Federal University Of Pernambuco
Center of Natural and Exact Sciences
Department of Statistics

Postgraduate Program in Statistics

THE HALF-NORMAL GENERALIZED FAMILY AND KUMARASWAMY NADARAJAH-HAGHIGHI DISTRIBUTION

Stênio Rodrigues Lima

Masters dissertation

Recife-PE
2015

Federal University Of Pernambuco
Center of Natural and Exact Sciences
Department of Statistics

Stênio Rodrigues Lima

THE HALF-NORMAL GENERALIZED FAMILY AND KUMARASWAMY NADARAJAH-HAGHIGHI DISTRIBUTION

Advisor: Teacher Dr. Gauss Moutinho Cordeiro

Masters dissertation submitted to the Post Graduate Program in Statistics, Department of Statistics, Federal University of Pernambuco as a partial requirement for obtaining a Master in Statistics.

Recife-PE
2015

Catalogação na fonte
Bibliotecária Jane Souto Maior, CRB4-571

L732h Lima, Stênio Rodrigues
The half-normal generalized family and kumaraswamy nadarajah-haghighi distribution / Stênio Rodrigues Lima – Recife: O Autor, 2015.
53 f.: il., fig., tab.

Orientador: Gauss Moutinho Cordeiro.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, 2015.
Inclui referências.

1. Probabilidade. 2. Distribuição de probabilidade. I. Cordeiro, Gauss Moutinho (orientador). II. Título.

519.2 CDD (23. ed.) UFPE- MEI 2016-004

STÊNIO RODRIGUES LIMA

THE HALF-NORMAL GENERALIZED FAMILY AND KUMARASWAMY NADARAJAH-HAGHIGHI DISTRIBUTION

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

Aprovada em: 03 de dezembro de 2015.

BANCA EXAMINADORA

Prof. PhD. Gauss Moutinho Cordeiro
UFPE

Prof. Dr. Abraão David Costa do Nascimento (Examinador Interno)
UFPE

Prof. Dr. Cláudio Tadeu Cristino (Examinador Externo)
UFRPE

"A luta enriquece-o de experiência, a dor aprimora-lhe as emoções e o sacrifício tempera-lhe o caráter. O Espírito encarnado sofre constantes transformações por fora, a fim de acrisolar-se e engrandecer-se por dentro."

Chico Xavier

Agradecimentos

Agradeço ao senhor meu Deus, por me fortalecer nessa caminhada e pela graça conquistada.

Agradeço a minha família que me ajudou na formação como pessoa. Meu pai, minha mãe e meu irmão que sempre estiveram comigo em todos os momentos (bons e ruins) por qual passei, e meu espelho de crescimento e conquista (meu grande irmão, Reneé). Estes que sempre me encheram de orgulho e isentivos, auxiliando-me e apoiando-me em todas atitudes e decisões que a vida nos ensinado.

Ao meu orientador Dr. Gauss Cordeiro pelo incentivo e confiança no meu trabalho e pelo vasto conhecimento que tem compartilhado.

Aos professores com os quais tive a oportunidade de conviver, entre eles: Valmária Rocha e Fernando Ferraz, que contribuíram no meu crescimento acadêmico e profissional, Audrey Cysneiros, Francisco Cribari, Leandro Chaves, dentre outro.

Agradeço aos meus colegas de turma que, direta ou indiretamente contribuíram nessa caminhada, dentre eles: Pedro Almeida, Francisco Hildemar, Rodrigo Alves, Telma de Sousa, Wanesa Weridiana, Laura Vicunã, Teresinha Késsia, Raphaela Belchior, Jonas Weverson, Diego Coelho, Jessica Rivas, dentre outros. Pela amizade e companheirismo durante esses anos de pós-graduação.

À CAPES, pelo apoio financeiro.

Resumo

As distribuições generalizadas têm sido amplamente estudadas na Estatística e diversos autores têm investigado novas distribuições de sobrevivência devido à sua flexibilidade para ajustar dados. Neste trabalho um novo método de compor distribuições é proposto: a família Half-Normal-G, em que G é chamada distribuição baseline. Demostramos que as funções densidades das distribuições propostas podem ser expressas como combinação linear de funções densidades das respectivas exponencializadas-G. Diversas propriedades dessa família são estudadas. Apresentamos também uma nova distribuição de probabilidade baseado na Família de Distribuições Generalizadas Kumaraswamy (kw-G), já conhecida na literatura. Escolhemos como baseline a distribuição Nadarajah-Haghighi, recentemente estudada por Nadarajah e Haghighi (2011) e que desenvolveram algumas propriedades interessantes. Estudamos várias propriedades da nova distribuição Kumaraswamu-Nadarajah-Haghighi (Kw-NH) e fizemos duas aplicações de bancos de dados mostrando empiricamente a flexibilidade do modelo.

Palavras-chave: Distribuição generalizadas. Distribuição Half-Normal. Distribuição Kumaraswamy. Distribuição Nadarajah-Haghighi. Estimativa de máxima verossimilhança. Função taxa de falha. Momentos. Tempo de vida.

Abstract

Generalized distributions have been widely studied in the Statistics, and several authors have investigated new distributions of survival because of their flexibility to adjust data. In this work, a new method of composing distributions is proposed: a Half-Normal-G family, where G is called the distribution baseline. We demonstrate that the new density functions can be expressed as a linear combination of exponentiated-G ("EG", for short) density functions. Several properties of this family are studied. We also present a new probability distribution based on the family of Kumaraswamy generalized distributions ("kw-G", for short), which known in the literature. Chosen as the baseline the Nadarajah-Haghighi distribution, recently proposed by Nadarajah and Haghighi (2011) that developed some interesting properties. We study various properties of the new distribution, Kumaraswamy-Nadarajah-Haghighi (Kw-NH), and make two applications of databases, empirically showing the flexibility of the model.

Keywords: Generalized distributions. Half-Normal distribution. Hazard function. kumaraswamy distribution. Lifetime. Maximum likelihood estimation. Moments. Nadarajah-Haghighi distribution.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Generalized distributions have been widely studied in Statistics and several authors have investigated new survival distributions because of their flexibility to fit data. For example, Lemonte (2013) proposed the exponentiated Nadarajah-Haghighi distribution by adding an extra parameter form in Nadarajah-Haghighi distribution. This construction method can be seen also in Mudholkar et al. (1995), Nadarajah and Kotz (2004), and Nadarajah and Gupta (2007), among others. Bourguignon et al. (2013) introduced the Kumaraswamy-Pareto distribution (Kw-P), based on a composition of Kumaraswamy distribution and Pareto distribution. The Kw-P distribution generalizes the Pareto distribution and can be more flexible. Many authors have generalized other distributions similarly to Kw-P. The kw-Weibull by Cordeiro et al. (2010),Kw-generalized gamma by de Pascoa et al. (2011), Kw-generalized half-normal by Cordeiro et al. (2012a), kw-exponentiated Pareto by Elbatal (2013), Kw-Gumbel by Cordeiro et al. (2012b), Kw-Birnbaum-Saunders by Saulo et al. (2012), Kw-normal by Correa et al. (2012), Kw-BurrXII by Paranaiba et al. (2013), Kw-Lomax by Shams (2013), and Kw-generalized Rayleigh by Gomes et al. (2014) distributions are some examples obtained by taking $G(x)$ to be the cdf of the Weibull, generalized gamma, generalized half-normal, exponentiated pareto, Gumbel, Birnbaum-Saunders, normal, Burr XII, Lomax and generalized Rayleigh distributions, respectively, among several others.

In this dissertation we study two families of probability distributions, we introduce a new family based on the generalized half-normal distribution, and we extended a family already known in the literature, generalized Kumaraswamy family. In the second chapter, the half-normal family of continuous distribution is introduced with additional parameters to generalize any continuous baseline distribution. We derive some mathematical properties. Finally, we performed two applications to real data. In the third chapter, we present a new continuous distribution based on Kumaraswamy family composed with the Nadarajah-Haghighi distribution. We study some mathematical properties, and maximum likelihood techniques are used to adjust the model and to show its potential. We show that the proposed probability density function can be expressed as a linear combination of the density function Nadarajah-Haghighi exponentiated distribution.

Chapter 2

The Half-Normal Generalized Family of Distributions

Resumo

Neste capítulo, propomos a família half-normal de distribuições com um parâmetro positivo adicional para generalizar qualquer distribuição *baseline* contínua. Apresentamos quatro modelos especiais: distribuições half-normal-Weibull, half-normal-Pareto, half-normal-Gumbel e half-normal-log-logistic. Derivamos algumas propriedades matemáticas da nova família: momentos ordinários e incompletos, função geradora, função quantil e desvios médios. Discutimos a estimação dos parâmetros do modelo por máxima verossimilhança. Duas aplicações a dados reais mostram que a nova família pode fornecer melhor ajuste do que outros modelos de vida importantes.

Palavras-chave: Distribuição generalizadas; distribuição Half-Normal; tempo de vida.

Abstract

In this chapter, we proposed the half-normal family of distributions with an extra positive parameter to generalize any continuous baseline distribution. Four special models, the half-normal-Weibull, half-normal-Pareto, half-normal-Gumbel and half-normal-log-logistic distributions are presented. Some mathematical properties of the new family such as ordinary and incomplete moments, quantile and generating functions, and mean deviations are investigated. We discuss the estimation of the model parameters by maximum likelihood. Two applications the real data show that the new family can provides better fits than other important lifetime models.

Key words: Generalized distribution. Half-Normal distribution. Lifetime.

2.1 Introdution

Numerous classical distributions have been extensively used over the past decades for modelling data in several areas such as engineering, actuarial, environmental and medical sciences, biological studies, demography, economics, finance, and insurance. However, in many applied areas, there is a clear need for extended forms of these distributions. For that reason, several methods for generating new families of distributions have been studied recently.

There has been an increased interest in defining new generators for univariate continuous families by introducing one or more additional shape parameter(s) into a baseline distribution. The well-known generators are the following: beta-G by Eugene et al. (2002), Kumaraswamy-G (Kw-G) by Cordeiro and de Castro (2011), McDonald-G (Mc-G) by Alexander et al. (2012), gamma-G (type 1) by Zografos and Balakrishnan (2009), gamma-G (type 2) by Ristič and Balakrishanan (2012), gamma-G (type 3) by Torabi and Montazari (2012), log-gamma-G by Amini et al. (2012), logistic-G by Torabi and Montazari (2013), exponentiated generalized-G by Cordeiro et al. (2013), Weibull-G by Bourguignon et al. (2014), and exponentiated half-logistic-G by Cordeiro et al. (2014).

The half-normal (HN) distribution is a special case of the folded normal distribution. Let Z be an ordinary normal distribution, $N(0, a^2)$, then $Y = |Z|$ has the HN distribution. Thus, this distribution is a fold at the mean of an ordinary normal distribution with zero mean. A previous study by Bland and Altman (1999) used the HN distribution to study the relationship between measurement error and magnitude. Bland (2005) extended their work by using this distribution to estimate the standard deviation as a function so that measurement error could be controlled. In his work, various exercise tests were analyzed and it was determined that the variability of performance does decline with practice. Manufacturing industries have used the HN distribution to model lifetime processes under fatigue. These industries often produce goods with a long lifetime need for customers, making the cost of the resources needed to analyze the product failure times very high. To save time and money, the HN distribution is used in this reliability analysis to study the probabilistic aspects of the product failure times de Castro et al. (2012).

Various generalizations of the HN distribution have been derived. These extensions include the generalized half-normal (GHN) (Cooray et al., 2008) beta-generalized half-normal (Pescrim et al., 2010), and Kumaraswamy generalized half-normal (de Castro et al., 2012) distributions. Several of the corresponding applications include the stress-rupture life of kevlar 49/epoxy strands placed under sustained pressure (Cooray et al., 2008) and failure times of mechanical components and flood data (de Castro et al., 2012).

Let $G(x)$ be the cumulative distribution function (cdf) of any random variable X and $r(t)$ be the probability density function (pdf) of a random variable T defined on $[0, \infty)$. The cdf of the T-X family of distributions, Alzaatreh et al. (2013), is given by

$$F(x) = \int_0^{\frac{G(x)}{\overline{G}(x)}} r(t)\, dt, \tag{2.1}$$

where $\bar{G}(x) = 1 - G(x)$. Differentiating (2.1) in relation the X gives

$$f(x) = \left(\frac{d}{dx}\frac{G(x)}{\overline{G}(x)}\right) r\left(\frac{G(x)}{\overline{G}(x)}\right). \tag{2.2}$$

If a random variable T has the HN distribution with parameter $a > 0$, its pdf is given by

$$r(t;a) = \frac{\sqrt{2}}{a\sqrt{\pi}}\exp\left(-\frac{t^2}{2a^2}\right), \quad t > 0, \tag{2.3}$$

and its cdf becomes

$$R(t;a) = \frac{2}{\sqrt{\pi}}\int_0^{t/(\sqrt{2}a)} \mathrm{e}^{-\mathrm{z}^2}\mathrm{dz} = \mathrm{erf}\left(\frac{\mathrm{t}}{\sqrt{2}\mathrm{a}}\right), \tag{2.4}$$

where $\mathrm{erf(x)} = (2/\sqrt{\pi})\int_0^{\mathrm{x}} \mathrm{e}^{-\mathrm{z}^2}\mathrm{dz}$ is the error functions. $\gamma(a,x)$ is denoted the incomplet gamma function.

The error function (also called the Gauss error function) is a function related to the normal distribution and gamma distribution. This function has as property, for $\nu > 0$,

- $erf(0) = 0$;
- $erf(\pm\infty) = \pm 1$;
- $erf(-\nu) = -erf(\nu)$. Is a impair function;
- $erf(\nu) = \frac{1}{\sqrt{\pi}}\gamma\left(\frac{1}{2},\nu^2\right)$. $\gamma()$ the incomplet gamma function.

The aim of this dissertation is to study a new family of continuous distributions, called the *half-normal*-G ("HNG" for short), where $r(t)$ in (2.1) is given by (2.3). Its generated cdf and pdf are

$$F(x) = \int_0^{\frac{G(x)}{\overline{G}(x)}} \frac{\sqrt{2}}{a\sqrt{\pi}}\exp\left(-\frac{t^2}{2a^2}\right)dt = \mathrm{erf}\left\{\frac{\mathrm{G(x)}}{\sqrt{2}\mathrm{a}\,\overline{\mathrm{G}}(\mathrm{x})}\right\} \tag{2.5}$$

and

$$f(x) = \sqrt{\frac{2}{\pi}}\frac{g(x)}{a\,\overline{G}(x)^2}\,\mathrm{e}^{\frac{-\mathrm{G(x)}^2}{2\mathrm{a}^2\,\overline{\mathrm{G}}(\mathrm{x})^2}}, \tag{2.6}$$

respectively, where $g(x) = dG(x)/dx$. The HNG family has the same parameters of the G distribution plus one additional scale parameter $a > 0$. The pdf (2.6) will be most tractable when $G(x)$ and $g(x)$ have simple analytic expressions. Henceforth, a random variable with density (2.6) is denoted by $X \sim$ HNG(a).

The hazard rate function (rhf) of X is

$$h(x) = \sqrt{\frac{2}{\pi}}\frac{g(x)\,\mathrm{e}^{\frac{-\mathrm{G(x)}^2}{2\mathrm{a}^2\overline{\mathrm{G}}(\mathrm{x})^2}}}{a\,\overline{G}(x)^2\left[1 - \mathrm{erf}\left(\frac{\mathrm{G(x)}}{\sqrt{2}\,\mathrm{a}\,\overline{\mathrm{G}}(\mathrm{x})}\right)\right]}. \tag{2.7}$$

2.2 Special HNG distributions

The HNG density function (2.6) allows for greater flexibility and can be applied in many areas of engineering and biology. Here, we present and study some special cases of this family because it extends several widely known distributions in the literature. In the following examples, a is the HN generator parameter.

2.2.1 Half-normal-Weibull (HNW) distribution

The HNW distribution is defined from (2.6) by taking $G(x) = 1 - \exp[-(\beta\, x)^{\alpha}]$ and $g(x) = \alpha\, \beta\, x^{\alpha-1} \exp[-(\beta\, x)^{\alpha}]$ to be the cdf and pdf of the Weibull(α, β) distribution. The HNW density fuction and hazard rate function are given, respectively, by

$$f_{HNW}(x) = \frac{\sqrt{2}\alpha\beta^{\alpha}x^{\alpha-1}}{\sqrt{\pi}a\exp[-(\beta x)^{\alpha}]}\ \mathrm{e}^{\frac{-\{1-\exp[-(\beta x)^{\alpha}]\}^2}{2a^2\{\exp[-(\beta x)^{\alpha}]\}^2}}. \tag{2.8}$$

and

$$\begin{aligned} h(x) \quad &= \quad \frac{\sqrt{2}\,\alpha\,\beta^{\alpha\,x^{\alpha-1}}\exp[-(\beta\,x)^{\alpha}]}{\sqrt{\pi}\,a\,\exp[-(\beta\,x)^{2\,\alpha}]}\exp\left(\frac{\{\exp[-(\beta\,x)^{\alpha}]-1\}^2}{2\,a^2\,\exp[-(\beta\,x)^{2\alpha}]}\right) \\ &\times \quad \left[1-\mathrm{erf}\left(\frac{1-\exp[-(\beta\,x)^{\alpha}]}{\sqrt{2}\,a\,\exp[-(\beta\,x)^{\alpha}]}\right)\right]^{-1} \end{aligned} \tag{2.9}$$

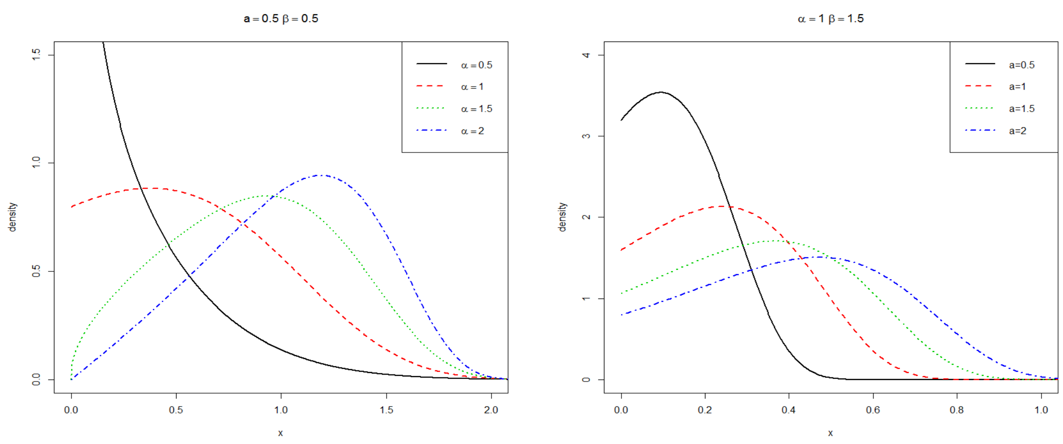


Figure 2.1: The HNW density function for some parameter values

2.2.2 Half-normal-Pareto (HNPa) distribution

Let $G(x)$ be the Pareto cdf with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, say $G(x) = 1 - (x/\beta)^{-\alpha}$. The HNPa density fuction and hazard rate function are given, respectively, by

$$f_{HNPa}(x) = \frac{\sqrt{2}\alpha x^{\alpha-1}}{\sqrt{\pi}a\beta^{\alpha}}\ \mathrm{e}^{\frac{(x^{-\alpha}-\beta^{-\alpha})^2}{2a^2x^{-2\alpha}}}. \tag{2.10}$$

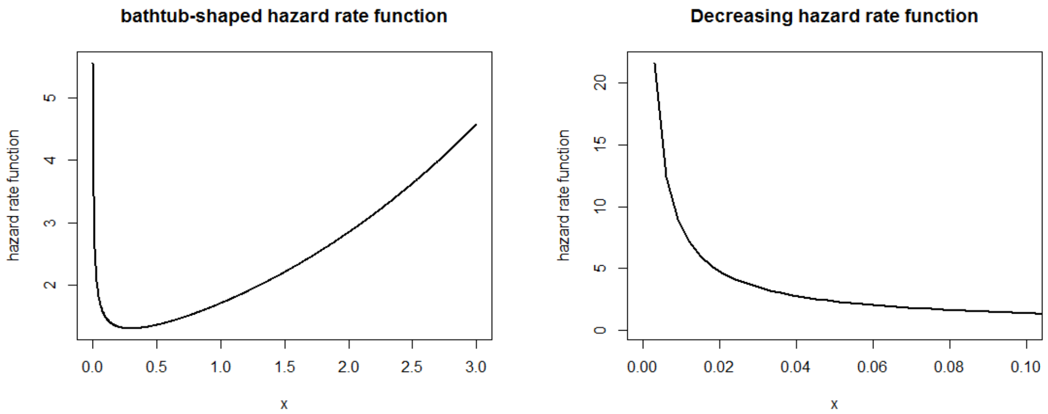


Figure 2.2: The HNW hrf for some parameter values

and

$$h(x) = \frac{\sqrt{2}\,\alpha,\,\beta^{\alpha}}{\sqrt{\pi}\,a\,x^{\alpha+1}}\,\exp\left\{\frac{-[(\beta/x)^{\alpha}-1]^2}{2\,a^2\,(\beta/x)^{2\alpha}}\right\} \times \left[1-\text{erf}\left(\frac{1-(\beta/x)^{\alpha}}{\sqrt{2}\,a\,(\beta/x)^{\alpha}}\right)\right]^{-1} \qquad (2.11)$$

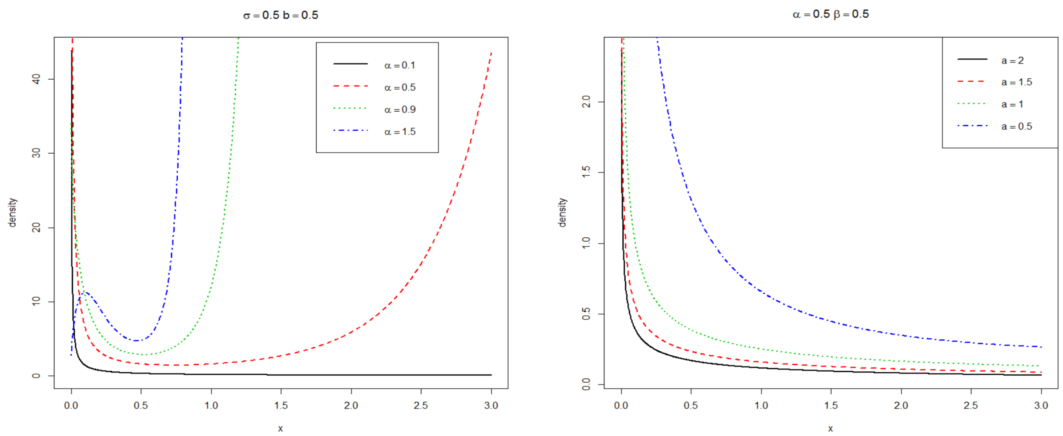


Figure 2.3: The HNPa density function for some parameter values

2.2.3 Half-normal-Gumbel (HNGu) distribution

Consider the Gumbel distribution with location parameter $\mu \in \mathrm{R}$ and scale parameter $\sigma > 0$, where the pdf and cdf (for $x \in \mathrm{R}$) are

$$g(x) = \frac{1}{\sigma}\exp\left\{\left(\frac{\mathrm{x}-\mu}{\sigma}\right) - \exp\left(\frac{\mathrm{x}-\mu}{\sigma}\right)\right\} \quad \text{and} \quad G(x) = \exp\left\{-\exp\left(\frac{\mathrm{x}-\mu}{\sigma}\right)\right\},$$

respectively. Inserting these expressions into (2.6) and (2.7) gives the HNGu density function and hrf, respectively

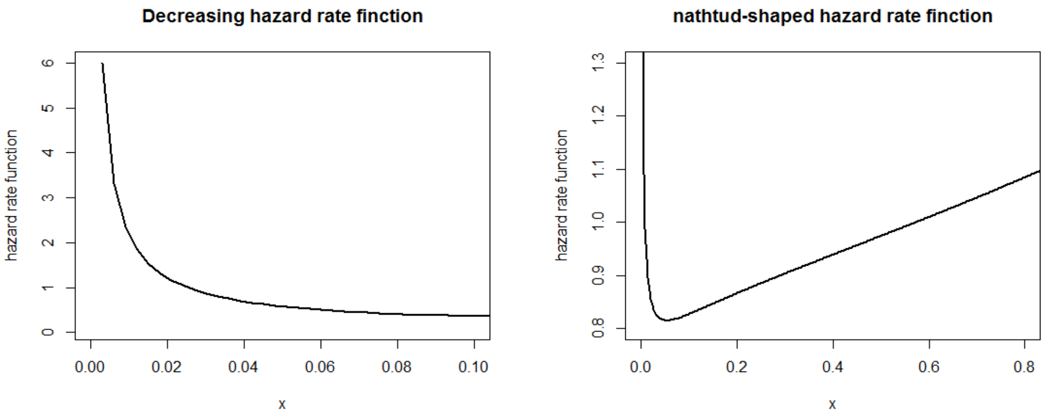


Figure 2.4: The HNPa hrf for some parameter values

$$f_{HNGu}(x) = \frac{\sqrt{2}\mathrm{e}^{\left(\frac{\mathrm{x}-\mu}{\sigma}\right)}}{\sqrt{\pi}a\sigma \mathrm{e}^{-\mathrm{e}^{\left(\frac{\mathrm{x}-\mu}{\sigma}\right)}}} \exp\left\{ \frac{\mathrm{e}^{-\mathrm{e}^{\left(\frac{\mathrm{x}-\mu}{\sigma}\right)}} - 1}{2\mathrm{a}^2\left[\mathrm{e}^{-\mathrm{e}^{\left(\frac{\mathrm{x}-\mu}{\sigma}\right)}}\right]^2} \right\}. \tag{2.12}$$

and

$$\begin{aligned} h(x) \quad &= \quad \frac{\sqrt{2}\exp[(\frac{(v-\mu)}{\sigma}) - \exp(\frac{(x-\mu)}{\sigma})]}{\sqrt{\pi}\,a\,\sigma\,(1-\exp\{-\exp[-\frac{(x-\mu)}{\sigma}]\})^2} \exp\left[\frac{-(\exp\{-\exp[-\frac{(x-\mu)}{\sigma}]\})^2}{2\,a^2(1-\exp\{-\exp[-\frac{(x-\mu)}{\sigma}]\})^2}\right] \\ &\times \left[1-\mathrm{erf}\left(\frac{\exp\{-\exp[-\frac{(x-\mu)}{\sigma}]\}}{\sqrt{2}\,a\,(1-\exp\{-\exp[-\frac{(x-\mu)}{\sigma}]\})}\right)\right]^{-1} \end{aligned} \tag{2.13}$$

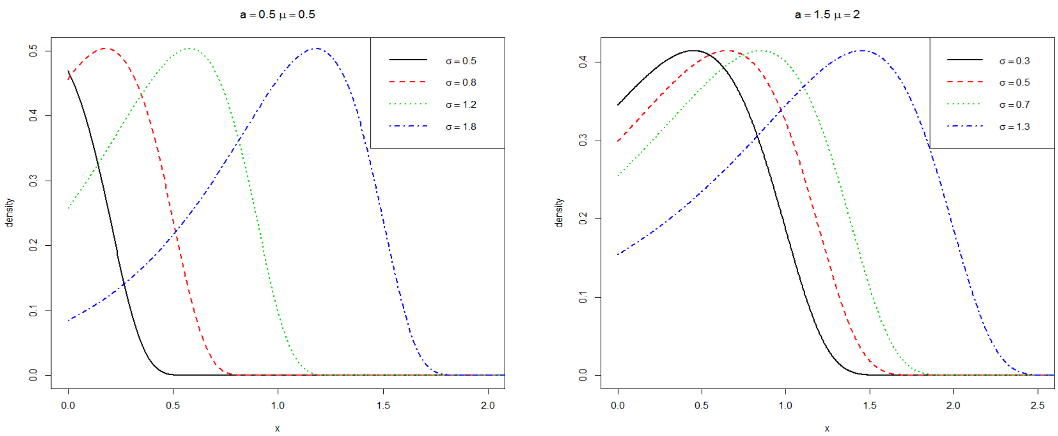


Figure 2.5: The HNGu density function for some parameter values

2.2.4 Half-normal-log-logistic (HNLL) distribution

The pdf and cdf of the log-logistc (LL) distribution are (for $x, \alpha, \beta > 0$)

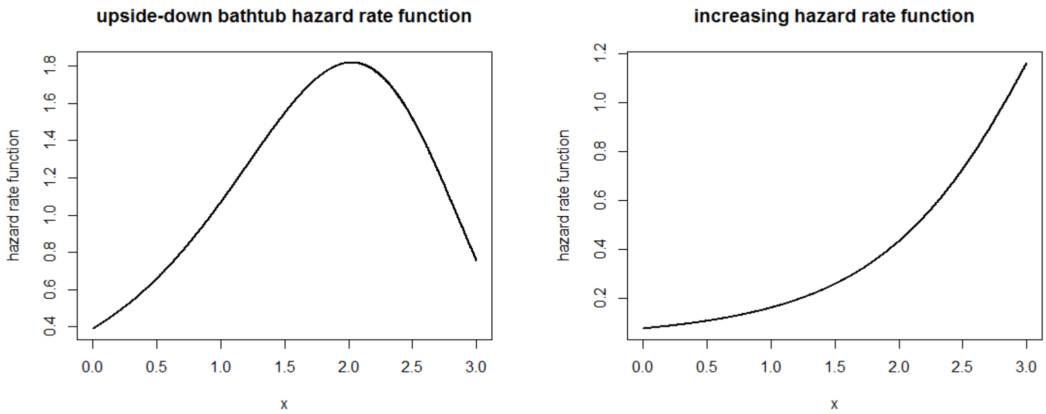


Figure 2.6: The HNGu hrf for some parameter values

$$g(x) = \frac{\beta}{\alpha^{\beta}} x^{\beta-1} \left[1 + \left(\frac{x}{\alpha}\right)\right]^{-2} \quad \text{and} \quad G(x) = 1 - \left[1 + \left(\frac{x}{\alpha}\right)^{\beta}\right]^{-1},$$

respectively. Inserting these expressions into (2.6) and (2.7) gives the HNLL density function and hrf (for $x > 0$), respectively

$$f_{HNLL}(x) = \frac{\sqrt{2}\beta x^{\beta-1} \left[1 + \left(\frac{x}{\alpha}\right)\right]^{-2}}{\sqrt{\pi} a \alpha^{\beta}} \exp \left\{ \frac{\left[\left(1 + \left(\frac{x}{\alpha}\right)^{\beta}\right)^{-1} - 1\right]^{2}}{2a^{2} \left[1 + \left(\frac{x}{\alpha}\right)^{\beta}\right]^{-2}} \right\}. \tag{2.14}$$

and

$$h(x) = \frac{\sqrt{2}\,\beta, x^{\beta+1}}{2\sqrt{\pi}\,a^{3}, \alpha^{\beta+2}} \text{erf} \left[\frac{x}{\sqrt{2}\,a\,\alpha}\right]^{-1} \tag{2.15}$$

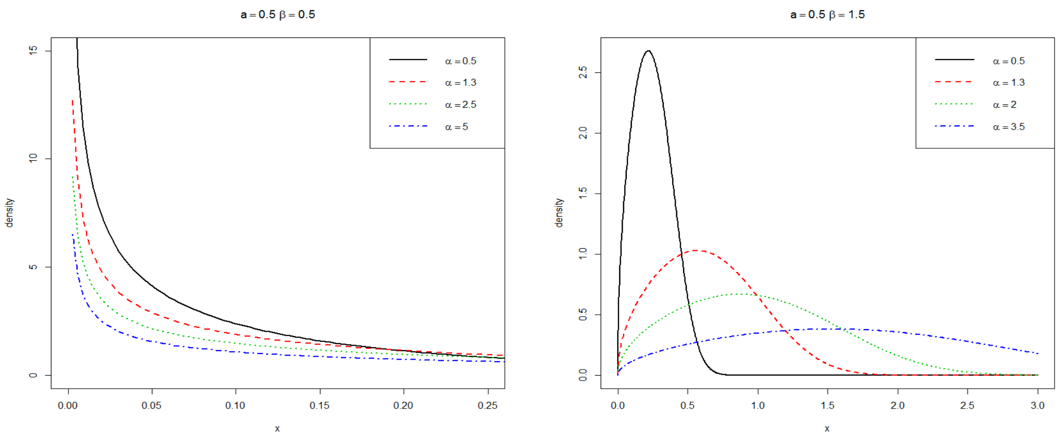


Figure 2.7: The HNLL density function for some parameter values

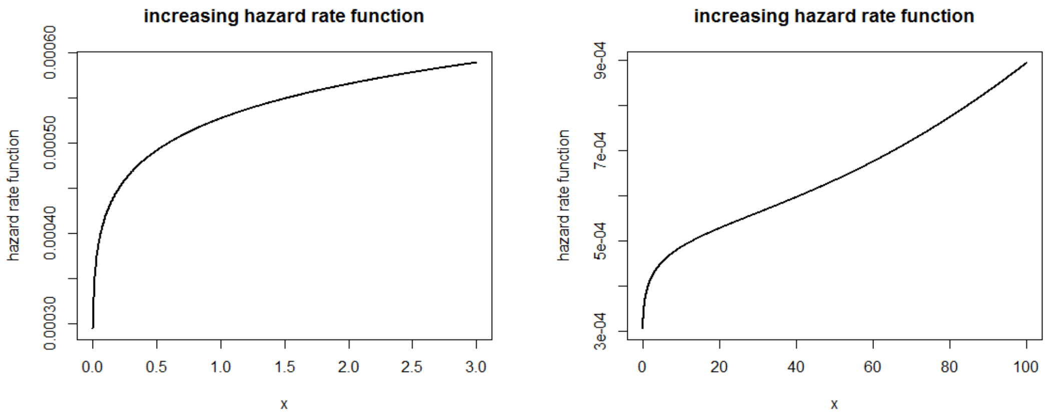


Figure 2.8: The HNLL hrf for some parameter values

2.3 Useful expansions

Some useful expansions for (2.5) and (2.6) can be derived using the concept of exponentiated distributions. For an arbitrary baseline cdf $G(x)$, a random variable is said to have the exponentiated-G (exp-G) distribution with parameter $a > 0$, say $Y_a \sim$ exp-$G(a)$, if its pdf and cdf are given by $h_a(x) = aG^{a-1}(x)g(x)$ and $H_a(x) = G^a(x)$, respectively. The properties of exponentiated distributions have been studied by many authors in recent years. See Mudholkar and Srivastava (1993) for exponentiated Weibull, Gupta et al. (1998) for exponentiated Pareto, Nadarajah (2005) for exponentiated Gumbel, and Nadarajah and Gupta (2007) for exponentiated gamma. By using the power series for the exponential function, we obtain

$$\exp\left\{-\frac{\mathrm{G(x)}^2}{2\sigma^2[\overline{\mathrm{G}}(\mathrm{x})]^2}\right\} = \sum_{\mathrm{k}=0}^{\infty} \frac{(-1)^{\mathrm{k}}}{2^{\mathrm{k}}\mathrm{k}!\sigma^{2\mathrm{k}}} \frac{\mathrm{G(x)}^{2\mathrm{k}}}{\overline{\mathrm{G}}(\mathrm{x})^{2\mathrm{k}}}.$$

Inserting this expansion in equation (2.6) gives

$$f(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}}\, g(x) \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!\sigma^{2k}} \frac{G(x)^{2k}}{[\overline{G}(x)]^{2k+2}}. \tag{2.16}$$

Now, using the generalized binomial theorem, $\overline{G}(x)^{2k+2}$ can be expressed as

$$\overline{G}(x)^{-(2k+2)} = \sum_{n=0}^{\infty}(-1)^n \binom{-2-2k}{n} G(x)^n \tag{2.17}$$

where $\binom{-2-2k}{n} = \frac{\Gamma(-2k-1)}{\Gamma(n+1)\Gamma(-2k-n-1)}$, and $\Gamma(x)$ is the gamma function.

Inserting (2.17) in equation (2.16), the density function of X can be expressed as a mixture of exp-G density functions

$$f(x) = \sum_{k,n=0}^{\infty} q_{k,n}\, h_{2k+n+1}(x), \tag{2.18}$$

where $q_{k,n} = \frac{\sqrt{2}(-1)^{k+n}}{(2k+n+1)\sqrt{\pi}2^k\sigma^{2k+1}k!}\binom{-2-2k}{n}$, and $h_{2k+n+1}(x)$ denotes the pdf of the exp-G (2k+n+1) distribution. The cdf correspoding (2.18) is given by

$$F(x) = \sum_{k,n=0}^{\infty} q_{k,n}\, H_{2k+n+1}(x), \tag{2.19}$$

where $H_{2k+n+1}(x)$ denotes the cdf of the exp-G(2k+n+1) distribution. Therefore, several properties of the HNG family can be obtained by using properties of the exp-G distribution; see for example, Mudholkar et al. (1995) and Nadarajah and Kotz (2006), among others.

2.4 Main properties

In this section, we obtain the quantile function (qf), ordinary and incomplete moments, moment generating function (mgf), and mean deviations of the HNG family. The formula derived throughout the dissertation can be easily handled in analytical software such as Maple and Mathematica which have the ability to deal with symbolic expressions of formidable size and complexity.

2.4.1 Quantile function

By inverting (2.5), an explicit expression for the qf of X is created, as shown below:

$$Q(u) = F^{-1}(u) = Q_G\left(\frac{\sqrt{2}\,a\,\mathrm{erf}^{-1}(\mathrm{u})}{1+\sqrt{2}\,a\,\mathrm{erf}^{-1}(\mathrm{u})}\right), \tag{2.20}$$

where $Q_G(u) = G^{-1}(u)$ is the qf of the baseline G distribution and $u \in (0,1)$.

Quantiles of interest can be obtained from (2.20) by substituting appropriate values for u. In particular, the median of X is

$$Median(X) = Q_G\left(\frac{\sqrt{2}\,a\,\mathrm{erf}^{-1}(0.5)}{1+\sqrt{2}\,a\,\mathrm{erf}^{-1}(0.5)}\right).$$

We can also use (2.20) for simulating HNG random variables: if U is a uniform random variable on the unit interval $(0,1)$, then

$$X = Q_G\left(\frac{\sqrt{2}\,a\,\mathrm{erf}^{-1}(\mathrm{U})}{1+\sqrt{2}\,a\,\mathrm{erf}^{-1}(\mathrm{U})}\right)$$

has the pdf (2.6).

Henceforth, we use an equation by Gradshteyn and Ryzhik (2007, Section 0.314) for a power series raised to a positive integer n

$$\left(\sum_{i=0}^{\infty} a_i\, u^i\right)^n = \sum_{i=0}^{\infty} c_{n,i}\, u^i, \tag{2.21}$$

where the constants $c_{n,i}$ (for $i = 1, 2, \ldots$) are determined from the recurrence equation

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

and $c_{n,0} = a_0^n$.

First, the expansion holds

$$\frac{z}{1+z} = \sum_{i=1}^{\infty} (-1)^{n+1}\, z^n. \tag{2.22}$$

Second, using (2.21) and (2.22) and the power series for the error function given at **http://mathworld.wolfram.com/InverseErf.html**, we can rewrite $Q(u)$ as

$$Q(u) = Q_G\left(\sum_{k=0}^{\infty} q_k\, u^{2k+1}\right), \tag{2.23}$$

where (for $k \geq 0$) $q_k = \sum_{n=1}^{\infty} (-1)^{n+1} (\sqrt{2}\sigma)^n c_{n,k}$, $c_{n,0} = \frac{\sqrt{\pi}}{2}$ and

$$c_{n,k} = \frac{\sqrt{\pi}}{2k(2k+1)} \sum_{j=1}^{k} [j(n+1) - k]\{[2^{-2(k+1)} \pi^{k+1/2}] + k\} c_{n,k-j}.$$

Let $W(\cdot)$ be any integrable function in the positive real line. We can write

$$\int_0^{\infty} W(x)\, f(x) dx = \int_0^1 W\left[Q_G\left(\sum_{k=0}^{\infty} q_k\, u^{2k+1}\right)\right] du. \tag{2.24}$$

Equations (2.23) and (2.24) are the main results of this section, since we can obtain from them various HNG mathematical quantities. In fact, several of them can follow by using the right-hand integral for special $W(\cdot)$ functions, which are usually more simple than if they are based on the left-hand integral. Established algebraic expansions to

determine these quantities based on equation (2.24) can sometimes be more efficient then using numerical integration of (2.6) and (2.18), which can be prone to rounding off errors among others.
For example, we can obtain easily the moments of X as $\mu'_n = E(X^n) = \int_0^1 (u \sum_{s=0}^{\infty} q_s\, u^{2s})^n\, du = \sum_{s=0}^{\infty} h_{n,s} \int_0^1 u^{n+2s} du = \sum_{s=0}^{\infty} h_{n,s}/(n+2s+1)$, where $h_{n,s}$ can be determined based on the quantities q_s from equation (2.21).

2.4.2 Moments

Some of the most important features and characteristics of a distribution can be studied through moments (e.g. tendency, dispersion, skewness and kurtosis). From now on, let $Y_{k,n} \sim \exp-\mathrm{G}(2\mathrm{k}+\mathrm{n}+1)$. A first formula for the rth moment of X can be obtained from (2.18) as

$$\mu'_r = E(X^r) = \sum_{k,n=0}^{\infty} q_{k,n}\, E(Y_{k,n}^r). \tag{2.25}$$

A second formula for $E(X^r)$ follows from (2.25) in terms of the baseline qf $Q_G(u)$. We have

$$\mu'_r = E(X^r) = \sum_{k,n=0}^{\infty} (2k+n+1)\, q_{k,n}\, \tau(r, 2k+n), \tag{2.26}$$

where $\tau(r,\sigma) = \int_{-\infty}^{\infty} x^r\, G(x)^{\sigma}\, g(x)dx = \int_0^1 Q_G(u)^r\, u^{\sigma} du$. The ordinary moments of several HNG distributions can be calculated directly from equations (2.26).

Further, the central moments (μ_r) and cumulants (κ_r) of X can be calculated as

$$\mu_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \mu_1'^k\, \mu'_{r-k} \qquad \text{and} \qquad \kappa_r = \mu'_r - \sum_{k=1}^{r-1} \binom{r-1}{k-1} \kappa_k\, \mu'_{r-k},$$

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

Plots of the skewness and kurtosis for some choices of the parameter a as function of α and β are given below.

- *The HNW distribution.* For $\beta = 5$, Figure 2.9 show that the skewness and kurtosis curves decrease with α (a fixed).

- *The HNPa distribution* For $\beta = 0.2$, Figure 2.10 show that the skewness and kurtosis curves decrease with α (a fixed).

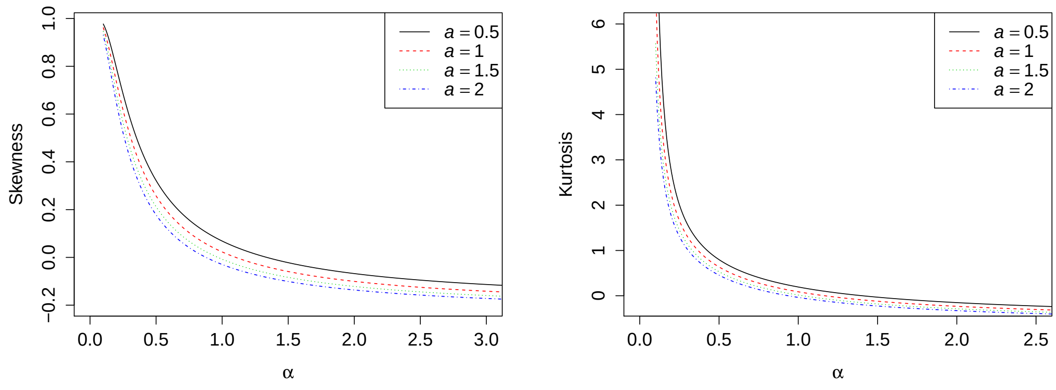


Figure 2.9: Skewness and kurtosis of the HNW distribution as a function of a, for some values of α with $\beta = 5$.

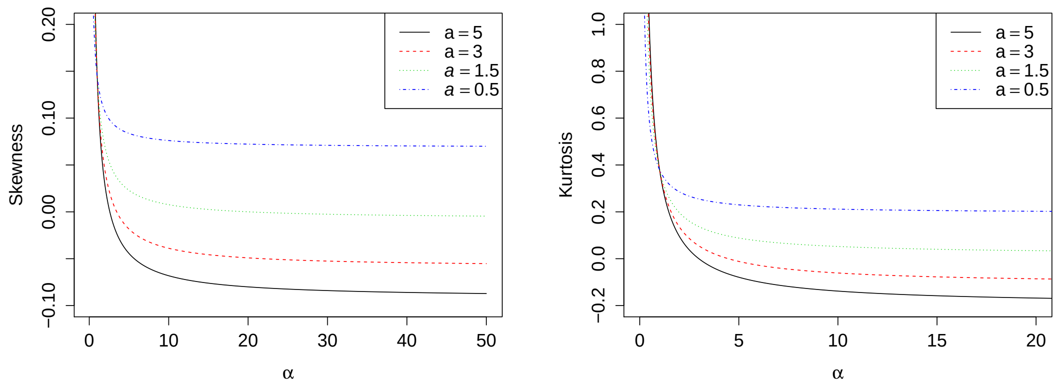


Figure 2.10: Skewness and kurtosis of the HNPa distribution as a function of a, for some values of β with $\alpha = 0.2$.

Incomplete Moments

For empirical purposes, the shape of many distributions can be usefully described by the incomplete moments. These types of moments play an important role for measuring inequality, for example, income quantiles. The main application of the first incomplete moment refers to the Bonferroni and Lorenz curves. These curves are very useful in economics, reliability, demography, insurance, and medicine.
The rth incomplete moment of X is determined as

$$m_r(y) = \int_{-\infty}^{y} x^r\, f(x)dx = \sum_{k,j=0}^{\infty} (2k+n+1)\, q_{k,n} \int_{0}^{G(y)} Q_G(u)^r\, u^{2k+n} du.$$

The last integral can be computed in most baseline G distributions.

2.4.3 Moments Generating function

We derive two formulae for the gf $M(t) = E(\mathrm{e}^{\mathrm{t\,X}})$ of X. The first formula comes from (2.18) as

$$M(t) = \sum_{k,n=0}^{\infty} q_{k,n}\, M_{k,n}(t), \tag{2.27}$$

where $M_{k,n}(t)$ is the mgf of $Y_{k,n}$. Hence, $M(t)$ can be immediately determined from the gf of the exp-G distribution.

A secund formula for $M(t)$ can be derived from (2.18) as

$$M(t) = \sum_{k,n=0}^{\infty} (2k+n+1)\, q_{k,n}\, \rho(t, 2k+n), \tag{2.28}$$

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

Therefore, we can obtain the gf's of several HNG distributions directly from equations (2.28).

2.5 Mean deviations

The mean deviations about the mean ($\delta_1 = E(|X - \mu_1'|)$) and about the median ($\delta_2 = E(|X - M|)$) of X can be expressed as $\delta_1 = 2\mu_1'\, F\,(\mu_1') - 2m_1\,(\mu_1')$ and $\delta_2(X) = \mu_1' - 2m_1(M)$, respectively, where $\mu_1' = E(X)$, $M = Median(X)$ is the median, $F(\mu_1')$ is calculated (2.5) and $m_1(z) = \int_{-\infty}^{z} xf(x)dx$ is the first incomplete moment.

We provide two alternative ways to compute $m_1(z)$. The first one comes from (2.18) as

$$m_1(z) = \sum_{k,n=0}^{\infty} q_{k,n}\, J_{k,n}(z), \tag{2.29}$$

where $J_{k,n}(z) = \int_{-\infty}^{z} x\, h_{2k+n+1}(x)dx$ is the basic quantity to compute the mean deviations of the exp-G distributions.

A second formula for $m_1(z)$ can be derived by setting $u = G(x)$ in equation (2.18)

$$m_1(z) = \sum_{k,n=0}^{\infty} (2k+n+1)\, q_{k,n}\, T_{k,n}(z), \tag{2.30}$$

where $T_{k,n}(z) = \int_0^{G(z)} Q_G(u)\, u^{2k+n} du$ is a simple integral based on qf of G. Hence, the mean deviations of the HNG family can be computed from (2.29) and (2.30).

Applications of these equations to obtain Bonferroni and Lorenz curves can be performed. For a given probability π, these curves are given by $B(\pi) = m_1(q)/\pi\mu_1'$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $\mu_1' = E(X)$, $q = Q_G\left[\frac{\sqrt{2}\sigma \text{erf}^{-1}(\pi)}{[1-\sqrt{2}\sigma \text{erf}^{-1}(\pi)]}\right]$ is the qf of X at π and $m_1(q)$ comes from the equations given before.

2.6 Maximum likelihood estimation

In this section, we determine the maximum likelihood estimates (MLEs) of the model parameters of the new family from complete samples only. Let $x_1, \ldots, x_n$ be observed values from the HNG distribution. Let $\underline{\boldsymbol{\theta}}$ be a p-vector parameter vector specifying $G(\cdot)$. The log-likelihood function $\log L = \log L(a, \underline{\boldsymbol{\theta}})$ is given by

$$\log L = \frac{n}{2}\log(2/\pi) - n\log a + \sum_{i=1}^{n}\log g(x_i) - 2\sum_{i=1}^{n}\log[\bar{G}(x_i)] - \frac{1}{2a^2}\sum_{i=1}^{n}\frac{G(x_i)^2}{\bar{G}(x_i)^2} \quad (2.31)$$

The first derivatives of log L with respect to the parameters a and $\underline{\boldsymbol{\theta}}$ are

$$\frac{\partial \log L}{\partial a} = \frac{1}{a^3}\sum_{i=1}^{n}\frac{G(x_i)^2}{\bar{G}(x_i)^2} - \frac{n}{a}$$

and

$$\begin{aligned}\frac{\partial \log L}{\partial \underline{\boldsymbol{\theta}}} &= \sum_{i=1}^{n}\frac{\partial g(x_i)/\partial\underline{\boldsymbol{\theta}}}{g(x_i)} - 2\sum_{i=1}^{n}\frac{\partial \bar{G}(x_i)/\partial\underline{\boldsymbol{\theta}}}{\bar{G}(x_i)} \\ &- \frac{1}{a^2}\sum_{i=1}^{n}\frac{G(x_i)\,\bar{G}(x_i)^2\,\partial G(x_i)/\partial\underline{\boldsymbol{\theta}} - G(x_i)^2\,\bar{G}(x_i)\,\partial\bar{G}(x_i)/\partial\underline{\boldsymbol{\theta}}}{\bar{G}(x_i)^4}\end{aligned}$$

The maximum likelihood estimates (MLEs) of $(a, \underline{\boldsymbol{\theta}})$, say $(\hat{a}, \hat{\underline{\boldsymbol{\theta}}})$, are the simultaneous solutions of the equations $\partial \log L/\partial a = 0$ and $\partial \log L \partial\underline{\boldsymbol{\theta}} = 0$. We estimate the unknown parameters of each model by maximum likelihood. There exist many maximization methods in R packages like NR (Newton-Raphson), BFGS (Broyden-FletcherGoldfarb-Shanno), BHHH (Berndt-Hall-Hall-Hausman), SANN (Simulated-Annealing) and NM (Nelder-Mead). The MLEs are calculated using Limited Memory quasi-Newton code for Bound-constrained optimization (L-BFGS-B) and the Anderson-Darling (A^*) and Cramér-Von Mises (W^*) statistics are computed to compare the fitted models. The computations are carried out using R-package *AdequacyModel* given freely from *http://cran.r-project.org/web/packages/AdequacyModel/AdequacyModel.pdf.*

2.7 Applications

2.7.1 Exceedances of Flood Peaks

The data refer to exceedances of flood peaks (in m^3/s) of the Wheaton River near Carcross in Yukon Territory, Canada (see, for instance, Choulakian and Stephens, 2001; Mahmoudi, 2011). In many applications, there is qualitative information about the hrf, which can help with selecting a particular model. In this context, a device called the total time on test (**TTT**) plot, Aarset (1987), is useful. The **TTT** plot is obtained by plotting $G(r/n) = [(\sum_{i=1}^{r} y_{i:n}) + (n-r)y_{r:n}]/\sum_{i=1}^{n} y_{i:n}$, where $r = 1, \ldots, n$ and $y_{i:n}$ $(i = 1, \ldots, n)$ are the order statistics of the sample, against r/n. It is a straight diagonal for constant failure rates, and it is convex for decreasing failure rates and concave for increasing failure rates. It is first convex and then concave if the failure rate is bathtub shaped. It is first concave and then convex if the failure rate is upside-down bathtub. The **TTT** plot for the exceedances of flood peaks data in Figure 2.11 indicates a bathtub-shaped failure rate and therefore the appropriateness of the HNW distribution to fit these data.

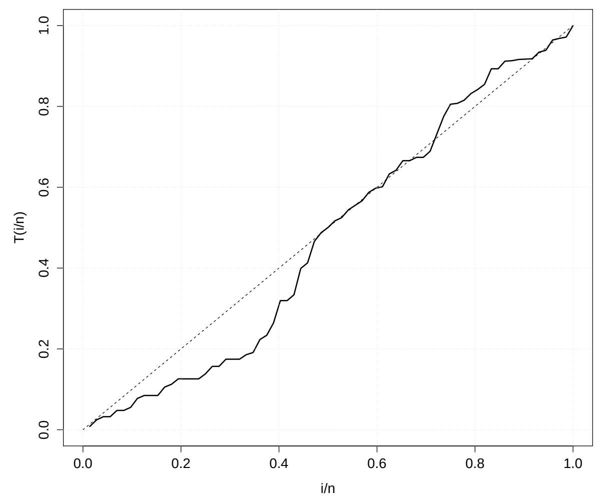


Figure 2.11: TTT-plot for the exceedances of flood peak data.

For these data, we fit the HNW distribution defined in (2.8). Its fit is compared with others models well-known in literature: gamma-weibull (GW), Nadarajah et al. (2012), exponentiated Weibull (EW), Mudholkar and Srivastava (1993), and Weibull (We), Gurvich et al. (1997), models with corresponding densities:

$$\begin{aligned} f_{GW}(x) &= \frac{\alpha\beta^{\alpha a}}{\Gamma(a)} x^{a\alpha-1} \mathrm{e}^{-(\beta \mathrm{x})^{\alpha}}, \\ f_{EW}(x) &= a\alpha\beta^{\alpha} x^{\alpha-1} \mathrm{e}^{-(\beta \mathrm{x})^{\alpha}} \left(1-\mathrm{e}^{-(\beta \mathrm{x})^{\alpha}}\right)^{\mathrm{a}-1}, \\ f_{We}(x) &= \alpha\beta x^{\beta-1} \mathrm{e}^{-\alpha \mathrm{x}^{\beta}}, \end{aligned}$$

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

Table 2.1 lists the MLEs of the parameters (the standard errors are given in parentheses) for the HNW, GW, EW and We distributions fitted to the exceedances of flood peak data.

Table 2.1: MLEs (standard errors in parentheses)

Model	Estimates		
HNW(a,α,β)	0.1873	0.6734	0.0048
	(0.0793)	(0.0766)	(0.0021)
GW(a,α,β)	1.4887	0.0411	0.4783
	(0.7231)	(0.0193)	(0.3207)
EW(a,α,β)	0.5194	0.0502	1.3852
	(0.3118)	(0.0209)	(0.5881)
We(α,β)	0.0859	0.9011	
	(0.0118)	(0.0855)	

Table 2.2: W^* and A^* statistics

Model	W^*	A^*
HNW	**0.09842**	**0.62177**
GW	0.10541	0.64624
EW	0.10509	0.64234
We	0.13799	0.78544

The statistics W^* and A^* are described in Chen and Balakrishnan (1995). In general, the smaller the values of the goodness-of-fit measures, the better the fit to the data. The statistics W^* and A^* for all models are listed in Table 2.2. From the figures in this table, we conclude that the HNW model fits the current data better than the other models. Therefore, the HNW model may be an interesting alternative to other models available in the literature for modeling positive real data with bathtub-shaped hrf.

The estimated pdf and cdf for the fitted HNW model to the current data and the histogram of the data and the empirical cumulative distribution are displayed in Figure 2.12.

2.7.2 Percentage of body fat data

Here, we consider the data referring to the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. A variety of popular health books suggest that the readers assess their health, at least in part, by estimating their percentage of body fat. In Bailey (1994), for instance, readers can estimate body fat from tables using their age and various skin-fold measurements obtained by using a caliper. Other texts give predictive equations for body fat using body circumference measurements (e.g. abdominal circumference) and/or skin-fold measurements. See,

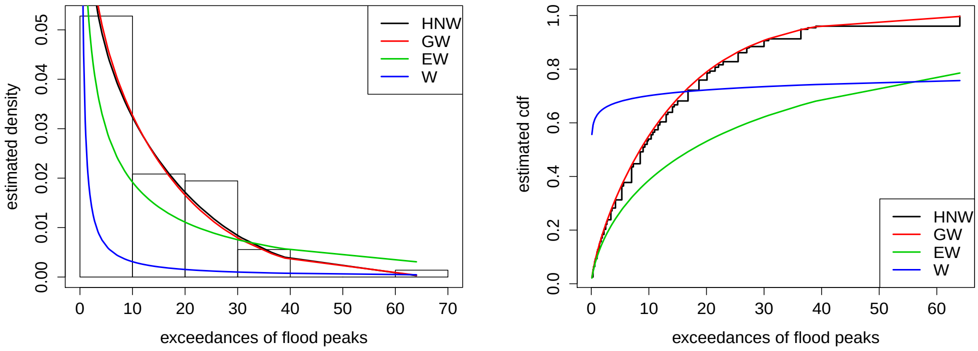


Figure 2.12: left: Estimated density of the HNW model; right: Empirical cdf and estimated cdf of the HNW model

for instance, Behnke and Wilmore (1974, pp. 66-67), Wilmore (1977) and Katch and McArdle (1983, pp. 120-132).

Percentage of body fat for an individual can be estimated once body density has been determined. Siri (1956) assumes that the body consists of two components: lean body tissue and fat tissue. The data can also be found on the site *http://lib.stat.cmu.edu/datasets/bodyfat*. For the TTT plot in Figure 2.13, we can verify a concave curve and therefore we can consider distributions with increasing hrf for these data.

Therefore, we fit the HNLL distribution defined in (2.14) and the exponentiated log-logistic (ELL) (Rosaiah et al. , 2006), McDonald log-logistic (McLL) (Tahir et al., 2014), beta log-logistic (BLL) (Lemonte, 2012) and log-logistic (LL) models with associated densities:

$$
\begin{aligned}
f_{ELL}(x) &= \frac{\alpha a}{\beta^{\alpha a}} x^{\alpha a-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-(a+1)}, \\
f_{McLL}(x) &= \frac{c}{B(ac^{-1},b)}\left(\frac{\alpha}{\beta^{a\alpha-1}}\right) x^{\alpha a-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-(a+1)} \\
&\times \left[1-\left\{1-\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-1}\right\}^{c}\right]^{b-1}, \\
f_{BLL}(x) &= \frac{1}{B(a,b)}\left(\frac{\alpha}{\beta^{a\alpha-1}}\right) x^{a\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-(\alpha+\beta)}, \\
f_{LL}(x) &= \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1}\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{-2},
\end{aligned}
$$

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

The MLEs of the parameters (with standard errors) of the fitted models are given in Table 2.3. The statistics W^* and A^* for all these fitted models are listed in Table 2.4.

Based on the statistics W^* and A^*, we conclude that the HNLL model fits the current data better than the other models. Therefore, it may be an interesting alternative to other lifetime models available in the literature for modeling positive real data with increased-shaped hrf.

The estimated pdf and cdf of the three best fitted models to the histogram of the data and the empirical cumulative distribution are displayed in Figure 2.14.

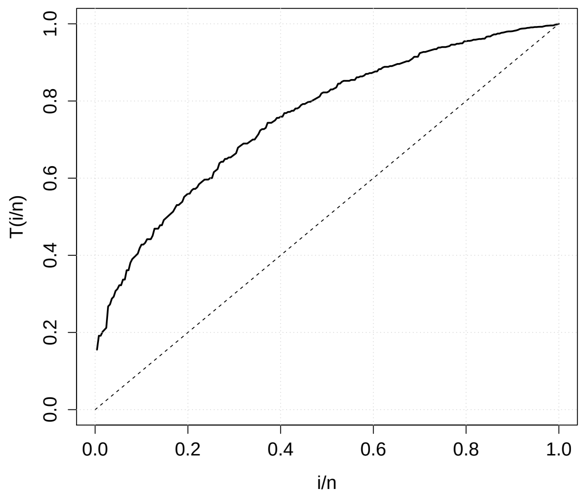


Figure 2.13: TTT plot for percent body fat from Siri (1956) equation data.

Table 2.3: MLEs (standard errors in parentheses)

Model	Estimates				
HNLL(a,α,β)	8.615931	1.978180	7.302491		
	(9.2474)	(0.1011)	(15.4808)		
ELL(a,α,β)	28.4033	8.9949	0.2093		
	(1.2439)	(1.5806)	(0.0486)		
McLL(a,b,c,α,β)	1.093647	23.7361	6.5262	2.2521	27.3499
	(1.3685)	(67.6007)	(5.6939)	(2.1495)	(8.2357)
BELL(a,b,α,β)	4.2583	41.2238	0.49373	6.0359	
	(1.9744)	(19.8646)	(0.2869)	(9.7253)	
LL(α,β)	18.1616	3.5074			
	(0.5707)	(0.1852)			

2.8 Concluding remarks

There has been a great interest among statisticians and applied researchers to construct flexible lifetime models to provide better fits to survival data. In this dissertation,

Table 2.4: W^* and A^* statistics

Model	W^*	A^*
HNLL	**0.02152**	**0.19394**
ELL	0.058609	0.42661
McLL	0.02655	0.19654
BELL	0.03223	0.24747
LL	0.60602	3.64363

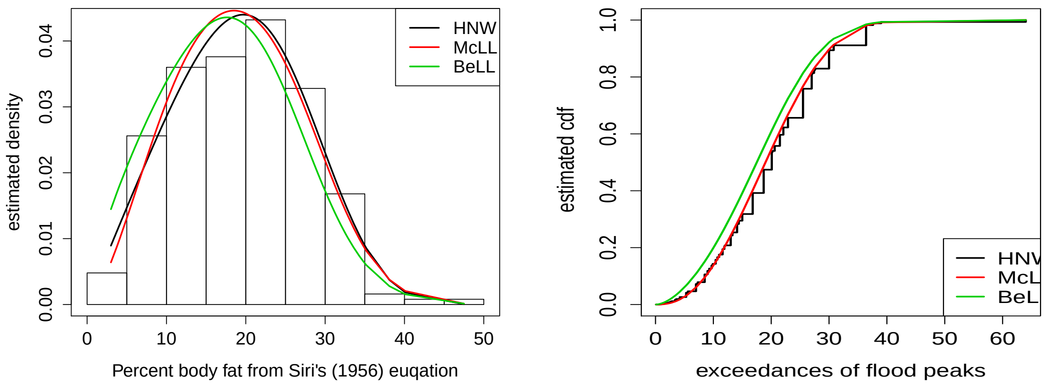


Figure 2.14: left: Estimated density of the HNLL model; right: Empirical cdf and estimated cdf of the HNLL model

we propose the new half normal family of distributions. We provide the density, cumulative, and hazard rate functions of the new model. Some of its structural properties include an expansion for the density function and explicit expressions for the quantile function, ordinary and incomplete moments, generating function and mean deviations. The maximum likelihood method is employed for estimating the model parameters. We fit two special models of the proposed family to real data sets to demonstrate the flexibility of the new family compared to other classes of distributions. These special models provide consistently better fits than other competing models. We hope that the proposed distribution will serve as an alternative model to other models available in the literature for modelling positive real data in many areas such as engineering, survival analysis, hydrology, and economics.

Chapter 3

The Kumaraswamy Nadarajah-Haghighi distribution

Resumo

Propomos novo modelo de tempo de vida de quatro parâmetros, chamado distribuição *Kumaraswamy Nadarajah-Haghighi*, para generalizar o modelo de dois parâmeros Nadarajah-Haghighi.A distribuição proposta é muito flexível para análise de dados positivos. A sua função de risco pode ser constante, decrescente, crescente, banheira invertida e forma de banheira dependendo dos valores dos parâmetros. Ele inclui como modelos especiais as distribuições Nadarajah-Haghighi e Nadarajah-Haghighi exponencializada (Lemonte, 2013). Apresentamos expressões explicitas para os momentos ordinário e incompleto, desvios médios, função quantil e estatísticas de ordem. A estimação dos parâmetros do modelo é realizada por máxima verossimilhança. A flexibilidade do novo modelo é ilustrado empiricamente por meio de duas aplicações a banco de dados reais. Esperamos que a nova distribuição servirá como um modelo alternativo para outras distribuições úteis para a modelagem de dados reais positivos em muitas áreas.

Palavras-chave: Distribuição Kumaraswamy; distribuição Nadarajah-Haghighi; estimação de máxima verossimilhança; função taxa de falha; momentos.

Abstract

We propose a new four-parameter lifetime model, called the Kumaraswamy Nadarajah-Haghighi distribution, to generalize the two-parameter Nadarajah-Haghighi model. The new model is quite flexible for analyzing positive data. Its hazard function can be constant, decreasing, increasing, upside-down bathtub, and bathtub-shaped depending on the parameter values. It includes as special models the Nadarajah-Haghighi and exponentiated Nadarajah-Haghighi (Lemonte, 2013) distributions. We provide explicit expressions for the ordinary and incomplete moments, mean deviations, quantile function, and order statistics. The estimation of the model parameters is performed by maximum likelihood using the `BFGS` algorithm. The flexibility of the new model is proved empirically by means of two applications to real data set. We hope that the new distribution will serve as an alternative model to other useful distributions for modeling positive real data in many areas.

Key words: Hazard function. Kumaraswamy distribution. Maximum likelihood estimation. Moment. Nadarajah and Haghighi distribution.

3.1 Introduction

In recent years, several ways of generating new continuous distributions have been proposed based on different modifications of the beta, gamma, and Weibull distributions, among others, to provide bathtub hazard rate functions (hrfs). The beta generated family was proposed by Eugene et al. (2002). Jones (2004) studied a family that arose naturally from the distribution of order statistics. Some researchers have suggested using other bounded distributions on $(0, 1)$ to obtain a generalization of any parent cumulative distribution function (cdf).

Cordeiro and de Castro (2011) proposed another generator called the Kumaraswamy-G (Kw-G for short) class. For any parent cdf $G(x)$, they defined the probability density function (pdf) $f(x)$ and cdf $F(x)$ of the Kw-G family by

$$f(x) = a\, b\, g(x)\, G(x)^{a-1}\, [1 - G(x)^a]^{b-1}, \tag{3.1}$$

and

$$F(x) = 1 - [1 - G(x)^a]^b, \tag{3.2}$$

respectively, where $g(x) = dG(x)/dx$ and $a > 0$ and $b > 0$ are additional shape parameters to the G model. If X is a random variable with density (3.1), we write $X \sim$Kw-G(a, b). Each new Kw-G distribution can be obtained from a parent G distribution.

One major benefit of the Kw-G family is its ability to fitting skewed data that can not be properly modeled by existing distributions. This fact was demonstrated recently by Cordeiro et al. (2010), who applied the Kumaraswamy Weibull distribution to failure data. The density family (3.1) has many of the same properties of the class of beta-G distributions, but has some advantages in terms of tractability, since it does not involve any special function such as the beta function. The Kw-Gumbel by Cordeiro et al. (2012b), Kw-Birnbaum-Saunders by Saulo et al. (2012), Kw-normal by Correa et al. (2012), Kw-Pareto by Bourguignon et al. (2013), Kw-BurrXII by Paranaiba et al. (2013), Kw-Lomax by Shams (2013), and Kw-generalized Rayleigh by Gomes et al. (2014) distributions are some examples obtained by taking $G(x)$ to be the cdf of the Gumbel, Birnbaum-Saunders, normal, Pareto, Burr XII, Lomax and generalized Rayleigh distributions, respectively, among several others. Hence, each new Kw-G distribution can be generated form a specified G distribution.

Nadarajah and Haghighi (2011) proposed a generalization of the exponential distribution as an alternative to the gamma, Weibull and exponentiated exponential (EE) distributions with cdf and pdf (for $x > 0$) given by

$$G(x;\alpha,\beta) = 1 - \exp[1 - (1 + \beta\, x)^{\alpha}] \tag{3.3}$$

and

$$g(x;\alpha,\beta) = \alpha\beta(1 + \beta\, x)^{\alpha-1} \exp[1 - (1 + \beta\, x)^{\alpha}], \tag{3.4}$$

respectively, where $\alpha > 0$ is a shape parameter and $\beta > 0$ is a scale parameter. If Y follows the Nadarajah-Haghighi (NH) model, we write $Y \sim$NH(α,β). This generalization always has its mode at zero and allows for increasing, decreasing and constant hrfs. Lemonte (2013) extended this distribution by applying the exponentiated class and studied the generated model called the exponentiated NH (ENH) distribution, whose hrf can exhibit the classical four shapes: increasing, decreasing, unimodal and bathtub-shaped.

Bourguignon et al. (2015) extended the NH distribution by defining the gamma NH (GNH) based on the class of generalized gamma-G distributions pioneered by Zografos and Balakrishnan (2009). The generalized distributions follow by taking any parent G distribution in the cdf of a gamma distribution with one additional shape parameter. In a similar way, many gamma-type distributions were introduced and studied; see, for example, the gamma-uniform investigated by Torabi and Montazari (2012).

In this chapter, we propose a four-parameter extension of the NH distribution named the *Kumaraswamy Nadarajah-Haghighi* (Kw-NH) distribution by combining the works of Cordeiro and de Castro (2011) and Nadarajah and Haghighi (2011). We derive some mathematical properties of the new distribution. It can have increasing, decreasing, unimodal and bathtub-shaped hazard functions and thus is quite flexible to analyze lifetime data.

3.2 The Kw-NH distribution

By inserting (3.3) and (3.4) in equation (3.1), we define the Kw-NH density function with positive shape parameters a, b, α and scale parameter $\beta > 0$, for $x > 0$ by

$$\begin{aligned} f(x;a,b,\alpha,\beta) = \quad & a\,b\,\alpha\,\beta\,(1+\beta\,x)^{\alpha-1}\{\exp[1-(1+\beta\,x)^{\alpha}]\}\,\{1-\exp[1-(1+\beta\,x)^{\alpha}]\}^{a-1} \\ & \times\{1-\{1-\exp[1-(1+\beta\,x)^{\alpha}]\}^{a}\}^{b-1}. \end{aligned} \tag{3.5}$$

Evidently, the above density function does not involve any complicated function, which is a positive point of the current generalization. Hereafter, a random variable X following (3.5) is denoted by $X \sim$Kw-NH(a, b, α, β).

The study of (3.5) is important since it extends very useful distributions. In fact, the

NH distribution is obtained when $a = b = 1$. The exponential distribution follows by taking $\alpha = \beta = 1$ and $a = 1$. The ENH distribution is a special case when $b = 1$. The cdf and hrf corresponding to (3.5) are given by

$$F(x; a, b, \alpha, \beta) = 1 - \{1 - \{1 - \exp\left[1 - (1 + \beta\, x)^{\alpha}\right]\}^{a}\}^{b} \tag{3.6}$$

and

$$\tau(x) = \frac{a\, b\, \alpha\, \beta\, (1 + \beta\, x)^{\alpha-1}\{\exp[1 - (1 + \beta\, x)^{\alpha}]\}\,\{1 - \exp[1 - (1 + \beta\, x)^{\alpha}]\}^{a-1}}{1 - \{1 - \exp\left[1 - (1 + \beta\, x)^{\alpha}\right]\}^{a}}, \tag{3.7}$$

respectively.

Figures 3.1 and 3.2 display some plots of the density and hrf of the Kw-NH distribution for selected parameter values, respectively. The plots reveal that the new distribution is very flexible and that its hrf can have decreasing, increasing, upside-down bathtub, and bathtub-shaped forms.

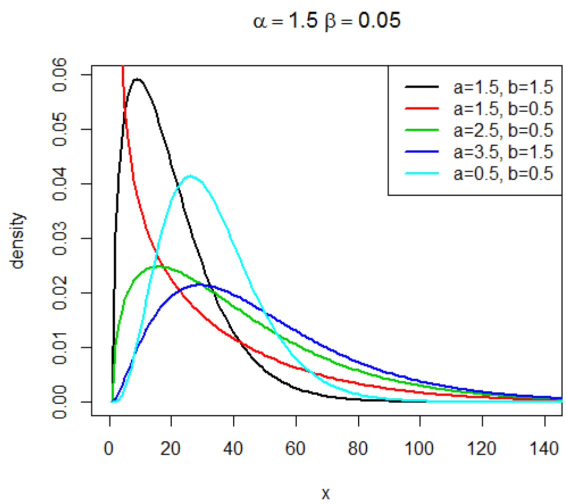


Figure 3.1: Plots of the Kw-NH density for some parameter values.

3.2.1 Linear representations

Expansions for equations (3.5) and (3.6) can be derived using the concept of exponentiated distributions. We obtain a linear representation for the Kw-NH cumulative distribution by using the generalized binomial theorem (for $|z| < 1$ and $\epsilon > 0$)

$$(1 - z)^{\epsilon} = \sum_{j=0}^{\infty} (-1)^{j} \binom{\epsilon}{j} z^{j}, \tag{3.8}$$

where $\binom{\epsilon}{j} = \epsilon(\epsilon - 1)(\epsilon - 2) \ldots (\epsilon - j + 1)/j!$.

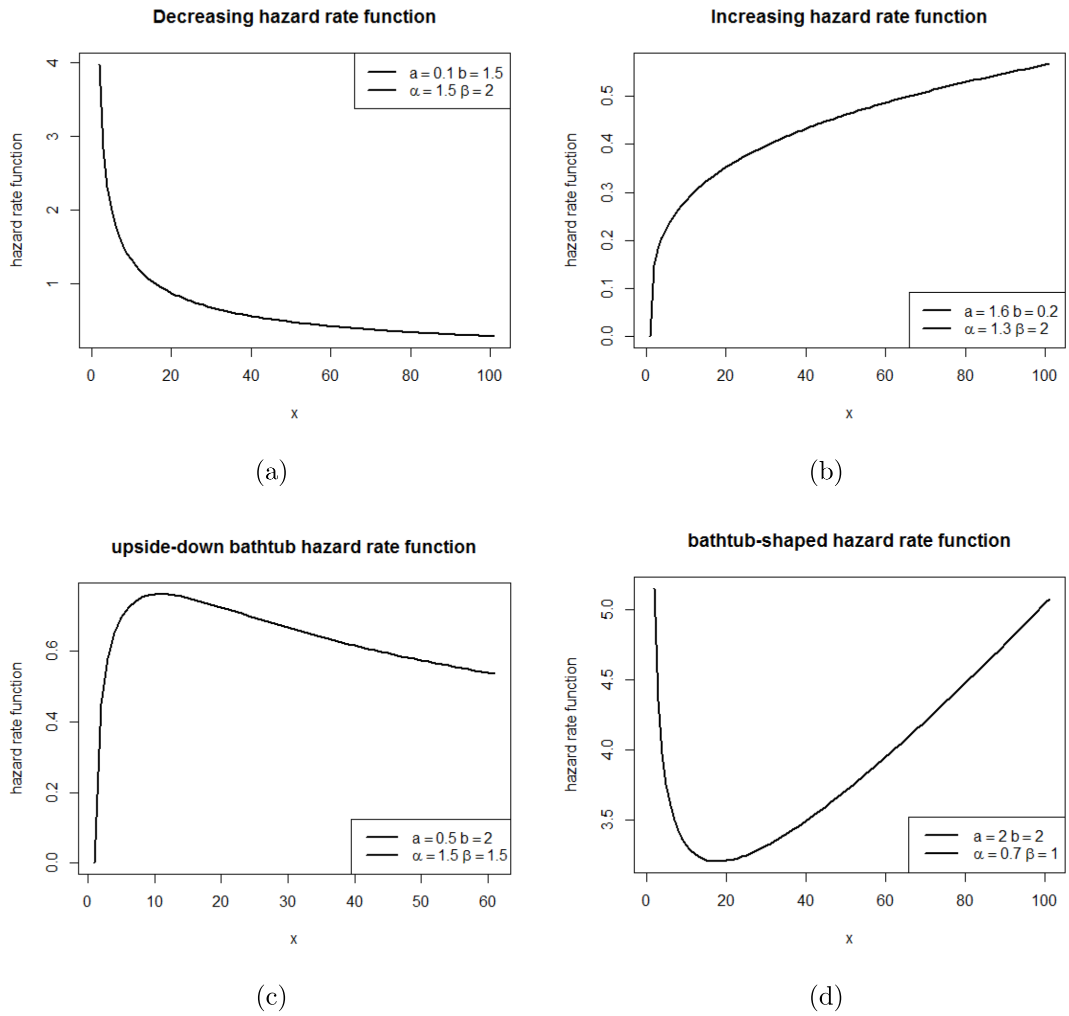


Figure 3.2: The Kw-NH hrf for some parameter values.

We can write (3.6) as

$$\begin{aligned} F(x;\alpha,\beta,a,b) &= \sum_{j=0}^{\infty}(-1)^j \binom{b}{j} \left\{1-\exp\left[1-(1-\beta\, x)^{\alpha}\right]\right\}^{ja} \\ &= \sum_{j=0}^{\infty} \omega_j\, H(x;\alpha,\beta,ja), \end{aligned} \tag{3.9}$$

where $\omega_j = (-1)^j \binom{b}{j}$ and $H(x;\alpha,\beta,ja)$ denotes the ENH cumulative distribution given by

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

By differentiating (3.9) with respect to x, we obtain

$$f(x;\alpha,\beta,a,b) = \sum_{j=0}^{\infty} \rho_j\, h(x;\alpha,\beta,ja) \tag{3.10}$$

where $\rho_j = \frac{(-1)^j\, b}{(j+1)}\binom{b-1}{j}$ and $h(x;\alpha,\beta,ja)$ denotes the ENH density function with parameters α, β and $(j+1)a$. Then, the Kw-NH density function can be expressed as an infinite linear combination of ENH densities and then some of its basic mathematical properties can be obtained from those ENH properties. For example, the ordinary, incomplete and factorial moments and moment generating function (mgf) of the Kw-NH distribution can follow from those ENH quantities.

3.2.2 Limiting behaviour of the density

Lemma 1. The limit of the density function of X when $x \to \infty$ is 0 and the limit as $x \to 0$ are

$$\lim_{x\to 0} f(x;a,b,\alpha,\beta) = \begin{cases} \infty, & a < 1, \\ b\,\alpha\,\beta, & a = 1, \\ 0, & a > 1. \end{cases}$$

Proof. It is easy to demonstrate the result from the density function (3.5)

3.3 Quantile function

The qf of X, say $x = Q(u)$, follows easily by inverting (3.6) as

$$x = Q(u) = \beta^{-1}\left\{1 - \log\left\{1 - \left[1-(1-u)^{1/b}\right]^{1/a}\right\}\right\}^{1/\alpha} - \beta^{-1}, \quad u \in (0,1). \tag{3.11}$$

Quantiles of interest can be obtained from (3.11) by substituting appropriate values for u. In particular, the median of X is $Q(0.5)$

$$Median(X) = \beta^{-1}\left\{1 - \log\left\{1 - \left[1-(1-0.5)^{1/b}\right]^{1/a}\right\}\right\}^{1/\alpha} - \beta^{-1}.$$

We can also use (3.11) for simulating Kw-NH random variables: if U is a uniform random variable on the unit interval $(0,1)$, then

$$X = Q(U) = \beta^{-1}\left\{1 - \log\left\{1 - \left[1-(1-U)^{1/b}\right]^{1/a}\right\}\right\}^{1/\alpha} - \beta^{-1}$$

has the pdf (3.5). Next, we use the qf given by (3.11) to determine the Bowley's skewness and the Moors' kurtosis. The Bowley's skewness is based on quartiles, see Kenney and Keeping (1962), given by

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

and the Moor's kurtosis, see Moors (1998), is based on octiles given by

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

where $Q(\cdot)$ is the qf given by (3.11). Plots of the skewness and kurtosis for selected values of b, as functions of a, and for selected values of a, as functions of b, for $\alpha = 1.5$, $\beta = 0.5$, are displayed in Figures 3.3. These plots reveal that the skewness increases or decreases for b fixed and decreases for a fixed, whereas the kurtosis decreases when b increases for fixed a and when a increases for fixed b.

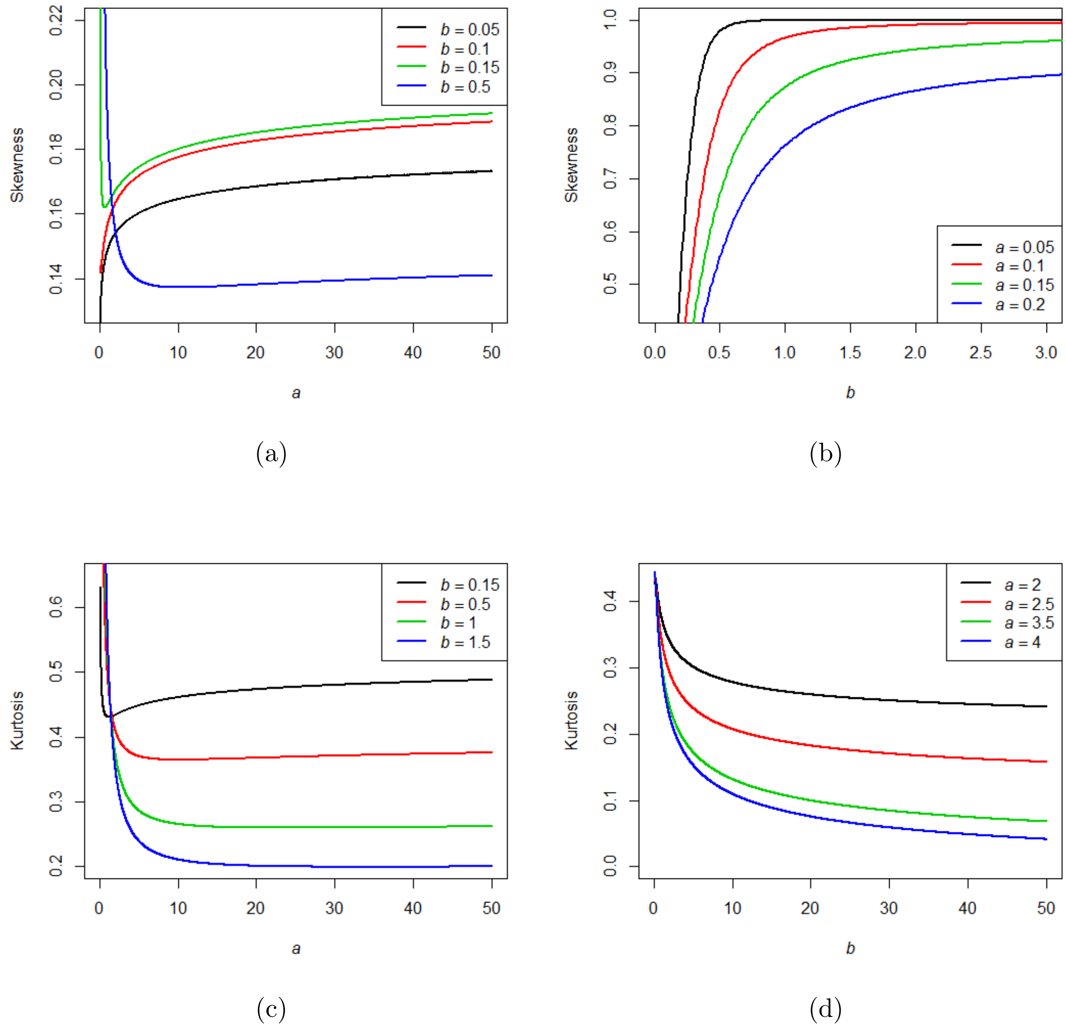


Figure 3.3: Skewness and kurtosis of the Kw-NH distribution.

3.4 Moments

The ordinary and incomplete moments of X can be determined from the moments of $Y_j \sim$ENH$((j+1)a)$. First, we can write from (3.10)

$$\mu_r' = E(X^r) = \sum_{j=0}^{\infty} \frac{\rho_j}{(j+1)}\, E(Y_j^r)$$

and then change variables to use the ENH qf

$$\mu_r' = \beta^{-r} \sum_{j=0}^{\infty} \frac{\rho_j}{(j+1)}\, T_r(\alpha, (j+1)a),$$

where $T_r(\alpha, (j+1)a) = \int_0^1 \{[1-\log(1-u)]^{1/\alpha} - 1\}^r\, u^{(j+1)a-1}\, du$ is an integral to be evaluated numerically.

Alternatively, the moments of X can be determined based on the quantity $E(Y_j^r)$ given by Lemonte (2013) as

$$\mu_r' = \beta^{-r} \sum_{k,j=0}^{\infty} \sum_{i=0}^{r} \frac{(-1)^{r+k-i}\,(j+1)a\, e^{k+1}\, \rho_j}{(k+1)^{1/\alpha+1}} \binom{(j+1)a-1}{k} \binom{r}{i}\, \Gamma\left(\frac{i}{\alpha}+1, k+1\right), \tag{3.12}$$

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

The central moments (μ_r) and cumulants (κ_r) of X can be determined from (3.4) or (3.12) as

$$\mu_r = \sum_{k=0}^{r} \binom{r}{k} (-1)^k\, \mu_1'^k\, \mu_{s-k}', \qquad \kappa_r = \mu_s' - \sum_{k=1}^{r-1} \binom{r-1}{k-1} \kappa_k\, \kappa_{r-k},$$

respectively, where $\kappa_1 = \mu_1'$. Thus, $\kappa_2 = \mu_2' - \mu_1'^2$, $\kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$, $\kappa_4 = \mu_4' - 4\mu_3'\mu_1' - 3\mu_2'^2 + 12\mu_2'\mu_1'^2 - 6\mu_1'^4$, etc. Additionally, the skewness and kurtosis can be obtained from the third and fourth standardized cumulants in the forms $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and $\gamma_2 = \kappa_4/\kappa_2^2$, respectively.

The rth incomplete moment of X is given by

$$m_r(z) = \int_0^z x^r\, f(x; a, b, \alpha, \beta) dx = \sum_{j=0}^{\infty} \frac{\rho_j}{(j+1)} \int_0^z x^r\, h(x; \alpha, \beta, (j+1)a) dx$$

where $\int_0^z x^r\, h(x; \alpha, \beta, (j+1)a)dx$ is the rth incomplete moment of the ENH distribution. Then,

$$m_r(z) = \beta^{-1} \sum_{j=0}^{\infty} \frac{\rho_j}{(j+1)} \int_{1-e^{1-(1+\beta\, z)^{\alpha}}}^{\infty} \{[1-\log(1-u)]^{1/\alpha} - 1\}^r\, u^{(j+1)a-1}\, du.$$

Alternatively, using the incomplete moments given by Lemonte (2013), $m_r(z)$ reduces to

$$\begin{aligned} m_r(z) &= \beta^{-r} \sum_{k,j=0}^{\infty} \sum_{i=0}^{r} \frac{(-1)^{r+k-i}\,(j+1)a\,e^{k+1}\,\rho_j}{(k+1)^{1/\alpha+1}} \binom{(j+1)a-1}{k} \binom{r}{i} \\ &\times\; \Gamma\left(\frac{i}{\alpha}+1,(k+1)[1-(1+\beta\,z)^{\alpha}]\right). \end{aligned} \tag{3.13}$$

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

$$\delta_1 = 2\,\mu_1'\,F(\mu_1') - 2\,m_1\,(\mu_1') \quad \text{and} \quad \delta_2 = \mu_1' - 2\,\mathrm{m}_1(\mathrm{M}),$$

where $\mu_1' = E(X)$, $M = Median(X)$ is the median of X and $m_1(z)$ is given by (3.13) with $r=1$.

Finally, we can construct Lorenz and Bonferroni curves, which are important in several fields such as economics, reliability, demography, insurance and medicine, based on the first incomplete moment. They are defined (for a given probability π) by $L(\pi) = T_1(q)/\mu_1'$ and $B(\pi) = T_1(q)/(\pi\mu_1')$, respectively, where $q = Q(\pi)$ is determined from (3.11).

3.5 Moments Generating function

A representation for the mgf $M(t)$ of X can follow from the ENH generating function. We can write $M(t) = \sum_{j=0}^{\infty} \rho_j\, M_j(t)$, where ρ_j is defined by (3.5) and $M_j(t)$ is the mgf of Y_j ∼ENH($(j+1)a$), see Bourguignon et al. (2015), given by

$$M_j(t) = \sum_{s,r=0}^{\infty} \frac{\eta_s\, g_{s,r}\, t^j}{r/\beta+1}. \tag{3.14}$$

Here, for $s \geq 0$, $\eta_s = \sum_{j=0}^{\infty} \frac{(-1)^s}{\lambda^j\, j!}\binom{j}{s}$, $g_{s,0} = \zeta_0^s$ and $g_{s,r} = (r\zeta_0)^{-1}\sum_{n=1}^{r}[n(s+1)-r]\,\zeta_n\, g_{s,r-n}$ (for $r \geq 1$), where $\zeta_r = \sum_{m=0}^{\infty} f_m\, d_{m,r}$ (for $r \geq 0$) and, for $m \geq 0$, $d_{m,0} = a_0^m$, $d_{m,r} = (r\,a_0)^{-1}\sum_{s=0}^{r}[s(m+1)-r]\,a_s\, d_{m,r-s}$ $(r \geq 1)$, $f_m = \sum_{l=m}^{\infty}(-1)^{l-m}\binom{l}{m}(\alpha^{-1})_l/l!$, and $(\alpha^{-1})_l = (\alpha^{-1}) \times (\alpha^{-1}-1)\ldots(\alpha^{-1}-l+1)$ is the descending factorial.

A second representation for $M(t) = E[\exp(tx)]$ of X is obtained from (3.10). We have

$$M(t) = \int_0^{\infty} \exp(tx)\, f(x) = \int_0^{\infty} \exp(tx) \sum_{j=0}^{\infty} \frac{\rho_j}{(j+1)}\, h(x)$$

By using the binomial series expansions, we obtain

$$\{1-\exp[1-(1+\beta\,x)^{\alpha}]\}^{(j+1)a} = \sum_{i=0}^{\infty} \binom{(j+1)a}{i} (-1)^i \exp\{i[1-(1+\beta\,x)^{\alpha}]\}$$

and

$$\exp(tx) = \sum_{s=0}^{\infty} \frac{t^s\, x^s}{s!}$$

Hence, we can write

$$\begin{aligned} M(t) &= \alpha\,\beta\,a \sum_{j,i=0}^{\infty} \rho_j \binom{(j+1)a}{i} (-1)^i \exp(i+1) \sum_{s=0}^{\infty} \frac{t^s}{s!} \\ &\times \int_0^{\infty} x^s (1+\beta\,x)^{[(j+1)a+1]-1} \exp[-(i+1)(1+\beta\,x)^{\alpha}] dx \end{aligned}$$

Since the inner quantities of the summation are absolutely integrable, the integration and summation can be interchanged.

For $s > 0$ integer, it follows that

$$\int_0^{\infty} x^s (1+\beta\,x)^{[(j+1)a+1]-1} \exp[-(i+1)(1+\beta\,x)^{\alpha}] =$$
$$\frac{\beta^{-n-1}}{[(j+1)a+1]} \sum_{r=0}^{s} \binom{s}{r} \frac{(-1)^{s-r}}{(i+1)^{\frac{i}{[(j+1)a+1]}+1}} \Gamma\left(\frac{i}{[(j+1)a+1]}+1, i+1\right)$$

where $\Gamma(a,x) = \int_x^{\infty} z^{a-1} \mathrm{e}^{-z}\, \mathrm{dz}$ denotes the complementary incomplete gamma function, which can be evaluated in `MATHEMATICA`, `R`, etc. Then, the mgf of X can be expressed as

$$\begin{aligned} M(t) &= \alpha\,a \sum_{j,i,s}^{\infty} \sum_{r=0}^{s} \binom{(j+1)a}{i} \binom{s}{r} \frac{\rho_j\,\beta^r\,t^s\,(-1)^{s+j-r}}{s![(j+1)a+1](i+1)^{\{i/[(j+1)a+1]+1\}}} \\ &\times \Gamma\left(\frac{i}{[(j+1)a+1]}+1, i+1\right) \end{aligned} \tag{3.15}$$

Equation (3.15) is the main result of this section.

3.6 Order Statistics

Order statistics make their appearance in many areas of statistical theory and practice. The density function $f_{i:n}(x)$ of the ith order statistic, for $i = 1, \ldots, n$, from i.i.d. random variables $X_1, \ldots, X_n$ following any Kw-G distribution, is simply given by

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

Using (3.9) and (3.10), the pdf of $X_{i:n}$ can be expressed as

$$\begin{aligned} f_{i:n}(x) &= \frac{n!}{(i-1)!(n-i)!}\sum_{k=0}^{n-i}(-1)^k\binom{n-i}{k}\left[\sum_{j=0}^{\infty}\rho_j\, a\, g(x)G(x)^{(j+1)a-1}\right] \\ &\times \left[\sum_{j=0}^{\infty}\omega_j\, G(x)^{j\,a}\right]^{i+k-1}, \end{aligned}$$

where the coefficients ρ_j and ω_j are given in section 3.2.1, $h(x) = (j+1)a\, g(x)G(x)^{(j+1)a-1}$ and $H(x) = G(x)^{ja}$.

We use throughout the paper an equation of Gradshteyn and Ryzhik (2007, Section 0.314) for a power series raised to a positive integer j

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

where the coefficients $c_{j,i}$ (for $i = 1, 2, \ldots$) are easily obtained from the recurrence equation

$$c_{j,i} = (ia_0)^{-1}\sum_{m=i}^{i}[m(j+1) - i]\, a_m\, c_{j,i-m}$$

and $c_{j,0} = a_0^j$. The coefficients $c_{j,i}$ can be determined from $c_{j,0}, \ldots, c_{j,i-1}$ and then from the quantities $a_0, \ldots, a_i$ listed above. In fact, c_j, i can be given explicitly in terms of the coefficients a_i, although it is not necessary for numerically programming our expansions in any algebraic or numerical software.

Based on the above power series, we obtain

$$\left[\sum_{j=0}^{\infty}\omega_j\, G(x)^{j\,a}\right]^{i+k-1} = \sum_{j=0}^{\infty}\eta_{i+k-1,j}G(x)^{ja}$$

where $\eta_{i+k-1,0} = \kappa_0^{i+k-j}$ and $\eta_{i+k-1,j} = (\kappa\,\omega_0)^{-1}\sum_{m=1}^{j}[m(i+k) - j]\omega_m\, \eta_{i+k-1,j-m}$.

Thus, the pdf of $X_{i:n}$ re reduces to

$$f_{i:n}(x) = g(x)\sum_{j=0}^{\infty} m_{k,j}G(x)^{(2j+1)a-1} \tag{3.16}$$

where

$$m_{k,j} = \sum_{k=0}^{n-i}\frac{(-1)^k\, a\, n!\, \rho_j\, \eta_{i+k-1,j}}{(i-1)!(n-i)!}\binom{n-i}{k}$$

Equation (3.16) can be expressed as

$$f_{i:n}(x) = \sum_{k=0}^{n-i}\sum_{j=0}^{\infty} p_{k,j}h_{(2j+1)a-1}(x) \tag{3.17}$$

where

$$p_{k,j} = \frac{m_{k,j}}{(2j+1)a}$$

Equation (3.17) is the main result of this section. It reveals that the odf of the Kw-NH order statistics is a double linear combination of ENH densities with parameters $(2j+1)a, \alpha$ and β.

3.7 Maximum likelihood estimation

Several approaches for parameter estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The maximum likelihood estimators (MLEs) enjoy desirable properties and can be used when constructing confidence intervals for the parameters. The normal approximation for these estimators in large sample distribution theory is easily handled either analytically or numerically. In this section, the parameters of the new model are estimated by maximum likelihood. Let $x_1, \ldots, x_n$ be a random sample of size n from $X \sim$ Kw-NH(a, b, α, β). The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (a, b, \alpha, \beta)^\top$ can be expressed as

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & \; n \quad \{\log(a) + \log(b) + \log(\alpha) + \log(\beta)\} \, + \, (\alpha - 1) \sum_{i=0}^{n} \log(1 + \beta\, x_i) \\ + & \; (a-1) \sum_{i=0}^{n} \log\{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\} \\ + & \; (b-1) \sum_{i=0}^{n} \log(1 - \{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\}^a). \end{aligned}$$

The components of the score vector are given by

$$\boldsymbol{U_\theta} = (U_a, U_b, U_\alpha, U_\beta)^\top = \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial a}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial b}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta} \right)^\top,$$

where

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial a} &= \frac{n}{a} - \sum_{i=0}^{n} \frac{(b-1)\,\{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\}^a \log\{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\}}{1 - \{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\}^a} \\ &+ \sum_{i=0}^{n} \log\{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\}, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial b} &= \frac{n}{b} + \sum_{i=0}^{n} \log(1 - \{1 - \exp[1 - (1 + \beta\, x_i)^\alpha]\}^a), \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} &= \frac{n}{\alpha} - \sum_{i=0}^{n} \frac{(a-1)\,(1 + \beta\, x_i)^\alpha \, \log(1 + \beta\, x_i)}{1 - \exp(x_i - 1)} + \sum_{i=0}^{n} \log(1 + \beta\, x_i) \end{aligned}$$

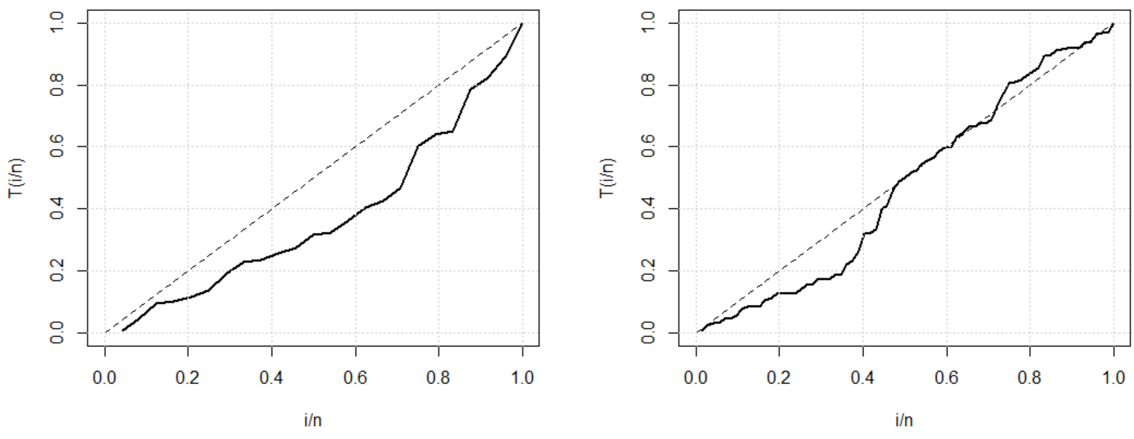
$$
\begin{aligned}
&- \sum_{i=0}^{n} \frac{a\,(1+\beta\,x_i)^{\alpha}\,\log(1+\beta\,x_i)\,\exp[1-(1+b\,x_i)^{\alpha}]}{1-\{1-\exp[1-(1+\beta\,x_i)^{\alpha}]\}^{a}} \\
&\times \{1-\exp[1-(1+\beta\,x_i)^{\alpha}]\}^{a-1}, \\
\frac{\partial\ell(\boldsymbol{\theta})}{\partial\beta} &= \frac{n}{\beta} + \sum_{i=0}^{n} \frac{x_i}{1+\beta\,x_i} - \sum_{i=0}^{n} \frac{(a-1)\,\alpha\,x_i\,(1+\beta\,x_i)^{\alpha-1}}{1-\exp[1-(1+\beta\,x_i)^{\alpha}]} \\
&- \sum_{i=0}^{n} \frac{a\,\alpha\,(1+\beta\,x_i)^{\alpha-1}\,\exp[1-(1+\beta\,x_i)^{\alpha}]\,\{1-\exp[1-(1+\beta\,x_i)^{\alpha}]\}^{a-1}}{1-\{1-\exp[1-(1+\beta\,x_i)^{\alpha}]\}^{a}}.
\end{aligned}
$$

Setting these equations to zero, $U(\boldsymbol{\theta}) = 0$ and solving them simultaneously yields the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. There exists many maximization methods in the R package like NR (Newton-Raphson), BFGS (Broyden-FletcherGoldfarb-Shanno), BHHH (Berndt-Hall-Hall-Hausman), SANN (Simulated-Annealing), and NM (Nelder-Mead). The MLEs are evaluated using the Limited Memory quasi-Newton code for Bound-constrained optimization (L-BFGS-B). Further, the Anderson-Darling (A^*) and Cramér-Von Mises (W^*) statistics are computed for comparing the fitted models. The computations are carried out using the R-package *AdequacyModel* given freely from
http://cran.r-project.org/web/packages/AdequacyModel/AdequacyModel.pdf.

3.8 Applications to real data

We perform two applications of the Kw-NH distribution to real data for illustrative purposes. We estimate the unknown parameters of the fitted distributions by the maximum-likelihood method as discussed in Section 3.7. The first example is a data set from Kus (2007) consisting of 24 observations on the period between successive earthquakes in the last century in the North Anatolia fault zone. For the second example, we consider the data corresponding to the exceedances of flood peaks (in m^3/s) of the Wheaton River near Carcross in Yukon Territoru, Canada. The data consist of 72 exceedances for the years 1958-1984, rounded to one decimal place. They are analysed by Choulakian and Stephens (2001). The use of the Kw-NH distribution for fitting these two data sets can be adequate. In many applications, there is qualitative information about the hrf, which can help in selecting a particular model. In this context, a device called the total time on test (**TTT**) plot by Aarset (1987) is useful. The **TTT** plot for the Earthquakes in North Anatolia fault zone data in Figure 3.4(a) indicates an upside-down bathtub hrf, whereas the **TTT** plot for the Oits IQ Scores data in Figure 3.4(b) reveals an increasing hrf. Therefore, these plots indicate the appropriateness of the Kw-NH distribution to fit these data, since the new model can present both forms of the hrf.

Table 3.1 provides some descriptive measures for the second data sets, which include central tendency statistics, standard deviation (SD), coefficient of variation (CV), skewness (S) and kurtosis (K), among others.



(a) (b)

Figure 3.4: TTT plots - (a) Earthquakes in North Anatolia fault zone data; (b) Flood Peaks Exceedances

Table 3.1: Descriptives statistics.

Statistic	Flood Peaks Exceedances	
	Earthquakes in North Anatolia fault zone	Oits IQ Scores
Mean	1430	12.2
Median	624.5	9.5
Mode	1125	6.23
SD	1980.7	12.3
CV	1.385	1.007
S	2.18	1.44
K	4.67	2.73
Minimum	9	0.1
Maximum	8592	64.0

We compare the fits of the Kw-NH distribution defined in (3.5), Kumaraswamy Weibull (Kw-We) by Cordeiro et al. (2010), gamma Nadarajah-Haghighi (GNH) by Bourguignon et al. (2015), ENH by Lemonte (2013), and NH by (3.3) to a real data set (for $x > 0$) with corresponding densities:

$$
\begin{aligned}
f_{Kw-We}(x) &= a\,b\,\alpha\,\beta^{\alpha}\,x^{\alpha-1}\exp[-(\beta\,x)^{\alpha}]\{1-\exp[-(\beta\,x)^{\alpha}]\}^{a-1}(1-\{1-\exp[-(\beta\,x)^{\alpha}]\}^{a})^{b-1},\\
f_{GNH}(x) &= \frac{a\,\beta}{\Gamma(a)}\,(1+\beta\,x)^{\alpha-1}\left[(1+\beta\,x)^{\alpha}-1\right]^{a-1}\exp\left\{1-(1+\beta\,x)^{\alpha}\right\},\\
f_{ENH}(x) &= a\,\alpha\,\beta(1+\beta\,x)^{\alpha-1}\exp\{1-(1+\beta\,x)^{\alpha}\}[1-\exp\{1-(1+\beta\,x)^{\alpha}\}]^{\beta-1},
\end{aligned}
$$

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

Table 3.2 lists the MLEs of the parameters (their standard errors are given in parentheses) of the unknown parameters of all lifetime models for the Earthquakes in North Anatolia fault zone data, whereas Table 3.3 does the some for the Exceedances of Wheaton River flood data.

The W^* and A^* statistics for all fitted models are presented in Table 3.4 for both data sets. First, note that the Kw-NH model fits the Earthquakes in North Anatolia fault zone data better than the other models according to the W^* and A^* statistics. On the other hand, the Kw-NH and GNH distributions provide better fits to the Exceedances of Wheaton River flood data according to the W^* statistic, although the Kw-NH distribution should be chosen according to he and A^* statistic. This implies that the Kw-NH could also be chosen as the best distribution for modeling both data sets.

Table 3.2: The MLEs (standard errors) of the model parameters for the Earthquakes in North Anatolia fault zone.

Model	Estimates			
Kw-NH(a,b,α,β)	1.3315	3.3696	0.2341	0.0037
	(0.3514)	(4.2522)	(0.1722)	(0.0020)
Kw-We(a,b,α,β)	13.9982	18.5321	0.1380	0.0357
	(19.2417)	(72.340)	(0.1487)	(0.2228)
GNH(a,α,β)	7.2396	0.2281	13.4234	
	(5.3929)	(0.0636)	(71.2482)	
ENH(a,α,β)	1.9491	0.3158	0.0183	
	(1.9026)	(0.1424)	(0.0471)	
NH(α,β)	0.5264	0.0026		
	(0.0909)	(0.0007)		

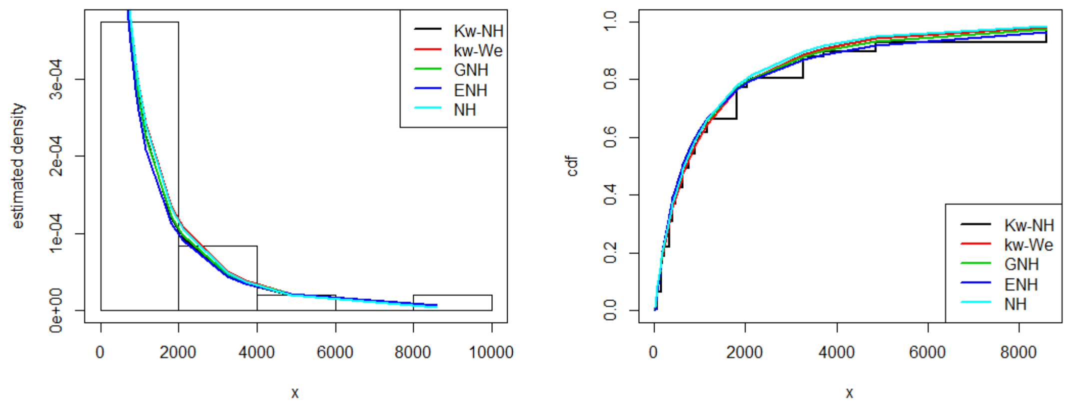
Table 3.3: The MLEs (standard errors) of the model parameters for the Flood Peaks Exceedances.

Model	Estimates			
Kw-NH(a,b,α,β)	0.7603	1.6358	2.2269	0.0143
	(0.1419)	(1.4156)	(1.7630)	(0.0177)
Kw-We(a,b,α,β)	0.5561	0.7458	1.2915	0.0685
	(0.3052)	(0.3394)	(0.4428)	(0.0375)
GNH(a,α,β)	0.7324	1.8714	0.0254	
	(0.1299)	(1.4536)	(0.028)	
ENH(a,α,β)	0.7309	1.6884	0.0316	
	(0.1373)	(1.1677)	(0.0318)	
NH(α,β)	0.8410	0.1094		
	(0.2599)	(0.0597)		

Plots of the estimated pdfs and cdfs of the Kw-NH, Kw-We, GNH, ENH and NH models fitted to both data sets are given in Figures 3.5 and 3.6. They indicate that the Kw-NH distribution is superior to the other distributions in terms of model fitting.

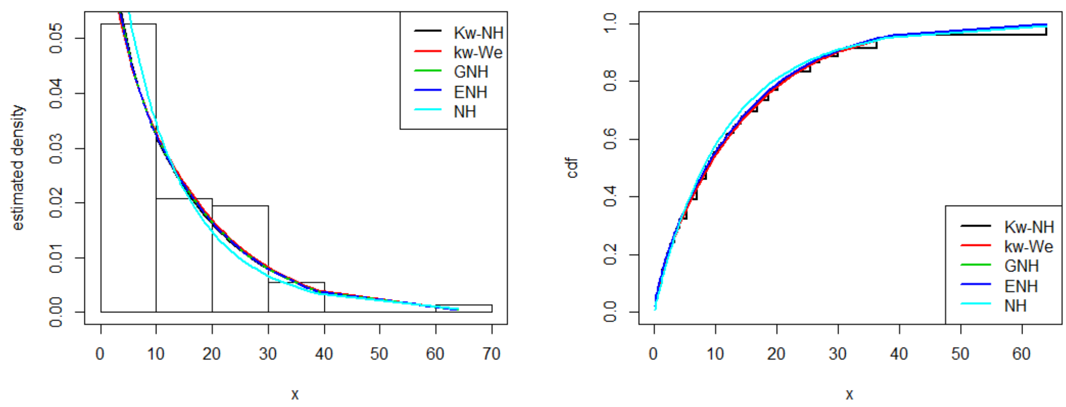
Table 3.4: W^* and A^* Statistics.

Distribution	Earthquakes in North Anatolia fault zone		Flood Peaks Exceedances	
	W^*	A^*	W^*	A^*
Kw-NH	**0.018**	**0.153**	**0.1026**	**0.621**
Kw-We	0.024	0.165	0.1076	0.656
GNH	0.023	0.193	0.1026	0.627
ENH	0.026	0.227	0.1027	0.628
NH	0.025	0.165	0.1442	0.817



(a) (b)

Figure 3.5: Estimated pdf and cdf from the fitted kw-NH, GNH, ENH and NH models for the earthquakes in North Anatolia fault zone data.



(a) (b)

Figure 3.6: Estimated pdf and cdf from the fitted kw-NH, GNH, ENH and NH models for the exceedances of flood peaks data.

3.9 Concluding remarks

The modeling and analysis of lifetimes is an important aspect of statistical work in a wide variety of scientific and technological fields. Continuous univariate distributions have been extensively used over the past decades for modeling data in several fields such as environmental and medical sciences, engineering, demography, biological studies, actuarial, economics, finance and insurance. However, in many applied areas such as lifetime analysis, finance and insurance, there is a clear need for extended forms of these distributions.

We introduced and studied the Kumaraswamy Nadarajah-Haghighi (Kw-NH) model to extend the Nadarajah-Haghighi (NH) and other distributions. We derive a linear representation for the density function and obtain explicit expressions for the ordinary and incomplete moments, quantile and generating function, mean deviations, and density function of the order statistics and their moments. The model parameters are estimated by maximum likelihood and the observed information matrix is determined. Two applications of the new model to real data sets reveal that the new model can be used quite effectively to provide better fits than its main sub-models. We hope that the proposed model may attract wider applications in Statistics.

References

AARSET, M. V. How to identify bathtub hazard rate. *IEEE Transactions Reliability*, v. 36, p. 106-108, 1987.

ALEXANDER, C. et al. Generalized beta-generated distributions. *Computational Statistics and Data Analysis*, v. 56, p. 1880-1897, 2012.

ALZAATREH, A.; FAMOYE, F.; LEE, C. Weibull-Pareto distribution and its applications. *Communications in Statistics: Theory & Methods*, v. 42, p. 1673-1691, 2013.

AMINI, M.; MIRMOSTAFAEE, S. M. T. K.; AHMADI, J. Log-gamma-generated families of distributions. *Statistics*, v. 48, p. 913-932, 2012.

BAILEY, C. *Smart Exercise: Burning Fat, Getting Fit.* 1rd ed., Boston, 1994, 304p.

BENKE, A.R.; WILMORE, J.H. *Evaluation and Regulation of Body Build and Composition.* Boston: University of Michigan, Prentice-Hall, 1974, 236p.

BLAND, J. M.; ALTMAN, D. G. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, v. 8, p. 135-160, 1999.

BLAND, J. M. *The half-normal distribution method for measurement error: two case studies.* Available in: <http://wwwu.sers.york.ac.uk/ mb55/talks/halfnor.pdf>. Access: 20 of Abr. 2005.

BOURGUIGNON, M., et al. A new generalized gamma distribution with applications, *American Journal of Mathematical and Management*, v. 34, p. 309-342, 2013.

BOURGUIGNON, M.; SILVA, R.B.; CORDEIRO, G.M. The Weibull-G family of probability distributions. *Journal of Data Science*, v. 12, p. 53-68, 2014.

BOURGUIGNON, M. et al. The Kumaraswamy Pareto distribution, *Journal of Statistics Theory and Applications*, v. 12, p. 129-144, 2013.

CHOULAKIAN, V.; STEPHENS, M.A. Goodness-of-fit for the generalized Pareto distribution. *Technometrics*, v. 43, p. 478-484, 2001.

CORDEIRO, G. M.; ORTEGA, E. M. M.; NADARAJAH, S. The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*. v. 347, P. 1174-1197, 2010.

CORDEIRO, G.M.; de CASTRO, M. A new family of generalized distributions. *Journal of Statistics Computation and Simulation*, v. 81, p. 883-898, 2011.

CORDEIRO, G.; PESCIM, R.; ORTEGA, E. The Kumaraswamy generalized half-normal distribution for skewed positive data. *Journal of Data Science*, v. 10, p. 195-224, 2012.

CORDEIRO, G. M.; NADARAJAH, S.; ORTEGA, E. M. M. The Kumaraswamy Gumbel distribution. *Statistical Methods and Applications*, v. 21, p. 139-168, 2012.

CORDEIRO, G. M.; ORTEGA E.M.M.; da CUNHA, D.C.C. The exponentiated generalized class of distributions. *Journal of Data Science*. v. 11, p. 1-27, 2013.

CORDEIRO, G.M.; ALIZADEH, M.; ORTEGA, E.M.M. The Exponentiated Half-Logistic Family of Distributions: Properties and Applications. *Journal of Probability and Statistics*, v. 2014, Article ID 864396, 21 p., 2014.

COORAY, K.;ANADA, M. A Generalization of the Half-Normal distribution with applications to lifetime data. *Communications in Statistics-Theory and Methods*, v. 37, p. 1323-1337, 2008.

CORREA, M. A.; NOGUEIRA, D. A.; FERREIRA, E. B. Kumaraswamy normal and Azzalini‘s skew normal modeling asymmetry. *Sigmae*, v. 1, p. 65-83, 2012.

de CASTRO, M.; et al. The Kumaraswamy generalized half-normal distribution for skewed positive data. *Journal of Data Science.*, v. 10, p. 195-224, 2012.

de PASCOA, M. A. R.; ORTEGA, E. M. M.; CORDEIRO, G. M.The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, v. 8, p. 411-433, 2011.

ELBATAL, I. Kumaraswamy exponentiated Pareto distribution. *Economic Quality Control*, v. 28, p. 1-8., 2013.

EUGENE ,N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. *Communications in Statistics. Theory and Methods*, v. 31, p. 497-512, 2002.

GRADSHTEYN, I.S.; RYZHIK, I.M. Table of Integrals, Series, and Products. *Academic Press, New York*, 2007.

GOMES, A. E., et al. A new lifetime model: the Kumaraswamy generalized Rayleigh distribution. *Journal of Statistical Computation and Simulation*, v. 84, p.290-309, 2014.

GUPTA, R.C.; GUPTA, P.L.; GUPTA R.D. Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and Methods*, v. 27, p. 887-904, 1998.

GURVICH, M.; DIBENEDETTO, A.; RANADE, S. A new statistical distribution for characterizing the random strength of brittle materials. *Journal of Materials Science*, v.32 , p. 2559-2564, 1997.

JONES, M. C. Families of distributions arising from distributions of order statistics. *Test*, v.13, p. 1 -43, 2004.

KATCH, F.; MCARDKE, W. *Nutrition, Weight Control, and Exercise.* Boston: University of Michigan, Lea & Febiger, 1983, 332p.

KENNEY, J.F.; KEEPING, E.S. *Mathematics of Statistics*, 3rd ed., Pt. 1. Princeton, NJ: Van Nostrand 1962, pp. 101-102.

KUS, C. A new lifetime distribution. *Computation Statistics and Data Analysis*, v. 9, p. 4497-4509, 2007.

LEMONTE, A.L. The beta log-logistic distribution. *Brazilian Journal of Probability and Statistics*, v. 28, p. 313-332, 2012.

LEMONTE, A.L. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, v. 62, p. 149-170, 2013.

MOORS, J.J.A. A quantile alternative for kurtosis. *Journal of the Royal Statistical Society D*, v. 37, p. 25-32, 1998.

MUDHOLKAR, G.S.; SRIVASTAVA, D.K. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, v. 42, p. 299-302, 1993.

MULDHOLKAR, G.S.; SRIVASTAVA, D.K.; FREIMER, M. The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, v. 37, p. 436-445, 1995.

NADARAJAH, S.; KOTZ,, S. The beta Gumbel distribution. *Mathematical Problems in engineering*, v.4, p. 323-332, 2006.

NADARAJAH, S. The exponentiated Gumbel distribution with climate application. *Environmetrics*, v. 17, p. 13-23, 2005

NADARAJAH, S.; KOTZ, S. The exponentiated type distributions. *Acta Applicandae Mathematicae*, v. 92, p. 97-111, 2006.

NADARAJAH, S.; GUPTA, A.K. The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin*, v. 59, p. 29-54, 2007.

NADARAJAH, S.; CORDEIRO, G.M.; ORTEGA, E.M.M. General results for the Kumaraswamy-G distribution. *Journal of Statistical Computation and Simulation*, v. 82, p. 951-979, 2011.

NADARAJAH, S.; HAGHIGHI, F. An extension of the expontial distribution, *Statistics*, v. 45, p. 543-558, 2011.

NADARAJAH, S.; CORDEIRO, G.M.; ORTEGA, E.M. The gamma-G family of distribution: Mathematical properties and applications. *Communications in Statistics: Theory* *and Methods*, To appear, 2012.

Paranaiba, P. F., et al. The Kumaraswamy Burr XII distribution: theory and pratice. *Journal of Statistical Computation and Simulation*, v. 83, p. 2217-2143, 2013.

RISTIC, M. M.; BALAHRISHNAN, N. The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, v. 82, p. 1191-1206, 2012.

ROBERTS, H.V. *Data analysis for managers with Minitab.* Redwood City, CA, Scientific Press, 1988, 366.

ROSAIAH, K.; KANTAM, R. R. L.; KUMAR, Ch. S. Reliability test plans for exponentiated log-logistic distribution. *Economic Quality Control*, v. 21, p. 279-289, 2006.

SAULO, H.; LEÃO, J.; BOURGUIGNON, M. The Kumaraswamy Birnbaum-Sanders distribution. *Journal Statistics Theory Practice*, v. 6, p. 754-759, 2012.

SHANS, T. M. The Kumaraswamy-Generalized Lomax distribution. *Middle-East Journal of Scientific Research*, v. 17, p. 641-646, 2013.

SIRI, W.E. *Advances in Biological and Medical Physics.* vol. IV, New York, Academic Press, 1956, 366p.

TAHIR, M. H. et al. McDonald Log-Logistic distribution. *Journal of Statistics Theory and Applications*, v. 13, p 65-82, 2014.

TORABI, H.; MONTAZARI, N. H. The gamma-uniform distribution and its application. *Kybernetika*, v. 48, p. 16-30, 2012.

TOBARI, H.; MONTAZARI, N.H. The logistic-uniform distribution and its application. *Communications Statistical Computation and Simulation*, v. 43, p. 2551-2569, 2013.

ZOGRAFOS, K; BALAKRISHNAN, N. On families of beta and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, v. 6, p. 344-362, 2009.

WILMORE, J. *Athletic Training and Physical Fitness: Physiological Principles of the Conditioning Process.* Boston: University of Michigan, Allyn and Bacon, 1977, 266p.