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

FERNANDO ARTURO PEÑA RAMÍREZ

NEW GENERALIZED NADARAJAH-HAGHIGHI DISTRIBUTIONS IN SURVIVAL ANALYSIS

RECIFE

2017

Fernando Arturo Peña Ramírez

NEW GENERALIZED NADARAJAH-HAGHIGHI DISTRIBUTIONS IN SURVIVAL ANALYSIS

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

Advisor: Professor Dr. Gauss Moutinho Cordeiro

RECIFE

2017

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

R173n Ramírez, Fernando Arturo Peña
New generalized Nadarajah-Haghighi distributions in survival analysis / Fernando Arturo Peña Ramírez. – 2017.
114 f.: il., fig., tab.

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

1. Estatística aplicada. 2. Probabilidade. I. Cordeiro, Gauss Moutinho (orientador). II. Título.

310 CDD (23. ed.) UFPE- MEI 2017-135

Fernando Arturo Peña Ramirez

New generalized Nadarajah-Haghighi distributions in survival analysis

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

Aprovada em: 30 de maio de 2017.

BANCA EXAMINADORA

Prof. Gauss Moutinho Cordeiro
UFPE

Prof.ª Maria do Carmo Soares de Lima
UFPE

Prof. Abraão David Costa do Nascimento
UFPE

Prof. Rodrigo Bernardo da Silva
UFPB

Prof. Eufrásio de Andrade Lima Neto
UFPB

Este trabalho é especialmente dedicado

aos meus pais, Cecilia e Gregorio

Agradecimentos

Mais uma meta atingida, que sem eles com certeza não teria sido alcançada: meus pais, Cecilia e Gregorio. Infinita gratidão com eles

Ao Professor Dr. Gauss Cordeiro, pela magnífica orientação, paciência e disponibilidade nesses anos de trabalho. Obrigado Professor por acreditar em mim e ser refêrencia como pesquisador e como pessoa.

Aos meus irmãos Miguel, Eduardo, Lenny e Laura. Obrigado por ser suporte e me apoiar apesar da distância. Agradeço aos meus sobrinhos Sergio, Santi, Juan Felipe e a minha cunhada Viviana.

Aos membros da banca: Professora Maria do Carmo, Abraão, Rodrigo e Eufrasio. Obrigado pelas valiosas sugestões que fizeram a este trabalho.

A Renata que foi minha grande parceira de pesquisa nos anos de doutorado. Foram muitas coisas que aprendí dela e que fizeram parte deste trabalho. Muito obrigado pelos momentos compartilhados.

Aos meus grandes amigos que de alguma forma fizeram parte deste processo: Marianita (e donha Rosinha, claro), Pachito, Luz Milena, Juan Felipe, Luciana, Bruna Palm, Diego Canterle, Olivia, Miguel, Marcela, Ivan Roa, Piña, Natalie Primo, Kelly, Gabi, Pamela, e todos os demais que por limitação de espaço não mencionarei.

Aos meus colegas do doutorado pelas horas compartilhadas, especialmente Cláudio Tablada e Giannini Italino.

A Valéria Bittencourt de quem tenho um profundo carinho e gratidão por estos anos de amizade.

À CAPES, pelo apoio financeiro.

Abstract

The interest in developing new continuous distributions a remain important in statistical analysis. This topic is also important in survival analysis and has been used in many applications in fields like biological sciences, economics, engineering, physics, social sciences, among others. One reason is that the time of life or survival time is a random variable which can take constant, decreasing, increasing, upside-down bathtub (unimodal) and bathtub-shaped hazard rate functions. These new models can be defined by adding parameters to an existing distribution and considering the compounding approach, among other techniques. In this thesis, we consider these methods to propose four new continuous distributions, namely the *exponentiated generalized power Weibull*, *Nadarajah-Haghighi Lindley*, *Weibull Nadarajah-Haghighi* and *logistic Nadarajah-Haghighi* distributions. We provide a comprehensive mathematical and statistical treatment of these distributions and illustrate their flexibility through applications to real data sets. They are useful alternatives to other classical lifetime models.

Keywords: Exponential distribution. Generalized Weibull distribution. Lindley distribution. Nadarajah-Haghighi distribution. Lifetime data. Survival function.

Resumo

A geração de novas distribuições contínuas constitui uma importante área de pesquisa em Estatística. Este tópico é, também, importante na área de análise de sobrevivência e tem aplicações em outros campos do conhecimento, tais como, ciências biológicas, economia, engenheria, física, ciências sociais, entre outras. Uma das razões para generalizar uma distribuição conhecida é que a função de risco em forma generalizada é mais flexível podendo assumir padrão constante, crescente, decrescente, banheira invertida (unimodal) e forma de banheira. Estes novos modelos podem ser definidos adicionando parâmetros usando como base uma distribuição já existente ou fazendo composição de duas ou mais distribuições, entre outras técnicas. Nesta tese, consideramos esses métodos para propor quatro novas distribuições contínuas: as distribuições *exponentiated generalized power Weibull*, *Nadarajah-Haghighi Lindley*, *Weibull Nadarajah-Haghighi* e *logistic Nadarajah-Haghighi*. Estudamos importantes propriedades matemáticas e estatísticas dessas distribuições e evidenciamos a flexibilidade delas por meio de aplicações usando conjuntos de dados reais. As quatro novas distribuições constituem uma alternativa competitiva para outras distribuições clássicas para descrever dados de sobrevivência.

Palavras-chave: Dados de sobrevivência. Distribuição exponencial. Distribuição Lindley. Distribuição Nadarajah-Haghighi. Distribuição Weibull. Função de sobrevida.

List of Figures

List of Tables

List of Abbreviations and Acronyms

cdf	cumulative distribution function
EE	exponentiated exponential
EGNH	exponentiated generalized Nadarajah-Haghighi distribution
EGPW	exponentiated generalized power Weibull distribution
ENH	exponentiated Nadarajah-Haghighi distribution
EW	exponentiated Weibull distribution
exp-G	exponentiated-G
FW	flexible Weibull
GNH	gamma Nadarajah-Haghighi distribution
GPW	generalized power Weibull distribution
hrf	hazard rate function
KNH	Kumaraswamy Nadarajah-Haghighi
KS	Kolmogorov-Smirnov statistic
Kw-W	Kumaraswamy Weibull
LL	log-logistic distribution
LNH	logistic Nadarajah-Haghighi distribution
LX	*logistic-X* family of distributions
mgf	moment generating function
MLEs	maximum likelihood estimators
MW	modified Weibul distribution
NH	Nadarajah-Haghighi distribution
NHL	Nadarajah-Haghighi Lindley distribution
pdf	probability density function
qf	quantile function
RMSEs	root mean squared errors
SANN	simulated-annealing
STDs	sexually transmitted diseases
TTT	total time on test
W	Weibull distribution
Wexp	Weibull exponential distribution

Contents

Chapter 1

Introduction

The interest in developing new probability distributions remains important in statistical analysis. According to Lee *et al.* (2013) since 1980 the methodologies of generating new distributions have been focusing in adding parameters or combining existing distributions. The authors discussed five generating methods developed since then: (1) methods of generating skewed distributions, (2) the beta-generated method, (3) the method of adding parameters, (4) the transformed-transformer (T-X) method, and (5) the composite method. Tahir and Nadarajah (2015) wrote a survey with 312 reference papers on different distributions or G-classes of distributions, most of these introduced in the recent years.

This topic is also important in survival analysis and has been used in many applications in fields like biological sciences, economics, engineering, physics, social sciences, among others. One reason is that generalizing a known distribution might allow to the resulting model to accommodate non-monotone forms for the hazard rate function (hrf). Lai (2013) pointed out that the time of life or failure can have different interpretations depending on the area of applications. So, we can obtain more flexible distributions for modeling this kind of random variables.

Many recent distributions have been introduced by using power transformation method. Let $G(x)$ be the baseline cumulative distribution function (cdf) of a random variable X. We can obtain a new distribution by exponentiating the cdf of X as follows

$$F(x) = G(x)^{\alpha},$$

where $\alpha > 0$ is an additional shape parameter. From such power transformation, we can define the exponentiated exponential (EE) (Gupta *et al.*, 1998), exponentiated Weibull (EW) (Mudholkar and Srivastava, 1993), the exponentiated Nadarajah-Haghighi (ENH) (Lemonte, 2013), among several others distributions.

Another technique that has been considered is the compounding approach. They allow for greater flexibility of the tails and are motivated for engineering and biological applications. Besides, compounding families might be suitable for complementary risk problems based in the presence of latent risks. Adamidis and Loukas (1998) pionnered this method by introducing the exponential geometric distribution as the minimum of N independent and identical by exponential random variables, where N has geometric distribution. Since then, many other composed distributions have been proposed by taking the minimum of two distributions. For a continuous and other discrete distribution, see for example, the exponential Poisson (Kus, 2007), Weibull geometric (Barreto-Souza *et al.*, 2011) and Pareto Poisson-Lindley (Asgharzadeh *et al.*, 2013).

Cordeiro *et al.* (2014a) studied the exponential-Weibull lifetime distribution as the minimum between the exponential and Weibull random variables. Asgharzadeh *et al.* (2016) introduced the Weibull Lindley (WL) distribution by taking the minimum between the Lindley and Weibull random variables. In this approach, we have a composition by taking the minimum of two continuous independent random variables. It might be useful in engineering for modeling systems composed of two independent components in series.

In this thesis, we consider the adding parameters, composition and T-X methods to propose four new probability distributions, namely the *exponentiated generalized power Weibull* (EGPW), *Nadarajah-Haghighi Lindley* (NHL), *Weibull Nadarajah-Haghighi* (WNH) and *logistic Nadarajah-Haghighi* (LNH) distributions. In Chapter 2, we provide a comprehensive mathematical and statistical treatment of the EGPW distribution and illustrate the flexibility of the new model by means of applications to real data sets. This model is a useful extension of the Weibull distribution based on the Nadarajah Haghighi (NH) model with a power parameter, such as the ENH model. Its failure rate function takes the most common types of hazard functions and the model also includes as special cases some important distributions discussed in the literature.

The NHL distribution is introduced in Chapter 3. This new model is based on compounding the Lindley and Nadarajah Haghighi distributions. We study general mathematical properties of this distribution, such as mean residual life, ordinary and incomplete moments, moment generating function, mean deviations, Bonferroni and Lorenz curves and entropy. The method of maximum likelihood is used for estimating the model parameters and the observed information matrix is derived. We provide a applications to a real data set to ilustrate empirically its flexibility and potentiality as an useful alternative for other classical lifetime models.

In Chapter 4, we introduce the four-parameter WNH distribution, that is obtained by considering the Nadarajah-Haghighi distribution as baseline model in the generated Weibull-G family proposed by Bourguignon *et al.* (2014). The density function of the new distribution can be expressed as a linear combination of exponentiated Nadarajah-Haghighi densities, which is why some structural properties of the new model can be easily derived from the properties of those latter. The maximum likelihood method is presented to estimate the model parameters. A simulation study is performed. The usefulness of the new distribution is illustrated using two applications to real data.

Chapter 5 introduces the new three-parameter model LNH. It is obtained by inserting the Nadarajah-Haghighi distribution in the *logistic-X* family pionnered by Tahir *et al.* (2016a). As in the previous model, given that the density function of the new distribution can be expressed as a linear combination of exponentiated Nadarajah-Haghighi densities, several properties of the new distribution can be derived from the known properties of the exponentiated baseline model. In addition, we present explicit expressions for some statistical quantities. A simulation study is carried out to verify the precision of the estimates and we illustrate the usefulness of the new distribution by means of two applications to real data. Finally, in Chapter 6, we present some concluding remarks and outline some future research lines.

Chapter 2

A new Nadarajah-Haghighi generalization

Resumo

Neste Capitulo, introduzimos um novo modelo chamado distribuição *exponentiated generalized power Weibull*, que constitui uma útil generalização das distribuições exponencial, Weibull, Nadarajah-Haghighi, entre várias outras. O modelo proposto é flexível em modelar as quatro formas mais comuns da função de taxa de risco: crescente, decrescente, unimodal e de banheira. Além disso, é um modelo bastante competitivo frente a outras distribuições amplamente utilizadas, como as distribuições Weibull, exponencial, Weibull exponencializada, entre outras. Algumas propriedades matemáticas são estudadas. Consideramos a estimação dos parâmetros do novo modelo pelo método de máxima verossimilhança e realizamos uma simulação de Monte Carlo com o objetivo de avaliar essas estimativas. Também ilustramos empiricamente a utilidade da nova distribuição por meio de uma aplicação a dados reais.

Palavras-chave: Dados de tempo de vida. Distribuição exponencial. Distribuição Nadarajah-Haghighi. Distribuição Weibull potencia generalizada. Função de sobrevida.

Abstract

In this Chapter, we propose a new lifetime model called the *exponentiated generalized power Weibull* distribution, which is a useful generalization of the Weibull and Nadarajah-Haghighi

distributions, among others. The model is flexible in modeling the most common types of hazard rate functions. It is a very competitive model to the well-known Weibull, exponentiated exponential and exponentiated Weibull distributions, among others. Some of its mathematical properties are investigated. We discuss estimation of the model parameters by maximum likelihood. Simulation studies are performed and we provide one application to real data to ilustrate empirically the flexibility of the proposed distribution.

Keywords: Exponential distribution. Generalized power Weibull distribution. Lifetime data. Nadarajah-Haghighi distribution. Survival function.

2.1 Introduction

There has been an increased interest in defining new continuous distributions by adding shape parameters to an existing baseline model. One of the most widely-accepted methods on this parameter induction is the exponentiated-G (exp-G) class. Let $G(y)$ and $g(y)$ be the baseline cumulative distribution function (cdf) and the probability density function (pdf) of a random variable Y, respectively. We obtain the exp-G cdf by raising $G(y)$ to a positive exponent, which adds an extra power shape parameter to the baseline model. Thus, a random variable Y has an exp-G distribution if its cdf is given by

$$F(y) = G(y)^{\beta},$$

for $y \in \mathcal{D} \subseteq \mathbb{R}$ and $\beta > 0$ represents the additional parameter. The corresponding pdf is given by

$$f(y) = \beta\, g(x)\, G(y)^{\beta-1}.$$

Tahir and Nadarajah (2015) traced this approach back to the first half of the nineteenth century and found twenty-eight different exp-G models published in the recent literature. Most of these models are motivated by their usefulness in exploring tail properties and also for improving the goodness-of-fit in comparison with their baselines. Another current reason for introducing exp-G distributions is their applications in lifetime data analysis.

Thus, the classical lifetime distributions have been received great attention as baselines on the exp-G class, among other generated families. Using the exponential lifetime model as baseline, Gupta *et al.* (1998) pioneered the exponentiated exponential (EE) distribution. Gupta and Kundu (2001), Zheng (2002), Gupta and Kundu (2007), Abdel-Hamid and AL-Hussaini (2009) and Nadarajah (2011) provided several properties and applications of the EE distribution.

The exponentiated Weibull (EW) distribution was introduced by Mudholkar and Srivastava (1993). The mathematical properties of the EW distribution has been studied extensively by Mudholkar and Hutson (1996), Nassar and Eissa (2003), Nadarajah and Gupta (2005) and Nadarajah and Kotz (2006), among several other. These authors also showed that in practical situations the EW distribution can provide better fits than the traditional lifetime models, including exponential and Weibull distributions.

Another model that has been considered for modeling lifetime data is the Nadarajah-Haghighi (NH) distribution. Introduced by Nadarajah and Haghighi (2011), the NH model is a generalization of the exponential distribution with cdf given by (for $z > 0$)

$$G(z) = 1 - \exp\{1 - (1 + \lambda z)^{\alpha}\}, \tag{2.1}$$

where λ and α are the scale and shape parameters, respectively. If Z has the cdf (2.1), we write $Z \sim \mathrm{NH}(\alpha, \lambda)$. The pdf of Z is given by

$$g(z) = \alpha\lambda(1 + \lambda z)^{\alpha-1} \exp\{1 - (1 + \lambda z)^{\alpha}\}. \tag{2.2}$$

The motivations for studying the NH model are: the relationship between the pdf (2.2) and its hrf, the ability (or inability) to model data with mode fixed at zero and the fact that it can be interpreted as a truncated Weibull distribution. Further details and general properties can be found in Nadarajah and Haghighi (2011). The exponentiated Nadarajah-Haghighi (ENH) was proposed by Lemonte (2013).

The exponential, NH and Weibull distributions are all special cases of the generalized power Weibull (GPW) distribution, proposed by Bagdonavicius and Nikulin (2002) in the context of accelerated failure time models. They presented the cdf, pdf and hrf for this distribution, which are given by (for $t > 0$)

$$G(t) = 1 - \exp\{1 - (1 + \lambda t^{\gamma})^{\alpha}\}, \tag{2.3}$$

$$g(t) = \alpha\lambda\gamma t^{\gamma-1}(1 + \lambda t^{\gamma})^{\alpha-1}\exp\{1 - (1 + \lambda t^{\gamma})^{\alpha}\} \tag{2.4}$$

and

$$h(t) = \alpha\lambda\gamma t^{\gamma-1}(1 + \lambda t^{\gamma})^{\alpha-1},$$

respectively. Lai (2013) described the GPW among the Weibull generalizations that are often required to prescribe the nonmonotonic nature of the empirical hazard rates.

Nikulin and Haghighi (2006) introduced a chi-square statistic for testing the validity of GPW distribution and presented an application to censored survival times of cancer patients. Nikulin and Haghighi (2009) presented shape analysis for the GPW pdf and hrf. They also obtained a series representation for the sth moment of the GPW distribution, but only for integer values of s/γ. They do not provide a general expression for the GPW ordinary moments. We also note that there is a lack of studies exploring other structural properties of the GPW distribution, such as incomplete moments, skewness, mean deviations Bonferroni and Lorenz curves and Rényi entropy.

In this Chapter, we use the concept of exponentiated distributions for introducing a new four-parameter Weibull-type family, so-called the *exponentiated generalized power Weibull* (EGPW) distribution. The proposed distribution is obtained considering the GPW model as baseline in the exp-G family. Thus, the EGPW cdf and pdf are given by (for $t > 0$)

$$F(t) = [1 - \exp\{1 - (1 + \lambda t^{\gamma})^{\alpha}\}]^{\beta}, \tag{2.5}$$

and

$$f(t) = \alpha\beta\lambda\gamma t^{\gamma-1}\frac{(1 + \lambda t^{\gamma})^{\alpha-1}\exp\{1 - (1 + \lambda t^{\gamma})^{\alpha}\}}{[1 - \exp\{1 - (1 + \lambda t^{\gamma})^{\alpha}\}]^{1-\beta}}, \tag{2.6}$$

respectively. Here, λ is the scale parameter and γ, α and β are shape parameters. Henceforth, we denote by T a random variable having cdf (2.5), say $T \sim \text{EGPW}(\alpha, \beta, \lambda, \gamma)$. Identifiability is a property which a model must satisfy in order for precise inference to be possible, which refers to whether the parameters unknown in the model can be uniquely estimated. Equation (2.5) is clearly identifiable.

The hrf of T is given by

$$\begin{aligned} h(t) =& \alpha\beta\lambda\gamma t^{\gamma-1}(1+\lambda t^{\gamma})^{\alpha-1} \\ & \times \frac{\exp\{1-(1+\lambda t^{\gamma})^{\alpha}\}\left[1-\exp\{1-(1+\lambda t^{\gamma})^{\alpha}\}\right]^{\beta-1}}{1-\left[1-\exp\{1-(1+\lambda t^{\gamma})^{\alpha}\}\right]^{\beta}}. \end{aligned} \tag{2.7}$$

By inverting (2.5), we obtain an explicit expression for the quantile function (qf), of the EGPW distribution, say $Q(u)$, as

$$Q(u) = \lambda^{-1/\gamma}\left\{\left[1-\log(1-u^{1/\beta})\right]^{1/\alpha}-1\right\}^{1/\gamma}, \quad u \in (0,1). \tag{2.8}$$

Its median follows by setting $u = 1/2$. The simulation of the EGPW random variable is straightforward. If $U \sim U(0,1)$, then the random variable $T = Q(U)$ follows the EGPW distribution given by (2.6).

Some motivations for introducing the EGPW distribution are:

- The new distribution is quite flexible because it contains several well-known lifetime distributions as special models, see Table 2.1. This feature is also suitable for testing the goodness of fit for these distributions.

Table 2.1: Some special models of the EGPW distribution.

α	β	λ	γ	Distribution
1	1	-	1	Exponential
1	-	-	1	Exponentiated exponential
1	1	-	2	Rayleigh
1	-	-	2	Burr type X
1	1	-	-	Weibull
1	-	-	-	Exponentiated Weibull
-	1	-	1	Nadarajah-Haghighi
-	-	-	1	Exponentiated Nadarajah-Haghighi
-	1	-	-	Generalized Power Weibull

- The current distribution can also be derived from a power transform on an ENH random variable. Let $Y \sim$ENH(α, β, λ), the cdf of Y is given by

$$F_Y(y) = [1 - \exp\{1 - (1 + \lambda y)^{\alpha}\}]^{\beta}, \quad y > 0,$$

where $\alpha > 0$, $\beta > 0$ and $\lambda > 0$. Now, consider the transform $T = Y^{1/\gamma}$, where $\gamma > 0$. Thus, the cdf of T has the form $F(t) = F_Y(t^{\gamma})$ given by (2.5). A similar approach was considered by Gomes *et al.* (2008). They proposed a new method of estimation for the generalized gamma distribution through the power transformation $W = X^c$, where X is a generalized gamma random variable and W has the gamma distribution.

- Once several structural properties on the GPW distribution have not been studied, they shall be obtained from those on the EGPW distribution.

- By pioneering a GPW generalization on the exp-G family, it is also possible to obtain several properties on other generated families based on linear combinations from those of the EGPW distribution. For example, for the beta-G family (Eugene *et al.*, 2002) the density function can be expressed as a mixture of exp-G pdfs for any baseline. Similar results can also be demonstrated for the Kumaraswamy-G introduced by Cordeiro and Castro (2011), the McDonald-G by Alexander *et al.* (2012) and the gamma-G by Zografos and Balakrishnan (2009), among several others generated families of distributions.

- Let $\beta > 0$ be an integer. Thus, $F(t)$ given in (2.5) represents the cdf of the maximum value on a β-variate random sample from the GPW distribution, say: $T = \max\{T_1, \ldots, T_\beta\}$. In other words, the EGPW distribution can be used to model the maximum lifetime of a random sample from the GPW distribution with size β. Further, as part of the exp-G family, the EGPW distribution has the following physical interpretation. Consider a parallel system consisting of $\beta = n$ components, which means that the system works if at least one of the n-components works. If the lifetime distributions of the components are independent and identically distributed GPW random variables, then the lifetime distribution of the system becomes the EGPW cdf with power parameter $\beta = n$.

- The EGPW may provide consistently 'better fits' than other Weibull generalizations, including its special models. This fact is well-demonstrated by fitting the proposed distribution for a data sets in Section 2.11. The applications also illustrate that the EGPW distribution can also be very competitive to other widely known lifetime models.

The Chapter is outlined as follows. Some mathematical properties of the new distribution are provided in Sections 2.3-2.8. They include ordinary and incomplete moments, mean deviations about the mean and the median, Bonferroni and Lorenz curves and Rényi entropy. In Section 2.9, we present the maximum likelihood method to estimate the model parameters. In Section 2.10, a simulation study evaluates the performance of the maximum likelihood estimators (MLEs). Applications for two real data sets are presented in Section 2.11. Section 2.12 presents some concluding remarks.

2.2 Density and hazard shapes

Note that the pdf (2.6) can be expressed in terms of the cdf and pdf given in (2.3) and (2.4), respectively, in the form $f(t) = \beta G(t)^{\beta-1} g(t)$. Thus, the multiplicative factor $\beta G(t)^{\beta-1}$ is greater (smaller) than one for $\beta > 1$, ($\beta < 1$) and for larger values of t, and the opposite occurs for smaller values of t. The inclusion of the new shape parameter β, provides greater flexibility in terms of skewness and kurtosis in the new distribution. The pdf (2.6) can takes various forms depending on the values of the α, β and γ shape parameters. It is easy to verify that

$$\lim_{t\to 0} f(t) = \begin{cases} \infty & \text{if } \beta < 1, \\ \alpha\lambda\gamma & \text{if } \beta = 1, \\ 0 & \text{if } \beta > 1, \end{cases}$$

and

$$\lim_{t\to\infty} f(t) = 0.$$

Seting $z = (1 + \lambda t^{\gamma})^{\alpha}$ we can rewrite the EGPW pdf as

$$\psi(z) = \alpha\beta\lambda^{1/\gamma}\gamma\, z^{(\alpha-1)/\alpha}(z^{1/\alpha} - 1)^{(\gamma-1)/\gamma}\mathrm{e}^{1-z}(1 - \mathrm{e}^{1-z})^{\beta-1}.$$

Differentiating twice $\log\psi(z)$ with respect to z, we arrive that

$$\frac{\mathrm{d}^2\log\psi(z)}{\mathrm{d}z^2}=-\left[\frac{\alpha-1}{\alpha z^2}+\frac{(\beta-1)\mathrm{e}^{1-z}}{(1-\mathrm{e}^{1-z})^2}+\frac{(\gamma-1)z^{[2(1-\alpha)/\alpha]}}{\alpha^2\gamma(z^{1/\alpha}-1)}\right].$$

Note that $z=(1+\lambda t^\gamma)^\alpha$ implies that $z>1$. Thus, we can verify that for $t>0$, $\alpha<1$, $\beta<1$ and $\gamma<1$, $[\mathrm{d}^2\log\psi(z)/\mathrm{d}z^2]>0$. It implies that the EGPW pdf is log-convex. Further, for $t>0$, $\alpha>1$, $\beta>1$ and $\gamma>1$, $[\mathrm{d}^2\log\psi(z)/\mathrm{d}z^2]<0$. It implies that the EGPW pdf is log-concave. Figure 2.1 displays plots of the pdf (2.5) for some parameter values. It illustrates the flexibility of the EGPW density, which allows modeling skewed and asymmetrical data.

Analogously, the EGPW hrf can be rewritten as

$$\phi(z)=\alpha\beta\lambda^{1/\gamma}\gamma\, z^{(\alpha-1)/\alpha}(z^{1/\alpha}-1)^{(\gamma-1)/\gamma}\,\frac{\mathrm{e}^{1-z}(1-\mathrm{e}^{1-z})^{-1}}{(1-\mathrm{e}^{1-z})^{-b}-1}.$$

The critical point are obtained from the equation

$$\frac{\mathrm{d}\log\phi(z)}{\mathrm{d}z}=\frac{\alpha-1}{\alpha z}+\frac{(\gamma-1)z^{(1-\alpha)/\alpha}}{\alpha\gamma\,(z^{1/\alpha}-1)}+\frac{\beta\mathrm{e}^{1-z}}{(1-\mathrm{e}^{1-z})[1-(1-\mathrm{e}^{1-z})]}-\frac{\mathrm{e}^{1-z}}{1-\mathrm{e}^{1-z}}-1=0.$$

For $z>1$ and $\alpha=\beta=\gamma=1$, $[\mathrm{d}\log\phi(z)/\mathrm{d}z]=0$, and failure rate function is constant. For $\alpha<1$, $\gamma<1$ and $\beta<1$, $\mathrm{d}\log\phi(z)/\mathrm{d}z<0$ and the hrf is decreasing. There may be more than one root to (2.7).

Figure 2.2 provides plots of the hrf (2.7) for some parameter values, revealing that the EGPW distribution can have decreasing, increasing, upside-down bathtub and bathtub-shaped hazard functions. This feature makes the new distribution very attractive to model lifetime data. For example, according to Nadarajah *et al.* (2011) most empirical life systems have bathtub shapes for their hrfs.

2.3 Moments

From equation (2.8), and after some algebra, we can write

$$\mu_s'=\mathbb{E}(T^s)=\beta\lambda^{-s/\gamma}I_s(\alpha,\beta,\gamma),$$

where $I_s(\alpha,\beta,\gamma)=\int_0^1\{[1-\log(1-u)]^{1/\alpha}-1\}^{s/\gamma}u^{\beta-1}\mathrm{d}u$ is an integral to be evaluated numerically.

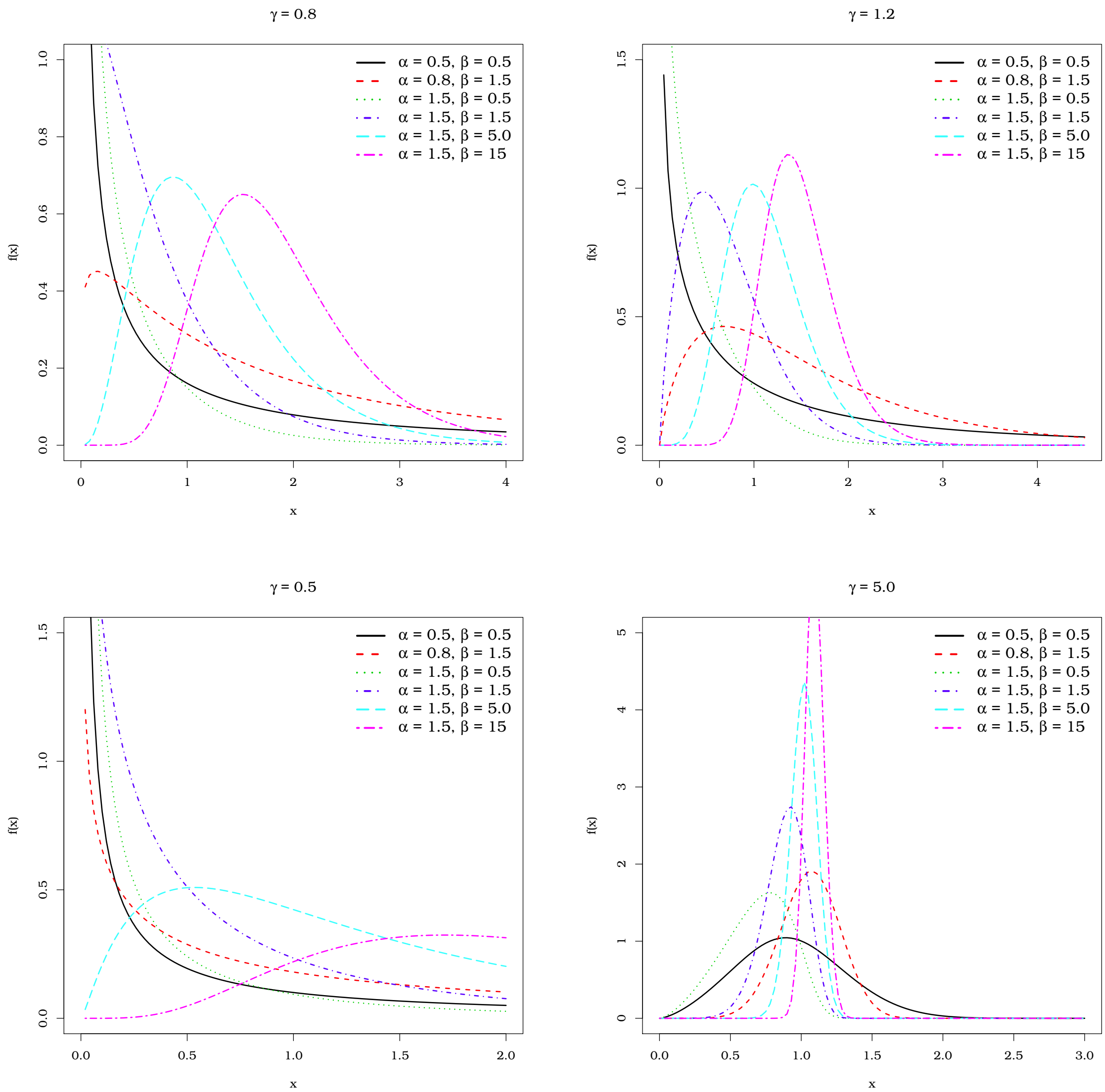


Figure 2.1: Plots of the EGPW density for $\lambda = 1$.

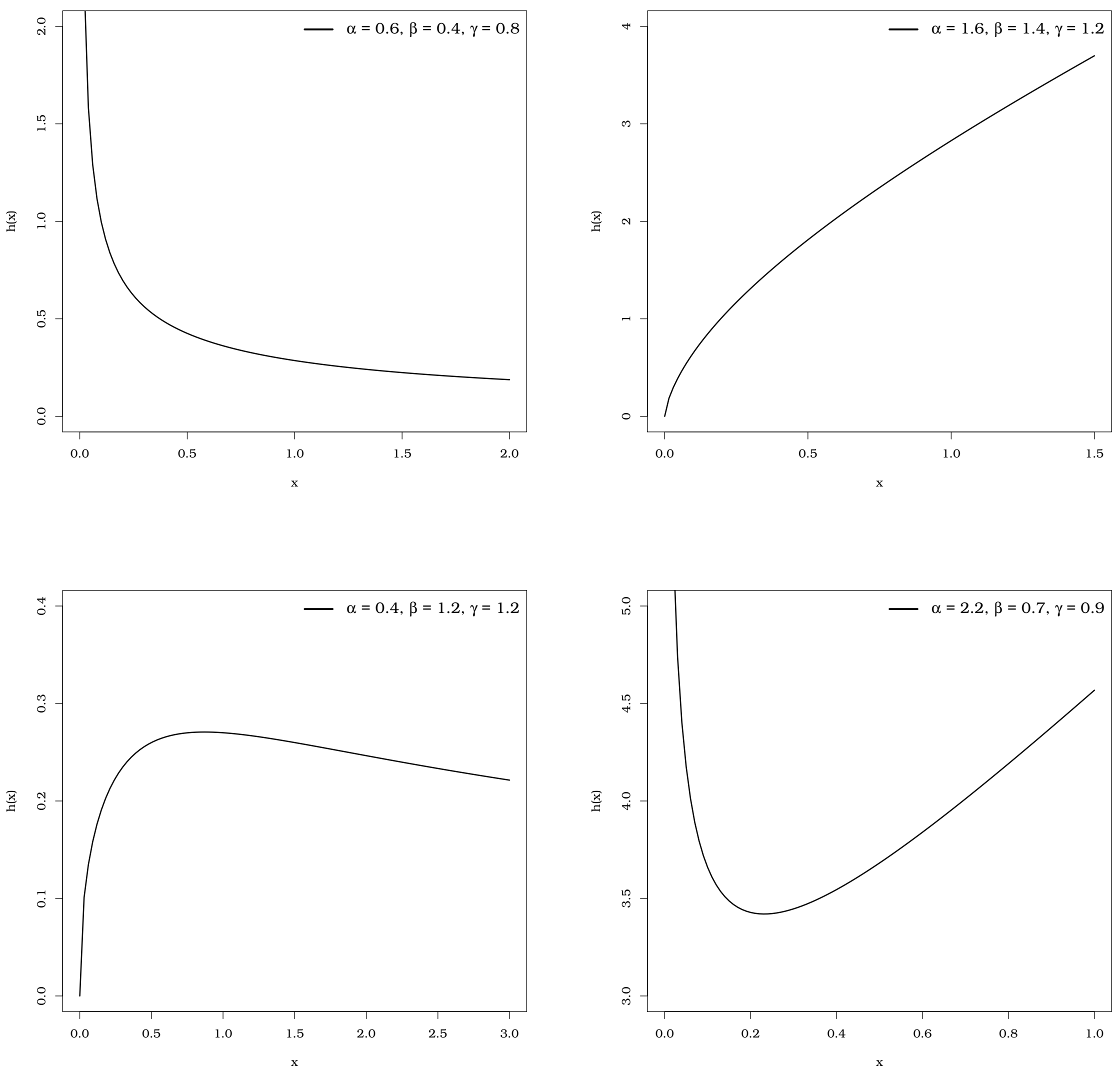


Figure 2.2: Plots of the EGPW hrf for $\lambda = 1$.

Using the binomial expansion since $0 < \mathrm{e}^{1-(1+\lambda t^{\gamma})^{\alpha}} < 1$, the denominator of (2.6) can be expressed as

$$[1-\exp\{1-(1+\lambda t^{\gamma})^{\alpha}\}]^{\beta-1} = \sum_{j=0}^{\infty}(-1)^{j}\binom{\beta-1}{j}\mathrm{e}^{j[1-(1+\lambda t^{\gamma})^{\alpha}]}.$$

We can rewrite μ_s' as

$$\mu_s' = \alpha\beta\lambda\gamma\sum_{j=0}^{\infty}(-1)^{j}\binom{\beta-1}{j}\mathrm{e}^{j+1}\int_0^{\infty} t^{s+\gamma-1}(1+\lambda t^{\gamma})^{\alpha-1}\mathrm{e}^{-(j+1)(1+\lambda t^{\gamma})^{\alpha}}\mathrm{d}t. \tag{2.9}$$

We consider the integral

$$J = \int_0^{\infty} t^{s+\gamma-1}(1+\lambda t^{\gamma})^{\alpha-1}\,\mathrm{e}^{-(j+1)(1+\lambda t^{\gamma})^{\alpha}}\mathrm{d}t.$$

Setting $u = (j+1)(1+\lambda t^{\gamma})^{\alpha}$, we have

$$t = \left\{\lambda^{-1}\left[\left(\frac{u}{j+1}\right)^{1/\alpha}-1\right]\right\}^{1/\gamma}.$$

Hence, after some algebra, we obtain

$$J = \left(\frac{1}{\lambda}\right)^{s/\gamma}\int_{j+1}^{\infty}\left[\left(\frac{u}{j+1}\right)^{1/\alpha}-1\right]^{s/\gamma}\frac{\mathrm{e}^{-u}}{\alpha\lambda\gamma(j+1)}\,\mathrm{d}u. \tag{2.10}$$

The most general case of the binomial theorem is the power series identity

$$(x+a)^{\nu} = \sum_{k=0}^{\infty}\binom{\nu}{k}x^{k}\,a^{\nu-k}, \tag{2.11}$$

where $\binom{\nu}{k}$ is a binomial coefficient and ν is a real number. This power series converges for $\nu \geq 0$ an integer, or $|x/a| < 1$. This general form is from Graham (1994). By using (2.11) in equation (2.10), since $\left|[u/(j+1)]^{1/\alpha}\right| < 1$, it follows from (2.9) that

$$\mu_s' = \beta\lambda^{-s/\gamma}\sum_{i,j=0}^{\infty}\frac{(-1)^{i+j}\,\mathrm{e}^{j+1}}{(j+1)^{[s-\gamma(i-\alpha)]/\alpha\gamma}}\binom{\beta-1}{j}\binom{s/\gamma}{i}\Gamma\left(\frac{s-\gamma(i-\alpha)}{\alpha\gamma},\ j+1\right), \tag{2.12}$$

where $\Gamma(a,x) = \int_x^{\infty} z^{a-1}\,\mathrm{e}^{-z}\mathrm{d}z$ denotes the complementary incomplete gamma function, which is defined for all real numbers except the negative integers.

2.4 Incomplete moments

The sth incomplete moment of T is defined by $m_s(y) = \int_0^y t^s f(t)\mathrm{d}t$. Using a suitable substitution of variable and (2.8), it follows that

$$m_s(y) = \beta\lambda^{-s/\gamma} \int_0^{1-\mathrm{e}^{1-(1+\lambda y^\gamma)^\alpha}} \{[1-\log(1-u)]^{1/\alpha} - 1\}^{s/\gamma} u^{\beta-1}\mathrm{d}u.$$

Substituting the upper limit of the integral in (2.9) for y, setting $u = (j+1)(1+\lambda t^\gamma)^\alpha$ and using (2.11), we have

$$m_s(y) = \beta\lambda^{-s/\gamma} \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}\,\mathrm{e}^{j+1}}{(j+1)^{[s-\gamma(i-\alpha)]/\alpha\gamma}} \binom{\beta-1}{j}\binom{s/\gamma}{i} \int_{j+1}^{(j+1)(1+\lambda y^\gamma)^\alpha} u^{(s-\gamma i)/(\alpha\gamma)}\mathrm{e}^{-u}\mathrm{d}u.$$

Hence, an alternative expression for $m_s(y)$ takes the form

$$\begin{aligned} m_s(y) = \beta\lambda^{-s/\gamma} &\sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}\,\mathrm{e}^{j+1}}{(j+1)^{[s-\gamma(i-\alpha)]/\alpha\gamma}} \binom{\beta-1}{j}\binom{s/\gamma}{i} \\ &\left[\Gamma\left(\frac{s-\gamma(i-\alpha)}{\alpha\gamma},\ j+1\right) - \Gamma\left(\frac{s-\gamma(i-\alpha)}{\alpha\gamma},\ (j+1)(1+\lambda\, y^\gamma)^\alpha\right)\right]. \end{aligned}$$

2.5 Skewness

The central moments (μ_s) and cumulants (κ_s) of T can be expressed recursively from equation (2.12) as

$$\mu_s = \sum_{k=0}^{s} (-1)^k \binom{s}{k} \mu_1'^k \mu_{s-k}' \qquad \text{and} \qquad \kappa_s = \mu_s' - \sum_{k=1}^{s-1} \binom{s-1}{k-1} \kappa_k\, \mu_{s-k}',$$

respectively, where $\kappa_1 = \mu_1'$. Thus, $\kappa_2 = \mu_2' - \mu_1'^2, \kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$, etc. The skewness $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and kurtosis $\gamma_2 = \kappa_4/\kappa_2^2$ can be determined from the third and fourth standardized cumulants.

The MacGillivray's skewness function of T is given by

$$\rho(u) = \rho(u;\alpha,\beta,\gamma) = \frac{\rho_{(1)}(u;\alpha,\beta,\gamma)}{\rho_{(2)}(u;\alpha,\beta,)} = \frac{Q(1-u) + Q(u) - 2Q(1/2)}{Q(1-u) - Q(u)},$$

where $u \in (0,\, 1)$, $Q(\cdot)$ is the qf defined in (2.8),

$$\begin{aligned}\rho_{(1)}(u;\alpha,\beta,\gamma) &= \left\{\left[1-\log(1-(1-u)^{1/\beta})\right]^{1/\alpha}-1\right\}^{1/\gamma}\\ &+ \left\{\left[1-\log(1-u^{1/\beta})\right]^{1/\alpha}-1\right\}^{1/\gamma}\\ &- 2\left\{\left[1+\beta^{-1}\log(2)-\log(2^{1/\beta}-1)\right]^{1/\alpha}-1\right\}^{1/\gamma}\end{aligned}$$

and

$$\begin{aligned}\rho_{(2)}(u;\alpha,\beta,\gamma) &= \left\{\left[1-\log(1-(1-u)^{1/\beta})\right]^{1/\alpha}-1\right\}^{1/\gamma}\\ &- \left\{\left[1-\log(1-u^{1/\beta})\right]^{1/\alpha}-1\right\}^{1/\gamma}.\end{aligned}$$

It is based on quantiles and illustrates the effects of the shape parameters α, β and γ on the skewness, see MacGillivray (1986). Plots of $\rho(u)$ for some parameter values are displayed in Figure 2.3. These plots reveal that when the parameters β and γ increase, the function $\rho(u)$ is close to zero. The closer $\rho(u)$ is to the horizontal line $\rho(u) = 0$, the density is more symmetrical. The quantity $\rho(u)$ does not depend on the parameter λ since it is a scale parameter.

2.6 Mean deviations

The mean deviations about the mean ($\delta_1 = \mathbb{E}(|T-\mu_1'|)$) and about the median ($\delta_2 = \mathbb{E}(|T-M|)$) of T can be expressed as

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

respectively, where $\mu_1' = \mathbb{E}(T)$, $M = \mathrm{Median}(T) = Q(0.5)$ is the median, $F(\mu_1')$ is easily determined from (2.5) and $m_1(y) = \int_0^y t\,f(t)\mathrm{d}t$ is the first incomplete moment. Hence, we can write

$$m_1(y) = \beta\lambda^{-1/\gamma}\int_0^{1-\mathrm{e}^{1-(1+\lambda y^{\gamma})^{\alpha}}} \{[1-\log(1-u)]^{1/\alpha}-1\}^{1/\gamma}u^{\beta-1}\mathrm{d}u.$$

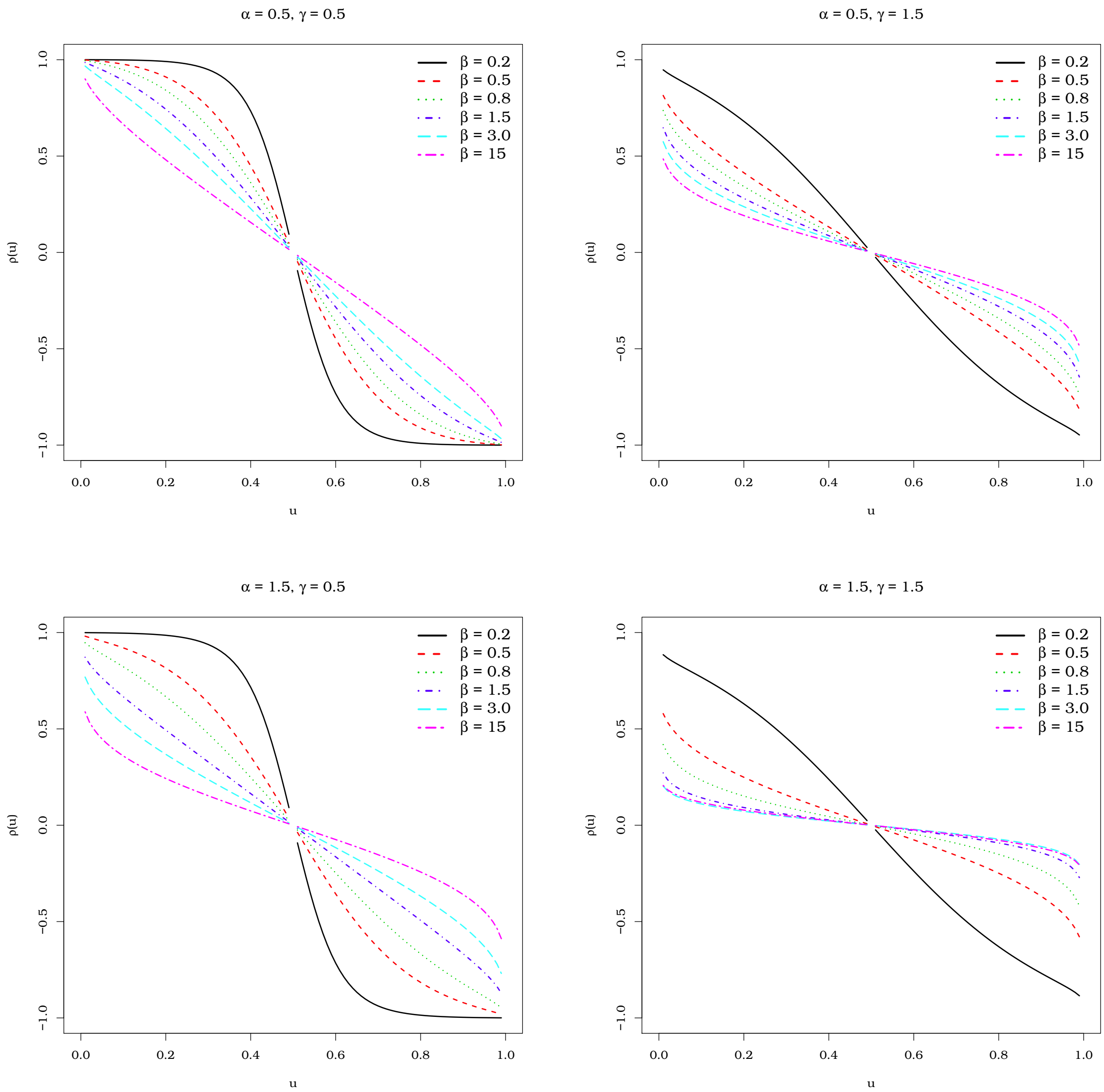


Figure 2.3: The MacGillivray's skewness of the EGPW distribution.

Alternatively, we can determine $m_1(y)$ as

$$m_1(y) = \beta\lambda^{-1/\gamma} \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}\,\mathrm{e}^{j+1}}{(j+1)^{[1-\gamma(i-\alpha)]/\alpha\gamma}} \binom{\beta-1}{j}\binom{1/\gamma}{i}$$
$$\left[\Gamma\left(\frac{1-\gamma(i-\alpha)}{\alpha\gamma},\ j+1\right) - \Gamma\left(\frac{1-\gamma(i-\alpha)}{\alpha\gamma},\ (j+1)(1+\lambda\, y^{\gamma})^{\alpha}\right)\right].$$

2.7 Bonferroni and Lorenz curves

Applications of the previous results to the Bonferroni and Lorenz curves are important in several fields such as economics, demography, insurance and medicine. They are defined, for a given probability π, by $B(\pi) = m_1(q)/(\pi\mu_1')$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $q = Q(\pi)$ follows from (2.8). The Gini concentration (C_G) is defined as the area between the curve $L(\pi)$ and the straight line. Hence,

$$C_G = 1 - 2\int_0^1 L(\pi)\mathrm{d}u.$$

An alternative expression is $C_G = (2\delta - \mu_1')/\mu_1'$, where $\delta = \mathbb{E}[TF(T)] = \int_{-\infty}^{\infty} tF(t)f(t)\mathrm{d}t$. Setting $u = F(t)$, and after some algebra, the quantity δ can be expressed as

$$\delta = \beta\lambda^{-1/\gamma}\int_0^1 u^{2\beta-1}\{[1-\log(1-u)]^{1/\alpha} - 1\}^{1/\gamma}\mathrm{d}u.$$

This integral can be easily evaluated numerically in software such as `R` and `Ox`, among others. An alternative expression for δ takes the form

$$\delta = \beta\lambda^{-1/\gamma} \sum_{i,j=0}^{\infty} \frac{(-1)^{i+j}\,\mathrm{e}^{j+1}}{(j+1)^{[1-\gamma(i-\alpha)]/\alpha\gamma}} \binom{2\beta-1}{j}\binom{1/\gamma}{i}$$
$$\times\,\Gamma\left(\frac{1-\gamma(i-\alpha)}{\alpha\gamma},\ j+1\right).$$

For $\gamma = 1$ we can prove that this expression reduces to that one obtained by Lemonte (2013).

2.8 Entropy

The entropy of a random variable is a measure of variation of the uncertainty. Entropy measure has been used in many applications in fields like physics, engineering, and economics.

Several measures of entropy have been studied in the literature. However, we consider the most popular entropy measure: Rényi entropy (Renyi, 1961). The Rényi entropy of a random variable with pdf $f(x)$ is defined by

$$I_R = I_R(\delta) = \frac{1}{1-\delta} \log \left[\int_{-\infty}^{\infty} f^{\delta}(x) \mathrm{d}x \right],$$

for $\delta > 0$ and $\delta \neq 1$. The Rényi entropy of T, setting $u = (1 + \lambda t^{\gamma})^{\alpha}$, can be expressed as

$$I_R = M + \frac{1}{1-\delta} \log \left(\int_1^{\infty} \frac{u^{\alpha^{-1}(\alpha-1)(\delta-1)} (u^{1/\alpha} - 1)^{\gamma^{-1}(\gamma-1)(\delta-1)} \mathrm{e}^{\delta(1-u)}}{[1 - \mathrm{e}^{1-u}]^{\delta(1-\beta)}} \mathrm{d}u \right),$$

where $M = -\log(\alpha\gamma\lambda^{\gamma}) + \frac{\delta}{1-\delta}\log(\beta)$. The above integral can be evaluated numerically. By expanding the denominator using the binomial expansion, we have

$$\begin{aligned} I_R &= M + \frac{1}{1-\delta} \log \left[\sum_{j=0}^{\infty} (-1)^j \mathrm{e}^{\delta+j} \binom{\delta(\beta-1)}{j} \right. \\ &\times \left. \left(\int_1^{\infty} u^{\alpha^{-1}(\alpha-1)(\delta-1)} (u^{1/\alpha} - 1)^{\gamma^{-1}(\gamma-1)(\delta-1)} \mathrm{e}^{-u(\delta+j)} \mathrm{d}u \right) \right]. \end{aligned}$$

Again, by using the binomial expansion, I_R can by expressed as

$$\begin{aligned} I_R &= M + \frac{1}{1-\delta} \log \left[\sum_{j,k=0}^{\infty} \frac{(-1)^{j+k} \mathrm{e}^{\delta+j}}{(j+\delta)^{[\delta(\gamma\alpha-1)+1]/\gamma\alpha}} \right. \\ &\times \left. \binom{\delta(\beta-1)}{j} \binom{\frac{(\gamma-1)(\delta-1)}{\gamma}}{k} \Gamma\left(\frac{\delta(\gamma\alpha-1)+1}{\gamma\alpha}, j+\delta \right) \right]. \end{aligned}$$

For $\gamma = 1$, the last expression reduces to

$$\begin{aligned} I_R = & - \log(\alpha\lambda) + \frac{\delta}{1-\delta}\log(\beta) + \frac{1}{1-\delta} \\ & \times \log \left[\sum_{j=0}^{\infty} \frac{(-1)^j \mathrm{e}^{\delta+j}}{(j+\delta)^{[\delta(\alpha-1)+1]/\alpha}} \binom{\delta(\beta-1)}{j} \Gamma\left(\frac{\delta(\alpha-1)+1}{\alpha}, j+\delta \right) \right], \end{aligned}$$

which agrees with the result by Lemonte (2013).

2.9 Maximum likelihood estimation

This section addresses the estimation of the unknown parameters of the EGPW distribution by the maximum likelihood method. Let $t_1, \ldots, t_n$ be a observed sample of size n from the

EGPW$(\alpha, \beta, \lambda, \gamma)$ distribution. Let $\boldsymbol{\theta} = (\alpha, \beta, \lambda, \gamma)^T$ be the parameter vector of interest. The log-likelihood function for $\boldsymbol{\theta}$ based on this sample is

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= n + n\log(\alpha\,\beta\,\lambda\,\gamma) + (\gamma-1)\sum_{i=1}^{n}\log(t_i) - \sum_{i=1}^{n}(1+\lambda\,t_i^{\gamma})^{\alpha} \\ &+ (\alpha-1)\sum_{i=1}^{n}\log(1+\lambda\,t_i^{\gamma}) + (\beta-1)\sum_{i=1}^{n}\log\left[1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}\right]. \end{aligned} \tag{2.13}$$

The components of the score vector $\boldsymbol{U(\theta)} = [U_\alpha(\boldsymbol{\theta}), U_\beta(\boldsymbol{\theta}), U_\lambda(\boldsymbol{\theta}), U_\gamma(\boldsymbol{\theta})]^\top$ are given by

$$\begin{aligned} U_\alpha(\boldsymbol{\theta}) = \frac{n}{\alpha} &+ \sum_{i=1}^{n}\log(1+\lambda\,t_i^{\gamma}) - \sum_{i=1}^{n}(1+\lambda\,t_i^{\gamma})^{\alpha}\log(1+\lambda\,t_i^{\gamma}) \\ &+ (\beta-1)\sum_{i=1}^{n}\frac{(1+\lambda\,t_i^{\gamma})^{\alpha}\log(1+\lambda\,t_i^{\gamma})\,\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}}{1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}}, \end{aligned}$$

$$U_\beta(\boldsymbol{\theta}) = \frac{n}{\beta} + \sum_{i=1}^{n}\log\left[1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}\right],$$

$$\begin{aligned} U_\lambda(\boldsymbol{\theta}) = \frac{n}{\lambda} &+ (\alpha-1)\sum_{i=1}^{n}t_i^{\gamma}\,(1+\lambda\,t_i^{\gamma})^{-1} - \alpha\sum_{i=1}^{n}t_i^{\gamma}\,(1+\lambda\,t_i^{\gamma})^{\alpha-1} \\ &+ \alpha\,(\beta-1)\sum_{i=1}^{n}\frac{t_i^{\gamma}\,(1+\lambda\,t_i^{\gamma})^{\alpha-1}\,\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}}{1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}}, \end{aligned}$$

and

$$\begin{aligned} U_\gamma(\boldsymbol{\theta}) = \frac{n}{\gamma} &+ \sum_{i=1}^{n}\log(t_i) - \alpha\lambda\sum_{i=1}^{n}t_i^{\gamma}\log(t_i)\,(1+\lambda\,t_i^{\gamma})^{\alpha-1} \\ &+ \lambda(\alpha-1)\sum_{i=1}^{n}t_i^{\gamma}\log(t_i)\,(1+\lambda\,t_i^{\gamma})^{-1} \\ &+ \lambda\alpha\,(\beta-1)\sum_{i=1}^{n}\frac{t_i^{\gamma}\log(t_i)\,(1+\lambda\,t_i^{\gamma})^{\alpha-1}\;\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}}{1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}}. \end{aligned}$$

Setting the above equations to zero, $\boldsymbol{U(\theta)} = \mathbf{0}$, and solving them simultaneously yields the MLEs of the four parameters. These equations can not be solved analytically. We have to use iterative techniques such as the quasi-Newton BFGS and Newton-Raphson algorithms, see Press *et al.* (2007). The initial values for the parameters are important but are not hard to obtain from fitting special EGPW sub-models. Note that, for fixed α, λ and γ, the MLE of β is given by

$$\hat{\beta}(\hat{\alpha},\hat{\lambda},\hat{\gamma}) = -\frac{n}{\sum_{i=1}^{n}\log[1-\mathrm{e}^{1-\left(1+\hat{\lambda}t_i^{\hat{\gamma}}\right)^{\hat{\alpha}}}]}.$$

Thus, it is easily observed that fixed on $t_1,\ldots,t_n$,

- $\hat{\beta}\to 0$ when $\hat{\alpha}\to 0$ and/or $\hat{\lambda}\to 0$
- $\hat{\beta}\to \infty$ when $\hat{\alpha}\to 0$ and/or $\hat{\lambda}\to \infty$
- $\hat{\beta}\to 0$ when $\hat{\gamma}\to\infty$ and $t_i<1$, for some $i\leq n$
- $\hat{\beta}\to \infty$ when $\hat{\gamma}\to\infty$ and $t_i<1$, $\forall i\leq n$.

This behavior anticipates that estimates for smaller α and/or λ may require improved estimation procedures.

By replacing β by $\hat{\beta}$ in equation (2.13) and letting $\boldsymbol{\theta_p}=(\alpha,\,\lambda,\,\gamma)$, the profile log-likelihood function for $\boldsymbol{\theta_p}$ can be expressed as

$$\begin{aligned}
\ell(\boldsymbol{\theta_p}) = \quad & n \quad \log(n)+n\log(\alpha\,\lambda\,\gamma)+(\gamma-1)\sum_{i=1}^{n}\log(t_i)-\sum_{i=1}^{n}(1+\lambda\,t_i^{\gamma})^{\alpha} \\
+ \quad & (\alpha-1)\sum_{i=1}^{n}\log(1+\lambda\,t_i^{\gamma})-\sum_{i=1}^{n}\log\left[1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}\right] \\
- \quad & n\log\left\{-\sum_{i=1}^{n}\log\left[1-\mathrm{e}^{1-\left(1+\lambda\,t_i^{\gamma}\right)^{\alpha}}\right]\right\}.
\end{aligned} \tag{2.14}$$

We assume that the standard regularity conditions for $\ell_p=\ell(\boldsymbol{\theta_p})$ hold: i) The parameter space, say Θ, is open and ℓ_p has a global maximum in Θ; ii) For almost all samples $t_1,\cdots,t_n$, the fourth-order log-likelihood derivatives with respect to the model parameters exist and are continuous in an open subset of Θ that contains the true parameter vector; iii) The expected information matrix is positive definite and finite. These regularity conditions are not restrictive and hold for the models cited in this thesis.

The corresponding score vector of (2.14), $\boldsymbol{U(\theta_p)}$, has the components

$$
\begin{aligned}
U_\alpha(\boldsymbol{\theta_p}) = {} & \frac{n}{\alpha} + \sum_{i=1}^{n} \log\left(1+\lambda\, t_i^\gamma\right) - \sum_{i=1}^{n} \left(1+\lambda\, t_i^\gamma\right)^\alpha \log\left(1+\lambda\, t_i^\gamma\right) \\
& - n \sum_{i=1}^{n} \frac{\left(1+\lambda\, t_i^\gamma\right)^\alpha \log\left(1+\lambda\, t_i^\gamma\right) \mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}}{1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}} \left\{ \sum_{i=1}^{n} \log\left[1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}\right] \right\}^{-1} \\
& - \sum_{i=1}^{n} \frac{\left(1+\lambda\, t_i^\gamma\right)^\alpha \log\left(1+\lambda\, t_i^\gamma\right) \mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}}{1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}}, \\
U_\lambda(\boldsymbol{\theta_p}) = {} & \frac{n}{\lambda} + (\alpha-1) \sum_{i=1}^{n} t_i^\gamma \left(1+\lambda\, t_i^\gamma\right)^{-1} - \alpha \sum_{i=1}^{n} t_i^\gamma \left(1+\lambda\, t_i^\gamma\right)^{\alpha-1} \\
& - n\alpha \sum_{i=1}^{n} \frac{t_i^\gamma \left(1+\lambda\, t_i^\gamma\right)^{\alpha-1} \mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}}{1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}} \left\{ \sum_{i=1}^{n} \log\left[1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}\right] \right\}^{-1} \\
& - \alpha \sum_{i=1}^{n} \frac{t_i^\gamma \left(1+\lambda\, t_i^\gamma\right)^{\alpha-1} \mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}}{1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}},
\end{aligned}
$$

and

$$
\begin{aligned}
U_\gamma(\boldsymbol{\theta_p}) = {} & \frac{n}{\gamma} + \sum_{i=1}^{n} \log(t_i) + \lambda(\alpha-1) \sum_{i=1}^{n} \frac{t_i^\gamma \log(t_i)}{1+\lambda t_i^\gamma} - \alpha\lambda \sum_{i=1}^{n} t_i^\gamma \log(t_i) \left(1+\lambda t_i^\gamma\right)^{\alpha-1} \\
& - n\alpha\lambda \sum_{i=1}^{n} \frac{t_i^\gamma \log(t_i) \left(1+\lambda t_i^\gamma\right)^{\alpha-1} \mathrm{e}^{1-\left(1+\lambda t_i^\gamma\right)^\alpha}}{1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}} \left\{ \sum_{i=1}^{n} \log\left[1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}\right] \right\}^{-1} \\
& - \alpha\lambda \sum_{i=1}^{n} \frac{t_i^\gamma \log(t_i) \left(1+\lambda t_i^\gamma\right)^{\alpha-1} \mathrm{e}^{1-\left(1+\lambda t_i^\gamma\right)^\alpha}}{1-\mathrm{e}^{1-\left(1+\lambda\, t_i^\gamma\right)^\alpha}}
\end{aligned}
$$

Solving the equations in $\boldsymbol{U(\theta_p)} = \boldsymbol{0}$ simultaneously yields the MLEs of α, λ and γ. The MLE of β is just $\hat{\beta}(\hat{\alpha}, \hat{\lambda}, \hat{\gamma})$. The maximization of the profile log-likelihood might be simpler since it involves only three parameters. Lemonte (2013) noted a similar result for the ENH model but mentioned that some of the properties that hold for a genuine likelihood do not hold for its profile version.

For interval estimation of the components of $\boldsymbol{\theta}$, we can adopt the observed information matrix $\boldsymbol{J(\theta)}$ given by

$$
\boldsymbol{J(\theta)} = -\frac{\partial^2 \ell(\theta)}{\partial\theta\, \partial\theta^T} = \begin{pmatrix} J_{\alpha\alpha} & J_{\alpha\beta} & J_{\alpha\lambda} & J_{\alpha\gamma} \\ . & J_{\beta\beta} & J_{\beta\lambda} & J_{\beta\gamma} \\ . & . & J_{\lambda\lambda} & J_{\lambda\gamma} \\ . & . & . & J_{\gamma\gamma} \end{pmatrix},
$$

whose elements can be obteined from the authors upon request.

Under the standard regularity conditions cited before, the multivariate normal $N_4(0, \boldsymbol{J}(\widehat{\boldsymbol{\theta}})^{-1})$ distribution can be used to construct approximate confidence intervals for the model parameters.

2.10 Simulation study

In this section, a Monte Carlo simulation experiment is performed in order to examine the accuracy of the MLEs of the model parameters. The simulations are carried out by generating observations from the EGPW distribution, using the inverse transformation method for different parameter combinations. The number of observations is set at n= 100, 300 and 500 and the number of replications at 10,000. For maximizing the log-likelihood function, we use the `Optim` function with analytical derivatives in `R`. From the results of the simulations given in Table 2.2, we can verify that the root mean squared errors (RMSEs) of the MLEs of α, β, λ and γ decay toward zero as the sample size increases, as expected. As the sample size n increases, the mean estimates of the parameters tend to be closer to the true parameter values.

Table 2.2: Mean estimates and RMSEs of the EGPW distribution for some parameter values.

					Mean estimates				**RMSEs**			
n	α	β	λ	γ	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\gamma}$
100	0.3	4.0	3.0	1.6	0.355	4.041	3.061	2.140	0.313	1.988	2.084	1.376
	1.7	0.8	0.1	0.2	1.670	0.795	0.105	0.253	0.833	0.434	0.109	0.117
	3.0	2.0	5.0	0.6	3.848	2.212	5.512	0.627	2.027	1.104	3.112	0.157
	3.5	0.9	0.2	0.1	2.729	0.781	0.284	0.166	1.674	0.493	0.276	0.128
	7.0	1.5	5.0	0.2	7.296	1.643	5.215	0.199	1.717	0.504	1.644	0.027
	7.5	1.3	4.0	0.5	8.034	1.436	4.786	0.510	2.488	0.589	2.386	0.100
300	0.3	4.0	3.0	1.6	0.316	4.035	2.990	1.843	0.145	1.515	1.534	0.807
	1.7	0.8	0.1	0.2	1.630	0.762	0.094	0.235	0.657	0.271	0.052	0.086
	3.0	2.0	5.0	0.6	3.454	2.096	5.083	0.603	1.277	0.573	2.189	0.086
	3.5	0.9	0.2	0.1	3.155	0.858	0.233	0.117	1.063	0.278	0.111	0.050
	7.0	1.5	5.0	0.2	7.093	1.556	5.121	0.199	1.130	0.285	1.141	0.018
	7.5	1.3	4.0	0.5	7.774	1.337	4.431	0.507	1.753	0.303	1.645	0.066
500	0.3	4.0	3.0	1.6	0.308	4.005	2.990	1.766	0.105	1.296	1.344	0.613
	1.7	0.8	0.1	0.2	1.639	0.766	0.093	0.226	0.572	0.226	0.041	0.069
	3.0	2.0	5.0	0.6	3.322	2.062	5.030	0.600	0.991	0.429	1.902	0.068
	3.5	0.9	0.2	0.1	3.301	0.883	0.219	0.107	0.789	0.204	0.073	0.025
	7.0	1.5	5.0	0.2	7.090	1.530	5.090	0.200	0.938	0.218	0.964	0.015
	7.5	1.3	4.0	0.5	7.724	1.323	4.272	0.504	1.462	0.231	1.327	0.053

2.11 Application

In this section, we present an application to illustrate the flexibility of the EGPW distribution, that indicates the potentiality of the new distribution for modeling positive data. The data set has size 101 and represents the stress-rupture life of kevlar 49/epoxy strands which are subjected to constant sustained pressure at the 90% stress level until all had failed such that we obtain complete data with exact failure times. This data set was studied by Andrews and Herzberg (1985). Table 2.3 gives a descriptive summary of the samples. Note that the data set present positive skewness that is a feature that can be well modeled by our proposed distribution.

Table 2.3: Descriptive statistics for the stress-rupture data.

Statistics	**Real data sets**
	Stress-rupture data
Mean	1.0248
Median	0.8000
Mode	0.5000
Variance	1.2529
Skewness	3.0017
Kurtosis	13.7089
Maximum	7.8900
Minimum	0.0100
n	101

We fit the EGPW distribution (2.6) to this data set and also present a comparative study with the fits of some nested and non-nested models. One of these models is the Kumaraswamy Weibull (Kw-W) distribution, whose pdf is given by

$$g(t) = \frac{a\,b\,c\,\beta^c\,t^{c-1}\exp\{-(\beta t)^c\}\,[1-\exp\{-(\beta t)^c\}]^{a-1}}{\{1-[1-\exp\{-(\beta t)^c\}]^a\}^{1-b}}, \quad t>0,$$

where $a>0$, $b>0$, $c>0$ and $\beta>0$. Another model is the EW distribution, whose pdf is given by

$$g(t) = \alpha\,\beta\,\lambda\,t^{\alpha-1}\exp(-\lambda\,t^{\alpha})\,[1-\exp(-\lambda\,t^{\alpha})]^{\beta-1}, \quad t>0,$$

where $\alpha>0$ and $\beta>0$ are shape parameters and $\lambda>0$ is a scale parameter. This distribution is quite flexible because its hrf presents the classic five forms (constant, decreasing, increasing, upside-down bathtub and bathtub-shaped). The Weibull model arises from the EW model when $\beta = 1$.

The ENH distribution can also have the same shapes for the hrf and therefore can be an interesting alternative to the EW distribution in modeling positive data.

Xie *et al.* (2002) proposed a modified Weibul density (MW) given by

$$g(t) = \lambda\beta\left(\frac{t}{\alpha}\right)^{1-\beta}\exp\left\{\left(\frac{t}{\alpha}\right)^{\beta} + \lambda\,\alpha\left(1-\exp\left\{\frac{t}{\alpha}\right\}^{\beta}\right)\right\},\quad t>0,$$

where $\lambda > 0$, $\beta > 0$ and $\alpha > 0$. For $\alpha = 1$, it becomes the Chen distribution (Chen, 2000). The MW and Chen distributions can have increasing or bathtub-shaped failure rate. An extension of the Weibull model proposed by Bebbington *et al.* (2007) has pdf in the form

$$g(t) = \left(\alpha + \frac{\beta}{t^2}\right)\exp\left(\alpha t - \frac{\beta}{t}\right)\exp\left\{-\exp\left(\alpha t - \frac{\beta}{t}\right)\right\},\ t>0,$$

where $\alpha > 0$ and $\beta > 0$. We shall use the same terminology by Lemonte (2013) for this distribution, i.e., denote the flexible Weibull (FW) density. The FW model can have increasing or modified bathtub-shaped failure rate.

We use the simulated-annealing (SANN), BFGS and Nelder-Mead methods for maximizing the log-likelihood function of the models in the application. The MLEs and goodness-of-fit statistics are evaluated using the `AdequacyModel` script in `R` software. Tables 2.4 list the MLEs and the corresponding standard errors in parentheses of the unknown parameters for the fitted models to stress-rupture failure times. In applications there is qualitative information about the failure rate shape, which can help for selecting some models. Thus, a device called the total time on test (TTT) plot (Aarset, 1987) is useful. The TTT plot is obtained by plotting

$$T\left(\frac{r}{n}\right) = \left[\sum_{i=1}^{r} y_{i:n} + (n-r)\,y_{r:n}\right] \Big/ \sum_{i=1}^{n} y_{i:n},$$

against r/n, where $r = 1,\ldots,n$ and $y_{i:n}\,(i = 1,\ldots,n)$ are the order statistics of the sample.

From the figures in Table 2.4, we note that with the parameters estimated using the EGPW model for the data set, it is evident decreasing-increasing-decreasing shape in hrf. This fact is in agreement with the TTT plot based on this data set. Figure 2.4 provides plots of the TTT plot and hrf for the fitted models for the stress-rupture times data sets.

Table 2.4: The MLEs of the model parameters for the stress-rupture data and the corresponding standard errors in parentheses.

Distributions	**Estimates**			
EGPW$(\alpha, \beta, \lambda, \gamma)$	0.1349	0.1022	0.0415	6.6681
	(0.0171)	(0.0104)	(0.0154)	(0.0136)
Kw-W(a, b, c, β)	0.7029	0.2175	1.0118	4.3625
	(0.1620)	(0.1038)	(0.0027)	(2.1072)
GPW$(\alpha, \lambda, \gamma)$	1.2659	0.7182	0.8696	
	(0.4483)	(0.3485)	(0.1039)	
EW(β, λ, γ)	0.8488	1.0419	0.8171	
	(0.2981)	(0.2511)	(0.3157)	
MW$(\alpha, \beta, \lambda\ )$	0.0027	0.2259	7.0190	
	(0.0008)	(0.0076)	(1.5244)	
ENH(α, β, λ)	1.0732	0.7762	0.8426	
	(0.2760)	(0.3582)	(0.1238)	
NH(α, λ)	0.8898	1.1810		
	(0.1853)	(0.4270)		
Chen(β, λ)	0.5410	0.5303		
	(0.0585)	(0.0321)		
Weibull(α, λ)	0.9919	0.9259		
	(0.1121)	(0.0726)		
FW(α, β)	0.3287	0.0838		
	(0.0246)	(0.0133)		

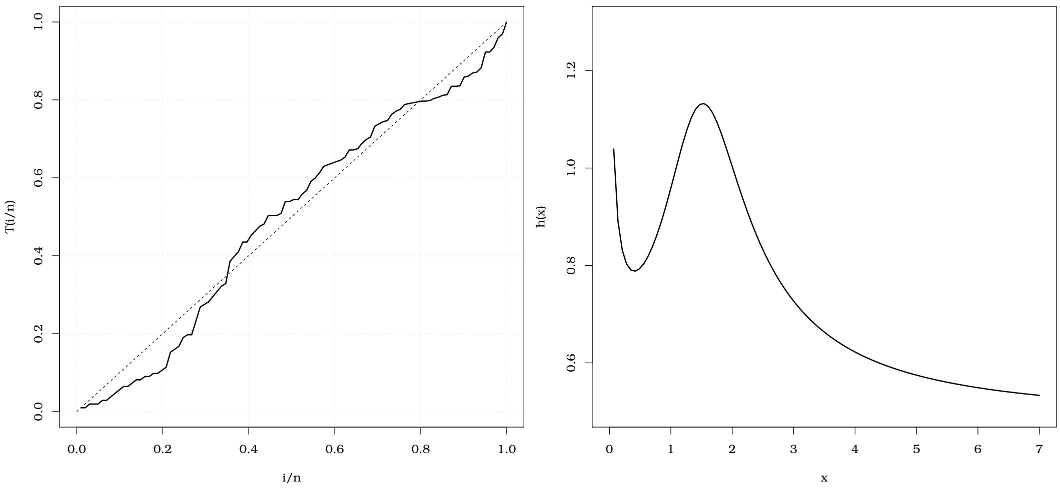
Chen and Balakrishnan (1995) constructed the Cramér-von Mises and Anderson-Darling corrected statistics. We adopt these statistics, where we have a random sample $x_1, \ldots, x_n$ with empirical distribution function $F_n(x)$, and require to test if the sample comes from a special distribution. The Cramér-von Mises (W^*) and Anderson-Darling (A^*) statistics are given by

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

and

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

respectively, where $F_n(x)$ is the empirical distribution function, $F(x;\hat{\theta}_n)$ is the postulated distribution function evaluated at the MLE $\hat{\theta}_n$ of θ. Note that the statistics W^* and A^* are given by



(a) (b)

Figure 2.4: The TTT plot and EGPW hrf for the stress-rupture failure times data.

the differences of $F_n(x)$ and $F(x;\hat{\theta}_n)$. Thus, the lower are these statistics, we have more evidence that $F(x;\hat{\theta}_n)$ generates the sample. The details to evaluate the statistics W^* and A^* are given by Chen and Balakrishnan (1995).

The Kolmogorov-Smirnov (KS), A^* and W^* statistics for these models are given in Table 2.5 for the data sets. We emphasize that the EGPW model fits the stress-rupture failure data better than the other models according to all these statistics. These goodness-of-fit statistics indicate that the EGPW distribution provides a good fit in the application.

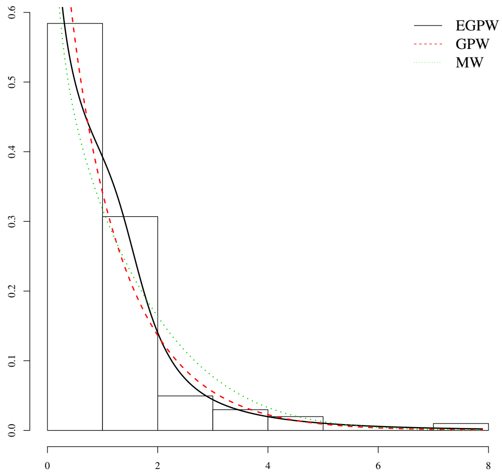
More information is provided by the histogram of the data with the fitted EGPW density function for the data set, which is displayed in Figure 2.5. Clearly the new distribution provides a closer fit to the histogram. The estimated and empirical cumulative function of the most competitive models are shown in Figure 2.5. From this plot, note that the EGPW model fits adequately and hence can be used to model these data. Finally, we can conclude that in the application presented the EGPW distribution is quite competitive to other well-known and widely used distributions, such as the Kw-W, EW and Weibull models.

Table 2.5: Goodness-of-fit statistics for the models fitted to the stress-rupture failure times data.

	Statistics		
Distributions	**W****	**A***	**KS**
EGPW	**0.0722**	**0.4672**	**0.0699**
Kw-W	0.1400	0.8478	0.1017
GPW	0.1730	0.9930	0.0833
EW	0.1686	0.9736	0.0875
MW	0.0980	0.7596	0.1292
ENH	0.1670	0.9667	0.0837
NH	0.2053	1.1434	0.0819
Chen	0.1207	0.8756	0.0973
Weibull	0.1987	1.1115	0.0900
FW	1.1130	5.9971	0.3054

2.12 Concluding remarks

In this Chapter, we introduce the exponentiated generalized power Weibull (EGPW) model to generalize the Weibull distribution. It has an exponentiated parameter and its hazard rate function allows constant, decreasing, increasing, upside-down bathtub or bathtub-shaped shapes. The new distribution contains several well-known lifetime distributions as special models. It can also be derived from a power transform on an exponentiated Nadarajah-Haghighi random variable. Since several structural properties on the generalized power Weibull (GPW) distribution have not been studied, they can be obtained from those on the EGPW distribution. It can also be useful to obtain the properties of other generated families under the GPW baseline. We give a physical motivation for introducing the EGPW distribution if β is an integer. We obtain some structural properties of the EGPW distribution, perform the estimation of parameters by maximum likelihood and prove empirically the flexibility of the new model in an application to real data. The new distribution yields a good adjustment in this application. We note that the EGPW distribution is quite competitive with other lifetime models and can be used effectively to provide better fits than the other usual lifetime distributions.



(a)

Figure 2.5: Histogram and estimated densities of the EGPW, GPW and MW model for the stress-rupture failure times data.

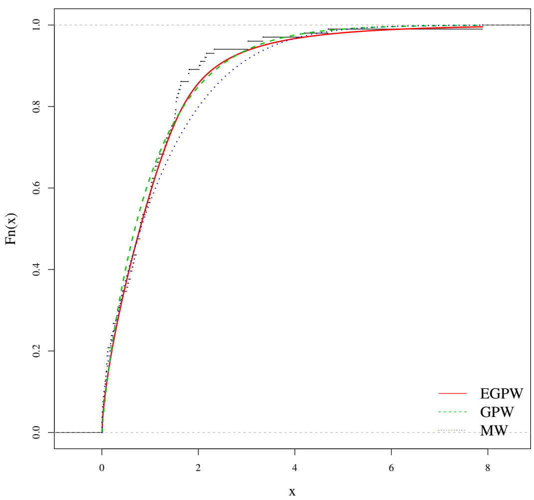


Figure 2.6: Estimated and empirical cdfs for EGPW, GPW and MW models for the stress-rupture failure times data.

Chapter 3

The Nadarajah-Haghighi Lindley distribution

Resumo

Neste capítulo, uma nova distribuição contínua é proposta baseada na composição entre as distribuições Lindley e Nadarajah-Haghighi, a qual pode ser utilizada em aplicações em engenharia, bem como em outras áreas do conhecimento. A distribuição proposta é competitiva com os demais modelos utilizados em análise de sobrevivência, tais como as distribuições Weibull, Weibull exponencializada, Nadarajah-Haghighi exponencializada e outras. Algumas das suas propriedades matemáticas são estudadas, incluindo a função de vida média residual, momentos, função geradora de momentos, desvios médios e curvas de Lorenz e Bonferroni. É também discutida a estimação dos parâmetros do modelo via máxima verossimilhança. Um estudo de simulação é realizado e apresentamos uma aplicação a dados reais para ilustrar empiricamente a utilidade da nova distribuição proposta, na cual se obteve um bom ajuste para o conjunto de dados utilizado, sendo uma alternativa útil aos modelos clássicos de análise de sobrevivência.

Palavras-chave: Distribuição exponencial. Dados de tempo de vida distribuição Lindley. Distribuição Nadarajah-Haghighi. Método de composição.

Abstract

In this Chapter, we propose a new continuous distribution based on compounding the Lindley and Nadarajah Haghighi distributions, which may be useful in engineering applications and other areas. The introduced distribution is a very competitive model to other lifetime models, such as the Weibull, exponentiated Weibull and exponentiated Nadarajah-Haghighi distributions, among others. Some of its properties are investigated including mean residual life, moments, generating function, mean deviations, Bonferroni and Lorenz curves. We discuss the estimation of the model parameters by maximum likelihood. A simulation study is performed and we provide two applications to real data for illustrative purposes. We show that the proposed distribution yields a good adjustment for both data sets, and it can be a useful alternative for other classical lifetime models.

Keywords: Compounding approach. Exponential distribution. Lifetime data. Lindley distribution. Nadarajah-Haghighi distribution.

3.1 Introduction

The Lindley distribution was introduced by Lindley (1958) in the context of fiducial and Bayesian inference. This distribution is a mixture of exponential and length-biased exponential distributions. Let Y be a Lindley random variable with parameter $\gamma > 0$ having pdf given by

$$g(y) = \frac{\gamma^2}{1+\gamma}(1+y)\ \mathrm{e}^{-\gamma y}, \qquad y > 0,$$

where the mixing proportion is $\gamma/(1+\gamma)$. The survival function of Y is

$$\overline{G}(y) = \frac{1+\gamma+\gamma y}{1+\gamma}\ \mathrm{e}^{-\gamma y}.$$

Various statistical properties of this distribution are discussed in details by Ghitany *et al.* (2008b). The authors also showed that the Lindley distribution is quite competitive with the exponential distribution. Gupta and Singh (2013) studied the parameter estimation of the Lindley distribution with hybrid censored data. Krishna and Kumar (2011) considered the estimation of the model parameters for progressively type II right censored sample and Mazucheli and Achcar (2011) applied this distribution to competing risks in lifetime data.

In distribution theory context, some generalizations are obtained based on transformations of the Lindley distribution. We refer the reader to Nadarajah and Bakouch (2011) for the generalized (or exponentiated) Lindley, Bakouch *et al.* (2012) for the extended Lindley, Ghitany *et al.* (2011) and Al-Mutairi *et al.* (2015) for the weighted Lindley, Ghitany *et al.* (2013) for the power Lindley and Ashour and Eltehiwy (2015) for the exponentiated power Lindley distributions.

Another technique that has been considered is the discrete-continuous compounding approach. It is defined as the minimum of N independent and identical continuous random variables, where N is a discrete random variable. Adamidis and Loukas (1998) pionnered this method and introduced the exponential geometric distribution. We also find in the literature some models obtained from compositions of Lindley and other discrete distributions. Sankaran (1970) introduced the discrete Poisson-Lindley by combining the Poisson and Lindley distributions. Zamani and Ismail (2010) presented the negative binomial Lindley distribution. The zero-truncated Poisson-Lindley and Pareto Poisson-Lindley distributions were introduced by Ghitany *et al.* (2008a) and Asgharzadeh *et al.* (2013), respectively.

The Nadarajah-Haghighi (NH) distribution was pioneered by Nadarajah and Haghighi (2011) as a generalization of the exponential distribution. Let Z denote a NH random variable with parameters $\alpha > 0$ and $\lambda > 0$. Its pdf and survival function are

$$q(z) = \alpha\lambda(1+\lambda z)^{\alpha-1}\mathrm{e}^{1-(1+\lambda z)^{\alpha}}$$

and

$$\overline{Q}(z) = \mathrm{e}^{1-(1+\lambda z)^{\alpha}},$$

respectively. Several generalizations of the NH distribution have been proposed in recent years, such as the exponentiated Nadarajah-Haghighi (Lemonte, 2013), the gamma Nadarajah-Haghighi (Bourguignon *et al.*, 2015), beta Nadarajah-Haghighi (Dias, 2016) distributions and the Nadarajah-Haghighi family of distributions (Dias, 2016), among others. By using the discrete-continuous compounding approach, we have the Poisson Gamma Nadarajah-Haghighi (Ortega *et al.*, 2015) and geometric Nadarajah-Haghighi (Marinho, 2016) distributions, for example.

A comprehensive review on compounding method for generating distributions can be found in Tahir *et al.* (2016a). They pointed a different compounding approach by taking the minimum between two continuous distributions. Note that in the discrete-continuous compositions we have that N is a discrete random variable that represents the number of identical elements that follows some continuous distribution. For the continuous-continuous compositions, we suppress the condition to be identically distributed and fix $N = 2$. Some well-known continuous-continuous compounded models are the additive Weibull (Xie and Lai, 1995; Lemonte *et al.*, 2014), exponential-Weibull (Cordeiro *et al.*, 2014a) and generalized exponential-exponential (Popovíc *et al.*, 2015) distributions, among others.

In this Chapter, we introduce a new continuous-continuous compounded model referred to as the *Nadarajah-Haghighi Lindley* (NHL) distribution. This three-parameter distribution is obtained by compounding the Lindley and NH distributions. We assume that Y and Z are independent random variables and define $X = \min(Y, Z)$ as a NHL random variable, whose survival function is given by

$$\overline{F}(x) = \overline{G}(x)\overline{Q}(x).$$

The cdf of X is

$$F(x) = 1 - \frac{1+\gamma+\gamma\,x}{1+\gamma}\,\mathrm{e}^{1-\gamma\,x-(1+\lambda\,x)^{\alpha}}, \qquad x > 0, \tag{3.1}$$

where $\alpha > 0$, $\lambda \geq 0$ and $\gamma \geq 0$. The pdf and hrf of X are given by

$$f(x) = \frac{(1+\gamma+\gamma\,x)\left[\gamma+\alpha\lambda(1+\lambda\,x)^{\alpha-1}\right]-\gamma}{\gamma+1}\,\mathrm{e}^{1-\gamma\,x-(1+\lambda\,x)^{\alpha}} \tag{3.2}$$

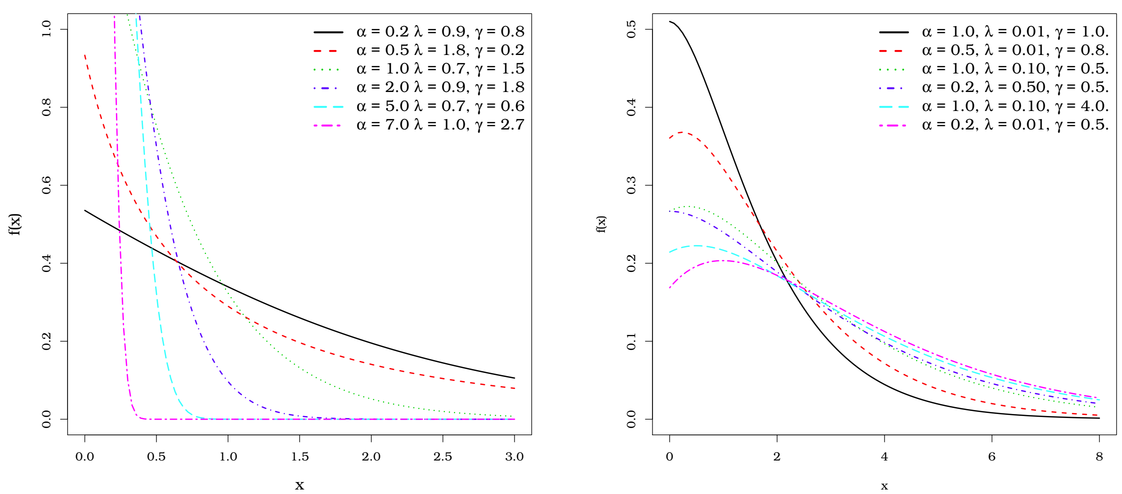
and

$$h(x) = \frac{(1+\gamma+\gamma\,x)\left[\gamma+\alpha\lambda(1+\lambda\,x)^{\alpha-1}\right]-\gamma}{\gamma+\gamma\,x+1},$$

respectively. Henceforth, we denote $X \sim \mathrm{NHL}(\alpha, \lambda, \gamma)$. The proposed distribution contains as special models well-known distributions. For $\gamma = 0$, the NHL reduces to the NH distribution. If

$\gamma = 0$ and $\alpha = 1$, we have the exponential distribution. For $\alpha = 0$ or $\lambda = 0$, we have the Lindley distribution.

Figure 3.1 provides plots of the pdf of X for some parameter values. The new distribution presents decreasing and reverse J shaped curve. Figure 3.2 reveals that the NHL distribution can have decreasing, increasing, upside-down bathtub and bathtub-shaped hazard functions. This feature makes the new distribution very competitive with the Weibull, gamma and exponential distributions that exhibit only monotonic hazard rates. According to Nadarajah and Bakouch (2011) this is a major weakness because most empirical life systems have bathtub shapes for their hrfs.



(a) (b)

Figure 3.1: Pdf plots for the NHL distribution.

This approach may be useful in engineering. For example, let Y and Z denote the lifetimes of two independent components of a system. Then, the lifetime of the system will be a NHL random variable. Cordeiro *et al.* (2014a) studied a similar situation for the exponential-Weibull lifetime distribution and presented some motivations that may be adapted for the current distribution, such as

- *Time to the first failure.* Consider a system with two sub-systems functioning in series

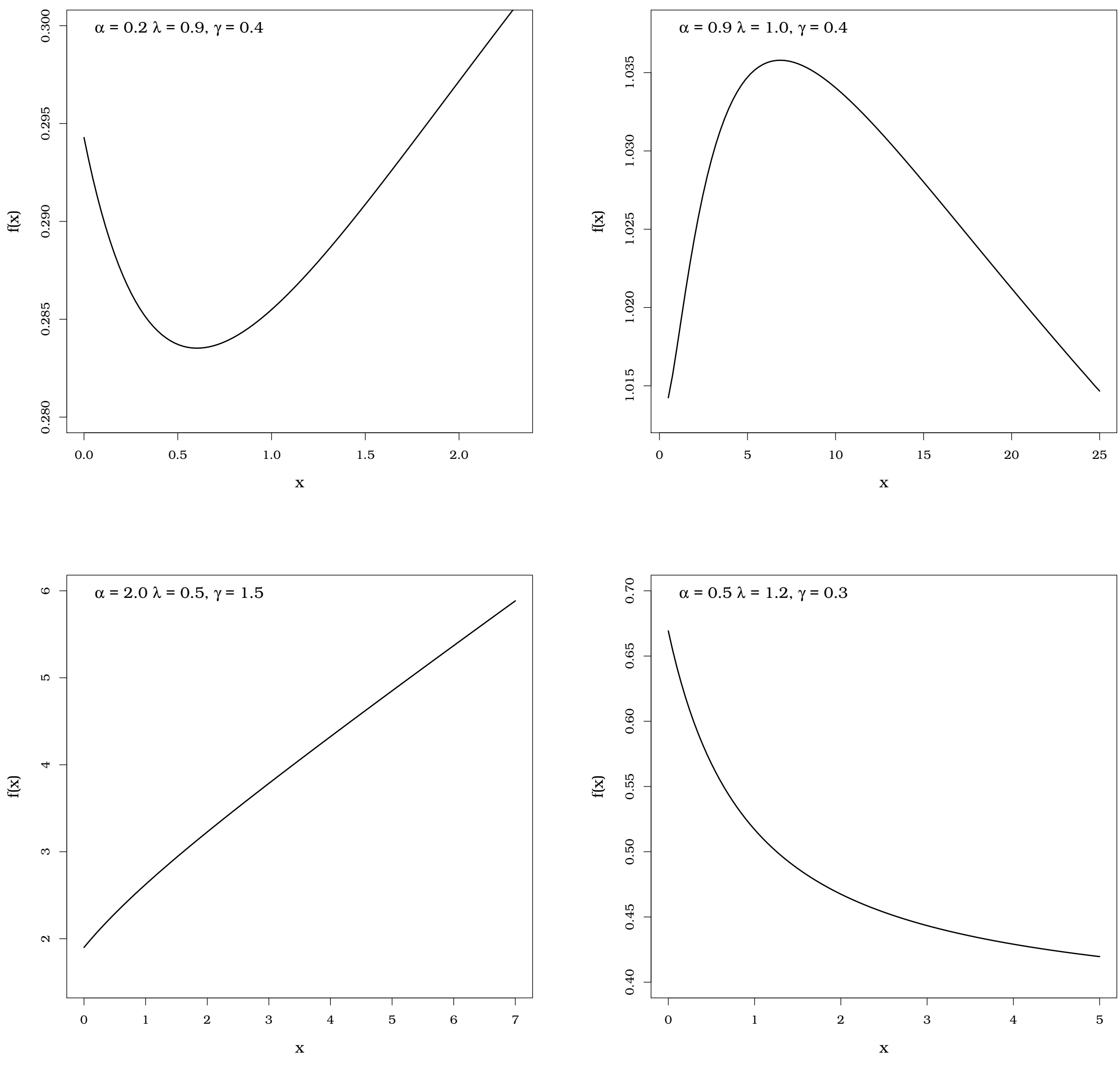


Figure 3.2: Hrf plots for the NHL distribution

independently at a given time. Suppose that the system fails if each or both sub-systems fail. Let Y and Z be their failure times. If Y and Z follow the Lindley and NH distributions, respectively. Thus, the NHL distribution can characterize the system lifetime.

- *Reliability.* From the stochastic representation $X = \min\{Y, Z\}$, we note that the NHL model can arise in series systems with two different components. This situation may appear in engineering applications and biological organisms.

Furthermore, Nadarajah and Haghighi (2011) mentioned some advantages of NH model, such as the ability to model data with mode fixed at zero and the fact that it can be interpreted as a truncated Weibull distribution. The NHL distribution also accumulates this advantages once it has the NH distribution as special model.

The rest of this chapter is outlined as follows. In Sections 3.2-3.4, we derive a range of mathematical properties of the NHL distribution. In Section 3.5, we adopt the maximum likelihood method to estimate the model parameters. We perform a simulation study in Section 3.6. A real data application is provided in Section 3.7. Some concluding remarks are offered in Section 3.8.

3.2 Mean residual life

The mean residual life is a relevant characteristic to the design of safe systems in a wide variety of applications in engineering and reliability. Given that a component survives up to time $x > 0$, the residual life is defined by

$$m(x) = \mathbb{E}(X - x | X > x) = \frac{1}{1 - F(x)} \int_x^\infty [1 - F(t)]\, \mathrm{d}t.$$

It represents the period beyond x until the time of failure. Note that

$$m(x) = \frac{\mathrm{e}^{x\gamma + (1+\lambda x)^\alpha}}{(1 + \gamma + \gamma\, x)} \int_x^\infty (1 + \gamma + \gamma\, t)\, \mathrm{e}^{-\gamma\, t - (1+\lambda\, t)^\alpha}\, \mathrm{d}t.$$

We consider the integral

$$J = \int_x^\infty (1 + \gamma + \gamma\, t)\, \mathrm{e}^{-\gamma\, t - (1+\lambda\, t)^\alpha}\, \mathrm{d}t.$$

Setting $u = (1+\lambda\, t)^{\alpha}$, we have $t = (u^{1/\alpha} - 1)/\lambda$. So, the above integral reduces to

$$J = \frac{1}{\alpha\,\lambda} \int\limits_{(1+\lambda\,x)^{\alpha}}^{\infty} u^{(1-\alpha)/\alpha}[1+\gamma+\gamma\,\lambda^{-1}(u^{1/\alpha}-1)]\,\mathrm{e}^{-u-\gamma\,\lambda^{-1}(u^{1/\alpha}-1)}\,\mathrm{d}u.$$

By expanding

$$\mathrm{e}^{-\gamma\,\lambda^{-1}(u^{1/\alpha}-1)} = \sum_{i=0}^{\infty} \frac{(-1)^i \gamma^i (u^{1/\alpha}-1)^i}{\lambda^i i!},$$

and using the binomial expansion for $(u^{1/\alpha}-1)^i$, we can write

$$\begin{aligned} m(x) &= \frac{\mathrm{e}^{\gamma\,x+(1+\lambda\,x)^{\alpha}}}{\alpha\,\lambda(1+\gamma+\gamma\,x)} \sum_{i=0}^{\infty}\sum_{j=0}^{i} \frac{(-1)^{2i-j}\gamma^i}{\lambda^i\, i!}\binom{i}{j} \left[\frac{\gamma}{\lambda}\,\Gamma\left(\frac{j+2}{\alpha}\,,\,(1+\lambda\,x)^{\alpha}\right)\right. \\ &+ \left.\left(1+\gamma-\frac{\gamma}{\lambda}\right)\,\Gamma\left(\frac{j+1}{\alpha}\,,\,(1+\lambda\,x)^{\alpha}\right)\right], \end{aligned}$$

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

3.3 Generating function and moments

We denote by $M_X(t)$ the moment generating function (mgf) of X. From Equation (3.2), we obtain

$$M_X(t) = \frac{\mathrm{e}}{1+\gamma}\int_0^{\infty} \left\{(\gamma+\gamma x+1)[\gamma+\alpha\lambda(1+\lambda x)^{\alpha-1}]-\gamma\right\}\,\mathrm{e}^{x(t-\gamma)-(1+\lambda x)^{\alpha}}\mathrm{d}x.$$

Setting $u = (1+\lambda x)^{\alpha}$, we have

$$\begin{aligned} M_X(t) = \frac{\mathrm{e}}{\alpha\lambda(1+\gamma)}\int_1^{\infty} u^{\frac{1-\alpha}{\alpha}}&\left\{[\gamma+\gamma\lambda^{-1}(u^{1/\alpha}-1)+1](\gamma+\alpha\lambda u^{\frac{\alpha-1}{\alpha}})-\gamma\right\} \\ \times \exp&\left\{\frac{(u^{1/\alpha}-1)(t-\gamma)}{\lambda}-u\right\}\mathrm{d}u. \end{aligned}$$

By expanding

$$\exp\left\{\frac{(u^{1/\alpha}-1)(t-\gamma)}{\lambda}\right\} = \sum_{i=0}^{\infty}\frac{(u^{1/\alpha}-1)^i(t-\gamma)^i}{\lambda^i i!},$$

using the binomial expansion for $(u^{1/\alpha}-1)^i$ and, after some algebra, we can write

$$\begin{aligned} M_X(t) = \frac{\mathrm{e}}{1+\gamma}&\sum_{i=0}^{\infty}\sum_{j=0}^{i}\frac{(-1)^{i-j}(t-\gamma)^i}{\lambda^i\, i!}\binom{i}{j}\left[\frac{\gamma}{\lambda}\,\Gamma\left(\frac{j+1}{\alpha}+1\,,\,1\right)\right. \\ &+\left(1+\gamma-\frac{\gamma}{\lambda}\right)\,\Gamma\left(\frac{j+\alpha}{\alpha}\,,\,1\right)+\frac{\gamma^2}{\alpha\lambda^2}\,\Gamma\left(\frac{j+2}{\alpha}\,,\,1\right) \\ &+\left.\frac{\gamma^2(\lambda-1)}{\alpha\lambda^2}\,\Gamma\left(\frac{j+1}{\alpha}\,,\,1\right)\right]. \end{aligned}$$

The expression above is useful for computing moments and cumulants of a random variable, among other interesting characteristics of a distribution (e.g., tendency, dispersion, skewness and kurtosis).

The sth ordinary moment of X is given by

$$\mu'_s = \frac{\mathrm{e}}{1+\gamma}\int_0^\infty x^s \left\{(\gamma+\gamma x+1)[\gamma+\alpha\lambda(1+\lambda x)^{\alpha-1}]-\gamma\right\} \mathrm{e}^{-\lambda x-(1+\lambda x)^\alpha}\mathrm{d}x.$$

Again, setting $u = (1+\lambda x)^\alpha$ in this integral and expanding $\exp(-\lambda x)$, we can write (for $s \geq 1$)

$$\begin{aligned}
\mu'_s &= \frac{\mathrm{e}}{\alpha\lambda^{s+2}(\gamma+1)}\sum_{i=0}^{s}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{(-1)^{2j-i-k+s}\gamma^j}{j!\lambda^j}\binom{s}{i}\binom{j}{k}\left[\gamma^2\Gamma\left(\frac{i+2}{\alpha}\,,\,1\right)\right.\\
&+ \alpha\lambda\gamma\Gamma\left(\frac{i+1}{\alpha}+1\,,\,1\right)\gamma^2(\lambda-1)\Gamma\left(\frac{i+1}{\alpha}\,,\,1\right)\\
&+ \left.\alpha\lambda(\lambda-\gamma)\Gamma\left(\frac{i}{\alpha}+1\,,\,1\right)\right].
\end{aligned}$$

The central moments (μ_s) and cumulants (κ_s) of X can be determined from these raw moments using well-known relationships. We have

$$\mu_s = \sum_{k=0}^{s}(-1)^k\binom{s}{k}\mu_1'^k\mu'_{s-k} \qquad \text{and} \qquad \kappa_s = \sum_{k=0}^{s-1}\binom{s-1}{k-1}\kappa_k\,\mu'_{s-k},$$

respectively, where $\kappa_1 = \mu'_1$. Thus, $\kappa_2 = \mu'_2 - \mu_1'^2, \kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^3$, etc.

The skewness $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ can be determined from the third standardized cumulant. Further, an alternative expression for the skewness was introduced by MacGillivray (1986). The MacGillivray's skewness function of X is expressed as

$$\rho(u) = \rho(u;\alpha,\beta,\gamma) = \frac{Q(1-u)+Q(u)-2Q(1/2)}{Q(1-u)-Q(u)},$$

where $u \in (0,\,1)$ and $Q(\cdot)$ is the qf of X, which can only be determined numerically. Plots of the MacGillivray skewness for some parameter values are displayed in Figure 3.3. Note that $\rho(u)$ presents more variability for small values of γ. Once the qf of the NHL distribution can not be solved analytically, we can use numerical techniques for inverting the cdf in (3.1). We use the `inverse` function from the `GoFKernel` script in the `R` software. It may also be helpful in random number generation by the inversion method.

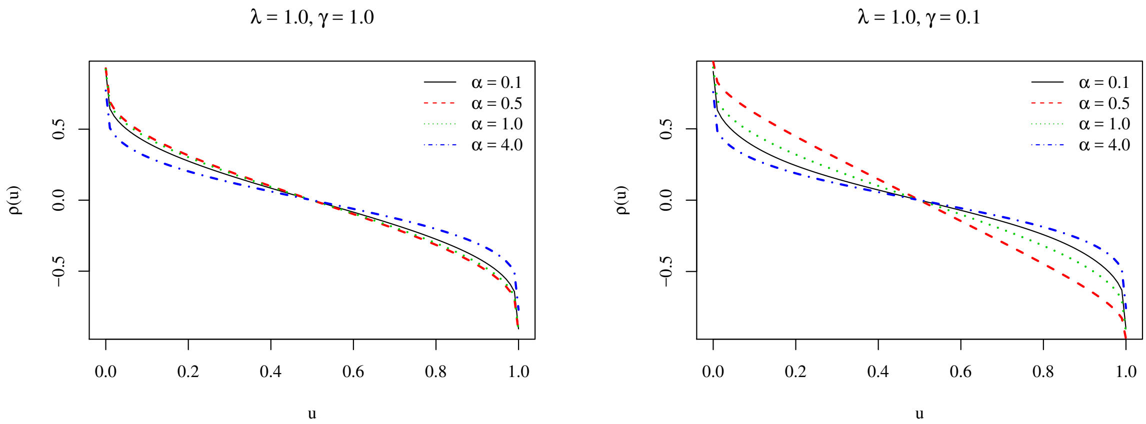


Figure 3.3: Skewness of the NHL model for some parameter values.

3.4 Incomplete moments

The sth incomplete moment of X is defined by $m_s(y) = \int_0^y x^n f(x)dx$. Thus, by inserting (3.2) in $m_s(y)$, we have

$$m_s(y) = \frac{\mathrm{e}}{1+\gamma}\int_0^y x^s \left\{(\gamma+\gamma x+1)[\gamma+\alpha\lambda(1+\lambda x)^{\alpha-1}]-\gamma\right\} \mathrm{e}^{-\lambda x-(1+\lambda x)^{\alpha}}\mathrm{d}x.$$

Using the exponential and binomial expansions, we obtain

$$\begin{aligned} m_s(y) = {} & \frac{\mathrm{e}}{\alpha\lambda^{s+2}(\gamma+1)}\sum_{i=0}^{s}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{(-1)^{2j-i-k+s}\gamma^j}{j!\lambda^j} \\ & \times \binom{s}{i}\binom{j}{k}\left\{\left[\gamma^2\Gamma\left(\frac{i+2}{\alpha}\,,1\right)+\alpha\lambda\gamma\Gamma\left(\frac{i+1}{\alpha}+1\,,1\right)\right.\right. \\ & \left. + \gamma^2(\lambda-1)\Gamma\left(\frac{i+1}{\alpha}\,,1\right)+\alpha\lambda(\lambda-\gamma)\Gamma\left(\frac{i}{\alpha}+1\,,1\right)\right] \\ & - \left[\gamma^2\Gamma\left(\frac{i+2}{\alpha}\,,(\lambda y+1)^{\alpha}\right)+\alpha\lambda\gamma\Gamma\left(\frac{i+1}{\alpha}+1\,,(\lambda y+1)^{\alpha}\right)\right. \\ & + \gamma^2(\lambda-1)\Gamma\left(\frac{i+1}{\alpha}\,,(\lambda y+1)^{\alpha}\right) \\ & \left.\left. + \alpha\lambda(\lambda-\gamma)\Gamma\left(\frac{i}{\alpha}+1\,,(\lambda y+1)^{\alpha}\right)\right]\right\}. \end{aligned} \tag{3.3}$$

In various practical situations, the shape of many distributions can be usefully described by the incomplete moments. For example, the amount of scatter in a population is measured to some extent by the totality of deviations from the mean and median, which are applications of the first incomplete moment. If X has the NHL distribution with pdf (3.2) we can derive the mean deviations about the mean ($\delta_1 = \mathbb{E}(|X-\mu'_1|)$) and about the median ($\delta_2 = \mathbb{E}(|T-M|)$) of X from the relations

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2\,m_1(\mu'_1) \qquad \text{and} \qquad \delta_2 = \mu'_1 - 2\,m_1(M),$$

respectively, where $\mu'_1 = \mathbb{E}(X)$, $M = \text{Median}(X)$ is the median, $F(\mu'_1)$ is easily determined from (3.1) and $m_1(y) = \int_0^y x\,f(x)\mathrm{d}x$ is the first incomplete moment given by (3.3) with $s = 1$.

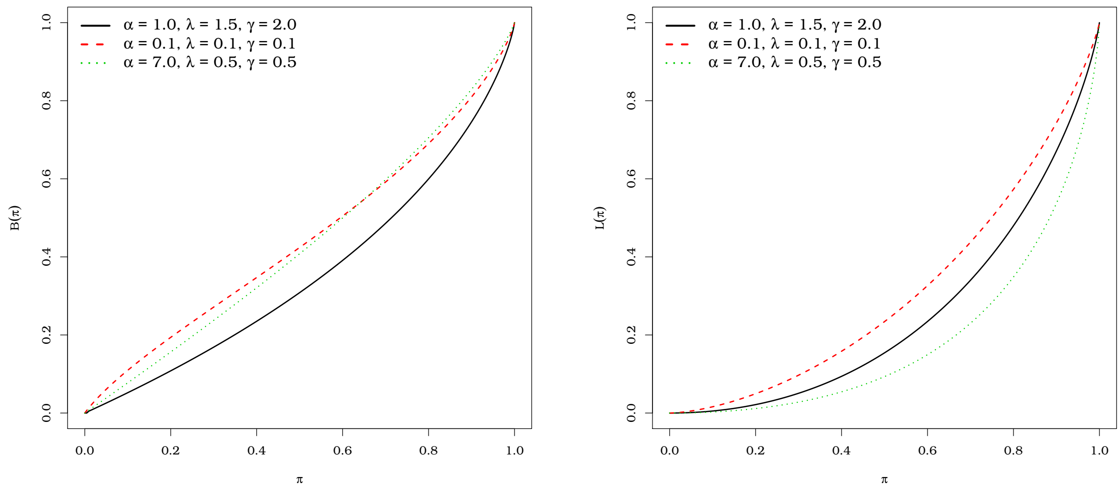
Another applications of the first incomplete moment refers to the Lorenz and Bonferroni curves defined, for a given probability π, by

$$B(\pi) = \frac{m_1(q)}{\pi\mu'_1} \qquad \text{and} \qquad L(\pi) = \frac{m_1(q)}{\mu'_1},$$

respectively, where $q = F^{-1}(\pi)$ is the inverse function of (3.1) evaluated at π. Hence,

$$\begin{aligned}
B(\pi) &= \frac{\mathrm{e}}{\alpha\lambda^{q+2}(\gamma+1)\,\pi\,\mu'_1}\sum_{i=0}^{1}\sum_{j=0}^{\infty}\sum_{k=0}^{j}\frac{(-1)^{2j-i-k+q}\gamma^j}{j!\lambda^j}\\
&\times \binom{q}{i}\binom{j}{k}\left\{\left[\gamma^2\Gamma\left(\frac{i+2}{\alpha}\,,\,1\right)+\alpha\lambda\gamma\Gamma\left(\frac{i+1}{\alpha}+1\,,\,1\right)\right.\right.\\
&+ \left.\gamma^2(\lambda-1)\Gamma\left(\frac{i+1}{\alpha}\,,\,1\right)+\alpha\lambda(\lambda-\gamma)\Gamma\left(\frac{i}{\alpha}+1\,,\,1\right)\right]\\
&- \left[\gamma^2\Gamma\left(\frac{i+2}{\alpha}\,,\,(\lambda y+1)^\alpha\right)+\alpha\lambda\gamma\Gamma\left(\frac{i+1}{\alpha}+1\,,\,(\lambda y+1)^\alpha\right)\right.\\
&+ \gamma^2(\lambda-1)\Gamma\left(\frac{i+1}{\alpha}\,,\,(\lambda y+1)^\alpha\right)\\
&+ \left.\left.\alpha\lambda(\lambda-\gamma)\Gamma\left(\frac{i}{\alpha}+1\,,\,(\lambda y+1)^\alpha\right)\right]\right\}. \qquad (3.4)
\end{aligned}$$

The Lorenz curve follows by multiplying (3.4) by π. For each $\pi = F(x)$, $L(\pi)$ is the proportion of the total volume of income that accumulates the set of units with income less than or equal to x. It is clear that $L(\pi) \le \pi$ for $0 \le \pi \le 1$, where $L(\pi) = \pi$ in case of equidistribution and $L(\pi) = 0$, $0 \le p < 1$, $L(1) = 1$, if the concentration is maximum. It is possible to show that $B(\pi) \le 1$, $0 \le \pi \le 1$. For an egalitarian distribution we have that $B(\pi) = 1$, $0 \le \pi \le 1$, whereas when the concentration is maximum, $B(\pi) = 0$ if $0 \le \pi < 1$ and $B(1) = 1$. Plots of the Lorenz and Bonferroni curves for some parameter values are displayed in Figure 3.4. Equations (3.3) and (3.4) are the main result of this section.



(a) (b)

Figure 3.4: (a) Bonferroni curves for the NHL distribution for some parameter values; (b) Lorenz curves for the NHL distribution for some parameter values

3.5 Maximum likelihood estimation

The maximum likelihood method for estimation of the three parameters of the NHL distribution is presented in this section. Let $x_1, \ldots, x_n$ be a observed sample of size n from the NHL$(\alpha, \lambda, \gamma)$ distribution and $\boldsymbol{\theta} = (\alpha, \lambda, \gamma)^T$ the parameter vector of interest. The log-likelihood function for $\boldsymbol{\theta}$ based on this sample is

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= n - \gamma \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} (1 + \lambda x_i)^{\alpha} - n \log(1 + \gamma) \\ &+ \sum_{i=1}^{n} \log\left\{(1 + \gamma + \gamma x_i)\left[\gamma + \alpha\lambda(1 + \lambda x_i)^{\alpha - 1}\right] - \gamma\right\}. \end{aligned} \tag{3.5}$$

The maximum likelihood estimates (MLEs) of the model parameters can be obtained by maximizing (3.5). Alternatively, we can differentiating (3.5) and solving the resulting nonlinear likelihood equations. We can express the components of the score vector $\boldsymbol{U(\theta)}$ as

$$
\begin{aligned}
U_\alpha(\boldsymbol{\theta}) = {} & \lambda \sum_{i=1}^{n} \frac{1}{R(x_i)} (1+\gamma+\gamma x_i)(1+\lambda x_i)^{\alpha-1}[1+\alpha\log(1+\lambda x_i)] \\
& + \sum_{i=1}^{n} (1+\lambda x_i)\log(1+\lambda x_i),
\end{aligned}
$$

$$
\begin{aligned}
U_\lambda(\boldsymbol{\theta}) = {} & \alpha \sum_{i=1}^{n} \frac{1}{R(x_i)} (1+\alpha\lambda x_i)(1+\gamma+\gamma x_i)(1+\lambda x_i)^{\alpha-2} \\
& - \alpha \sum_{i=1}^{n} x_i (1+\lambda x_i)^{\alpha-1},
\end{aligned}
$$

and

$$
U_\gamma(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{1}{R(x_i)} (1+x_i) \left[2\gamma + \alpha\lambda(1+\lambda x_i)^{\alpha-1}\right] - \sum_{i=1}^{n} x_i - \frac{n}{1+\gamma},
$$

where $R(x) = (1+\gamma+\gamma x)\left[\gamma + \alpha\lambda(1+\lambda x_i)^{\alpha-1}\right] - \gamma$.

The MLEs of the three unknown parameters are obtained by setting the equations above to zero, $\boldsymbol{U(\theta)} = \mathbf{0}$, and solving them simultaneously. Once these equations can not be solved analytically, we have to use iterative techniques. The quasi-Newton BFGS and Newton-Raphson algorithms are well-known alternatives for solving these equations, see Press *et al.* (2007). The initial values for the parameters are not hard to obtain by fitting special NHL models.

The observed information matrix given by

$$
\boldsymbol{J(\theta)} = -\frac{\partial^2 \ell(\theta)}{\partial\theta\,\partial\theta^T} = \begin{pmatrix} J_{\alpha\alpha} & J_{\alpha\lambda} & J_{\alpha\gamma} \\ . & J_{\lambda\lambda} & J_{\lambda\gamma} \\ . & . & J_{\gamma\gamma} \end{pmatrix},
$$

can be used to construct approximate confidence intervals for the model parameters. Under standard regularity conditions and for large n, the distribution of $(\hat{\alpha}-\alpha, \hat{\lambda}-\lambda, \hat{\gamma}-\gamma)$ can be approximated by a trivariate normal distribution $N_3(0, \boldsymbol{J(\widehat{\theta})}^{-1})$. The elements of $\boldsymbol{J(\widehat{\theta})}$ can be available from the authors upon request.

3.6 Simulation study

In this section, a Monte Carlo simulation experiment is conducted to evaluate the MLEs of the parameters of the NHL distribution. The simulations are performed by generating observations from six different parameter combinations of the NHL distribution using the inverse transformation method. The number of Monte Carlo replications is $N = 10,000$. We use the subroutine `maxBFGS` with analytical derivatives in `R` for maximizing the log-likelihood function.

The results of the simulations are given in Table 3.1. The RMSEs are estimated from N Monte Carlo replications. We set the sample size at n= 100, 300 and 500. Note that for all parametrizations the mean estimates of the parameters tend to be closer to the true parameter values when the sample size n increases as expected under first-order asymptotic theory.

Table 3.1: Monte Carlo simulation results: mean estimates and RMSEs of the NHL distribution for some parameter values.

				Mean estimates			**RMSEs**		
n	α	λ	γ	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\gamma}$
100	0.1	0.5	2.5	0.430	0.491	2.522	2.781	0.179	0.065
	0.1	2.5	0.5	0.123	2.489	0.507	0.035	0.010	0.011
	0.1	0.2	0.3	0.277	0.184	0.286	0.446	0.028	0.050
	1.3	2.0	2.5	1.466	1.933	2.564	1.425	0.264	0.332
	0.5	2.5	0.1	0.525	2.475	0.106	0.039	0.093	0.015
	2.5	0.5	0.1	3.207	0.438	0.125	1.982	0.126	0.068
300	0.1	0.5	2.5	0.211	0.490	2.515	0.356	0.011	0.020
	0.1	2.5	0.5	0.115	2.490	0.509	0.020	0.010	0.009
	0.1	0.2	0.3	0.168	0.189	0.300	0.153	0.012	0.022
	1.3	2.0	2.5	1.391	1.935	2.571	0.443	0.256	0.321
	0.5	2.5	0.1	0.515	2.484	0.108	0.020	0.044	0.011
	2.5	0.5	0.1	2.720	0.477	0.120	0.849	0.070	0.059
500	0.1	0.5	2.5	0.210	0.490	2.513	1.069	0.013	0.020
	0.1	2.5	0.5	0.112	2.490	0.509	0.014	0.010	0,009
	0.1	0.2	0.3	0.145	0.189	0.304	0.150	0.011	0.018
	1.3	2.0	2.5	1.382	1.916	2.603	0.310	0.317	0.432
	0.5	2.5	0.1	0.513	2.486	0.109	0.015	0.033	0.011
	2.5	0.5	0.1	2.643	0.479	0.133	0.510	0.051	0.082

3.7 Application

In this section, we fit the NHL distribution for a real data set to illustrate the potentiality of this distribution for modeling positive data. It represents the times to reinfection of sexually transmitted diseases (STDs) for eight hundred and seventy seven patients. These data are taken from Section 1.12 of Klein and Moeschberger (1997).

Table 3.2 provides a descriptive summary for the reinfection times data. We have large amplitude and variance. The measures of central tendency, such as mean, median and mode, are quite distant when compared among them. Besides, it presents positive values for the skewness and kurtosis. It also exhibit large amplitude and variance.

Table 3.2: Descriptive statistics for the reinfection times data set.

Statistics	**Reinfection times**
Mean	369.5268
Median	247.0000
Mode	5.000
Variance	136940.7000
Skewness	1.15262
Kurtosis	0.51036
Maximum	1529.0000
Minimum	1.0000
n	877

For modeling these data, we fit the NHL distribution and also considered the fits of six related distributions. They are described as follows: the ENH (Lemonte, 2013) with pdf

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

where $\alpha > 0$ and $\beta > 0$ are shape parameters and $\lambda > 0$ is a scale parameter; EW, whose pdf is given by

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

where $\alpha > 0$ and $\beta > 0$ are shape parameters and $\lambda > 0$ is a scale parameter; the Weibull model arises from the EW model when $\beta = 1$; and the NH, Lindley and exponential distributions, which are NHL special models.

We estimate the model parameters of the NHL distribution and the above competitive models by maximum likelihood. The goodness-of-fit statistics considered are: KS, W^* and A^*. The lower are these statistics, the better is the adjustment to the data. The MLEs and goodness-of-fit statistics are evaluated using the `AdequacyModel` script in the `R` software (Marinho *et al.*, 2016).

Table 3.3 lists the MLEs (and the corresponding standard errors in parentheses) of the unknown parameters for the fitted models. We note that all distributions present reasonable estimates for the standard errors. Table 3.4 presents the goodness-of-fit statistics for reinfection times. According to these statistics, the NHL distribution provides a good fit and is quite competitive with the other current distributions.

The introduced distribution yields the best fit under all goodness-of-fit statistics. Figure 3.5 displays the plots of the cumulative and empirical cumulative functions for the most competitive models to both data sets. These plots illustrate the good adjustment of the NHL distribution. Therefore, the results reveal that the proposed distribution can be an effective alternative to the Weibull, EW and ENH distributions, among others.

Table 3.3: The MLEs of the model parameters for reinfection data and corresponding standard errors in parentheses.

Distributions	**Estimates**		
NHL$(\alpha, \lambda, \gamma)$	0.1247	0.0718	0.0031
	(0.0121)	(0.0130)	(0.0001)
ENH(α, λ, β)	0.8445	0.0030	0.7908
	(0.0347)	(0.0001)	(0.0324)
EW(α, λ, β)	0.9819	0.0026	0.7447
	(0.0100)	(0.0001)	(0.0293)
Weibull(α, λ)	0.8464	0.0073	
	(0.0206)	(0.0009)	
NH(α, λ)	0.7950	0.0039	
	(0.0363)	(0.0003)	
Lindley(γ)	0.0053		
	(0.0001)		
Exp(λ)	0.0026		
	(7.568e-05)		

3.8 Concluding remarks

We introduce the Nadarajah-Haghighi Lindley (NHL) model by compounding the Lindley and Nadarajah-Haghighi distributions. Once we have a composition by taking the minimum of two continuous independent random variables, the proposed distribution might be useful in engineering for modeling the failure time of systems composed of two independent components in series. The NHL distribution has the Lindley, Nadarajah-Haghighi and exponential distributions as submodels. Besides, it is a competitive model to the Weibull, exponentiated Weibull

Table 3.4: Goodness-of-fit statistics for the models fitted to reinfection times.

	Statistics		
Distributions	W^*	A^*	KS
NHL	**0.0904**	**0.9826**	**0.0332**
Lindley	0.3532	2.8802	0.2076
NH	0.5782	4.2894	0.0677
ENH	0.4790	3.6459	0.0477
Exp	0.3572	2.8505	0.0844
Weibull	0.5082	3.8380	0.0552
EW	0.3422	2.7487	0.0576

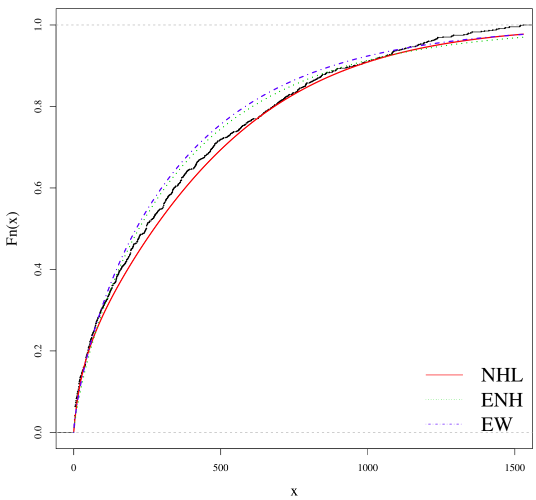


Figure 3.5: Estimated cdfs of the NHL, ENH and EW models for the reinfection times.

and exponentiated Nadarajah-Haghighi distributions, among others. We obtain some structural properties of the proposed distribution, perform the estimation of the parameters by maximum likelihood and provide an application to a real lifetime data set. The new distribution yields a good adjustment for the times to reinfection of sexually transmitted diseases and is quite competitive to other classical lifetime models.

Chapter 4

The Weibull Nadarajah-Haghighi distribution

Resumo

Neste capítulo é introduzida uma nova distribuição de quatro parâmetros, denominada *Weibull Nadarajah-Haghighi*. O novo modelo é obtido considerando a distribuição Nadarajah-Haghighi como *baseline* na família *Weibull-G*, definida por Bourguignon *et al.* (2014). A distribuição proposta pode fornecer as formas constante, crescente, decrescente, banheira e banheira invertida para sua função de risco. Estes são os formatos de maior importância e utilidade em análise de sobrevivência. São exploradas algumas propriedades matemáticas da nova distribuição, tais como a função quantílica, momentos ordinários e incompletos e a entropia de Rényi. O método de máxima verossimilhaça é utilizado para obter as estimativas dos parâmetros do modelo. Um estudo de simulação é conduzido. Além disso, a aplicabilidade da distribuição proposta é ilustrada através de dois conjuntos de dados reais. Para ambas as bases de dados, a nova distribuição apresentou melhores ajustes que algumas distribuições amplamente conhecidas em análise de sobrevivência.

Palavras-chave: Dados de tempo de vida. Distribuição Nadarajah-Haghighi. Família Weibull-G. Função taxa de falha. Máxima verossimilhança.

Abstract

In this chapter, we introduce a new four-parameter model called the Weibull Nadarajah-Haghighi distribution. It is obtained by inserting the Nadarajah-Haghighi distribution in the *Weibull-G* family pioneered by Bourguignon *et al.* (2014). The proposed distribution can produce constant, increasing, decreasing, bathtub and upsidedown-bathtub hazard rate shapes, which are the most important and useful hazard shapes in lifetime analysis. We explore some of its structural properties including the quantile function, ordinary and incomplete moments and Rényi entropy. The maximum likelihood method is presented to estimate the model parameters. A simulation study is performed. The applicability of the new distribution is illustrated by means of two real data sets. The new model provides better fits than some widely known lifetime distributions.

Keywords: Hazard rate function. Lifetime data. Maximum likelihood. Nadarajah-Haghighi distribution. Weibull-G family.

4.1 Introduction

The Weibull, gamma and exponential are three of the most popular distributions in lifetime data analysis. The first one has an advantage because it has survival and hazard functions in closed-form and has the exponential distribution as a special model, see Murthy *et al.* (2004) for details. An alternative model is the EE distribution, discussed by Gupta *et al.* (1998). The EE model is important because it has several properties similar to those of the gamma distribution. Thus, it is a generalization of the exponential model and has closed-form survival and hazard functions.

Gupta and Kundu (2001), Zheng (2002), Gupta and Kundu (2007), Abdel-Hamid and AL-Hussaini (2009) and Nadarajah (2011) provided properties and applications of the EE distribution. For some generalizations of the Weibull distribution, we refer the reader to Mudholkar and Srivastava (1993) for the EW, Xie *et al.* (2002) for the modified Weibull, Cordeiro *et al.* (2010) for the Kw-W and Bagheri *et al.* (2016) for the generalized modified Weibull power series distributions.

Another generalization of the exponential distribution that may be an alternative to those classical lifetime models was introduced by Nadarajah and Haghighi (2011). The Nadarajah-Haghighi (NH) distribution has cdf and pdf given by (2.1) and (2.2) and has the exponential distribution is a special case for $\alpha = 1$. The motivations for studying the NH distribution are the relationship between the pdf (2.2) and the hrf, the ability (or inability) to model data with mode fixed at zero and the fact that it can be interpreted as a truncated Weibull distribution. Further details and general properties can be found in Nadarajah and Haghighi (2011).

Many authors developed extensions of the NH distribution in recent years. For example, Lemonte (2013) pioneered the ENH and Bourguignon *et al.* (2015) introduced the gamma Nadarajah-Haghighi (GNH) distributions. Lima (2015) studied the Kumaraswamy Nadarajah-Haghighi (KNH) and VedoVatto *et al.* (2016) investigated the exponentiated generalized Nadarajah-Haghighi (EGNH) distributions. The NH model is also a special case of the GPW distribution proposed by Bagdonavicius and Nikulin (2002) and discussed by Nikulin and Haghighi (2006, 2009).

Further, Dias (2016) introduced the Nadarajah-Haghighi-G family of distributions. This family is a sub-family of the T-X class defined by Alzaatreh *et al.* (2013). The beta-G family pioneered by Eugene *et al.* (2002) and the gamma-G family defined by Zografos and Balakrishnan (2009) are two well-known generated families in this class. Bourguignon *et al.* (2014) proposed another useful family in the T-X class called the *Weibull-G* family.

Let $G(x)$ and $g(x)$ denote the cdf and pdf of a baseline model with parameter vector $\boldsymbol{\xi}$. Consider the Weibull cdf $F(x) = 1 - \mathrm{e}^{-a\,x^b}$ for $x > 0$, $a > 0$ and $b > 0$. The *Weibull-G* family is obtained by replacing the argument x with $G(x)/\overline{G}(x)$ in the Weibull cdf, where $\overline{G}(x) = 1 - G(x)$. Then, for $x \in \mathcal{D} \subseteq \mathbb{R}$, the cdf and pdf of the *Weibull-G* family are given by

$$F(x; a, b, \boldsymbol{\xi}) = \int_0^{\frac{G(x;\boldsymbol{\xi})}{1-G(x;\boldsymbol{\xi})}} a\,b\,t^{b-1}\mathrm{e}^{-a\,t^b}\mathrm{d}t = 1 - \exp\left\{-a\left[\frac{G(x;\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})}\right]^b\right\} \tag{4.1}$$

and

$$f(x; a, b, \boldsymbol{\xi}) = a\,b\,g(x;\boldsymbol{\xi})\frac{G(x;\boldsymbol{\xi})^{b-1}}{\overline{G}(x;\boldsymbol{\xi})^{b+1}}\exp\left\{-a\left[\frac{G(x;\boldsymbol{\xi})}{\overline{G}(x;\boldsymbol{\xi})}\right]^b\right\}, \tag{4.2}$$

respectively. If $b = 1$, we have the exp-G family. Note that the *Weibull-G* family does not have as a special case the baseline G distribution. However, we can consider the distributions of this family as a compounding between the Weibull and the baseline distributions (Tahir *et al.*, 2016b).

In this Chapter, we introduce a new four-parameter distribution called *the Weibull Nadarajah-Haghighi* (WNH) distribution. Inserting Equation (2.1) in (4.1) yields the WNH cdf given by

$$F(x) = 1 - \exp\left\{-a\,[\exp\{(1+\lambda\,x)^{\alpha} - 1\} - 1]^{b}\right\}. \tag{4.3}$$

The corresponding pdf reduces to

$$\begin{aligned} f(x) &= a\,b\,\alpha\,\lambda\,(1+\lambda\,x)^{\alpha-1}\,[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{b-1} \\ &\quad\times \exp\left\{-b\,[1-(1+\lambda\,x)^{\alpha}] - a\,[\exp\{(1+\lambda\,x)^{\alpha}-1\}-1]^{b}\right\}. \end{aligned} \tag{4.4}$$

Hereafter, a random variable with density function (4.4) is denoted by $X \sim \mathrm{WNH}(a, b, \alpha, \lambda)$. The hrf of X becomes

$$\begin{aligned} \tau(x) &= a\,b\,\alpha\,\lambda\,(1+\lambda\,x)^{\alpha-1}\,[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{b-1} \\ &\quad\times \exp\{-b\,[1-(1+\lambda\,x)^{\alpha}]\}. \end{aligned}$$

Some motivations for proposing the WNH distribution are:

1. The new distribution allows for greater flexibility of its pdf than the baseline density. The NH density is only monotonically decreasing, but the WNH density can be unimodal and is quite flexible for skewness and kurtosis. Figure 4.1 displays plots of the WNH pdf for some parameter values and reveals that the proposed density allows to fit left and right skewed data. More details about the skewness and kurtosis can also be found in Section 4.3.2.

2. The NH distribution can only have monotonic types of hrf, but the WNH hrf presents decreasing, increasing, upside-down bathtub and bathtub-shaped forms. This feature makes the new distribution quite competitive with other popular lifetime distributions and very attractive to be used to model lifetime data. Figure 4.2 displays plots showing all mentioned hrf forms.

3. The WNH distribution contains as special models some known distributions. For $\alpha = 1$, it reduces to the Weibull-Exponential distribution introduced by Oguntunde *et al.* (2015). If $a = \theta/\lambda$ (for $\theta > 0$), $b = 1$ and $\alpha = 1$, it becomes the Gompertz distribution.

4. In practical situations, the WNH distribution may provide 'better fits' than other generated models under the NH baseline. See the results of Section 4.6. They reveal that the WNH model can be a useful alternative not only to other NH generated distributions but also to other widely known lifetime models.

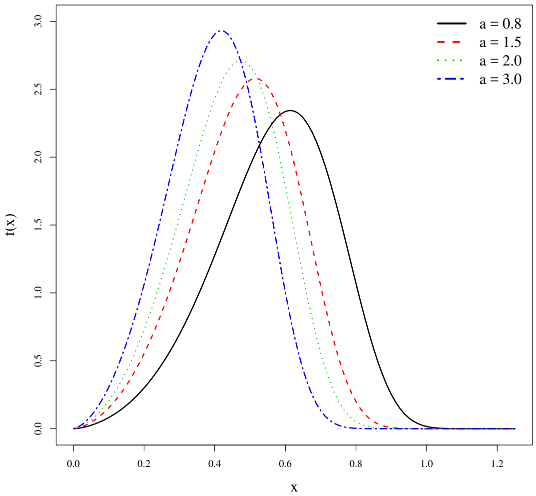
The Chapter is outlined as follows. In Section 4.2, we derive a linear representation for the WNH density function. In Section 4.3, we explore some structural properties of the new distribution. The estimation of the model parameters by maximum likelihood is presented in Section 4.4. A simulation study is performed in Section 4.5. Section 4.6 provides two applications to real data for illustrative purposes. Section 4.7 offers some concluding remarks.

4.2 Useful expansion

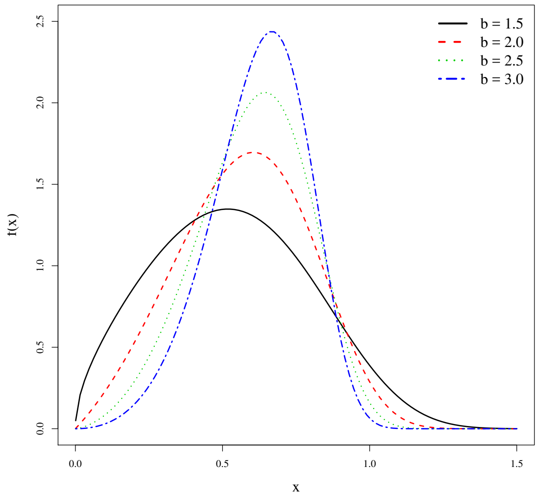
Bourguignon *et al.* (2014) demonstrated that the *Weibull-G* density function can be expressed in terms of the exp-G densities. Let $G(y)$ be the baseline cdf of a random variable Y. The exp-G cdf is obtained by a power transformation of $G(y)$ given by $H_c(y) = G(y)^c$, where $c > 0$ is an additional shape parameter. Then, the exp-G density function is given by

$$h_c(y) = c\, g(y) G(y)^c.$$

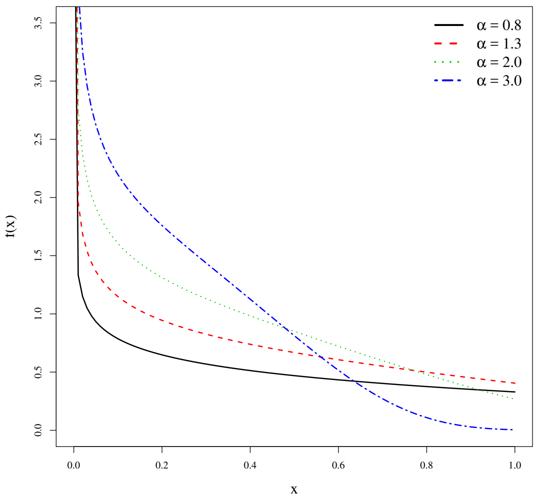
Using such method, we can define the exponentiated exponential (Gupta *et al.*, 1998) and exponentiated Weibull (Mudholkar and Srivastava, 1993) models, among several others distributions. Tahir and Nadarajah (2015) wrote a survey with other different ways for obtaining generated continuous distributions and listed twenty eight different exp-G models already published in the literature.



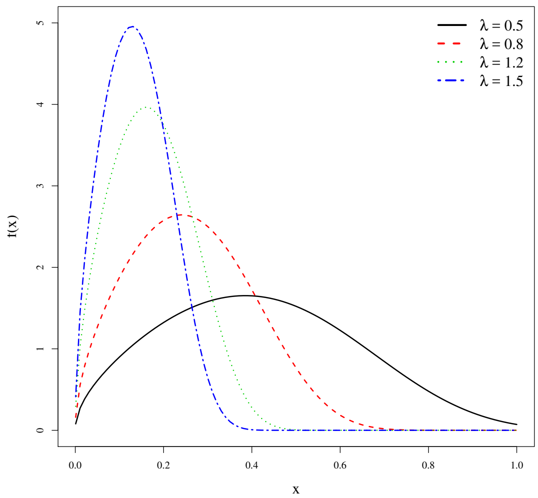
(a) $b = 2.5, \alpha = 2$ and $\lambda = 0.5$



(b) $a = 1.5, \alpha = 2$ and $\lambda = 0.4$



(c) $a = 2.0, b = 1.5$ and $\alpha = 1.8$



(d) $a = 1.4, b = 0.8$ and $\lambda = 0.5$

Figure 4.1: Plots of the WNH density.

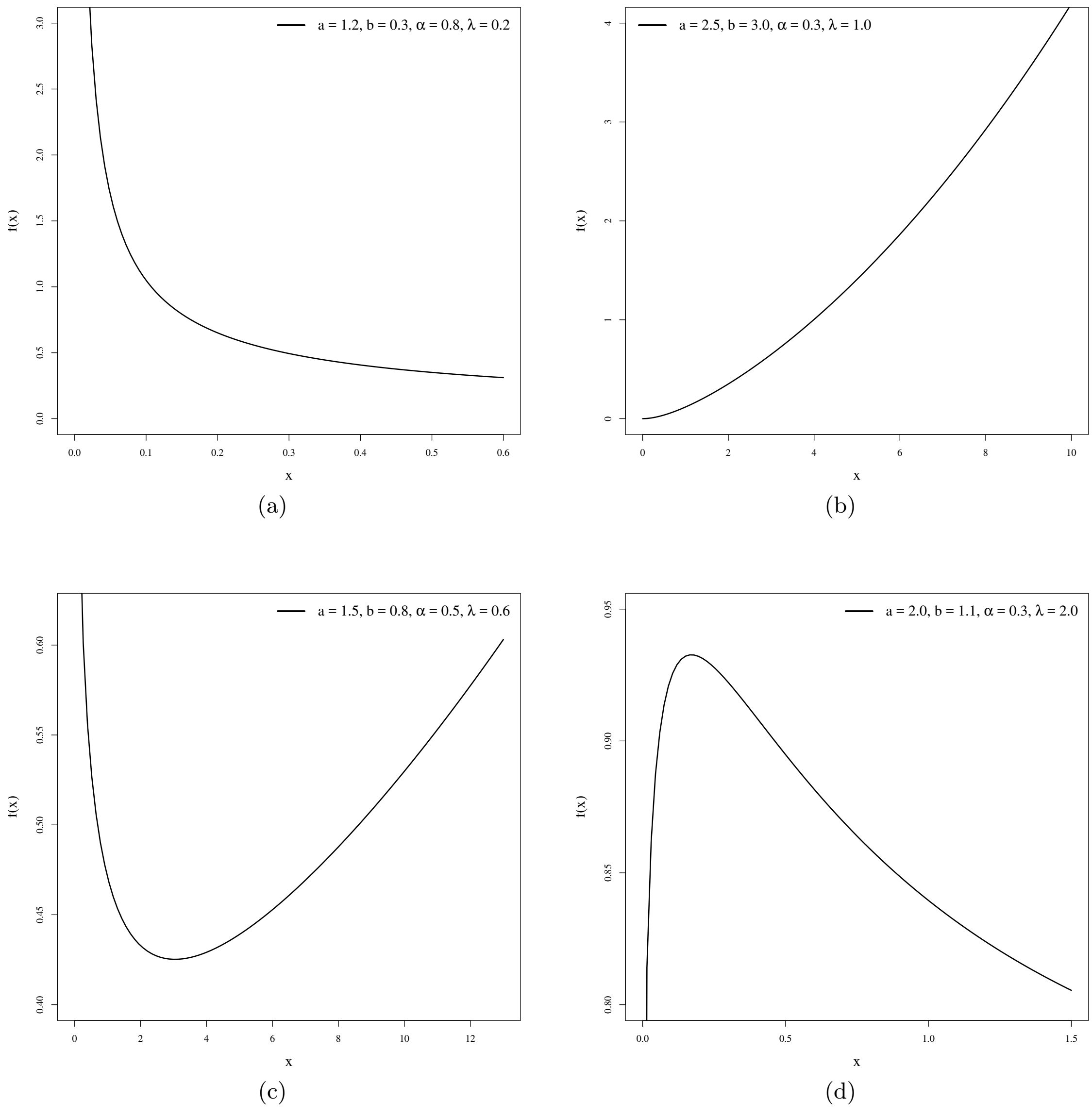


Figure 4.2: Plots of the WNH hrf.

By replacing equations (2.1) and (2.2) in (4.2), we obtain

$$f(x) = a\,b\,\alpha\,\lambda\,(1+\lambda\,x)^{\alpha-1}\exp\{1-(1+\lambda\,x)^{\alpha}\}\frac{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{b-1}}{[\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{b+1}}$$
$$\times\exp\left\{-a\left[\frac{1-\exp\{1-(1+\lambda\,x)^{\alpha}\}}{\exp\{1-(1+\lambda\,x)^{\alpha}\}}\right]^{b}\right\}. \tag{4.5}$$

By expanding the exponential function in the last quantity in (4.5), we have

$$\exp\left\{-a\left[\frac{1-\exp\{1-(1+\lambda\,x)^{\alpha}\}}{\exp\{1-(1+\lambda\,x)^{\alpha}\}}\right]^{b}\right\} = \sum_{i=0}^{\infty}\frac{(-1)^{i}a^{i}}{i!}\,\frac{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{i\,b}}{[\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{i\,b}}.$$

Inserting the above expansion in (4.5) and after some algebra, we obtain

$$f(x) = a\,b\,\alpha\,\lambda\,(1+\lambda\,x)^{\alpha-1}\exp\{1-(1+\lambda\,x)^{\alpha}\}$$
$$\times\sum_{i=0}^{\infty}\frac{(-1)^{i}a^{i}}{i!}\,\frac{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{(i+1)b-1}}{[\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{(i+1)b+1}}. \tag{4.6}$$

By using the generalized binomial theorem, we can rewrite the quantity $[\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{-[(i+1)b+1]}$ as

$$\{1-[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]\}^{-[(i+1)b+1]} = \sum_{j=0}^{\infty}\frac{\Gamma([i+1]b+j+1)}{j!\,\Gamma([i+1]b+1)}$$
$$\times\,[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{j}.$$

By inserting the last equation in (4.6) and after some simplifications, the WNH density function can be expressed as an infinite linear combination of exp-NH densities, namely

$$f(x) = \sum_{i,j=0}^{\infty}\omega_{i,j}\,h_{(i+1)b+j}(x), \tag{4.7}$$

where

$$\omega_{i,j} = \frac{(-1)^{i}\,b\,a^{i+1}\,\Gamma([i+1]b+j+1)}{i!\,j!\,[(i+1)b+j]\,\Gamma([i+1]b+1)}.$$

As mentioned before, the ENH model (Lemonte, 2013) is the exp-G distribution by taking for the baseline the NH model. Figure 4.3 reveals the convergence of $S=\sum_{i,j=0}^{n}\omega_{i,j}$ for $n=1,2,\ldots,15$ and $a=b=0.5$. Equation (4.7) is the main result of this section.

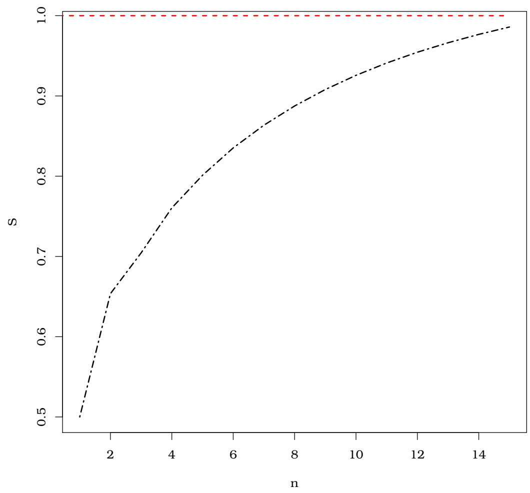


Figure 4.3: Sum of the coefficients $S = \sum_{i,j=0}^{n} \omega_{i,j}$ of the linear combination in (4.7).

4.3 Some structural properties

In this section, we obtain some structural properties of the WNH distribution from those of the ENH model. Our investigation includes the qf, ordinary and incomplete moments, mean deviations, Bonferroni and Lorenz curves and Rényi entropy.

4.3.1 Quantile function

The qf of X is determined by inverting equation (4.3). Thus, for $u \in (0,1)$, we have

$$Q(u) = \frac{1}{\lambda}\left\{\left[1 + \log\left(1 + \left[\frac{-\log(1-u)}{a}\right]^{\frac{1}{b}}\right)\right]^{\frac{1}{\alpha}} - 1\right\}. \tag{4.8}$$

Setting $u = 0.5$ gives the median $M = Q(0.5)$ of X. The qf is a useful tool to obtain skewness and kurtosis measures and for simulating WNH random variables using the inverse transformation method. Let U be a standard uniform random variable. Thus, the random variable $X = Q(U)$ has pdf given by (4.4).

4.3.2 Ordinary and central moments

The sth ordinary moments of X follows from (4.7) as

$$\mu'_s = E(X^s) = \sum_{i,j=0}^{\infty} w_{i,j} \int_0^{\infty} x^s \, h_{(i+1)b+j}(x) \, \mathrm{d}x.$$

Using a result in Lemonte (2013), we obtain

$$\mu'_s = \lambda^{-s} \sum_{i,j,l=0}^{\infty} \sum_{k=0}^{s} \frac{(-1)^{s+l-k} \, [(i+1)b+j] \, \mathrm{e}^{l+1} \, w_{i,j}}{(l+1)^{k/\alpha+1}} \times \binom{(i+1)b+j-1}{l} \binom{s}{k} \Gamma\left(\frac{k}{\alpha}+1, \; l+1\right),$$

where $\Gamma(a,x) = \int_x^{\infty} z^{a-1} \mathrm{e}^{-z} \mathrm{d}z$ denotes the complementary incomplete gamma function.

An alternative representation for the ordinary moments of X can be based on the NH qf. We can write

$$\mu'_s = \lambda^{-s} \sum_{i,j=0}^{\infty} w_{i,j} [(i+1)b+j] I_{i,j}^{(s)}(\alpha, b), \tag{4.9}$$

where $I_{i,j}^{(s)}(\alpha, b) = \int_0^1 u^{(i+1)b+j-1} \{[1-\log(1-u)]^{1/\alpha} - 1\}^s \, \mathrm{d}u$ can be evaluated numerically.

The nth central moment of X, say μ_n, follows as

$$\mu_n = E(X-\mu)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \mu_1'^k \mu'_{n-k}.$$

The cumulants (κ_n) of X can be obtained from (4.9) as

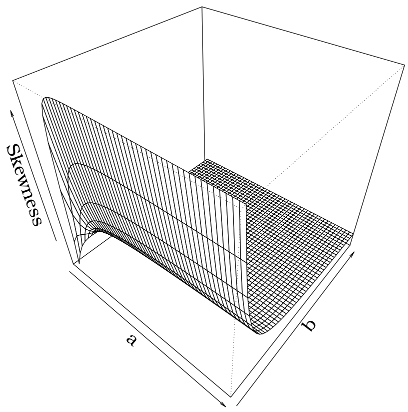
$$\kappa_n = \mu'_s - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \kappa_k \, \mu'_{n-k},$$

where $\kappa_1 = \mu'_1$. The standard skewness and kurtosis measures of X can be determined from the ordinary moments using well-known relationships. The Bowley skewness and the Moors kurtosis of X can also be defined in terms of the qf, respectively, by

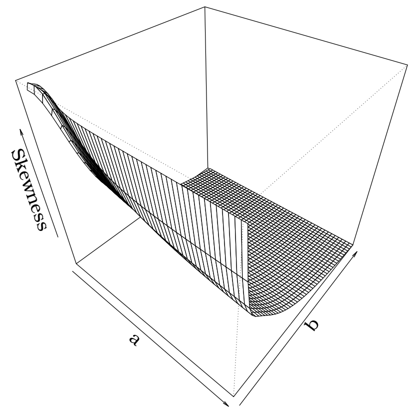
$$B = \frac{Q(3/4) - 2Q(1/2) + Q(1/4)}{Q(3/4) - Q(1/4)} \quad \text{and} \quad M = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(3/4) - Q(1/4)},$$

where $Q(\cdot)$ is given by (4.8). See Kenney and Keeping (1962) and Moors (1988).

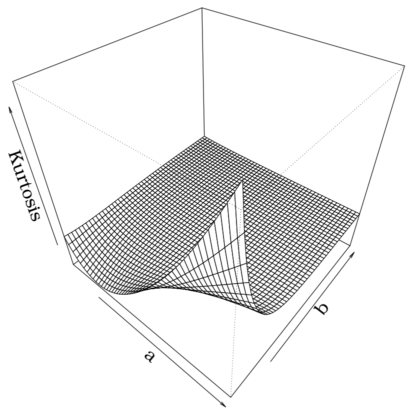
Plots of the measures B and M for some parameter values are displayed in Figure 4.4. These plots indicate that both measures are very sensitive on the shape parameters, thus indicating the flexibility of the model.



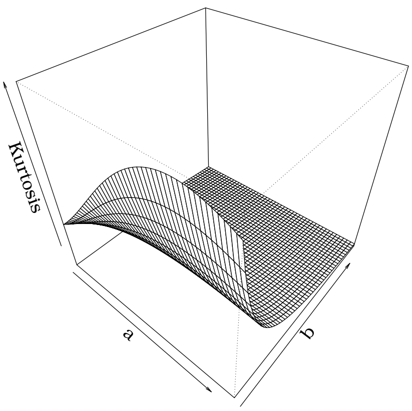
(a) $\alpha = 1.5$, $\lambda = 1.5$



(b) $\alpha = 0.1$, $\lambda = 1.5$



(c) $\alpha = 1.5$, $\lambda = 0.2$



(d) $\alpha = 0.1$, $\lambda = 1.5$

Figure 4.4: The Bowley skewness and Moors kurtosis of the WNH distribution for given α and λ.

4.3.3 Incomplete moments

Let $m_s(y)$ denote the sth incomplete moment of X, say $m_s(y) = \int_0^y x^s f(x)\mathrm{d}x$. From Equation (4.7), we can write

$$m_s(y) = E(X^s) = \sum_{i,j=0}^{\infty} w_{i,j} \int_0^y x^s\, h_{(i+1)b+j}(x)\, \mathrm{d}x.$$

We can show that $m_s(y)$ is given by

$$m_s(y) = \lambda^{-s} \sum_{i,j=0}^{\infty} w_{i,j}[(i+1)b+j] \int_0^{1-\mathrm{e}^{1-(1+\lambda y)^{\alpha}}} \{[1-\log(1-u)]^{1/\alpha}-1\}^s u^{(i+1)b+j-1}\, \mathrm{d}u.$$

Alternatively, we can determine $m_s(y)$ as

$$m_s(y) = \lambda^{-s} \sum_{i,j,l=0}^{\infty} \sum_{k=0}^{s} \frac{(-1)^{s+l-k}\,[(i+1)b+j]\,\mathrm{e}^{l+1}\,w_{i,j}}{(l+1)^{k/\alpha+1}} \times \binom{(i+1)b+j-1}{l}\binom{s}{k} \left[\Gamma\left(\frac{k}{\alpha}+1,\ l+1\right) - \Gamma\left(\frac{k}{\alpha}+1,\ (l+1)(1+\lambda y)^{\alpha}\right)\right].$$

4.3.4 Mean deviations

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

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

respectively, where $\mu_1' = \mathrm{E}(X)$, $M = \mathrm{Median}(X) = Q(0.5)$ is the median, $F(\mu_1')$ is easily determined from (4.3) and $m_1(y) = \int_0^y x\, f(x)\mathrm{d}x$ is the first incomplete moment. Hence, we can write

$$m_1(y) = \lambda^{-1} \sum_{i,j=0}^{\infty} [(j+1)b+i]\, w_{i,j} \int_0^{1-\mathrm{e}^{1-(1+\lambda y)^{\alpha}}} \{[1-\log(1-u)]^{1/\alpha}-1\}\, u^{(i+1)b+j-1}\, \mathrm{d}u.$$

Alternatively, we can determine $m_1(y)$ from the sums for $m_s(y)$ given in Section 4.3.3 by taking $s = 1$. An important application of the previous result refers to the Bonferroni and Lorenz curves, which have not only use in economics but also in other fields like reliability, insurance, and medicine. They are defined, for a given probability π, by $B(\pi) = m_1(q)/(\pi\mu_1')$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $q = Q(\pi)$ follows from (4.8).

4.3.5 Rényi entropy

An important measure of variation of the uncertainty of a random variable is the Rényi entropy. The theory of entropy has been successfully used in a wide variety of applications in fields like physics, engineering, and economics. For the density function $f(x)$, the Rényi entropy is defined by

$$I_R(\delta) = \frac{1}{1-\delta}\,\log\left[\int_{-\infty}^{\infty} f^{\delta}(x)\mathrm{d}x\right], \tag{4.10}$$

for $\delta > 0$ and $\delta \neq 1$. By inserting (4.5) in equation (4.10), we obtain the Rényi entropy of X as

$$\begin{aligned} I_R = C + \frac{1}{1-\delta}\log\Bigg\{\int_0^{\infty} (1+\lambda\,x)^{\delta(\alpha-1)}\exp\{\delta[1-(1+\lambda\,x)^{\alpha}]\} \qquad\qquad (4.11)\\ \times\,\frac{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{\delta(b-1)}}{[\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{\delta(b+1)}}\exp\left\{-a\delta\left[\frac{1-\exp\{1-(1+\lambda\,x)^{\alpha}\}}{\exp\{1-(1+\lambda\,x)^{\alpha}\}}\right]^{b}\right\}\mathrm{d}x\Bigg\}, \end{aligned}$$

where $C = [\delta/(1-\delta)]\log(a\,b\,\delta\,\lambda)$. The above integral can be evaluated numerically. By expanding the exponential function in equation (4.11), we can write

$$\begin{aligned} I_R = C + \frac{1}{1-\delta}\log\Bigg\{\sum_{i=0}^{\infty}\frac{(-1)^i(\delta a)^i}{i!}\int_0^{\infty}(1+\lambda\,x)^{\delta(\alpha-1)}\exp\{\delta[1-(1+\lambda\,x)^{\alpha}]\}\\ \times\,\frac{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{\delta(b-1)+ib}}{[\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{\delta(b+1)+ib}}\,\mathrm{d}x\Bigg\}. \end{aligned}$$

Rewriting $[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}]^{-\delta(b+1)-ib}$ as $\left\{1-[1-\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}]\right\}^{-\delta(b+1)-ib}$, using the generalized binomial expansion and inserting in the last equation, we have

$$\begin{aligned} I_R = C + \frac{1}{1-\delta}\log\Bigg\{\sum_{i,j=0}^{\infty}\frac{(-1)^i(\delta a)^i\,\Gamma(\delta(b+1)+ib+j)}{i!\,j!\,\Gamma(\delta(b+1)+ib)}\\ \times\int_0^{\infty}(1+\lambda\,x)^{\delta(\alpha-1)}\exp\{\delta[1-(1+\lambda\,x)^{\alpha}]\}[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{\delta(b-1)+ib+j}\mathrm{d}x\Bigg\}. \end{aligned}$$

Using the binomial expansion, since $0 < \mathrm{e}^{1-(1+\lambda x)^{\alpha}} < 1$, in the previous expression, setting $u = (1+\lambda x)^{\alpha}(i+\delta)$ and after some algebra, we obtain

$$\begin{aligned} I_R = \frac{\delta}{1-\delta}\log(a\,b) - \log(\alpha\,\lambda) + \frac{1}{1-\delta}\log\Bigg\{\sum_{i,j,k=0}^{\infty}\frac{(-1)^{i+k}\,(\delta a)^i\,\mathrm{e}^{\delta+k}\,\Gamma(\delta(b+1)+ib+j)}{i!\,j!\,(k+\delta)^{[\delta(\alpha-1)+1]/\alpha}\Gamma(\delta(b+1)+ib)}\\ \times\binom{\delta(b-1)+ib+j}{k}\Gamma\left(\frac{\delta(\alpha-1)+1}{\alpha}\,,\,k+\delta\right)\Bigg\}. \end{aligned} \tag{4.12}$$

Equation (4.12) can be computed numericaly.

4.4 Maximum likelihood estimation

In this section, we consider the maximum likelihood method for estimating the unknown parameters of the WNH distribution. Let $x_1, \ldots, x_n$ be a observed sample of size n from the WNH(a, b, α, λ) distribution given by (4.4). Based on this sample, the log-likelihood function for the parameter vector $\boldsymbol{\theta} = (a, b, \alpha, \lambda)$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}) =& n \log(a\, b\, \alpha\, \lambda) + b \sum_{i=1}^{n} (1 + \lambda\, x_i)^{\alpha} + (\alpha - 1) \sum_{i=1}^{n} \log(1 + \lambda\, x_i) \\ &+ (b - 1) \sum_{i=1}^{n} \log\left[1 - \mathrm{e}^{1-(1+\lambda\, x_i)^{\alpha}}\right] - a \sum_{i=1}^{n} \left[\mathrm{e}^{(1+\lambda\, x_i)^{\alpha}-1} - 1\right]^{b} - n\, b. \qquad (4.13)\end{aligned}$$

Equation (4.13) can be maximized either directly or by solving the nonlinear likelihood equations obtained by differentiating (4.13).

The components of the score vector $\boldsymbol{U(\theta)} = [U_a(\boldsymbol{\theta}), U_b(\boldsymbol{\theta}), U_\alpha(\boldsymbol{\theta}), U_\lambda(\boldsymbol{\theta})]^{\top}$ can be expresssed as

$$U_a(\boldsymbol{\theta}) = \frac{n}{a} - \sum_{i=1}^{n} \left[\mathrm{e}^{(1+\lambda\, x_i)^{\alpha}-1} - 1\right]^{b},$$

$$\begin{aligned}U_b(\boldsymbol{\theta}) =& \frac{n}{b} + \sum_{i=1}^{n} (1 + \lambda\, x_i)^{\alpha} + \sum_{i=1}^{n} \log\left[1 - \mathrm{e}^{1-(1+\lambda\, x_i)^{\alpha}}\right] \\ &- a \sum_{i=1}^{n} \left[\mathrm{e}^{(1+\lambda\, x_i)^{\alpha}-1} - 1\right]^{b} \log\left[\mathrm{e}^{(1+\lambda\, x_i)^{\alpha}-1} - 1\right] - n,\end{aligned}$$

$$\begin{aligned}U_\alpha(\boldsymbol{\theta}) =& \frac{n}{\alpha} + b \sum_{i=1}^{n} (1 + \lambda\, x_i)^{\alpha} \log[(1 + \lambda\, x_i)^{\alpha}] + \sum_{i=1}^{n} \log(1 + \lambda\, x_i) \\ &+ (b - 1) \sum_{i=1}^{n} \frac{(1 + \lambda\, x_i)^{\alpha} \log(1 + \lambda\, x_i)\, \mathrm{e}^{1-(1+\lambda\, x_i)^{\alpha}}}{1 - \mathrm{e}^{1-(1+\lambda\, x_i)^{\alpha}}} \\ &- a\, b \sum_{i=1}^{n} (1 + \lambda\, x_i)^{\alpha} \log(1 + \lambda\, x_i)\, \mathrm{e}^{(1+\lambda\, x_i)^{\alpha}-1} \left[\mathrm{e}^{(1+\lambda\, x_i)^{\alpha}-1} - 1\right]^{b-1}\end{aligned}$$

and

$$U_\lambda(\boldsymbol{\theta}) = \frac{n}{\lambda} + (\alpha-1)\sum_{i=1}^{n} x_i\,(1+\lambda x_i)^{-1} - b\,\alpha\sum_{i=1}^{n} x_i\,(1+\lambda x_i)^{\alpha-1}$$
$$- a\,b\,\alpha\sum_{i=1}^{n} x_i(1+\lambda\,x_i)^{\alpha-1}\,\mathrm{e}^{(1+\lambda\,x_i)^\alpha-1}\left[\mathrm{e}^{(1+\lambda\,x_i)^\alpha-1}-1\right]^{b-1}$$
$$+\,\alpha\,(b-1)\sum_{i=1}^{n}\frac{x_i(1+\lambda\,x_i)^{\alpha-1}\,\mathrm{e}^{1-(1+\lambda\,x_i)^\alpha}}{1-\mathrm{e}^{1-(1+\lambda\,x_i)^\alpha}}.$$

The MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can also be obtained by setting $U_a(\boldsymbol{\theta})$, $U_b(\boldsymbol{\theta})$, $U_\alpha(\boldsymbol{\theta})$ and $U_\lambda(\boldsymbol{\theta})$ equal to zero and solving these equations simultaneously. Once they can not be solved analytically, we may use iterative techniques such as Newton-Raphson algorithm for the maximization (Press *et al.*, 2007).

Alternatively, we have from $U_a(\boldsymbol{\theta}) = 0$ that a semi-closed MLE for a is given by

$$\hat{a}(\hat{b},\hat{\alpha},\hat{\lambda}) = \frac{n}{\sum_{i=1}^{n}\left[\mathrm{e}^{(1+\hat{\lambda}\,x_i)^{\hat{\alpha}}-1}-1\right]^{\hat{b}}}.$$

Letting $\boldsymbol{\theta_p} = (b,\alpha,\lambda)$ and replacing $\hat{a}$ in (4.13), we can use the profile log-likelihood to obtain the MLEs for the other three parameters. It is given by

$$\ell(\boldsymbol{\theta_p}) = n\log\left(\frac{n\,b\,\alpha\,\lambda}{\sum_{i=1}^{n}\left[\mathrm{e}^{(1+\lambda\,x_i)^\alpha-1}-1\right]^b}\right) + b\sum_{i=1}^{n}(1+\lambda\,x_i)^\alpha + (\alpha-1)\sum_{i=1}^{n}\log\,(1+\lambda\,x_i)$$
$$+\,(b-1)\sum_{i=1}^{n}\log\left[1-\mathrm{e}^{1-(1+\lambda\,x_i)^\alpha}\right] - n\,(b+1). \tag{4.14}$$

The maximization of (4.14) may be simpler than of (4.13) because it involves only three parameters.

The MLEs have interesting asymptotic properties that allow to construct approximate confidence intervals and testing hypotheses for the model parameters. For n large, and under standard regularity conditions, the distribution of $(\hat{a}-a, \hat{b}-b, \hat{\alpha}-\alpha, \hat{\lambda}-\lambda)$ can be approximated by a multivariate normal distribution $N_4(0, \boldsymbol{J}(\widehat{\boldsymbol{\theta}})^{-1})$, where $\boldsymbol{J}(\widehat{\boldsymbol{\theta}})$ is the observed information matrix given by

$$\boldsymbol{J(\theta)} = -\frac{\partial^2 \ell(\theta)}{\partial\theta\, \partial\theta^T} = \begin{pmatrix} J_{a\,a} & J_{a\,b} & J_{a\,\alpha} & J_{a\,\lambda} \\ \cdot & J_{b\,b} & J_{b\,\alpha} & J_{b\,\lambda} \\ \cdot & \cdot & J_{\alpha\alpha} & J_{\alpha\lambda} \\ \cdot & \cdot & \cdot & J_{\lambda\lambda} \end{pmatrix}.$$

.

The elements of $\boldsymbol{J(\widehat{\theta})}$ can be available from the authors upon request.

4.5 Simulation study

A Monte Carlo experiment is conducted to examine the performance of the MLEs of the WNH parameters. The simulations are performed as follows:

- The data are generated using the inverse transformation method, where $X = Q(U)$ is obtained from equation (4.8).
- The parameter values are set at $\alpha = 0.1$, $\lambda = 1.1$ and a and b vary in the set $\{1.5, 3.0, 6.5\}$.
- The sample sizes are $n = 100, 300$ and 500.
- Each sample size is replicated 10,000 times.
- The average estimates and RMSEs are calculated.

The simulation results are given in Table 4.1. As expected, the RMSEs decay when the sample size increases. Moreover, the mean estimates of the parameters tend to be closer to the true parameter values for larger values of n. These results are usually expected under first-order asymptotic theory.

4.6 Applications

In this section, two applications to real survival data are presented. In order to illustrate the potentiality of the new distribution, we compare the WNH model with ten other related distributions in terms of model fitting. The considered models are (for $x > 0$):

Table 4.1: Mean estimates and RMSEs of the WNH distribution for some parameter values (with $\alpha = 0.1$ and $\lambda = 1.1$).

			Mean estimates				RMSEs			
n	a	b	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{a}$	$\hat{b}$	$\hat{\alpha}$	$\hat{\lambda}$
100	6.5	3.0	6.361	2.959	0.111	1.430	1.711	0.630	0.045	1.113
		1.5	6.979	1.474	0.127	1.389	2.416	0.349	0.090	1.210
	3.0	6.5	3.007	6.320	0.106	1.200	1.096	0.877	0.019	0.702
		1.5	3.591	1.552	0.105	1.637	1.659	0.350	0.035	1.291
	1.5	6.5	1.361	6.308	0.109	1.211	0.876	0.873	0.023	0.713
		3.0	1.666	2.743	0.124	1.242	1.112	0.717	0.057	1.182
300	6.5	3.0	6.360	2.803	0.115	1.030	1.075	0.532	0.041	0.645
		1.5	6.493	1.523	0.103	1.450	1.294	0.235	0.020	0.940
	3.0	6.5	2.898	6.503	0.101	1.163	0.744	0.501	0.008	0.411
		1.5	3.347	1.520	0.101	1.424	1.052	0.267	0.022	1.047
	1.5	6.5	1.607	6.512	0.101	1.195	0.687	0.531	0.010	0.499
		3.0	1.565	2.834	0.111	1.188	0.901	0.532	0.029	0.722
500	6.5	3.0	6.487	2.973	0.103	1.228	1.015	0.385	0.019	0.627
		1.5	6.446	1.535	0.102	1.396	1.162	0.213	0.015	0.839
	3.0	6.5	3.009	6.464	0.102	1.106	0.634	0.528	0.010	0.421
		1.5	3.161	1.526	0.101	1.396	1.030	0.203	0.015	0.896
	1.5	6.5	1.413	6.451	0.102	1.109	0.568	0.418	0.007	0.388
		3.0	1.593	2.954	0.105	1.222	0.687	0.390	0.022	0.611

- The EGNH density given by

$$f(x) = a\,b\,\alpha\,\lambda\,\frac{(1+\lambda\,x)^{\alpha-1}\,\left[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}}{\left\{1-\left[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}\right\}^{1-b}}.$$

- The KwNH density given by

$$f(x) = a\,b\,\alpha\,\lambda\,\frac{(1+\lambda\,x)^{\alpha-1}\,\left[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}\left\{1-\left[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}\right\}^{a-1}}{\left\{1-\left[1-\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}\right\}^{1-b}}.$$

- The GNH density given by

$$f(x) = \frac{\alpha\,\lambda}{\Gamma(a)}\,(1+\lambda\,x)^{\alpha-1}\,[(1+\lambda\,x)^{\alpha}-1]^{a-1}\,\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}.$$

- The WExp density given by

$$f(x) = a\,b\,\lambda\,\left(1-\mathrm{e}^{\lambda\,x}\right)^{b-1}\,\mathrm{e}^{-b\,\lambda\,x-a\left(\mathrm{e}^{\lambda\,x}-1\right)^{b}}.$$

- The GPW density given by

$$f(x) = \alpha\,\lambda\,\gamma\,(1+\lambda\,x^{\gamma})^{\alpha-1}\,\mathrm{e}^{1-(1+\lambda\,x^{\gamma})^{\alpha}}.$$

- The EW density given by

$$f(x) = \alpha\,\beta\,\lambda\,x^{\alpha-1}\,\exp\left(-\lambda\,x^{\alpha}\right)\left[1-\exp\left(-\lambda\,x^{\alpha}\right)\right]^{\beta-1}.$$

- The ENH density given by

$$f(x) = \alpha\,\beta\,\lambda\,\frac{(1+\lambda\,x)^{\alpha-1}\exp\{1-(1+\lambda\,x)^{\alpha}\}}{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{1-\beta}}.$$

- The NH density given by (2.2).

- The Weibull density given by

$$f(x) = \alpha\,\lambda\,(\lambda x)^{\alpha-1}\,\mathrm{e}^{-(\lambda x)^{\alpha}}.$$

- The Gompertz density given by

$$f(x) = \theta\,\mathrm{e}^{\lambda\,x}\,\mathrm{e}^{\frac{\theta}{\lambda}\,(\mathrm{e}^{\lambda\,x}-1)}.$$

The parameters of the above densities are all positive real numbers.

The first data set refers to the times of successive failures of the air conditioning system of a fleet with 213 Boeing 720 jet airplanes (Proschan, 1963). The data were also analyzed by Adamidis and Loukas (1998) and Ristić and Balakrishnan (2012), among others. The second data set consists of the service times of 63 Aircraft Windshield (Murthy *et al.*, 2004). The unit of measurement is 1000 h. These data were studied by Tahir *et al.* (2015) for fitting the Weibull Lomax distribution.

Table 4.2 presents the descriptive statistics for both data sets. For the two samples, the mean is larger than the median and the mode is outlier. The Boeing 720 data have great variance and amplitude, showing more variability than the Aircraft Windshield. The first data set presents positive kurtosis, while the second data set has negative one. Moreover, both data sets present positive skewness, indicating right-skewed data. Note that the WNH model allows fitting skewed-data, as explained in Section 1.

Tables 4.3 and 4.5 list the MLEs (and the corresponding standard errors in parentheses) of the unknown parameters for all fitted models to the first and second data sets, respectively.

Table 4.2: Descriptive statistics for the Boeing 720 and Aircraft Windshield data sets.

Statistics	**Real data sets**	
	Boeing 720	**Aircraft Windshield**
Mean	93.14	2.09
Median	57.00	2.06
Mode	14.00	2.50
Variance	11398.47	1.55
Skewness	2.11	0.44
Kurtosis	4.92	-0.27
Minimum	1.00	0.05
Maximum	603.00	5.14
n	213	63

These results are obtained using the `AdequacyModel` script in `R` software (Marinho *et al.*, 2016). We use the simulated-annealing algorithm for maximizing the log-likelihood function for these models.

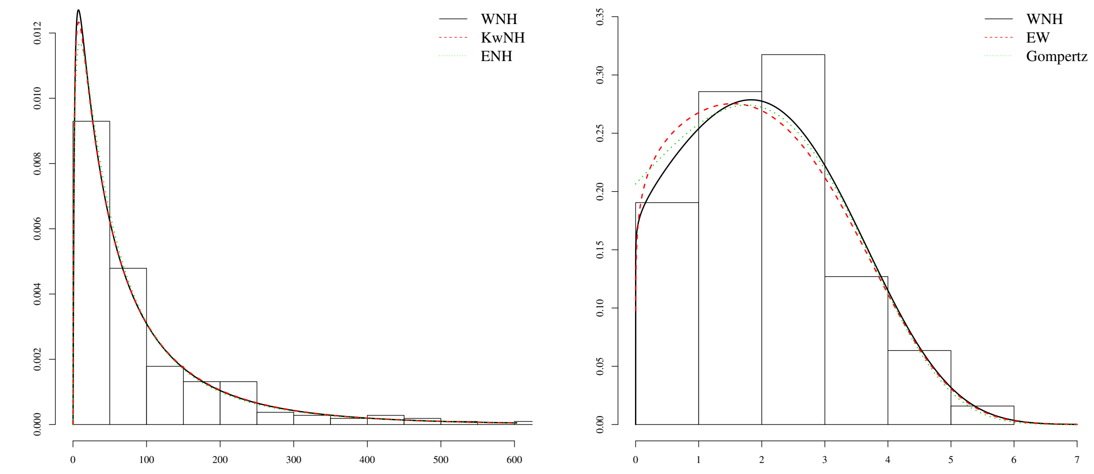
We consider the following goodness-of-fit statistics to select the most appropriate model: the KS statistic and the A^* and W^* corrected statistics. A^* and W^* are tests based on the empirical cdf, and to obtain them, we can proceed as follows: (i) compute $\eta_i = F(x_i, \boldsymbol{\theta})$ where F is a cdf with known form, $\boldsymbol{\theta}$ is a k-dimensional parameter vector unknown and the x_i's are in ascending order; (ii) compute $y_i = \Phi^{-1}(\eta_i)$, where $\Phi(\cdot)$ is the standard normal cdf and $\Phi(\cdot)^{-1}$ its qf; (iii) compute $u_i = \Phi\{(y_i - \bar{y})/s_y\}$, where $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ and $s_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$; (iv) calculate $W^2 = \sum_{i=1}^{n}\{u_i - (2i-1)/(2n)\}^2 + 1/(12n)$ and $A^2 = -n - n^{-1}\sum_{i=1}^{n}\{(2i - 1)\log(u_i) + (2n+1-2i)\log(1-u_i)\}$ and (v) modify W^2 into $W^*(1+0.5/n)$ and A^2 into $A^* = A^2(1 + 0.75/n + 2.25/n^2)$ The lower are them, the better is the model adjustment to the data. For further details, the reader is referred to Chen and Balakrishnan (1995).

Tables 4.4 and 4.6 provide the values of the goodness-of-fit statistics for all the fitted models to the Boeing 720 and the Aircraft Windshield data sets, respectively. The WNH model gives the lowest values of KS, A^* and W^* for both data sets. So, it could be chosen as the best model among the other known lifetime models, including the generated distributions from the NH baseline model.

Figure 4.5 displays the histogram and the plots of estimated densities of the three more competitive models according to the goodness-of-fit statistics, whereas the plots of estimated cdfs for these models are displayed in Figure 4.6. The plots confirm that the WNH distribution yields an effective alternative to other NH generated distributions, such as the KwNH and ENH models in the first application. It also can be useful against other widely known lifetime models, such as the EW and Gompertz models in the second data set.

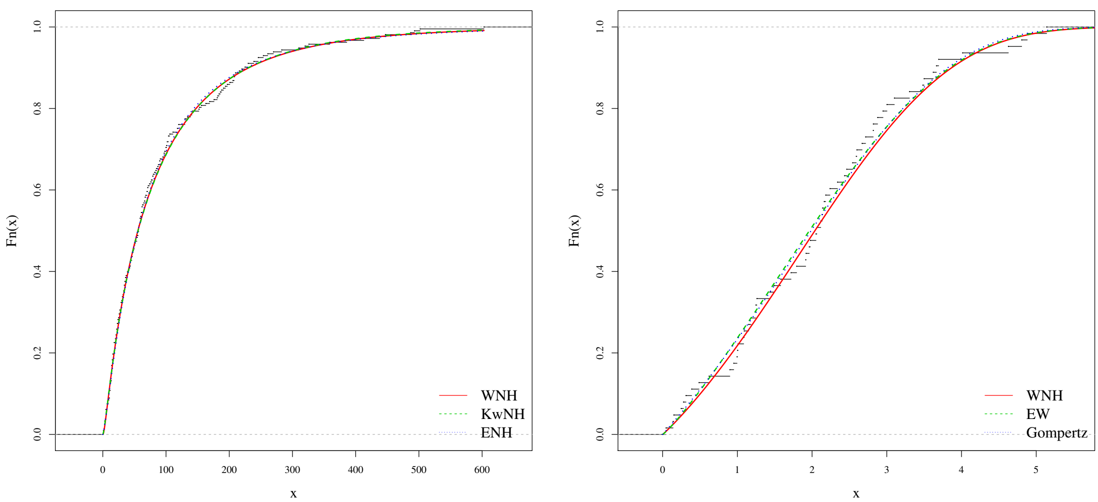
Table 4.3: The MLEs of the model parameters to the Boeing 720 jet airplanes data set.

Distributions	**Estimates**			
WNH(a, b, α, λ)	5.6708	2.6781	0.0793	0.9853
	(2.4485)	(0.2825)	(0.0083)	(0.4853)
EGNH(a, b, α, λ)	0.0119	9.1082	0.9174	1.4956
	(0.0058)	(0.6526)	(0.0692)	(0.4088)
KwNH(a, b, α, λ)	1.5797	0.2124	0.7244	0.1344
	(0.3358)	(0.0891)	(0.0793)	(0.0517)
GNH(a, α, λ)	1.3541	0.5726	0.0440	
	(0.2213)	(0.0775)	(0.0217)	
Wexp(a, b, λ)	2.1559	0.7478	0.0030	
	(0.2971)	(0.0402)	(0.0003)	
GPW$(\alpha, \lambda, \gamma)$	1.5918	0.2756	14.8971	
	(0.1918)	(0.0466)	(2.534)	
EW(α, β, λ)	0.0279	0.6187	2.2841	
	(0.0111)	(0.0883)	(0.6961)	
ENH(α, β, λ)	0.5402	0.0431	1.3862	
	(0.0633)	(0.0152)	(0.2059)	
NH(α, λ)	0.7256	0.0188		
	(0.0891)	(0.0044)		
Weibull(α, λ)	51.9386	0.7593		
	(4.2992)	(0.0429)		
Gompertz(θ, λ)	0.0147	0.0021		
	(0.0035)	(0.0009)		



(a) (b)

Figure 4.5: Histogram and estimated densities of the (a) WNH, KwNH and ENH models for the Boeing 720 data set; (b) WNH, EW and Gompertz models for the Aircraft Windshield data set



(a) (b)

Figure 4.6: Estimated and empirical cdfs of the (a) WNH, KwNH and ENH models for the Boeing 720 data set; (b) WNH, EW and Gompertz models for the Aircraft Windshield data set

Table 4.4: Goodness-of-fit statistics for the models fitted to the Boeing 720 jet airplanes data set.

	Statistics		
Distributions	**KS**	$\mathbf{A^*}$	$\mathbf{W^*}$
WNH	**0.0284**	**0.2226**	**0.0323**
EGNH	0.0417	0.3102	0.6932
KwNH	0.0311	0.2285	0.0337
GNH	0.0367	0.2672	0.0373
Wexp	0.2913	1.7808	0.0954
GPW	0.0827	0.5692	0.0850
EW	0.0411	0.2926	0.0389
ENH	0.0335	0.2534	0.0336
NH	0.0782	0.5091	0.0466
Weibull	0.1046	0.6634	0.1775
Gompertz	0.4733	2.8589	0.1397

Table 4.5: The MLEs of the model parameters to the Aircraft Windshield data set.

Distributions	**Estimates**			
WNH(a, b, α, λ)	0.2135	1.0479	0.6657	1.3431
	(0.0881)	(0.2106)	(0.1176)	(0.4799)
EGNH(a, b, α, λ)	0.0067	9.9244	2.1092	3.9386
	(0.0022)	(1.3329)	(0.1545)	(0.6356)
KwNH(a, b, α, λ)	1.2302	0.2046	2.6044	0.4097
	(0.3722)	(0.0941)	(0.4623)	(0.1126)
GNH$(a\,\alpha\,\lambda)$	1.38304	4.5360	0.0916	
	(0.1867)	(1.7030)	(0.0412)	
WExp(a, b, λ)	2.2251	1.3864	0.1815	
	(1.0921)	(0.1460)	(0.0474)	
GPW$(\alpha, \lambda, \gamma)$	1.311	4.251	8.979	
	(0.1753)	(4.5984)	(9.2221)	
EW(α, β, λ)	0.2829	3.3454	0.3407	
	(0.0373)	(1.0606)	(0.1439)	
ENH(α, β, λ)	5.19228	0.07274	1.42357	
	(1.7410)	(0.0278)	(0.2221)	
NH(α, λ)	6.08583	0.05458		
	(2.4754)	(0.0242)		
Weibull(α, λ)	2.308	1.625		
	(0.1869)	(0.1680)		
Gompertz(θ, λ)	0.2065	0.4882		
	(0.0498)	(0.0984)		

Table 4.6: Goodness-of-fit statistics for the models fitted to the Aircraft Windshield data set.

	Statistics		
Distributions	**KS**	$\mathbf{A}^*$	$\mathbf{W}^*$
WNH	**0.0425**	**0.2812**	**0.0628**
EGNH	0.1114	0.6844	0.7383
KwNH	0.0651	0.3975	0.0799
GNH	0.0697	0.4247	0.0926
WExp	0.0639	0.3907	0.0698
GPW	0.0576	0.3561	0.0877
EW	0.0503	0.3237	0.0739
ENH	0.0763	0.4633	0.0939
NH	0.0754	0.4574	0.1518
Weibull	0.1046	0.6339	0.1095
Gompertz	0.0449	0.3019	0.0675

4.7 Concluding remarks

We introduce the *Weibull Nadarajah-Haghighi* (WNH) model by inserting the Nadarajah-Haghighi distribution in the *Weibull-G* family (Bourguignon *et al.*, 2014). The proposed distribution allows for greater flexibility of the density function than the Nadarajah-Haghighi density and presents constant, increasing, decreasing, bathtub and upsidedown-bathtub hazard rate shapes. It has the Weibull exponential and Gompertz distributions as sub-models. We explore some structural properties of the WNH distribution, estimate its parameters by maximum likelihood, perform a simulation study and provide two applications to real lifetime data sets. The WNH model is quite competitive not only than other NH generated distributions but also to other widely known lifetime models.

Chapter 5

The Logistic Nadarajah-Haghighi distribution

Resumo

Introduzimos uma nova distribuição de três parametros, chamada logistic Nadarajah-Haghighi. Esta é obtida inserindo a distribuição Nadarajah-Haghighi na família *logistic-X*, proposta por Tahir *et al.* (2016a). A nova distribuição pode acomodar densidade unimodal, superando uma limitação da distribuição Nadarajah-Haghighi, a qual possui apenas densidade monótona decrescente. A distribuição proposta também possui função taxa de falha mais flexível que sua baseline, permitindo formas monótonas e de banheira invertida. São obtidas algumas propriedades adicionais da nova distribuição, incluindo a função quantílica, momentos ordinários e incompletos e também os desvios médios e curvas de Lorenz e Bonferroni. O método de máxima verossimilhança é considerado para estimar os parâmetros do modelo. Um estudo de simulação é realizado para verificar a precisão das estimativas e a utilidade da nova distribuição é ilustrada através de duas aplicações a dados reais. O modelo proposto apresenta melhores ajustes que outros modelos sob a mesma distribuição *baseline* e também com relação a outros modelos amplamente utilizados no contexto de análise de sobrevivência.

Palavras-chave: Dados de tempo de vida. Distribuição Nadarajah-Haghighi. Família logistic-X. Função taxa de falha. Máxima verossimilhança.

Abstract

We introduce a new three-parameter model called the logistic Nadarajah-Haghighi distribution. It is obtained by inserting the Nadarajah-Haghighi distribution in the *logistic-X* family pionnered by Tahir *et al.* (2016a). The proposed distribution can produce an unimodal density overcoming a Nadarajah-Haghighi limitation that can only have monotonic decreasing density. It is also more flexible than the baseline hazard rate function, allowing constant, increasing, decreasing and upsidedown-bathtub forms. We obtain some structural properties for the new distribution, including the quantile function, ordinary and incomplete moments, mean deviations and Bonferroni and Lorenz curves. We present the maximum likelihood estimators for the model parameters. A simulation study is carried out to verify the precision of the estimates and we illustrate the usefulness of the new distribution by means of two applications to real data. It provides consistently better fits than other generated models under the same baseline distribution and some other widely known lifetime distributions.

Keywords: Hazard rate function. Lifetime data. Logistic-X family. Maximum likelihood. Nadarajah-Haghighi distribution.

5.1 Introduction

Several authors propose continuous models by adding shape parameter(s) to a baseline distribution. Tahir and Nadarajah (2015) provide an extensive review of the well-established and widely-accepted generated families, such as the exp-G, the beta-G by Eugene *et al.* (2002), the Kumaraswamy-G by Cordeiro and Castro (2011) and the McDonald-G by Alexander *et al.* (2012), among others. According to Tahir and Nadarajah (2015), this parameter induction is useful because it may allow to improve the goodness-of-fit and explore tail properties of the generated family.

In this context, Alzaatreh *et al.* (2013) defined the transformed transform (T-X) class, which includes all the above-mentioned generators as sub-families. Since then, several other generated families have been introduced. For example, the exponentiated T-X family by Alzaghal *et al.*

(2013), Weibull-G by Bourguignon *et al.* (2014), Lomax-G by Cordeiro *et al.* (2014b), the new Weibull-G by Tahir *et al.* (2016c) and the odd Burr-G by Alizadeh *et al.* (2017).

Tahir *et al.* (2016a) pioneered another competitive class of continuous distributions in the T-X family, called the *logistic-X* class ("LX" for short). Let $G(x)$ and $g(x)$ denote, respectively, cdf and pdf of a baseline model with parameter vector $\boldsymbol{\xi}$. Consider the logistic cdf $H(t) = (1+\mathrm{e}^{-\lambda t})^{-1}$, for $t \in I\!\!R$ and $\lambda > 0$. The cdf of the LX family is obtained by replacing the argument t with $\log\{-\log[1-G(x)]\}$ in the logistic cdf. Thus, the cdf and pdf of the LX family are given by

$$F(x;\gamma,\boldsymbol{\xi}) = \frac{1}{1+\{-\log[1-G(x;\boldsymbol{\xi})]\}^{-\gamma}} \tag{5.1}$$

and

$$\begin{aligned} f(x;\gamma,\boldsymbol{\xi}) = \frac{\gamma\, g(x;\boldsymbol{\xi})}{1-G(x;\boldsymbol{\xi})} &\left[1+[-\log(1-G(x;\boldsymbol{\xi})]^{-\gamma}\right]^{-(\gamma+1)} \\ &\times \left\{1+[-\log(1-G(x;\boldsymbol{\xi})]^{-\gamma}\right\}^{-2}, \end{aligned}$$

respectively. Tahir *et al.* (2016a) present some general properties and applications of the LX family. However, there is no much contributed work addressed to specific baselines on such parameter induction. We can only refer the reader to Tahir *et al.* (2016a) for a description of the mathematical properties and applications of the logistic-Fréchet distribution, which follows from the LX family under the Fréchet baseline.

In this chapter, we introduce a new three-parameter distribution so-called the *logistic Nadarajah-Haghighi* (LNH) distribution. It is based on the NH distribution, which is considered as baseline in the LX family. The NH distribution was introduced by Nadarajah and Haghighi (2011) and has cdf and pdf given by (2.1) and (2.2), respectively.

Note that the exponential distribution arises as a special case for $\alpha = 1$. Nadarajah and Haghighi (2011) emphasize some characteristics of the NH model such as: i) its density is always monotonically decreasing; ii) its hrf can only have monotonic types; iii) the ability to model data with mode fixed at zero; iv) it can be interpreted as a truncated Weibull distribution. Some NH generalizations have been introduced in recent years, such as the GPW (Bagdonavicius and Nikulin, 2002), the ENH (Lemonte, 2013) and GNH (Bourguignon *et al.*, 2015) distributions.

Inserting equation (2.1) in (5.1) gives the LNH cdf

$$F(x) = \frac{[(1+\lambda\, x)^{\alpha} - 1]^{\gamma}}{1 + [(1+\lambda\, x)^{\alpha} - 1]^{\gamma}}. \tag{5.2}$$

The corresponding pdf reduces to

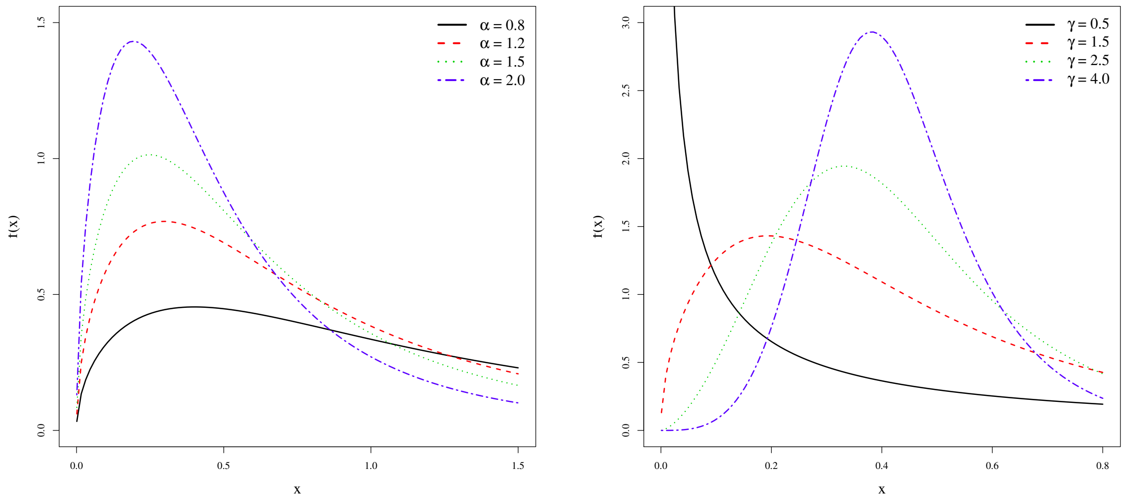
$$f(x) = \frac{\gamma\, \alpha\, \lambda\, (1+\lambda\, x)^{\alpha-1}\, [(1+\lambda\, x)^{\alpha} - 1]^{\gamma-1}}{\{1 + [(1+\lambda\, x)^{\alpha} - 1]^{\gamma}\}^{2}}. \tag{5.3}$$

Hereafter, a random variable with density function (5.3) is denoted by $X \sim \mathrm{LNH}(\gamma, \alpha, \lambda)$. The hrf of X becomes

$$h(x) = \frac{\gamma\, \alpha\, \lambda\, (1+\lambda\, x)^{\alpha-1}\, [(1+\lambda\, x)^{\alpha} - 1]^{\gamma-1}}{1 + [(1+\lambda\, x)^{\alpha} - 1]^{\gamma}}.$$

We provide at least three possible motivations for introducing the LNH distribution. First, the NH distribution can only have monotonic decreasing pdf, but the LNH density overcomes this limitation. Figure 5.1 displays some plots of the LNH pdf for some parameter values and reveal that it can be unimodal, being more flexible than the baseline density. The second motivation is based on the ability to accommodating upside-down bathtub shape for the hrf, which is not allowed under the baseline model. Figure 5.2 illustrates this feature by providing plots of the LNH hrf for some parameter values. The third motivation is about to provide consistently better fits than other generated models under the same baseline distribution, see the results in Section 5.9. Considering these motivations, our purpose by introducing the LNH model is to define a wide flexible distribution with applications to survival analysis and possibly to other fields like biological sciences, economics, engineering, physics, etc.

The rest of the Chapter is outlined as follows. We derive a linear representation for the LNH density function in Section 5.2. Some structural properties of the new distribution are explored in Sections 5.3-5.6. In Section 5.7, we present the maximum likelihood estimators for the model parameters. A simulation study is performed in Section 5.8. In Section 5.9, we provide two applications of the LNH distribution and compare it with others related distributions. Finally, Section 5.10 offers some concluding remarks.



(a) $\lambda = 1$ and $\gamma = 1.5$ (b) $\lambda = 1$ and $\alpha = 2$

Figure 5.1: Plots of the LNH density.

5.2 Useful expansions

Tahir *et al.* (2016a) demonstrated that the *logistic-X* density function can be expressed as an infinite linear combination of exp-G densities. It is a desirable feature since the exp-G family has been widely explored. Tahir and Nadarajah (2015) traced back this family of distributions to the first half of the nineteenth century, but we can also refer the reader to Gupta *et al.* (1998), Mudholkar and Srivastava (1993), Gupta and Kundu (2001) and Sarhan *et al.* (2013) for other studies.

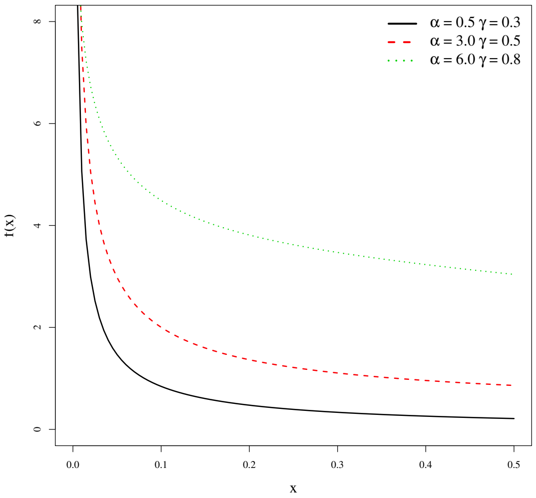
By replacing Equation (2.1) in (5.1), we can write the LNH cdf as

$$F(x) = \frac{1}{1 + \left\{-\log\left[1 - \left(1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}\right)\right]\right\}^{-\gamma}}. \tag{5.4}$$

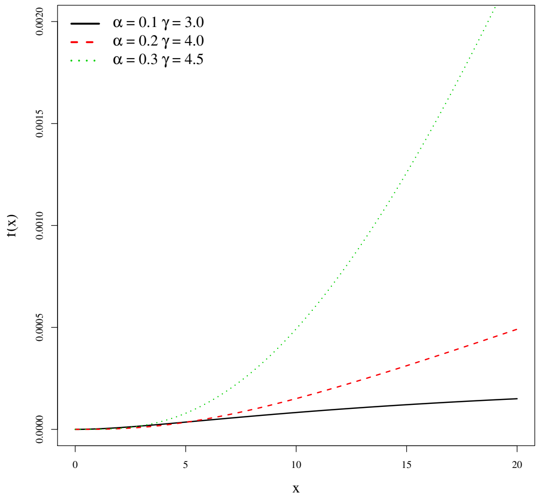
Let $z = 1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}$ and consider the following power series (for $a < 0$)

$$\begin{aligned} 1 + [-\log(1-z)]^a = 1 + \Bigg[1 + \frac{a}{2}\,z + \frac{1}{24}(3a^2 + 5a)\,z^2 + \frac{1}{48}(a^3 + 5a^2 + 6a)\,z^3 \\ + \frac{1}{5760}(15a^4 + 150a^3 + 485a^2 + 502a)\,z^4\Bigg]\,z^a + O(z^{a+5}), \end{aligned}$$

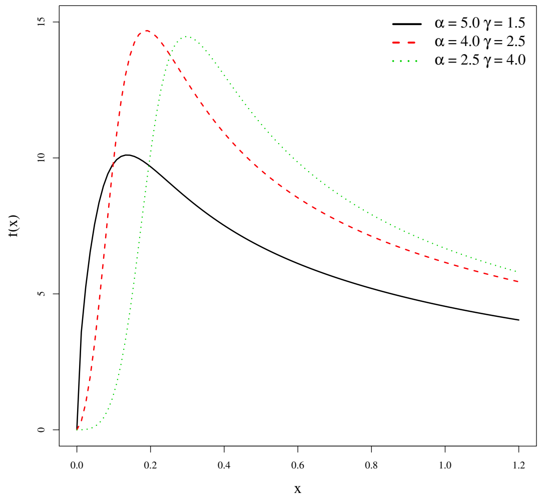
which can be obtained in the `Mathematica` software.



(a) For $\lambda = 1$



(b) For $\lambda = 0.1$



(c) For $\lambda = 1.5$

Figure 5.2: Plots of the LNH hrf.

Using the last expression in (5.4) and after some algebra, we have

$$F(x) = \frac{[1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{\gamma}}{[1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{\gamma} + \sum_{i=0}^{\infty} p_i \, [1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{i}}, \tag{5.5}$$

where the p_i's are given by $p_0 = 1, p_1 = \gamma/2$, $p_2 = \gamma\,(3\gamma + 5)/24$, $p_3 = \gamma\,(\gamma^2 + 5\gamma + 6)/48$, $p_4 = \gamma\,(15\gamma^3 + 150\gamma^2 + 485\gamma + 502)/5760$, etc.

Note that for any $\gamma > 0$ real non-integer, the following expansion holds

$$[1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{\gamma} = \sum_{i=0}^{\infty} q_i \, [1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{i}, \tag{5.6}$$

with $q_i = \sum_{r=i}^{\infty} (-1)^{i+r} \binom{\gamma}{r} \binom{r}{i}$. Thus, inserting (5.6) in equation (5.5) gives

$$F(x) = \frac{\sum_{i=0}^{\infty} q_i \, [1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{i}}{\sum_{i=0}^{\infty} \upsilon_i \, [1 - \mathrm{e}^{1-(1+\lambda x)^{\alpha}}]^{i}} = \sum_{i=0}^{\infty} c_i \, H_i(x), \tag{5.7}$$

where $\upsilon_i = q_i + p_i$ and the coefficients c_i's are determined from the recurrence equation (for $i \geq 0$)

$$c_i = \frac{1}{\upsilon_0} \left(q_i - \frac{1}{\upsilon_0} \sum_{l=0}^{i} \upsilon_r \, c_{i-l} \right)$$

and $H_i(x)$ denotes the exp-G cdf with the NH model as baseline and power parameter i.

Thus, by differentiating (5.7), we can rewrite the LNH density as an infinite linear combination of exp-NH density functions

$$f(x) = \sum_{i=0}^{\infty} c_{i+1} \, h_{i+i}(x), \tag{5.8}$$

where $h_{i+1}(x)$ is the exp-G pdf with the NH model as baseline and power parameter $i + 1$. As mentioned before, the exp-NH distribution was pionnered by Lemonte (2013) and is refereed to as the ENH model. Equations (5.7) and (5.8) are de main results of this section.

5.3 Quantile function

By inverting equation (5.2), we determine the qf of X . Thus, for $u \in (0, 1)$, we have

$$.Q(u) = \frac{1}{\lambda} \left\{ \left[1 + \left(\frac{u}{1-u} \right)^{\frac{1}{\gamma}} \right]^{\frac{1}{\alpha}} - 1 \right\}. \tag{5.9}$$

If U has a uniform distribution in $(0, 1)$, then $Q(U)$ has pdf given by (5.3). Then, simulating LNH random variable using the inverse transformation method is straightforward. The qf is also useful to obtain any quantiles of interest. For example, setting $u = 0.5$ gives the median $M = Q(0.5)$ of X.

The Bowley skewness and the Moors kurtosis are defined in terms of the qf, respectively, by

$$B = \frac{Q(3/4) - 2Q(1/2) + Q(1/4)}{Q(3/4) - Q(1/4)} \quad \text{and} \quad M = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(3/4) - Q(1/4)},$$

where $Q(\cdot)$ is given by (5.9). See Kenney and Keeping (1962) and Moors (1988).

Figure 5.3 displays the measures B and M for some parameter values of the LNH distribution. These plots indicate that the both measures are very sensitive to different values of the shape parameters α and γ, thus indicating the usefullness of the proposed distribution in many practical situations.

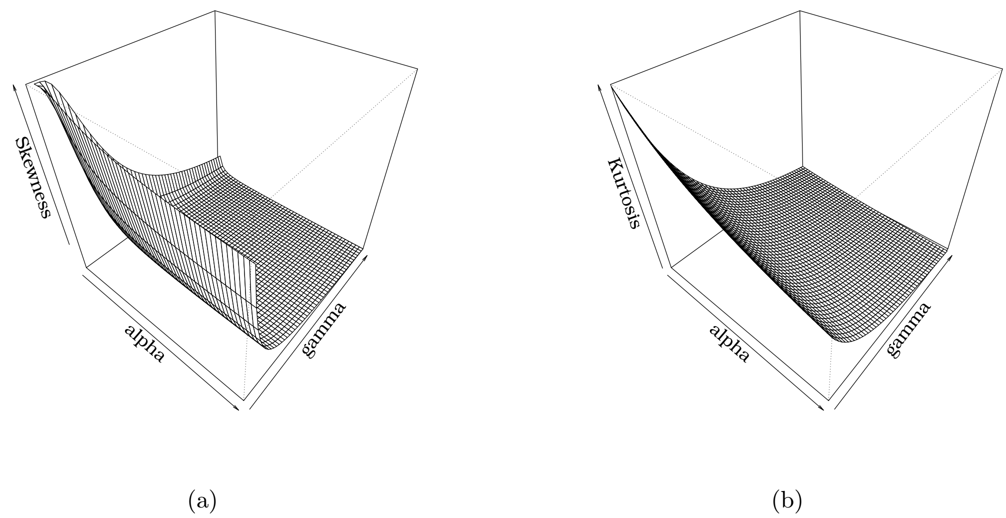


Figure 5.3: The Bowley skewness (a) and Moors kurtosis (b) of the LNH distribution.

5.4 Ordinary and central moments

Let X be a random variable having the LNH distribution. Using the result in (5.8), it is easy to obtain the sth moment of X from the known properties of the exp-G distribution as

$$\mu_s' = E(X^s) = \sum_{i=0}^{\infty} c_{i+1} \int_0^{\infty} x^s\, h_{i+1}(x)\, \mathrm{d}x.$$

Hence, the moments of X can be obtained in closed-form using the result reported in Lemonte (2013). Then, we obtain (for $s = 1, 2, 3, \ldots$)

$$\mu_s' = \lambda^{-s} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{k=0}^{s} \frac{(-1)^{s+j-k}\,(i+1)\,\mathrm{e}^{j+1}\, c_{i+1}}{(j+1)^{k/\alpha+1}} \binom{i}{j} \binom{s}{k} \Gamma\left(\frac{k}{\alpha}+1,\ j+1\right), \tag{5.10}$$

where $\Gamma(a, x) = \int_x^{\infty} z^{a-1} \mathrm{e}^{-z} \mathrm{d}z$ denotes the complementary incomplete gamma function.

The moments of X can also be obtained from (5.9) as

$$\mu_s' = \lambda^{-s} \int_0^1 \left\{ \left[1 + \left(\frac{u}{1-u} \right)^{\frac{1}{\gamma}} \right]^{\frac{1}{\alpha}} - 1 \right\}^s \mathrm{d}u.$$

Using the binomial expansion, since $[1 + u^{1/\gamma}(1-u)^{-1/\gamma}]^{1/\alpha} > 1$, and after some algebra, we obtain

$$\mu_s' = \lambda^{-s} \sum_{i=0}^{s} (-1)^{s-i} \binom{s}{i} I_i(\alpha, \gamma),$$

where $I_i(\alpha, \gamma) = \int_0^{\infty} (1 + u^{1/\gamma})^{i/\alpha} (1+u)^{-2} \mathrm{d}u$. Thus, it is easily observed that

- $\mu_1' = \frac{1}{\lambda(\alpha-1)}$ for $\gamma = 1$ and $\alpha > 1$
- $\mu_1' \to 0$ when $\alpha \to \infty$
- $\mu_1' \to \frac{2^{1/\alpha}-1}{\lambda}$ when $\gamma \to \infty$
- $\mu_s' \to 0$ when $\lambda \to \infty$, $\forall s$.

In general, $I_i(\alpha, \gamma)$ can be evaluated numerically.

The nth central moment of X, say μ_n, follows as

$$\mu_n = E(X - \mu)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \mu_1'^k \mu_{n-k}'.$$

The cumulants (κ_n) of X can be obtained from (5.10) as

$$\kappa_n = \mu'_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \kappa_k \, \mu'_{n-k},$$

where $\kappa_1 = \mu'_1$. The standard skewness and kurtosis measures of X can be determined from the ordinary moments using well-known relationships.

5.5 Incomplete moments

For lifetime models, it is of interest to know the sth incomplete moment of X, defined as $m_s(y) = \int_0^y x^s f(x)\mathrm{d}x$. It follows that

$$m_s(y) = \lambda^{-s} \sum_{i=0}^{s} (-1)^{s-i} \binom{s}{i} \, I_i^*(\alpha, \gamma; y),$$

where

$$I_i^*(\alpha, \gamma; y) = \int_0^{[(1+\lambda y)^\alpha - 1]^\gamma} (1 + u^{1/\gamma})^{i/\alpha} (1+u)^{-2} \mathrm{d}u.$$

An alternative expression for $m_s(y)$ takes the form

$$m_s(y) = \lambda^{-s} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{k=0}^{s} \frac{(-1)^{s+j-k}\,(i+1)\,\mathrm{e}^{j+1}\,c_{i+1}}{(j+1)^{k/\alpha+1}} \binom{i}{j} \binom{s}{k} \left[\Gamma\left(\frac{k}{\alpha}+1,\ j+1\right) - \Gamma\left(\frac{k}{\alpha}+1,\ (j+1)(1+\lambda y)^\alpha\right) \right]. \tag{5.11}$$

It is of some interest to know the sth upper incomplete moment of X defined by $T_s(y) = E(X^s|X > y)$. We can show that $T_s(y)$ is given by

$$T_s(y) = \lambda^{-s} \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{k=0}^{s} \frac{(-1)^{s+j-k}\,(i+1)\,\mathrm{e}^{j+1}\,c_{i+1}}{(j+1)^{k/\alpha+1}} \binom{i}{j} \binom{s}{k} \times \Gamma\left(\frac{k}{\alpha}+1,\ (j+1)(1+\lambda y)^\alpha\right).$$

An application of $T_s(y)$ is the mean residual lifetime function defined by $T_1(y) - y$, which represents the expected additional life length for a unit which is alive at age y.

5.6 Mean deviations

The deviations from the mean and the median are used as measures of the amount of scatter in a population. The mean deviations about the mean and about the median of X can be determined as

$$\delta_1 = \int_0^\infty |X - \mu_1'|\, f(x)\, \mathrm{d}x = 2\mu_1' F(\mu_1') - 2\, m_1(\mu_1')$$

and

$$\delta_2 = \int_0^\infty |X - M|\, f(x)\, \mathrm{d}x = \mu_1' - 2\, m_1(M),$$

respectively, where $F(\mu_1')$ follows from (5.2), $m_1(\mu_1')$ denotes the first incomplete moment and $M = Q(0.5)$ is the median. We can show that $\mu_1' = [I_1(\alpha, \gamma) - 1]$ and

$$m_1(M) = \int_0^{(2/\gamma - 1)^\gamma} (1 + u^{1/\gamma})^{i/\alpha} (1 + u)^{-2} \mathrm{d}u.$$

Expressions for μ_1' and M in closed-form can be easily obtained from equation (5.10) and (5.11), respectively.

Important applications of the previous results are the Bonferroni and Lorenz curves, which have not only use in economics but also in other fields like reliability, insurance, and medicine. They are defined, for a given probability π, by $B(\pi) = m_1(q)/(\pi\mu_1')$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $q = Q(\pi)$ follows from (5.9).

5.7 Maximum likelihood estimation

The maximum likelihood method is the most commonly technique employed for parameter estimation. The interesting asymptotic properties and usefulness for constructing confidence intervals are some advantages of using this procedure. In this section, we provide the procedure to obtain the MLEs of the unknown parameters for the LNH distribution.

Let $x_1, \ldots, x_n$ be a observed sample of size n from the LNH$(\gamma, \alpha, \lambda)$ distribution given by (5.3) and let $\boldsymbol{\theta} = (\gamma, \alpha, \lambda)$ be the parameter vector. The log-likelihood for $\boldsymbol{\theta}$ based on this sample is given by

$$\ell(\boldsymbol{\theta}) = n\,\log(\gamma\,\alpha\,\lambda) + (\alpha-1)\sum_{i=1}^{n}\log(1+\lambda\,x_i) + (\gamma-1)\sum_{i=1}^{n}\log[(1+\lambda\,x_i)^{\alpha}-1] \qquad (5.12)$$
$$-\,2\sum_{i=1}^{n}\log\left\{1+[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma}\right\}.$$

The MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be determined by maximizing (5.12) . Alternatively, they can be obtained by setting all components of the score vector $\boldsymbol{U(\theta)} = [U_{\gamma}(\boldsymbol{\theta}), U_{\alpha}(\boldsymbol{\theta}), U_{\lambda}(\boldsymbol{\theta})]^{\top}$ equal to zero and solving these equations simultaneously. These components are expressed as

$$U_{\gamma}(\boldsymbol{\theta}) = \frac{n}{\gamma} + \sum_{i=1}^{n}\log[(1+\lambda\,x_i)^{\alpha}-1] - 2\sum_{i=1}^{n}\frac{[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma}\log[(1+\lambda\,x_i)^{\alpha}-1]}{1+[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma}},$$

$$U_{\alpha}(\boldsymbol{\theta}) = \frac{n}{\alpha} + \sum_{i=1}^{n}\log(1+\lambda\,x_i) + (\gamma-1)\sum_{i=1}^{n}\frac{(1+\lambda x_i)^{\alpha}\log(1+\lambda\,x_i)}{\log\{1+(1+\lambda\,x_i)^{\alpha}-1]\}}$$
$$-\,2\,\gamma\sum_{i=1}^{n}\frac{(1+\lambda\,x_i)^{\alpha}\,[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma-1}\,\log(1+\lambda\,x_i)^{\alpha}}{\log\{1+[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma}\}}$$

and

$$U_{\lambda}(\boldsymbol{\theta}) = \frac{n}{\lambda} + (\alpha-1)\sum_{i=1}^{n}x_i\,(1+\lambda x_i)^{-1} + \alpha\,(\gamma-1)\sum_{i=1}^{n}\frac{x_i\,(1+\lambda\,x_i)^{\alpha-1}}{(1+\lambda\,x_i)^{\alpha}-1}$$
$$-\,2\,\alpha\,\gamma\sum_{i=1}^{n}\frac{x_i\,(1+\lambda\,x_i)^{\alpha-1}[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma-1}}{1+[(1+\lambda\,x_i)^{\alpha}-1]^{\gamma}}.$$

Note that these equations cannot be solved in closed-form. Thus, we may apply iterative techniques such as Newton-Raphson type algorithms for the maximization (Press *et al.*, 2007). Under standard regularity conditions, we have that $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ can be approximated by a multivariate normal distribution $N_3(0, \boldsymbol{J(\widehat{\theta})}^{-1})$, where $\boldsymbol{J(\widehat{\theta})}$ is the observed information matrix given by

$$\boldsymbol{J(\theta)} = -\frac{\partial^2\,\ell(\theta)}{\partial\theta\,\partial\theta^T} = \begin{pmatrix} J_{\gamma\gamma} & J_{\gamma\alpha} & J_{\gamma\lambda} \\ . & J_{\alpha\alpha} & J_{\alpha\lambda} \\ . & . & J_{\lambda\lambda} \end{pmatrix}.$$

.

The elements of $\boldsymbol{J(\widehat{\theta})}$ can be available from the authors upon request.

5.8 Simulation study

A Monte Carlo experiment is conducted to examine the performance of the MLEs of the LNH parameters. The simulations are performed as follows:

- The parameter values are chosen.
- The data are generated using the inverse transformation method, where $X = Q(U)$ is obtained from equation (5.9).
- The sample sizes are $n = 100, 300$ and 500.
- Each sample size is replicated 10,000 times.
- The average estimates and RMSE are calculated.

The simulation results are given in Table 5.1. As expected, the RMSEs decay when the sample size increases. Moreover, the mean estimates of the parameters tend to be closer to the true parameter values for larger values of n. These results are usually expected under first-order asymptotic theory.

5.9 Applications

In this section, we present two applications to real data sets to illustrate that, in practical situations, the LNH model can provide "better fits" than other related distributions. The first data set was reported by Bjerkedal (1960). It is about the survival times of 72 guinea pigs infected with virulent tubercle bacilli. The data (measured in days) were previous considered by Gupta *et al.* (1997) for fitting the lognormal distribution. The second data set represents the failure times of 20 mechanical components and was reported by Murthy *et al.* (2004). Cordeiro and Bourguignon (2016) also considered these data in order to illustrate the importance of the Ristić-Balakrishnan family of distributions.

Table 5.2 presents the descriptive statistics for the two data sets. For both cases, the mean is greater than the median and mode. We have positive skewness and kurtosis. The first data

Table 5.1: Mean estimates and RMSEs for some LNH parameter values ($\alpha = 0.1$ and $\lambda = 1.1$).

				Mean estimates			**RMSEs**		
n	γ	α	λ	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\lambda}$
100	0.5	1.8	0.2	0.503	2.030	0.208	0.062	0.719	0.123
	0.5	3.2	0.3	0.506	3.692	0.329	0.058	1.718	0.180
	0.7	1.3	0.5	0.720	1.416	0.583	0.110	0.524	0.581
	1.2	0.8	3.0	1.235	0.941	3.761	0.238	0.736	2.865
	1.2	1.5	0.9	1.249	1.896	1.160	0.241	1.395	1.274
300	0.5	1.8	0.2	0.504	1.838	0.208	0.036	0.343	0.067
	0.5	3.2	0.3	0.501	3.357	0.307	0.031	0.751	0.099
	0.7	1.3	0.5	0.708	1.319	0.532	0.056	0.246	0.174
	1.2	0.8	3.0	1.227	0.818	3.377	0.143	0.207	1.538
	1.2	1.5	0.9	1.216	1.591	0.943	0.110	0.461	0.377
500	0.5	1.8	0.2	0.503	1.812	0.208	0.028	0.249	0.051
	0.5	3.2	0.3	0.500	3.253	0.308	0.025	0.560	0.078
	0.7	1.3	0.5	0.705	1.314	0.517	0.044	0.197	0.128
	1.2	0.8	3.0	1.215	0.815	3.197	0.109	0.155	1.123
	1.2	1.5	0.9	1.203	1.566	0.918	0.078	0.344	0.268

set presents more variability than the second one. It can be noted by observing their variances and amplitudes.

In terms of model fitting, we compare the LNH distribution with the widely known Weibull, EW, GPW and NH distributions. We also consider some other generated models under the NH baseline. These models and their corresponding densities are listed bellow (for $x > 0$):

Table 5.2: Descriptive statistics for the guinea pigs and mechanical components.

Statistics	**Real data sets**	
	Guinea pigs	**Mechanical components**
Mean	1.77	0.12
Median	1.50	0.10
Mode	1.25	0.07
Variance	1.07	0.01
Skewness	1.34	3.59
Kurtosis	1.99	12.20
Minimum	0.10	0.07
Maximum	5.55	0.48
n	72.00	20.00

- The EGNH density given by

$$f(x) = a\,b\,\alpha\,\lambda\,\frac{(1+\lambda\,x)^{\alpha-1}\,\left[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}}{\left\{1-\left[\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}}\right]^{a}\right\}^{1-b}},$$

where $a>0$, $b>0$, $\alpha>0$ and $\lambda>0$.

- The GNH density given by

$$f(x) = \frac{\alpha\,\lambda}{\Gamma(a)}\,(1+\lambda\,x)^{\alpha-1}\,[(1+\lambda\,x)^{\alpha}-1]^{a-1}\,\mathrm{e}^{1-(1+\lambda\,x)^{\alpha}},$$

where $a>0$, $\alpha>0$ and $\lambda>0$.

- The GPW density given by

$$f(x) = \alpha\,\lambda\,\gamma\,(1+\lambda\,x^{\gamma})^{\alpha-1}\,\mathrm{e}^{1-(1+\lambda\,x^{\gamma})^{\alpha}},$$

where $\alpha>0$, $\lambda>0$ and $\gamma>0$.

- The EW density given by

$$f(x) = \alpha\,\beta\,\lambda\,x^{\alpha-1}\,\exp\left(-\lambda\,x^{\alpha}\right)\left[1-\exp\left(-\lambda\,x^{\alpha}\right)\right]^{\beta-1},$$

where $\alpha>0$, $\beta>0$ and $\lambda>0$.

- The ENH density given by

$$f(x) = \alpha\,\beta\,\lambda\,\frac{(1+\lambda\,x)^{\alpha-1}\exp\{1-(1+\lambda\,x)^{\alpha}\}}{[1-\exp\{1-(1+\lambda\,x)^{\alpha}\}]^{1-\beta}},$$

where $\alpha>0$, $\beta>0$ and $\lambda>0$.

- The NH density given by (2.2).

- The Weibull density given by

$$f(x) = \alpha\,\lambda\,(\lambda x)^{\alpha-1}\,\mathrm{e}^{-(\lambda x)^{\alpha}},$$

where $\alpha>0$ and $\lambda>0$.

Table 5.3: The MLEs of the model parameters to the guinea pigs data set.

Distributions	Estimates			
EGNH(a, b, α, λ)	0.0093	13.5260	1.9726	4.7183
	(0.0032)	(1.6167)	(0.1442)	(0.8815)
LNH$(\gamma, \alpha, \lambda)$	2.3455	3.9430	0.1227	
	(0.2456)	(1.4014)	(0.0479)	
GNH(a, α, λ)	3.0463	0.9915	1.7350	
	(0.6225)	(0.1704)	(0.8591)	
GPW$(\alpha, \lambda, \gamma)$	2.8845	0.3791	1.0563	
	(0.6568)	(0.1363)	(0.1998)	
EW(α, β, λ)	0.8700	1.1610	2.6060	
	(0.2835)	(0.2467)	(1.2181)	
ENH(α, β, λ)	1.2151	0.7615	3.1451	
	(0.2866)	(0.3381)	(0.7583)	
NH(α, λ)	4.7831	0.0836		
	(1.9613)	(0.0385)		
Weibull(α, λ)	1.9960	1.8280		
	(0.1361)	(0.1588)		

Tables 5.3 and 5.5 give the MLEs (with corresponding standard errors in parentheses) for all the fitted models to the guinea pigs and the mechanical components data sets, respectively. All these results are carried out using the simulated-annealing algorithm for maximizing the log-likelihood function, which is disposable in `AdequacyModel` script in `R` software (Marinho *et al.*, 2016).

Tables 5.4 and 5.6 present the A^*, W^* and KS statistics. These goodness-of-fit measures are helpful for selecting the most appropriate model. Since the values of the A^*, W^* and KS statistics are smaller for the LNH distribution compared with those from the other models, we can conclude that the new distribution fits better than the other models for both data sets.

The histogram and the estimated densities of the three more competitive models according to the goodness-of-fit statistics are plotted in Figure 5.4. Figure 5.5 displays the plots of estimated cdfs for these models. Based on these plots, it is possible to confirm the conclusion that the LNH distribution yields a very competitive alternative to other NH generalizations, such as the GPW distribution in both data sets and ENH model for the guinea pigs data. The new distribution also overcomes other widely known lifetime models, such as the EW model in the second data

Table 5.4: Goodness-of-fit statistics for the models fitted to the guinea pigs data set.

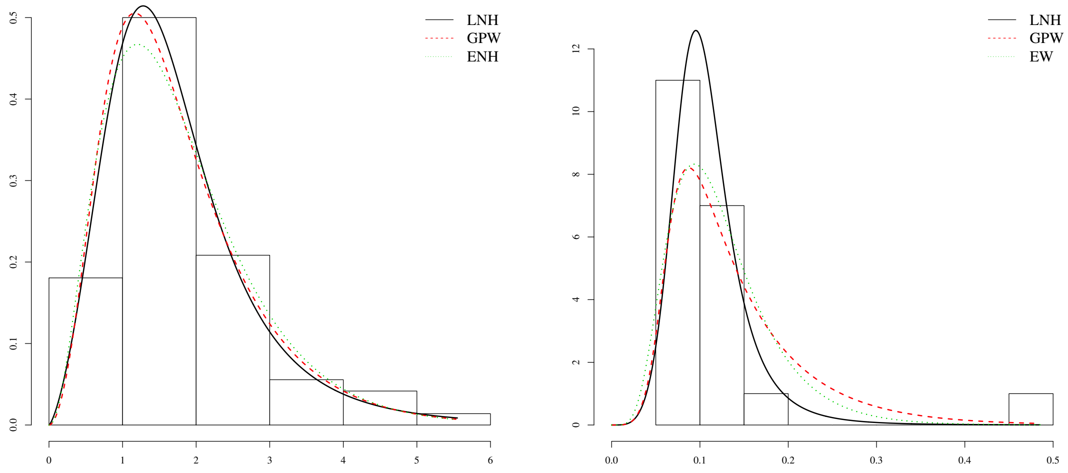
	Statistics		
Distributions	**KS**	**A***	**W***
EGNH	0.0967	0.5701	0.8340
LNH	**0.0646**	**0.3948**	**0.0808**
GNH	0.0965	0.5954	0.0903
GPW	0.0719	0.4587	0.0823
EW	0.0907	0.5628	0.0877
ENH	0.0906	0.5626	0.0853
NH	0.1927	1.1347	0.2354
Weibull	0.1652	0.9727	0.1044

Table 5.5: The MLEs of the model parameters to the mechanical components data set.

Distributions	**Estimates**			
EGNH	0.0696	16.0231	2.2595	15.6816
	(0.0268)	(3.6816)	(0.3236)	(3.7837)
LNH$(\gamma, \alpha, \lambda)$	4.9800	1.0150	9.4760	
	(1.0774)	(0.3102)	(4.0127)	
GNH(a, α, λ)	2.4220	1.4000	10.4650	
	(0.7096)	(0.2990)	(5.1827)	
GPW$(\alpha, \lambda, \gamma)$	5.2867	0.1836	0.06962	
	(1.3534)	(0.0524)	(0.0086)	
EW(α, β, λ)	20.7950	1.0060	7.1765	
	(7.4255)	(0.1603)	(3.0213)	
ENH(α, β, λ)	1.0850	17.4880	6.45220	
	(0.2162)	(8.0323)	(3.2188)	
NH(α, λ)	2.301	2.7750		
	(0.8734)	(1.3102)		
Weibull(α, λ)	0.1372	1.6263		
	(0.0201)	(0.2306)		

Table 5.6: Goodness-of-fit statistics for the models fitted to the mechanical components data set.

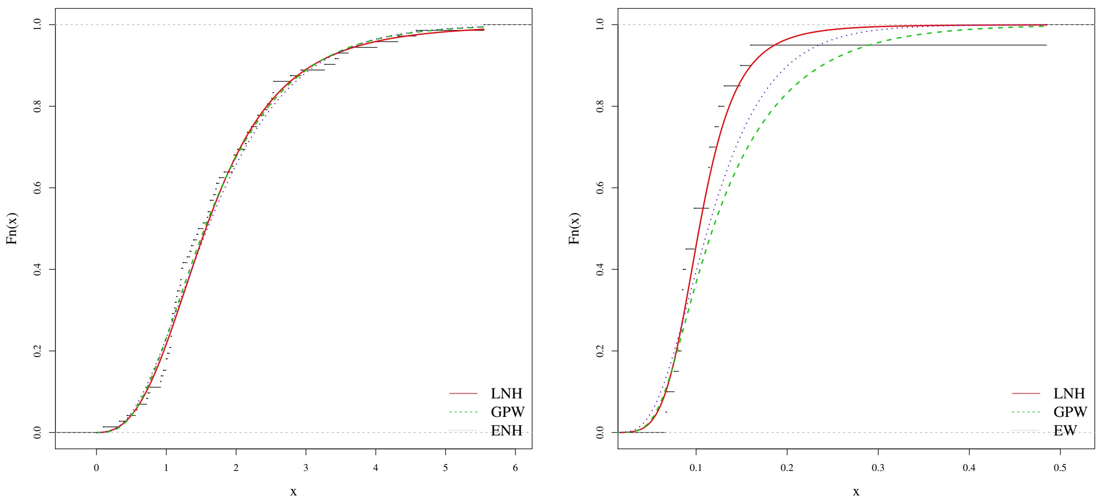
	Statistics		
Distributions	**W***	**A***	**KS**
EGNH(a, b, α, λ)	0.2815	1.8464	0.9482
LNH	**0.0912**	**0.6825**	**0.1291**
GNH	0.3534	2.2297	0.2833
GPW	0.1381	1.0225	0.2736
EW	0.2078	1.4387	0.2242
ENH	0.2221	1.5210	0.2226
NH	0.4078	2.5059	0.3771
Weibull	0.3943	2.4380	0.2677



(a) (b)

Figure 5.4: Histogram and estimated densities of the (a) LNH, GPW and ENH models for the guinea pigs data set; (b) LNH, GPW and EW models for the mechanical components data set.

set.



(a) (b)

Figure 5.5: Estimated and empirical cdfs of the (a) LNH, GPW and ENH models for the guinea pigs data set; (b) LNH, GPW and EW models for the mechanical components data set.

5.10 Concluding remarks

We introduce a Nadarajah-Haghighi generated model, which is part of the *logistic-X* family pioneered by Tahir *et al.* (2016a), called the logistic Nadarajah-Haghighi distribution. The new three-parameter model is more flexible than its baseline distribution because it presents unimodal density shape and the hazard rate function allows constant, increasing, decreasing and upsidedown-bathtub forms. We demonstrate that the proposed density can be rewritten as an infinite linear combination of exponentiated Nadarajah-Haghighi densities and derive structural properties from this relationship. They include the ordinary and incomplete moments and mean deviations. We also present the quantile function for the proposed distribution. The maximum likelihood estimators for the model parameters are obtained and a simulation study is performed. The usefulness of the new distribution is illustrated by means of two applications to real data. In both cases, the new model provides consistently better fits than other generated models under the same baseline distribution and than some other known lifetime distributions. Thus, it may be a useful alternative in practical situations.

Chapter 6

Final conclusion

In this thesis, we consider different generating methods for introducing four new distributions that generalize the Nadarajah-Haghighi (NH) model. They are proposed as alternatives for modeling lifetime data, although they can also be applied in other fields. Some general motivations for introducing these models are i) to overcome a limitation of the NH density, which can only be monotonic decreasing; ii) to obtain distributions that accommodate non-monotonic hazard shapes, since the NH hazard rate function only presents monotonic types and iii) to propose distributions that might provide 'better fits' than some widely known lifetime models and other NH generalizations.

In Chapter 2, we introduce a new four parameter distribution using the concept of exponentiated (exp-G) models. The called *exponentiated generalized power Weibull* distribution is obtained by taking the generalized power Weibull as baseline in the exp-G family. It has nine known lifetime distributions as special models, including the NH distribution, and can also be derived from a power transform in the exponentiated Nadarajah-Haghighi distribution. The motivations for introducing this new model are presented and some mathematical properties are studied. Furthermore, we obtain the maximum likelihood estimators, perform a simulation study and illustrate the flexibility of this model by means of an application to a lifetime data set.

In Chapter 3, we use the continuous-continuous compounding approach to define the Nadarajah-Haghighi Lindley distribution. It is obtained from the minimum between a NH and a Lindley random variables. Some properties of this distribution are derived, the maximum likelihood method is considered to obtain the parameter estimators, a simulation study is carried out and an application is presented.

Chapters 4 and 5 define the Weibull Nadarajah-Haghighi and the logistic Nadarajah-Haghighi distributions, respectively. They are obtained by taking the NH distribution as baseline in the *Weibull-G* and *logistic-X* generated families, respectively. Some structure properties of the both distributions are obtained from the exp-NH properties, such as the moments, incomplete moments, mean deviations and Bonferroni and Lorenz curves. We also derive explicit expressions for the quantile function of the models, estimate their parameters by maximum likelihood method, conduct simulation studies and considere applications to real lifetime data sets.

Due to the NH distribution was recently introduced in the literature, there are several other generated methods that could be developed considering this model. Additionaly, we list some studies to be investigated:

- Alizadeh *et al.* (2017) defined the Odd-Burr generalized family of distributions. We propose the investigation of the Odd-Burr Nadarajah-Haghighi distribution, obtained by inserting the Nadarajah-Haghighi model as baseline in this family.

- Using similar approach as Adamidis and Loukas (1998), another future research line is to introduce the Poisson logistic Nadarajah-Haghighi model by compounding the discrete Poisson distribution withe the logistic Nadarajah-Haghighi distribution presented in Chapter 5. This distribution could be used as a cure rate survival model, being competitive with the Poisson gamma Nadarajah-Haghighi distribution (Ortega *et al.*, 2015), among others.

References

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

Abdel-Hamid, A. and AL-Hussaini, E. (2009). Estimation in step-stress accelerated life tests for the exponentiated exponential distribution with type-I censoring. *Computational Statistics and Data Analysis* **53**, 1328–1338.

Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate. *Statistics and Probability Letters* **39**, 35–42.

Al-Mutairi, D. K., Ghitany, M. E. and Kundu, D. (2015). Inferences on stress-strength reliability from weighted Lindley distributions. *Communications in Statistics - Theory and Methods* **44**, 4096–4113.

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

Alizadeh, M., Cordeiro, G. M., Nascimento, A. D. C., do Carmo S. Lima, M. and Ortega, E. M. M. (2017). Odd-Burr generalized family of distributions with some applications. *Journal of Statistical Computation and Simulation* **87**, 367–389.

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

Alzaghal, A., Famoye, F. and Lee, C. (2013). Exponentiated T-X family of distributions with some applications. *International Journal of Probability and Statistics* **2**, 31–49.

Andrews, D. F. and Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer Series in Statistics, New York.

Asgharzadeh, A., Bakouch, H. S. and Esmaeili, L. (2013). Pareto Poisson–Lindley distribution with applications. *Journal of Applied Statistics* **40**, 1717–1734.

Asgharzadeh, A., Sharafi, F. and Nadarajah, S. (2016). Weibull Lindley distribution. *Preprint* .

Ashour, S. K. and Eltehiwy, M. A. (2015). Exponentiated power Lindley distribution. *Journal of Advanced Research* **6**, 895 – 905.

Bagdonavicius, V. and Nikulin, M. (2002). *Accelerated life models: modeling and statistical analysis*. Chapman and Hall/CRC, Boca Raton.

Bagheri, S., Samani, E. B. and Ganjali, M. (2016). The generalized modified Weibull power series distribution: Theory and applications. *Computational Statistics and Data Analysis* **94**, 136–160.

Bakouch, H. S., Al-Zahrani, B. M., Al-Shomrani, A. A., Marchi, V. A. and Louzada, F. (2012). An extended Lindley distribution. *Journal of the Korean Statistical Society* **41**, 75–85.

Barreto-Souza, W., Morais, A. and Cordeiro, G. (2011). The Weibull-geometric distribution. *Journal of Statistical Computation and Simulation* **81**, 645–657.

Bebbington, M., Lai, C. D. and Zitikis, R. (2007). A flexible Weibull extension. *Reliability Engineering and System Safety* **92**, 719–726.

Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene* **72**, 130–148.

Bourguignon, M., Silva, R. B. and Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science* **12**, 53–68.

Bourguignon, M., Lima, M. d. C. S., Leão, J., Nascimento, A. D. C., Pinho, L. G. B. and Cordeiro, G. M. (2015). A new generalized gamma distribution with applications. *American Journal of Mathematical and Management Sciences* **34**, 309–342.

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

Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters* **49**, 155–161.

Cordeiro, G., Ortega, E. and Lemonte, A. J. (2014a). The exponential–Weibull lifetime distribution. *Journal of Statistical Computation and Simulation* **47**, 626–653.

Cordeiro, G., Ortega, E., Popović, B. and Pescim, R. (2014b). The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation* **247**, 465–486.

Cordeiro, G. M. and Bourguignon, M. (2016). New results on the Ristić-Balakrishnan family of distributions. *Communications in Statistics - Theory and Methods* .

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

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

Dias, C. R. B. (2016). New continuous distributions applied to lifetime data and survival analysis. Doctoral thesis, Universidade Federal de Pernambuco.

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

Ghitany, M., Al-Mutairi, D. and Nadarajah, S. (2008a). Zero-truncated Poisson–Lindley distribution and its application. *Mathematics and Computers in Simulation* **79**, 279 – 287.

Ghitany, M., Atieh, B. and Nadarajah, S. (2008b). Lindley distribution and its application. *Mathematics and Computers in Simulation* **78**, 493–506.

Ghitany, M., Alqallaf, F., Al-Mutairi, D. and Husain, H. (2011). A two-parameter weighted Lindley distribution and its applications to survival data. *Mathematics and Computers in Simulation* **81**, 1190–1201.

Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N. and Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. *Computational Statistics and Data Analysis* **64**, 20–33.

Gomes, O., Combes, C. and Dussauchoy, A. (2008). Parameter estimation of the generalized gamma distribution. *Mathematics and Computers in Simulation* **79**, 963–995.

Graham, R. L. (1994). *Concrete mathematics: a foundation for computer science*. Addison-Wesley, Boston.

Gupta, P. K. and Singh, B. (2013). Parameter estimation of Lindley distribution with hybrid censored data. *International Journal of System Assurance Engineering* **4**, 378–385.

Gupta, R. and Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal* **43**, 117–130.

Gupta, R. and Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference* **137**, 3537–3547.

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

Gupta, R. C., Kannan, N. and Raychaudhuri, A. (1997). Analysis of lognormal survival data. *Mathematical biosciences* **139**, 103–115.

Kenney, J. and Keeping, E. (1962). *Mathematics of Statistics*, 3º edn. Chapman and Hall Ltda, New Jersey.

Klein, J. and Moeschberger, M. (1997). *Survival Analysis Techniques for Censored and Truncated Data*. Springer Verlag, New York.

Krishna, H. and Kumar, K. (2011). Reliability estimation in Lindley distribution with progressively type II right censored sample. *Mathematics and Computers in Simulation* **82**, 281–294.

Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis,* **51**, 4497–4509.

Lai, C. D. (2013). Constructions and applications of lifetime distributions. *Applied Stochastic Models in Business and Industry* **29**, 127–140.

Lee, C., Famoye, F. and A, A. (2013). Overview methods for generating families of univariate continuous distributions in the recent decades. *WIREs Computational Stat* **5**, 219–238.

Lemonte, A., Cordeiro, G. and Ortega, E. (2014). On the additive Weibull distribution. *Communication in Statistics - Theory and Methods* **43**, 2066–2080.

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

Lima, S. R. L. (2015). The half-normal generalized family and Kumaraswamy Nadarajah-Haghighi distribution. Master thesis, Universidade Federal de Pernambuco.

Lindley, D. (1958). Fiducial distributions and bayes' theorem. *Journal of the Royal Statistical Society, A* **20**, 102–107.

MacGillivray, H. (1986). Skewness and asymmetry: measures and orderings. *Annals of Statistics 14, 994–1011* .

Marinho, P. R. D. (2016). Some new families of continuous distributions. Doctoral thesis, Universidade Federal de Pernambuco.

Marinho, P. R. D., Bourguignon, M. and Dias, C. R. B. (2016). *AdequacyModel: Adequacy of Probabilistic Models and General Purpose Optimization.* URL https://CRAN.R-project.org/package=AdequacyModel, r package version 2.0.0.

Mazucheli, J. and Achcar, J. (2011). The Lindley distribution applied to competing risks lifetime data. *Computer Methods and Programs in Biomedicine* **104**, 188–192.

Moors, J. (1988). A quantile alternative for kurtosis. *The statistician* pp. 25–32.

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

Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated Weibull family: some properties and a flood data application. *Communications in Statistics-Theory and Methods* **25**, 3059–3083.

Murthy, D., Xie, M. and R, J. (2004). *Weibull models*. Wiley, New York.

Nadarajah, S. (2011). The exponentiated exponential distribution: a survey. *AStA Advances in Statistical Analysis* **95**, 219–251.

Nadarajah, S. and Bakouch, H. S. (2011). A generalized Lindley distribution. *Sankhya B* **73**, 331–359.

Nadarajah, S. and Gupta, A. K. (2005). On the moments of the exponentiated Weibull distribution. *Communications in Statistics - Theory and Methods* **34**, 253–256.

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

Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematica* **92**, 97–111.

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

Nassar, M. M. and Eissa, F. H. (2003). On the exponentiated Weibull distribution. *Communications in Statistics - Theory and Methods* **32**, 1317–1336.

Nikulin, M. and Haghighi, F. (2006). A chi-squared test for the generalized power Weibull family for the head-and-neck cancer censored data. *Journal of Mathematical Sciences* **133**, 1333 – 1341.

Nikulin, M. and Haghighi, F. (2009). On the power generalized Weibull family. *Metron* **67**, 75–86.

Oguntunde, P., Balogun, O., Okagbue, H. and Bishop, S. (2015). The Weibull-exponential distribution: Its properties and applications. *Journal of Applied Sciences* **15**, 1305–1311.

Ortega, E. M., Lemonte, A. J., Silva, G. O. and Cordeiro, G. M. (2015). New flexible models generated by gamma random variables for lifetime modeling. *Journal of Applied Statistics* **42**, 2159–2179, http://dx.doi.org/10.1080/02664763.2015.1021669.

Popovíc, B., Ristíc, M. and Cordeiro, G. (2015). A two parameter distribution obtained by compounding the generalized exponential and exponential distributions. *Mediterranean Journal of Mathematics* **15**, 2935–2949.

Press, W. H., Teukolsky, A. A., Vetterling, W. T. and Flanner Y, B. P. (2007). *Numerical Recipes in C: The Art of Scientific Computing*, 3º edn. Cambridge University Press, New York.

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

Renyi, A. (1961). On measures of entropy and information. Em: *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561.

Ristić, M. and Balakrishnan, N. (2012). The gamma exponentiated exponential distribution. *Journal of Statatistics Computation and Simulation* **82**, 1191–1206.

Sankaran, M. (1970). The discrete Poisson–Lindley distribution. *Biometrics* **26**, 145–149.

Sarhan, A., Ahmed, A. B. and Alasbahi, I. (2013). Exponentiated generalized linear exponential distribution. *Appl Math Model* **37**, 2838–2849.

Tahir, M. H. and Nadarajah, S. (2015). Parameter induction in continuous univariate distributions: Well-established G families. *Annals of the Brazilian Academy of Sciences* **87**, 539–568.

Tahir, M. H., Cordeiro, G. M., Mansoor, M. and Zubair, M. (2015). The Weibull-Lomax distribution: properties and applications. *Hacettepe Journal of Mathematics and Statistics* **44**, 455–474.

Tahir, M. H., Cordeiro, G. M., Alzaatreh, A., Mansoor, M. and Zubair, M. (2016a). The logistic-X family of distributions and its applications. *Communications in Statistics - Theory and Methods* **45**, 7326–7349.

Tahir, M. H., Cordeiro, G. M., Mansoor, M., Zubair, M. and Alizadeh, M. (2016b). The Weibull-Dagum distribution: Properties and applications. *Communications in Statistics - Theory and Methods* **45**, 3548–3567.

Tahir, M. H., Zubair, M., Mansoor, M., Cordeiro, G. M., Alizadeh, M. and Hamedani, G. G. (2016c). A new Weibull-G family of distributions. *Hacettepe Journal of Mathematics and Statistics* **45**, 629–647.

VedoVatto, T., Nascimento, A. D. C., Miranda Filho, W. R., Lima, M. C. S., Pinho, L. G. and Cordeiro, G. M. (2016). Some computational and theoretical aspects of the exponentiated generalized Nadarajah-Haghighi distribution. *arxivorg/abs/161008876v1* .

Xie, M. and Lai, C. (1995). Reliability analysis using an additive weibull model with bathtub shaped failure rate function. *Reliability Engineering & System Safety* **52**, 87–93.

Xie, M., Tang, Y. and Goh, T. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and System Safety* **76**, 279–285.

Zamani, H. and Ismail, N. (2010). Negative binomial-Lindley distribution and its application. *Journal of Mathematics and Statistics* **6**, 4–9.

Zheng, G. (2002). On the Fisher information matrix in type II censored data from the exponentiated exponential family. *Biometrical Journal* **44**, 353–357.

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