4.4) in (4.10) and after some algebraic manipulations we obtain $\tau_{s,r}$ as

$$\tau_{s,r} = \frac{c_1}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \sum_{m=0}^{2s_j+j} A(k_1,\ldots,k_j) \binom{2s_j+j}{m} \alpha^{2s_j+j-m} (-\beta)^m \, I(s-m-2,\alpha,\beta), \tag{4.11}$$

where $c_1 = \frac{\beta}{\sqrt{2\pi}\,\Phi(\alpha)^{r+1}}$, $a_k = \frac{(-1)^k\, 2^{(1-2k)/2}}{\sqrt{\pi}(2k+1)\,k!}$, $A(k_1,\ldots,k_j) = a_{k_1}\ldots a_{k_j}$ and $s_j = k_1+\ldots+k_j$. Here, $I(p,\alpha,\beta)$ is the integral given by

$$\begin{aligned}
I(p,\alpha,\beta) &= \int_0^\infty x^p \exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\} dx \\
&= \frac{1}{\sqrt{\pi}}\left\{2^{(-3/2-p/2)}\,\beta^{(p+1)} \exp^{-\alpha^2/2}\left[-\frac{1}{2}\frac{\sqrt{2}\pi^2\alpha\beta\, L_{p/2}^{1/2}\left(\frac{\alpha^2}{2}\right)}{\sqrt{\beta^2}\sin\left(\frac{p}{2}\pi\right)\,\Gamma\left(\frac{3}{2}+\frac{p}{2}\right)}\right.\right. \\
&\quad \left.\left.-\frac{1}{2}\frac{\pi^2(\alpha^2+1)\, L_{p/2}^{1/2}\left(\frac{\alpha^2}{2}\right)}{\cos\left(\frac{p}{2}\pi\right)\,\Gamma\left(2+\frac{p}{2}\right)} + \frac{1}{2}\frac{\pi^2\alpha^2\, L_{p/2}^{3/2}\left(\frac{\alpha^2}{2}\right)}{\cos\left(\frac{1}{2}\pi\right)\,\Gamma\left(1+\frac{p}{2}\right)}\right]\right\},
\end{aligned}$$

where $L_{p/2}^{1/2}\left(\frac{\alpha^2}{2}\right)$, $\Gamma(\cdot)$, $\sin(\cdot)$ and $\cos(\cdot)$ are the Laguerre, gamma, sine and cosine functions, respectively.

Finally, the moments of X can be expressed from (4.8) as

$$\begin{aligned}
E(X^s) &= \int_0^\infty x^s\, f(x)\, dx \\
&= a^{-1}\sum_{k=0}^{\infty} \frac{w_k}{k+1} \int_0^\infty x^s\, h_{(k+1)a}(x)\, dx \\
&= \sum_{k=0}^{\infty} w_k\, \tau_{s,(k+1)a-1},
\end{aligned}$$

where $w_k = (-1)^k\, ab\binom{b-1}{k}$ and $\tau_{s,(k+1)a-1}$ is defined in (4.10).

Then, we obtain the moments of the Kw-Alpha distribution as an infinite linear combination of convenient PWMs of the Alpha distribution.

4.3.4 Generating function

In this section, we derive the moment generating function (mgf) $M(t) = E(e^{tX})$ of the Kw-Alpha distribution. The mgf of X can be obtained using the fact that the Kw-Alpha density function is a linear combination of Exp-Alpha densities. Thus,

$$\begin{aligned}
M(t) &= \int_0^\infty \mathrm{e}^{tx}\, a^{-1} \sum_{k=0}^{\infty} \frac{w_k}{k+1}\, h_{(k+1)a}(x)\, dx \\
&= \sum_{k=0}^{\infty} w_k \int_0^\infty \mathrm{e}^{tx}\, G^{(k+1)a-1}(x)\, g(x)\, dx,
\end{aligned}$$

where $w_k = (-1)^k\, ab\, \binom{b-1}{k}$.

Using the power series for the exponential function, the Kw-Alpha mgf reduces to

$$M(t) = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{w_k\, t^\ell}{\ell!}\, \tau_{\ell,(k+1)a-1},$$

where $\tau_{\ell,(k+1)a-1}$ is defined in (4.10).

4.3.5 Mean Deviations

If X has the Kw-Alpha distribution with cdf $F(x)$ given by (4.5), we can derive the mean deviations about the mean $\mu = E(X)$ and about the median M from the relations

$$\delta_1 = \int_0^\infty |x-\mu|\, f(x)\, dx \qquad \text{and} \qquad \delta_2 = \int_0^\infty |x-M|\, f(x)\, dx,$$

respectively.

Defining the first incomplete moment of X by $I(s) = \int_0^s x\, f(x)\, dx$, these measures can be determined from

$$\delta_1 = 2\,\mu\, F(\mu) - 2\, I(\mu) \qquad \text{and} \qquad \delta_2 = E(X) + 2\, M\, F(M) - M - 2\, I(M), \tag{4.12}$$

where $F(\mu)$ and $F(M)$ are easily obtained from equation (4.5).

Setting

$$\rho(s,r) = \int_0^s x\, G(x)^r\, g(x)\, dx,$$

we obtain from equation (4.8)

$$I(s) = \sum_{k=0}^{\infty} w_k \int_0^s x\, G^{(k+1)a-1}(x)\, g(x)\, dx = \sum_{k=0}^{\infty} w_k\, \rho(s,(k+1)a-1).$$

Then,

$$I(\mu) = \sum_{k=0}^{\infty} w_k\, \rho(\mu,(k+1)a-1).$$

Combining (4.3) and (4.11) and defining

$$J(s,-m-1) = \int_0^s x^{-m-1} \exp\left\{ -\frac{1}{2} \left(\alpha - \frac{\beta}{x} \right)^2 \right\} dx,$$

we can write

$$\rho(s,r)=\frac{\beta}{\sqrt{2\pi}\,2^r\,\Phi^{r+1}(\alpha)}\sum_{j=0}^{r}\binom{r}{j}\sum_{k_1,\ldots,k_j=0}^{\infty}\sum_{m=0}^{2s_j+j}A(k_1,\ldots,k_j)$$
$$\times\binom{2s_j+j}{m}\alpha^{2s_j+j-m}\,(-\beta)^m\,J(s,-m-1).$$

Letting $t=\alpha-\beta/x$, the last integral reduces to

$$J(s,-m-1)=\int_{-\infty}^{\alpha-\beta/s}\frac{1}{\beta^m}\,(\alpha-t)^{m-1}\,\mathrm{e}^{-t^2/2}\,dt$$

and using the binomial expansion and interchanging terms, we have

$$J(s,-m-1)=\sum_{\ell=0}^{\infty}\frac{(-1)^{\ell}\,\alpha^{m+1-\ell}}{\beta^{m+1}}\binom{m+1}{\ell}\int_{-\infty}^{\alpha-\beta/s}t^{\ell}\,\mathrm{e}^{-t^2/2}\,dt.$$

We now define

$$G(\ell)=\int_0^{\infty}x^{\ell}\,\mathrm{e}^{-x^2/2}dx=2^{(\ell-1)/2}\,\Gamma(\ell+1/2).$$

For evaluating the integral in $J(s,-m-1)$, it is required to consider two cases. If $\alpha-\beta/s<0$, we have

$$\int_{-\infty}^{\alpha-\beta/s}t^{\ell}\,\mathrm{e}^{-t^2/2}\,dt=(-1)^{\ell}\,G(\ell)+(-1)^{\ell+1}\int_0^{-(\alpha-\beta/s)}t^{\ell}\,\mathrm{e}^{-t^2/2}\,dt.$$

If $\alpha-\beta/s>0$, we have

$$\int_{-\infty}^{\alpha-\beta/s}t^{\ell}\,\mathrm{e}^{-t^2/2}\,dt=(-1)^{\ell}\,G(\ell)+\int_0^{\alpha-\beta/s}t^{\ell}\,\mathrm{e}^{-t^2/2}\,dt.$$

Further, the integrals of the type $\int_0^q x^{\ell}\,\mathrm{e}^{-x^2/2}\,dx$ can be calculated easily as

$$\int_0^q x^{\ell}\,\mathrm{e}^{-x^2/2}\,dx=\frac{2^{\ell/4+1/4}\,q^{\ell/2+1/2}\,\mathrm{e}^{-q^2/4}}{(\ell/2+1/2)(\ell+3)}\,M_{\ell/4+1/4,\ell/4+3/4}(q^2/4)$$
$$+\frac{2^{\ell/4+1/4}\,q^{\ell/2-3/2}\,\mathrm{e}^{-q^2/4}}{\ell/2+1/2}\,M_{\ell/4+5/4,\ell/4+3/4}(q^2/4),$$

where $M_{k,m}(x)$ is the Whittaker function, which can be expressed in terms of the confluent hypergeometric function ${}_1F_1$ by $M_{k,m}(x)=\mathrm{e}^{-x^2/2}\,x^{m+1/2}\,{}_1F_1(\frac{1}{2}+m-k;1+2m;x)$, where ${}_1F_1(\alpha_1;\beta_1;x)=\sum_{k=0}^{\infty}\frac{(\alpha_1)_k}{(\beta_1)_k}\frac{x^k}{k!}$, for $\alpha_1>0$, $\beta_1>0$.

Hence, we have all quantities to obtain $J(s,-m-1)$, $\rho(s,r)$, $I(\mu)$ and then the mean deviations δ_1 and δ_2 in (4.12).

4.3.6 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 size n from the Kw-Alpha distribution. It is well-known that

$$
\begin{aligned}
f_{i:n}(x) &= \frac{f(x)}{B(i,n-i+1)} F(x)^{i-1}[1-F(x)]^{n-i} \\
&= \frac{1}{B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x) F(x)^{i+j-1}, \qquad (4.13)
\end{aligned}
$$

where $B(\cdot)$ is the beta function.

Substituting (4.1) and (4.2) in equation (4.13), we can write

$$
\begin{aligned}
f_{i:n}(x) = \frac{ab}{B(i,n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} G^{a-1}(x)\,[1-G^a(x)]^{b-1} \\
\times \{1-[1-G^a(x)]^b\}^{i+j-1}\, g(x). \qquad (4.14)
\end{aligned}
$$

Applying the binomial expansion (4.7) twice in equation (4.14) gives

$$
f_{i:n}(x) = \frac{ab}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} \delta_\ell \, h_{(\ell+1)a}(x), \qquad (4.15)
$$

where

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

and $h_{(\ell+1)a}(x)$ is given by (4.9).

Equation (4.15) reveals that the density function of $X_{i:n}$ is a linear combination of Exp-Alpha densities. A direct application of (4.15) is to obtain the moments and the mgf of the Kw-Alpha order statistics.

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

$$
E(X_{i:n}^r) = \frac{ab}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} \delta_\ell^* \, \tau_{r,(\ell+1)a-1},
$$

where $\delta_\ell^* = [(\ell+1)a]\,\delta_\ell$ and $\tau_{r,(\ell+1)a-1}$ is defined in (4.10).

Finally, the mgf of $X_{i:n}$ can be written as

$$
M(t) = \frac{ab}{B(i,n-i+1)} \sum_{\ell=0}^{\infty} \sum_{r=0}^{\infty} \frac{\delta_\ell^* \, t^r}{r!} \, \tau_{r,(\ell+1)a-1}.
$$

4.4 Estimation

Here, we provide an analytic procedure to determine the maximum likelihood estimators (MLEs) for the Kw-Alpha parameters. The MLEs are employed due to their interesting asymptotic properties.

Let $(x_1, \ldots, x_n)$ be a random sample of size n from the Kw-Alpha distribution with unknown parameter vector $\boldsymbol{\theta} = (a, b, \alpha, \beta)^\top$. The log-likelihood function for $\boldsymbol{\theta}$ based on a given sample can be expressed as

$$\begin{aligned}\ell(\boldsymbol{\theta}) =& n\,\log\left(\frac{ab\beta}{\sqrt{2\pi}}\right) - n\,a\log[\Phi(\alpha)] - 2\sum_{i=1}^{n}\log(x_i) - \frac{n\alpha^2}{2} + \alpha\beta\sum_{i=1}^{n}\frac{1}{x_i} - \frac{\beta^2}{2}\sum_{i=1}^{n}\frac{1}{x_i^2}\\ &+ (a-1)\sum_{i=1}^{n}\log[\Phi(t_i)] + (b-1)\sum_{i=1}^{n}\log\left\{1 - \left[\frac{\Phi(t_i)}{\Phi(\alpha)}\right]^a\right\},\end{aligned} \tag{4.16}$$

where $t_i = \left(\alpha - \frac{\beta}{x_i}\right)$, $i = 1, \ldots, n$.

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

$$U_a(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n}\log\left[\frac{\Phi(t_i)}{\Phi(\alpha)}\right]\left\{1 - (b-1)\,\frac{\Phi(t_i)^a}{\Phi(\alpha)^a - \Phi(t_i)^a}\right\},$$

$$U_b(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n}\log\left\{1 - \left[\frac{\Phi(t_i)}{\Phi(\alpha)}\right]^a\right\}, \tag{4.17}$$

$$\begin{aligned}U_\alpha(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\alpha} =& -n\,\alpha - n\,a\frac{\phi(\alpha)}{\Phi(\alpha)} + \beta\sum_{i=1}^{n}\frac{1}{x_i} + (a-1)\sum_{i=1}^{n}\frac{\phi(t_i)}{\Phi(t_i)}\\ &- a(b-1)\sum_{i=1}^{n}\frac{\Phi(t_i)^a}{\Phi(\alpha)^a - \Phi(t_i)^a}\left[\frac{\phi(t_i)}{\Phi(t_i)} - \frac{\phi(\alpha)}{\Phi(\alpha)}\right],\end{aligned}$$

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

The MLEs, $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{\alpha}, \hat{\beta})^\top$, can be obtained numerically by solving $U(\hat{\boldsymbol{\theta}}) = 0$.

It is possible note that from (4.17) we can obtain a semiclosed estimator of b. For $\frac{\partial\ell(\boldsymbol{\theta})}{\partial b} = 0$, the estimator for b is given by

$$\hat{b}_{\hat{a},\hat{\alpha},\hat{\beta}} = -\frac{n}{\sum_{i=1}^{n} \log\left\{1-\left[\frac{\Phi(t_i^*)}{\Phi(\hat{\alpha})}\right]^{\hat{a}}\right\}}, \tag{4.18}$$

where $t_i^* = (\hat{\alpha} - \frac{\hat{\beta}}{x_i})$.

Another way to obtain the MLEs is using the profile log-likelihood. We suppose b fixed and rewrite the log-likelihood (4.16) as $\ell(\boldsymbol{\theta}) = \ell_b(a,\alpha,\beta)$. In this manner, we show that b is fixed but $\boldsymbol{\theta}_b = (a,\alpha,\beta)^\top$ varies. To estimate $\boldsymbol{\theta}_b$ we maximize $\ell_b(a,\alpha,\beta)$ with respect to $\boldsymbol{\theta}_b$, i.e.,

$$\hat{\boldsymbol{\theta}}_b = \arg\max_{\boldsymbol{\theta}_b} \ell_b(a,\alpha,\beta).$$

The estimative of b can be obtained analytically by (4.18). By replacing b by $\hat{b}_{a,\alpha,\beta}$ in (4.16), we can obtain the profile log-likelihood function for a, α and β as

$$\ell_b(a,\alpha,\beta) = n\left\{-1-\frac{\alpha}{2} - a\log[\Phi(\alpha)] + \log\left[-\frac{n\,a\,\beta}{\sqrt{2\pi}\sum_{i=1}^{n}\log\left\{1-\left[\frac{\Phi(t_i)}{\Phi(\alpha)}\right]^{a}\right\}}\right]\right\}$$
$$-2\sum_{i=1}^{n}\log(x_i) + \beta\sum_{i=1}^{n}\frac{1}{2\,x_i} + (a-1)\sum_{i=1}^{n}\log[\Phi(t_i)] - \sum_{i=1}^{n}\log\left\{1-\left[\frac{\Phi(t_i)}{\Phi(\alpha)}\right]^{a}\right\}. \tag{4.19}$$

Hence, the MLEs of a, α and β can be also determined by maximizing the profile log-likelihood function (4.19) with respect to a, α and β.

We can see that the maximization of (4.19) is more simpler than the maximization of (4.16) with respect to their respective parameters. This fact is due to the number of parameters to the profile log-likelihood function, which is smaller than the number of parameters to the log-likelihood function. However, some of the new properties that hold for a true likelihood do not hold for its profile version, because the profile log-likelihood is not a real likelihood function.

For interval estimation and hypothesis tests on the model parameters, we require the 4×4 observed information matrix $J = J(\boldsymbol{\theta})$ given in the Appendix. Under general regularity conditions, $(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is asymptotically normal $N_4(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. In practice, we can substitute $I(\boldsymbol{\theta})$ by the observed information matrix evaluated at $\hat{\boldsymbol{\theta}}$, say $J(\hat{\boldsymbol{\theta}})$. The asymptotic multivariate normal $N_4(0, J(\hat{\boldsymbol{\theta}})^{-1})$ distribution of $\hat{\boldsymbol{\theta}}$ can be used to construct approximate confidence intervals for the parameters.

4.5 Simulation study and applications

In this section, we perform a Monte Carlo simulation study and we provide two applications to real data to illustrate the flexibility of the Kw-Alpha distribution.

4.5.1 Simulation study

Here, we provide a Monte Carlo simulation study (with $1,000$ replications) to quantify bias of the MLEs of the Kw-Alpha parameters. The simulations are based on the following scenarios:

(i) $a = (0.5, 1, 2), b = \alpha = \beta = 1$ and

(ii) $b = (1, 2, 4), a = \alpha = \beta = 1$.

We chose to vary the values of parameters a and b since they are additional parameters of the proposed distribution.

We evaluate the means and standard deviations of the bias for all point estimators based on artificially generated samples of sizes $n = 10, 20, \dots, 100$. These results are given for every pair (a, n) and (b, n) in Figure 4.4 and Figure 4.5, respectively. The plots indicate that the means and standard deviations of the bias decrease when the sample increases as expected.

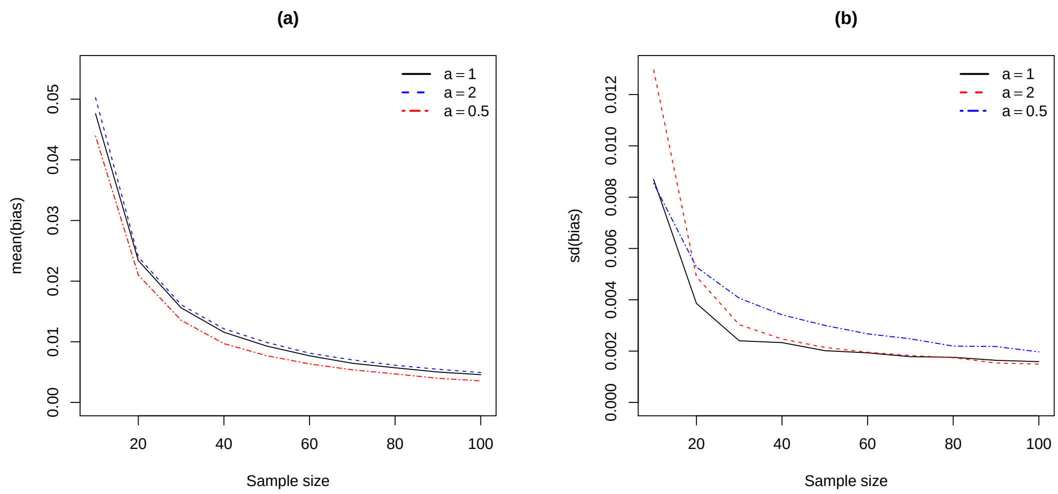


Figure 4.4: (a) Mean of the bias for $b = \alpha = \beta = 1$. (b) Standard deviation of the bias for $b = \alpha = \beta = 1$.

4.5.2 Applications to real data

Two applications to real data sets are used to illustrate the flexibility of the Kw-Alpha distribution, with pdf given by (4.6). Here, we compare the Kw-Alpha model with competitive distributions well-known to fit real data set. These distributions and their probability density functions are listed in Table 4.2. It is important to emphasize that all these distributions have support in the positive real set and their parameters are also real positives, except the location parameter $\mu \in \mathrm{IR}$ of the Log-Normal distribution.

The MLEs of the parameters are computed (as discussed in Section 4.4) and the goodness-of-fit statistics for Kw-Alpha model are compared with other distributions, as mencioned above. We consider the Anderson-Darling (A*) and Cramér-von Misses (W*), which are described by Chen and Balakrishnan (1995), and Kolmogorov-Smirnov (KS) statistics. Are also calculated other

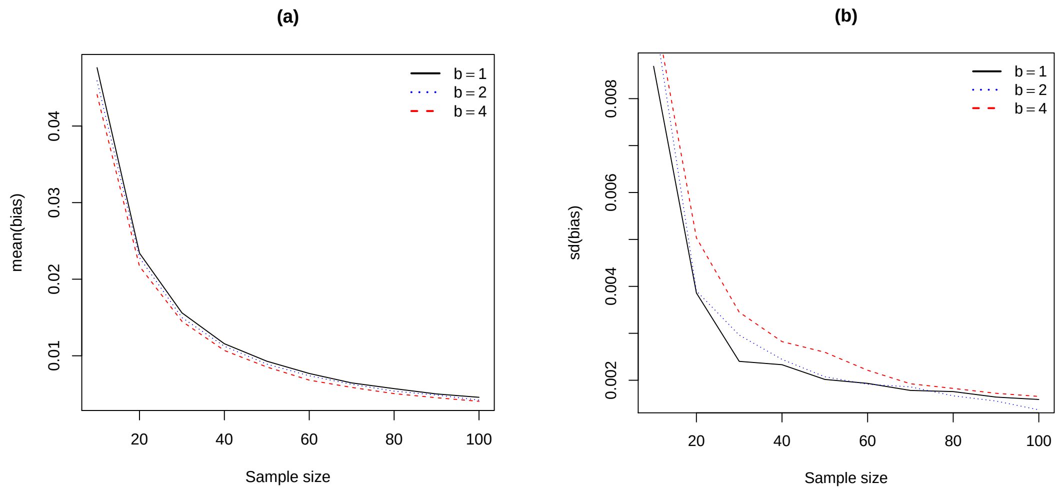


Figure 4.5: (a) Mean of the bias for $a = \alpha = \beta = 1$. (b) Standard deviation of the bias for $a = \alpha = \beta = 1$.

measures of goodness of fit, such as: Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criteria (CAIC) and Hannan-Quinn Information Criterion (HQIC).

The W* and A* statistics are used to verify which distribution fits better to the data. Since the values of W* and A* are smaller for the Kw-Alpha model when compared with those values of the other models, the new model seems to be a very competitive model for the data. We use these statistics, where we have a random sample $(x_1, \ldots, x_n)$ with empirical distribution function $F_n(x)$ and require to test if the sample comes from a special distribution. The W^* and A^* statistics are given by

$$W^* = \left\{ n \int_{-\infty}^{+\infty} \{F_n(x) - F(x; \hat{\theta}_n)\}^2 \, dF(x; \hat{\theta}_n) \right\} \left(1 + \frac{0.5}{n}\right) = W^2 \left(1 + \frac{0.5}{n}\right),$$

$$A^* = \left\{ n \int_{-\infty}^{+\infty} \frac{\{F_n(x) - F(x; \hat{\theta}_n)\}^2}{\{F(x; \hat{\theta})\,(1 - F(x; \hat{\theta}_n))\}} \, dF(x; \hat{\theta}_n) \right\} \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$$
$$= A^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right),$$

respectively, where $F_n(x)$ is the empirical distribution function and $F(x; \hat{\theta}_n)$ is the specified distribution function evaluated at the MLE $\hat{\theta}_n$ of θ. Note that the W^* and A^* statistics are given by difference distances of $F_n(x)$ and $F(x; \hat{\theta}_n)$. Thus, the lower are the W^* and A^* statistics more evidence we have that $F(x; \hat{\theta}_n)$ generates the sample.

The package `GenSA` available in the `R` programming language is used to obtain the initial values of the model parameters and the computations to the fit model are performed using the function `goodness.fit` of `AdequacyModel` package. The numerical BFGS (Broyden-Fletcher-Goldfarb-Shanno) procedure is used for minimization of the −log-likelihood function.

Table 4.2: Distribution, probability density function and reference.

Distribution	Probability density function	Reference
EG-Inverse Gaussian (λ, μ, a, b)	$a\,b\left[1-\Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right)\right]^{a-1}\left\{1-\left[1-\Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right)\right]\right\}^{b-1}$	Lemonte and Cordeiro, 2011
Beta-Exponential (a, b, λ)	$\frac{\lambda}{B(a,b)}\,e^{-b\,\lambda\,x}\,[1-e^{-\lambda\,x}]^{a-1}$	Nadarajah and Kotz, 2006
Exp-Alpha (a, α, β)	$\frac{a\beta\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)^a}\,\Phi\left(\alpha-\frac{\beta}{x}\right)^{a-1}$	Capítulo 3
Exp-Gompertz (α, β, a)	$\alpha\,\beta\,a\,\exp[\beta+\alpha\,x-\beta\,\exp(\alpha\,x)]\,\{1-\exp[-\beta(\exp(\alpha\,x)-1)]\}^{a-1}$	Abu-Zinadah and Aloufi, 2014
Exp-Weibull (a, λ, γ)	$a\frac{\gamma}{\lambda}\left(\frac{x}{\lambda}\right)^{\gamma-1}\exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\}\left\{1-\exp\left[-\left(\frac{x}{\lambda}\right)^{\gamma}\right]\right\}^{a-1}$	Mudholkar and Srivastava, 1993
Alpha (α, β)	$\frac{\beta}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)}\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}$	Salvia, 1985
Birnbaum-Saunders (BS) (α, β)	$(2\alpha e^{\alpha^2}\sqrt{2\pi\beta})^{-1}\,x^{-3/2}\,(x+\beta)\,\exp\left[-\frac{(x/\beta)-(x/\beta)^{-1}}{2\alpha^2}\right]$	Birnbaum and Saunders, 1969
Burr XII (α, β)	$\alpha\,\beta\,\frac{x^{\alpha-1}}{(1+x^{\alpha})^{\beta+1}}$	Burr, 1942
Exp-Half-Normal (σ, a)	$\frac{a\sqrt{2}}{\sigma\sqrt{\pi}}\,\exp\left\{-\frac{x^2}{2\,\sigma^2}\right\}\,\{2\,\Phi(x)-1\}^{a-1}$	Gupta et al., 1998; Cooray and Ananda, 2008
Exp-Lindley (a, θ)	$\frac{a\,\theta^2}{\theta+1}\,(1+x)\,e^{-\theta\,x}\left\{1-\frac{e^{-\theta\,x}(1+\theta+\theta\,x)}{1+\theta}\right\}^{a-1}$	Nadarajah et al., 2012
Gamma (k, θ)	$\frac{x^{k-1}\,e^{-x/\theta}}{\theta^k\,\Gamma(k)}$	Khodabin and Ahmabadi, 2010[1]
Inverse Gaussian (λ, μ)	$\left[\frac{\lambda}{2\pi\,x^3}\right]^{1/2}\exp\left\{-\frac{\lambda\,(x-\mu)^2}{2\,\mu^2\,x}\right\}$	Chhikara and Folks, 1989
Log-Logistic II (α, β)	$\frac{\alpha\beta\,(\beta\,x)^{\alpha-1}}{[1+(\beta\,x)^{\alpha}]^2}$	Singh, 1998
Log-Normal (μ, σ)	$\frac{1}{\sqrt{2\pi}\sigma\,x}\,\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}$	Johnson, Kotz and Balakrishnan, 1994
Rayleigh II (μ, λ)	$2\lambda\,(x-\mu)e^{-\lambda(x-\mu)^2}$	Johnson, Kotz and Balakrishnan, 1994
Chi-Squared (μ)	$\frac{1}{2^{\mu/2}\Gamma(\mu/2)}x^{\mu/2-1}\,e^{-x/2}$	Johnson, Kotz and Balakrishnan, 1994
Exponential (λ)	$\lambda\,e^{-\lambda\,x}$	Gupta and Kundu, 1999, 2001
Half-Normal (σ^2)	$\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Cooray and Ananda, 2008
Log-Logistic I (α)	$\frac{\alpha\,x^{\alpha-1}}{(1+x^{\alpha})^2}$	Singh, 1998[2]
Rayleigh I (σ)	$\frac{x}{\sigma^2}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Hoffman and Karst, 1975

[1] with new parametrization and $\tau = 1$

[2] with $\beta = 1$

The data sets are:

(i) *Glass fibres data*

These data are studied by Smith and Naylor (1987), which represent the strengths of 1.5 cm glass fibres, measured at the National Physical Laboratory, England. Unfortunately, the units of measurement are not given in the paper.

(ii) *PBCseq protime data*

This data set contains variables referent to multiple laboratory results for each patient. The variable used for this study is Protime, which is the standardised blood clotting time. An analysis based on the data can be found in Murtagh et al., 1994.

Table 4.3 gives a descriptive summary of each data set, which includes central tendency statistics, standard deviation, among others. The distribution of the glass fibres data has negative skewness (left skewed) and positive kurtosis, whereas the distribution of the PBCseq protime data has positive skewness (right skewed) and positive kurtosis. We note that the considered data sets have both positive and negative skewness, whereas only positive values of the kurtosis were found. This fact can be confirmed in Section 4.3.2.

Table 4.3: Descriptive statistics.

Statistic	Data	
	glass fibres	PBCseq protime
Mean	1.51	11.00
Median	1.59	10.80
Std. Dev.	0.32	1.48
Minimum	0.55	9.00
Maximum	2.24	36.00
Skewness	−0.88	6.23
Kurtosis	0.80	76.73
n	63	1945

One of the important device which can help selecting a particular model is the total time on test (TTT) plot (for more details see Aarset, 1987). This plot is constructed through the quantities

$$T(i/n) = \left[\sum_{j=1}^{i} X_{j:n} + (n-i)\, X_{i:n}\right] \Big/ \sum_{j=1}^{n} X_{j:n} \; versus \; i/n,$$

where $i = 1, \ldots, n$ and $X_{j:n}$ is the j-th order statistics of the sample (Mudholkar et al., 1995).

The Figure 4.6 shows the TTT plot for glass fibres and PBCseq protime data, which denote an increasing hrf. This plot indicates the adequacy of the Kw-Alpha distribution to fit these data.

Tables 4.4 and 4.5 give the MLEs (with standard errors in parentheses) for some fitted models for glass fibres data and PBCseq protime data, respectively. Table 4.6 lists the values of the W*,

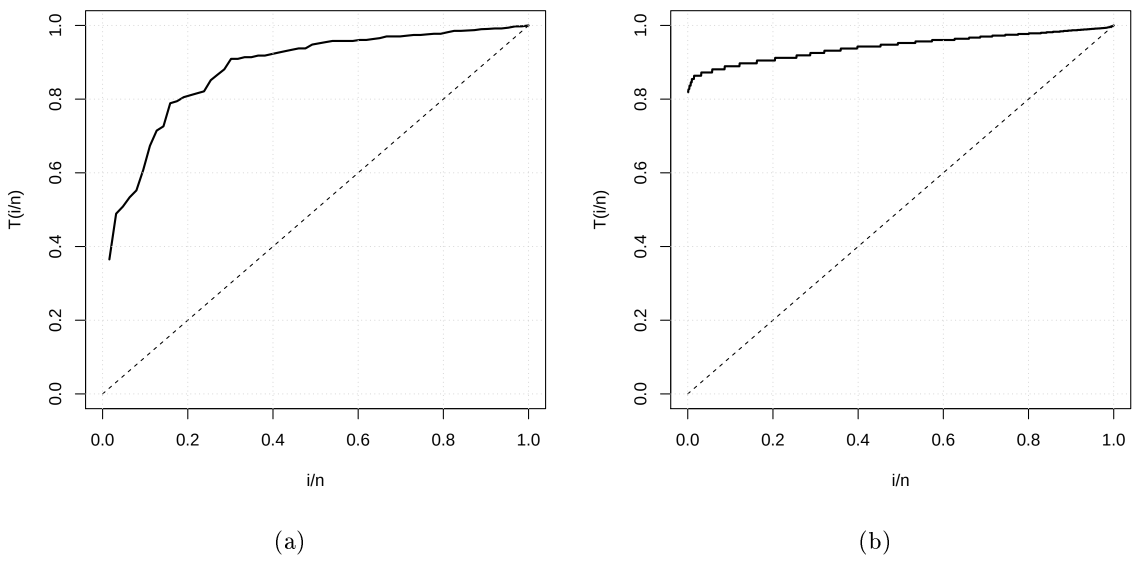


Figure 4.6: TTT plots (a) glass fibres data; (b) PBCseq protime data.

A*, KS, AIC, CAIC, BIC and HICQ for all fitted models for the considered two data sets. The Table 4.6 reveal that the Kw-Alpha model corresponds to the best fit for all data set, even when compared to distributions (with one, two and three parameters) well-known in the literature for fit to real data set. Note that we use different scenarios to the data set in concerning the compared distributions. This fact proves that, in differents situations, the proposed model have a upper performance to the other distributions analyzed.

Table 4.4: Maximum likelihood estimates (standard errors in parentheses) for the glass fibres data.

Distribution	MLEs (standard errors)			
Kw-Alpha (a, b, α, β)	0.008	0.112	19.476	27.674
	(0.003)	(0.014)	(0.013)	(0.013)
EG-Inverse-Gaussian (λ, μ, a, b)	33.133	7.523	988.648	0.345
	(0.180)	(0.170)	(5.658)	(0.044)
Exp-Alpha (a, α, β)	0.016	9.890	25.841	
	(0.005)	(2.147)	(4.651)	
Exp-Gompertz (α, β, a)	0.007	358.014	31.758	
	(0.001)	(56.717)	(9.813)	
Alpha (α, β)	3.021	4.261		
	(0.309)	(0.392)		
Birnbaum-Saunders (α, β)	0.262	1.457		
	(0.023)	(0.048)		
Burr XII (α, β)	7.482	0.320		
	(1.285)	(0.065)		
Exp-Half-Normal (σ, a)	0.779	11.592		
	(0.045)	(2.588)		
Exp-Lindley (a, θ)	26.171	2.990		
	(7.985)	(0.245)		
Gamma (k, θ)	17.439	11.573		
	(3.078)	(2.072)		
Inverse Gaussian (λ, μ)	21.560	1.506		
	(3.841)	(0.050)		
Log-Normal (μ, σ)	0.381	0.257		
	(0.032)	(0.022)		
Rayleigh II (μ, λ)	0.526	0.938		
	(0.024)	(0.125)		
Chi-Squared (μ)	2.364			
	(0.221)			
Exponential (λ)	0.664			
	(0.084)			
Half-Normal (σ)	1.540			
	(0.137)			
Log-Logistic I (α)	3.451			
	(0.341)			
Rayleigh I (σ)	1.089			
	(0.068)			

Table 4.5: Maximum likelihood estimates (standard errors in parentheses) for the PBCseq protime data.

Distribution	MLEs (standard errors)			
Kw-Alpha (a, b, α, β)	5.628	0.515	11.142	102.708
	(0.144)	(0.020)	(0.011)	(0.388)
EG-Inverse-Gaussian (λ, μ, a, b)	31.494	1001.475	70.272	543.833
	(1.267)	(455.149)	(4.146)	(96.185)
Beta-Exponential (a, b, λ)	1000.000	3.000	0.544	
	(112.797)	(0.177)	(0.015)	
Exp-Weibull (a, λ, γ)	998.638	0.310	1.634	
	(152.971)	(0.012)	(0.038)	
Alpha (α, β)	10.315	112.083		
	(0.166)	(1.797)		
Exp-Lindley (a, θ)	1370.890	0.852		
	(115.474)	(0.008)		
Log-Logistic II (α, β)	18.709	0.092		
	(0.355)	(0.0001)		

Table 4.6: Goodness-of-fit statistics.

Distribution	W^*	A^*	KS	AIC	CAIC	BIC	HQIC
Glass fibres data							
Kw-Alpha	**0.264**	**1.459**	**0.212**	**41.689**	**42.378**	**50.261**	**45.060**
EG-Inverse-Gaussian	0.305	1.669	0.185	41.881	42.571	50.454	45.253
Exp-Alpha	0.896	4.820	0.312	75.994	76.401	82.423	78.523
Exp-Gompertz	0.785	4.286	0.232	68.548	68.955	74.978	71.077
Alpha	1.234	6.544	0.271	95.207	95.407	99.493	96.893
BS	0.718	3.915	0.238	61.061	61.261	65.347	62.746
Burr XII	1.132	6.128	0.330	101.442	101.642	105.728	103.128
Exp-Half-Normal	0.587	3.223	0.217	53.628	53.828	57.914	55.314
Exp-Lindley	0.761	4.158	0.226	65.239	65.439	69.525	66.925
Gamma	0.568	3.117	0.216	51.903	52.103	56.189	53.589
Inverse Gaussian	0.722	3.939	0.238	61.341	61.541	65.627	63.027
Log-Normal	0.700	3.830	0.231	60.009	60.209	64.296	61.695
Rayleigh II	0.670	3.590	0.289	67.027	67.227	71.313	68.713
Chi-Squared	0.580	3.181	0.427	181.347	181.413	183.490	182.190
Exponential	0.570	3.127	0.418	179.661	179.726	181.804	180.504
Half-Normal	0.470	2.581	0.436	147.915	147.981	150.058	148.758
Log-Logistic I	0.804	4.403	0.534	138.944	139.010	141.088	139.787
Rayleigh I	0.465	2.553	0.333	101.581	101.647	103.724	102.424
PBCseq protime data							
Kw-Alpha	**0.499**	**3.983**	**0.048**	**5516.175**	**5516.195**	**5538.467**	**5524.372**
EG-Inverse-Gaussian	0.703	5.565	0.048	5569.767	5569.787	5592.059	5577.963
Beta-Exponential	1.967	13.953	0.673	5858.173	5858.185	5874.892	5864.320
Exp-Weibull	1.557	11.573	0.073	5743.132	5743.144	5759.851	5749.279
Alpha	2.385	16.604	0.076	5767.879	5767.885	5779.025	5771.977
Exp-Lindley	1.859	13.382	0.163	6066.103	6063.109	6077.249	6070.201
Log-Logistic II	2.426	16.760	0.067	5752.309	5752.315	5763.455	5756.408

4.6 Concluding remarks

We introduce the Kumaraswamy Alpha (Kw-Alpha) model, which generalizes the Alpha distribution discussed by Salvia (1985). The new distribution is constructed using Cordeiro and de Castro's (2011) generator. The Kw-Alpha model has three shape parameters and one scale parameter, and includes as special models the Exponentiated Alpha, Lehmann type II Alpha and Alpha distributions. The Kw-Alpha density and hazard rate functions take various forms depending on its shape parameters. We provide some mathematical properties of the new distribution including explicit expressions for the ordinary moments, generating function, mean deviations and density function of the order statistics. These results are valid for certain regions of the parametric space. We estimate the model parameters by maximum likelihood method and obtain a semiclosed estimator for one of the parameters. We provide a simulation study and two applications to real data, which indicate that the new model is a competitive distribution to fit real data.

Appendix

The elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters (a, b, α, β) are

$$J_{a,a} = -\,\frac{n}{a^2} - (b-1)\sum_{i=1}^{n} \log^2\left[\frac{\Phi(t_i)}{\Phi(\alpha)}\right]\left\{\frac{\Phi(t_i)^a\,\Phi(\alpha)^a}{[\Phi(\alpha)^a - \Phi(t_i)^a]^2}\right\},$$

$$J_{a,b} = -\sum_{i=1}^{n} \frac{\Phi(t_i)^a}{\Phi(\alpha)^a - \Phi(t_i)^a},$$

$$J_{a,\alpha} = \sum_{i=1}^{n} \frac{\phi(t_i)\Phi(\alpha) - \Phi(t_i) - \phi(\alpha)}{\Phi(t_i)\,\Phi(\alpha)}\left\{1 - a\,(b-1)\,\frac{\Phi(t_i)^a\Phi(\alpha)}{[\Phi(\alpha)^a - \Phi(t_i)^a]^2}\left[\frac{\phi(t_i)}{\Phi(t_i)} - \frac{\phi(\alpha)}{\Phi(\alpha)}\right]\right\},$$

$$J_{a,\beta} = \sum_{i=1}^{n} \frac{\phi(t_i)}{x_i\,\Phi(t_i)}\left\{\frac{a\,(b-1)}{x_i}\,\frac{\Phi(\alpha)^a\,\Phi(t_i)^{a-1}\,\phi(t_i)}{[\Phi(\alpha)^a - \Phi(t_i)^a]^2} - 1\right\},$$

$$J_{b,b} = -\,\frac{n}{b^2},$$

$$J_{b,\alpha} = a\sum_{i=1}^{n} \frac{\Phi(t_i)^{a-1}}{\Phi(\alpha)}\left[\frac{\phi(t_i)\,\Phi(\alpha) - \Phi(t_i)\,\phi(\alpha)}{\Phi(\alpha)^a - \Phi(t_i)^a}\right],$$

$$J_{b,\beta} = a \sum_{i=1}^{n} \frac{\Phi(t_i)^{a-1}\, \phi(t_i)}{x_i\, [\Phi(\alpha)^a - \Phi(t_i)^a]},$$

$$\begin{aligned}
J_{\alpha,\alpha} =& - n \left\{ 1 + \, a \frac{[\dot{\phi}(\alpha)]_\alpha\, \Phi(\alpha) - \phi(\alpha)^2}{\Phi(\alpha)^2} \right\} \\
&+ (a-1) \sum_{i=1}^{n} \frac{[\dot{\phi}(t_i)]_\alpha\, \Phi(t_i) - \phi(t_i)^2}{\Phi(t_i)^2} \\
&- a\,(b-1) \sum_{i=1}^{n} \left\{ \frac{a\, \Phi(t_i)^a\, \Phi(\alpha)^a}{[\Phi(\alpha)^a - \Phi(t_i)^a]^2} \left[\frac{\phi(t_i)}{\Phi(t_i)} - \frac{\phi(\alpha)}{\Phi(\alpha)} \right]^2 \right. \\
&\qquad \left. + \frac{\Phi(t_i)^a}{\Phi(\alpha)^a - \Phi(t_i)^a} \left[\frac{[\dot{\phi}(t_i)]_\alpha\, \Phi(t_i) - \phi(t_i)^2}{\Phi(t_i)^2} - \frac{[\dot{\phi}(\alpha)]_\alpha\, \Phi(\alpha) - \phi(\alpha)^2}{\Phi(\alpha)^2} \right] \right\},
\end{aligned}$$

$$\begin{aligned}
J_{\alpha,\beta} =& - (a-1) \sum_{i=1}^{n} \frac{[\dot{\phi}(t_i)]_\beta\, \Phi(t_i) - \phi(t_i)^2}{x_i\, \Phi(t_i)^2} + \sum_{i=1}^{n} \frac{1}{x_i} \\
&+ a^2\,(b-1) \sum_{i=1}^{n} \left\{ \frac{\phi(t_i)\, \Phi(t_i)^{a-1}\, \Phi(\alpha)^a}{x_i\, [\Phi(\alpha)^a - \Phi(t_i)^a]^2} \left[\frac{\phi(t_i)}{\Phi(t_i)} - \frac{\phi(\alpha)}{\Phi(\alpha)} \right] \right\} \\
&+ a\,(b-1) \sum_{i=1}^{n} \left\{ \frac{\Phi(t_i)^a}{\Phi(\alpha)^a - \Phi(t_i)^a} \left[\frac{[\dot{\phi}(t_i)]_\beta\, \Phi(t_i) - \phi(t_i)^2}{x_i\, \Phi(t_i)^2} \right] \right\},
\end{aligned}$$

$$\begin{aligned}
J_{\beta,\beta} =& - \frac{n}{\beta^2} - \sum_{i=1}^{n} \frac{1}{x_i^2} - (a-1) \sum_{i=1}^{n} \left\{ \frac{[\dot{\phi}(t_i)]_\beta}{x_i\, \Phi(t_i)} + \frac{1}{x_i^2} \left[\frac{\phi(t_i)}{\Phi(t_i)} \right]^2 \right\} \\
&- a\,(b-1) \sum_{i=1}^{n} \left\{ \frac{[\dot{\phi}(t_i)]_\beta\, \Phi(t_i)^{a-1}}{x_i\, [\Phi(\alpha)^a - \Phi(t_i)^a]} \right\} \\
&+ a\,(a-1)\,(b-1) \sum_{i=1}^{n} \left\{ \frac{\phi(t_i)^2\, \Phi(t_i)^{a-2}}{x_i^2\, [\Phi(\alpha)^a - \Phi(t_i)^a]} \right\} \\
&+ a^2\,(b-1) \sum_{i=1}^{n} \left\{ \frac{\phi(t_i)^2\, [\Phi(t_i)^{a-1}]^2}{x_i^2\, [\Phi(\alpha)^a - \Phi(t_i)^a]^2} \right\},
\end{aligned}$$

where $t_i = \left(\alpha - \frac{\beta}{x_i}\right)$, $[\dot{\phi}(\alpha)]_\alpha = \frac{\partial\, \phi(\alpha)}{\partial\, \alpha}$, $[\dot{\phi}(t_i)]_\alpha = \frac{\partial\, \phi(t_i)}{\partial\, \alpha}$ and $[\dot{\phi}(t_i)]_\beta = \frac{\partial\, \phi(t_i)}{\partial\, \beta}$.

Chapter 5

The Exponentiated Generalized Exponentiated-Generated Family

Resumo

Neste capítulo, apresentamos o gerador de distribuições da exponencializada generalizada exponencializada (EG-Exp-G), que é uma extensão da classe exponencializada generalizada proposta por Cordeiro et al. (2013). Novas distribuições são propostas como submodelos desta família, incluindo a distribuição EG-Exp-Alpha. Propriedades da classe de distribuições EG-Exp-G são derivadas, a saber: desvios médios, entropias de Rényi e Shannon, estatísticas de ordem, função geratriz de momentos, função quantílica, momentos ordinários e incompletos. Discutimos a estimação dos parâmetros do modelo por máxima verossimilhança e fornecemos um estimador em forma semifechada para um dos parâmetros adicionais. Um estudo de simulação de Monte Carlo é realizado para avaliar as propriedades assintóticas dos estimadores de máxima verossimilhança. Quatro aplicações a dados reais são utilizados para ilustrar a utilidade da nova classe.

Palavras-chave: Desvios médios. Distribuição Alpha. Estatísticas de ordem. Exponencializada-G. Exponencializada Generalizada-G. Função quantílica. Kumaraswamy-G. Máxima verossimilhança. Momentos.

Abstract

In this chapter, we present the exponentiated generalized exponentiated-generated (EG-Exp-G) family of distributions, which is an extension of the exponentiated generalized class proposed by Cordeiro et al. (2013). Some new distributions are proposed as submodels of this family, including the EG-Exp-Alpha distribution. Some of its mathematical properties are derived such as: mean deviations, moment generating function, order statistics, ordinary and incomplete moments, quantile function, Rényi and Shannon entropies. We discuss the estimation of the model parameters by maximum likelihood and provide a semiclosed estimator for one of the additional parameters. A Monte Carlo simulation study is performed to evaluate the assymptotic properties of the maximum likelihood estimators. Four applications to real data are used to illustrate the usefulness of the new family.

Keywords: Alpha Distribution. Exponentiated class. Exponentiated Generalized class. Kumaraswamy class. Maximum Likelihood Estimation. Mean deviations. Moments. Order statistics. Quantile function.

5.1 Introduction

In recent years, some different generalizations of well-known distributions have received increased attention in the literature. For example, we can refer Eugene et al. (2002) for the *beta* class, Zografos and Balakrishnan (2009) for the *gamma* class, Cordeiro and de Castro (2011) for the *Kumaraswamy* (Kw) class and the more recent ones by Alexander et al. (2012) and Cordeiro et al. (2013), who defined the generalized beta-generated and *exponentiated generalized* classes, respectively. Generalizing distributions provide some advantages, allowing us to define new families and to extend well-known distributions and provide great flexibility in modeling real data, which can be applied in several fields.

The generator proposed by Cordeiro et al. (2013) is defined as follows. For a baseline continuous cumulative distribution function (cdf), say $G(x)$, they introduced the *exponentiated generalized* ("EG" for short) class of distributions with two additional shape parameters $a > 0$ and $b > 0$ with cdf and probability density function (pdf) given by

$$F_1(x; a, b) = \{1 - [1 - G(x)]^a\}^b$$

and

$$f_1(x; a, b) = a\, b\, [1 - G(x)]^{a-1}\, \{1 - [1 - G(x)]^a\}^{b-1}\, g(x),$$

respectively.

In this chapter, we propose a new class of distributions that extends the EG class. We call this new class of distributions as the *exponentiated generalized exponentiated-generated* (EG-Exp-G) family with cdf and pdf given by

$$F(x; a, b, c) = \{1 - [1 - G(x)^a]^b\}^c \tag{5.1}$$

and

$$f(x; a, b, c) = \; a\, b\, c\; G(x)^{a-1}[1 - G(x)^a]^{b-1}\{1 - [1 - G(x)^a]^b\}^{c-1} g(x), \tag{5.2}$$

respectively, where $a > 0$, $b > 0$ and $c > 0$ are three additional shape parameters.

We note that there is not complicated function in equations (5.1) and (5.2) in contrast with the generalized beta-generated (GBG) class of distributions defined by Alexander et al. (2012), which also includes three extra shape parameters but involves the incomplete beta function.

The baseline distribution $G(x)$ is clearly a special case of (5.1) when $a = b = c = 1$. Setting $a = b = 1$ or $b = c = 1$, we obtain the exponentiated type distributions defined by Gupta et al. (1998). For $a = b = 1$, we obtain the so-called Lehmann type I distribution, say $\mathrm{Exp}^a(\mathrm{G})$. For $a = c = 1$, we have the Lehmann type II distribution corresponding to the cdf $F(x) = 1 - [1 - G(x)]^b$. So, the family (5.1) generalizes both Lehmann types I and II alternative distributions. In additional, the EG and Kw-G classes are special cases of the EG-Exp-G family when $a = 1$ and $c = 1$, respectively.

The rest of this chapter is organized as follows. In Section 5.2, we present special models of the EG-Exp-G family. Some general useful expansions for the EG-Exp-G density and quantile functions and explicit expressions for the moments, generating function, mean deviations, Rényi and Shannon entropy measures and order statistics are derived in Section 5.3. Maximum likelihood estimation of the model parameters is discussed in Section 5.4. A simulation study is provided to evaluate the maximum likelihood estimates of the EG-Exp-Alpha model in Section 5.5 and four applications to real data sets show the usefulness of the new family for data modeling. Concluding remarks are provided in Section 5.6.

5.2 Special EG-Exp-G distributions

In this section, we present some distributions that can arise as special models of the EG-Exp-G family. In all special cases, the additional parameters a, b and c are positive real numbers.

5.2.1 The EG-Exp-Alpha Model

A random variable X has an Alpha distribution (Katsev, 1968; Wager and Barash, 1971; Sherif, 1983; Salvia, 1985) with shape $\alpha > 0$ and scale $\beta > 0$ parameters, if its cdf is given by

$$G(x;\alpha,\beta) = \frac{\Phi\left(\alpha - \frac{\beta}{x}\right)}{\Phi(\alpha)},$$

for $x > 0$, where $\Phi(\cdot)$ denote the standard normal cumulative function.
Then, the EG-Exp-Alpha cumulative distribution is given by

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

We can express the EG-Exp-Alpha density function as

$$\begin{aligned} f(x;a,b,c,\alpha,\beta) =& \frac{a\,b\,c\,\beta}{\sqrt{2\pi}x^2}\,\frac{\Phi^{a-1}\left(\alpha - \frac{\beta}{x}\right)}{\Phi^a(\alpha)}\left[1 - \frac{\Phi^a\left(\alpha - \frac{\beta}{x}\right)}{\Phi^a(\alpha)}\right]^{b-1} \\ &\times \left\{1 - \left[1 - \frac{\Phi^a\left(\alpha - \frac{\beta}{x}\right)}{\Phi^a(\alpha)}\right]^b\right\}^{c-1} \exp\left\{-\frac{1}{2}\left(\alpha - \frac{\beta}{x}\right)^2\right\}. \end{aligned} \tag{5.3}$$

The hazard rate function (hrf) corresponding to (5.3) is given by

$$\begin{aligned} h(x;a,b,c,\alpha,\beta) =& \frac{a\,b\,c\,\beta}{\sqrt{2\pi}x^2}\,\frac{\Phi^{a-1}\left(\alpha-\frac{\beta}{x}\right)}{\Phi^a(\alpha)}\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\} \\ &\times\left[1-\frac{\Phi^a\left(\alpha-\frac{\beta}{x}\right)}{\Phi^a(\alpha)}\right]^{b-1}\left\{1-\left[1-\frac{\Phi^a\left(\alpha-\frac{\beta}{x}\right)}{\Phi^a(\alpha)}\right]^b\right\}^{c-1} \\ &\times\frac{1}{1-\left\{1-\left[1-\frac{\Phi^a\left(\alpha-\frac{\beta}{x}\right)}{\Phi^a(\alpha)}\right]^b\right\}^c}. \end{aligned}$$

Plots of the EG-Exp-Alpha hrf for some parameter values are displayed in Figure 5.1, which shows that the hrf has upside-down bathtub, increasing and decreasing shapes for different values of the parameters.

5.2.2 The EG-Exp-Gamma Model

The gamma cumulative distribution (for $x > 0$) with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ is given by

$$G(x;\alpha,\beta) = \frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)},$$

where $\gamma(\alpha,x) = \int_0^x \omega^{\alpha-1}\mathrm{e}^{-\omega}\mathrm{d}\omega$ is the incomplete gamma function and $\Gamma(\alpha) = \int_0^\infty \omega^{\alpha-1}\mathrm{e}^{-\omega}\mathrm{d}\omega$ is the gamma function. The EG-Exp-Gamma cumulative distribution becomes

$$F(x;a,b,c,\alpha,\beta) = \left\{1-[1-(\gamma_1(\alpha,\beta x))^a]^b\right\}^c,$$

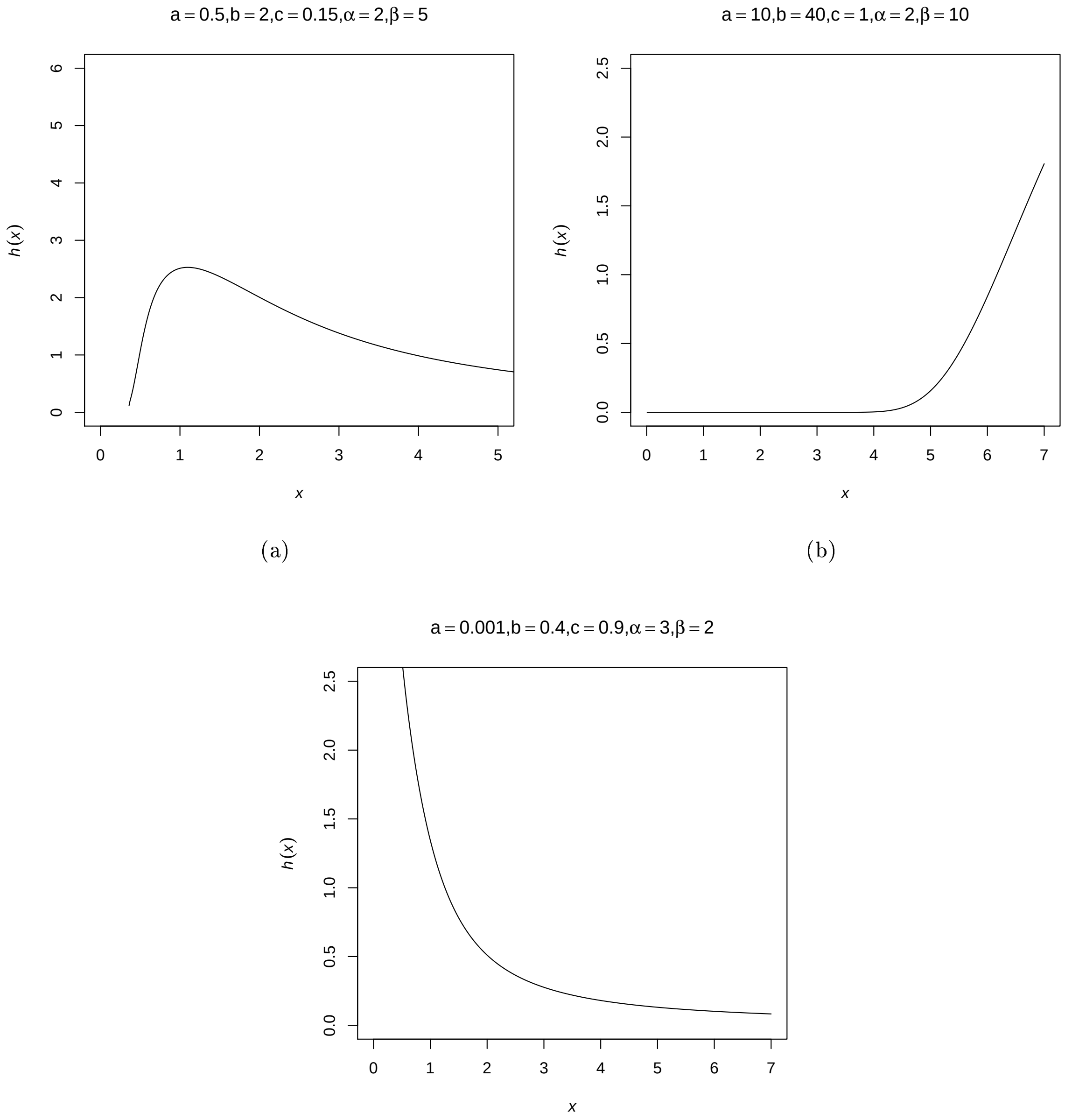
where $\gamma_1(\alpha,\beta x) = \gamma(\alpha,\beta x)/\Gamma(\alpha)$ is the incomplete gamma ratio function.
The corresponding density function reduces to

$$\begin{aligned} f(x;a,b,c,\alpha,\beta) =& \frac{a\,b\,c\;\beta^\alpha x^{\alpha-1}\mathrm{e}^{-\beta x}}{\Gamma(\alpha)}\,[\gamma_1(\alpha,\beta x)]^{a-1}\,\{1-[\gamma_1(\alpha,\beta x)]^a\}^{b-1} \\ &\times\left\{1-\{1-[\gamma_1(\alpha,\beta x)]^a\}^b\right\}^{c-1}. \end{aligned} \tag{5.4}$$

The hrf corresponding to (5.4) is given by

$$h(x;a,b,c,\alpha,\beta) = \frac{a\,b\,c\;[\gamma_1(\alpha,\beta x)]^{a-1}\;\{1-[\gamma_1(\alpha,\beta x)]^a\}^{b-1}\left\{1-\{1-[\gamma_1(\alpha,\beta x)]^a\}^b\right\}^{c-1}}{\beta^{-\alpha}\,x^{1-\alpha}\,\mathrm{e}^{\beta x}\,\Gamma(\alpha)\,\left\{1-\left\{1-[1-[\gamma_1(\alpha,\beta x)]^a]^b\right\}^c\right\}}.$$

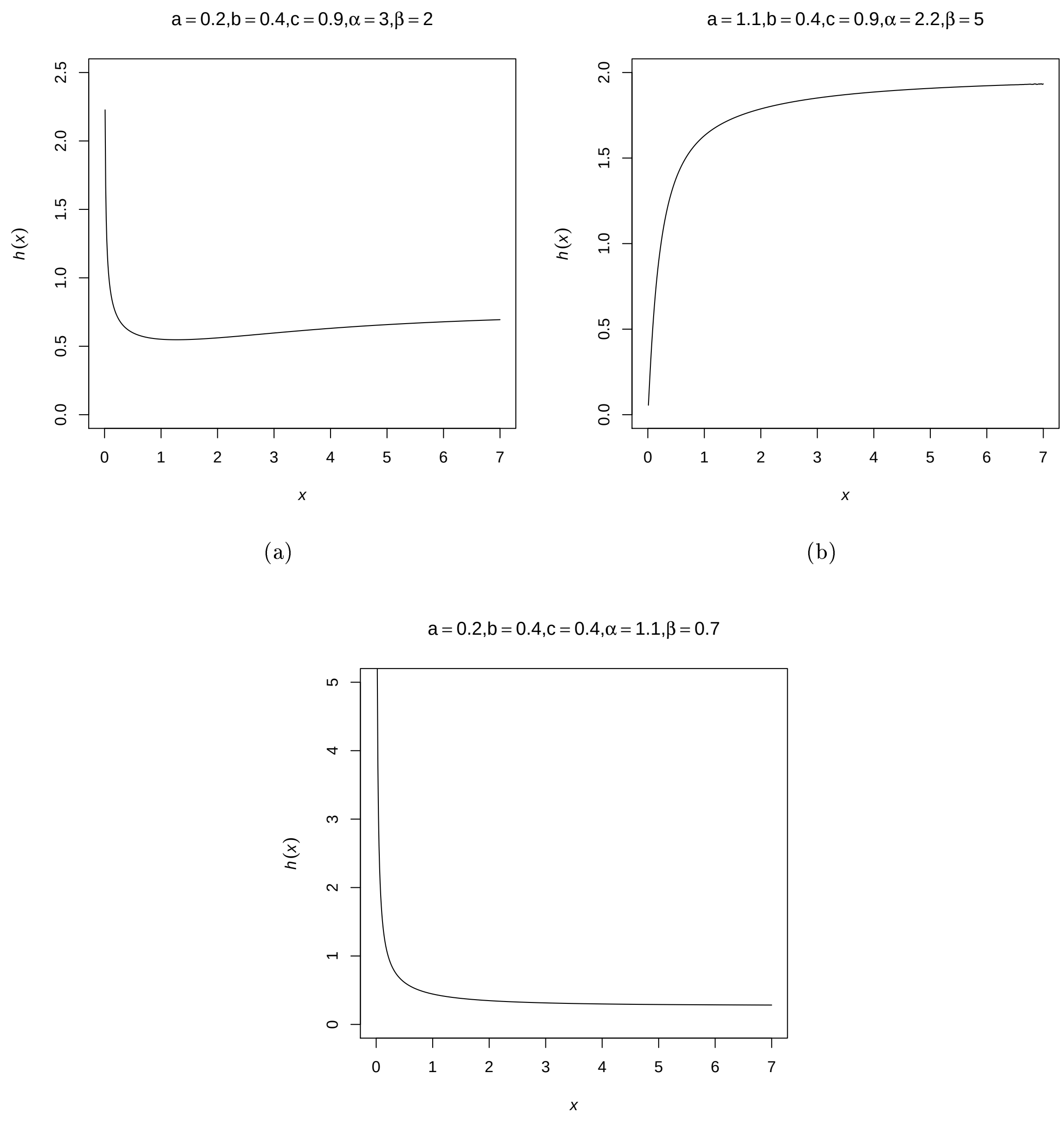
Plots of the EG-Exp-Gamma hrf for some parameter values are displayed in Figure 5.2, which reveals that the hrf has bathtub, increasing and decreasing shapes for different values of the parameters.



(a) (b)

(c)

Figure 5.1: Plots of the EG-Exp-Alpha hazard rate function for some parameter values. (a) Upside-down bathtub hazard rate function. (b) Increasing hazard rate function. (c) Decreasing hazard rate function.



(c)

Figure 5.2: Plots of the EG-Exp-Gamma hazard rate function for some parameter values. (a) Bathtub-shaped hazard rate function. (b) Increasing hazard rate function. (c) Decreasing hazard rate function.

5.2.3 The EG-Exp-Gumbel Model

The Gumbel cumulative distribution with parameters $\mu \in \mathrm{IR}$ and $\sigma > 0$ is given by

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

The EG-Exp-Gumbel (EG-Exp-Gu) cumulative distribution becomes

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

and the corresponding pdf reduces to

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

5.2.4 The EG-Exp-Log-Normal Model

Let $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cumulative and density functions, respectively. The EG-Exp-Log-Normal (EG-Exp-LogN) cumulative distribution is given by

$$F(x;a,b,c,\mu,\sigma) = \left\{1-\left[1-\Phi^a\left(\frac{\log x-\mu}{\sigma}\right)\right]^b\right\}^c,$$

where $x > 0$, $\mu \in \mathrm{IR}$ is a location parameter and $\sigma > 0$ is a scale parameter. The EG-Exp-LogN density function becomes

$$\begin{aligned}f(x;a,b,c,\mu,\sigma) = {} & \frac{a\,b\,c}{x}\,\Phi^{a-1}\left(\frac{\log x-\mu}{\sigma}\right)\left[1-\Phi^a\left(\frac{\log x-\mu}{\sigma}\right)\right]^{b-1} \\ & \times\left\{1-\left[1-\Phi^a\left(\frac{\log x-\mu}{\sigma}\right)\right]^b\right\}^{c-1}\phi\left(\frac{\log x-\mu}{\sigma}\right).\end{aligned} \tag{5.5}$$

The hrf corresponding to (5.5) is given by

$$h(x;a,b,c,\mu,\sigma) = \frac{a\,b\,c\,\Phi^{a-1}\left(\frac{\log x-\mu}{\sigma}\right)\left[1-\Phi^a\left(\frac{\log x-\mu}{\sigma}\right)\right]^{b-1}\left\{1-\left[1-\Phi^a\left(\frac{\log x-\mu}{\sigma}\right)\right]^b\right\}^{c-1}\phi\left(\frac{\log x-\mu}{\sigma}\right)}{x\left\{1-\left\{1-\left[1-\Phi^a\left(\frac{\log x-\mu}{\sigma}\right)\right]^b\right\}^c\right\}}.$$

Plots of the EG-Exp-Log-Normal hrf for some parameter values are displayed in Figure 5.3, which shows that the hrf has upside-down bathtub, increasing and decreasing shapes for different values of the parameters.

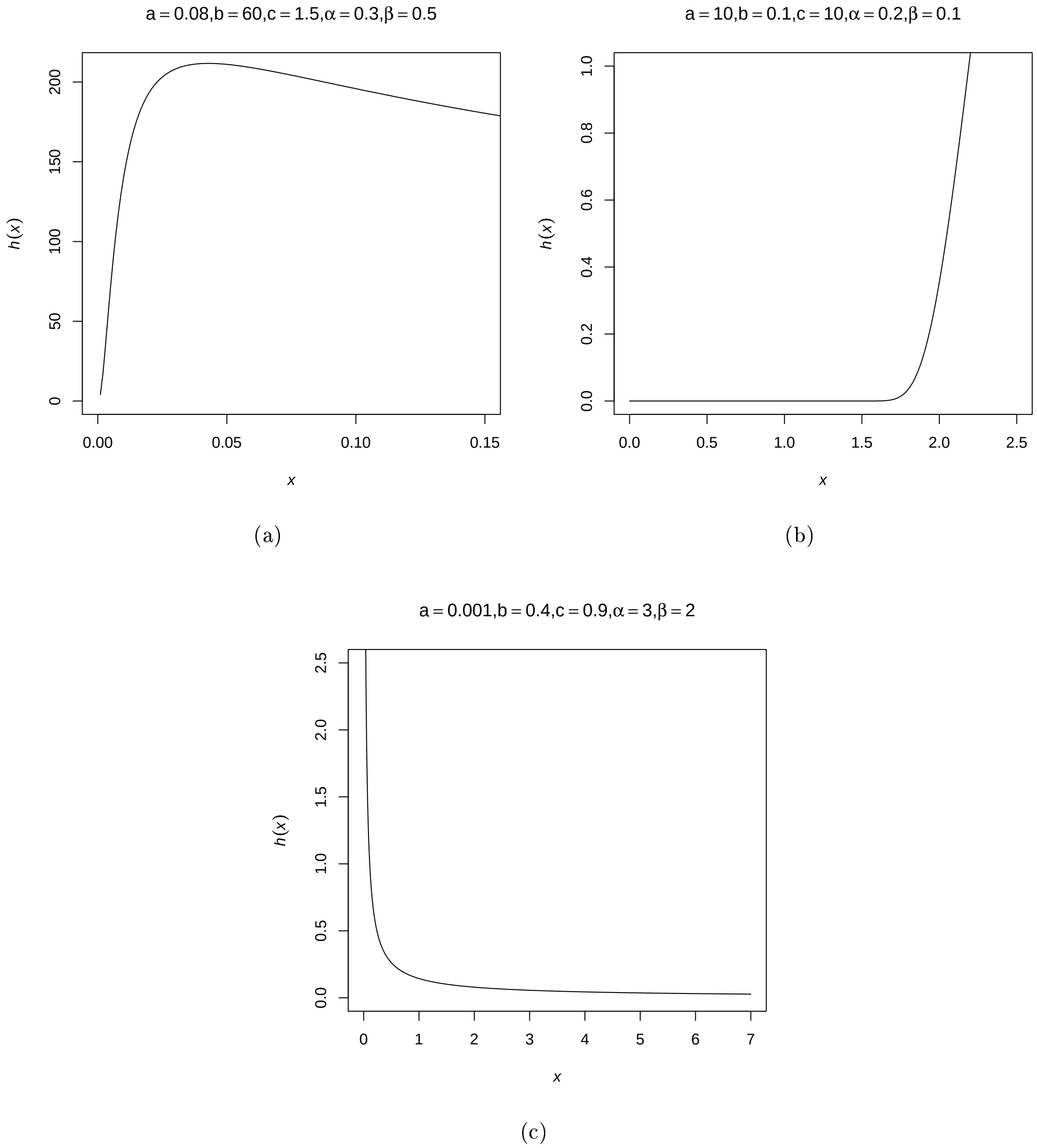


Figure 5.3: Plots of the EG-Exp-Log-Normal hazard rate function for some parameter values. (a) Upside-down bathtub hazard rate function. (b) Increasing hazard rate function. (c) Decreasing hazard rate function.

5.2.5 The EG-Exp-Normal Model

The EG-Exp-Normal (EG-Exp-N) cumulative distribution is given by

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

where $x \in \mathrm{IR}$, $\mu \in \mathrm{IR}$ is a location parameter and $\sigma > 0$ is a scale parameter. The EG-Exp-N density function becomes

$$f(x;a,b,c,\mu,\sigma) = \frac{a\,b\,c}{\sigma}\,\Phi^{a-1}\left(\frac{x-\mu}{\sigma}\right)\left[1-\Phi^{a}\left(\frac{x-\mu}{\sigma}\right)\right]^{b-1}$$
$$\times\left\{1-\left[1-\Phi^{a}\left(\frac{x-\mu}{\sigma}\right)\right]^{b}\right\}^{c-1}\phi\left(\frac{x-\mu}{\sigma}\right),$$

where $\phi(\cdot)$ denote the standard normal density function.

5.2.6 The EG-Exp-Weibull Model

The cdf of the Weibull distribution is

$$G(x;\lambda,\gamma) = 1-\exp\{-(\lambda x)^{\gamma}\},$$

for $x,\lambda,\gamma > 0$, where λ is the scale parameter and γ is the shape parameter. Then, the EG-Exp-Weibull (EG-Exp-W) cdf and pdf are given, respectively, by

$$F(x;a,b,c,\lambda,\gamma) = \Big\{1-\{1-[1-\exp\left[-\left(\lambda x\right)^{\gamma}\right]]^{a}\}^{b}\Big\}^{c}$$

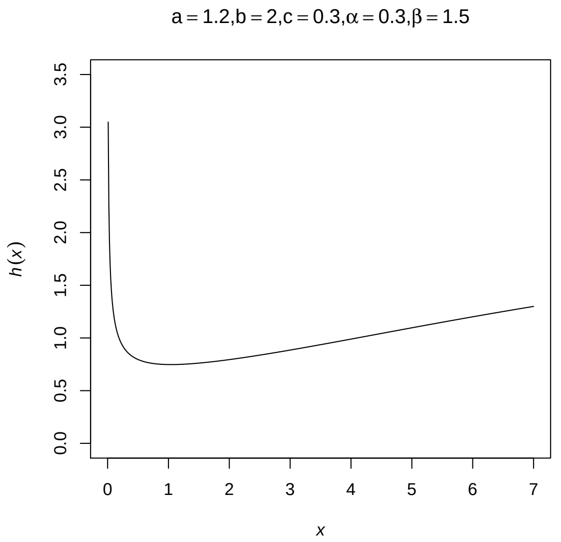
and

$$f(x;a,b,c,\lambda,\gamma) = a\,b\,c\,\gamma\,\lambda^{\gamma}x^{\gamma-1}\exp\{-(\lambda x)^{\gamma}\}\;\{1-\exp[-(\lambda x)^{\gamma}]\}^{a-1}$$
$$\times\{1-[1-\exp[-(\lambda x)^{\gamma}]]^{a}\}^{b-1}$$
$$\times\{1-\{1-[1-\exp[-(\lambda x)^{\gamma}]]^{a}\}^{b}\}^{c-1}. \tag{5.6}$$

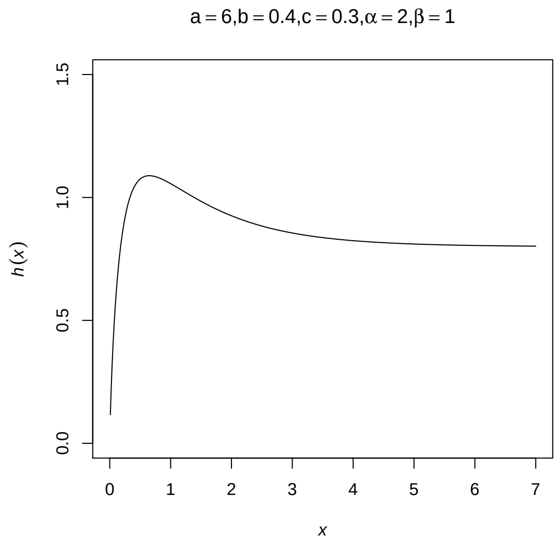
The hrf corresponding to (5.6) is given by

$$h(x;a,b,c,\lambda,\gamma) = \left\{\frac{a\,b\,c\,\gamma\,\lambda^{\gamma}x^{\gamma-1}\exp\{-(\lambda x)^{\gamma}\}}{1-\Big\{1-\{1-[1-\exp\left[-\left(\lambda x\right)^{\gamma}\right]]^{a}\}^{b}\Big\}^{c}}\right\}$$
$$\times\{1-\exp[-(\lambda x)^{\gamma}]\}^{a-1}$$
$$\times\{1-[1-\exp[-(\lambda x)^{\gamma}]]^{a}\}^{b-1}$$
$$\times\{1-\{1-[1-\exp[-(\lambda x)^{\gamma}]]^{a}\}^{b}\}^{c-1}.$$

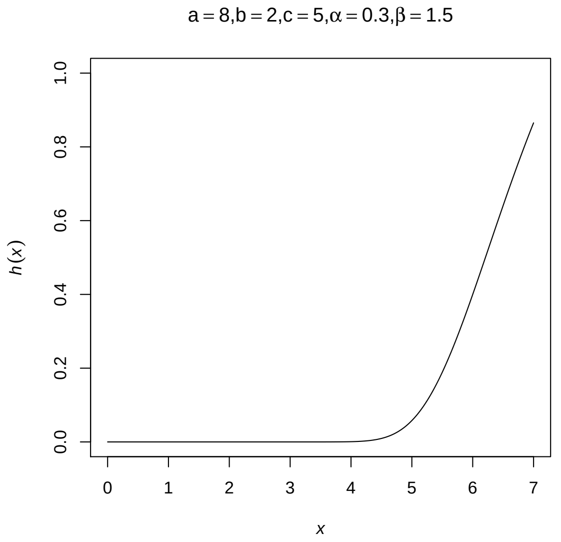
Plots of the EG-Exp-Weibull hrf for some parameter values are displayed in Figure 5.4, which reveals that the hrf has bathtub, upside-down bathtub, increasing and decreasing shapes for different values of the parameters.



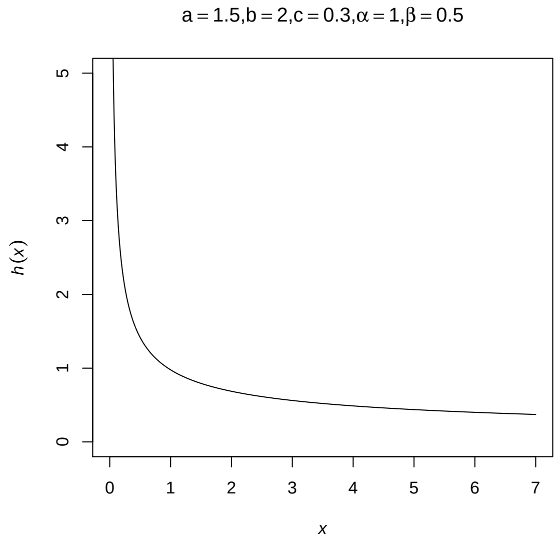
(a)



(b)



(c)



(d)

Figure 5.4: Plots of the EG-Exp-Weibull hazard rate function for some parameter values. (a) Bathtub-shaped hazard rate function. (b) Upside-down bathtub hazard rate function. (c) Increasing hazard rate function. (d) Decreasing hazard rate function.

5.3 Properties

In this section, we provide some mathematical properties of the proposed class. Henceforth, we denote by X a random variable having pdf (5.2).

5.3.1 A Useful Representation

For an arbitrary baseline cdf $G(x)$, a random variable is said to have the Exp-G distribution with power parameter $a > 0$, say $Y \sim \text{Exp}^a(\text{G})$, if its cdf and pdf are $H_a(x) = G(x)^a$ and $h_a(x) = a\, g(x)G(x)^{a-1}$, respectively.

We consider the generalized binomial expansion

$$(1-z)^b = \sum_{k=0}^{\infty} (-1)^k \binom{b}{k} z^k, \tag{5.7}$$

which holds for any real non-integer b and $|z| < 1$. Using expansion (5.7) twice in equation (5.1), we can express the cumulative function of X as

$$F(x) = \sum_{j=0}^{\infty} w_{j+1}\, H_{(j+1)a}(x), \tag{5.8}$$

where $w_{j+1} = \sum_{m=1}^{\infty} (-1)^{m+j+1} \binom{c}{m} \binom{bm}{j+1}$ and $H_{(j+1)a}(x) = G(x)^{(j+1)a}$ denotes (for $j \geq 0$) the cdf of the $\text{Exp}^{(j+1)a}(\text{G})$ distribution. By differentiating (5.8), we obtain

$$f(x) = \sum_{j=0}^{\infty} w_{j+1}\, h_{(j+1)a}(x), \tag{5.9}$$

where $h_{(j+1)a}(x) = (j+1)a\, g(x)G(x)^{[(j+1)a]-1}$ is the Exp-G pdf with power parameter $(j+1)a$.

Equation (5.9) reveals that the EG-Exp-G density function is a linear combination of Exp-G densities.

5.3.2 Quantile function

The quantile function (qf) of X can be obtained by inverting the parent cdf G

$$x_u = Q(u) = G^{-1}\{1 - [1 - u^{1/c}]^{1/b}\}^{1/a}, \quad u \in (0,1). \tag{5.10}$$

The qf of EG-Exp-Alpha distribution (5.3), for example, is given by

$$x = Q(u) = \frac{\beta}{\alpha - \Phi^{-1}\left\{\Phi(\alpha)\left[1 - (1 - u^{1/c})^{1/b}\right]^{1/a}\right\}}, \quad u \in (0,1).$$

The median of X is simply $x_{1/2} = Q(1/2)$. Further, it is possible to generate EG-Exp-Alpha variates by $X = Q(U)$, where U is a uniform variate on the unit interval $(0,1)$.

The effect of the additional shape parameters a, b and c on the skewness and kurtosis of the EG-Exp-Alpha model can be based on quantile measures. In this sense, two important measures are the Bowley skewness (Kenney and Keeping, 1962) and the Moors kurtosis (Moors, 1988).

These measures are given by

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

and

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

respectively, where the Bowley skewness (B) is based on quartiles, while the Moors kurtosis (M) is based on octiles.

For the normal distribution, $B = M = 0$. Plots of these skewness and kurtosis measures for some choices of the parameters are displayed in Figure 5.5. These plots indicate that, in general, both skewness and kurtosis increase when a (c) increases for fixed values of b, c, α and β (a, b, α and β), and decrease when b increases for fixed values of a, c, α and β. Further, the kurtosis is negative for small values of b when $a = c = \alpha = \beta = 10$ and skewness is negative for small values of c when $a = b = \alpha = \beta = 10$.

If the baseline qf, say $Q_G(u)$, does not have a closed-form expression, it can usually be written in terms of a power series

$$Q_G(u) = \sum_{i=0}^{\infty} a_i\, u^i, \tag{5.11}$$

where the coefficients a_i's are suitably chosen real numbers depending on the parameters of the G distribution.

From now on, we use a result by Gradshteyn and Ryzhik (2000) (Section 0.314) for a power series raised to a positive integer n (for $n \leq 1$)

$$Q_G(u)^n = \left(\sum_{i=0}^{\infty} a_i\, u^i \right)^n = \sum_{i=0}^{\infty} c_{n,i}\, u^i, \tag{5.12}$$

where the coefficients $c_{n,i}$ (for $i = 1, 2, \ldots$) are easily obtained from the recurrence equation (which $c_{n,0} = a_0^n$)

$$c_{n,i} = (i\, a_0)^{-1} \sum_{m=1}^{i} [m\,(n+1) - i]\, a_m\, c_{n,i-m}. \tag{5.13}$$

Clearly, the quantity $c_{n,i}$ can be determined from $c_{n,0}, \ldots, c_{n,i-1}$ and then from the quantities $a_0, \ldots, a_i$.

Next, we derive an expansion for the argument of $Q_G(\cdot)$ in (5.10). Using the generalized binomial expansion twice since $u \in (0,1)$, the qf of X can be expressed as

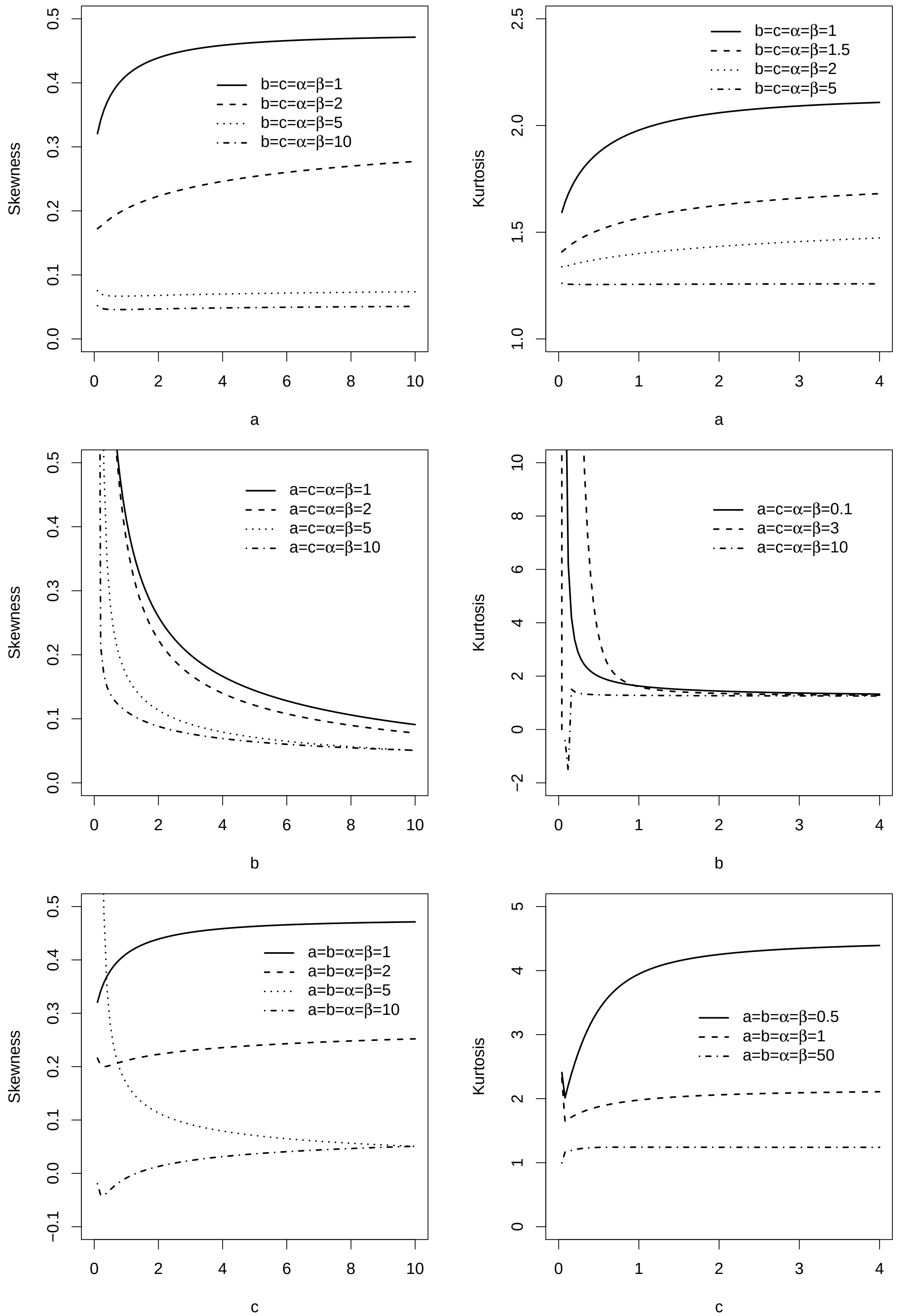


Figure 5.5: Plots of the EG-Exp-Alpha skewness and kurtosis as functions of a for some values of b, c, α and β, as function of b for some values of a, c, α and β and as function of c for some values of a, b, α and β.

$$Q(u) = Q_G\left(\sum_{j=0}^{\infty} \delta_j\, u^{j/c}\right), \tag{5.14}$$

where

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

For any baseline G distribution, we combine (5.11) and (5.14) to obtain

$$Q(u) = Q_G\left(\sum_{j=0}^{\infty} \delta_j\, u^{j/c}\right) = \sum_{i=0}^{\infty} a_i \left(\sum_{j=0}^{\infty} \delta_j\, u^{j/c}\right)^i,$$

and then using (5.12) and (5.13), we can write

$$Q(u) = \sum_{j=0}^{\infty} e_j\, u^{j/c}, \tag{5.15}$$

where $e_j = \sum_{i=0}^{\infty} a_i\, c_{i,j}$, $c_{i,0} = \delta_0^i$ and (for $j > 1$)

$$c_{i,j} = (j\,\delta_0)^{-1} \sum_{m=1}^{j} [m\,(i+1) - j]\, \delta_m\, c_{i,j-m}.$$

Equation (5.15) is the main result of this section and allows to obtain various mathematical quantities for the EG-Exp-G family in terms of integrals in $(0,1)$.

5.3.3 Moments

Henceforward, let $Y_j \sim \text{Exp}^{(j+1)a}(G)$ (for $j \geq 0$), i.e. having pdf $h_{(j+1)a}(x)$. A first formula for the nth moment of X can be determined from (5.9) as

$$\mu'_n = E(X^n) = \sum_{j=0}^{\infty} w_{j+1}\, E(Y_j^n). \tag{5.16}$$

Expressions for moments of several exponentiated distributions are given by Nadarajah and Kotz (2006), which can be used to obtain $E(X^n)$. We now provide an application of (5.16) by taking the baseline Weibull with cdf given by

$$G(x;\lambda,\gamma) = 1 - \exp\{-(\lambda x)^\gamma\}$$

where $\gamma > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. The corresponding Exp-Weibull (Exp-W) density function with power parameter $a > 0$ is given by

$$\pi(x;\lambda,\gamma) = a\gamma\,\lambda^\gamma\, x^{\gamma-1}\, \exp\{-(\lambda x)^\gamma\}\, \{1 - \exp[-(\lambda x)^\gamma]\}^{a-1}. \tag{5.17}$$

The nth moment of (5.17), say ρ_n, becomes (by Cordeiro et al., 2012)

$$\rho_n = \lambda^{-n}\,\Gamma(n/\gamma+1)\sum_{r=0}^{\infty}\frac{w_r}{(r+1)^{n/\gamma}}, \tag{5.18}$$

where

$$w_r = \frac{a}{(r+1)}\sum_{i=0}^{\infty}(-1)^{i+r}\binom{(i+1)a-1}{r}.$$

Combining (5.16) and (5.18), we obtain the nth moment of the EG-Exp-Weibull distribution given by (5.6) as

$$\mu_n' = \lambda^{-n}\,\Gamma(n/\gamma+1)\sum_{j,r,i=0}^{\infty}\frac{(-1)^{i+r}[(j+1)a]\binom{(i+1)(j+1)a-1}{r}w_{j+1}}{(r+1)^{n/\gamma+1}}.$$

A second formula for $E(X^n)$ follows from (5.9) as

$$\mu_n' = E(X^n) = \sum_{j=0}^{\infty}[(j+1)a]\,w_{j+1}\,\tau(n,(j+1)a-1), \tag{5.19}$$

where $\tau(n,a)$ is given in terms of the baseline qf $Q_G(u) = G^{-1}(u)$ as

$$\tau(n,a) = \int_{-\infty}^{\infty} x^n G(x)^a g(x)dx = \int_0^1 Q_G(u)^n u^a du.$$

The ordinary moments of several EG-Exp-G distributions can be determined directly from (5.19). Next, we provide two examples. First, the moments of the EG-Exp-Exponential (with parameter $\lambda > 0$) distribution are given by

$$\mu_n' = n!\lambda^n\sum_{j,m=0}^{\infty}\frac{(-1)^{n+m}\,[(j+1)a]\,\binom{[(j+1)a]-1}{m}\,w_{j+1}}{(m+1)^{n+1}}.$$

Second, for the EG-Exp-Standard logistic distribution, where the baseline cdf is $G(x) = (1+e^{-x})^{-1}$, we obtain (for $t < 1$) using a result by Prudnikov et al. (1986) (Section 2.6.13, equation 4)

$$\mu_n' = \sum_{j=0}^{\infty}[(j+1)a]\,w_{j+1}\left(\frac{\partial}{\partial t}\right)^n B(t+(j+1)a, 1-t)|_{t=0},$$

where $B(a,b) = \int_0^1 t^{a-1}\,(1-t)^{b-t}\,dt$ is the beta function.

The central moments (μ_n) and cumulants (κ_n) of X can be easily obtained from the ordinary moments using well-know relationships.

For empirical purposes, the shapes of many distributions can be usefully described by the incomplete moments. The nth incomplete moment of X follows as

$$m_n(y) = \int_{-\infty}^{y} x^n\,f(x)\,dx = \sum_{j=0}^{\infty}[(j+1)a]\,w_{j+1}\int_0^{G(y)} Q_G(u)^n\,u^{[(j+1)a]-1}du. \tag{5.20}$$

This integral in (5.20) can be evaluated for most baseline G distributions.

5.3.4 Generating function

In this section, we provide two general formulae for the moment generating function (mgf) $M(t) = E(\mathrm{e}^{\mathrm{tX}})$ of X. A first formula for $M(t)$ comes from (5.9) as

$$M(t) = \sum_{j=0}^{\infty} w_{j+1} M_j(t),$$

where $M_j(t)$ is the mgf of Y_j. Hence, $M(t)$ can be determined from the generating function of the Exp-G distribution.

A second formula for $M(t)$ can be derived from (5.9) as

$$M(t) = \sum_{j=0}^{\infty} [(j+1)a]\, w_{j+1}\, \rho(t, (j+1)a - 1), \tag{5.21}$$

where $\rho(t, a)$ is given by

$$\rho(t, a) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} G(x)^a g(x) dx = \int_0^1 \exp\{t\, Q_G(u)\}\, u^a du.$$

We can obtain the mgfs for several EG-Exp-G distributions directly from equation (5.21). For example, the generating functions of the EG-Exp-Exponential (with parameter λ and $t < \lambda^{-1}$) and EG-Exp-Standard logistic (for $t < 1$) distributions are given by

$$M(t) = \sum_{j=0}^{\infty} [(j+1)a]\, B((j+1)a, 1 - \lambda t)\, w_{j+1}$$

and

$$M(t) = \sum_{j=0}^{\infty} [(j+1)a]\, B(t + (j+1)a, 1 - t)\, w_{j+1},$$

respectively.

5.3.5 Mean Deviations

The mean deviations about the mean ($\delta_1 = E(|X - \mu_1'|)$ and about the median ($\delta_2 = E(|X - M|)$) of X can be expressed as

$$\delta_1 = 2\mu_1' F(\mu_1') - 2m_1(\mu_1') \qquad \text{and} \qquad \delta_2 = \mu_1' - 2m_1(M), \tag{5.22}$$

respectively, where $\mu_1' = E(X)$, $M = Median(X)$ is the median given by $M = Q(1/2)$, $F(\mu_1')$ is easily obtained from the cdf (5.1) and $m_1(z) = \int_{-\infty}^{z} x f(x) dx$ is the first incomplete moment.

Here, we provide two formulae to compute δ_1 and δ_2. The first equation for $m_1(z)$ can be derived from (5.9) as

$$m_1(z)=\sum_{j=0}^{\infty}w_{j+1}J_{j+1}(z), \tag{5.23}$$

where

$$J_{j+1}(z)=\int_{-\infty}^{z}x\,h_{(j+1)a}(x)dx. \tag{5.24}$$

Equation (5.24) is the basic quantity to compute the mean deviations of the Exp-G distributions. Hence, the mean deviations in (5.22) depend only on the mean deviations of the Exp-G distribution. So, alternative representations for δ_1 and δ_2 are

$$\delta_1=2\mu_1'F(\mu_1')-2\sum_{j=0}^{\infty}w_{j+1}J_{j+1}(\mu_1')\qquad\text{and}\qquad\delta_2=\mu_1'-2\sum_{j=0}^{\infty}w_{j+1}J_{j+1}(M).$$

A simple application of (5.23) and (5.24) refers to the EG-Exp-Weibull distribution. The Exp-W density function (for $x>0$) with power parameter $(j+1)a$, shape parameter γ and scale parameter λ is

$$h_{(j+1)a}(x)=\gamma\,\lambda^{\gamma}\,(j+1)a\,x^{\gamma-1}\,\exp\{-(\lambda x)^{\gamma}\}\,[1-\exp\{-(\lambda x)^{\gamma}\}]^{[(j+1)a]-1}$$

and then

$$\begin{aligned}J_{j+1}(z)&=\gamma\,\lambda^{\gamma}\,(j+1)a\int_0^z x^{\gamma}\exp\{-(\lambda x)^{\gamma}\}[1-\exp\{-(\lambda x)^{\gamma}\}]^{[(j+1)a]-1}dx\\&=\gamma\,\lambda^{\gamma}\,(j+1)a\sum_{r=0}^{\infty}(-1)^r\binom{(j+1)a-1}{r}\int_0^z x^{\gamma}\exp\{-(r+1)(\lambda x)^{\gamma}\}dx.\end{aligned}$$

The last integral is just the incomplete gamma function and then the mean deviations for the EG-Exp-Weibull distribution can be determined from

$$m_1(z)=\lambda^{-1}\sum_{j,r=0}^{\infty}\frac{(-1)^r[(j+1)a]\,w_{j+1}}{(r+1)^{1+1/\gamma}}\binom{(j+1)a-1}{r}\gamma(1+\gamma^{-1},(r+1)(\lambda z)^{\gamma}).$$

A second general formula for $m_1(z)$ can be derived setting $u=G(x)$ in equation (5.9)

$$m_1(z)=\sum_{j=0}^{\infty}(j+1)a\,w_{j+1}\,T_{j+1}(z), \tag{5.25}$$

where

$$T_{j+1}(z)=\int_0^{G(z)}Q_G(u)\,u^{[(j+1)a]-1}du. \tag{5.26}$$

In a similar way, the mean deviations of any EG-Exp-G distribution can be determined from equations (5.25) and (5.26). For example, the mean deviations of the EG-Exp-Exponential distribution (with parameter λ) follow by using the generalized binomial expansion from the function

$$T_{j+1}(z) = \lambda^{-1}\,\Gamma((j+1)a-1)\sum_{k=0}^{\infty}\frac{(-1)^k\{1-\exp(-k\lambda z)\}}{\Gamma((j+1)a-1-k)(k+1)!}.$$

These equations can be addressed to obtain Bonferroni and Lorenz curves, which represent another important application of the first incomplete moment. For a given probability π, these curves are defined by $B(\pi) = m_1(q)/(\pi\mu_1')$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $\mu_1' = E(X)$ and $q = Q(\pi)$ is given by (5.10).

5.3.6 Entropy measure

The entropy of a random variable is a measure of variation of the uncertainty. Recently, numerous measures of entropy have been studied and compared in the literature. Here, we will derive the Rényi and Shannon entropies (Shannon, 1948; Rényi, 1961).

The Rényi entropy of a random variable X with pdf $f(x)$ is given by

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

for $\lambda > 0$ and $\lambda \neq 1$.

Using the binomial expansion (5.7) twice in equation (5.2), we can write

$$f(x)^{\lambda} = (abc)^{\lambda}\sum_{\ell=0}^{\infty}\delta_{\ell}\,G(x)^{\lambda(a-1)+a\ell}\,g(x)^{\lambda},$$

where

$$\delta_{\ell} = \sum_{k=0}^{\infty}(-1)^{k+\ell}\binom{\lambda(c-1)}{k}\binom{bk+\lambda(b-1)}{\ell}.$$

Then, the Rényi entropy of X reduces to

$$I_R(\lambda) = \frac{1}{1-\lambda}\log\left\{(abc)^{\lambda}\sum_{\ell=0}^{\infty}\delta_{\ell}\,I_{\ell}\right\}, \tag{5.27}$$

where

$$I_{\ell} = \int_{-\infty}^{\infty} G(x)^{\lambda(a-1)+a\ell}\,g(x)^{\lambda}\,dx$$

can be determined from the baseline G distribution, at least numerically.

The Shannon entropy of a random variable X is defined by $I_S = E\{-\log[f(X)]\}$. It is a special case of the Rényi entropy when $\lambda \uparrow 1$, i.e., the Shannon entropy can be obtained by limiting $\lambda \uparrow 1$ in (5.27). However, it is easier to derive an expression for I_S from its definition. We have

$$I_S = -\log(abc) - (a-1)E\{\log[G(X)]\} - E\{\log[g(X)]\}$$
$$- (b-1)E\{\log[1-G(X)^a]\} - (c-1)E\{\log[1-[1-G(X)^a]^b]\}. \quad (5.28)$$

The expectations in (5.28) can be evaluated numerically for given $G(\cdot)$ and $g(\cdot)$. By using the linear representation (5.9) and Prudnikov et al. (1986) (Section 2.6.3, equation 2), the first two expectations in (5.28) can be expressed as

$$E\{\log[G(X)]\} = \sum_{j=0}^{\infty}(j+1)a\, w_{j+1}\int_{-\infty}^{\infty}\log[G(x)]G(x)^{[(j+1)a]-1}g(x)dx$$
$$= -\sum_{j=0}^{\infty}\frac{w_{j+1}}{(j+1)a}$$

and

$$E\{\log[g(X)]\} = \sum_{j=0}^{\infty}(j+1)a\, w_{j+1}\int_{-\infty}^{\infty}\log[g(x)]G(x)^{[(j+1)a]-1}g(x)dx,$$

respectively. The last equation can be also be given in terms of the baseline qf as

$$E\{\log[g(X)]\} = \sum_{j=0}^{\infty}(j+1)a\, w_{j+1}\int_{-\infty}^{\infty}\log[g(Q_G(u))]u^{[(j+1)a]-1}du,$$

where this integral can be evaluated for most baseline distributions at least numerically.

Changing variables and calculating the integral by Prudnikov et al. (1986) (Section 2.6.9, equation 7 and Section 2.6.3, equation 2), we can write the remaining expectations in (5.28) as

$$E\{\log[1-G(X)^a]\} = -\frac{1}{bc}[\mathcal{C}+\psi(c+1)]$$

and

$$E\{\log[1-[1-G(X)^a]^b]\} = -\frac{1}{c^2},$$

respectively, where $\mathcal{C}$ is Euler's constant.

Then, the Shannon entropy can be expressed as

$$I_S = -\log(abc) + (a-1)\sum_{j=0}^{\infty}\frac{w_{j+1}}{(j+1)a}$$
$$-\sum_{j=0}^{\infty}(j+1)a\, w_{j+1}\int_{-\infty}^{\infty}\log[g(x)]G(x)^{[(j+1)a]-1}g(x)dx$$
$$+\frac{b-1}{bc}[\mathcal{C}+\psi(c+1)] + \frac{c-1}{c^2}.$$

5.3.7 Order statistics

Suppose $X_1, \ldots, X_n$ is a random sample from the EG-Exp-G family. Let $X_{i:n}$ denote the ith order statistic. It is well-known that the density of $X_{i:n}$ is given by

$$\begin{aligned} f_{i:n}(x) =& \frac{f(x)}{B(i, n-i+1)} F(x)^{i-1}[1-F(x)]^{n-i} \\ =& \frac{1}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x) F(x)^{i+j-1}. \end{aligned} \tag{5.29}$$

Substituting (5.1) and (5.2) in equation (5.29), we can write

$$\begin{aligned} f_{i:n}(x) =& \frac{abc}{B(i, n-i+1)} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} G(x)^{a-1}[1-G(x)^a]^{b-1} \\ & \times \{1-[1-G(x)^a]^b\}^{c(j+1)-1} g(x). \end{aligned} \tag{5.30}$$

Applying the binomial expansion (5.7) twice in (5.30), $f_{i:n}(x)$ reduces to

$$f_{i:n}(x) = \sum_{\ell=0}^{\infty} \rho_\ell \, h_{(\ell+1)a}(x), \tag{5.31}$$

where

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

$h_{(\ell+1)a}(x)$ is the Exp-G density function with power parameter $(\ell+1)a$. It reveals that the density of $X_{i:n}$ is a linear combination of Exp-G density functions.

Equation (5.31) is the main result of this section. So, several mathematical properties like generating function, mean deviations, ordinary and incomplete moments of the EG-Exp-G order statistics can be obtained from those quantities of the Exp-G distribution. The rth moment of $X_{i:n}$, for example, is given by

$$E(X_{i:n}^r) = \sum_{\ell=0}^{\infty} [(\ell+1)a] \rho_\ell \, \tau(r, (\ell+1)a-1),$$

where $\tau(r, a)$ is defined in (5.19).

Further, the mgf of $X_{i:n}$ can be expressed as

$$M(t) = \sum_{\ell=0}^{\infty} [(\ell+1)a] \rho_\ell \, \rho(t, (\ell+1)a-1),$$

where $\rho(t, a)$ is given by (5.21).

5.4 Estimation

In this section, we determine the maximum likelihood estimates (MLEs) of the parameters of the EG-Exp-G family. Let $(x_1, \ldots, x_n)$ be a random sample of size n from this family with vector of parameters $\boldsymbol{\theta}^\top = (a, b, c, \boldsymbol{\tau}^\top)$, where $\boldsymbol{\tau}$ is a $p \times 1$ vector of unknown parameters of the parent cdf $G(x; \boldsymbol{\tau})$. The log-likelihood function for the vector of parameters $\boldsymbol{\theta}$ can be expressed as

$$\begin{aligned}\ell(\boldsymbol{\theta}) =& n \log(a\,b\,c) + \sum_{i=1}^{n} \log[g(x_i;\boldsymbol{\tau})] + (a-1)\sum_{i=1}^{n} \log[G(x_i;\boldsymbol{\tau})] \\ &+ (b-1)\sum_{i=1}^{n} \log[1 - G(x_i;\boldsymbol{\tau})^a] + (c-1)\sum_{i=1}^{n} \log\{1 - [1 - G(x_i;\boldsymbol{\tau})^a]^b\}. \end{aligned} \tag{5.32}$$

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

$$U_a(\boldsymbol{\theta}) = \frac{n}{a} + \sum_{i=1}^{n} \log[G(x_i;\boldsymbol{\tau})] \left\{ 1 - \frac{(b-1)G(x_i;\boldsymbol{\tau})^a}{1 - G(x_i;\boldsymbol{\tau})^a} + \frac{b(c-1)G(x_i;\boldsymbol{\tau})^a[1 - G(x_i;\boldsymbol{\tau})^a]^{b-1}}{1 - [1 - G(x_i;\boldsymbol{\tau})^a]^b} \right\},$$

$$U_b(\boldsymbol{\theta}) = \frac{n}{b} + \sum_{i=1}^{n} \log[1 - G(x_i;\boldsymbol{\tau})^a] \left\{ 1 - \frac{(c-1)[1 - G(x_i;\boldsymbol{\tau})^a]^b}{1 - [1 - G(x_i;\boldsymbol{\tau})^a]^b} \right\},$$

$$U_c(\boldsymbol{\theta}) = \frac{n}{c} + \sum_{i=1}^{n} \log\{1 - [1 - G(x_i;\boldsymbol{\tau})^a]^b\}, \tag{5.33}$$

$$\begin{aligned} U_{\tau_j}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \Bigg\{ & \frac{[\dot{g}(x_i;\boldsymbol{\tau})]_{\tau_j}}{g(x_i;\boldsymbol{\tau})} + \frac{(a-1)[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}}{G(x_i;\boldsymbol{\tau})} - \frac{a(b-1)G(x_i;\boldsymbol{\tau})^{a-1}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}}{1 - G(x_i;\boldsymbol{\tau})^a} \\ & + \frac{ab(c-1)G(x_i;\boldsymbol{\tau})^{a-1}[1 - G(x_i;\boldsymbol{\tau})^a]^{b-1}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}}{1 - [1 - G(x_i;\boldsymbol{\tau})^a]^b} \Bigg\}, \end{aligned}$$

where $[\dot{g}(x_i;\boldsymbol{\tau})]_{\tau_j} = \frac{\partial g(x_i;\boldsymbol{\tau})}{\partial \tau_j}$ and $[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j} = \frac{\partial G(x_i;\boldsymbol{\tau})}{\partial \tau_j}$ for $j = 1, \ldots, p$.

The MLEs, $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c}, \hat{\boldsymbol{\tau}}^\top)^\top$, can be obtained numerically by solving $U(\hat{\boldsymbol{\theta}}) = 0$.

It is possible note from (5.33) that we can obtain a semiclosed estimator of c. For $\frac{\partial \ell(\boldsymbol{\theta})}{\partial c} = 0$, the estimator for c is given by

$$\hat{c}_{\hat{a},\hat{b},\hat{\boldsymbol{\tau}}} = -\frac{n}{\sum_{i=1}^{n} \log\left\{1 - [1 - G(x_i;\hat{\boldsymbol{\tau}})^{\hat{a}}]^{\hat{b}}\right\}}. \tag{5.34}$$

Another way to obtain the MLEs is using the profile log-likelihood. We suppose c fixed and rewrite the log-likelihood (5.32) as $\ell(\boldsymbol{\theta}) = \ell_c(a, b, \boldsymbol{\tau})$. In this manner, we show that c is fixed but $\boldsymbol{\theta}_c = (a, b, \boldsymbol{\tau}^\top)^\top$ varies. To estimate $\boldsymbol{\theta}_c$, we maximize $\ell_c(a, b, \boldsymbol{\tau})$ with respect to $\boldsymbol{\theta}_c$.

The estimative of c can be obtained analytically by (5.34). By replacing c by $\hat{c}_{a,b,\boldsymbol{\tau}}$ in (5.32), we can obtain the profile log-likelihood function for a, b and $\boldsymbol{\tau}$ as

$$\begin{aligned}\ell_c(a,b,\boldsymbol{\tau}) &= n\log\left\{-\frac{n\,a\,b}{\sum_{i=1}^{n}\log\{1-[1-G(x_i;\boldsymbol{\tau})^a]^b\}}\right\}+\sum_{i=1}^{n}\log[g(x_i;\boldsymbol{\tau})]\\ &+(a-1)\sum_{i=1}^{n}\log[G(x_i;\boldsymbol{\tau})]+(b-1)\sum_{i=1}^{n}\log[1-G(x_i;\boldsymbol{\tau})^a]\\ &-n-\sum_{i=1}^{n}\log\{1-[1-G(x_i;\boldsymbol{\tau})^a]^b\}. \qquad (5.35)\end{aligned}$$

Hence, the MLEs of a, b and $\boldsymbol{\tau}$ can be also determined by maximizing the profile log-likelihood function (5.35) with respect to a, b and $\boldsymbol{\tau}$, respectively.

We can see that the maximization of (5.35) is more simpler than the maximization of (5.32) with respect to their respective parameters. This fact is due to the number of parameters to the profile log-likelihood function, which is smaller than the number of parameters to the log-likelihood function. However, some of the new properties that hold for a true likelihood do not hold for its profile version, because the profile log-likelihood is not a real likelihood function.

For interval estimation and hypothesis tests on the model parameters, we require the $(p+3)\times(p+3)$ observed information matrix $J=J(\boldsymbol{\theta})$ given in the Appendix. Under standard regularity conditions $(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is asymptotically normal $N_{p+3}(0, I(\boldsymbol{\theta})^{-1})$, where $I(\boldsymbol{\theta})$ is the expected information matrix. In practice, we can substitute $I(\boldsymbol{\theta})$ by the observed information matrix evaluated at $\hat{\boldsymbol{\theta}}$, say $J(\hat{\boldsymbol{\theta}})$. The multivariate normal distribution $N_{p+3}(0, J(\hat{\boldsymbol{\theta}})^{-1})$ can be used to construct approximate confidence intervals for the parameters.

5.5 Simulation study and applications

In this section, we perform a Monte Carlo simulation study and provide four applications to real data to illustrate the flexibility of the EG-Exp-G class through the EG-Exp-Alpha (5.3), EG-Exp-Gamma (5.4) and EG-Exp-Log-Normal (5.5) models.

5.5.1 Simulation study

Here, we provide a Monte Carlo simulation study (with 1,000 replications) to quantify bias of the MLEs of the EG-Exp-Alpha parameters. The simulations are based on the following scenarios:

(i) $a=(7,9,12), c=\beta=1, b=5, \alpha=2,$

(ii) $b=(2,4,5), a=c=\alpha=\beta=1$ and

(iii) $c=(1,2,5), a=\alpha=\beta=1, b=5.$

We chose to vary the values of parameters a, b and c since they are extra parameters of the EG-Exp-G class.

We evaluate the means and standard deviations of the bias for all point estimators based on artificially generated samples of sizes $n = 10, 20, \ldots, 100$. These results are given for every pair (a, n), (b, n) and (c, n) in Figure 5.6, Figure 5.7 and Figure 5.8, respectively. The plots indicate that, in general, the means and standard deviations of the bias decrease when the sample increases as expected.

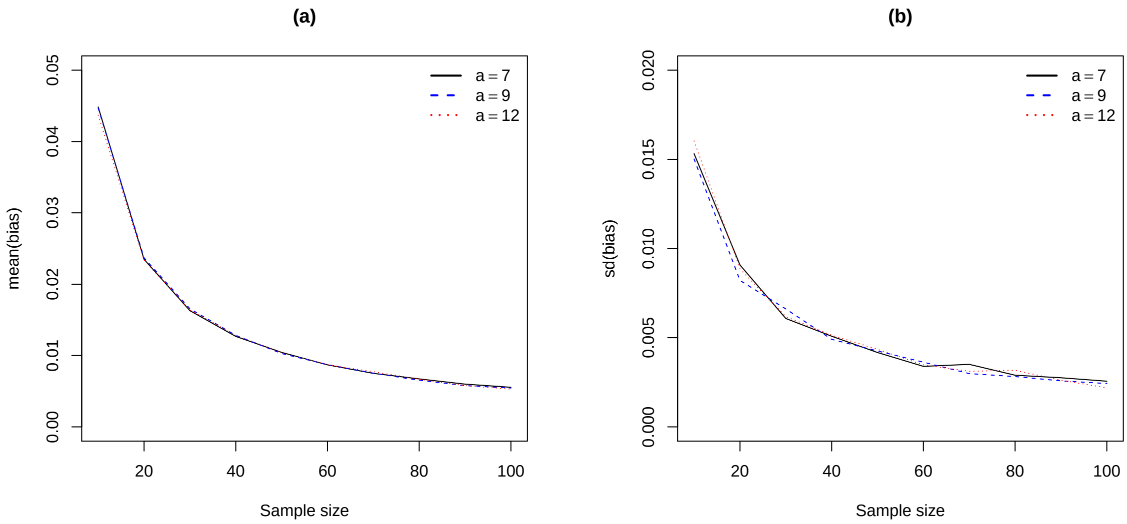


Figure 5.6: (a) Mean of the bias for $c = \beta = 1, b = 5, \alpha = 2$. (b) Standard deviation of the bias for $c = \beta = 1, b = 5, \alpha = 2$.

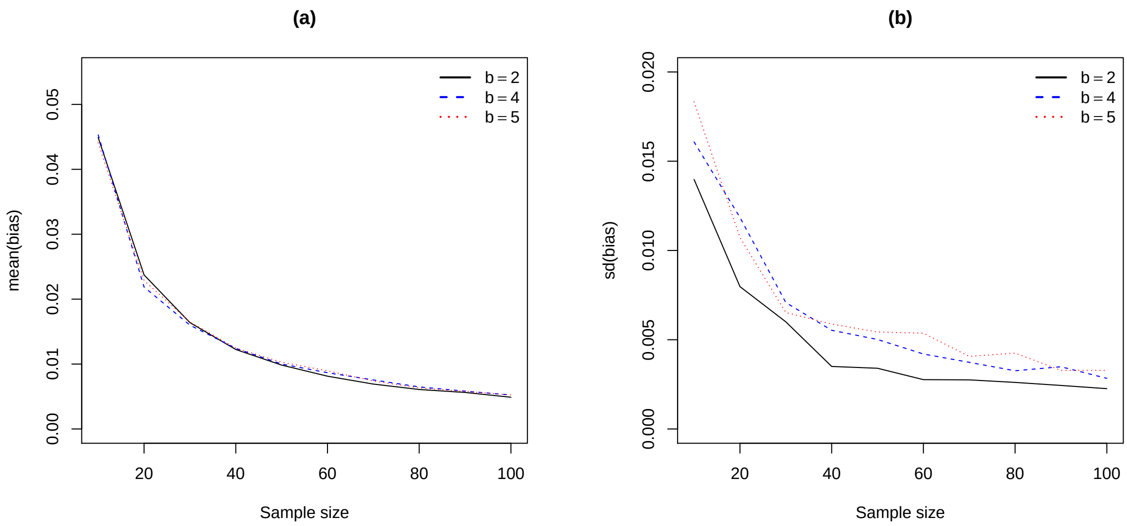


Figure 5.7: (a) Mean of the bias for $a = b = \alpha = \beta = 1$. (b) Standard deviation of the bias for $a = b = \alpha = \beta = 1$.

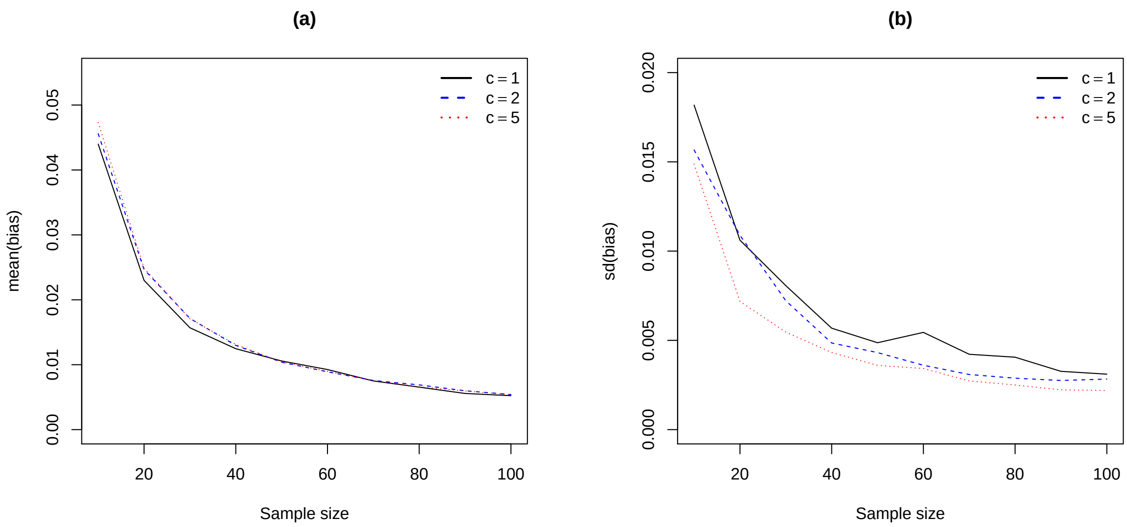


Figure 5.8: (a) Mean of the bias for $a = \alpha = \beta = 1, b = 5$. (b) Standard deviation of the bias for $a = \alpha = \beta = 1, b = 5$.

5.5.2 Applications to real data

In this section, we provide four applications to real data sets to illustrate the flexibility of the EG-Exp-G family and we use as example the EG-Exp-Alpha, EG-Exp-Gamma and EG-Exp-Log-Normal distributions with pdfs given by (5.3), (5.4) and (5.5), respectively. Here, we compare the mencioned EG-Exp-G models with competitive distributions well-known to fit real data set, as shown in Table 5.1. It is important to emphasize that all these distributions have support in the positive real set and their parameters are also positive reals, except the location parameter $\mu \in \mathrm{IR}$ of the Log-Normal distribution.

The MLEs of the parameters are computed (as discussed in Section 5.4) and the goodness-of-fit statistics for EG-Exp-Alpha, EG-Exp-Gamma and EG-Exp-Log-Normal models are compared with other distributions, as mencioned above. We consider the Cramér-von Misses (W*) and Anderson-Darling (A*), which are described by Chen and Balakrishnan (1995), and Kolmogorov-Smirnov (KS) statistics. Other measures of goodness of fit are also calculated, such as: Akaike Information Criteria (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criteria (CAIC) and Hannan-Quinn Information Criterion (HQIC).

The W* and A* statistics are used to verify which distribution fits better to the data. Since the values of W* and A* are smaller for the EG-Exp-G models when compared with those values of the other distributions, the proposed model seems to be a very competitive model for the data. We use these statistics, where we have a random sample $(x_1, \ldots, x_n)$ with empirical distribution function $F_n(x)$ and require to test if the sample comes from a special distribution. The W^* and A^* statistics are given by

Table 5.1: Distribution, probability density function and reference.

Distribution	Probability density function	Reference
EG-Log-Logistic II (α, β, a, b)	$\frac{a\,b\,\alpha\beta\,(\beta\,x)^{\alpha-1}}{[1+(\beta\,x)^{\alpha}]^2}\left\{1-\frac{(\beta\,x)^{\alpha}}{1+(\beta\,x)^{\alpha}}\right\}^{a-1}\left\{1-\left[1-\frac{(\beta\,x)^{\alpha}}{1+(\beta\,x)^{\alpha}}\right]^{a}\right\}^{b-1}$	Cordeiro et al., 2013; Singh, 1998
Exp-Alpha (a, α, β)	$\frac{a\beta\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)^a}\,\Phi\left(\alpha-\frac{\beta}{x}\right)^{a-1}$	Capítulo 3
Exp-Lomax (a, λ, α)	$a, \lambda\alpha\,[1+\lambda\,x]^{-(\alpha-1)}\,\{1-[1+\lambda\,x]^{-\alpha}\}^{a-1}$	Abdul-Moniem and Abdel-Hameed, 2012
Exp-Weibull (a, λ, γ)	$a\frac{\gamma}{\lambda}\left(\frac{x}{\lambda}\right)^{\gamma-1}\exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\}\left\{1-\exp\left[-\left(\frac{x}{\lambda}\right)^{\gamma}\right]\right\}^{a-1}$	Mudholkar and Srivastava, 1993
Alpha (α, β)	$\frac{\beta}{\sqrt{2\pi}\,x^2\,\Phi(\alpha)}\,\exp\left\{-\frac{1}{2}\left(\alpha-\frac{\beta}{x}\right)^2\right\}$	Salvia, 1985
Birnbaum-Saunders (BS) (α, β)	$(2\alpha e^{\alpha^2}\sqrt{2\pi\beta})^{-1}\,x^{-3/2}\,(x+\beta)\,\exp\left[-\frac{(x/\beta)-(x/\beta)^{-1}}{2\alpha^2}\right]$	Birnbaum and Saunders, 1969
Burr XII (α, β)	$\alpha\,\beta\,\frac{x^{\alpha-1}}{(1+x^{\alpha})^{\beta+1}}$	Burr, 1942
Exp-Half-Normal (σ, a)	$\frac{a\,\sqrt{2}}{\sigma\,\sqrt{\pi}}\,\exp\left\{-\frac{x^2}{2\,\sigma^2}\right\}\,\{2\,\Phi(x)-1\}^{a-1}$	Gupta et al., 1998; Cooray and Ananda, 2008
Exp-Lindley (a, θ)	$\frac{a\,\theta^2}{\theta+1}\,(1+x)\,e^{-\theta\,x}\left\{1-\frac{e^{-\theta\,x}(1+\theta+\theta\,x)}{1+\theta}\right\}^{a-1}$	Nadarajah et al., 2012
Gamma (k, θ)	$\frac{x^{k-1}\,e^{-x/\theta}}{\theta^k\,\Gamma(k)}$	Khodabin and Ahmabadi, 2010[1]
Inverse Gaussian (λ, μ)	$\left[\frac{\lambda}{2\pi\,x^3}\right]^{1/2}\,\exp\left\{-\frac{\lambda\,(x-\mu)^2}{2\,\mu^2\,x}\right\}$	Chhikara and Folks, 1989
Log-Logistic II (α, β)	$\frac{\alpha\beta\,(\beta\,x)^{\alpha-1}}{[1+(\beta\,x)^{\alpha}]^2}$	Singh, 1998
Log-Normal (μ, σ)	$\frac{1}{\sqrt{2\pi}\sigma\,x}\,\exp\left\{-\frac{(\log x-\mu)^2}{2\sigma^2}\right\}$	Johnson, Kotz and Balakrishnan, 1994
Rayleigh II (μ, λ)	$2\lambda\,(x-\mu)e^{-\lambda(x-\mu)^2}$	Johnson, Kotz and Balakrishnan, 1994
Chi-Squared (μ)	$\frac{1}{2^{\mu/2}\Gamma(\mu/2)}x^{\mu/2-1}\,e^{-x/2}$	Johnson, Kotz and Balakrishnan, 1994
Exponential (λ)	$\lambda\,e^{-\lambda\,x}$	Gupta and Kundu, 1999, 2001
Half-Normal (σ)	$\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\,\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Cooray and Ananda, 2008
Log-Logistic I (α)	$\frac{\alpha\,x^{\alpha-1}}{(1+x^{\alpha})^2}$	Singh, 1998
Rayleigh I (σ)	$\frac{x}{\sigma^2}\,\exp\left\{-\frac{x^2}{2\sigma^2}\right\}$	Hoffman and Karst, 1975

[1] with $\beta = 1$

$$
\begin{aligned}
W^* &= \left\{ n \int_{-\infty}^{+\infty} \{F_n(x) - F(x;\hat{\theta}_n)\}^2 \, dF(x;\hat{\theta}_n) \right\} \left(1 + \frac{0.5}{n}\right) = W^2 \left(1 + \frac{0.5}{n}\right), \\
A^* &= \left\{ n \int_{-\infty}^{+\infty} \frac{\{F_n(x) - F(x;\hat{\theta}_n)\}^2}{\{F(x;\hat{\theta})\,(1 - F(x;\hat{\theta}_n))\}} \, dF(x;\hat{\theta}_n) \right\} \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right) \\
&= A^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right),
\end{aligned}
$$

respectively, where $F_n(x)$ is the empirical distribution function and $F(x;\hat{\theta}_n)$ is the specified distribution function evaluated at the MLE $\hat{\theta}_n$ of θ. Note that the W^* and A^* statistics are given by difference distances of $F_n(x)$ and $F(x;\hat{\theta}_n)$. Thus, the lower are the W^* and A^* statistics more evidence we have that $F(x;\hat{\theta}_n)$ generates the sample.

The package `GenSA` available in the `R` programming language is used to obtain the initial values of the model parameters and the computations to the fit model are performed using the function `goodness.fit` of `AdequacyModel` package. The numerical BFGS (Broyden-Fletcher-Goldfarb-Shanno) procedure is used for minimization of the −log-likelihood function.

The data sets are:

(i) *Duration data*

The duration data refers to eruption time (in minutes) from the 'Old Faithful' geyser in Yellowstone National Park, Wyoming, United States. The version of the eruptions data comes from Azzalini and Bowman (1990) and is of continuous measurement from August 1 to August 15, 1985. These data, called 'geyser' in software `R`, are avaliable in package `MASS`.

(ii) *Sitka size data*

This data set gives repeated measurements on the log-size (height times diameter squared) of 79 Sitka spruce trees, 54 of which were grown in ozone-enriched chambers and 25 of which were controls. The size was measured eight times in 1989, at roughly monthly intervals. These data, called 'Sitka89' in software `R`, are avaliable in package `MASS`.

(iii) *Cancer data*

These data represent remission times (in months) of a random sample of 128 bladder cancer patients reported in Lee and Wang (2003).

(iv) *Percentage of body fat data*

This data set refers to the percentages of body fat determined by underwater weighting and various body circumference measurements for 250 men. For details about these data, see `http://lib.stat.cmu.edu/datasets/bodyfat`.

Table 5.2 gives a descriptive summary of each data set, which includes central tendency statistics, standard deviation, among others. The distribution of the duration data has negative skewness (left skewed) and negative kurtosis, whereas the distribution of the sitka size data has negative skewness (left skewed) and positive kurtosis. Furthermore, the distribution of the cancer data has positive skewness and positive kurtosis and the distribution of the body fat data is right skewed and presents negative kurtosis.

Table 5.2: Descriptive statistics.

Statistic	Data			
	duration	sitka size	cancer	body fat
Mean	3.461	5.991	9.366	19.300
Median	4.000	6.060	6.395	19.250
Std. Dev.	1.147	0.704	10.508	8.230
Minimum	0.833	3.610	0.080	3.000
Maximum	5.450	7.560	79.050	47.500
Skewness	−0.450	−0.527	3.248	0.194
Kurtosis	−1.431	0.061	15.195	−0.402
n	299	632	128	250

One of the important device which can help selecting a particular model is the total time on test (TTT) plot (for more details see Aarset, 1987). This plot is constructed through the quantities

$$T(i/n) = \left[\sum_{j=1}^{i} X_{j:n} + (n-i)\, X_{i:n}\right] \Big/ \sum_{j=1}^{n} X_{j:n} \ versus \ i/n,$$

where $i = 1, \ldots, n$ and $X_{j:n}$ is the j-th order statistics of the sample (Mudholkar et al., 1995).

The Figure 5.9 shows the TTT plot for the studied data, which indicates an increasing hrf for the duration data, sitka data and body fat data, and an upside-down bathtub rate function for the cancer data. This fact can be confirmed through the Figures 5.1, 5.2 and 5.3 and then indicate the adequacy of the EG-Exp-Alpha, EG-Exp-Gamma and EG-Exp-Log-Normal distributions to fit these data.

Tables 5.3, 5.4, 5.5 and 5.6 give the MLEs (with standard errors in parentheses) for some fitted models for duration data, sitka size data, cancer data and body fat data, respectively. Table 5.7 lists the values of the W*, A*, KS, AIC, CAIC, BIC and HICQ for all fitted models for all data sets. This table reveals that the EG-Exp model corresponds to the best fit for all data set, even when compared to distributions (with one and two parameters) well-known in the literature for fit to real data set. Note that we use different scenarios to the data set in concerning the compared distributions, shown that the proposed class have a competitive performance when compared to the other distributions analyzed.

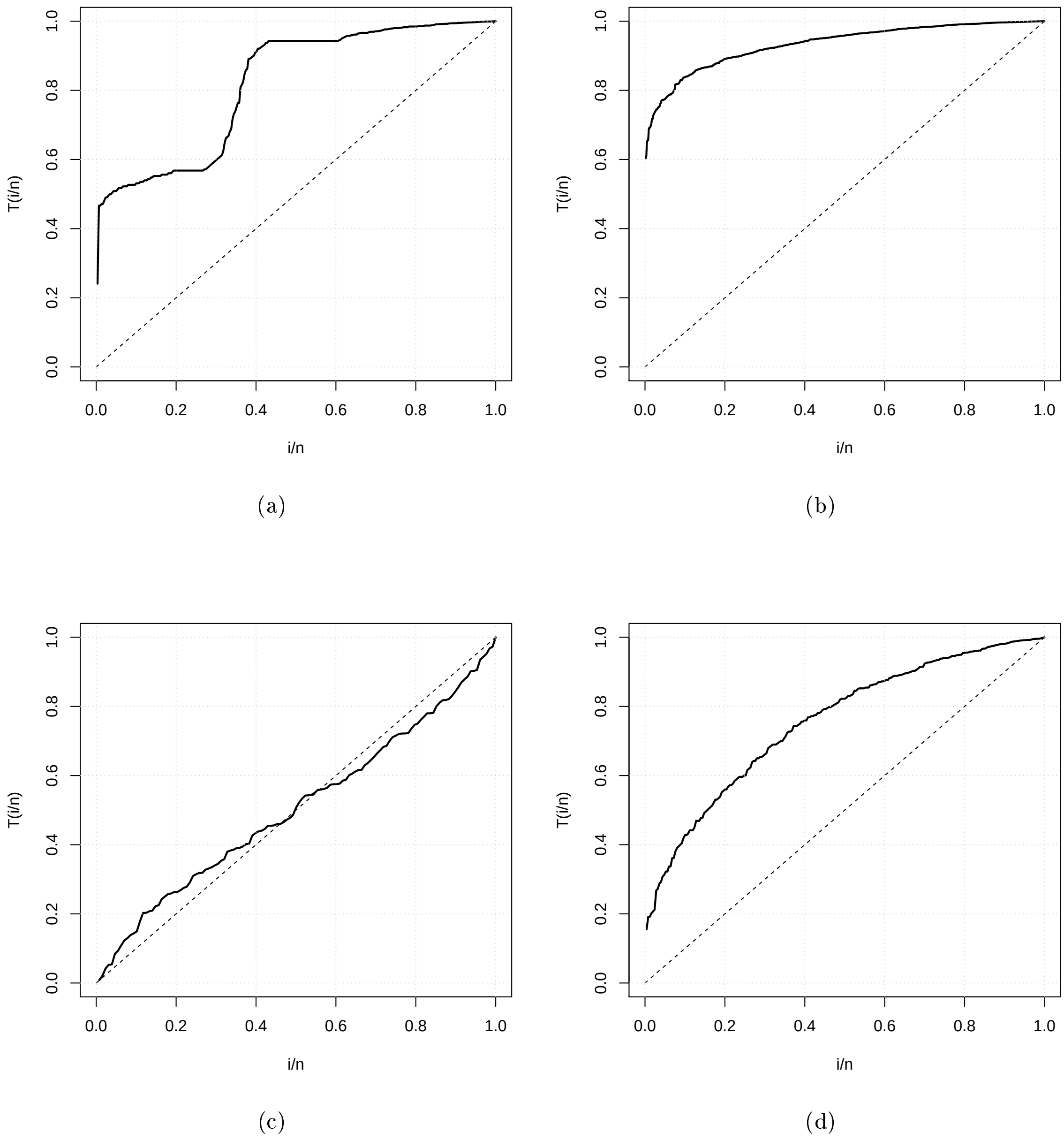


Figure 5.9: TTT plots (a) duration data; (b) sitka size data; (c) cancer data; (d) body fat data.

Table 5.3: Maximum likelihood estimates (standard errors in parentheses) for the duration data.

Distribution	MLEs (standard errors)				
EG-Exp-Alpha (a, b, c, α, β)	156.993	234.751	0.281	0.154	0.230
	(0.007)	(50.329)	(0.023)	(0.008)	(0.008)
EG-Log-Logistic II (α, β, a, b)	3.012	227.068	0.316	432.863	
	(1.324)	(0.986)	(0.139)	(61.522)	
Exp-Alpha (a, α, β)	0.444	2.179	8.888		
	(0.123)	(0.160)	(1.072)		
Alpha (α, β)	2.189	6.652			
	(0.130)	(0.316)			
BS (α, β)	0.390	3.214			
	(0.015)	(0.071)			
Burr XII (α, β)	9.632	0.088			
	(3.157)	(0.029)			
Inverse Gaussian (λ, μ)	21.860	3.460			
	(1.787)	(0.079)			
Log-Normal (μ, σ)	1.174	0.385			
	(0.022)	(0.015)			
Chi-Squared (μ)	4.186				
	(0.148)				
Exponential (λ)	0.288				
	(0.016)				
Half-Normal (σ)	3.645				
	(0.149)				
Log-Logistic I (α)	1.278				
	(0.057)				

Table 5.4: Maximum likelihood estimates (standard errors in parentheses) for the sitka size data.

Distribution	MLEs (standard errors)				
EG-Exp-Alpha (a, b, c, α, β)	854.159	491.839	2.165	1.406	0.249
	(111.684)	(169.734)	(0.305)	(0.108)	(0.032)
Exp-Lomax (a, λ, α)	719.902	125.468	0.009		
	(107.046)	(3.459)	(0.0002)		
Exp-Weibull (a, λ, γ)	1775.536	1.697	0.889		
	(441.300)	(0.129)	(0.020)		
Alpha (α, β)	7.620	44.958			
	(0.217)	(1.264)			
BS (α, β)	0.123	5.945			
	(0.003)	(0.029)			
Exp-Half-Normal (σ, a)	2.169	109.608			
	(0.034)	(12.594)			
Exp-Lindley (a, θ)	732.353	1.432			
	(135.584)	(0.036)			
Inverse Gaussian (λ, μ)	388.989	5.990			
	(21.882)	(0.029)			
Log-Normal (μ, σ)	1.782	0.123			
	(0.004)	(0.003)			
Rayleigh II (μ, λ)	3.603	0.161			
	(0.005)	(0.006)			
Chi-Squared (μ)	6.919				
	(0.137)				
Exponential (λ)	0.166				
	(0.006)				
Log-Logistic I (α)	0.864				
	(0.026)				

Table 5.5: Maximum likelihood estimates (standard errors in parentheses) for the cancer data.

Distribution	MLEs (standard errors)				
EG-Exp-Gamma (a, b, c, α, β)	6.928	0.405	0.183	1.158	0.195
	(0.317)	(0.005)	(0.022)	(0.026)	(0.025)
BS (α, β)	1.374	4.571			
	(0.086)	(0.446)			
Burr XII (α, β)	2.334	0.233			
	(0.354)	(0.039)			
Inverse Gaussian (λ, μ)	3.381	9.365			
	(0.422)	(1.377)			
Chi-Squared (μ)	6.745				
	(0.301)				
Half-Normal (σ)	14.045				
	(0.877)				
Log-Logistic I (α)	0.789				
	(0.055)				
Rayleigh I (σ)	9.931				
	(0.438)				

Table 5.6: Maximum likelihood estimates (standard errors in parentheses) for the body fat data.

Distribution	MLEs (standard errors)				
EG-Exp-Log-Normal (a, b, c, μ, σ)	0.187	9.110	0.181	4.164	0.162
	(0.012)	(0.170)	(0.031)	(0.014)	(0.011)
Exp-Lomax (a, λ, σ)	6.396	13.790	0.010		
	(0.784)	(2.396)	(0.002)		
BS (α, β)	0.532	16.870			
	(0.023)	(0.548)			
Exp-Lindley (a, θ)	2.871	0.155			
	(0.316)	(0.008)			
Gamma (k, θ)	4.609	0.238			
	(0.398)	(0.021)			
Inverse Gaussian (λ, μ)	63.454	19.301			
	(5.675)	(0.673)			
Log-Logistic II (α, β)	3.507	0.055			
	(0.185)	(0.001)			
Log-Normal (μ, σ)	2.847	0.512			
	(0.032)	(0.022)			
Chi-Squared (μ)	18.239				
	(0.371)				
Exponential (λ)	0.051				
	(0.003)				
Half-Normal (σ)	20.976				
	(0.938)				
Log-Logistic I (α)	0.537				
	(0.026)				

Table 5.7: Goodness-of-fit statistics.

Distribution	W^*	A^*	KS	AIC	CAIC	BIC	HQIC
			Duration data				
EG-Exp-Alpha	**3.934**	**21.545**	**0.279**	**918.696**	**918.900**	**937.198**	**926.101**
EG-Log-Logistic II	4.928	26.726	1.000	1385.753	1385.889	1400.555	1391.677
Exp-Alpha	5.507	29.089	0.306	1079.644	1079.725	1090.745	1084.087
Alpha	5.564	29.290	0.289	1091.063	1091.104	1098.464	1094.025
BS	4.817	26.121	0.284	981.445	981.486	988.846	984.407
Burr XII	5.096	26.849	0.344	1404.967	1405.007	1412.368	1407.929
Inverse Gaussian	4.841	26.234	0.285	983.533	983.574	990.934	986.495
Log-Normal	4.853	26.396	0.280	985.132	985.172	992.533	988.094
Chi-Squared	4.532	24.891	0.299	1161.955	1161.969	1165.656	1163.436
Exponential	4.523	24.849	0.369	1342.419	1342.433	1346.120	1343.900
Half-Normal	4.199	23.261	0.339	1209.551	1209.565	1213.252	1211.032
Log-Logistic I	5.014	27.163	0.645	1718.910	1718.923	1722.610	1720.391
			Sitka size data				
EG-Exp-Alpha	**0.617**	**4.121**	**0.068**	**1377.567**	**1377.662**	**1399.811**	**1386.206**
Exp-Lomax	2.190	13.751	0.148	1572.675	1572.713	1586.021	1577.858
Exp-Weibull	2.375	14.832	0.146	1568.763	1568.801	1582.109	1573.946
Alpha	2.016	12.730	0.104	1494.123	1494.142	1503.020	1497.578
BS	1.041	6.806	0.080	1408.677	1408.696	1417.574	1412.132
Exp-Half-Normal	1.668	10.643	0.090	1469.040	1469.060	1477.938	1472.496
Exp-Lindley	2.381	14.869	0.110	1536.512	1536.531	1545.410	1539.967
Inverse Gaussian	1.043	6.818	0.080	1408.850	1408.869	1417.748	1412.305
Log-Normal	1.033	6.760	0.079	1407.923	1407.942	1416.821	1411.379
Rayleigh II	1.350	8.651	0.222	1682.724	1682.743	1691.622	1686.180
Chi-Squared	0.803	5.310	0.383	2738.493	2738.500	2742.942	2740.221
Exponential	0.800	5.290	0.498	3528.822	3528.829	3533.271	3530.550
Log-Logistic I	1.119	7.294	0.765	4880.325	4880.331	4884.773	4882.052
			Cancer data				
EG-Exp-Gamma	**0.035**	**0.238**	**0.050**	**830.707**	**831.198**	**844.967**	**836.501**
BS	0.413	2.561	0.168	864.083	864.179	869.787	866.401
Burr XII	0.749	4.554	0.250	911.033	911.129	916.737	913.350
Inverse Gaussian	0.655	3.990	0.192	884.610	884.706	890.314	886.927
Chi-Squared	0.137	0.819	0.164	1006.399	1006.431	1009.251	1007.558
Half-Normal	0.441	2.591	0.201	864.232	864.264	867.084	865.391
Log-Logistic I	0.193	1.277	0.526	1011.721	1011.752	1014.573	1012.879
Rayleigh I	0.466	2.729	0.352	984.531	984.563	987.383	985.690
			Body fat data				
EG-Exp-Log-Normal	**0.063**	**0.405**	**0.046**	**1761.320**	**1761.566**	**1778.927**	**1768.406**
Exp-Lomax	0.627	3.764	0.091	1799.209	1799.306	1809.773	1803.460
BS	0.758	4.561	0.115	1806.801	1806.850	1813.844	1809.636
Exp-Lindley	0.403	2.431	0.079	1779.858	1779.906	1786.901	1782.692
Gamma	0.323	1.958	0.076	1773.375	1773.424	1780.418	1776.210
Inverse Gaussian	0.821	4.930	0.118	1811.724	1811.773	1818.767	1814.559
Log-Logistic II	0.605	3.642	0.077	1799.264	1799.312	1806.307	1802.098
Log-Normal	0.714	4.288	0.098	1802.968	1803.017	1810.011	1805.803
Chi-Squared	0.320	1.939	0.146	1857.217	1857.233	1860.738	1858.634
Exponential	0.327	1.980	0.261	1982.084	1982.100	1985.605	1983.501
Half-Normal	0.131	0.809	0.221	1886.590	1886.606	1890.111	1888.007
Log-Logistic I	0.819	4.904	0.684	2703.389	2703.405	2706.911	2704.806

5.6 Concluding remarks

We propose the *exponentiated generalized exponentiated-generated* (EG-Exp-G) family of distributions, which is an extension of the exponentiated generalized (EG) class proposed by Cordeiro et al. (2013). The EG-Exp-G class includes as special cases the two classes of Lehmann's alternatives and the EG and Kw-G (Cordeiro and de Castro, 2011) classes of distributions. Some new distributions are proposed as submodels of this family, including the EG-Exp-Alpha distribution. We provide some mathematical properties of the new class including explicit expressions for the quantile function, moments, generating function, mean deviations, Rényi entropy, Shannon entropy and order statistics. We estimate the model parameters by maximum likelihood method and provide a semiclosed estimator for one of the extra parameters. A Monte Carlo simulation study and four applications to real data show that the new family is a competitive model to fit real data when compared with well-known distributions used in survival analysis.

Appendix

The elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters $(a,b,c,\boldsymbol{\tau})$ are given by

$$J_{a,a} = -\frac{n}{a^2} - \sum_{i=1}^{n}[\log G(x_i;\boldsymbol{\tau})]^2 \times \left\{\frac{b(c-1)G^a(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^b\{[1-G^a(x_i;\boldsymbol{\tau})]^b + b\,G^a(x_i;\boldsymbol{\tau})-1\}}{[1-G^a(x_i;\boldsymbol{\tau})]^2\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}^2} + \frac{(b-1)G^a(x_i;\boldsymbol{\tau})}{[1-G^a(x_i;\boldsymbol{\tau})]^2}\right\},$$

$$J_{a,b} = -\sum_{i=1}^{n}\log G(x_i;\boldsymbol{\tau}) \times \left\{\frac{(c-1)G^a(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^b\{b\log[1-G^a(x_i;\boldsymbol{\tau})]-[1-G^a(x_i;\boldsymbol{\tau})]^b+1\}}{[1-G^a(x_i;\boldsymbol{\tau})]\{[1-G^a(x_i;\boldsymbol{\tau})]^b-1\}^2} + \frac{G^a(x_i;\boldsymbol{\tau})}{1-G^a(x_i;\boldsymbol{\tau})}\right\},$$

$$J_{a,c} = b\sum_{i=1}^{n}\frac{G^a(x_i;\boldsymbol{\tau})\log G(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^{b-1}}{\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}},$$

$$\begin{aligned}J_{a,\tau_j} =& \sum_{i=1}^{n}\frac{[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}}{G(x_i;\boldsymbol{\tau})} \\ &-(b-1)\sum_{i=1}^{n}\frac{[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}\,G^{a-1}(x_i;\boldsymbol{\tau})}{[1-G^a(x_i;\boldsymbol{\tau})]^2}\times\{[1+a\log G(x_i;\boldsymbol{\tau})][1-G^a(x_i;\boldsymbol{\tau})]+a\,G^a(x_i;\boldsymbol{\tau})\log G(x_i;\boldsymbol{\tau})\} \\ &+b(c-1)\sum_{i=1}^{n}\frac{[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}G^{a-1}(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^{b-1}}{\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}^2}\times \\ &\left\{\left[1+a\log G(x_i;\boldsymbol{\tau})-a(b-1)G^a(x_i;\boldsymbol{\tau})\log G(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^{-1}\right]\times\right. \\ &\left.\left[1-[1-G^a(x_i;\boldsymbol{\tau})]^b\right]-ab\,G^a(x_i;\boldsymbol{\tau})\log G(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^{b-1}\right\},\end{aligned}$$

$$J_{b,b} = -\frac{n}{b^2} - (c-1)\sum_{i=1}^{n} \frac{[1-G^a(x_i;\boldsymbol{\tau})]^b \{\log[1-G^a(x_i;\boldsymbol{\tau})]\}^2}{\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}^2},$$

$$J_{b,c} = -\sum_{i=1}^{n} \frac{[1-G^a(x_i;\boldsymbol{\tau})]^b \log[1-G^a(x_i;\boldsymbol{\tau})]}{1-[1-G^a(x_i;\boldsymbol{\tau})]^b},$$

$$\begin{aligned} J_{b,\tau_j} =& -a\sum_{i=1}^{n} \frac{[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j} G^{a-1}(x_i;\boldsymbol{\tau})}{1-G^a(x_i;\boldsymbol{\tau})} \\ &+ a(c-1)\sum_{i=1}^{n} \frac{G^{a-1}(x_i;\boldsymbol{\tau})[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[1-G^a(x_i;\boldsymbol{\tau})]^{b-1}}{\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}^2} \times \\ &\Big\{\{1+b\log[1-G^a(x_i;\boldsymbol{\tau})]\}\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\} + b[1-G^a(x_i;\boldsymbol{\tau})]^b \log[1-G^a(x_i;\boldsymbol{\tau})]\Big\}, \end{aligned}$$

$$J_{c,c} = -\frac{n}{c^2}, \quad J_{c,\tau_j} = ab\sum_{i=1}^{n} \frac{G^{a-1}(x_i;\boldsymbol{\tau})[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[1-G^a(x_i;\boldsymbol{\tau})]^{b-1}}{1-[1-G^a(x_i;\boldsymbol{\tau})]^b},$$

$$\begin{aligned} J_{\tau_j,\tau_s} =& \sum_{i=1}^{n} \frac{[\ddot{g}(x_i;\boldsymbol{\tau})]_{\tau_j\tau_s} g(x_i;\boldsymbol{\tau}) - [\dot{g}(x_i;\boldsymbol{\tau})]_{\tau_j}[\dot{g}(x_i;\boldsymbol{\tau})]_{\tau_s}}{[g(x_i;\boldsymbol{\tau})]^2} \\ &+ (a-1)\sum_{i=1}^{n} \frac{\{[\ddot{G}(x_i;\boldsymbol{\tau})]_{\tau_j\tau_s} G(x_i;\boldsymbol{\tau}) - [\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_s}\}}{[G(x_i;\boldsymbol{\tau})]^2} \\ &+ a(b-1)\sum_{i=1}^{n} \frac{G^{a-1}(x_i;\boldsymbol{\tau})}{[1-G^a(x_i;\boldsymbol{\tau})]^2} \Big\{[\ddot{G}(x_i;\boldsymbol{\tau})]_{\tau_j\tau_s} - (a-1)\,G^{-1}(x_i;\boldsymbol{\tau})[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_s} \\ &\qquad - aG^{a-1}(x_i;\boldsymbol{\tau})[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_s}\Big\} \times [1-G^a(x_i;\boldsymbol{\tau})] \\ &+ ab(c-1)\sum_{i=1}^{n} \frac{G^{a-1}(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^{b-1}\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}}{\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}^2} \times \\ &\qquad \Big[(a-1)G^{-1}(x_i;\boldsymbol{\tau})[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_s} \\ &\qquad - a(b-1)G^{a-1}(x_i;\boldsymbol{\tau})[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[1-G^a(x_i;\boldsymbol{\tau})]^{-1} + [\ddot{G}(x_i;\boldsymbol{\tau})]_{\tau_j\tau_s}\Big] \\ &- (ab)^2(c-1)\sum_{i=1}^{n} \frac{G^{2(a-1)}(x_i;\boldsymbol{\tau})[1-G^a(x_i;\boldsymbol{\tau})]^{2(b-1)}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j}[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_s}}{\{1-[1-G^a(x_i;\boldsymbol{\tau})]^b\}^2}, \end{aligned}$$

where $[\dot{g}(x_i;\boldsymbol{\tau})]_{\tau_j} = \frac{\partial g(x_i;\boldsymbol{\tau})}{\partial \tau_j}$, $[\dot{G}(x_i;\boldsymbol{\tau})]_{\tau_j} = \frac{\partial G(x_i;\boldsymbol{\tau})}{\partial \tau_j}$, $[\ddot{g}(x_i;\boldsymbol{\tau})]_{\tau_j\tau_s} = \frac{\partial^2 g(x_i;\boldsymbol{\tau})}{\partial \tau_j \tau_s}$, $[\ddot{G}(x_i;\boldsymbol{\tau})]_{\tau_j\tau_s} = \frac{\partial^2 G(x_i;\boldsymbol{\tau})}{\partial \tau_j \partial \tau_s}$, for $j, s = 1, \ldots, p$.

Chapter 6

General Conclusions

In this thesis, we discuss the Alpha distribution (Katsev, 1968; Wager and Barash, 1971; Sherif, 1983; Salvia, 1985). The authors provided the characterization and a number of structural properties of this distribution, but do not perform simulation study and applications to real data sets. The Alpha model have one shape parameter and one scale parameter and its cumulative distribution function (cdf) and probability density function (pdf) involves the standard normal cumulative function. We study the shapes of the pdf and hazard rate function (hrf) of the Alpha distribution and we show that its hrf has increasing and decreasing shapes for different values of the parameters. Applications to real data give strong evidence that the Alpha distribution is a competitive model, comparing to distributions already well-known in the literature, as for example Birnbaum-Saunders (BS), Burr XII, Gamma, Inverse Gaussian, Nadarajah-Haghighi (NH), Weibull, among others. In this sense, we show that the Alpha distribution is a flexible model and then its use as baseline applied to known generators becomes interesting for survival area.

It is important to emphatize that until now no extension of the Alpha distribution has been studied. In this way, we introduce the Exponentiated Alpha (Exp-Alpha) and the Kumaraswamy Alpha (Kw-Alpha) distributions, which both generalizes the Alpha model. The Exp-Alpha distribution is constructed using the exponentiated class (Gompertz, 1825; Verhulist, 1838, 1845, 1847), which is widely applied to baseline distributions in applications to real data set because this class provide a greater flexibility. The Exp-Alpha model has two shape parameters, one scale parameter and its hrf takes various forms depending on its shape parameters. The Kw-Alpha model is constructed using Cordeiro and de Castro's (2011) generator and has three shape parameters and one scale parameter, including as special models the Exp-Alpha, Lehmann type II Alpha and Alpha distributions.

We also propose the Exponentiated Generalized Exponentiated-Generated (EG-Exp-G) family of distributions, which is an extension of the Exponentiated Generalized (EG) class proposed by Cordeiro et al. (2013). The EG-Exp-G class includes as special cases the two classes of Lehmann's alternatives and the EG and Kw-G (Cordeiro and de Castro, 2011) classes of distributions. Some new distributions are proposed as submodels of this family, including the EG-Exp-Alpha distribution. Applications to real data sets indicate that, although the EG-Exp-G class introduced three additional parameters, this class is more flexible with respect to distributions already used to fit real data, such as Birnbaum-Saunders, gamma and Weibull models.

For all chapters present in this thesis, we study the shapes of the pdf and hrf of the proposed models, estimate the model parameters by maximum likelihood method, present a Monte Carlo simulation study to evaluate the assymptotic properties of the maximum likelihood estimators and provide applications to real data, which indicate that the proposed models are competitive distributions to fit real data. In addition, for Chapters 3, 4 and 5, we provide some mathematical properties, such as mean deviations, moments, moment generating function, order statistics and quantile function. The properties presented in Chapters 3 and 4 are valid for certain regions of the parametric space due to non existence of moments for the Alpha distribution. For EG-Exp-G class, we also provide a study for Rényi and Shannon entropies.

The main contribution of this thesis is to discuss the Alpha distribution, which was very competitive when compared with models well-known in the literature, and also show that the Alpha distribution can be used as baseline model for several classes of distributions already exist in the survival area, introducing new distributions that will provide great flexibility in modeling real data, which can be applied in several fields.

References

Aarset, M.V. (1987). How to identify bathtub hazard rate. *IEEE Transactions on Reliability*, **36**, 106-108.

Abdul-Moniem, I., Abdel-Hameed, H.F. (2012). On exponentiated Lomax distribution. *International Journal of Mathematical Education*, **33 (5)**: 1-7.

Abu-Zinadah, H.H., Aloufi, A.S. (2014). Some Characterizations of the Exponentiated Gompertz Distribution. *International Mathematical Forum*, **9**, 1427-1439.

Alexander, C., Cordeiro, G.M., Ortega, E.M.M., Sarabia, J.M. (2012). Generalized beta-generated distributions. *Computational Statistics and Data Analysis*, **56**, 1880-1897.

AL-Hussaini, E.K., Ahsanullah, M. (2015). Exponentiated Distributions, Vol. 5. Atlantis Studies in Probability and Statistics, USA.

Azzalini, A., Bowman, A. W. (1990). A look at some data on the Old Faithful geyser. *Applied Statistics*, **39**, 357-365.

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

Bourguignon, M., Silva, R.B., Zea, L.M., Cordeiro, G.M. (2013). The Kumaraswamy Pareto distribution. *Journal of Statistical Theory and Applications*, **12**, 129-144.

Burr, I.W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, **13**, 215-232.

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

Chhikara, R., Folks, L. (1989). The inverse Gaussian distribution: theory, methodology and applications. New York: Marcel Dekker.

Cooray, K., Ananda, M.M.A. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communication in Statistics - Theory and Methods*, **37**, 1323-1337.

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

Cordeiro, G.M., Lemonte, A.J. (2014). The exponentiated generalized Birnbaum-Saunders. *Applied Mathematics and Computation*, **247**, 762-779.

Cordeiro, G.M., Nadarajah, S., Ortega, E.M.M. (2012b). The Kumaraswamy Gumbel distribution. *Statistical Methods & Applications*, **21**, 139-168.

Cordeiro, G.M., Ortega E.M.M., Silva, G. (2011). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, **81**, 827-842.

Cordeiro, G.M., Ortega, E.M.M., Cunha, D.C.C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, **11**, 1-27.

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

Cordeiro, G.M., Ortega, E.M.M., Silva, G.O. (2012c). The Kumaraswamy modified Weibull distribution: theory and applications. *Journal of Statistical Computation and Simulation*, **84**, 1387-1411.

Cordeiro, G.M., Pescim, R.R., Ortega, E.M.M. (2012a). The Kumaraswamy Generalized Half-Normal Distribution for Skewed Positive Data. *Journal of Data Science (Online)*, **10**, 195-224.

Cordeiro, G.M., Saboor, A., Khan, M.N., Ozel, G., de Pascoa, M.A.R. (2016). The Kumaraswamy exponential-Weibull distribution: theory and applications. *Hacettepe Journal of Mathematics and Statistics*, **45**, 1203-1229.

de Pascoa, M.A.R., Ortega, E.M.M., Cordeiro, G.M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, **8**, 411-433.

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

Gomes, A.E., DA-SILVA, C.Q., Cordeiro, G.M., Ortega, E.M.M. (2014). A new lifetime model: the Kumaraswamy generalized Rayleigh distribution. *Journal of Statistical Computation and Simulation*, **84**, 290-309.

Gompertz, B. (1825). On the nature of the function of the law of human mortality and on a new model of determining the value of life contingencies. Philos. *Philosophical Transactions of the Royal Society of London*, **115**, 513-585.

Gradshteyn, I.S., Ryzhik, I.M. (2000). Table of integrals, series, and products. Academic Press, San Diego.

Gross, A.J., Clark, V.A. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. New York: John Wiley and Sons.

Gupta, R.C., Gupta, P.L., Gupta, R.D. (1998). Modeling failure time data by Lehmann alternatives. *Communications in Statistics - Theory and Methods*, **27**, 887-904.

Gupta, R.D., Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, **41**, 173-188.

Gupta, R.D., Kundu, D. (2001). Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, **43**, 117-130.

Hand, D.J., Daly, F., McConway, K., Lunn, D., Ostrowski, E. (1993). A Handbook of Small Data Sets. Chapman and Hall/CRC.

Hoffman, D., Karst, O. (1975). The Theory of the Rayleigh Distribution and Some of Its Applications. *Journal of Ship Research*, **19**, 172-191.

Johnson, N.L., Kotz, S., Balakrishnan, N. (1994). Continuous univariate distributions, Vol. 1. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2 edn), New York: John Wiley & Sons.

Katsev, P.G. (1968). Statistical methods in the study of cutting tools (Russian), Mashinostroyeniye, Moscow.

Kenney, J.F., Keeping, E.S. (1962). *Mathematics of Statistics*, 3 edn, Chapman & Hall Ltd, New

Jersey.

Khodabin, M., Ahmadabadi, A. (2010). Some properties of generalized gamma distribution. *Mathematical Sciences*, **4**, 9-28.

Klein and Moeschberger (1997). Survival Analysis Techniques for Censored and truncated data, Springer.

Lee, E.T., Wang, J.W. (2003). Statistical Methods for Survival Data Analysis, Third Edition, Wiley Series in Probability and Statistics.

Lemonte, A.J., Cordeiro, G.M. (2011). The exponentiated generalized inverse Gaussian distribution. *Statistics & Probability Letters*, **81**, 506-517.

Lima, S.R. (2015). The Half-Normal Generalized Family and Kumaraswamy Nadarajah-Haghighi Distribution. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco, Recife.

McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, **52**, 647-663.

Moors, J.J. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D*, **37**, 25-32.

Mudholkar, G., Srivastava, D. (1993). Exponentiated Weibull family for analyzing bathtub failure-real data. *IEEE Transaction on Reliability*, **42**, 299-302.

Mudholkar, G.S., Srivastava, D.K., Freimer, M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, **37**, 436-445.

Murtaugh, P.A., Dickson, E.R., Van Dam, G.M., Malinchoc, M., Grambsch, P.M. Langworthy, A.L., Gips, C.H. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*, **20(1.1)**, 126-34.

Nadarajah, S., Bakouch, H.S., Tahmasbi, R. (2011). A generalized lindley distribution. *Sankhya B*, **73**, 331-359.

Nadarajah, S., Gupta, A.K. (2007). The Exponentiated Gamma Distribution with Application to Drought Data. *Calcutta Statistical Association Bulletin*, **59**, 29-54.

Nadarajah, S., Haghighi, F. (2011). An extension of the exponential distribution. *Statistics*, **45**, 543-558.

Nadarajah, S., Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*, **92**, 97-111.

Paranaíba, P.F., Ortega, E.M.M., Cordeiro, G.M., de Pascoa, M.A.R. (2013). The Kumaraswamy Burr XII distribution: theory and practice. *Journal of Statistical Computation and Simulation*, **83**, 2117-2143.

Prudnikov, A.P., Brychkov, Y.A., Marichev, O.I. (1986). *Integrals and Series*, volumes 1, 2 and 3. Gordon and Breach Science Publishers, Amsterdam.

RÃ©nyi, A. (1961). On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 547-561.

Salvia, A.A. (1985). Reliability application of the Alpha distribution. *IEEE Transactions on Reliability*, **34**, 251-252.

Santana, T.V., Ortega, E.M., Cordeiro, G.M., Silva, G. (2012). The Kumaraswamy-Log-Logistic Distribution. *Journal of Statistical Theory and Applications*, **11**, 265-291.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-432.

Sherif, Y.S. (1983). Models for accelerated life-testing. Proc. Ann. Reliability and Maintainability Symp.

Silva, R.C., Sanchez, J.J.D., Lima, F.P., Cordeiro, G.M. (2015). The Kumaraswamy Gompertz Distribution. *Journal of Data Science (Print)*, **13**, 241-260.

Singh, V.P. (1998). Two-Parameter Log-Logistic Distribution. *Water Science and Tecnology Library*, **30**, 284-296.

Smith, R.L., Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, **36**, 358-369.

Verhulst, P.F. (1838). Notice sur la loi que la population pursuit dans son accroissement. *Correspondance Mathématique et physique, publiée par L.A.L. Quetelet*, **10**, 113-121.

Verhulst, P.F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* (i.e. Mémoires, Series 2), **18**, 1-42.

Verhulst, P.F. (1847). Deuxiéme mémoire sur la loi d'accroissement de la population. *Mémoires de l'Académie Royale des Sciences, des Lettres et de Beaux-Arts de Belgique*, **20(2)**, 1-32.

Wager, J.G., Barash, M.M. (1971). Study of the distribution of the life of HSS tools. *ASME Transactions*, **93**, 1044-1050.

Wagner, S.S., Altman, S.A. (1973). What time do the baboons come down from the trees? (An estimation problem). *Biometrics*, **29**, 623-35.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18 (3)**: 293-297.

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