FEDERAL UNIVERSITY OF PERNAMBUCO
CENTRE FOR NATURAL AND EXACT SCIENCES
POSTGRADUATE PROGRAM IN STATISTICS

MATHEMATICAL PROPERTIES OF SOME GENERALIZED GAMMA MODELS

MARIA DO CARMO SOARES DE LIMA

Doctoral thesis

Recife
2015

Federal University of Pernambuco
Centre for Natural and Exact Sciences

Maria do Carmo Soares de Lima

MATHEMATICAL PROPERTIES OF SOME GENERALIZED GAMMA MODELS

Doctoral thesis submitted to the Post 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
2015

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da Silva, CRB4-1217

L732m Lima, Maria do Carmo Soares
Mathematical properties of some generalized Gamma models / Maria do Carmo Soares Lima. – Recife: O Autor, 2015.
151 f.: il., fig.. tab.

Orientador: Gauss Moutinho Cordeiro.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN. Estatística, 2015.
Inclui referências e apêndices.

1. Estatística matemática. 2. Distribuições exponencializadas. 3. Momentos ordinários. 4. Estatísticas de ordem. Cordeiro, Gauss Moutinho (orientador). II. Título.

519.5 CDD (23. ed.) UFPE- MEI 2015-07

MARIA DO CARMO SOARES DE LIMA

MATHEMATICAL PROPERTIES OF SOME GENERALIZED GAMMA MODELS

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

Aprovada em: 14/01/2015.

BANCA EXAMINADORA

____
Prof. PhD. Gauss Moutinho Cordeiro (Orientador)
Universidade Federal de Pernambuco

____
Prof. Dr. Manoel Raimundo Sena Junior (Examinador Interno)
Universidade Federal de Pernambuco

____
Prof. Dr. Abraão David Costa do Nascimento (Examinador Interno)
Universidade Federal de Pernambuco

____
Prof. Dr. Rodrigo Bernardo da Silva (Examinador Externo)
Universidade Federal da Paraíba

____
Prof. Dr. Cláudio Tadeu Cristino (Examinador Externo)
Universidade Federal Rural de Pernambuco

Dedico esta tese à minha mãe guerreira, ao meu irmão, à minha querida avó e (não poderia esquecer) ao meu amado esposo.

Agradecimentos

Quero agradecer primeiramente à Deus por mais essa vitória, numa longa carreira de anos de dedicação.

À minha amada e guerreira mãe, da qual tenho muito orgulho de ser filha, por sempre ter me apoiado na minha caminhada, estando ao meu lado em todas as minhas decisões. E ao me irmão pela confiança em diversos aspectos e pelo carinho. À minha avó, pelo seu amor, rezas e noites de preocupação.

Ao meu amado companheiro e esposo Flávio, a quem admiro como homem, marido e, principalmente como pai. Agradeço pelos incentivos e palavras certas, nas horas em que pensei em jogar tudo ao alto. Espero um dia poder retribuir um pouquinho do carinho e compreensão que você me deu.

Ao meu orientador Gauss Cordeiro pelo incentivo e confiança no meu trabalho.

Aos membros da banca, pelas críticas construtivas.

Aos professores com os quais tive a oportunidade de conviver, entre eles: Sérgio Santa Cruz, Airton Castro, Antônio Carlos, Adriano Pedrosa, Jalila Rios, Cleide Martins, Marcus Vinícius, Francisco Brito, William, Antônio (Tony), Marcos Barone, Audrey Cysneiros, Francisco Cribari, Francisco Cysneiros, Sérgio Bezerra, Klauss Vasconcellos, dentre outros.

Aos meus colaboradores: Edwin Ortega, Abraão Nascimento, Rodrigo Pescim, Marcelo Bourguignon, Cibele da Silva, Antônio Gomes, Morah Alizadeh, Jeremias Leão e Marcelino Pascoa.

Aos colegas que, direta ou indiretamente contribuíram nessa caminhada, dentre eles: Thiago Fiel, Luiz Silva, Marlon, João, Gilson Simões, Thiago Tanaka, Cláudia, Luiz Filipe, Vinícius, Roberto, Luz Milena, dentre outros.

Á Valéria pelo carinho, paciência e pela amizade que construímos juntas.

À CAPES, pelo apoio financeiro.

Resumo

Modelagem e análise de tempos de vida são aspectos importantes do trabalho estatístico, em uma ampla variedade de áreas científicas e tecnológicas. Estudamos algumas propriedades matemáticas de uma família recente chamada gama-G [Zografos and Balakrishnan (2009) and Ristić and Balakrishnan (2012)], denotada aqui por GG, em que G é chamada distribuição baseline. Escolhemos, como baselines, cinco distribuições amplamente conhecidas: Birnbaum-Saunders, Normal, Lindley, Nadarajah-Haghighi e uma extensão da Weibull. A mais recente, Nadarajah-Haghighi, foi estudada por Nadarajah e Haghighi (2011), que desenvolveram algumas propriedades interessantes. Demonstramos que as funções densidades das distribuições propostas podem ser expressas como combinação linear de funções densidades das respectivas exponencializadas-G. Para uma baseline arbitrária com cdf $G(x)$, uma variável aleatória é dita ter distribuição exponencializada-G, com parâmetro $a > 0$, digamos $X \sim \exp{-G(a)}$, se sua pdf e cdf são $h_a(x) = aG^{a-1}(x)g(x)$ and $H_a(x) = G^a(x)$, respectivamente. As propriedades de algumas exponecializadas têm sido estudadas por muitos autores, veja Mudholkar e Srivastava (1993) e Mudholkar *et al.* (1995) para Weibull exponencializada (exp-W), Gupta *et al.* (1998) para Pareto exponencializada, Gupta and Kundu (2001) para exponencial exponencializada (exp-E) e Nadarajah e Gupta (2007) para gama exponencializada (exp-G). Mais recentimente, Cordeiro *et al.* (2011a) investigaram algumas propriedades matemáticas para a distribuição gama generalizada exponencializada (exp-GG). Além disso, várias de suas propriedades estruturais são derivadas, incluindo expressões explícitas para os momentos, as funções quantílica e geratriz de momentos, desvios médios e dois tipos de entropia. Também investigamos as estatísticas de ordem e de seus momentos. Técnicas de máxima verossimilhança são usadas para ajustar os novos modelos e para mostrar a sua potencialidade.

Palavras-chave: Desvios médios. Distribuição Birnbaum-Saunders. Distribuição Extended Weibull. Distribuição Gamma-G. Distribuição Lindley. Distribuição Nadarajah-Haghighi. Distribuição Normal. Estimação por máxima verossimilhança. Função quantílica.

Abstract

The modeling and analysis of lifetimes are important aspects of statistical work in a wide variety of scientific and technological fields. We study some relevant mathematical properties of the recent family called gamma-*G* family [Zografos and Balakrishnan (2009) and Ristić and Balakrishnan (2012)], denoted here by "GG", for short, where *G* is called the baseline distribution. As baseline distributions, we choose five widely-known models: Birnbaum-Saunders, Normal, Lindley, Nadarajah-Haghighi and an extended Weibull. The fourth one is the most recent, studied by Nadarajah and Haghighi (2011), who developed some of its interesting properties. We demonstrate that the new density functions can be expressed as linear combination of exponentiated-G ("EG", for short) density functions. For an arbitrary baseline cdf $G(x)$, a random variable is said to have the exponentiated-*G* distribution with parameter $a > 0$, say $X \sim$ exp-G(a), if its pdf and cdf are $h_a(x) = aG^{a-1}(x)g(x)$ and $H_a(x) = G^a(x)$, respectively. The properties of some exponentiated distributions have been studied by several authors, see Mudholkar and Srivastava (1993) and Mudholkar *et al.* (1995) for exponentiated Weibull (exp-W), Gupta *et al.* (1998) for exponentiated Pareto, Gupta and Kundu (2001) for exponentiated exponential (exp-E) and Nadarajah and Gupta (2007) for exponentiated gamma (exp-G) distributions. More recently, Cordeiro *et al.* (2011a) investigated some mathematical properties for exponentiated generalized gamma (exp-GG) distribution. Further, various of their structural properties are derived, including explicit expressions for the moments, quantile and generating functions, mean deviations, probability weighted moments and two types of entropy. We also investigate the order statistics and their moments. Maximum likelihood techniques are used to fit the new models and to show their potentiality on real data set.

Keywords: Birnbaum-Saunders distribution. Extended Weibull distribution. Gamma-G distribution. Lindley distribution. Maximum likelihood estimation. Mean deviation. Nadarajah-Haghighi distribution. Normal distribution. Quantile function.

List of Figures

3.1 Plots of the GBS density for some parameter values. 30
3.2 Some types of the GBS hrf: (a) Increasing and decreasing hrf. (b) Unimodal hrf. (c) Bathtub hrf. 31
3.3 Skewness and kurtosis of the GBS distribution as functions of a for some values of α. 35
3.4 Skewness and kurtosis of the GBS distribution as functions of a for some values of β. 36
3.5 Bonferroni and Lorenz curves for the GBS distribution for some parameter values. 37

4.1 Plots of the new density function for some parameter values. 3(a) For different values of a with $\mu = 0$ and $\sigma = 1$. (b) For different values of a and σ with $\mu = 0$. (c) For different values of a, μ and σ. 53
4.2 (a) Skewness of X as function of a for some values of μ. (b) Skewness of X as function of a for some values of σ. 59
4.3 (a) Kurtosis of X as function of a for some values of μ. (b) Kurtosis of X as function of a for some values of σ. 59

5.1 Plots of the GL density and hazard functions for some parameter values. 80
5.2 Plots of the GL qf for $\lambda = a = k \in \{1/4, 1/2, 1, 4\}$. 83
5.3 Plots of the skewness and kurtosis measures for the GL distribution. 84
5.4 Plots of the Bonferroni curve for $\lambda = a = k \in \{1/4, 1/2, 1, 4\}$. 85
5.5 Plots of the theoretical and empirical densities for the GL(3,2) distribution. 90
5.6 Plots of the bias estimates for a in terms of λ in the GL distribution. 92
5.7 Application to real data. 93
5.8 Application to AIRSAR image. 96

6.1 Plots for the GNH pdf for some parameter values; $\lambda = 1$. 107
6.2 The GNH hrf for some parameter values; $\lambda = 1$. 109

6.3 Bonferroni and Lorenz curves for some parameter values. 112
6.4 Galton's skewness and Moor's kurtosis for the GNH distribution. 113
6.5 TTT plots – (a) stress carbon fibres data; (b) number of sucessive failures air conditioning system data. 118
6.6 (a) QQ plot with envelope for the GNH distribution and (b) fitted densities of the GNH (solid line), GG (dashed line), exp-NH (dotted line) and GEE (dotdash line) distributions for fibre data. 120
6.7 (a) QQ plot with envelope for the GNH distribution and (b) fitted densities of the GNH (solid line), GG (dashed line) exp-NH (dotted line) and GEE (dotdash line) distributions for number of successive failures for the air conditioning system. . 121

7.1 Plots of the GEW density. 132
7.2 Plots of the GEW hazard rate function. 133
7.3 Skewness of the GEW distribution for several choices of the parameters. 139
7.4 Kurtosis of the GEW distribution for several choices of the parameters. 140
7.5 Plots of the GEW density and sub-models for the warp breakage rates data (Tippett, 1950). 146

List of Tables

3.1 Descriptive statistics. 40
3.2 MLEs of the model parameters for the three data sets and the AIC, CAIC and BIC statistics. 42
3.3 LR tests. 42

4.1 MLEs and information criteria. 65
4.2 MLEs and information criteria. 66

5.1 Average of MLEs and their corresponding estimates for the MSE 91
5.2 Descriptive statistics from the real data set 92
5.3 MLEs and their standard errors based on the first real data set 92
5.4 Godness-of-fit measures based on a real data set 93
5.5 Descriptive statistics from the second real data set 94
5.6 MLEs and their standard errors based on the second real data set 94
5.7 Godness-of-fit measures based on a real data set 95

6.1 Descriptives statistics. 117
6.2 MLEs (standard erros in parenthesis). 119
6.3 Goodness-of-fit tests. 120

7.1 MLEs of the model parameters for the warp breakage rates data (Tippett, 1950), the corresponding SEs (given in parentheses) and the AIC measure. 145
7.2 LR tests for the warp breakage rates data (Tippett, 1950). 145

Contents

1 Introduction **14**

2 The gamma-G family of distributions **18**
- 2.1 Introduction 19
- 2.2 Expansions 20
- 2.3 Quantile function 21
- 2.4 Moments 21
- 2.5 Generating functions 22
- 2.6 Entropies 22
- 2.7 Order statistics 23

3 An extended Birnbaum-Saunders distribution **27**
- 3.1 Introduction 28
- 3.2 The new distribution 29
- 3.3 Useful expansions 30
- 3.4 Quantile Function 32
- 3.5 Moments 33
- 3.6 Generating function 34
- 3.7 Other Measures 35
  - 3.7.1 Mean deviations 35
  - 3.7.2 Reliability 36
- 3.8 Order statistics 37
- 3.9 Inference and estimation 38
  - 3.9.1 Maximum likelihood estimation 39
- 3.10 Applications 40
- 3.11 Concluding remarks 42
- 3.12 Appendix 43

4 A new extension of the normal distribution **50**
4.1 Introduction . . . 51
4.2 The GN distribution . . . 51
4.3 Useful expansions . . . 52
4.4 Quantile Function . . . 54
4.5 Moments . . . 55
4.6 Generating function . . . 58
4.7 Entropies . . . 60
4.8 Order statistics . . . 61
4.9 Estimation . . . 63
4.10 Applications . . . 64
4.10.1 Application 1: Carbohydrates data . . . 64
4.10.2 Application 2: Carbon monoxide data . . . 65
4.11 Concluding remarks . . . 66
4.12 Appendix . . . 67

5 The gamma Lindley distribution **77**
5.1 Introduction . . . 78
5.2 The gamma Lindley distribution . . . 78
5.3 Useful expansions . . . 79
5.4 Quantile Function . . . 81
5.5 Moments . . . 83
5.6 Generating function . . . 86
5.7 Order statistics . . . 86
5.8 Estimation . . . 88
5.9 Applications . . . 89
5.9.1 Simulation study . . . 89
5.9.2 Applications to real data . . . 91
5.10 Concluding remarks . . . 95
5.11 Appendix . . . 96

6 A new generalized gamma distribution **104**
6.1 Introduction . . . 105
6.2 The new distribution . . . 106
6.3 Useful expansions . . . 108
6.4 Quantile Function . . . 110
6.5 Moments . . . 111
6.6 Generating function . . . 113
6.7 Order statistics . . . 114
6.8 Entropies . . . 114
6.9 Maximum likelihood estimation . . . 115
6.10 Applications to real data . . . 117

6.11 Concluding remarks . . . 121
6.12 Appendix . . . 121

7 The Gamma Extended Weibull Distribution **128**

7.1 Introduction . . . 129
7.2 The GEW Distribution . . . 130
7.3 Useful expansions . . . 131
7.4 Quantile Function . . . 134
7.5 Generating Function . . . 136
7.6 Moments . . . 138
7.7 Mean Deviations . . . 141
7.8 Entropy . . . 141
7.9 Estimation . . . 143
7.10 Application . . . 144
7.11 Concluding remarks . . . 146
7.12 Appendix . . . 147

CHAPTER 1

Introduction

For any continuous baseline G distribution, Zografos and Balakrishnan (2009) proposed a generalized gamma-generated distribution (denoted here with the prefix "GG", for short) with an extra positive parameter. The central ideia is: let $X_{(U_1)}, \ldots, X_{(U_n)}$ be upper record values arising from a sequence of i.i.d. continuous random variables from a population with cumulative density function (cdf) $G(x)$ and probability density function (pdf) $g(x)$, where $1 \leq U_1 < \ldots < U_n \leq n$. Then, the pdf of the n$th$ upper record value, $X_{(U_n)}$, is given by

$$g_{X_{(U_n)}}(x) = \frac{g(x)}{(n-1)!} \left\{-\log\left[1 - G(x)\right]\right\}^{n-1}, \quad -\infty < x < \infty.$$

In the literature, the quantity $n \in \mathbb{N}$ such that $n \geq 1$ has been replaced with $a \in \mathbb{R}_+$, resulting (Zografos and Balakrishnan, 2009)

$$f(x) = \frac{g(x)}{\Gamma(a)} \left\{-\log\left[1 - G(x)\right]\right\}^{a-1}, \quad -\infty < x < \infty.$$

To that end, we employ the generator proposed by Zografos and Balakrishnan (2009). They studied some of its mathematical properties and presented some special cases. Here, we provide a comprehensive treatment of general mathematical properties of GG distributions by taking four different baselines. We discuss density expansions, quantile function, moments, incomplete moments, generating functions, entropies, order statistics, estimation of the model parameters by maximum likelihood and provide applications to real data sets. We present mathematical properties of four different models: Gamma Birnbaum-Saunders, Gamma Normal, Gamma Lindley and Gamma Nadarajah-Haghighi.

The first one uses the Birnbaum-Saunders as baseline distribution. Birnbaum and Saunders (1969a) pioneered a lifetime model which is commonly used in reliability studies. Based on this distribution, a new model called the gamma Birnbaum-Saunders distribution is proposed for describing fatigue life data. Several properties of this distribution including explicit expressions for the ordinary and incomplete moments, generating and quantile functions, mean deviations, density function of the order statistics and their moments are derived. We discuss the estimation of the model parameters by the method of maximum likelihood. The superiority of the new model is illustrated by means of three failure real data sets.

In the second chapter, we discuss the gamma normal distribution which has the normal distribution as the baseline. We study some of its mathematical properties too and maximum likelihood techniques are used in order to fit the new model and to show its potenciality by means of two examples of real data. Based on three criteria, the proposed distribution provides a better fit than the skew-normal distribution.

The third chapter is about a gamma Lindley (GL) distribution, which generalizes the Lindley model. We advance under three different aspects. First, the proposed model from the application of the Lindley distribution to the gamma generator, and the study of its structural properties are addressed. Secondly, the performance of the GL additional parameter estimation is quantified under variation of the Lindley (L) parameter by means of a simulation study. The new distribution can strongly be more flexible than the L model, when one wishes to analyze time between failures for repairable item. Thirdly, we consider the GL model for describing intensity data extracted from synthetic aperture radar (SAR) images. The *empirical distributions* (for more details, see, Gao (2010)) obtained from SAR data require specialized models since these data are corrupted by an interference pattern, called *speckle noise*(Oliver and Quegan, 1998). Results present meaningful evidence in favor of the GL model compared to the baseline Lindley, and, more important, the Weibull distribution – which has been indicated as a well-accepted model to describe empirical SAR distributions (Oliver and Quegan, 1998; Fernandes, 1998; de Fatima and Fernandes, 2000; Gao, 2010) – and the complementary exponential geometric (CEG) distribution (Louzada *et al.*, 2011)– like a recent bi-parametric extended model.

In the fourth chapter, we propose the gamma Nadarajah-Haghighi model, which is a new generalized gamma distribution. This distribution can be interpreted as a truncated generalized gamma distribution (Stacy, 1962). It can have a constant, decreasing, increasing, upside-down bathtub or bathtub-shaped hazard rate function depending on the values of its parameters. We demons-

trate that the new density function can be expressed as a linear combination of exponentiated Nadarajah-Haghighi density functions (Lemonte, 2013). Various of its structural properties are derived, including some explicit expressions for the moments, quantile and generating functions, skewness, kurtosis, mean deviations, Bonferroni and Lorenz curves, probability weighted moments and two types of entropy. We also obtain the order statistics. The method of maximum likelihood is used for estimating the model parameters and the observed information matrix is derived.

Finally, we propose, in the last chapter, the gamma extended Weibull model, which extend the Weibull and extended Weibull distributions among several other distributions. We obtain explicit expressions for the ordinary incomplete moments, generating and quantile functions, mean deviations, entropies and reliability. The method of maximum likelihood is used for estimating the model parameters. The applicability of the new model is illustrated by means of a real data set.

References

BIRNBAUM, Z.; SAUNDERS, S. A new family of life distributions. *Journal of Applied Probability*, v. 6, p. 319-327, 1969a.

DE FÁTIMA CARVALHO FERREIRA, A.; FERNANDES, D. Speckle filter for weibull-distributed sar images. *IEEE 2000 international geoscience and remote sensing symposium (igarss 2000)*, v. 2, p. 642–644, 2000.

FERNANDES, D. Segmentation of SAR images with Weibull distribution. *IEEE international geoscience and remote sensing symposium proceedings*, v. 1, 24–26, 1998.

GAO, G. Statistical modeling of SAR images: A survey. *Sensors*, v. 10, p. 775-795, 2010.

LEMONTE, A. J. 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.

LOUZADA, F.; ROMAN, M.; CANCHO, V. The complementary exponential geometric distribution: Model, properties, and a comparisonwith its counterpart. *Computational Statistics & Data Analysis*, v. 55, p. 2516–2524, 2011.

OLIVER, C.; QUEGAN, S. Understanding synthetic aperture radar images. The SciTech radar and defense series. *SciTech Publishing*, 1998

STACY, E. A generalization of the gamma distribution. *Annals of Mathematical Statistics*, v. 33, p. 1187–1192, 1962.

ZOGRAFOS, K.; BALAKRISHNAN, N. On Families of beta- and Generalized Gamma-generated Distributions and Associated Inference. *Statistical Methodology*, v. 6, p. 344–362, 2009.

CHAPTER 2

The gamma-G family of distributions

Resumo

Neste capítulo, falamos sobre o gerador gamma-G, com um parâmetro positivo adicional, proposto por Zografos e Balakrishnan (2009). Eles estudaram algumas propriedades matemáticas e apresentaram alguns casos especiais. Aqui, apresentamos algumas propriedades matemáticas gerais desse gerador, com o objetivo de, nos capítulos seguintes, mostrar todas essas propriedades sendo aplicadas a uma distribuição gamma-G, em que a *baseline* G é escolhida como uma determinada distribuição conhecida na literatura.

Palavras-chave: Baseline. Distribuição gama.

Abstract

In this chapter, we discuss about the gamma-G generator, with an extra positive parameter, proposed by Zografos e Balakrishnan (2009). They studied some of its mathematical properties and presented some special cases. Here, we present some general mathematical properties of gamma-G distributions, in order to, in the next chapters, show all these properties being applied to a gamma-G distribution, where the *baseline* G is chosen as a given distribution known in the literature.

Keywords: Baseline. Gamma distribution.

2.1 Introduction

Recently, attempts have been made to define new classes of lifetime distributions that provide greater flexibility in modeling skewed data in practice. In 2012, Torabi and Hedesh proposed a new general class of distributions (with two additional parameters), generated from the logit of the gamma random variable and they discussed some mathematical properties. Motivated by Torabi and Hedesh (2012), Cordeiro et al. introduced a new sub-family of the Zografos-Balakrishnan's family of distributions (paper submited in Applied Mathematical Modelling). In this paper, they introduced a gamma extended family of distributions with two extra generator parameters and studied some particular cases e properties of the new class. Amini *et al.*(2013) discussed two new families of distribution with two additional parameters, called Log Gamma-G I and Log Gamma-G II. In this paper the authors developed some mathematical properties of those new families.

Zografos and Balakrishnan (2009) and Ristić and Balakrishnan (2011) proposed a family of univariate distributions generated by gamma random variables. For any baseline cdf $G(x)$, $x \in \mathbb{R}$, they defined the gamma-G ("GG" for short) distribution with pdf $f(x)$ and cdf $F(x)$ given by

$$f(x) = \frac{g(x)}{\Gamma(a)} \left\{-\log\left[1 - G(x)\right]\right\}^{a-1} \tag{2.1}$$

and

$$F(x) = \frac{1}{\Gamma(a)} \int_0^{-\log[1-G(x)]} t^{a-1}\, \mathrm{e}^{-\mathrm{t}}\, \mathrm{dt}, \tag{2.2}$$

respectively, for $a > 0$, where $g(x) = dG(x)/dx$, $\Gamma(a) = \int_0^\infty t^{a-1}\,\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$ is the gamma function, $\gamma(a,z) = \int_0^z t^{a-1}\,\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$ denotes the incomplete gamma function and $\gamma_1(a,z) = \gamma(a,z)/\Gamma(a)$ is the incomplete gamma function ratio. Its hrf $h(x)$ is given by

$$h(x) = g(x)\, \left\{-\log\left[1 - G(x)\right]\right\}^{a-1} / \Gamma(a, -\log\left[1 - G(x)\right]),$$

where $\Gamma(a,z) = \int_z^\infty t^{a-1}\,\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$ denotes the complementary incomplete gamma function. The GG distribution has the same parameters of the G distribution plus an additional shape parameter $a > 0$. Each new GG distribution can be obtained from a specified G distribution. For $a = 1$, the G distribution is a basic exemplar with a continuous crossover towards cases with different shapes (for example, a particular combination of skewness and kurtosis). Nadarajah *et al.* (2013) derived several structural properties of the GG family of distributions, which hold for any G such as the asymptotic properties of (2.1) and (2.2), quantile function, ordinary and incomplete moments, generating function, mean deviations, asymptotic distribution of the extreme values, reliability and order statistics. Here, we intend to show some properties of GG distributions using four different baselines.

2.2 Expansions

Some useful expansions for (2.1) and (2.2) 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 distribution with parameter $a > 0$, say $X \sim$ exp-G(a), if its pdf and cdf are

$$h_a(x) = aG^{a-1}(x)g(x)$$

and

$$H_a(x) = G^a(x),$$

respectively. The properties of some exponentiated distributions have been studied by several authors, see Mudholkar and Srivastava (1993) and Mudholkar *et al.* (1995) for exponentiated Weibull (exp-W), Gupta *et al.* (1998) for exponentiated Pareto, Gupta and Kundu (2001) for exponentiated exponential (exp-E) and Nadarajah and Gupta (2007) for exponentiated gamma (exp-G) distributions. More recently, Cordeiro *et al.* (2011a) investigated some mathematical properties for exponentiated generalized gamma (exp-GG) distribution.

Nadarajah *et al.* (2013) used an expansion for the quantity $\{-\log\left[1-\Phi(v)\right]\}^{a-1}$ given by

$$\{-\log\left[1-\Phi(v)\right]\}^{a-1} = (a-1)\sum_{k=0}^{\infty}\binom{k+1-a}{k}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)}\,\Phi(v)^{a+k-1},$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,\mathrm{dt},$$

$a > 0$ is any real number and the quantities $p_{j,k}$ can be determined (for $j = 0, 1, 2, \ldots$ and $k = 1, 2, \ldots$) recursively by

$$p_{j,k} = k^{-1}\sum_{m=1}^{k}\frac{(-1)^m\left[m(j+1)-k\right]}{(m+1)}\,p_{j,k-m}, \tag{2.3}$$

and $p_{j,0} = 1$. For any real parameter $a > 0$, we define

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)} \tag{2.4}$$

and then (2.1) can be expressed as

$$f(x) = \sum_{k=0}^{\infty} b_k\,h_{a+k}(x), \tag{2.5}$$

where $h_{a+k}(x)$ denotes the exp-G density function with parameter $a+k$. The cdf corresponding to (2.5) becomes

$$F(x) = \sum_{k=0}^{\infty} b_k\,H_{a+k}(x), \tag{2.6}$$

where $H_{a+k}(x)$ denotes the exp-G cdf with parameter with parameter $a+k$. Based on equation (2.5), several structural properties of the GG distribution can be obtained by knowing those of the exp-G distribution.

2.3 Quantile function

Here, we use a result by Gradshteyn and Ryzhik (2007, Section 0.314) for a power series raised to a positive integer j

$$\left(\sum_{i=0}^{\infty} a_i\, x^i\right)^j = \sum_{i=0}^{\infty} c_{j,i}\, x^i, \tag{2.7}$$

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

$$c_{j,i} = (ia_0)^{-1} \sum_{m=1}^{i} [m(j+1) - i]\, a_m\, c_{j,i-m} \tag{2.8}$$

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

The GG qf, say $Q(u) = F^{-1}(u)$, can be expressed in terms of the G quantile function ($Q_G(\cdot)$). Inverting equation (2.2), it follows the qf of X as

$$F^{-1}(u) = Q_{GG}(u) = Q_G\left\{1 - \exp[-Q^{-1}(a, 1-u)]\right\}, \tag{2.9}$$

for $0 < u < 1$, where $Q^{-1}(a, u)$ is the inverse function of $Q(a, z) = 1 - \gamma(a, z)/\Gamma(a)$. Quantities of interest can be obtained from (2.9) by substituting appropriate values for u.

2.4 Moments

Let $Y \sim$exp-$G(a+k)$. A first formula for the nth moment of X can be obtained from (2.5) as

$$E(X^n) = \sum_{k=0}^{\infty} b_k\, E(Y^n).$$

Expressions for moments of several exponentiated distributions are given by Nadarajah and Kotz (2006), which can be used to produce $E(X^n)$.

A second formula for $E(X^n)$ can be obtained in terms of the baseline quantile function $Q_G(x) = G^{-1}(x)$. We obtain

$$E(X^n) = \sum_{k=0}^{\infty} (a+k)\, b_k\, \tau(n, a+k-1), \tag{2.10}$$

where the integral

$$\tau(n, a) = \int_{-\infty}^{\infty} x^n\, G(x)^a\, g(x) dx$$

can be expressed in terms of the G qf

$$\tau(n, a) = \int_0^1 Q_G(u)^n\, u^a du.$$

The nth incomplete moment of X is calculated as

$$m_n(y) = E\left(X^n | X < y\right) = \sum_{k=0}^{\infty} (a+k) b_k \int_0^{G(y)} Q_G(u)^n u^{a+k-1} du.$$

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

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 $\kappa_1 = \mu_1'$. Then, $\kappa_2 = \mu_2' - \mu_1'^2$, $\kappa_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$, $\kappa_4 = \mu_4' - 4\mu_3'\mu_1' - 3\mu_2'^2 + 12\mu_2'\mu_1'^2 - 6\mu_1'^4$, etc. 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. Other kinds of moments such L-moments may also be obtained in closed-form, but we consider only the previous moments for reasons of space.

2.5 Generating functions

A first formula for the mgf $M(t)$ of X comes from (2.5) as

$$M(t) = \sum_{k=0}^{\infty} b_k M_k(t),$$

where $M_k(t)$ is the mgf of Y_k. Hence, $M(t)$ can be immediately determined from the generating function of the exp-G distribution.

A second formula for $M(t)$ can be derived from (2.5) as

$$M(t) = \sum_{i=0}^{\infty} (a+k)\, b_k\, \rho(t, a+k-1), \tag{2.11}$$

where

$$\rho(t,a) = \int_{-\infty}^{\infty} \exp(tx)\, G(x)^a\, g(x) dx$$

can be calculated from the baseline qf $Q_G(x) = G^{-1}(x)$ by

$$\rho(t,a) = \int_0^1 \exp\left\{t\, Q_G(u)\right\} u^a du. \tag{2.12}$$

From equations (2.11) and (2.12) we can obtain the mgf′s of several gamma-G distributions directly.

2.6 Entropies

An entropy is a measure of variation or uncertainty of a random variable X. Two popular entropy measures are the Rényi and Shannon entropies (Shannon, 1948; Rached *et al.*,2001).

Following Nadarajah *et al.* (2013), the Rényi entropy, when X is a gamma-G random variable, is defined as

$$I_R(\gamma) = -\frac{\gamma \log \Gamma(a)}{1-\gamma} + \frac{1}{1-\gamma} \log \left\{ \sum_{k=0}^{\infty} \binom{k-\gamma a+\gamma}{k} \sum_{j=0}^{k} \frac{(-1)^{j+k} \binom{k}{j} p_{j,k}}{[\gamma(a-1)-j]} I_k \right\}, \tag{2.13}$$

where I_k comes from the baseline distribution as

$$I_k \; = \; \mathrm{E}\{G(Z)^{[(a-1)\gamma+k]} \, g(Z)^{\gamma-1}\} \; = \; \int_0^\infty G(x)^{[\gamma(a-1)+k]} \, g^{\gamma}(x) \, \mathrm{d}x,$$

where Z represents any continuous random variable with pdf and cdf given by $g(x)$ and $G(x)$, respectively.

Next, the Shannon entropy of a random variable X is defined by $E\{-\log[f(X)]\}$. It is a special case of the Rényi entropy when $\gamma \uparrow 1$. Equation (2.13) is very complicated for limiting, and then we derive an explicit expression for the Shannon entropy from its definition. We can write

$$\begin{aligned} \mathrm{E}\left[-\log f(X)\right] \;\; &= \;\; \log \Gamma(a) + (1-a) \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} \sum_{r=0}^{\infty} e_{j,r} \, \mathrm{E}\left[G^{r+j}(X)\right] \\ &\quad + \mathrm{E}\left\{\log\left[G(X)\right]\right\} - \mathrm{E}\left\{\log\left[g(X)\right]\right\}, \end{aligned}$$

where

$$\begin{aligned} \mathrm{E}\left[G^{r+j}(X)\right] \;\; &= \;\; \sum_{k=0}^{\infty} (a+k) \, b_k \int_0^\infty G^{a+r+j+k-1}(x) \, g(x) \, \mathrm{d}x \\ &= \;\; \sum_{k=0}^{\infty} \frac{(a+k) \, b_k}{(a+r+j+k)}, \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left\{\log\left[G(X)\right]\right\} \;\; &= \;\; \sum_{k=0}^{\infty} (a+k) \, b_k \int_0^\infty \log[G(x)] \, G^{a+k-1}(x) g(x) \, \mathrm{d}x \\ &= \;\; -\sum_{k=0}^{\infty} \frac{b_k}{a+k} \end{aligned}$$

and

$$\mathrm{E}\left\{\log\left[g(X)\right]\right\} \;\; = \;\; \sum_{k=0}^{\infty} (a+k) \, b_k \int_0^\infty \log[g(x)] \, G^{a+k-1}(x) g(x) \, \mathrm{d}x.$$

This last integral can be computed numerically for most baseline distributions.

2.7 Order statistics

Order statistics have been used in a wide range of problems, including robust statistical estimation and detection of outliers, characterization of probability distributions and goodness-of-fit tests, entropy estimation, analysis of censored samples, reliability analysis, quality control and strength of materials.

Suppose $X_1, \ldots, X_n$ is a random sample from the standard GG distribution and let $X_{1:n} < \cdots < X_{i:n}$ denote the corresponding order statistics. Using (2.5) and (2.6), the pdf of $X_{i:n}$ can be expressed as

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

Using (2.7) and (2.8), we can write

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

where $f_{j+i-1,0} = b_0^{j+i-1}$ and

$$f_{j+i-1,k} = (k\, b_0)^{-1} \sum_{m=1}^{k} [m(j+i) - k]\, b_m\, f_{j+i-1,k-m}.$$

Hence,

$$f_{i:n}(x) = \sum_{j=0}^{n-i} \sum_{r,k=0}^{\infty} m_{j,r,k}\, h_{a(j+i)+r+k}(x), \tag{2.14}$$

where

$$m_{j,r,k} = \frac{(-1)^j\, n!}{(i-1)!\,(n-i-j)!\, j!}\; \frac{(a+r)\, b_r\, f_{j+i-1,k}}{[a(j+i)+r+k]}.$$

Equation (2.14) is the main result of this section. It reveals that the pdf of the gamma-G order statistics is a triple linear combination of exp-G density functions. So, several mathematical quantities of the gamma-G order statistics like ordinary, incomplete and factorial moments, mgf, mean deviations and several others can be obtained from those quantities of gamma-G distributions. Clearly, the cdf of $X_{i:n}$ can be expressed as

$$F_{i:n}(x) = \sum_{j=0}^{n-i} \sum_{r,k=0}^{\infty} m_{j,r,k}\, H_{a(j+i)+r+k}(x).$$

References

AMINI, M.; MIRMOSTAFAEE, S.; AHMADI, J. Log-gamma-generated families of distributions. *Statistics: A Journal of Theoretical and Applied Statistics Publication details, including*, p. 37–41, 2013.

CORDEIRO, G. M.; ORTEGA, E.; SILVA, G. The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, v. 81, p. 827–842, 2011a.

CORDEIRO, G. M.; NADARAJAH, S.; ORTEGA, E. M. M. A New Subfamily of the Zografos-Balakrishnan's Family of Distributions. *Submited in Applied Mathematical Modelling*, 2014.

GRADSHTEYN, I.; RYZHIK, I. Table of integrals, series, and products, 2007. *New York: Academic*.

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

GUPTA, R.; KUNDU, D. Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal*, v. 43, p. 117–130, 2001.

MUDHOLKAR, G.; SRIVASTAVA, D. Exponentiated Weibull family for analyzing bathtub failure-real data. *IEEE Transaction on Reliability*, v. 42, p. 299–302, 1993.

MUDHOLKAR, G.; SRIVASTAVA, D.; FRIEMER, M. The exponential weibull family: A reanalysis of the bus-motor failure data. *Technometrics*, v. 37, p. 436–445, 1995.

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

Theory and Methods, 2013.

NADARAJAH, S.; GUPTA, A. A generalized gamma distribution with application to drought data. *Mathematics and Computer in Simulation*, v. 74, p. 1–7, 2007.

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

RACHED, Z.; ALAJAJI, F.; LORNE CAMPBELL, L. Renyi's divergence and entropy rates for finite al-phabet Markov sources. *IEEE Transactions on Information Theory*, v. 47, p. 1553–1561, 2001.

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

SHANNON, C. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, p. 379–423, 1948.

TORABI, H.; HEDESH, N. M. The gamma-uniform distribution and its applications. *Kybernetika*, v. 48, p. 16–30, 2012.

ZOGRAFOS, K.; BALAKRISHNAN, N. On Families of beta- and Generalized Gamma-generated Distributions and Associated Inference. *Statistical Methodology*, v. 6, p. 344–362, 2009.

CHAPTER 3

An extended Birnbaum-Saunders distribution

Resumo

Birnbaum e Saunders (1969a) foram os pioneiros de um modelo de vida que é comumente utilizado em estudos de confiabilidade. Com base nesta distribuição, um novo modelo chamado de distribuição gama-Birnbaum-Saunders é proposto para descrever dados de fadiga. Várias propriedades da nova distribuição, incluindo expressões explícitas para momentos ordinários incompletos, funções geradoras e de quantis, desvios médios, função densidade das estatísticas de ordem e os seus momentos são derivados. Discute-se o método de máxima verossimilhança para estimar os parâmetros do modelo. A superioridade do novo modelo é ilustrada por meio de três conjuntos de dados reais.

Palavras-chave: Dados de tempo de vida. Distribuição Birnbaum-Saunders. Distribuição de fadiga. Distribuição gama. Estimação de máxima verossimilhança.

Abstract

Birnbaum and Saunders (1969a) pioneered a lifetime model which is commonly used in reliability studies. Based on this distribution, a new model called the gamma-Birnbaum-Saunders distribution is proposed for describing fatigue life data. Several properties of the new distribution including explicit expressions for the ordinary and incomplete moments, generating and quantile functions, mean deviations, density function of the order statistics and their moments are derived. We discuss the method of maximum likelihood approach to estimate the model parameters. The superiority of the new model is illustrated by means of three failure real data sets.

Keywords: Birnbaum-Saunders distribution. Fatigue life distribution; Gamma distribution. Lifetime data. Maximum likelihood estimation.

3.1 Introduction

Motivated by problems of vibration in commercial aircraft that caused fatigue in the materials, Birnbaum and Saunders (1969a, 1969b) proposed the two-parameter Birnbaum-Saunders (BS) model, also known as the fatigue life distribution, with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, say BS(α, β). This distribution can be used to model lifetime data and it is widely applicable for modelling failure times of fatiguing materials. A random variable W having the BS(α, β) distribution is defined by

$$W = \beta \left[\frac{\alpha Z}{2} + \left\{ \left(\frac{\alpha Z}{2} \right)^2 + 1 \right\}^{1/2} \right]^2,$$

where Z is a standard normal random variable. Its cumulative distribution function (cdf) is given by

$$G(x) = \Phi(v), \ x > 0, \tag{3.1}$$

where $v(x) \equiv \alpha^{-1}\rho(x/\beta)$, $\rho(z) = z^{1/2} - z^{-1/2}$ and $\Phi(\cdot)$ is the standard normal cumulative function. The parameter β is the median of the distribution, i.e. $G(\beta) = \Phi(0) = 1/2$. For any $k > 0$, $k\,W \sim$ BS$(\alpha, k\beta)$. Kundu *et al.* (2008) investigated the shape of the BS hazard rate function (hrf). Results on improved statistical inference for this model are discussed by Wu and Wong (2004) and Lemonte *et al.* (2007, 2008). Further, Díaz-Garcia and Leiva (2005) proposed a new class of generalized BS distributions based on contoured elliptical distributions, whereas Guiraud *et al.* (2009) introduced a non-central version of the BS distribution. The probability density function (pdf) corresponding to (3.1) is given by

$$g(x) = k(\alpha, \beta)\, x^{-3/2}\, (x + \beta)\, \exp\left[-\frac{\tau(x/\beta)}{2\alpha^2} \right], \ x > 0, \tag{3.2}$$

where $k(\alpha, \beta) = (2\alpha e^{\alpha^2} \sqrt{2\pi\beta})^{-1}$ and $\tau(z) = z - z^{-1}$. The fractional moments of (3.2) are given by Rieck (1999) $\mathrm{E}(\mathrm{W}^p) = \beta^p\ \mathrm{I}(p, \alpha)$, where

$$\mathrm{I}(p, \alpha) = \frac{\kappa_{p+1/2}(\alpha^{-2}) + \kappa_{p-1/2}(\alpha^{-2})}{2\kappa_{1/2}(\alpha^{-2})}, \tag{3.3}$$

and $\kappa_\nu(z) = 0.5 \int_{-\infty}^{\infty} \exp\{-z\cosh(t) - \nu t\} \mathrm{d}t$ denotes the modified Bessel function of the third kind with ν representing its order and z the argument. A discussion of this function can be found in Watson (1995).

Here, we propose a new lifetime model called the *gamma Birnbaum-Saunders* (GBS) distribution to extend the BS model. We know the hrf plays an important role in lifetime data analysis.

The shape of the hazard function of Birnbaum-Saunders distribution is unimodal. Thus, we will propose a new model that has four forms of hrf, as follows. We provide properties of the new distribution, discuss maxi-
mum likelihood estimation of the model parameters and derive the observed information matrix. The rest of the chapter is outlined as follows. In Section 3.2, we discuss about the new distribution. We derive useful expansions in Section 3.3. The qf is thoroughly discussed in Section 3.4. In Sections 3.5 and 3.6, we obtain the ordinary moments and two representations for the mgf of X, respectively. In Section 3.7, we derive the mean deviations, Bonferroni and Lorenz curves and reliability. The order statistics are investigated in Section 3.8. In Section 3.9, we present the method of maximum likelihood. In Section 3.10, one application to real data sets is presented to demonstrate the potentiality of the new distribution for fatigue life modeling and the flexibility and practical relevance. Finally, Section 3.11 ends with some concluding remarks.

3.2 The new distribution

The pdf and cdf of the GBS distribution are defined (for $x > 0$) by applying (3.1) and (3.2) in equations (2.1) and (2.2)

$$f(x) = \frac{k(\alpha, \beta)}{\Gamma(a)}\, x^{-3/2}\,(x+\beta)\,\exp\left[-\frac{\tau(x/\beta)}{2\alpha^2}\right]\{-\log\left[1-\Phi(v)\right]\}^{a-1} \tag{3.4}$$

and

$$F(x) = \frac{1}{\Gamma(a)} \int_0^{-\log[1-\Phi(v)]} t^{a-1}\, \mathrm{e}^{-\mathrm{t}}\, \mathrm{dt}, \tag{3.5}$$

respectively. Evidently, the density function (3.4) does not involve any complicated function and the BS distribution arises as the basic exemplar for $a = 1$. The fact that the GBS distribution extends the BS distribution is also a positive point. Figure 3.1 displays some possible shapes of the density function (3.4) for selected parameter values. It is evident that the GBS distribution is much more flexible than the BS distribution.

The hazard rate function (hrf) corresponding to (3.4) is given by

$$h(x) = \frac{k(\alpha,\beta)\, x^{-3/2}\,(x+\beta)\,\exp\left[-\frac{\tau(x/\beta)}{2\alpha^2}\right]\{-\log\left[1-\Phi(v)\right]\}^{a-1}}{\Gamma(a)\,\Gamma(a, -\log\left[1-\Phi(v)\right])}.$$

Hereafter, a random variable X having density function (3.4) is denoted by $X \sim \mathrm{GBS}(\alpha, \beta, a)$.

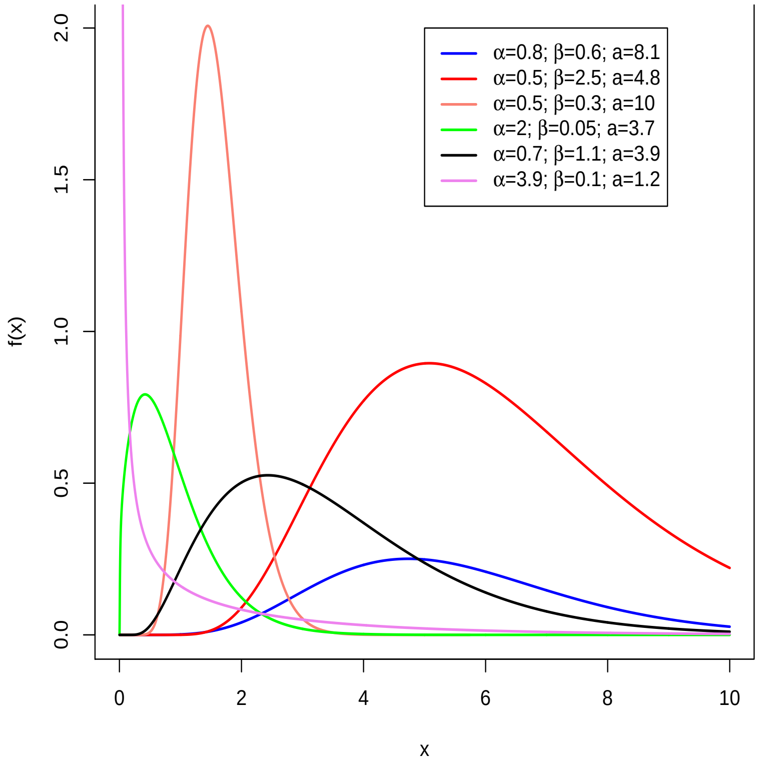


Figure 3.1: Plots of the GBS density for some parameter values.

Plots of the hrf $h(x)$ for selected parameter values are displayed in Figure 3.2. We note that these plots illustrate the four types of hazard shapes. The new model is easily simulated as follows: if V is a gamma random variable with parameter $a > 0$, then

$$X = \beta \left\{ \frac{\alpha\, \Phi^{-1}(V)}{2} + \left[\left(\frac{\alpha\, \Phi^{-1}(V)^2}{4} \right) + 1 \right]^{1/2} \right\}^2$$

has the GBS(α, β, a) distribution. This scheme is useful because of the existence of fast generators for gamma random variables and the standard normal quantile function is available in several statistical packages.

3.3 Useful expansions

Expansions for (3.4) and (3.5) can be derived using the concept of exponentiated distributions. Cordeiro *et al.* (2013) defined the *exponentiated Birnbaum-Saunders* (exp-BS) distribution with positive parameters α, β and c, say $Y \sim$ exp-BS(α, β, c), if its cdf and pdf are given by $H(y; \alpha, \beta, c) = \Phi(v)^c$ and $h(y; \alpha, \beta, c) = c\, g(y)\, \Phi(v)^{c-1}$, respectively, where v is defined in (3.1) and g is given in (3.2).

From (3.4), we can write

$$f(x) = \sum_{k=0}^{\infty} b_k\, h_{a+k}(x), \tag{3.6}$$

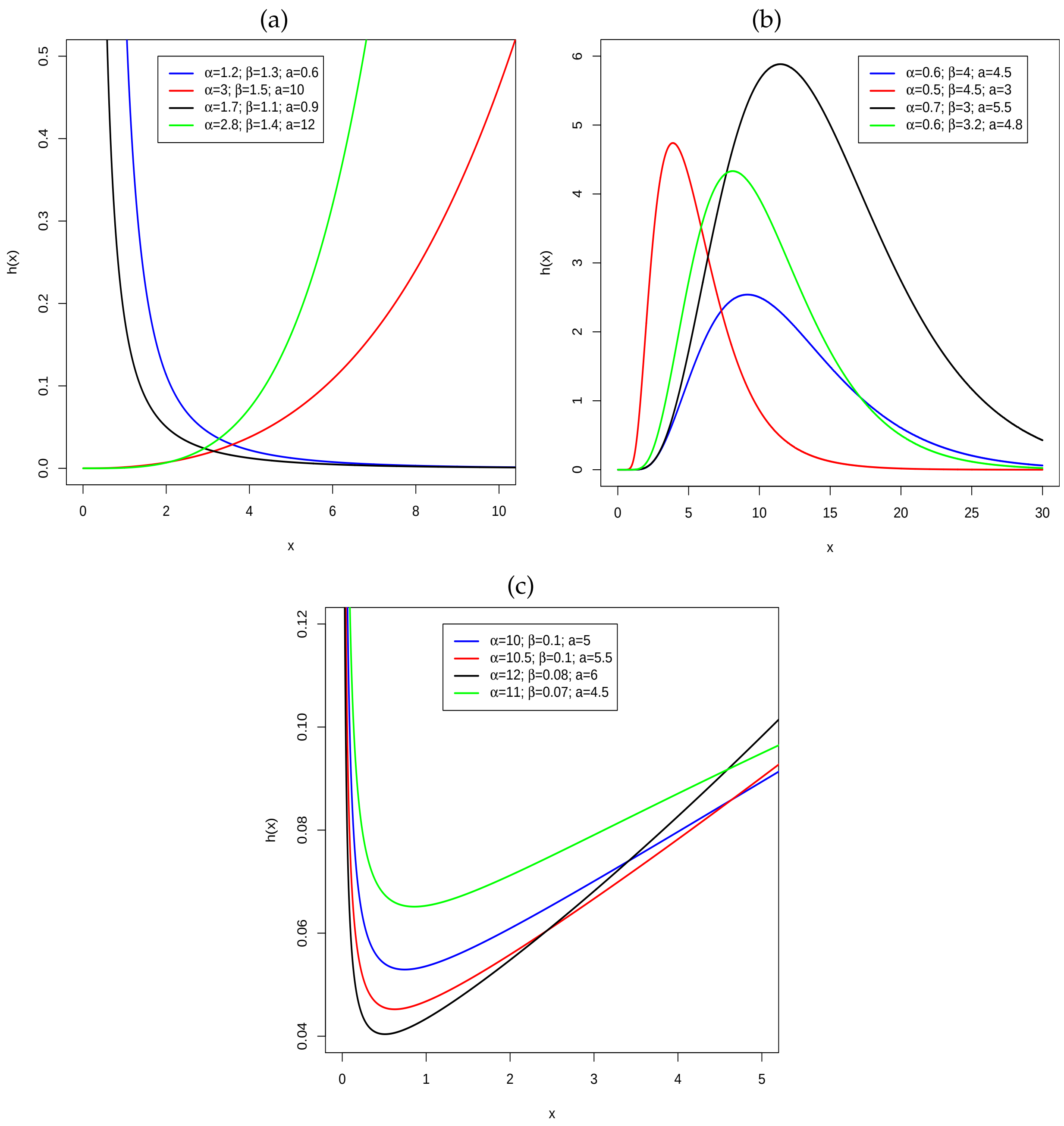


Figure 3.2: Some types of the GBS hrf: (a) Increasing and decreasing hrf. (b) Unimodal hrf. (c) Bathtub hrf.

where

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)} \sum_{j=0}^{k} \frac{(-1)^{j+k}\binom{k}{j}\, p_{j,k}}{(a-1-j)}$$

and $h_{a+k}(x)$ denotes the exp-BS$(\alpha, \beta, a+k)$ density function. The cdf corresponding to (3.6) becomes

$$F(x) = \sum_{k=0}^{\infty} b_k\, H_{a+k}(x) = \sum_{k=0}^{\infty} b_k\, \Phi(v)^{a+k}, \tag{3.7}$$

where $H_{a+k}(x) = \Phi(v)^{a+k}$ denotes the exp-BS cdf with parameters α, β and $(a+k)$. Based on equation (3.6), several structural properties of the GBS distribution can be obtained by knowing those of the exp-BS distribution.

If a is a nonnegative integer, we can expand $\Phi(v)^{a+k}$ as

$$\Phi(v)^{a+k} = \sum_{r=0}^{\infty} s_r(a+k)\, \Phi(v)^r, \tag{3.8}$$

where

$$s_r(a) = \sum_{l=r}^{\infty} (-1)^{r+l} \binom{a}{l} \binom{l}{r}.$$

Thus, from equations (3.2), (3.6) and (3.8), we obtain

$$f(x) = g(x) \sum_{r=0}^{\infty} d_r\, \Phi(v)^r, \tag{3.9}$$

where $d_r = \sum_{k=0}^{\infty} b_k\, s_r(a+k)$. Equations (3.6) and (3.9) are the main results of this section.

3.4 Quantile Function

The GBS quantile function (qf), say $Q(u) = F^{-1}(u)$, can be expressed in terms of the BS qf ($Q_{BS}(\cdot)$) and beta qf ($Q_\beta(\cdot)$). The BS qf is straightforward computed from the standard normal qf $x = Q_N(u) = \Phi^{-1}(u)$ (Cordeiro and Lemonte, 2011)

$$Q_{BS}(u) = \frac{\beta}{2} \left\{ 2 + \alpha^2\, Q_N(u)^2 + \alpha\, Q_N(u) \left[4 + \alpha^2\, Q_N(u)^2\right]^{1/2} \right\}.$$

Inverting $F(x) = u$, we obtain the qf of X as

$$F^{-1}(u) = Q_{GBS}(u) = Q_{BS} \left\{ 1 - \exp[-Q^{-1}(a, 1-u)] \right\}, \tag{3.10}$$

for $0 < u < 1$, where $Q^{-1}(a, u)$ is the inverse function of $Q(a, z) = 1 - \gamma(a, z)/\Gamma(a)$. Quantities of interest immediately follow from (3.10) by substituting appropriate values for u. Further, the BS qf can be expressed as

$$Q_{BS}(u) = \sum_{i,j=0}^{\infty} \overline{p}_j\, h_{j,i} u^{(i+j)/a}, \tag{3.11}$$

where $\overline{p}_j = \sum_{k=j}^{\infty} m_k \left(\frac{-1}{2}\right)^{k-j} \binom{k}{j}$, $h_{j,i} = (iv_0)^{-1} \sum_{m=0}^{i} [m(j+1) - i]\, v_m\, h_{j,i-m}$, $v_i = q_{i+1}$. Here, $q_0 = 0$, $q_1 = 1$, $q_2 = (\beta - 1)/(\alpha + 1), \ldots$, and the quantities q_i's (for $i \geq 2$) can be derived from a cubic recursive formula given in Appendix A.

We can obtain the inverse function $Q^{-1}(a, u)$ in the Wolfram website as

$$z = Q^{-1}(a, 1-u) = \sum_{i=0}^{\infty} a_i\, u^{i/a},$$

where $a_0 = 0$, $a_1 = \Gamma(a+1)^{1/a}$, $a_2 = \frac{\Gamma(a+1)^{2/a}}{(a+1)}$, $a_3 = \frac{(3a+5)\Gamma(a+1)^{3/a}}{2(a+1)^2(a+2)}, \ldots$, etc. By expanding the exponential function and using (2.7), we have

$$1 - \exp\left(-\sum_{r=0}^{\infty} a_r u^{r/a}\right) = 1 - \sum_{r=0}^{\infty} p_r\, u^{r/a},$$

where the p_r's are defined in Appendix A. Then,

$$Q_{GBS}(u) = Q_{BS}\left(1 - \sum_{r=0}^{\infty} p_r\, u^{r/a}\right).$$

After some algebra, using the power series expansion (2.7), we obtain from (3.11)

$$Q_{GBS}(u) = \sum_{i,j=0}^{\infty} \overline{p}_j\, h_{j,i} \sum_{s=0}^{\infty} (-1)^s \binom{(i+j)/a}{s} \sum_{r=0}^{\infty} d_{s,r}\, u^{r/a} = \sum_{r=0}^{\infty} \overline{c}_r\, u^{r/a}, \tag{3.12}$$

where the coefficients $d_{s,r}$ can be determined from (2.8) as $d_{s,r} = (rp_0)^{-1} \sum_{v=1}^{r} [v(s+1) - r]\, p_v\, d_{s,r-v}$ for $s \geq 0$, $r \geq 1$, $d_{s,0} = p_0^s$ and $\overline{c}_r$ is defined by

$$\overline{c}_r = \sum_{i,j,s=0}^{\infty} (-1)^s \binom{(i+j)/a}{s} \overline{p}_j\, h_{j,i}\, d_{s,r}.$$

Some algebraic details about (3.12) are given in Appendix A. Equation (3.12) can be used to obtain some mathematical measures of X such as the moments and generating function by integration over $[0, 1]$.

3.5 Moments

The ordinary moments of X can be derived from the PWMs (Greenwood *et al.*, 1979) of the BS distribution defined for p and r non-negative integers by

$$\tau_{p,r} = \int_0^{\infty} x^p\, \Phi(v)^r\, g(x) dx. \tag{3.13}$$

The integral (3.13) can be computed numerically in softwares such as MAPLE, MATLAB, MATHEMATICA, Ox or R. Cordeiro and Lemonte (2011) proposed an alternative representation to obtain $\tau_{p,r}$ as

$$\begin{aligned}\tau_{p,r} &= \frac{\beta^p}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j}^{\infty} A(k_1, \ldots, k_j) \sum_{m=0}^{2s_j+j} (-1)^m \binom{2s_j + j}{m} \\ &\times\ I\left(p + \frac{(2s_j + j - 2m)}{2}, \alpha\right), \end{aligned} \tag{3.14}$$

where $s_j = k_1 + \ldots + k_j$, $A(k_1, \ldots, k_j) = \alpha^{-2s_j - j}\,\overline{a}_{k_1}, \ldots, \overline{a}_{k_j}$, $\overline{a}_k = (-1)^k\, 2^{(1-2k)/2}\,[\sqrt{\pi}(2k+1)]^{-1}$ and $I(p + (2s_j + j - 2m)/2, \alpha)$ is determined from equation (3.3). Those algebraic softwares have currently the ability to deal with analytic expressions of formidable size and complexity.

The sth moment of X can be expressed from (2.5) as

$$\mu'_s = \sum_{r=0}^{\infty} d_r\, \tau_{s,r}, \tag{3.15}$$

where $\tau_{s,r}$ comes from (3.14) and d_r is given by (3.9). Equation (3.15) can be computed numerically in any symbolic mathematical software. Plots of the skewness and kurtosis, for selected values of α and β, as functions of a are displayed in Figures 3.3 and 3.4, respectively.

Next, we obtain the nth incomplete moment of X defined as $T_n(y) = \int_0^y x^n f(x) dx$. The incomplete moments play an important role for measuring inequality, for example, income quantiles and Lorenz and Bonferroni curves. Substituting (2.5) in the last equation, we obtain

$$T_n(y) = k(\alpha, \beta) \sum_{r=0}^{\infty} d_r \int_0^y (x + \beta)\, x^{n-3/2} \exp\left[-\frac{\tau(x/\beta)}{2\alpha^2}\right] \Phi(v)^r dx.$$

Using the expansion for $\Phi(v)^r$ given by Cordeiro and Lemonte (2011), $T_n(y)$ can be expressed as

$$\begin{aligned} T_n(y) \;=\; & k(\alpha, \beta) \sum_{r=0}^{\infty} \frac{d_r}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \beta^{-(2s_j+j)/2} A(k_1, \ldots, k_j) \sum_{m=0}^{2s_j+j} (-\beta)^m \binom{2s_j + j}{m} \\ & \times \left[D\left(n + \frac{2s_j + j - 2m - 1}{2}, y\right) + \beta\, D\left(n + \frac{2s_j + j - 2m - 3}{2}, y\right)\right], \end{aligned} \tag{3.16}$$

where the quantity $D(p, q)$ is defined in Appendix B. Equation (3.16) is the main result of this section.

The Bonferroni and Lorenz curves of X are defined as $B(\pi) = T_1(q)/[\pi\mu'_1]$ and $L(\pi) = T_1(q)/\mu'_1$, respectively, where $q = Q_{GBS}(\pi)$ comes from the qf (3.12) for a given probability π and (3.16) with $n = 1$. These two curves have applications in economics, reliability, demography, insurance and medicine; see for example Ordu *et al.* (2011) and Cuena and Seidl (2007). Plots of the Bonferroni and Lorenz curves versus π for some choices of a, $\alpha = 0.5$ and $\beta = 1.1$ are displayed in Figures 3.5a and 3.5b, respectively.

3.6 Generating function

We provide two representations for the mgf of X, say $M(s) = \mathrm{E}(\mathrm{e}^{sX})$. From equation (3.6), we obtain a first expansion

$$M(s) = \sum_{k=0}^{\infty} b_k\, (a + k) \int_0^{\infty} \mathrm{e}^{st} G(t)^{a+k-1} g(t) dt.$$

By expanding the exponential function and the qf, $M(s)$ can be rewritten as

$$M(s) = \sum_{n,s,k=0}^{\infty} \frac{(a + k)\, b_k\, \overline{v}_{s,n}}{(s/a + a + k)}\, \frac{s^n}{n!},$$

where $\overline{v}_{s,n} = (n\overline{q}_0)^{-1}\sum_{m=0}^{s}[m(s+1)-j]\,\overline{q}_m\,\overline{v}_{s,n-m}$ follows from (2.7) and (2.8) with $\overline{v}_{s,0} = \overline{q}_0^s$. Further, $\overline{q}_k$ comes from the expansion of $Q_{BS}(u)$ given by $Q_{BS}(u) = \sum_{s=0}^{\infty}\overline{q}_s u^{s/a}$, whose coefficients and the details of the proof are given in Appendix C.

A second representation for $M(s)$ is determined from the exp-BS generating function. We can write $M(s) = \sum_{k=0}^{\infty} b_k\, M_k(t)$, where b_k is given by (2.4) and $M_k(t)$ is the mgf of Y_k ∼exp-BS$(a+k)$ given by

$$M_k(t) = \sum_{r,k=0}^{\infty} \frac{(-1)^r\,\Gamma(a+1)\,\tau_{k,a+r-1}}{k!}\,t^k.$$

3.7 Other Measures

In this section, we derive the means deviations and the reliability of X.

3.7.1 Mean deviations

The mean deviations about the mean μ_1' and about the median M can be written as

$$\delta_1 = 2\big[\mu_1'\,F(\mu_1') - T_1(\mu_1')\big] \qquad \text{and} \qquad \delta_2 = \mu_1' - 2T_1(M),$$

where $T_1(\omega)$ is the first incomplete moment of X computed from (3.16) with $n = 1$. Then, the measures δ_1 and δ_2 are determined from this equation.

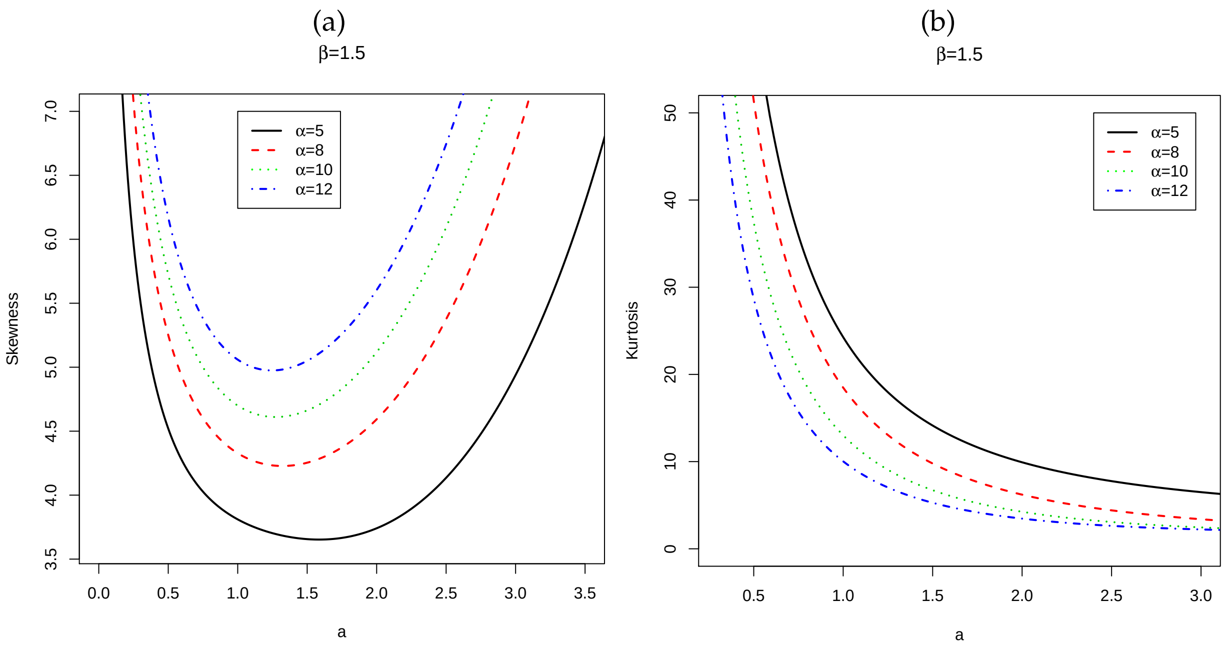


Figure 3.3: Skewness and kurtosis of the GBS distribution as functions of a for some values of α.

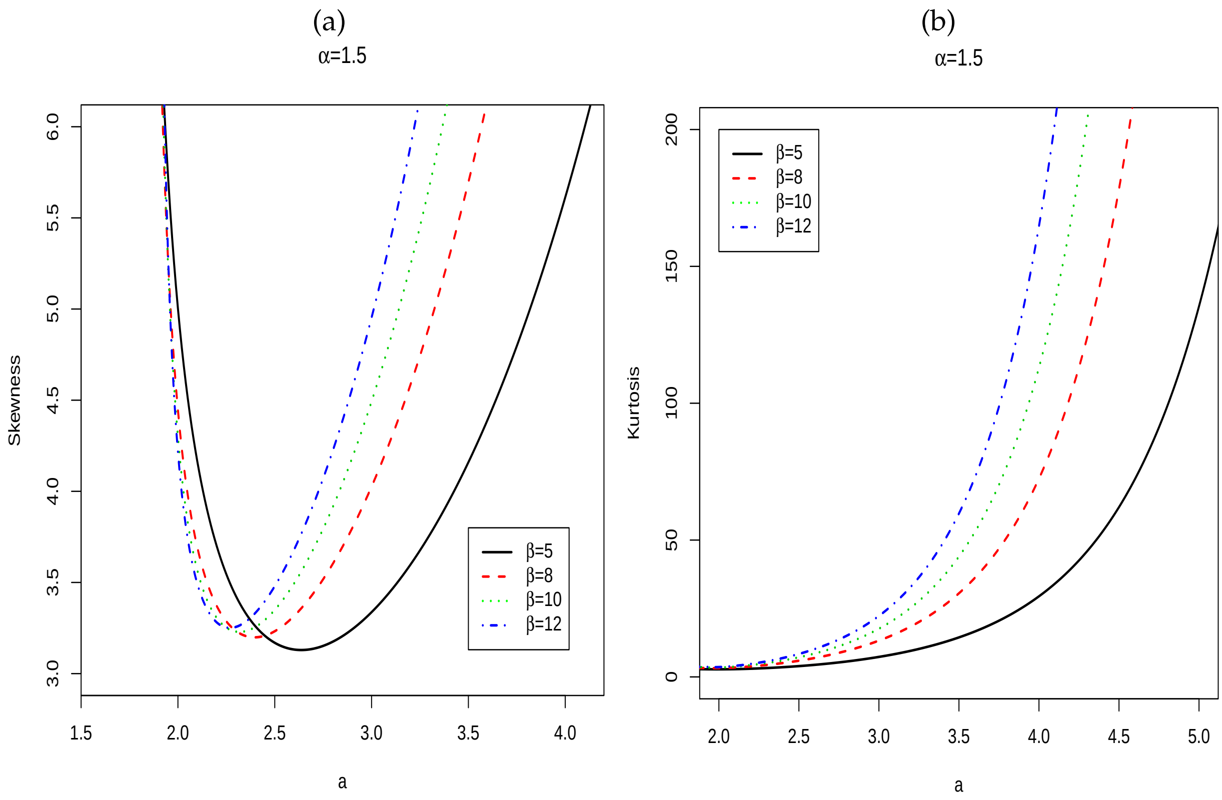


Figure 3.4: Skewness and kurtosis of the GBS distribution as functions of a for some values of β.

3.7.2 Reliability

Consider the life of a component which has a random strength X_1 subjected to a random stress X_2. The component fails at the instant that the stress applied to it exceeds the strength, and then a measure of component reliability is $R = Pr(X_1 < X_2) = \int_0^\infty f_1(x)\,F_2(x)dx$. We derive R when X_1 and X_2 have independent GBS(α, β, a_1) and GBS(α, β, a_2) distributions with the same shape parameters α and β. The pdf of X_1 and the cdf of X_2 can be obtained from (3.6) and (3.7) as

$$f_1(x) = g(x) \sum_{k=0}^{\infty} b_{1k}\,\Phi(v)^{a_1+k} \quad \text{and} \quad F_2(x) = \sum_{j=0}^{\infty} b_{2j}\,\Phi(v)^{a_2+j},$$

respectively, where

$$b_{1k} = \frac{\binom{k+1-a_1}{k}}{(a_1+k)\,\Gamma(a_1-1)} \sum_{i=0}^{k} \frac{(-1)^{i+k}\binom{k}{i}\,p_{i,k}}{(a_1-1-i)}, \quad b_{2j} = \frac{\binom{j+1-a_2}{j}}{(a_2+j)\,\Gamma(a_2-1)} \sum_{i=0}^{j} \frac{(-1)^{i+j}\binom{j}{i}\,p_{i,j}}{(a_2-1-i)},$$

and $p_{i,j}$ is defined in (2.3). Then, the reliability becomes

$$R = \sum_{j,k=0}^{\infty} b_{1k}\,b_{2j} \int_0^\infty \Phi(v)^{a_1+a_2+j+k}\,g(x)dx.$$

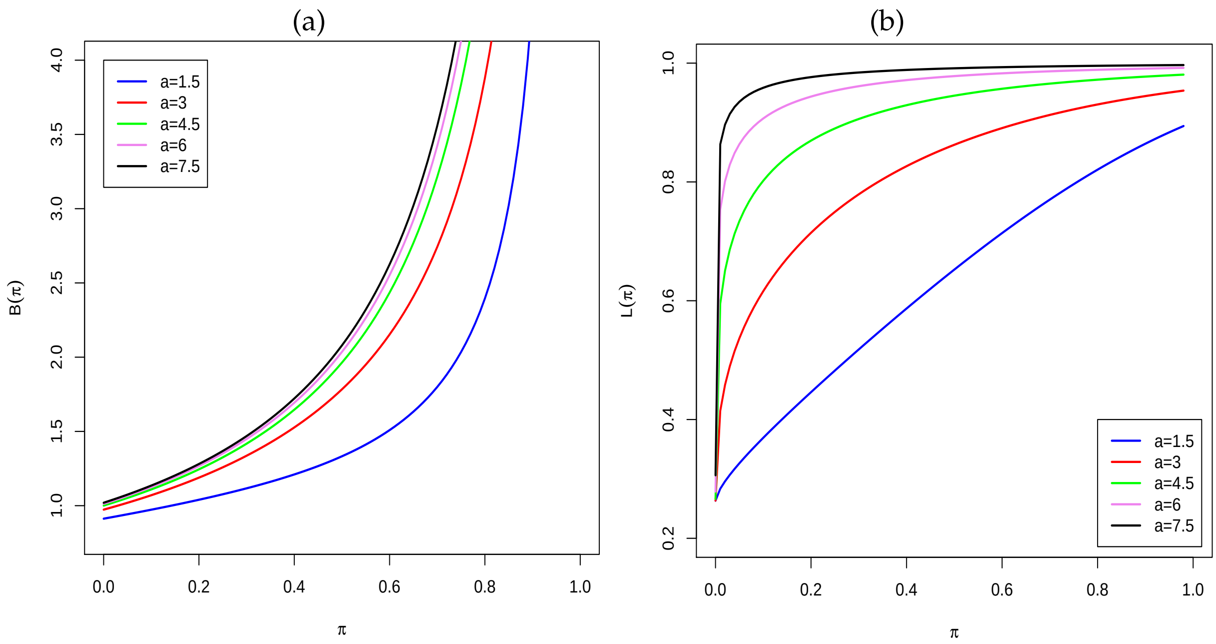


Figure 3.5: Bonferroni and Lorenz curves for the GBS distribution for some parameter values.

From equation (3.8), we can write

$$\Phi(v)^{a_1+a_2+j+k} = \sum_{r=0}^{\infty} (a_1 + a_2 + j + k)\, s_r\, \Phi(v)^r,$$

and R reduces to

$$R = \sum_{j,k=0}^{\infty} b_{1k}\, b_{2j} \sum_{r=0}^{\infty} s_r(a_1 + a_2 + j + k)\, \tau_{0,r},$$

where $\tau_{0,r}$ is obtained from (3.13).

3.8 Order statistics

Suppose $X_1, \ldots, X_n$ is a random sample from the GBS distribution and let $X_{1:n} < \cdots < X_{n:n}$ be the order statistics. Using (2.5) and (2.6), the pdf of $X_{i:n}$ can be expressed as

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

Based on equations (2.7) and (2.8), we obtain

$$\left[\sum_{k=0}^{\infty} b_k\, \Phi(v)^{a+k} \right]^{i+j-1} = \sum_{k=0}^{\infty} c_{j+i-1,k}\, \Phi(v)^{a(j+i-1)+k},$$

where $c_{j+i-1,0} = b_0^{j+i-1}$ and $c_{j+i-1,k} = (kb_0)^{-1}\sum_{m=1}^{k}[m(j+i)-k]\, b_m\, c_{j+i-1,k-m}$. Hence, the pdf of the ith order statistic of X reduces to

$$f_{i:n}(x) = g(x)\sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} m_{j,k,r}\,\Phi(v)^{a(j+i)+k+r}, \tag{3.17}$$

where

$$m_{j,k,r} = \frac{(-1)^j\,(a+r)\,\binom{n-i}{j}\,n!\,b_r\,c_{j+i-1,k}}{(i-1)!\,(n-i)!}.$$

Equation (3.17) can be expressed as

$$f_{i:n}(x) \;=\; \sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} f_{j,k,r}\,h_{r+k+a(j+i)}(x), \tag{3.18}$$

where $f_{j,k,r} = m_{j,k,r}/[r+k+a(j+i)]$.

Equation (3.18) is the main result of this section. It reveals that the pdf of the GBS order statistics is a triple linear combination of exp-BS distributions with parameters α, β and $[k+r+a(j+i)]$. So, several mathematical quantities of the GBS order statistics such as ordinary and incomplete moments, mgf, mean deviations and others can be obtained immediately from those quantities of the exp-BS distribution.

As a simple application of (3.17), the sth moment of X is given by

$$E(X_{i:n}^s) = \sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} m_{j,k,r}\,\tau_{s,k+r+a(j+i)}.$$

Another closed-form expression for $E(X_{i:n}^s)$ can be obtained using a result due to Barakat and Abdelkader (2004) applied to the independent and identically distributed case. Thus,

$$E(X_{i:n}^s) = s\sum_{j=n-i+1}^{n}(-1)^{j-n+i-1}\binom{j-1}{n-i}\binom{n}{j}J_j(s),$$

where $J_j(s) = \int_0^\infty x^{s-1}\,[1-F(x)]^j dx$.

By expanding $[1-F(x)]^j$ and using (3.7), we can obtain $J_j(s)$. For a real non-integer a, we can write from (3.7) and (2.8)

$$\begin{aligned} J_j(s) &= \sum_{m=0}^{j}(-1)^m\binom{j}{m}\int_0^\infty x^{s-1}\left(\sum_{k=0}^{\infty} b_k\,G(x)^{a+k}\right)^m dx \\ &= \sum_{m=0}^{j}(-1)^m\binom{j}{m}\sum_{k=0}^{\infty} d_{m,k}\,\tau_{s-1,am+k} = \sum_{k=0}^{\infty}\sum_{m=0}^{j}(-1)^m\binom{j}{m}d_{m,k}\,\tau_{s-1,am+k}, \end{aligned}$$

where $d_{m,k} = (kt_0)^{-1}\sum_{j=1}^{k}[j(m+1)-k]\,b_m\,d_{m,k-j}$ and $d_{m,0} = b_0^m$.

3.9 Inference and estimation

In this section, we discuss the maximum likelihood method and a Bayesian approach for infe-
rence and estimation of the model parameters of the GBS distribution.

3.9.1 Maximum likelihood estimation

First, the estimation of the parameters of the GBS model is investigated by maximum likelihood. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a random sample taken from X. The total log-likelihood function for $\theta = (\alpha, \beta, a)$ is

$$\begin{aligned}\ell(\theta) &= n\log\{k(\alpha,\beta)\} - n\log\Gamma(a) - \frac{3}{2}\sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log(x_i+\beta) + n\alpha^{-2} \\ &- \frac{1}{2\alpha^2}\sum_{i=1}^{n}\tau(x_i/\beta) + (a-1)\sum_{i=1}^{n}\log\{-\log[1-\Phi(v_i)]\}. \qquad (3.19)\end{aligned}$$

The elements of the score vector are given by

$$U_\alpha = -\frac{n}{\alpha}\left(1+\frac{1}{\alpha^2}\right) + \frac{1}{\alpha^3}\sum_{i=1}^{n}\left(\frac{x_i}{\beta}+\frac{\beta}{x_i}\right) - \frac{(a-1)}{\alpha}\sum_{i=1}^{n}\frac{v_i\,\phi(v_i)\,[1-\Phi(v_i)]^{-1}}{\{\log[1-\Phi(v_i)]\}},$$

$$\begin{aligned}U_\beta &= -\frac{n}{2\beta} + \sum_{i=1}^{n}\frac{1}{x_i+\beta} + \frac{1}{2\alpha^2\beta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta}+\frac{\beta}{x_i}\right) \\ &- \frac{(a-1)}{2\alpha\beta}\sum_{i=1}^{n}\frac{\tau(\sqrt{x_i/\beta})\,\phi(v_i)}{[1-\Phi(v_i)]\,\{\log[1-\Phi(v_i)]\}},\end{aligned}$$

$$U_a = \sum_{i=1}^{n}\log\{-\log[1-\Phi(v_i)]\} - n\,\psi(a),$$

where $\phi(\cdot)$ is the pdf of the standard normal, $\psi(p) = \Gamma'(p)/\Gamma(p)$ is the digamma function, $v_i = \alpha^{-1}\{(x_i/\beta)^{1/2} - (x_i/\beta)^{-1/2}\}$ and $\tau(\sqrt{x_i/\beta}) = (x_i/\beta)^{1/2} + (\beta/x_i)^{1/2}$ for $i = 1, \ldots, n$.

Maximization of (3.19) can be performed using well established functions such as nlm or optimize in the R statistical package. Setting these equations to zero, $\mathbf{U}(\theta) = \mathbf{0}$, and solving them simultaneously gives the maximum likelihood estimate (MLE) $\widehat{\theta}$ of θ, where $\mathbf{U}(\theta)$ is the score vector. These equations cannot be solved analytically and statistical software can be used to evaluate them numerically using iterative techniques such as the Newton-Raphson algorithm.

For interval estimation and tests of hypotheses on the parameters in θ, we require the 3×3 unit observed information matrix $\mathbf{J} = \mathbf{J}(\theta) = \{j_{r,s}\}$ whose elements $j_{r,s}$ for $r, s = \alpha, \beta, a$ are given in Appendix D. Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the estimated approximate multivariate normal $\mathrm{N}_3(\mathbf{0}, n^{-1}\mathbf{J}(\widehat{\theta})^{-1})$ can be used to construct approximate confidence intervals for the model parameters.

The likelihood ratio (LR) statistics is useful for comparing the new distribution with some special models. For example, we may use the LR statistic to check if the fit using the GBS distribution is statistically "superior" to a fit using the BS distribution for a given data set. In any case, considering the partition ` $= (\theta_1^T, \theta_2^T)^T$, with $\theta_1 = a$ and $\theta_2 = (\alpha, \beta)$, tests of hypotheses of the type $H_0 : \theta_1 = \theta_1^{(0)}$ versus $H_A : \theta_1 \neq \theta_1^{(0)}$ can be performed using the LR statistic $w = 2\{\ell(\widehat{\theta}) - \ell(\widetilde{\theta})\}$, where $\widehat{\theta}$ and $\widetilde{\theta}$ are the estimates of $\boldsymbol{\theta}$ unrestricted and restricted under H_0, respectively. Under the null hypothesis H_0, $w \xrightarrow{d} \chi^2_q$, where q is the dimension of the parameter vector $\boldsymbol{\theta}_1$ of interest. The LR test rejects H_0 if $w > \xi_\gamma$, where ξ_γ denotes the upper $100\gamma\%$ point of the χ^2_q distribution.

3.10 Applications

In this section, we compare the fits of the GBS, BS and beta Birnbaum-Saunders (βBS) (Cordeiro and Lemonte, 2011) distributions to three real uncensored data sets from Murthy *et al.* (2004) . The computations are performed using the procedure NLMixed in SAS and the R statistical software. We describe three data sets reported by Murthy *et al.* (2004):

- *Shocks data*

  We consider an uncensored data ($n = 20$) representing the number of shocks before failure. The data are: 2, 3, 6, 6, 7, 9, 9, 10, 10, 11, 12, 12, 12, 13, 13, 13, 15, 16, 16, 18.

- *Repairable data*

  The following data refer to the time between failures for repairable item ($n = 30$): 1.43, 0.11, 0.71, 0.77, 2.63, 1.49, 3.46, 2.46, 0.59, 0.74, 1.23, 0.94, 4.36, 0.40, 1.74, 4.73, 2.23, 0.45, 0.70, 1.06, 1.46, 0.30, 1.82, 2.37, 0.63, 1.23, 1.24, 1.97, 1.86, 1.17.

- *Stress data*

  These data refer to accelerated life testing of ($n = 40$) items with change in stress from 100 to 150 at $t = 15$. The data are: 0.13, 0.62, 0.75, 0.87, 1.56, 2.28, 3.15, 3.25, 3.55, 4.49, 4.50, 4.61, 4.79, 7.17, 7.31, 7.43, 7.84, 8.49, 8.94, 9.40, 9.61, 9.84, 10.58, 11.18, 11.84, 13.28, 14.47, 14.79, 15.54, 16.90, 17.25, 17.37, 18.69, 18.78, 19.88, 20.06, 20.10, 20.95, 21.72, 23.87.

Table 3.1 provides a summary of these data. The shocks data have negative skewness and kurtosis. The repairable data have positive skewness and kurtosis, and have less variability in the data. The stress data have positive skewness and negative kurtosis, larger values of these sample moments.

Table 3.1: Descriptive statistics.

Data	Mean	Median	Mode	Std. Dev.	Skewness	Kurtosis	Min.	Max.
Shocks	10.65	11.5	12.0[a]	4.28	-0.39	-0.28	2	18
Repairable	1.54	1.24	1.23	1.13	1.37	1.80	0.11	4.73
Stress	10.45	9.51	0.13[a]	6.99	0.23	-1.19	0.13	23.87

[a]There are various modes.

In order to estimate the model parameters, we consider the maximum likelihood estimation method discussed in Section 3.9. We take the estimates of α and β from the fitted BS distribution as starting values for the numerical iterative procedure.

Recently, Cordeiro and Lemonte (2011) proposed the β-BS distribution with four positive parameters a, b, α and β by extending the BS distribution which provides more flexibility to fit various types of lifetime data. Its cdf is given by $F(x) = I_{\Phi(v)}(a,b)$, where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function, $\Gamma(\cdot)$ is the gamma function, $I_y(a,b) = B_y(a,b)/B(a,b)$ is the incomplete beta function ratio and $B_y(a,b) = \int_0^y \omega^{a-1}(1-\omega)^{b-1}d\omega$ is the incomplete beta function. The corresponding pdf (for $x > 0$) is

$$f(x) = \frac{\kappa(\alpha,\beta)}{B(a,b)}\, x^{-3/2}\,(x+\beta)\, \exp\left\{-\tau(x/\beta)/(2\alpha^2)\right\}\, \Phi(v)^{a-1}\,\{1-\Phi(v)\}^{b-1},$$

where $\kappa(\alpha,\beta) = \exp(\alpha^{-2})/(2\alpha\sqrt{2\pi\beta})$ and $\tau(z) = z + z^{-1}$, $v = \alpha^{-1}\rho(t/\beta)$ and $\rho(z) = z^{1/2} - z^{-1/2}$. For more details, see Cordeiro and Lemonte (2011).

Table 3.2 lists the MLEs of the parameters (with standard errors in parentheses) and the values of the following statistics for some models: Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Bayesian Information Criterion (BIC). The figures in this table indicate that the GBS model has the smallest values of these statistics among all fitted models. So, it could be chosen as the more suitable model.

A formal test for the third skewness parameter in the GBS distribution is based on LR statistics described in Section 3.9. Applying these tests to the three data sets, we obtain the results listed in Table 3.3. For the repairable data, the additional parameter of the GBS distribution may not, in fact, be necessary because the LR test provides no indications against the BS model when compared with the GBS model. However, for the shocks and stress data, we reject the null hypotheses of the two LR tests in favor of the GBS distribution. The rejection is extremely highly significant for the stress data, and highly or very highly significant for the shocks data. This gives clear evidence of the potential need for three skewness parameter when modeling real data.

Table 3.2: MLEs of the model parameters for the three data sets and the AIC, CAIC and BIC statistics.

Shocks	a	b	α	β	AIC	CAIC	BIC
GBS	7.9526	-	21.8829	0.00193	123.1	124.6	126.1
	(1.7316)	-	(49.6893)	(0.0088)			
BS	1	-	0.5752	9.1023	127.2	127.9	129.2
	-	-	(0.0909)	(1.1225)			
β-BS	188.72	112.52	85.5110	0.01318	124.5	127.1	128.5
	(0.1596)	(0.4004)	(76.4559)	(0.0235)			
Repairable	a	b	α	β	AIC	CAIC	BIC
GBS	3.4530	-	1.5556	0.1448	85.3	86.2	89.5
	(1.3948)	-	(0.8757)	(0.2043)			
BS	1	-	0.8885	1.0938	87.1	87.6	89.9
	-	-	(0.1147)	(0.1606)			
β-BS	8.4879	1.0608	2.9364	0.0698	87.2	88.8	92.8
	(2.0876)	(0.3031)	(1.8261)	(0.0915)			
Stress	a	b	α	β	AIC	CAIC	BIC
GBS	3.6998	-	3.7682	0.1764	268.3	268.9	273.3
	(0.6388)	-	(2.5877)	(0.2627)			
BS	1	-	1.4908	4.4303	291.8	292.1	295.1
	-	-	()	()			
β-BS	450.32	384.67	69.2962	0.1805	269.6	270.7	276.3
	(0.6437)	(0.0397)	(39.1129)	(0.1999)			

Table 3.3: LR tests.

Shocks	Hypotheses	Statistic w	p-value
GBS vs BS	$H_0 : a = 1$ vs $H_1 : H_0$ is false	6.1	0.0135
Repairable	Hypotheses	Statistic w	p-value
GBS vs BS	$H_0 : a = 1$ vs $H_1 : H_0$ is false	3.8	0.0512
Stress	Hypotheses	Statistic w	p-value
GBS vs BS	$H_0 : a = 1$ vs $H_1 : H_0$ is false	25.5	<0.00001

3.11 Concluding remarks

The Birnbaum-Saunders (BS) distribution is widely used to model times to failure for materials subject to fatigue. We propose the gamma Birnbaum-Saunders (GBS) distribution to extend the BS distribution pioneered by Birnbaum and Saunders (1969a). We provide a mathematical treatment of the new distribution including expansions for the cumulative and density functions. We derive explicit expressions for the ordinary and incomplete moments, generating and quantile functions, mean deviations and moments of the order statistics. The estimation of the model parameters is approached by the method of maximum likelihood and the observed information matrix is derived. We consider the likelihood ratio (LR) statistic and other criteria to compare the GBS model with its baseline model and other non-nested model. Applications of the GBS distribution to three real data sets show that the new distribution provides consistently better fits than the BS distribution. We hope that this generalization may attract wider applications in the literature of the fatigue life distributions.

3.12 Appendix

Appendix A: Quantile function

We derive a power series for the $Q_{GBS}(u)$ in the following way. First, we use a known power series for $Q^{-1}(a, 1-u)$. Second, we obtain a power series for the argument $1 - \exp[-Q^{-1}(a, 1-u)]$. Third, we use the power series for the BS qf (Cordeiro and Lemonte, 2011) to obtain a power series for $Q_{GBS}(u)$. We introduce the following quantities defined by Cordeiro and Lemonte (2011). Let $Q^{-1}(a, z)$ be the inverse function of $Q(a, z) = 1 - \gamma(a, z)/\Gamma(a) = \Gamma(a, z)/\Gamma(a) = u$. The inverse incomplete gamma function in the Wolfram website[1] is given by

$$Q^{-1}(a, 1-u) \quad = \quad w + \frac{w^2}{a+1} + \frac{(3a+5)\,w^3}{2(a+1)^2(a+2)} + \frac{[a(8a+33)+31]\,w^4}{3(a+1)^3(a+2)(a+3)} + O(w^5),$$

where $w = [u\,\Gamma(a+1)]^{1/a}$. Thus, we can write the last equation as

$$z = Q^{-1}(a, 1-u) = \sum_{r=0}^{\infty} a_r\, u^{r/a}, \tag{3.20}$$

where the a_i's are given in Section 3.5. They can be expressed as $a_i = \bar{b}_i \Gamma(a+1)^{i/a}$, where $\bar{b}_0 = 0$, $\bar{b}_1 = 1$ and any coefficient $\bar{b}_{i+1}$ for $i \geq 1$ is determined by the cubic recurrence equation

$$\begin{aligned} \bar{b}_{i+1} \quad &= \quad \frac{1}{i(a+i)} \left\{ \sum_{r=1}^{i} \sum_{s=1}^{i-s+1} \bar{b}_r\, \bar{b}_s\, \bar{b}_{i-r-s+2}\, s\,(i-r-s+2) \times \right. \\ &\times \quad \left. \sum_{r=2}^{i} \bar{b}_r\, \bar{b}_{i-r+2}\, r\,[r-a-(1-a)(i+2-r)] \right\}. \end{aligned}$$

The first coefficients are $\bar{b}_2 = 1/(a+1)$, $\bar{b}_3 = (3a+5)/[2(a+1)^2(a+2)]$, ... Here, we present the algebraic details of the calculation of the GBS qf. The cdf of X is given by (3.5) and inverting $u = F(x)$, we obtain (3.10). The BS qf can be expressed as (Cordeiro and Lemonte, 2011)

$$Q_{BS}(u) = \sum_{i,j=0}^{\infty} \sum_{k=j}^{\infty} m_k \left(\frac{-1}{2}\right)^{k-j} \binom{k}{j} h_{j,i}\, u^{(i+j)/a} = \sum_{i,j=0}^{\infty} \bar{p}_j\, h_{j,i}\, u^{(i+j)/a}, \tag{3.21}$$

[1]http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/06/01/03/

where $\overline{p}_j = \sum_{k=j}^{\infty}(-1)^{k-j}\, 2^{j-k}\, m_k \binom{k}{j}$ and $m_k = (2\pi)^{k/2} \sum_{s=0}^{\infty} p_s\, e_{s,k}$. Here, $p_0 = \beta$, $p_2 = \beta\alpha^2/2$, $p_{2j+1} = \beta\,\alpha^{2j+1}\, 2^{-2j} \binom{1/2}{j}$ for $j \geq 0$ and $p_{2j} = 0$ for $j \geq 2$. The quantities $e_{s,k}$ come from the constants d_k's in (2.8) by $e_{s,0} = d_0^s$ and (for $k \geq 1$) $e_{s,k} = (kd_0)^{-1} \sum_{m=1}^{k} [m(j+1) - k]\, d_m\, e_{j,k-m}$, where $d_k = 0$ (for $k = 0, 2, 4, \ldots$), $d_k = c_{(k-1)/2}$ (for $k = 1, 3, 5, \ldots$) and the c_k's are calculated recursively from $c_{k+1} = \frac{1}{2(2k+3)} \sum_{r=0}^{k} \frac{(2r+1)\,(2k-2r+1)\, c_r\, c_{k-r}}{(r+1)(2r+1)}$.

Further, the basic quantities (Cordeiro and Lemonte, 2011) $h_{j,i}$'s come from (2.8) as

$$h_{j,i} = (iv_0)^{-1} \sum_{m=0}^{i} [m(j+1) - i]\, v_m\, h_{j,i-m},$$

where $v_i = q_{i+1}$, $q_0 = 0$, $q_1 = 1$, $q_2 = (\beta - 1)/(\alpha + 1), \ldots$, and the q_i's (for $i \geq 2$) can be derived from the cubic recursive formula

$$\begin{aligned} q_i \;&=\; \frac{1}{[i^2 + (a-2)i + (1-a)]} \Big\{ (1 - \delta_{i,2}) \textstyle\sum_{r=2}^{i-1} q_r\, q_{i+1-r} [r(1-\alpha)(i-r) - r(r-1)] \\ &\quad + \textstyle\sum_{r=1}^{i-1} \sum_{s=1}^{i-r} q_r\, q_s\, q_{i+1-r-s} [r(r-\alpha) + s(\alpha + \beta - 2)(i + 1 - r - s)] \Big\}, \end{aligned}$$

where $\delta_{i,2} = 1$ if $i = 2$ and $\delta_{i,2} = 0$ if $i \neq 2$.

The BS qf (3.21) holds for $-2 < (t/\beta)^{1/2} - (\beta/t)^{1/2} < 2$. Thus, replacing (3.20) in (3.10), we obtain

$$Q_{GBS}(u) = Q_{BS} \left[1 - \exp\left(- \sum_{r=0}^{\infty} a_r\, u^{r/a} \right) \right].$$

By expanding the exponential function and using (2.7), we have

$$\begin{aligned} 1 - \exp\left(- \sum_{r=0}^{\infty} a_r\, u^{r/a} \right) &= 1 - \sum_{l=0}^{\infty} \frac{(-1)^l \left(\sum_{r=0}^{\infty} a_r\, u^{r/a} \right)^l}{l!} \\ &= 1 - \sum_{l=0}^{\infty} \frac{(-1)^l \sum_{r=0}^{\infty} f_{l,r}\, u^{r/a}}{l!} = 1 - \sum_{r=0}^{\infty} p_r u^{r/a}, \qquad (3.22) \end{aligned}$$

where $p_r = \sum_{l=0}^{\infty} \frac{(-1)^l f_{l,r}}{l!}$, $f_{l,r} = (ra_0)^{-1} \sum_{q=1}^{r} [q(l+1) - r]\, a_m\, f_{l,r-q}$ and $f_{l,0} = a_0^l$. The coefficient $f_{l,r}$ can be obtained from $f_{l,0}, \ldots, f_{l,r-1}$ and then from $a_0, \ldots, a_r$. Substituting (3.21) in (3.22) gives

$$Q_{GBS}(u) = \sum_{i,j=0}^{\infty} \overline{p}_j\, h_{j,i} \left(1 - \sum_{r=0}^{\infty} p_r\, u^{r/a} \right)^{(i+j)/a}.$$

By expanding the binomial term for a non-integer power, we have

$$\begin{aligned} Q_{GBS}(u) \;&=\; \sum_{i,j=0}^{\infty} \overline{p}_j\, h_{j,i} \sum_{s=0}^{\infty} (-1)^s \binom{(i+j)/a}{s} \left(\sum_{r=0}^{\infty} p_r\, u^{r/a} \right)^s \\ &=\; \sum_{i,j=0}^{\infty} \sum_{s,r=0}^{\infty} (-1)^s \binom{(i+j)/a}{s} \overline{p}_j\, h_{j,i}\, d_{s,r}\, u^{r/a}, \end{aligned}$$

where the coefficients $d_{s,r}$ can be determined from (2.8) as $d_{s,r} = (r p_0)^{-1} \sum_{v=1}^{r} [v(s+1) - r]\, p_v\, d_{s,r-v}$. Finally, we can rewrite the last equation as

$$Q_{GBS}(u) = \sum_{r=0}^{\infty} \overline{c}_r\, u^{r/a},$$

where $\overline{c}_r = \sum_{i,j,s=0}^{\infty} (-1)^s \binom{(i+j)/a}{s} \overline{p}_j\, h_{j,i}\, d_{s,r}$.

Appendix B: Incomplete moments

From Cordeiro and Lemonte (2011), we can write

$$\Phi(v)^r = \frac{1}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \beta^{-(2s_j+j)/2} A(k_1,\ldots,k_j) \sum_{m=0}^{2s_j+j} (-\beta)^m \binom{2s_j+j}{m} x^{(2s_j+j-2m)/2},$$

where s_j and $A(k_1,\ldots,k_j)$ are defined in equation (3.14). Thus,

$$\begin{aligned} T_n(y) &= k(\alpha,\beta) \sum_{r=0}^{\infty} \frac{d_r}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \beta^{-(2s_j+j)/2} A(k_1,\ldots,k_j) \sum_{m=0}^{2s_j+j} (-\beta)^m \binom{2s_j+j}{m} \\ &\quad \times \int_0^y x^{n+(2s_j+j-2m-3)/2} (x+\beta) \exp\left[-\frac{\tau(x/\beta)}{2\alpha^2}\right] dx. \end{aligned} \tag{3.23}$$

Let

$$D(p,q) = \int_0^q x^q \exp\left[-\frac{(x/\beta+\beta/x)}{2\alpha^2}\right] dx = \beta^{p+1} \int_0^{q/\beta} u^q \exp\left[-\frac{(u+u^{-1})}{2\alpha^2}\right] du.$$

From Terras (1981), we can write $D(p,q) = \beta^{p+1} \kappa_{p+1}(\alpha^{-2}) - q^{p+1} \kappa_{p+1}\left(\frac{q}{2\alpha^2\beta}, \frac{\beta}{2\alpha^2 q}\right)$, where $\kappa_\nu(x_1,x_2)$ denotes the incomplete Bessel function with arguments x_1 and x_2 and order ν. For further details, see Jones (2007a, 2007b) and Harris (2008). By inserting the above quantities in (3.23), we obtain

$$\begin{aligned} T_n(y) &= k(\alpha,\beta) \sum_{r=0}^{\infty} \frac{d_r}{2^r} \sum_{j=0}^{r} \binom{r}{j} \sum_{k_1,\ldots,k_j=0}^{\infty} \beta^{-(2s_j+j)/2} A(k_1,\ldots,k_j) \sum_{m=0}^{2s_j+j} (-\beta)^m \binom{2s_j+j}{m} \\ &\quad \times \left[D\left(n+\frac{2s_j+j-2m-1}{2}, y\right) + \beta\, D\left(n+\frac{2s_j+j-2m-3}{2}, y\right)\right]. \end{aligned}$$

Appendix C: Generating function

Here, we prove the first result in Section 3.6. Using the expansion of the qf, we obtain

$$M(s) = \sum_{k=0}^{\infty} \int_0^{\infty} e^{st}\, b_k (a+k)\, G(t)^{a+k-1} g(t) dt = \sum_{k=0}^{\infty} (a+k)\, b_k \int_0^1 e^{s\, Q_{BS}(u)}\, u^{a+k-1} du.$$

Let $J_s = \{(i,j) \in \mathbf{N} \times \mathbf{N}; i+j=s\}$ and $\overline{q}_s = \sum_{(i,j)\in J_s} \overline{p}_j\, h_{j,i}$. We can rewrite (3.21) based on the set J_s as

$$Q_{BS}(u) = \sum_{i,j=0}^{\infty} \overline{p}_j\, h_{j,i}\, u^{(i+j)/a} = \sum_{s=0}^{\infty} \overline{q}_s\, u^{s/a},$$

By expanding the exponential function and using (2.7) in the last integral, we have

$$\int_0^1 \mathrm{e}^{\mathrm{s}\,\mathrm{Q}_{\mathrm{BS}}(\mathrm{u})}\,\mathrm{u}^{\mathrm{a+k-1}}\mathrm{du} = \sum_{\mathrm{n=0}}^{\infty}\frac{\mathrm{s}^{\mathrm{n}}}{\mathrm{n!}}\int_0^1\left(\sum_{\mathrm{s=0}}^{\infty}\overline{\mathrm{q}}_{\mathrm{s}}\,\mathrm{u}^{\mathrm{s/a}}\right)^{\mathrm{n}}\mathrm{u}^{\mathrm{a+k-1}}\mathrm{du} = \sum_{\mathrm{n=0}}^{\infty}\frac{\mathrm{s}^{\mathrm{n}}}{\mathrm{n!}}\int_0^1\sum_{\mathrm{s=0}}^{\infty}\overline{\mathrm{v}}_{\mathrm{s,n}}\,\mathrm{u}^{\mathrm{s/a+a+k-1}}\mathrm{du},$$

where $\overline{v}_{s,n} = (n\overline{q}_0)^{-1}\sum_{m=0}^{s}[m(s+1)-j]\,\overline{q}_m\,\overline{v}_{s,n-m}$ follows from (2.7), (2.8) and $\overline{v}_{s,0} = \overline{q}_0^{s}$. Finally, the mgf of X reduces to

$$M(s) = \sum_{n,s,k=0}^{\infty}\frac{(a+k)\,b_k\,\overline{v}_{s,n}}{(s/a+a+k)}\frac{s^n}{n!}.$$

Appendix D: Elements of the observed information matrix

The elements of the observed information matrix $J(\theta)$ for the parameters α, β and a are given by

$$\begin{aligned} j_{\alpha,\alpha} &= \frac{n}{\alpha^2}+\frac{3n}{\alpha^4}-\frac{3}{\alpha^4}\sum_{i=1}^{n}\left(\frac{x_i}{\beta}+\frac{\beta}{x_i}\right)+\frac{(a-1)}{\alpha^2}\sum_{i=1}^{n}\left\{\frac{\nu_i^3\,\phi(\nu_i)\,[1-\Phi(\nu_i)]^{-1}}{\{\log[1-\Phi(\nu_i)]\}}\right.\\ &\quad\left.+2\,\frac{\nu_i\,\phi(\nu_i)\,[1-\Phi(\nu_i)]^{-1}}{\{\log[1-\Phi(\nu_i)]\}}+\frac{\nu_i^2\,\phi(\nu_i)^2\,[1-\Phi(\nu_i)]^{-2}}{\{\log[1-\Phi(\nu_i)]\}^2}-\frac{\nu_i^2\,\phi(\nu_i)\,[1-\Phi(\nu_i)]^{-2}}{\{\log[1-\Phi(\nu_i)]\}}\right\}, \end{aligned}$$

$$\begin{aligned} j_{\alpha,\beta} &= -\frac{1}{\alpha^3\beta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta}-\frac{\beta}{x_i}\right)-\frac{(a-1)}{2\alpha^2\beta}\\ &\quad\sum_{i=1}^{n}\left\{\frac{\nu_i^2\,\phi(\nu_i)\,\tau(\sqrt{x_i/\beta})}{\{\log[1-\Phi(\nu_i)]\}\,[1-\Phi(\nu_i)]}-\frac{\phi(\nu_i)\,\tau(\sqrt{x_i/\beta})}{\{\log[1-\Phi(\nu_i)]\}\,[1-\Phi(\nu_i)]}\right.\\ &\quad\left.-\frac{\nu_i\,\phi(\nu_i)^2\,\tau(\sqrt{x_i/\beta})}{\{\log[1-\Phi(\nu_i)]\}^2\,[1-\Phi(\nu_i)]^2}-\frac{\nu_i\,\phi(\nu_i)^2\,\tau(\sqrt{x_i/\beta})}{\{\log[1-\Phi(\nu_i)]\}\,[1-\Phi(\nu_i)]^2}\right\}, \end{aligned}$$

$$\begin{aligned} j_{\beta,\beta} &= \frac{n}{2\beta^2}-\sum_{i=1}^{n}\frac{1}{(x_i+\beta)^2}-\frac{1}{\alpha^2\beta^3}\sum_{i=1}^{n}x_i-\frac{(a-1)}{4\alpha\beta^2}\\ &\quad\sum_{i=1}^{n}\left\{\frac{\alpha\,\nu_i\,\phi(\nu_i)\,[1-\Phi(\nu_i)]^{-1}}{\{\log[1-\Phi(\nu_i)]\}}+\frac{\alpha^{-1}\,\nu_i\,\phi(\nu_i)\;\tau(\sqrt{x_i/\beta})^2}{\{\log[1-\Phi(\nu_i)]\}\,[1-\Phi(\nu_i)]}\right.\\ &\quad\left.+\frac{\alpha^{-1}\,\phi(\nu_i)^2\,\tau(\sqrt{x_i/\beta})^2}{\{\log[1-\Phi(\nu_i)]\}\,[1-\Phi(\nu_i)]^2}+\frac{\alpha^{-1}\,\phi(\nu_i)^2\,\tau(\sqrt{x_i/\beta})^2}{\{\log[1-\Phi(\nu_i)]\}^2\,[1-\Phi(\nu_i)]^2}\right\}, \end{aligned}$$

$$j_{a,\beta} = -\sum_{i=1}^{n}\frac{(2\alpha\beta)^{-1}\,\tau(\sqrt{x_i/\beta})\,\phi(\nu_i)}{\{\log[1-\Phi(\nu_i)]\}\,[1-\Phi(\nu_i)]},$$

where $\psi^{'}(.)$ is the trigamma function.

References

BARAKAT, H. M.; ABDELKADER, Y. H. Computing the moments of order statistics from non-identical random variables. *Statistical Methods and Applications*, v. 13, p. 13–24, 2004.

BIRNBAUM, Z. W.; SAUNDERS, S. C. A new family of life distributions. *Journal of Applied Probability*, v. 6, p. 319–327, 1969a.

BIRNBAUM, Z. W.; SAUNDERS, S. C. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, v. 6, p. 328–377, 1969b.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. (2011a). The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, v. 81, p. 827–842, 2011a.

CORDEIRO, G. M.; LEMONTE, A. J. The beta Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, v. 55, p. 1445–1461, 2011b.

CORDEIRO, G. M.; LEMONTE, A. J.; ORTEGA, E. M. M. An extended fatigue life distribution. *Statistics*, 2011, doi:10.1080/02331888.2011.617447.

COWLES, M. K.; CARLIN, B. P. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, v. 91, p. 133–169, 1996.

DÍAZ-GARCÍA, J. A.; LEIVA, V. (2005). A new family of life distributions based on elliptically contoured distributions. *Journal of Statistical Planning and Inference*, v. 137, p. 1512–1513, 2005.

FLEMING, T. R.; O'FALLON, J. R.; O'BRIEN, P. C.; HARRINGTON, D. P. Modified Kolmogorov-

Smirnov test procedures with application to arbitrarily right censored data. *Biometrics*, v. 36, p. 607–626, 1980.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science*, v. 7, 457–472, 1992.

GRADSHTEYN, I. S.; RYZHIK, I. M. Table of integrals, series, and products. New York: Academic Press, 2007.

GREENWOOD, J. A.; LANDWEHR, J. M.; MATALAS, N. C.; WALLIS, J. R. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, v. 15, p. 1049–1054, 1979.

GUIRAUD, P.; LEIVA, V.; FIERRO, R. (2009). A non-central version of the Birnbaum-Saunders distribution for reliability analysis. *IEEE Transactions on Reliability*, v. 58, p. 152–160, 2009.

GUPTA, R. D.; KUNDU, D. Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, v. 43, p. 117–130, 2001.

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

HARRIS, F. E. Incomplete Bessel, generalized incomplete gamma, or leaky aquifer functions. *Journal of Computational and Applied Mathematics*, v. 215, p. 260–269, 2008.

JONES, D. S. Incomplete Bessel functions. *Proceedings of the Edinburgh Mathematical Society*, v. 50, p. 173–183, 2007a.

JONES, D. S. Incomplete Bessel functions. Asymptotic expansions for large argument. *Proceedings of the Edinburgh Mathematical Society*, v. 50, p. 711–723, 2007b.

KUNDU, D.; KANNAN, N.; BALAKRISHNAN, N. On the function of Birnbaum-Saunders distribution and associated inference. *Computational Statistics and Data Analysis*, v. 52, p. 2692–2702, 2008.

LEMONTE, A. J.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved statistical inference for the two-parameter Birnbaum-Saunders distribution, *Computational Statistical and Data Analysis*, v. 51, p. 4656–4681, 2007.

LEMONTE, A. J.; SIMAS, A. B.; CRIBARI-NETO, F. Bootstrap-based improved estimators for the two-parameter Birnbaum-Saunders distribution, *Journal Statistical Computation Simulation*,

v. 78, p. 37–49, 2008.

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

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

MURTHY, D. N. P.; XIE, M.; JIANG, R. Weibull models. New Jersey: John Wiley, 2004.

NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. The Zografos-Balakrishnan-G family of distributions: Mathematical properties and applications. *Communication in Statistics - Theory and Methods*, to appear, 2013.

NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. *Mathematics and Computer in Simulation*, v. 74, p. 1–7, 2007.

RIECK, J. R. A moment-generating function with application to the Birnbaum-Saunders distribution, *Communications in Statistics-Theory and Methods*, v. 28, p. 2213–2222, 1999.

RISTIĆ, M. M.; BALAKRISHNAN, N. The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 2011, doi: 10.1080/00949655.2011.574633.

TERRAS, R. A Miller algorithm for an incomplete Bessel function. *Journal of Computational Physics*, v. 39, p. 233–240, 1981.

WATSON, G. N. A Treatise on the Theory of Bessel Functions. Cambridge University Press, 1995.

WU, J.; WONG, A. C. M. Improved interval estimation for the two-parameter Birnbaum-Saunders distribution, *Computational Statistics and Data Analysis*, v. 47, p. 809–821, 2004.

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

CHAPTER 4

A new extension of the normal distribution

Resumo

Fornecer uma nova distribuição é sempre algo precioso para os estatísticos. Uma nova distribuição de três parâmetros chamada distribuição gama normal é definida e estudada. Várias propriedades estruturais da nova distribuição são derivadas, incluindo algumas expressões explícitas para o momentos, funções quantílica e geradora, desvios médios, probabilidade ponderada de momentos e dois tipos de entropia. Investigamos também as estatísticas de ordem e seus momentos. Técnicas de máxima verossimilhança são usadas para ajustar o novo modelo e para mostrar sua potencialidade através de dois exemplos de dados reais. Com base em três critérios, a proposta de distribuição fornece um melhor ajuste quando comparada com a distribuição skew-normal.

Palavras-chave: Distribuição gama. Distribuição normal. Estimação de máxima verossimilhança. Média desvio padrão. Quantil.

Abstract

Providing a new distribution is always precious for statisticians. A new three-parameter distribution called the gamma normal distribution is defined and studied. Various structural properties of the new distribution are derived, including some explicit expressions for the moments, quantile and generating functions, mean deviations, probability weighted moments and two types of entropy. We also investigate the order statistics and their moments. Maximum likelihood techniques are used to fit the new model and to show its potentiality by means of two examples of real data. Based on three criteria, the proposed distribution provides a bet-

ter fit then the skew-normal distribution.

Keywords: Gamma distribution; Maximum likelihood estimation; Mean deviation; Normal distribution; Quantile.

4.1 Introduction

In statistics, the normal distribution is the most popular model in applications to real data. When the number of observations is large, it can serve as an approximate distribution for other models. The pdf (for $x \in \mathbb{R}$) of the normal $\mathrm{N}(\mu,\sigma)$ distribution becomes

$$g(x;\mu,\sigma) = \frac{1}{\sqrt{2\,\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right), \qquad (4.1)$$

where $-\infty < \mu < \infty$ is a location parameter and $\sigma > 0$ is a scale parameter. Its cdf is given by

$$G(x;\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right). \qquad (4.2)$$

Here, we study some structural properties of the *gamma normal* (GN) distribution, which generalizes the normal distribution. We introduce the GN distribution and provide plots of its pdf. We derive expansions for the pdf and cdf (Section 4.3) and explicit expressions for the qf (Section 4.4), ordinary and incomplete moments and Bonferroni and Lorenz curves (Section 4.5), generating function (Section 4.6) and entropies (Section 4.7). In Section 4.8, we investigate the order statistics and their moments. The estimation of the model parameters is performed by maximum likelihood in Section 4.9 and two applications are provided in Section 4.10. Concluding remarks are addressed in Section 4.11.

4.2 The GN distribution

By taking the pdf (4.1) and cdf (4.2) of the normal distribution with location parameter $\mu \in \mathbb{R}$ and dispersion parameter $\sigma > 0$, the pdf and cdf of the GN distribution are obtained from equations (2.1) and (2.2) (for $x \in \mathbb{R}$) as

$$f(x) = \frac{1}{\sigma\Gamma(a)}\,\phi\left(\frac{x-\mu}{\sigma}\right)\left\{-\log\left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right]\right\}^{a-1} \qquad (4.3)$$

and

$$F(x) = \frac{1}{\Gamma(a)}\int_0^{-\log\left[1-\Phi\left(\frac{x-\mu}{\sigma}\right)\right]} t^{a-1}\,\mathrm{e}^{-t}\,\mathrm{d}t. \qquad (4.4)$$

Evidently, the GN distribution is defined by a simple transformation: if $Z \sim \mathrm{G}(a,1)$, then the random variable $X = \Phi^{-1}(1-\mathrm{e}^{-Z})$ has the density function (4.3), with $\mu = 0$ and $\sigma = 1$. Hereafter, a random variable X following (4.3) is denoted by $X \sim \mathrm{GN}(a,\mu,\sigma)$. The density function (4.3) does not involve any complicated function and the normal distribution arises as the basic exemplar for $a = 1$. It is a positive point of the current generalization. We motivate the paper by comparing the performances of the GN, normal and skew-normal models applied to two real data sets.

In Figure 4.1, we display some possible shapes of the density function (4.3) for some parameter values. It is evident that the GN distribution is much more flexible than the normal distribution.

The new distribution is easily simulated as follows: if V is a gamma random variable with parameter a, then

$$X = \sigma\,\Phi^{-1}[1-\exp(-V)] + \mu$$

has the GN(a,μ,σ) distribution. This scheme is useful because of the existence of fast generators for gamma random variables and for the standard normal quantile function in most statistical packages.

4.3 Useful expansions

Expansions for equations (4.3) and (4.4) can be derived using the concept of exponentiated distributions. Consider the *exponentiated normal* (exp-N) distribution with power parameter $a > 0$ defined by $Y \sim$ exp-N(a,μ,σ), with cdf and pdf given by $H_a(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)^a$ and $h_a(y) = \frac{a}{\sigma}\,\phi\left(\frac{y-\mu}{\sigma}\right)\,\Phi(\frac{y-\mu}{\sigma})^{a-1}$, respectively. Following Nadarajah *et al.* (2013), equation (4.3) can be expressed as

$$f(x) = \sum_{k=0}^{\infty} b_k\, h_{a+k}(x), \tag{4.5}$$

where

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)} \sum_{j=0}^{k} \frac{(-1)^{j+k}\binom{k}{j}\, p_{j,k}}{(a-1-j)}$$

and $h_{a+k}(x) = \left(\frac{a+k}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{x-\mu}{\sigma}\right)^{a+k-1}$ denotes the exp-N$(a+k,\mu,\sigma)$ density function. The cdf corresponding to (4.5) becomes

$$F(x) = \sum_{k=0}^{\infty} b_k\, H_{a+k}(x) = \sum_{k=0}^{\infty} b_k\,\Phi\left(\frac{x-\mu}{\sigma}\right)^{a+k}, \tag{4.6}$$

where $H_{a+k}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)^{a+k}$ denotes the exp-N cdf with parameters $a+k$, μ and σ.

If $a > 0$ is a real number, we can expand $\Phi\left(\frac{x-\mu}{\sigma}\right)^{a+k}$ as

$$\Phi\left(\frac{x-\mu}{\sigma}\right)^{a+k} = \sum_{r=0}^{\infty} s_r(a+k)\,\Phi\left(\frac{x-\mu}{\sigma}\right)^{r}, \tag{4.7}$$

where

$$s_r(a) = \sum_{l=r}^{\infty}(-1)^{r+l}\binom{a}{l}\binom{l}{r}. \tag{4.8}$$

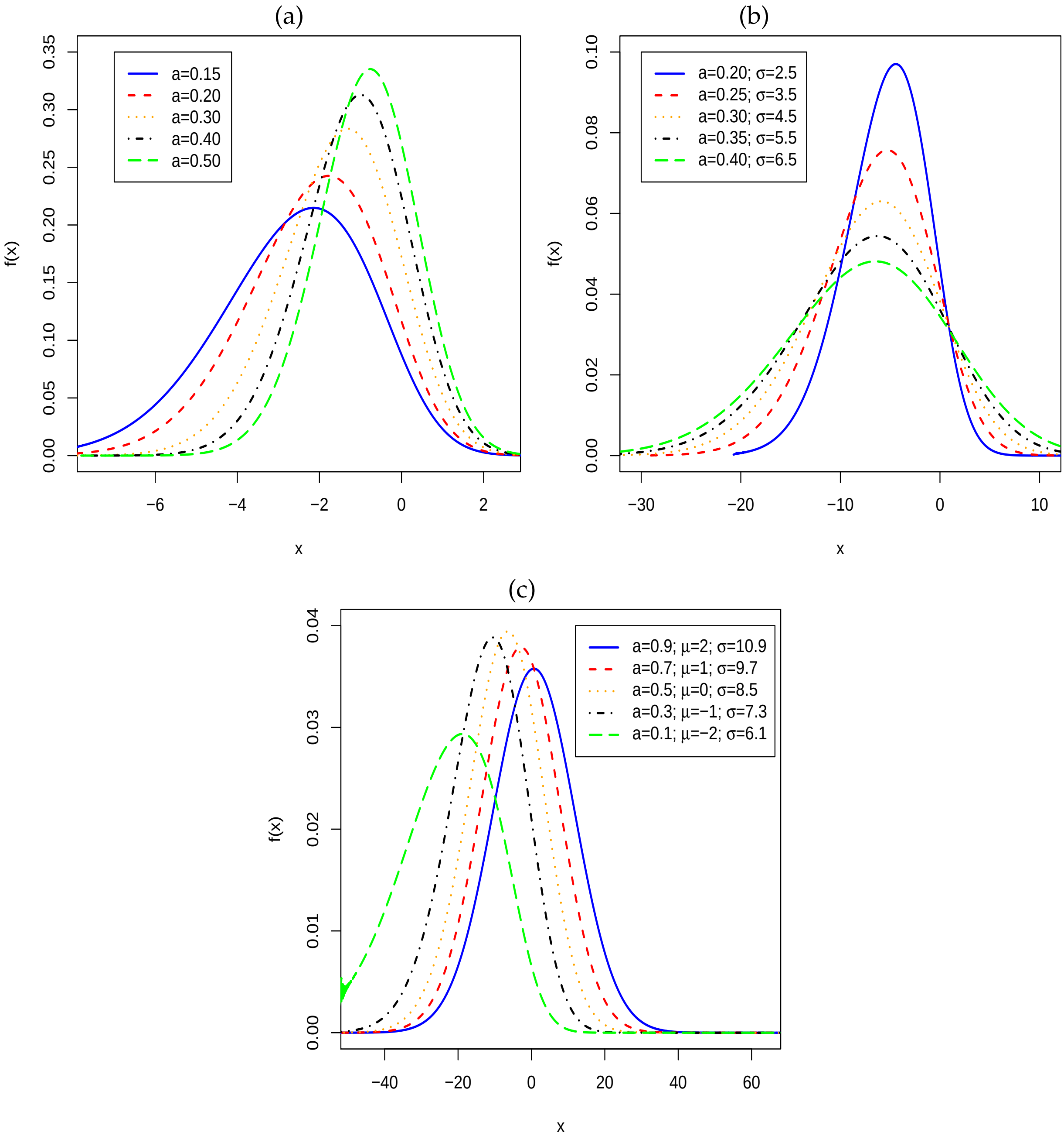


Figure 4.1: Plots of the new density function for some parameter values. 3(a) For different values of a with $\mu = 0$ and $\sigma = 1$. (b) For different values of a and σ with $\mu = 0$. (c) For different values of a, μ and σ.

Combining equations (4.6) and (4.7), we obtain

$$F(x)=\sum_{k=0}^{\infty}\sum_{r=0}^{\infty} b_k\, s_r(a+k)\,\Phi\left(\frac{x-\mu}{\sigma}\right)^r.$$

By differentiating the previous equation and changing indices, we can write

$$f(x)=\sum_{r=0}^{\infty} d_r\, h_{r+1}(x), \tag{4.9}$$

where $d_r = \sum_{k=0}^{\infty} b_k\, s_{r+1}(a+k)$. Clearly, by integrating both sides of the previous equation, $\sum_{r=0}^{\infty} d_r = 1$. Equation (4.9) is the main result of this section. It reveals that the GN density function is a linear combination of exp-N densities. So, several properties of the GN distribution can be obtained by knowing those properties of the exp-N distribution.

4.4 Quantile Function

The GN qf, say $Q(u) = F^{-1}(u)$, can be expressed in terms of the normal qf ($Q_N(\cdot)$). The normal qf is given by $x = Q_N(u) = \sigma\Phi^{-1}(u)+\mu$. Inverting equation (4.4), we obtain the qf of X as

$$F^{-1}(u)=Q_{GN}(u)=\mu+\sigma\, Q_N\left\{1-\exp[-Q^{-1}(a,1-u)]\right\}, \tag{4.10}$$

for $0<u<1$, where $Q^{-1}(a,u)$ is the inverse function of $Q(a,z)=1-\gamma(a,z)/\Gamma(a)$. Quantities of interest can be obtained from (4.10) by substituting appropriate values for u. Further, the normal qf can be expressed as (Steinbrecher, 2002) in equation (7.27), see Appendix A. Further, after some algebra (see Appendix A), we obtain

$$Q_N(u)=\sum_{s=0}^{\infty} w_s\, u^s, \tag{4.11}$$

where $w_s = \sum_{k=s}^{\infty}(-2)^{s-k}\,(\sqrt{2\pi})^k\binom{k}{s}\, d_k$ and the quantity d_k was defined in Section 4.3.

We can obtain the inverse function $Q^{-1}(a,u)$ in the Wolfram website as

$$z=Q^{-1}(a,1-u)=\sum_{i=0}^{\infty} a_i\, u^{i/a},$$

where $a_0 = 0$, $a_1 = \Gamma(a+1)^{1/a}$, $a_2 = \Gamma(a+1)^{2/a}/(a+1)$, $a_3 = (3a+5)\Gamma(a+1)^{3/a}/[2(a+1)^2(a+2)]$, etc.

By expanding the exponential function and using (2.7), we have (see Appendix A)

$$1-\exp\left(-\sum_{r=0}^{\infty} a_r u^{r/a}\right)=1-\sum_{r=0}^{\infty} p_r\, u^{r/a},$$

where the p_r's are defined there. We can write

$$Q_{GN}(u)=\mu+\sigma\, Q_N\left(1-\sum_{r=0}^{\infty} p_r u^{r/a}\right).$$

By using equations (4.11) and (2.7), we can obtain from (4.10)

$$Q_{GN}(u) = \mu + \sigma \sum_{r=0}^{\infty} \tau_r \, u^{r/a}, \tag{4.12}$$

where $\tau_r = h_{j,r} \, \overline{p}_j$, $\overline{p}_j = \sum_{s=0}^{\infty} \sum_{j=0}^{s} (-1)^j \, w_s \binom{s}{j}$ and $h_{j,i} = (i \, p_0)^{-1} \sum_{m=0}^{i} [m(j+1) - i] \; p_m \, h_{j,i-m}$. Some algebraic details about (4.12) and others quantities of interest are given in Appendix A. Equations (4.11) and (4.12) are the main results of this section.

4.5 Moments

Here, we obtain the ordinary and incomplete moments of X. They can be immediately derived from the moments of Y following the exp-N(a, μ, σ) distribution. Hereafter, let Z be the standard GN$(a, 0, 1)$ random variable. First, we obtain the moments of Z. Thus, we can write from (4.5)

$$\mu'_n = E(Z^n) = \sum_{k=0}^{\infty} b_k \int_{-\infty}^{\infty} x^n \, \Phi(x)^{a+k-1} \phi(x) \, dx.$$

Further, we can express μ'_n in terms of $Q_N(u)$ as

$$\mu'_n = \sum_{k=0}^{\infty} b_k \int_0^1 Q_N(u)^n \, u^{a+k-1} du.$$

Using (2.7) and (4.11), we can rewrite μ'_n as

$$\mu'_n = \sum_{k,s=0}^{\infty} \frac{b_k \, e_{n,s}}{(a+k+s)}, \tag{4.13}$$

where the quantities $e_{n,s}$ are determined from (2.8) and (4.11) as $e_{n,s} = (i \, w_0)^{-1} \sum_{m=1}^{s} [m(n+1) - s] \, w_m \, e_{n,s-m}$ for $s \geq 1$, $e_{n,0} = w_0^n$, $w_m = \sum_{k=m}^{\infty} (-2)^{m-k} \, (\sqrt{2\pi})^k \binom{k}{m} d_k$ and the quantity d_k was defined in Section 4.3. The moments of X immediately follow from the moments of Z as $E(X^n) = \sum_{k=0}^{n} \binom{n}{k} \, \mu^{n-k} \, \sigma^k \, \mu'_k$.

The second representation for μ'_n is based on (n, r)th PWM (for n and r positive integers) of the standard normal distribution given by

$$\mu'_n = \sum_{k,r=0}^{\infty} b_k \, s_{r+1}(a+k) \, \tau_{n,r}, \tag{4.14}$$

where b_k was defined previously and $s_r(a)$ is given by (4.8) and $\tau_{n,r}$ can be expressed as (Nadarajah, 2008)

$$\begin{aligned} \tau_{n,r} \;=\; & 2^{n/2} \pi^{-(r+1/2)} \sum_{\substack{p=0 \\ (n+r-p) \text{ even}}}^{r} \left(\frac{\pi}{2}\right)^p \binom{r}{p} \Gamma\left(\frac{n+r-p+1}{2}\right) \times \\ & F_A^{(r-p)}\left(\frac{n+r-p+1}{2}; \frac{1}{2}, \ldots, \frac{1}{2}; \frac{3}{2}, \ldots, \frac{3}{2}; -1, \ldots, -1\right), \end{aligned} \tag{4.15}$$

where

$$F_A^{(n)}(a,b_1,\ldots,b_n;c_1,\ldots,c_n;x_1,\ldots,c_n)=\sum_{m_1=0}^{\infty}\cdots\sum_{m_n=0}^{\infty}\frac{a_{m_1+\cdots+m_n}(b_1)_{m_1}\cdots(b_n)_{m_n}}{(c_1)_{m_1}\cdots(c_n)_{m_n}}\frac{x_1^{m_1}\cdots x_n^{m_n}}{m_1!\cdots m_n!}$$

is the Lauricella function of type A (Exton, 1978) and the Pochhammer symbol $(a)_k = a(a+1)\ldots(a+k-1)$ indicates the kth rising factorial power of a with the convention $(a)_0 = 1$.

We derive three formulae for the nth incomplete moment of Z given by $T_n(y) = P(Z < y) = \int_0^y x^n f(x)dx$. First, based on equation (4.5), with $\mu = 0$ and $\sigma = 1$, $T_n(y)$ reduces to

$$T_n(y)=\sum_{r=0}^{\infty} d_r \int_{-\infty}^{y} x^n\,\phi(x)\,\Phi(x)^r dx. \tag{4.16}$$

We can write $\Phi(x)$ as a power series $\Phi(x) = \sum_{j=0}^{\infty} a_j\, x^j$, where $a_0 = (1+\sqrt{2/\pi})^{-1}/2$, $a_{2j+1} = (-1)^j/[\sqrt{2\pi}\,2^j(2j+1)j!]$ for $j = 0,1,2\ldots$ and $a_{2j} = 0$ for $j = 1,2,\ldots$ Further, using (2.7), we have

$$\Phi(x)^r=\sum_{j=0}^{\infty} c_{r,j}\,x^j, \tag{4.17}$$

where the coefficients $c_{r,j}$ can be determined from the recurrence equation (2.8) with these a_i's. Thus, using (4.17) and changing variable in the last integral, it follows from (4.16)

$$T_n(y)=\frac{1}{\sqrt{2\,\pi}}\sum_{j,r=0}^{\infty} 2^{n+j-1}\,d_r\,c_{r,j}\,\gamma\left(\frac{n+j+1}{2},\frac{y^2}{2}\right). \tag{4.18}$$

Next, we derive a second representation for the moments. The integral $A(j,q) = \int_{-\infty}^{q} x^j\,\mathrm{e}^{-x^2/2}\,dx$ can be determined for $q > 0$ and $q < 0$. We define

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

For $q < 0$ and $q > 0$, we have

$$A(j,q)=(-1)^j\,G(j)+(-1)^{j+1}H(j,q)$$

and

$$A(j,q)=(-1)^j G(j)+H(j,q),$$

respectively, where the integral $H(j,q) = \int_0^q x^j\,\mathrm{e}^{-x^2/2}\,dx$ can be easily computed (Whittaker and Watson, 1990). The details are given in Appendix B. After some algebra, we can write $T_n(y)$ as

$$T_n(y)=\frac{1}{\sqrt{2\pi}}\sum_{k,r,j=0}^{\infty} b_k\,c_{r,j}\,s_{r+1}(a+k)A(j+n,y), \tag{4.19}$$

where $c_{r,j} = (j\,a_0)^{-1}\sum_{m=1}^{j}[m(r+1)-j]\,a_m\,c_{r,j-m}$, for $j \geq 1$, $c_{r,0} = a_0^r$, $c_{0,0} = 1$, $s_r(a)$ is given by (4.8) and the quantities a_i's are defined in Section 4.4. Some details about (4.19) are given in Appendix B.

A third representation for $T_n(y)$ is based on the normal qf. Thus, equation (4.17) becomes

$$T_n(y) = \sum_{r=0}^{\infty} d_r \int_0^{\Phi(y)} Q_N(u)^n u^r du.$$

After some algebra, using (2.7) and (4.11), we have

$$T_n(y) = \sum_{r,s=0}^{\infty} d_r\, e_{n,s}\, \frac{\Phi(y)^{s+r+1}}{(s+r+1)}, \tag{4.20}$$

where $e_{n,s}$ is given before. More details about (4.20) are addressed in Appendix B.

The nth incomplete moment of X follows after a binomial expansion

$$E(X^n|X<y) = \sum_{k=0}^{n} \mu'_{n-k}\, \sigma^k \binom{n}{k} T_k\left(\frac{y-\mu}{\sigma}\right).$$

We can derive the mean deviations of Z about the mean μ'_1 and about the median M in terms of its first incomplete moment. They can be expressed as

$$\delta_1 = 2\big[\mu'_1\, F(\mu'_1) - T_1(\mu'_1)\big] \qquad \text{and} \qquad \delta_2 = \mu'_1 - 2T_1(M), \tag{4.21}$$

where $\mu'_1 = E(Z)$ and $T_1(q) = \int_{-\infty}^{q} x\, f(x)\, dx$. The quantity $T_1(q)$ can be obtained from (4.18) (or (4.19) or (4.20)) with $n = 1$ and the measures δ_1 and δ_2 in (4.21) are immediately determined by setting $q = \mu'_1$ and $q = M$, respectively.

For a positive random variable X, the Bonferroni and Lorenz curves are defined by $B(\pi) = T_1(q)/(\pi\mu'_1)$ and $L(\pi) = T_1(q)/\mu'_1$, respectively, where $q = F^{-1}(\pi) = Q_{GN}(\pi)$ comes from the qf (4.10) for a given probability π.

Next, we obtain the PWMs of Z. They cover the summarization and description of theoretical probability distributions. The primary use of these moments is to estimate the parameters of a distribution whose inverse cannot be expressed explicitly. The $(s,p)th$ PWM of Z is formally defined as

$$\xi_{s,p} = E[Z^s\, F(Z)^p] = \int_0^{\infty} z^s\, F(z)^p\, f(z)\, \mathrm{d}z.$$

Using (4.6), (4.5) and (2.7), we obtain

$$\xi_{s,p} = \frac{1}{\sqrt{\pi}} \sum_{j,n,r=0}^{\infty} 2^{\frac{(j+n+s)}{2}-1}\; d_r\, \overline{f}_{p,j}\, c_{r,n}\, \Gamma\left(\frac{j+n+s+1}{2}\right), \tag{4.22}$$

where d_r is defined in Section 4.3, $\overline{f}_{p,j} = (j\, e_0)^{-1} \sum_{v=1}^{j} [v(p+1) - j]\, e_v\, \overline{f}_{p,j-m}$ for $j \geq 1$, $\overline{f}_{p,0} = e_0^p$, $e_j = \sum_{t=0}^{\infty} q_t\, c_{t,j}$ and $q_t = \sum_{t=0}^{\infty} b_t\, s_t(a+k)$. The quantity $c_{t,j}$ was just defined after equation (4.19).

Equations (4.13)-(4.15), (4.18)-(4.20) and (4.22) are the main results of this section. Some algebraic details are given in Appendix B.

The skewness and kurtosis measures can be calculated from the ordinary moments using well- known relationships. Plots of the skewness and kurtosis for selected parameters values as function of a are displayed in Figure 4.2 and 4.3, respectively . In the plots of Figures 4.2(a) and 4.3(c), $\sigma = 10.50$, whereas in those of Figures 4.2(b) and 4.3(d), $\mu = 2.50$.

4.6 Generating function

The generating function $M(-t) = E(\mathrm{e}^{-tZ})$ of $Z \sim$GN$(a, 0, 1)$ is given by

$$M(-t) = \frac{1}{\sqrt{2}\pi} \sum_{k,r=0}^{\infty} b_k\, s_{r+1}(a+k) \int_{-\infty}^{\infty} \Phi(x)^r \exp\left(-tx - \frac{x^2}{2}\right) dx.$$

Inserting equation (4.17), we obtain

$$M(-t) = \frac{1}{\sqrt{2}\pi} \sum_{k,r,j=0}^{\infty} b_k\, s_{r+1}(a+k)\, c_{r,j} \int_{-\infty}^{\infty} x^j \exp\left(-tx - \frac{x^2}{2}\right) dx.$$

Based on Prudnikov *et al.* (1986, Eq.2.3.15.8), the above integral can be rewritten as

$$J(s,j) = \int_{-\infty}^{\infty} x^j \exp\left(-sx - \frac{x^2}{2}\right) dx = (-1)^j \sqrt{2\pi}\, \frac{\partial^j}{\partial s^j} \left(\mathrm{e}^{s^2/2}\right).$$

Thus, the mgf of Z becomes

$$M(-t) = \frac{1}{\sqrt{2}\pi} \sum_{k,r,j=0}^{\infty} b_k\, s_{r+1}(a+k)\; c_{r,j}\, J(s,j). \tag{4.23}$$

A second representation for $M(t)$ can be based on the qf. We have

$$M(t) = \int_0^1 \exp\left[t\, Q_{GN}(u)\right]\, du.$$

Expanding the exponential function, using (4.12) and after some algebra, we obtain

$$M(t) = \sum_{k,r=0}^{\infty} \frac{d_{k,r}}{\left(\frac{r}{a}+1\right)} \frac{t^k}{k!}, \tag{4.24}$$

where $d_{k,r} = (r\, g_0)^{-1} \sum_{m=1}^{r} [m(k+1) - r]\, g_m\, d_{k,r-m}$ for $r \geq 1$, $d_{k,0} = g_0^k$, $d_{0,r} = 1$, $g_j = \overline{p}_j\, h_{j,r}$ and the quantities $\overline{p}_j$ and $h_{j,r}$ are given in Section 4.4.

Equations (4.23) and (4.24) are the main results of this section. The mgf of X is simply given by $M_X(t) = \mathrm{e}^{\mu}\, M(\sigma t)$. The cf has many useful and important properties which gives it a central role in statistical theory. Its approach is particularly useful in analysis of linear combination of independent random variables. Clearly, a simple representation for the cf $\phi_X(t) = M_X(\mathrm{i}t)$ of X, where $\mathrm{i} = \sqrt{-1}$, is given by

$$\phi_X(t) = \int_0^{\infty} \cos(tx)\, f(x) dx + \mathrm{i} \int_0^{\infty} \sin(tx)\, f(x) dx.$$

From the expansions $\cos(tx) = \sum_{r=0}^{\infty} \frac{(-1)^r}{(2r)!} (tx)^{2r}$ and $\sin(tx) = \sum_{r=0}^{\infty} \frac{(-1)^r}{(2r+1)!} (tx)^{2r+1}$, we obtain

$$\phi_x(t) = \sum_{r=0}^{\infty} \frac{(-1)^r\, t^{2r}}{(2r)!}\, E(X^{2r}) + \mathrm{i} \sum_{r=0}^{\infty} \frac{(-1)^r\, t^{2r+1}}{(2r+1)!}\, E(X^{2r+1}).$$

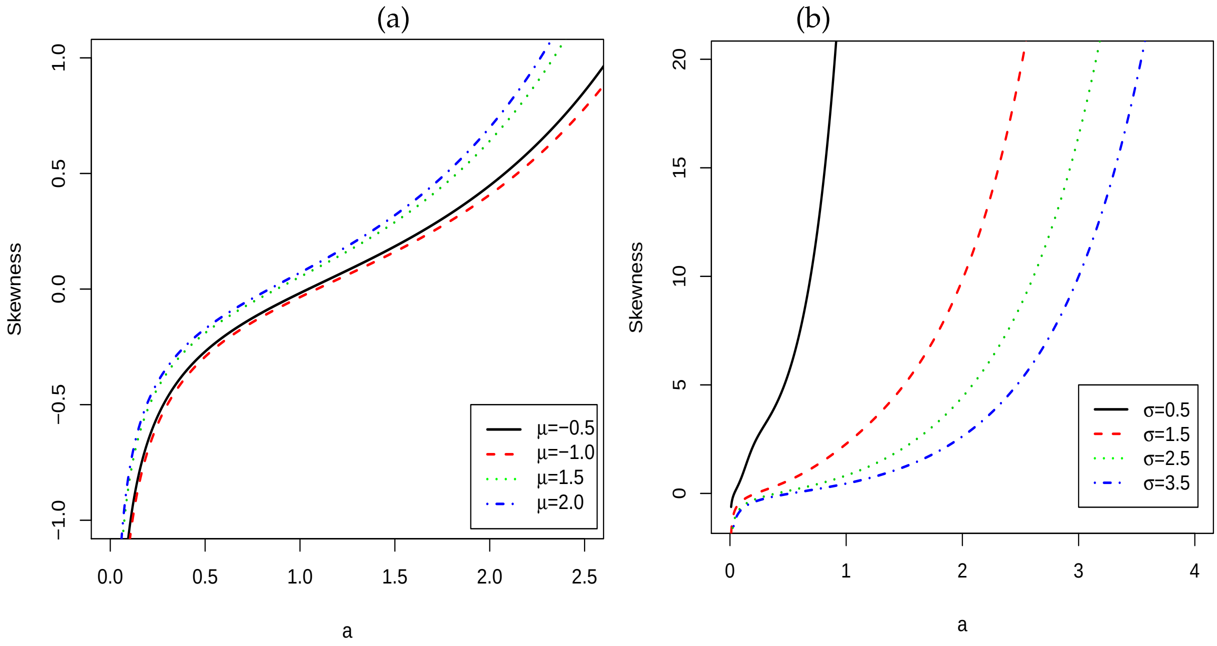


Figure 4.2: (a) Skewness of X as function of a for some values of μ. (b) Skewness of X as function of a for some values of σ.

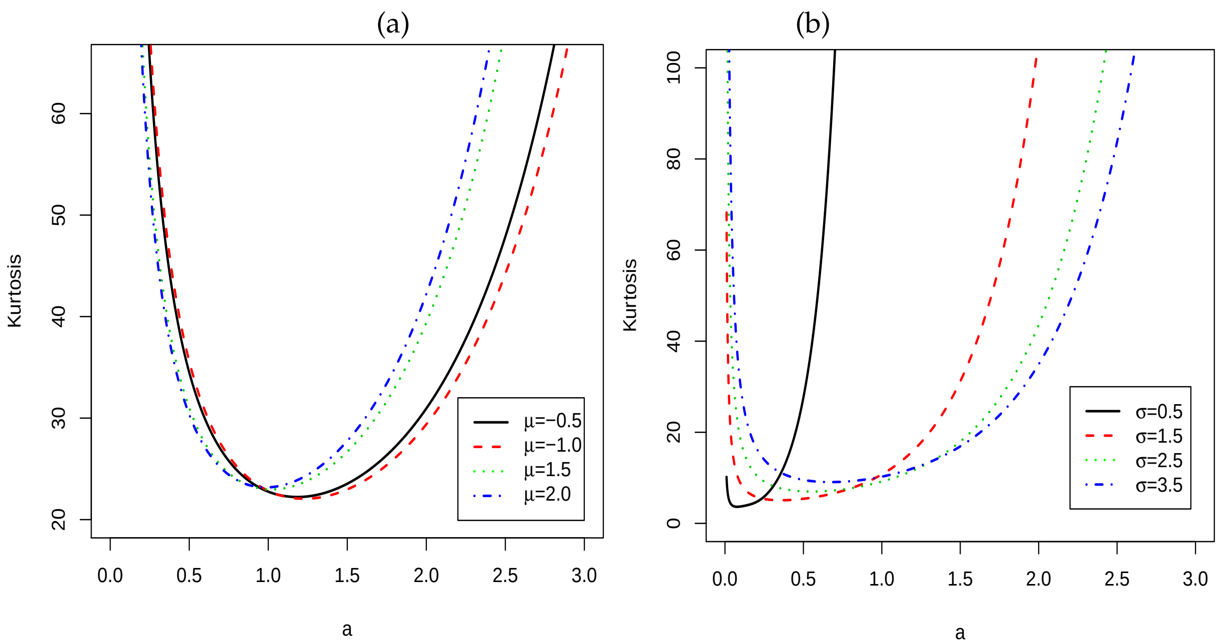


Figure 4.3: (a) Kurtosis of X as function of a for some values of μ. (b) Kurtosis of X as function of a for some values of σ.

4.7 Entropies

Here, we consider the random variable $Z \sim$GN(a,0,1). Thus, the Rényi entropy is defined as

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \int_{-\infty}^{+\infty} f^{\gamma}(x)dx$$

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

First, we consider $\gamma = n = 2, 3, \ldots$, $\mu = 0$, $\sigma = 1$ and the rth moment of the standard normal distribution given by

$$m'_r = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^r \mathrm{e}^{-x^2/2}. \tag{4.25}$$

We have two cases: $m'_r = 0$, if r is odd, and $m'_r = 1 \times 3 \ldots \times (r-1)$, if r is even.

Using (4.17), we can write from (4.5) and (4.17)

$$\begin{aligned} I_R(n) &= \frac{1}{1-n} \log \left\{ \left(\frac{1}{\sqrt{2\pi}} \right)^n \sum_{j=0}^{n} \wp_{n,j} \int_{-\infty}^{+\infty} x^j \, \mathrm{e}^{-\frac{n x^2}{2}} dx \right\} \\ &= \frac{1}{1-n} \left\{ -\frac{(n-1)}{2} \log(2\pi) - \frac{(j+1)}{2} \log(n) + \log \left[\sum_{j=0}^{n} \wp_{n,j} m j' \right] \right\}, \end{aligned} \tag{4.26}$$

where $\wp_{n,j} = (j\,\bar{e}_0)^{-1} \sum_{m=1}^{j} [m(n+1)-j]\, \bar{e}_m \, \wp_{n,j-m}$, $\bar{e}_j = \sum_{j=0}^{\infty} d_r \, c_{r,j}$, $\wp_{n,0} = \bar{e}_0^n$ and the m_j's are given by (4.25). The quantities d_r's are defined in Section 4.3, whereas the $c_{r,j}$'s and the a_v's are given in Section 4.5.

We can write $I_R(\gamma) = (1-\gamma)^{-1} E\{f(Z)^{\gamma-1}\}$. Let $\delta = E(Z)$. For γ real positive, we have

$$E\{f(Z)^{\gamma-1}\} = \delta^{\gamma-1} E\{1 + \theta\,[f(Z) - \delta]\}^{\gamma-1},$$

where $\theta = \delta^{-1}$. From the generalized binomial expansion, we obtain

$$\{1 + \theta\,[f(Z) - \delta]\}^{\gamma-1} = 1 + \sum_{n=1}^{\infty} \frac{\theta^n \, \Im_n}{n!} \, [f(Z) - \delta]^n,$$

where $\Im_n = \prod_{j=0}^{n-1} (\gamma - 1 - j)$. Further,

$$E\{f(Z)^{\gamma-1}\} = \delta^{\gamma-1} \left(1 + \sum_{n=2}^{\infty} \frac{\theta^n \, \Im_n}{n!} \, E\{[f(Z) - \delta]^n\} \right). \tag{4.27}$$

We now obtain $E\{[f(Z)]^n\}$ for $n \geq 2$. From equation (4.5) and using the binomial expansion, we can write

$$\rho_n = E\{[f(Z)]^n\} = \sum_{j=0}^{\infty} \wp_{n,j} \, \psi_{n,j},$$

where $\psi_{n,j} = E\{Z^j \, \phi(Z)^n\}$. Thus,

$$\psi_{n,j} = \int_{-\infty}^{\infty} x^j \, \phi(x)^{n+1} \, \mathrm{d}x.$$

Setting $\sqrt{(n+1)}\,x = y$, we can easily determine the last integral and then rewrite ρ_n as

$$\rho_n = \left(\frac{1}{\sqrt{2\pi}}\right)^n \sum_{j=0}^{\infty} \wp_{n,j} \left(\frac{1}{\sqrt{n+1}}\right)^{j+1} m_j{}'. \tag{4.28}$$

By expanding the binomial term in (4.27), we can obtain an explicit expression for $I_R(\gamma)$, which holds for any γ real positive and $\gamma \neq 1$, given by

$$I_R(\gamma) = (1-\gamma)^{-1}\,\delta^{\gamma-1}\left[1 + \sum_{n=2}^{\infty} \frac{\theta^n\,\Im_n}{n!} \sum_{k=0}^{n} (-\delta)^{n-k} \binom{n}{k} \rho_k\right], \tag{4.29}$$

where ρ_k is determined from (4.28). Algebraic details can be found in Appendix D.

Next, the Shannon entropy of a random variable Z is defined by $E\{-\log[f(Z)]\}$. It is a special case of the Rényi entropy when $\gamma \uparrow 1$. Equation (4.26) is very complicated for limiting, and then we derive an explicit expression for the Shannon entropy from its definition. We can write

$$\begin{aligned}\log[f(x)] &= \log\left\{\frac{1}{\sigma\Gamma(a)}\phi(x)\left\{-\log\left[1-\Phi(x)\right]\right\}^{a-1}\right\}\\ &= -\log[\sigma\Gamma(a)] + \log[\phi(x)] + (a-1)\log\left\{-\log[1-\Phi(x)]\right\}.\end{aligned} \tag{4.30}$$

So, we first calculate $E\left\{\log[\phi(X)]\right\}$ and $E\left[\log\left\{-\log[1-\Phi(X)]\right\}\right]$. Setting $\mu = 0$ and $\sigma = 1$, the first quantity is easily calculated as follows

$$E\left\{\log[\phi(X)]\right\} = -\frac{1}{2}\log(2\pi) - E\left(\frac{X^2}{2}\right) = -\frac{1}{2}\left[\log(2\pi) + \mu_2{}'\right], \tag{4.31}$$

where $\mu_2{}'$ comes from (4.13) or (4.14) with $n = 2$.

The second quantity $E\left[\log\left\{-\log[1-\Phi(x)]\right\}\right]$ is obtained from the expansion of $\log\left\{-\log[1-\Phi(x)]\right\}$. We can write (for $0 < u < 1$) from MATHEMATICA

$$\begin{aligned}\log\left\{-\log[1-u]\right\} &= \log(u) + \frac{u}{2} + \frac{5u^2}{24} + \frac{u^3}{8} + \frac{251u^4}{2880} + \frac{19u^5}{288} + \frac{19087u^6}{362880} + \frac{751u^7}{17280} + \frac{1070017u^8}{29030400}\\ &\quad + \frac{2857u^9}{89600} + \frac{26842253u^{10}}{958003200} + O(u^{11}).\end{aligned} \tag{4.32}$$

From equations (4.30)-(4.32), we obtain the Shanon entropy $E\{-\log[f(Z)]\}$ using the ordinary moments given by (4.13), (4.14) and (4.17). Equations (4.26), (4.29)-(4.32) are the main results of this section.

4.8 Order statistics

Suppose $Z_1, \ldots, Z_n$ is a random sample from the standard GN distribution and let $Z_{1:n} < \cdots < Z_{i:n}$ denote the corresponding order statistics. Using (4.5) and (4.6), the pdf of $Z_{i:n}$ can be expressed as

$$\begin{aligned}f_{i:n}(z) &= \frac{n!}{(i-1)!(n-i)!}\sum_{j=0}^{n-i}(-1)^j\binom{n-i}{j}\left[\sum_{r=0}^{\infty} b_r\,(a+r)\,\Phi(z)^{a+r-1}\,\phi(z)\right]\times\\ &\quad\times\left[\sum_{k=0}^{\infty} b_k\,\Phi(z)^{a+k}\right]^{i+j-1}.\end{aligned}$$

Based on equations (2.7) and (2.8), we obtain

$$\left[\sum_{k=0}^{\infty} b_k\,\Phi(z)^{a+k}\right]^{i+j-1} = \sum_{k=0}^{\infty} \eta_{i+j-1,k}\,\Phi(z)^{(i+j-1)a+k},$$

where $\eta_{i+j-1,0} = b_0^{i+j-1}$ and $\eta_{i+j-1,k} = (kb_0)^{-1}\sum_{m=1}^{k}\left[m(i+j)-k\right] b_m\,\eta_{i+j-1,k-m}$. Hence, the pdf of $Z_{i:n}$ reduces to

$$f_{i:n}(z) \;=\; \phi(z)\sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} m_{j,k,r}\,\Phi(z)^{(i+j)a+k+r-1}, \tag{4.33}$$

where

$$m_{j,k,r} = \frac{(-1)^j\,(a+r)\,n!\,b_r\,\eta_{i+j-1,k}}{(i-1)!\,(n-i-j)!\,j!}.$$

Equation (4.33) can be expressed as

$$f_{i:n}(z) \;=\; \sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} f_{j,k,r}\,h_{(i+j)a+k+r}(z), \tag{4.34}$$

where

$$f_{j,k,r} = \frac{m_{j,k,r}}{[(i+j)a+k+r]}\,.$$

Equation (4.34) is the main result of this section. It reveals that the pdf of the standard GN order statistics is a triple linear combination of exp-N densities with parameters $(i+j)a+k+r$, $\mu = 0$ and $\sigma = 1$. So, several mathematical quantities of the GN order statistics such as ordinary and incomplete moments, mgf and mean deviations can be immediately obtained from those quantities of the exp-N distribution. It gives the pdf of the GN order statistics as a power series of the standard normal cdf multiplied by the standard normal density function.

As an application of (4.33), the s-*th* ordinary moment of $Z_{i:n}$ becomes

$$E(Z_{i:n}^s) = \sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} m_{j,k,r}\tau_{s,(i+j)a+k+r-1},$$

where $\tau_{s,(i+j)a+k+r-1}$ can be obtained from (4.15).

Another closed-form expression for $E(Z_{i:n}^s)$ can be derived using a result due to Barakat and Abdelkader (2004) applied to the independent and identically distributed case. Thus,

$$E(Z_{i:n}^s) = s\sum_{j=n-i+1}^{n}(-1)^{i+j-n-1}\binom{j-1}{n-i}\binom{n}{j}J_j(s),$$

where $J_j(s) = \int_0^\infty z^{s-1}\,[1-F(z)]^j dx$. By expanding $[1-F(z)]^j$ and using (4.6) , we obtain $J_j(s)$. For any real $a > 0$, we can write from equations (4.6) and (2.8)

$$\begin{aligned}
J_j(s) \;&=\; \sum_{m=0}^{j}(-1)^m\binom{j}{m}\int_0^\infty z^{s-1}\left(\sum_{k=0}^{\infty} b_k\,\Phi(z)^{a+k}\right)^m dx \\
&=\; \sum_{m=0}^{j}(-1)^m\binom{j}{m}\sum_{k=0}^{\infty} d_{m,k}\,\tau_{s-1,ma+k} = \sum_{k=0}^{\infty}\sum_{m=0}^{j}(-1)^m\binom{j}{m}d_{m,k}\,\tau_{s-1,ma+k},
\end{aligned}$$

where $d_{m,k}$ is defined in Section 4.5 and the quantities $\tau_{n,r}$ are given in equation (4.15).

4.9 Estimation

Here, we consider estimation of the unknown parameters of the GL distribution by the method of maximum likelihood. Let $x_1, \ldots, x_n$ be a random sample of size n from the GN(a, μ, σ) distribution. The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (a, \mu, \sigma)^T$ can be expressed as

$$\begin{aligned} l(\boldsymbol{\theta}) &= -n\log(\sigma) - n\log[\Gamma(a)] + \sum_{i=1}^{n}\log\left[\phi\left(\frac{x_i-\mu}{\sigma}\right)\right] \\ &\quad +(a-1)\sum_{i=1}^{n}\log\left\{-\log\left[1-\Phi\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\}. \end{aligned} \tag{4.35}$$

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

$$\begin{aligned} U_a(\boldsymbol{\theta}) &= -n\psi(a) + \sum_{i=1}^{n}\log\left\{-\log\left[1-\Phi\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\}, \\ U_\mu(\boldsymbol{\theta}) &= \frac{1}{\sigma} + \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) + \frac{(a-1)}{\sigma}\sum_{i=1}^{n}\frac{\phi\left(\frac{x_i-\mu}{\sigma}\right)}{[1-\Phi\left(\frac{x_i-\mu}{\sigma}\right)]\log[1-\Phi\left(\frac{x_i-\mu}{\sigma}\right)]}, \\ U_\sigma(\boldsymbol{\theta}) &= -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \frac{(a-1)}{\sigma}\sum_{i=1}^{n}\frac{\left(\frac{x_i-\mu}{\sigma}\right)\phi\left(\frac{x_i-\mu}{\sigma}\right)}{[1-\Phi\left(\frac{x_i-\mu}{\sigma}\right)]\log[1-\Phi\left(\frac{x_i-\mu}{\sigma}\right)]}, \end{aligned}$$

where $\psi(\cdot)$ is the digamma function.

Setting these expressions to zero and solving them simultaneously yields the maximum likelihood estimates (MLEs) of the three parameters, under some regularity conditions. For more details, see Nocedal and Wright (1999, chapter 8) We use the matrix programming language Ox (MaxBFGS subroutine), see for example, Doornik (2006) and the procedure NLMixed in SAS to compute the MLE $\widehat{\theta}$. For interval estimation of the model parameters, we require the expected information matrix. The 3×3 total observed information matrix $J(\boldsymbol{\theta})$ is given by

$$J(\boldsymbol{\theta}) = \begin{pmatrix} J_{aa} & J_{a\mu} & J_{a\sigma} \\ . & J_{\mu\mu} & J_{\mu\sigma} \\ . & . & J_{\sigma\sigma} \end{pmatrix},$$

whose elements are listed in Appendix E. Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ is $N_3(\mathbf{0}, K(\boldsymbol{\theta})^{-1})$, where $K(\boldsymbol{\theta}) = E\{J(\boldsymbol{\theta})\}$ is the expected information matrix. The multivariate normal $N_3(0, J(\boldsymbol{\theta})^{-1})$ distribution can be used to construct approximate confidence intervals for the parameters.

The LR can be used for testing the goodness of fit of the GL distribution and for comparing this distribution with the normal model. We can compute the maximum values of the unrestricted and restricted log-likelihoods to construct LR statistics for testing some sub-models of the GL distribution. For example, we may use the LR statistic to check if the fit using the new distribution is statistically "superior" to a fit using the normal distribution for a given data set. In any case, hypothesis tests of the type $H_0 : \psi = \psi_0$ versus $H : \psi \neq \psi_0$, where ψ is a vector formed with some components of θ and ψ_0 is a specified vector, can be performed using LR statistics. For example, the test of $H_0 : a = 1$ versus $H :$ H_0 *is not true* is equivalent to compare the GN and normal distributions and then the LR statistic reduces to $w = 2\{\ell(\widehat{a},\widehat{\mu},\widehat{\sigma}) - \ell(1,\widetilde{\mu},\widetilde{\sigma})\}$, where $\widehat{a}$, $\widehat{\mu}$ and $\widehat{\sigma}$ are the MLEs under H and $\widetilde{\mu}$ and $\widetilde{\sigma}$ are the estimates under H_0.

4.10 Applications

In this section, the potentiality of the GN model is illustrated in two applications to real data. An alternative analysis of these data can be performed using the normal distribution. The beta-normal (BN) (Eugene *et al.*, 2002) and Kumaraswamy-normal (KwN) models extend the normal model and they can also be used to fit data that come from a distribution with heavy tails reducing the influence of aberrant observations.

The BN distribution

The BN pdf with parameters μ and σ and two extra shape parameters $\alpha > 0$ and $\beta > 0$ is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\sigma\Gamma(\alpha)\Gamma(\beta)} \left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1} \left[1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\beta-1} \phi\left(\frac{x-\mu}{\sigma}\right), \qquad (4.36)$$

$-\infty < x < \infty$. For $\alpha = \beta = 1$, we obtain the normal distribution. Recently, Alexander *et al.* (2012) and Cordeiro *et al.* (2012) proposed the generalized beta-generated and McDonald normal distributions, respectively. The first generated model contains, as special cases, several important distributions discussed in the literature such as the normal, exponentiated normal, BN and KwN distributions, among others.

Kumaraswamy-normal (KwN) distribution

The KwN pdf with parameters μ and σ and two extra shape parameters $a > 0$ and $b > 0$ is given by

$$f(x) = \frac{ab}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{a-1} \left[1-\Phi^{a}\left(\frac{x-\mu}{\sigma}\right)\right]^{b-1}, \quad -\infty < x < \infty. \qquad (4.37)$$

For $a = b = 1$, we have the normal distribution. Clearly, equation (4.37) is much simpler than (4.36).

4.10.1 Application 1: Carbohydrates data

The first example refers to the data from agronomic experiments (Matsuo, 1986) conducted at the Federal University of Paraná. The main objective was to verify the content of carbohydrates (in %) of the corn farms. Some summary statistics for the CO data are: mean=66.34, median=66.64, minimum=62.35 and maximum=68.46.

The parameters of each model are estimated by maximum likelihood (Section 4.7) using the subroutine NLMixed in SAS. We report the MLEs (and the corresponding standard errors in parentheses) of the parameters and the values of the Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Bayesian Information Criterion (BIC) in Table 4.1. The lower the values of these criteria, the better the fit. Since the values of these statistics are smaller for the GN distribution compared to their values for the other three models, we can conclude that the new distribution is the best model among the four to explain the current data. An analysis under the GN model also provides a check on the appropriateness of the normal model and indicates the extent for which inferences depend upon the model. For example, the LR statistic for testing the hypothesis $H_0 : a = 1$ versus $H : H_0$ is not true, i.e. to compare the GN and normal models, is $w = 2\{-63.05 - (65.20)\} = 4.30$ (p-value $= 0.0381$), which provides support toward the new model.

Table 4.1: MLEs and information criteria.

Carbohydrate		a	μ	σ	AIC	CAIC	BIC
GN		0.1454	68.3276	0.7443	132.1	132.9	136.9
		(0.0277)	(0.2963)	(0.0388)			
Normal		1	66.3379	1.4800	134.3	134.8	137.6
		-	(0.2467)	(0.1744)			
	α	β	μ	σ			
BN	0.1167	0.0678	65.5745	0.3683	137.3	138.6	143.6
	(0.0471)	(0.0129)	(0.3649)	(0.0475)			
	a	b	μ	σ			
KwN	0.1859	0.0309	66.6857	0.3460	132.2	133.3	138.3
	(0.2023)	(0.0281)	(1.1912)	(0.1034)			

4.10.2 Application 2: Carbon monoxide data

Here, we work with carbon monoxide (CO) measurements made in several brands of cigarettes in 1994. The data have been collected by the Federal Trade Commission (FTC), an independent agency of the United States government, whose main mission is the promotion of consumer protection. For three decades the FTC regularly has released reports on the nicotine and tar content of cigarettes. The reports indicate that nicotine levels, on average, had remained stable since 1980, after falling in the preceding decade. The report entitled "Tar, Nicotine, and Carbon Monoxide of the Smoke of 1206 Varieties of Domestic Cigarettes for the year of 1994" at `https://www.erowid.org/plants/tobacco/tobacconic.shtml` includes some information about the source of the data, smoker's behavior and beliefs about nicotine, tar and carbon monoxide contents in cigarettes. The data are in Appendix F.

The data include $n = 384$ records of CO measurements, in milligrams, in cigarettes of several brands. Some summary statistics for the CO data are: mean=11.34, median=12.00,

Table 4.2: MLEs and information criteria.

Carbon monoxide		a	μ	σ	AIC	CAIC	BIC
GN		0.1432	16.9819	2.0889	1931.8	1931.9	1943.3
		(0.0085)	(0.2476)	(0.0378)			
Normal		1	11.3425	4.0626	1950.4	1950.5	1958.1
		-	(0.2187)	(0.1547)			
	α	β	μ	σ			
BN	0.2143	3.1422	18.5092	2.8673	1932.9	1933.0	1948.3
	(0.0906)	(0.4851)	(0.4680)	(0.5866)			
	a	b	μ	σ			
KwN	0.2242	0.0730	11.8209	1.2921	1929.1	1929.2	1944.5
	(0.0420)	(0.0262)	(1.1516)	(0.1306)			

minimum=0.05 and maximum=22.00. In each case, the parameters are estimated by maximum likelihood using the subroutine NLMixed in SAS. We report the MLEs (and the corresponding standard errors in parentheses) of the parameters and the values of the AIC, CAIC and BIC statistics in Table 4.2. Since the values of these statistics are smaller for the GN and KwN distributions compared to those values for the other models, the new distribution is a very competitive model to explain these data and it is more parsimonious. The LR statistic for comparing the GN and normal models is $w = 2\{-962.9 - (-1946.4)\} = 20.6$ (p-value $=< 0.0001$), which yields favorable support toward to the first model.

4.11 Concluding remarks

In this chapter, we propose a new model called the gamma-normal distribution which extends the normal distribution. The proposed distribution is very versatile to fit real data and could be a good alternative to the normal and two recent generalizations of this distribution. We study some of its structural properties. We provide explicit expressions for the ordinary and incomplete moments, quantile and generating functions, mean deviations, Rényi entropy, Shannon entropy, order statistics and their moments. We derive a power series expansion for its quantile function which is useful to obtain alternative formulae for several mathematical measures. The model parameters are estimated by maximum likelihood and the observed information matrix is determined. The potentiality of the new model is illustrated by means of two examples.

4.12 Appendix

Appendix A: Quantile function

We derive a power series for the $Q_{GN}(u)$ in the following way. First, we use a known power series for $Q^{-1}(a, 1-u)$. Second, we obtain a power series for the argument $1 - \exp[-Q^{-1}(a, 1-u)]$. Third, we consider the power series for the normal quantile function given in Steinbrecher (2002) to obtain a power series for $Q_{GN}(u)$.

We introduce the following quantities defined by Cordeiro and Lemonte (2011). Let $Q^{-1}(a, z)$ be the inverse function of

$$Q(a,z) = 1 - \frac{\gamma(a,z)}{\Gamma(a)} = \frac{\Gamma(a,z)}{\Gamma(a)} = u.$$

The inverse quantile function $Q^{-1}(a, 1-u)$ is determined in the Wolfram website [1] as

$$\begin{aligned} Q^{-1}(a,1-u) &= w + \frac{w^2}{a+1} + \frac{(3a+5)w^3}{2(a+1)^2(a+2)} + \frac{[a(8a+33)+31]w^4}{3(a+1)^3(a+2)(a+3)} \\ &+ \frac{\{a(a[a(125a+1179)+3971]+5661)+2888\}\, w^5}{24(a+1)^4(a+2)^2(a+3)(a+4)} + O(w^6), \end{aligned}$$

where $w = [u\Gamma(a+1)]^{1/a}$. We can write the last equation as

$$z = Q^{-1}(a, 1-u) = \sum_{r=0}^{\infty} \delta_r\, u^{r/a}, \tag{4.38}$$

where δ_i's is given by $\delta_i = \overline{b}_i\, \Gamma(a+1)^{i/a}$. Here, $\overline{b}_0 = 0$, $\overline{b}_1 = 1$ and any coefficient $\overline{b}_{i+1}$ (for $i \geq 1$) can be obtained from the cubic recurrence equation

$$\overline{b}_{i+1} = \frac{1}{i(a+i)} \left\{ \sum_{r=1}^{i} \sum_{s=1}^{i-s+1} \overline{b}_r \overline{b}_s \overline{b}_{i-r-s+2}\, s\,(i-r-s+2) \sum_{r=2}^{i} \overline{b}_r \overline{b}_{i-r+2}\, r\,[r-a-(1-a)(i+2-r)] \right\}.$$

The first coefficients are $\overline{b}_2 = 1/(a+1)$, $\overline{b}_3 = (3a+5)/[2(a+1)^2(a+2)]$, ... Now, we present some algebraic details for the GN qf, say $Q_{GN}(u)$. The cdf of X is given by (4.4). The normal quantile function can be expressed as (Steinbrecher, 2002)

$$Q_N(u) = \Phi^{-1}(x) = \sum_{k=0}^{\infty} d_k \left[\sqrt{2\,\pi}\,(u-1/2) \right]^k, \tag{4.39}$$

where the coefficients d_k's are defined by $d_k = 0$ for $k = 0, 2, 4, \ldots$ and $d_k = e_{(k-1)/2}$ for $k = 1, 3, 5, \ldots$ The quantities e_k's are determined recursively from

$$e_{k+1} = \frac{1}{2(2k+3)} \sum_{r=0}^{k} \frac{(2r+1)(2k-2r+1)\, e_r\, e_{k-r}}{(r+1)(2r+1)}.$$

[1] http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/06/01/03/

Expanding the binomial term in (7.27), we obtain

$$Q_N(u) = \sum_{k=0}^{\infty}(\sqrt{2\pi})^k\, d_k \sum_{s=0}^{k}(-2)^{s-k}\binom{k}{s}u^s = \sum_{k=0}^{\infty}\sum_{s=0}^{k}(\sqrt{2\pi})^k\,(-2)^{s-k}\,d_k\binom{k}{s}u^s.$$

Changing $\sum_{k=0}^{\infty}\sum_{s=0}^{k}$ by $\sum_{s=0}^{\infty}\sum_{k=s}^{\infty}$, we have

$$Q_N(u) = \sum_{s=0}^{\infty}\sum_{k=s}^{\infty}(\sqrt{2\pi})^k(-2)^{s-k}\,d_k\binom{k}{s}u^s,$$

and then $Q_N(u) = \sum_{s=0}^{\infty} w_s\, u^s$, where $w_s = \sum_{k=s}^{\infty}(-2)^{s-k}\,(\sqrt{2\pi})^k\binom{k}{s}d_k$ and the quantity d_k was defined above.

By replacing (4.38) in equation (4.10), we can write

$$Q_{GN}(u) = \mu + \sigma\, Q_N\left\{1 - \exp\left[-\sum_{r=0}^{\infty}\delta_r\, u^{r/a}\right]\right\}.$$

By expanding the exponential function and using (2.7), we have

$$\begin{aligned} 1-\exp\left(-\sum_{r=0}^{\infty}\delta_r\, u^{r/a}\right) &= 1-\sum_{l=0}^{\infty}\frac{(-1)^l\,\left(\sum_{r=0}^{\infty}\delta_r\, u^{r/a}\right)^l}{l!} \\ &= 1-\sum_{l=0}^{\infty}\frac{(-1)^l\,\sum_{r=0}^{\infty}f_{l,r}\,u^{r/a}}{l!} = 1-\sum_{r=0}^{\infty}p_r\,u^{r/a}, \end{aligned} \qquad (4.40)$$

where $p_r = \sum_{l=0}^{\infty}\frac{(-1)^l f_{l,r}}{l!}$, $f_{l,r} = (r\,\delta_0)^{-1}\sum_{q=1}^{r}[q(l+1)-r]\,\delta_m\, f_{l,r-q}$ for $r \geq 1$ and $f_{l,0} = \delta_0^l$. Combining (4.10) and (4.40), we obtain

$$Q_{GN}(u) = \mu + \sigma\, Q_N\left(1-\sum_{r=0}^{\infty}p_r\,u^{r/a}\right).$$

Using the know result for $Q_N(u)$ in the last equation and expanding the binomial term, we have

$$\begin{aligned} Q_{GN}(u) &= \mu+\sigma\left\{\sum_{s=0}^{\infty}w_s\left(1-\sum_{r=0}^{\infty}p_r\,u^{r/a}\right)^s\right\} \\ &= \mu+\sigma\left\{\sum_{s=0}^{\infty}w_s\sum_{j=0}^{s}(-1)^j\binom{s}{j}\left(\sum_{r=0}^{\infty}p_r u^{r/a}\right)^j\right\}. \end{aligned}$$

Now, using (2.7), we obtain

$$\begin{aligned} Q_{GN}(u) &= \mu+\sigma\left\{\sum_{s=0}^{\infty}\sum_{j=0}^{s}(-1)^j\binom{s}{j}w_s\sum_{r=0}^{\infty}h_{j,r}\,u^{r/a}\right\} \\ &= \mu+\sigma\left\{\sum_{s,r=0}^{\infty}\sum_{j=0}^{s}(-1)^j\,w_s\,h_{j,r}\binom{s}{j}u^{r/a}\right\}, \end{aligned}$$

where $h_{j,r} = (r\,p_0)^{-1}\sum_{m=0}^{r}[m(j+1)-r]\,p_m\,h_{j,r-m}$. Finally,

$$Q_{GN}(u) = \mu+\sigma\sum_{r=0}^{\infty}\overline{p}_j\,h_{j,r}\,u^{r/a},$$

where $\overline{p}_j = \sum_{s=0}^{\infty}\sum_{j=0}^{s}(-1)^j\,w_s\binom{s}{j}$.

Appendix B: Moments

Here, we use equation (2.7) and the power series $\Phi(x) = \sum_{j=0}^{\infty} a_j\, x^j$ given in Section 4.5. We have

$$T_n(y) = \int_0^y x^n\, f(x) dx.$$

Inserting (4.5) (with $\mu = 0$ and $\sigma = 1$) in the last equation gives

$$T_n(y) = \sum_{r=0}^{\infty} d_r \int_0^y x^n\, \phi(x)\, \Phi(x)^r dx.$$

From the power series for $\Phi(x)$ and equation (2.7), we have

$$T_n(y) = \sum_{j,r=0}^{\infty} d_r\, c_{r,j} \int_0^y x^{n+j}\, \phi(x) dx = \frac{1}{\sqrt{2\,\pi}} \sum_{j,r=0}^{\infty} d_r\, c_{r,j} \int_0^y x^{n+j}\, \mathrm{e}^{-x^2/2}\, dx,$$

where d_r is defined in Section 4.3 and the quantities $c_{r,j}$ are obtained from (2.8) using the a_i's of the power series for $\Phi(x)$. Setting $z = x^2/2$, we obtain

$$T_n(y) = \frac{1}{\sqrt{2\,\pi}} \sum_{j,r=0}^{\infty} d_r\, c_{r,j} \int_0^{y^2/2} (2z)^{\frac{n+j-1}{2}}\, \mathrm{e}^{-z}\, dz = \frac{1}{\sqrt{2\,\pi}} \sum_{j,r=0}^{\infty} 2^{n+j-1}\, d_r\, c_{r,j}\, \gamma\left(\frac{n+j+1}{2}, \frac{y^2}{2}\right),$$

where $\gamma(\cdot,\cdot)$ is the gamma incomplete function.

The second representation for $T_n(y)$ is based on the integral $A(j,q) = \int_{-\infty}^{q} x^j\, \mathrm{e}^{-x^2/2} dx$, which is determined for $q > 0$ and $q < 0$. We define

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

For $q < 0$ and $q > 0$, we have

$$A(j,q) = (-1)^j\, G(j) + (-1)^{j+1} H(j,q) \quad \text{and} \quad A(j,q) = (-1)^j\, G(j) + H(j,q),$$

respectively, where the integral $H(j,q) = \int_0^j x^j\, \mathrm{e}^{-x^2/2} dx$ can be easily determined as (Whittaker and Watson, 1990)

$$\begin{aligned} H(j,q) &= \frac{2^{j/4+1/4} q^{j/2+1/2} e^{-q^2/4}}{(j/2+1/2)(j+3)} N_{j/4+1/4,j/4+3/4}(q^2/2) \\ &+ \frac{2^{j/4+1/4} q^{j/2-3/2} e^{-q^2/4}}{j/2+1/2} N_{j/4+5/4,j/4+3/4}(q^2/2), \end{aligned}$$

where $N_{k,m}(x)$ is the Whittaker function (Abramowitz and Stegun, 1972, p. 505; Whittaker and Watson 1990, pp. 339-351) given, in terms of the confluent hypergeometric function ${}_1F_1(a;b;z) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{z^k}{k!}$, or in terms of the Kummer's function $U(a,b;z) = z^{-a}\, {}_2F_0(a, 1+a-b; -z^{-1})$, where $(a)_k$ was defined in Section 4.5. We have

$$N_{k,m}=\frac{x^{m+1/2}}{\mathrm{e}^{-x/2}}\,{}_1F_1\left(\frac{1}{2}+m-k,1+2m;x\right)\quad\text{and}\quad N_{k,m}=\frac{x^{m+1/2}}{\mathrm{e}^{-x/2}}\,U\left(\frac{1}{2}+m-k,1+2m;x\right).$$

Combining (4.5) and (4.17), we can write

$$\begin{aligned}T_n(y) &= \frac{1}{\sqrt{2\pi}}\sum_{k,r=0}^{\infty} b_k\,s_{r+1}(a+k)\int_0^y x^n\,\mathrm{e}^{-x^2/2}\sum_{j=0}^{\infty} c_{r,j}\,x^j dx\\ &= \frac{1}{\sqrt{2\pi}}\sum_{j,k,r=0}^{\infty} b_k\,s_{r+1}(a+k)\,c_{r,j}\int_{-\infty}^{y} x^{j+n}\,\mathrm{e}^{-x^2/2}dx,\end{aligned}$$

where $c_{r,j}=(ja_0)^{-1}\sum_{m=1}^{j}[m(r+1)-j]\,a_m\,c_{r,j-m}$, for $j\geq 1$, $c_{r,0}=a_0^r$ and $c_{0,0}=1$ and the quantities a_i's are defined in Section 4.5.

Computing the last integral, we have

$$T_n(y)=\frac{1}{\sqrt{2\pi}}\sum_{k,r,j=0}^{\infty} b_k\,c_{r,j}\,s_{r+1}(a+k)\,A(j+n,y),$$

where $A(\cdot,\cdot)$ is determined as before and $s_r(a)$ is given by (4.8).

The third representation for $T_n(y)$ is based on the normal qf. We have

$$T_n(y)=\sum_{r=0}^{\infty} d_r\int_{-\infty}^{y} x^n\,\phi(x)\,\Phi(x)^r dx$$

The last integral can be rewritten according to the normal qf $Q_N(u)$ given in Section 4.4. Thus, using equations (2.7) and (4.11), we have

$$T_n(y)=\sum_{r=0}^{\infty} d_r\int_0^{\Phi(y)}\left(\sum_{s=0}^{\infty} w_s\,u^s\right)^n u^r du=\sum_{r=0}^{\infty} d_r\int_0^{\Phi(y)}\sum_{s=0}^{\infty} e_{n,s}\,u^{r+s}du,$$

where $e_{n,s}=(s\,w_0)^{-1}\sum_{m=1}^{s}[m(n+1)-s]\,w_m\,e_{n,s-m}$ (for $s\geq 1$),$e_{n,0}=w_0^n$ and the quantities w_m's are given in Section 4.4. Finally, we obtain

$$T_n(y)=\sum_{r,s=0}^{\infty} d_r\,e_{n,s}\,\frac{\Phi(y)^{r+s+1}}{(r+s+1)}.$$

Appendix C: Generating function

Here, we present the algebraic details of the second representation for $M(t)$ based on the quantile power series of X. Using (4.12) with $\mu=0$ and $\sigma=1$, we obtain

$$M(t)=\int_0^1 \exp\left[t\,Q_{GN}(u)\right]\,du=\int_0^1 \exp\left[t\left(\sum_{r=0}^{\infty}\overline{p}_j\,h_{j,r}\,u^{r/a}\right)\right]\,du,$$

where $\overline{p}_j=\sum_{s=0}^{\infty}\sum_{j=0}^{s}(-1)^j\binom{s}{j}\,w_s$, $w_s=\sum_{k=s}^{\infty}(\sqrt{2\pi})^k(-2)^{s-k}\,d_k\binom{k}{s}$ and
$h_{j,i}=(i\,p_0)^{-1}\sum_{m=0}^{i}[m(j+1)-i]\,p_m\,h_{j,i-m}$. Other quantities are well-defined in Section 4.4.

Expanding the exponential function, we have

$$M(t) = \int_0^1 \sum_{k=0}^{\infty} \frac{t^k \left(\sum_{r=0}^{\infty} \overline{p}_j \, h_{j,r} \, u^{r/a}\right)^k}{k!} \, du = \sum_{k,r=0}^{\infty} \frac{d_{k,r}}{\left(\frac{r}{a}+1\right)} \frac{t^k}{k!},$$

where $d_{k,r} = (r\, g_0)^{-1} \sum_{m=1}^{r} [m(k+1)-r]\, g_m \, d_{k,r-m}$ (for $r \geq 1$), $d_{k,0} = g_0^r$, $d_{0,0} = 1$, the quantities g_j's are given by $g_j = \overline{p}_j \, h_{j,r}$ and the other quantities $\overline{p}_j$ and $h_{j,r}$ are defined before.

Appendix D: Rényi entropy

The Rényi entropy of a random variable with pdf $f(x)$ is defined as

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \int_{-\infty}^{+\infty} f^{\gamma}(x) dx$$

for $\gamma > 0$ and $\gamma \neq 1$. We provide details about the Rényi entropy for γ positive integer first and then for positive real.

First, assuming $\gamma = n = 2, 3, \ldots$, $\mu = 0$ and $\sigma = 1$, we can write from (4.5) and (4.17)

$$\begin{aligned}
I_R(n) &= \frac{1}{1-n} \log \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} \sum_{r=0}^{\infty} d_r \, \Phi(x)^r \right)^n dx \\
&= \frac{1}{1-n} \log \int_{-\infty}^{+\infty} \left\{ \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-nx^2/2} \left(\sum_{r=0}^{\infty} d_r \sum_{j=0}^{\infty} c_{r,j} \, x^j \right)^n dx \right\} \\
&= \frac{1}{1-n} \log \int_{-\infty}^{+\infty} \left\{ \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-nx^2/2} \left(\sum_{j=0}^{\infty} \overline{e}_j \, x^j \right)^n dx \right\} \\
&= \frac{1}{1-n} \log \left\{ \left(\frac{1}{\sqrt{2\pi}} \right)^n \sum_{j=0}^{n} \wp_{n,j} \int_{-\infty}^{+\infty} x^j \, e^{-\frac{n x^2}{2}} dx \right\}
\end{aligned}$$

Letting $y = \sqrt{n}x$ and using equation (4.25), we have

$$I_R(n) = \frac{1}{n-1} \left\{ \frac{(n-1)}{2} \log(2\pi) + \frac{(j+1)}{2} \log(n) - \log \left[\sum_{j=0}^{n} \wp_{n,j} m_j{}' \right] \right\},$$

where $\wp_{n,j} = (j\,\overline{e}_0)^{-1} \sum_{m=1}^{j} [m(n+1)-j]\, \overline{e}_m \, \wp_{n,j-m}$ (for $j \geq 1$), $\wp_{n,0} = \overline{e}_0^n$, $\overline{e}_j = \sum_{j=0}^{\infty} d_r \, c_{r,j}$ and $m_j{}'$ is the jth moment of the normal distribution. The quantities d_r's are defined in Section 4.3 and the a_v's and $c_{r,j}$'s are given in Section 4.5.

We can write $I_R(\gamma) = (1-\gamma)^{-1} E\{f(Z)^{\gamma-1}\}$. Let $\delta = E(Z)$. For γ real positive, we can write

$$E\{f(Z)^{\gamma-1}\} = \delta^{\gamma-1} E\{1 + \theta\, [f(Z) - \delta]\}^{\gamma-1},$$

where $\theta = \delta^{-1}$. From the generalized binomial expansion, we obtain

$$\{1 + \theta\, [f(Z) - \delta]\}^{\gamma-1} = 1 + \sum_{n=1}^{\infty} \frac{\theta^n \, \Im_n}{n!} \, [f(Z) - \delta]^n,$$

where $\Im_n = \prod_{j=0}^{n-1}(\gamma - 1 - j)$. Further, we have

$$E\{f(Z)^{\gamma-1}\} = \delta^{\gamma-1}\left(1 + \sum_{n=2}^{\infty} \frac{\theta^n \Im_n}{n!} E\{[f(Z) - \delta]^n\}\right).$$

We now calculate $E\{[f(Z)]^n\}$ for $n \geq 2$. From equation (4.5) and using the binomial expansion, we can write

$$\rho_n = E\{[f(Z)]^n\} = \sum_{j=0}^{\infty} \wp_{n,j}\, \psi_{n,j}$$

where $\psi_{n,j} = E\{Z^j\, \phi(Z)^n\}$. Then,

$$\psi_{n,j} = \int_{-\infty}^{\infty} x^j\, \phi(x)^{n+1}\, \mathrm{d}x.$$

Setting $\sqrt{(n+1)}x = y$, we can easily determine the last integral and write ρ_n as

$$\rho_n = \left(\frac{1}{\sqrt{2\pi}}\right)^n \sum_{j=0}^{\infty} \wp_{n,j} \left(\frac{1}{\sqrt{n+1}}\right)^{j+1} {m_j}'.$$

By expanding the binomial term in (4.27), we can obtain an explicit expression for $I_R(\gamma)$, which holds for any γ real positive and $\gamma \neq 1$, given by

$$I_R(\gamma) = (1-\gamma)^{-1}\, \delta^{\gamma-1}\left[1 + \sum_{n=2}^{\infty} \frac{\theta^n \Im_n}{n!} \sum_{k=0}^{n} (-\delta)^{n-k} \binom{n}{k} \rho_k\right],$$

where ρ_j is determined from (4.28).

Appendix E: The observed information matrix

The elements of the observed information matrix $J(\boldsymbol{\theta})$ for the three parameters (a, μ, σ) are given by:

$$J_{aa} = -n\psi'(a), \quad J_{a\mu} = \frac{1}{\sigma}\sum_{i=1}^{n} \frac{\phi(z_i)}{[1-\Phi(z_i)]\log[1-\Phi(z_i)]},$$

$$J_{a\mu} = \frac{1}{\sigma}\sum_{i=1}^{n} \frac{z_i\phi(z_i)}{[1-\Phi(z_i)]\log[1-\Phi(z_i)]},$$

$$J_{\mu\mu} = -\frac{n}{\sigma^2} + \frac{(a-1)}{\sigma^2}\sum_{i=1}^{n} \frac{z_i\phi(z_i)}{[1-\Phi(z_i)]\log[1-\Phi(z_i)]} - \frac{(a-1)}{\sigma}\sum_{i=1}^{n} \frac{\phi^2(z_i)\{1+\log[1-\Phi(z_i)]\}}{[1-\Phi(z_i)]^2\{\log[1-\Phi(z_i)]\}^2},$$

$$\begin{aligned} J_{\mu\sigma} &= -\frac{2}{\sigma^2}\sum_{i=1}^{n} z_i + \frac{(a-1)}{\sigma^2}\sum_{i=1}^{n} \frac{(z_i^2-1)\phi(z_i)}{[1-\Phi(z_i)]\log[1-\Phi(z_i)]} - \\ &- \frac{(a-1)}{\sigma^2}\sum_{i=1}^{n} \frac{z_i\phi^2(z_i)\{1+\log[1-\Phi(z_i)]\}}{[1-\Phi(z_i)]^2\{\log[1-\Phi(z_i)]\}^2}, \end{aligned}$$

$$J_{\sigma\sigma} = \frac{n}{\sigma^2} - \frac{3}{\sigma^2}\sum_{i=1}^{n} z_i^2 + \frac{2(a-1)}{\sigma^2}\sum_{i=1}^{n}\frac{z_i\phi(z_i)}{[1-\Phi(z_i)]\log[1-\Phi(z_i)]} - \\ - \frac{(a-1)}{\sigma^2}\sum_{i=1}^{n}\frac{z_i^2\phi^2(z_i)\{1+\log[1-\Phi(z_i)]\}}{[1-\Phi(z_i)]^2\{\log[1-\Phi(z_i)]\}^2},$$

where $z_i = \left(\frac{x_i-\mu}{\sigma}\right)$ and $\psi'(\cdot)$ is the trigamma function.

Appendix F: Data of application 2

The data are:

0.05, 0.05, 0.05, 1, 1, 2, 2, 3, 2, 3, 3, 3, 3, 3, 4, 5, 4, 6, 6, 7, 4, 4, 6, 6, 6, 6, 6, 6, 7, 5, 5, 6, 7, 8, 8, 5, 4, 4, 4, 4, 5, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 5, 6, 6, 6, 7, 7, 7, 8, 8, 8, 8, 8, 5, 6, 6, 6, 8, 8, 8, 8, 8, 8, 9, 9, 10, 10, 11, 11, 11, 11, 11, 12, 12, 12, 5, 9, 10, 10, 5, 6, 8, 9, 9, 9, 10, 11, 10, 10, 12, 15, 10, 11, 11, 11, 11, 11, 12, 12, 12, 13, 14, 9, 6, 9, 9, 10, 10, 10, 10, 10, 10, 11, 11, 12, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13, 13, 11, 12, 12, 12, 13, 13, 15, 16, 16, 17, 17, 18, 15, 10, 10, 10, 11, 11, 12, 12, 9, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 13, 13, 14, 14, 14, 14, 15, 12, 13, 13, 14, 14, 14, 16, 14, 15, 15, 15, 17, 18, 14, 15, 16, 15, 14, 11, 11, 11, 12, 13, 13, 14, 15, 15, 9, 12, 12, 12, 12, 19, 12, 13, 14, 14, 14, 15, 16, 16, 14, 15, 15, 16, 14, 14, 17, 9, 11, 12, 12, 13, 13, 13, 13, 14, 14, 15, 16, 18, 13, 13, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 16, 16, 16, 17, 17, 14, 14, 14, 15, 16, 17, 9, 13, 13, 14, 14, 15, 16, 13, 13, 14, 14, 14, 14, 14, 14, 14, 14, 15, 17, 17, 12, 15, 22, 12, 17, 17, 15, 14, 15, 15, 16, 16, 17, 17, 17, 15, 16, 20, 20, 13, 15, 15, 15, 12, 18, 16, 16, 16, 14, 16, 15, 15, 16,18,16,16,18,16

References

ALEXANDER, A.; CORDEIRO, G. M.; ORTEGA, E. M. M.; SARABIA, J. M. Generalized beta-generated distributions. *Computational Statistics and Data Analysis* , v. 56, p. 1880–1897, 2012.

AZZALINI, A. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* , v. 12, p. 171–178, 1985.

ABRAMOWITZ, M.; STEGUN, I. A. "Confluent Hypergeometric Functions". Ch. 13 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover,1972. p. 503–515.

BARAKAT, H. M.; ABDELKADER, Y.H. Computing the moments of order statistics from non-identical random variables. *Statistical Methods and Application*, v. 13, p. 15–26, 2004.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, v. 81, p. 827–842, 2011.
-
CORDEIRO, G. M.; CINTRA, R. J.; RÊGO, L. C; ORTEGA, E. M. M. The McDonald normal distribution. *Pakistan Journal of Statistics and Operation Research*, v. 8, 301–329, 2012.

CORDEIRO, G. M.; LEMONTE, A. J. The β-Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling.*Computational Statistics and Data Analysis*, v. 55, p. 1445–1461, 2011.

DOORNIK, J. A. An Object-Oriented Matrix Language - Ox 4, 5th ed. London: Timberlake Consultants Press, 2006.

EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. *Commu-*

nications in Statistics: Theory and Methods, v. 31, p. 497–512, 2002.

EXTON, H. *Handbook of Hypergeometric Integrals: Theory, Applications, Tables, Computer Programs*. New York: Halsted Press, 1978.

GRADSHTEYN, I. S.; RYZHIK, I. M. Table of integrals, series, and products. New York: Academic Press, 2007.

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

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

Matsuo, T. O. (1986). *Uso da regressão de cumeeira em esperimentos agronômicos.* Master Thesis, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, São Paulo, 1986.

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

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; FRIEMER, M. The exponential Weibull family: A reanalysis of the bus-motor failure data. *Technometrics*, v. 37, 436–445, 1995.

NADARAJAH, S. Explicit expressions for moments of order statistics. *Statistics and Probability Letters*, v. 78, p. 196–205, 2008.

NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. The gamma-G family of distributions. Mathematical properties and applications. To appear in Communication in Statistics. Theory - Methods, 2013.

NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. *Mathematics and Computer in Simulation*, v. 74, p. 1–7, 2007.

PRUDNIKOV, A. P.; BRYCHKOV, Y. A.; MARICHEV, O. I. *Integrals and series: Elementary functions*. New York: Gordon & Breach Science Publishers, 1986.

RÉNYI, A. On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*. p. 547–561, 1961.

RISTIĆ, M. M.; BALAKRISHNAN, N. The gamma exponentiated exponential distribution.

Journal of Statistical Computation and Simulation, 2011, doi: 10.1080/00949655.2011.574633.

SHANNON, C. E. Prediction and Entropy of Printed English. *Bell System Technical Journal*, v. 30, p. 50–64, 1951.

STEINBRECHER, G. Taylor expansion for inverse error function around origin. (Working paper at University of Craiova).2002.

WEISBERG, S. *Applied linear regression (3rd ed.)*. New York: Wiley, 2005.

WHITTAKER, E.T.; WATSON, G.N. *A Course in Modern Analysis*. Cambridge: Cambridge University Press, 1990.

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

CHAPTER 5

The gamma Lindley distribution

Resumo

Uma nova distribuição de dois parâmetros chamada de modelo gama Lindley é definida e estudada. Várias de suas propriedades estruturais são derivadas, incluindo expressões explícitas para o momentos, funções quantílica e geradora de momentos, desvios médios e probabilidade ponderada de momentos. Também investigamos as estatísticas de ordem e de seus momentos. Técnicas de máxima verossimilhança são usadas para ajustar o novo modelo e para mostrar sua potencialidade. Com base em três critérios, o modelo proposto prevê um melhor ajuste para dois conjuntos de dados reais do que os modelos Lindley e exponencial geométrico complementar.

Palavras-chave: Desvios médios. Distribuição gama. Distribuição Lindley. Estimação de máxima verossimilhança. Função quantílica.

Abstract

A new two-parameter distribution called the gamma Lindley model is defined and studied. Various of its structural properties are derived, including explicit expressions for the moments, quantile and generating functions, mean deviations and probability weighted moments. We also investigate the order statistics and their moments. Maximum likelihood techniques are used to fit the new model and to show its potentiality. Based on three criteria, the proposed model provides a better fit to two real datasets than the Lindley and complementary exponential geometric distributions.

Keywords: Gamma distribution. Lindley distribution. Maximum likelihood estimation. Mean deviation. Quantile function.

5.1 Introduction

The Lindley distribution is an important uniparametric law with support on the positive real line. The pdf (for $x > 0$) of the Lindley $\mathrm{L}(\lambda)$ distribution with a scale parameter $\lambda > 0$ is given by

$$g(x;\lambda) = \frac{\lambda^2}{1+\lambda}(1+x)\,\mathrm{e}^{-\lambda x}. \tag{5.1}$$

Its cdf becomes

$$G(x;\lambda) = 1 - \mathrm{e}^{-\lambda x}\left(1+\frac{\lambda x}{1+\lambda}\right). \tag{5.2}$$

This model pioneered in Lindley (1958) is denoted by $X \sim \mathrm{L}(\lambda)$. Such distribution has been strongly used in several contexts. For instance, the Lindley distribution was firstly used in order to measure the difference between Fiducial and posterior distributions (related to Bayesian analysis) (Lindley, 1958) Subsequently, Sankaran (1970) used such law as the mixing distribution of a Poisson parameter, resulting in the distribution known as the *Poisson-Lindley* distribution. Recently, Ghitany *et al.* (2008) discussed and studied various properties of the Lindley pdf (5.1).

Although the Lindley distribution is frequently used, it involves only one parameter, what makes its application to the lifetime context a hard task. Various parametric models have been developed recently to cope with the wide variety of patterns in survival data. Some of the proposed parametric models have incorporated one or two shape parameters in a classical distribution to account for additional possible hazard shapes. This chapter aims to build up a new model with one additional parameter.

This chapter is organized as follows. In Section 5.2, we introduce the GL distribution and provide plots of its density and hrf. We derive expansions for the pdf and cdf (Section 5.3), explicit expressions for the qf (Section 5.4), ordinary and incomplete moments and Bonferroni and Lorenz curves (Section 5.5), and generating function (Section 5.6).In Section 5.7, we investigate the order statistics and their moments. The estimation of the model parameters is performed by maximum likelihood in Section 5.8 and a simulation study and two applications to real data are provided in Section 5.9. Concluding remarks are addressed in Section 5.10.

5.2 The gamma Lindley distribution

By taking the pdf (5.1) and cdf (5.2) of the Lindley distribution with scale parameter $\lambda > 0$, the pdf and cdf of the GL distribution are obtained from equations (2.1) and (2.2) (for $x \in \mathbb{R}$) as

$$f(x) \;=\; \frac{\lambda^2}{(1+\lambda)\Gamma(a)}(1+x)\,\mathrm{e}^{-\lambda x}\left[\lambda x - \log\left(1+\frac{\lambda x}{1+\lambda}\right)\right]^{a-1} \tag{5.3}$$

and

$$F(x) = \frac{1}{\Gamma(a)} \int_0^{\lambda x - \log\left(1+\frac{\lambda x}{1+\lambda}\right)} t^{a-1}\, \mathrm{e}^{-t}\, dt = \gamma_1\left(a, \lambda x - \log\left[1 + \frac{\lambda x}{1+\lambda}\right]\right). \tag{5.4}$$

In this chapter, a random variable X having density function (5.3) is denoted by $X \sim$GL(λ, a). Evidently, equation (5.3) does not involve any complicated function and the Lindley distribution arises as the basic exemplar for $a = 1$.

The GL survival function, $S(t) = 1 - F(t)$, can not be expressed in closed-form. However, it may be obtained in terms of the incomplete gamma function ratio.

By means of the *Inverse Transform Method* (for more details see Gentle, 2003), we can define the GL random number generator (RNG) from the well-known gamma cdf and Lindley cdf that can be easily implemented using common statistical packages. We motivate the new model by comparing the performances of the GL, Lindley, Weibull and complementary exponential geometric (CEG) distributions applied to two real data sets.

Figure 5.1 displays possible shapes of the density function (5.3) for selected parameter values. It is evident that the GL distribution is much more flexible than the Lindley distribution.

5.3 Useful expansions

Expansions for (5.3) and (5.4) can be derived using the concept of exponentiated distributions. Consider the *exponentiated Lindley* (exp-L) distribution with power parameter $a > 0$ defined by $Y \sim$ exp-L(λ, a) with cdf and pdf given by

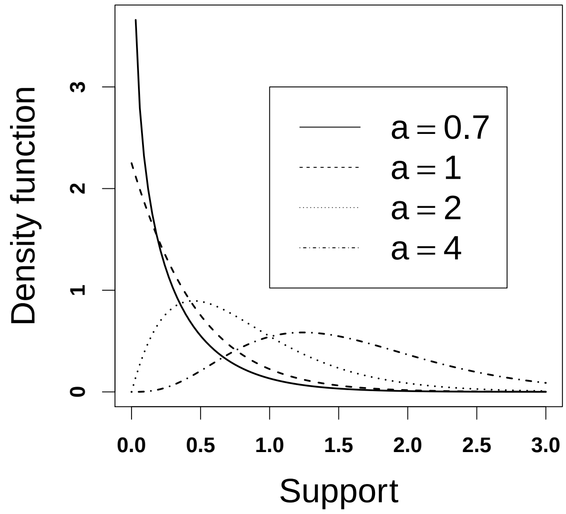
$$H_a(y) \;=\; \left\{1 - \mathrm{e}^{-\lambda x}\left(1 + \frac{\lambda x}{1+\lambda}\right)\right\}^a$$

and

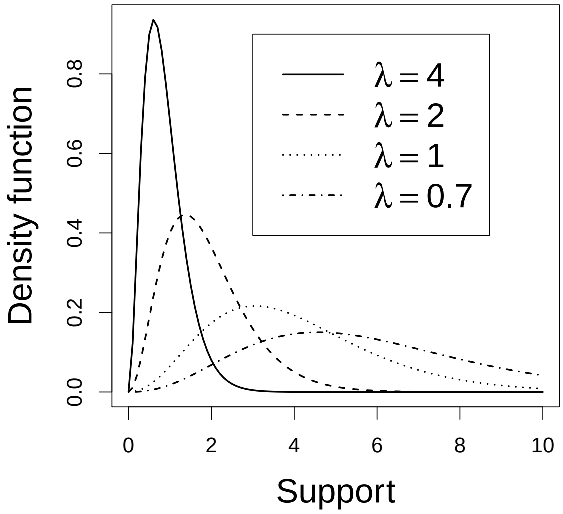
$$h_a(y) \;=\; \frac{a\lambda^2}{1+\lambda}(1+x)\,\mathrm{e}^{-\lambda x}\left\{1 - \mathrm{e}^{-\lambda x}\left(1 + \frac{\lambda x}{1+\lambda}\right)\right\}^{a-1},$$

respectively.

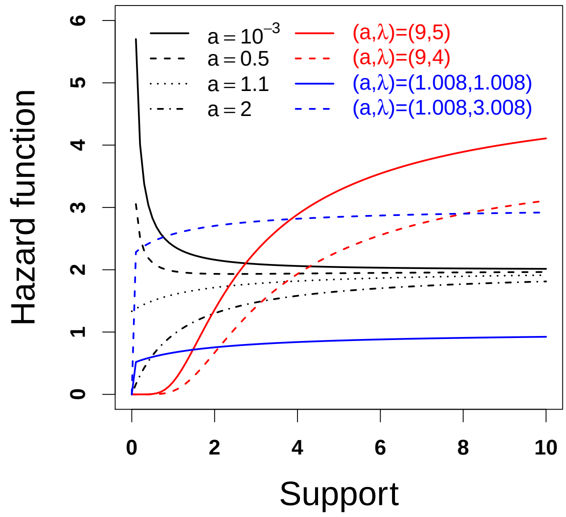
In this section, we present expansions for (5.3) and (5.4) in the form of a theorem and a corollary.



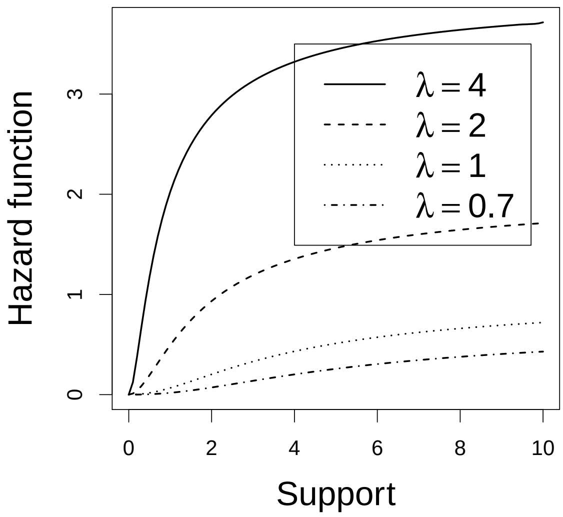
(a) The GL$(3, a)$ density



(b) The GL$(\lambda, 3)$ density



(c) The GL$(2, a)$ hazard for black curves



(d) The GL$(\lambda, 3)$ hazard

Figure 5.1: Plots of the GL density and hazard functions for some parameter values.

Theorem 5.3.1. *Let $X \sim GL(\lambda, a)$. The pdf of X can be expressed by the linear combination*

$$f(x) = \sum_{k=0}^{\infty} b_k \, h_{a+k}(x), \tag{5.5}$$

where $h_{a+k}(x)$ denotes the exp-L$(\lambda, a+k)$ pdf given by

$$h_{a+k}(x) = \frac{(a+k)\lambda^2}{1+\lambda}(1+x)\,\mathrm{e}^{-\lambda x}\left\{1-\mathrm{e}^{-\lambda x}\left(1+\frac{\lambda x}{1+\lambda}\right)\right\}^{a+k-1},$$

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)}$$

and $p_{j,k}$ is defined as follows.

Proof: Based on an expansion due to Nadarajah *et al.* (2013), we can write (for $a > 0$)

$$\{-\log(1-G(x))\}^{a-1} = (a-1)\sum_{k=0}^{\infty}\binom{k+1-a}{k}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)}\,G(x)^{a+k-1},$$

where the constants $p_{j,k}$ can be calculated recursively by

$$p_{j,k} = k^{-1}\sum_{m=1}^{k}\frac{(-1)^m\,[m(j+1)-k]}{(m+1)}\,p_{j,k-m},$$

for $k = 1, 2, \ldots$ and $p_{j,0} = 1$. Let

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)}.$$

Then, equation (5.3) can be expressed as

$$f(x) = \sum_{k=0}^{\infty} b_k\,h_{a+k}(x),$$

where $h_{a+k}(x) = \frac{(a+k)\lambda^2}{1+\lambda}(1+x)\,\mathrm{e}^{-\lambda x}\left\{1-\mathrm{e}^{-\lambda x}\left(1+\frac{\lambda x}{1+\lambda}\right)\right\}^{a+k-1}$ denotes the exp-L$(\lambda, a+k)$ density function.

Corollary 5.3.2. *The cdf of X can be expressed as*

$$F(x) = \sum_{k=0}^{\infty} b_k\,H_{a+k}(x), \tag{5.6}$$

where $H_{a+k}(x) = \left\{1-\mathrm{e}^{-\lambda x}\left(1+\frac{\lambda x}{1+\lambda}\right)\right\}^{a+k}$ denotes the exp-L cdf with parameters λ and $a+k$.

Theorem 5.3.1 and Corollary 5.3.2 are the main results of this section. They reveal that the GL pdf and cdf are linear combinations of the exp-L pdf′s and cdf′s, respectively. So, several mathematical properties of the GL distribution can be obtained by knowing those properties of the exp-L distribution.

5.4 Quantile Function

The GL qf, say $Q_{GL}(u) = F^{-1}(u)$, can be expressed in terms of the Lindley qf ($Q_L(\cdot)$). Inver-
ting equation (5.4), the qf of X (for $0 < u < 1$) reduces to

$$Q_{GL}(u) = F^{-1}(u) = Q_L\left\{1-\exp[-Q^{-1}(a, 1-u)]\right\}, \tag{5.7}$$

where $Q^{-1}(a,u)$ is the inverse function of $Q(a,z) = 1-\gamma(a,z)/\Gamma(a)$. Quantities of interest can be obtained from (5.7) by substituting appropriate values for u. The simulation of the GL distribution is addressed in Section 5.9. Figure 5.2 displays some plots of the GL qf for $\lambda = a = k \in \{1/4, 1/2, 1, 4\}$. From a theoretical point of view, some intractable statistical quantities of X can be derived from a power series expansion for $Q_{GL}(u)$. To that end, we have the following theorem.

Theorem 5.4.1. *The Lindley qf can be expanded as*

$$Q_L(u) = \sum_{n=0}^{\infty} t_n\, u^n, \tag{5.8}$$

where $t_n = \sum_{k=n+1}^{\infty} (-1)^{k-n} \binom{k}{n} \pi_k$.

Proof: We can determine the Lindley qf using the Lagrange theorem.

We assume that the power series expansion holds

$$x \;=\; G(u) \;=\; x_0 + \sum_{k=1}^{\infty} f_k\,(u-u_0)^k,\;\; f_1 = G'(u) \neq 0,$$

where $G(u)$ is analytic at a point u_0 that gives a simple x_0-point.

Then, the inverse function $G^{-1}(x)$ exists and is single-valued in the neighborhood of the point $x = x_0$. The power series inverse $x = Q_L(u)$ is given by

$$x \;=\; Q_L(u) \;=\; u_0 + \sum_{k=1}^{\infty} \pi_k\,(u-u_0)^k,$$

where

$$\pi_k = \frac{1}{k!}\,\frac{\partial^{k-1}}{\partial x^{k-1}} \left\{ [\psi(x)]^k \right\} \Big|_{x=x_0} \qquad \text{and} \qquad \psi(x) \;=\; \frac{x-x_0}{G(x)-u_0}.$$

Then, we can write the GL qf as follows

$$G(x) \;=\; 1 - \left(1 + \frac{\lambda x}{1+\lambda}\right) \mathrm{e}^{-\lambda x} \;=\; \mathrm{u}_0 \;+\; \mathrm{x} \sum_{\mathrm{i}=0}^{\infty} \mathrm{f}_\mathrm{i}\; \mathrm{x}^\mathrm{i},$$

where $u_0 = 1$ and $f_i = (-\lambda)^{i+1} \left[\frac{1}{(i+1)!} - \frac{1}{(1+\lambda)i!}\right]$.

Thus, we have

$$\psi(x) = \frac{x-x_0}{G(x)-u_0} = \frac{1}{\sum_{i=0}^{\infty} f_i\, x^i} = \frac{1}{\lambda\left(\frac{-\lambda}{1+\lambda}\right)} \sum_{i=0}^{\infty} \varrho_i\, x^i = -\left(\frac{1+\lambda}{\lambda^2}\right) \sum_{i=0}^{\infty} \overline{\varrho}_i\, x^i, \tag{5.9}$$

where $\overline{\varrho}_0 = -1$, $\overline{\varrho}_i = -\varrho_i$, $\varrho_0 = 1$ and $\varrho_i = \frac{1}{f_0} \sum_{j=1}^{\infty} f_j\, \varrho_{i-j}$.

So, we obtain from equation (5.9)

$$\frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}} \left\{ [\psi(x)]^k \right\} \Big|_{x=x_0} = \frac{\nu_{k,k-1}\,(1+\lambda)^k\,(k-1)!}{\lambda^{2k}}, \tag{5.10}$$

where $\nu_{k,i} = (k\overline{\varrho}_0)^{-1} \sum_{m=1}^{k} [m(i+1)-k]\, \overline{\varrho}_m\, \nu_{k,i-m}$ and $\nu_{k,0} = \overline{\varrho}_0^{\,i} = 1$.

From equations (5.9) and (5.10), π_k is given by

$$\pi_k = \frac{1}{k!}\,\frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}} \left\{ [\psi(x)]^k \right\} \Big|_{x=x_0} = \frac{\nu_{k,k-1}\,(1+\lambda)^k}{k\lambda^{2k}}.$$

Hence, the Lindley qf reduces to

$$Q_L(u) = \sum_{k=1}^{\infty} \frac{\nu_{k,k-1}\,(1+\lambda)^k}{k\lambda^{2k}}\;(u-1)^k.$$

An alternative expression for $Q_L(x)$ is

$$Q_L(u) = \sum_{n=0}^{\infty} t_n\, u^n,$$

where $t_n = \sum_{k=n+1}^{\infty} (-1)^{k-n} \binom{k}{n} \pi_k$.

Corollary 5.4.2. *The GL qf can be expanded as*

$$Q_{GL}(u) = \sum_{r=0}^{\infty} \tau_r\, u^{r/a}, \tag{5.11}$$

where $\tau_r = h_{j,r}\, \overline{p}_j$, $\overline{p}_j$ *and* $h_{j,i}$ *are defined and discussed in Appendix A.*

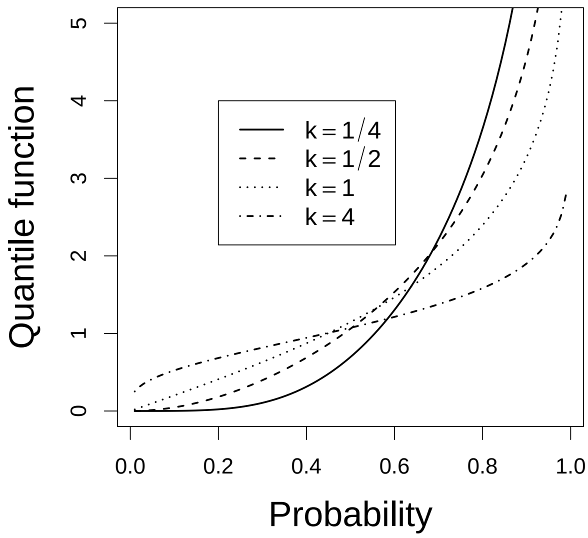


Figure 5.2: Plots of the GL qf for $\lambda = a = k \in \{1/4, 1/2, 1, 4\}$.

5.5 Moments

The ordinary and incomplete moments of X can be derived from the moments of Y_k following the exp-L$(\lambda, a+k)$ distribution. From Theorem 5.3.1, we can write

$$\mu'_n = E(X^n) = \sum_{k=0}^{\infty} b_k\, E(Y_k^n).$$

Using the moments of $Y_k \sim \exp - \mathrm{L(a+k)}$ (Nadarajah *et al.*, 2012), the following corollary is obtained.

Corollary 5.5.1. *Suppose that* $\mu'_n = E(X^n)$ *exists. Then,*

$$\mu'_n = \frac{\lambda^2}{\lambda+1} \sum_{k=0}^{\infty} (a+k)\, b_k\, K(a+k, \lambda, n, \lambda), \tag{5.12}$$

where $K(a,b,c,d) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} \sum_{k=0}^{j+1} \frac{(-1)^i\, b^j}{(1+b)^i\, (ib+d)^{c+k+1}} \binom{a-1}{i} \binom{j+1}{k}\, \Gamma(c+k+1)$.

Further, we can express μ'_n in terms of $Q_L(u)$ as

$$\mu'_n = \sum_{k=0}^{\infty} b_k\,(a+k)\int_0^1 Q_L(u)^n\,u^{a+k-1}\,\mathrm{d}u.$$

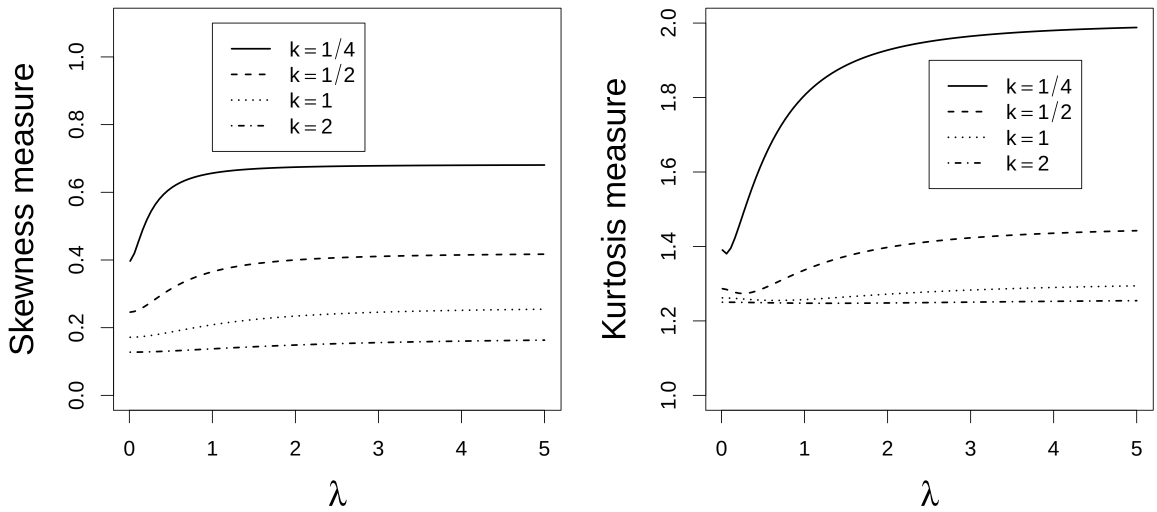
Thus, an alternative expansion for μ'_n can be obtained from Theorem 5.4.1 in the form of the following corollary.

Corollary 5.5.2. *Suppose that $\mu'_n = E(X^n)$ exists. Then,*

$$\mu'_n = \sum_{i,k=0}^{\infty} \frac{b_k\,(a+k)\,e_{n,i}}{(a+i+k+1)}, \tag{5.13}$$

where the quantities $e_{n,i}$ are determined from (2.8) *and* (5.8) *as $e_{n,i} = (it_0^{-1})\sum_{m=1}^{i}[m(n+1)-i]\,t_m\,e_{n,i-m}$, for $i \geq 1$, $e_{n,0} = t_0^n$, and the quantities t_i and d_k are defined in Appendices A and B, respectively.*

The skewness and kurtosis measures can be calculated from the ordinary moments using well-known relationships. Plots of these quantities for some parameter values as functions of a are displayed in Figure 5.3.



(a) $\text{Skewness}_{\text{GL}}$: $(\lambda, a) = (\bullet, k)$

(b) $\text{Kurtosis}_{\text{GL}}$: $(\lambda, a) = (\bullet, k)$

Figure 5.3: Plots of the skewness and kurtosis measures for the GL distribution.

The nth incomplete moment of X can be based on the Lindley qf as

$$m_n(y) = \int_0^y x^n\,f(x)\,dx = \sum_{k=0}^{\infty}(a+k)\,b_k\int_0^{1-\mathrm{e}^{-\lambda x}\left(1+\frac{\lambda x}{1+\lambda}\right)} Q_L(u)^n\,u^{a+k-1}\,du.$$

Let $J_s = \{(l,r) \in \mathbf{N}\times\mathbf{N}; \mathrm{l}+\mathrm{r}=\mathrm{s}\}$. After some algebra, using (2.7) and (5.8) and based on the set J_s, we obtain the subsequent corollary.

Corollary 5.5.3. *Suppose that* $m_n(y)$ *exists. Then,*

$$m_n(y) = \sum_{l,r=0}^{\infty} \overline{\zeta}_{l,r}\, y^{l+r+2} = \sum_{s=0}^{\infty} q_s\, y^{s+2}, \tag{5.14}$$

where $\overline{\zeta}_{l,r} = \sum_{i,j,k=0}^{\infty} \frac{(-1)^j\,(a+k)\,b_k\,m_{j,l}\,\overline{q}_{j,r}}{a+i+k+1} \binom{j}{s}\binom{a+i+k+1}{j}$ *and* $q_s = \sum_{(l,r)\in J_s} \overline{\zeta}_{l,r}$, *for* $s \geq 0$.

More details about (5.14) and other quantities are addressed in Appendix B.

We can derive the mean deviations of X about the mean μ_1' and about the median M in terms of its first incomplete moment. They can be expressed as

$$\delta_1 = 2\big[\mu_1'\,F(\mu_1') \;-\; m_1(\mu_1')\big] \qquad \text{and} \qquad \delta_2 = \mu_1' - 2m_1(M), \tag{5.15}$$

where $\mu_1' = E(X)$ and $m_1(q) = \int_0^q x\, f(x)\,\mathrm{d}x$.

The quantity $m_1(q)$ can be obtained from (5.14) with $n = 1$ and the measures δ_1 and δ_2 in (5.15) are determined by setting $q = \mu_1'$ and $q = M$, respectively.

For a positive random variable X, the Bonferroni and Lorenz curves are defined by $B(\pi) = m_1(q)/(\pi\mu_1')$ and $L(\pi) = m_1(q)/\mu_1'$, respectively, where $q = Q_{GL}(\pi)$ comes from (5.7) for a given probability π. Figure 5.4 displays plots of the GL Bonferroni function for $\lambda = a = k \in \{1/4, 1/2, 1, 4\}$.

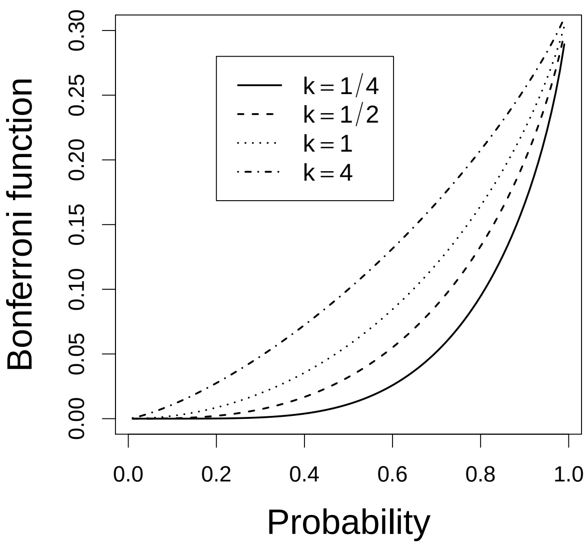


Figure 5.4: Plots of the Bonferroni curve for $\lambda = a = k \in \{1/4, 1/2, 1, 4\}$.

Next, we obtain the probability weighted moments (PWMs) of X. They cover the summarization and description of theoretical probability distributions. The primary use of these moments is to estimate the parameters of a distribution whose inverse cannot be expressed explicitly. The $(s, p)th$ PWM of X is formally defined as

$$\xi_{s,p} = E[X^s\,F(X)^p] = \int_0^{\infty} x^s\,F(x)^p\,f(x)\,\mathrm{d}x.$$

Using (2.7), (5.5) and (5.6), we obtain the subsequent corollary.

Corollary 5.5.4. *The (s,p)th PWM of X is given by*

$$\xi_{s,p} = \sum_{k,r,t}^{\infty} \frac{\alpha_{p,r}\,\omega_{k,r,t}}{(\lambda\,+\,\lambda t)^s}\left[\Gamma(s+k+1)\,+\,\frac{\Gamma(s+k+2)}{\lambda+\lambda t}\right], \tag{5.16}$$

where $\omega_{k,r,t} = \sum_{l=0}^{\infty} \frac{(-1)^t\, d_l\, \lambda^{k+2}}{(1+\lambda)^{k+1}\,(\lambda+t\lambda)^k}\binom{t}{k}\binom{l+r}{t}$, $d_l = \sum_{k=0}^{\infty} b_k\, s_{r+1}(a+k)$, $\alpha_{p,r} = (r v_0)^{-1}$ $\sum_{m=1}^{r}[m(p+1)-r]\, v_m\, c_{p,r-m}$, $v_m = \sum_{k=0}^{\infty} b_k\, s_m(a+k)$ *and* $s_m(a)$ *is given by (5.21) (Section F). Other quantities and some details about* (5.16) *can be found in Appendix B.*

Equations (5.12), (5.14) and (5.16) are the main results of this section. Some algebraic details are given in Appendix B.

5.6 Generating function

A first representation for the mgf $M(t)$ of X can be based on its qf. We have

$$M(t) = \int_0^1 \exp\left[t\, Q_{GL}(u)\right]\, \mathrm{d}u.$$

Expanding the exponential function using (5.11) and after some algebra, we obtain:

Corollary 5.6.1. *The mgf of X can be expressed as*

$$M(t) = \sum_{r,s=0}^{\infty} \frac{d_{s,r}}{\left(\frac{r}{a}+1\right)}\,\frac{t^s}{s!}, \tag{5.17}$$

where $d_{s,j} = (j\, g_0)^{-1}\sum_{m=1}^{j}[m(s+1)-j]\, g_m\, d_{s,j-m}$ *for* $s \geq 1$, $d_{s,0} = g_0^s$, $d_{0,j} = 1$, $g_j = \overline{p}_j\, h_{j,r}$ *and the quantities* $\overline{p}_j$ *and* $h_{j,r}$ *are given in Section 5.4.*

More details about this corollary can be found in Appendix C.

A second representation for $M(t)$ is determined from the exp-L generating function. We can write $M(t) = \sum_{k=0}^{\infty} b_k\, M_k(t)$, where b_k is defined in Section 5.3 and $M_k(t)$ is the mgf of $Y_k \sim \exp-\mathrm{L}(\mathrm{a}+\mathrm{k})$:

Corollary 5.6.2. *The mgf of X is given by*

$$M_k(t) = \frac{(a+k)\lambda^2}{1+\lambda}\, K(a+k,\lambda,0,\lambda-t), \tag{5.18}$$

where $K(a,b,c,d)$ *is defined in Section 5.5. Equations* (5.17) *and* (5.18) *are the main results of this section.*

5.7 Order statistics

Suppose $X_1,\ldots,X_n$ is a random sample from the GL distribution and let $X_{1:n} < \cdots < X_{n:n}$ denote the corresponding order statistics. The following theorem gives an expansion for the pdf of the ith order statistic $X_{i:n}$, say $f_{k:n}(z)$.

Theorem 5.7.1. *The pdf of $X_{i:n}$ can be expanded as*

$$f_{i:n}(z) \quad = \quad \sum_{j=0}^{n-i} \sum_{k,r=0}^{\infty} f_{j,k,r}\, h_{(i+j)a+k+r}(z), \tag{5.19}$$

where

$$f_{j,k,r} = \frac{m_{j,k,r}}{[(i+j)a+k+r]}.$$

Proof: Here, we use the follows relationship: for $z \in (0,1)$ and any real non-integer α, we have

$$z^{\alpha} = \sum_{r=0}^{\infty} s_r(\alpha)\, z^r, \tag{5.20}$$

where

$$s_r(\alpha) = \sum_{l=r}^{\infty} (-1)^{r+l} \binom{\alpha}{l} \binom{l}{r}. \tag{5.21}$$

Combining equations (5.6) and (5.20), we obtain

$$\begin{aligned} F(x) &= \sum_{k,r=0}^{\infty} b_k\, s_r(a+k) \left\{1 - \mathrm{e}^{-\lambda x} \left(1 + \frac{\lambda x}{1+\lambda}\right)\right\}^r \\ &= \sum_{k,r=0}^{\infty} b_k\, s_r(a+k)\, G(x)^r. \end{aligned}$$

By differentiating the previous equation, we can write

$$f(x) = g(x) \sum_{r=0}^{\infty} d_r\, G(x)^r, \tag{5.22}$$

where $d_r = \sum_{k=0}^{\infty} b_k\, s_r(a+k-1)$.

Using (5.6) and (5.22), the pdf of $X_{i:n}$ can be expressed as

$$\begin{aligned} f_{i:n}(z) =& \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(z)\, F(z)^{i+j-1} \\ =& \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} \left[\sum_{r=0}^{\infty} b_r\, (a+r)\, G(z)^{a+r-1}\, g(z)\right] \\ & \times \left[\sum_{k=0}^{\infty} b_k\, G(z)^{a+k}\right]^{i+j-1}. \end{aligned}$$

Based on equations (2.7), (2.8) and after some algebra, we obtain

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

where $\eta_{i+j-1,0} = b_0^{i+j-1}$ and $\eta_{i+j-1,k} = (kb_0)^{-1} \sum_{m=1}^{k} [m(i+j)-k]\, b_m\, \eta_{i+j-1,k-m}$. Hence, the pdf of $X_{i:n}$ reduces to

$$f_{i:n}(z) \quad = \quad g(z) \sum_{j=0}^{n-i} \sum_{k,r=0}^{\infty} m_{j,k,r}\, G(z)^{(i+j)a+k+r-1}, \tag{5.23}$$

where

$$m_{j,k,r} = \frac{(-1)^j\,(a+r)\,n!\,b_r\,\eta_{i+j-1,k}}{(i-1)!\,(n-i-j)!\,j!}.$$

Equation (5.23) can be expressed as

$$f_{i:n}(z) \;=\; \sum_{j=0}^{n-i}\sum_{r,k=0}^{\infty} f_{j,k,r}\,h_{(i+j)a+k+r}(z),$$

where

$$f_{j,k,r} = \frac{m_{j,k,r}}{[(i+j)a+k+r]}\,.$$

Equation (5.19) reveals that the pdf of the GL order statistics is a triple linear combination of exp-L densities with parameters $(i+j)a+k+r$ and λ. Then, several mathematical quantities of the GL order statistics such as ordinary and incomplete moments, mgf and mean deviations can be obtained from those exp-L quantities.

5.8 Estimation

Consider a random variable $X \sim \mathrm{GL}(\lambda, a)$ and let $\boldsymbol{\theta} = (\,\lambda,\, a\,)^\top$ be the model parameters. Thus, the associated log-likelihood function for one observation x reduces to

$$\begin{aligned}\ell(\boldsymbol{\theta};x) \;=\; &[2\,\log(\lambda) \,-\, \log(1\,+\,\lambda)] \,-\, \log[\Gamma(a)] \,+\, \log(1\,+\,x) \,-\, \lambda\,x\\ &+\,(a\,-\,1)\,\log\Big[\lambda\,x \,-\, \log\Big(1\,+\,\frac{\lambda\,x}{1+\lambda}\Big)\Big].\end{aligned} \tag{5.24}$$

The maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ is determined by maximizing
$\ell_n(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n}\,\ell(\boldsymbol{\theta};x_i)$ from a data set $x_1,\ldots,x_n$. The score function comes from (5.24) as

$$\boldsymbol{U}_\theta \;=\; (\,U_\lambda,\,U_a\,)^\top = \Big(\frac{\partial\,\ell(\boldsymbol{\theta};x)}{\partial\lambda},\;\frac{\partial\,\ell(\boldsymbol{\theta};x)}{\partial a}\Big)^\top.$$

After some algebra, two identities hold:

$$U_\lambda \;=\; \frac{\partial\,\ell(\boldsymbol{\theta};x)}{\partial\,\lambda} \tag{5.25}$$

$$\;=\; \frac{2\,+\,\lambda}{(\lambda\,+\,1)\lambda} \,-\, x \,+\, \left[\frac{a\,-\,1}{\lambda\,x\,-\,\log\Big(1+\frac{\lambda\,x}{1+\lambda}\Big)}\right]\left[x \,-\, \frac{x}{(1\,+\,\lambda)(\lambda\,+\,1\,+\,\lambda\,x)}\right] \tag{5.26}$$

and

$$U_a \;=\; \frac{\partial\,\ell(\boldsymbol{\theta};x)}{\partial\,a} \;=\; \log\Big[\lambda\,x \,-\, \log\Big(1\,+\,\frac{\lambda\,x}{1+\lambda}\Big)\Big] \,-\, \Psi(a),$$

where $\Psi(\cdot)$ is the digamma function.

Additionally, in order to make inference on the model parameters, we obtain the *observed information matrix*,

$$H(\boldsymbol{\theta};x) \;=\; \begin{pmatrix} \frac{\partial^2\,\ell(\boldsymbol{\theta};x)}{\partial\,\lambda^2} & \frac{\partial^2\,\ell(\boldsymbol{\theta};x)}{\partial\,\lambda\,\partial\,a}\\ \frac{\partial^2\,\ell(\boldsymbol{\theta};x)}{\partial\,a\,\partial\,\lambda} & \frac{\mathrm{d}^2\,\ell(\boldsymbol{\theta};x)}{\mathrm{d}\,a^2}\end{pmatrix},$$

and the *expected information matrix*

$$\mathcal{K}(\boldsymbol{\theta}) = \mathrm{E}[-H(\boldsymbol{\theta}; X)] = \begin{pmatrix} \kappa_{\lambda\lambda} & \kappa_{\lambda a} \\ \kappa_{a\lambda} & \kappa_{aa} \end{pmatrix}.$$

By differentiating the score function, we have $\kappa_{aa} = -\psi'(a)$,

$$\kappa_{\lambda\lambda} = \frac{\lambda^2 + 4\lambda + 2}{(\lambda+1)^2\,\lambda^2} - (a - 1)\,\mathrm{E}\left\{\frac{\mathrm{d}^2}{\mathrm{d}\,\lambda^2}\log\left[\lambda X - \log\left(1 + \frac{\lambda X}{1+\lambda}\right)\right]\right\}$$

and

$$\kappa_{\lambda a} = \kappa_{a\lambda} = \mathrm{E}\left[\frac{X - \frac{X}{(\lambda+1)(1+\lambda+\lambda X)}}{\lambda X - \log(1 + \frac{\lambda X}{\lambda+1})}\right]. \tag{5.27}$$

From (5.25), we have

$$\mathrm{E}(U_\lambda) = 0 \Leftrightarrow \mathrm{E}\left[\frac{X - \frac{X}{(\lambda+1)(1+\lambda+\lambda X)}}{\lambda X - \log(1 + \frac{\lambda X}{\lambda+1})}\right] = (a - 1)^{-1}\left\{\mathrm{E}(X) - \frac{2 + \lambda}{(\lambda + 1)\lambda}\right\},$$

where $\mathrm{E}(X)$ is given by (5.14) with $n = 1$.

Thus, substituting the last result in (5.27), we obtain

$$\kappa_{\lambda a} = (a - 1)^{-1}\left\{\mathrm{E}(X) - \frac{2 + \lambda}{(\lambda + 1)\lambda}\right\}.$$

5.9 Applications

5.9.1 Simulation study

This section aims to provide a simulation study in order to assess the accuracy of the MLEs described in Section 5.8. One of the advantages of the GL distribution is that its cdf has tractable analytic form. This fact provides a simple random number generator (RNG) given by the Algorithm 1. The use of this algorithm is illustrated in Figure 5.5 for the $\mathrm{GL}(3, 2)$ distribution.

A RNG FOR THE GL DISTRIBUTION [1]
Generate a value u from $U \sim \Gamma(a, 1)$. A possible outcome x from $X \sim \mathrm{GL}(\lambda, a)$ is then given by the solution of the non-linear equation:

$$\lambda x - \log\left(1 + \frac{\lambda x}{1 + \lambda}\right) = u.$$

Now, we perform a simulation study in order to assess the influence of the Lindley parameter, λ, on the GL additional parameter, a. To that end, we consider 5,000 Monte Carlo replications and, on each generated data, we compute the MLEs and (i) the average, (ii) bias and (iii) mean square error (MSE) to be quantified like an assessment criterion.

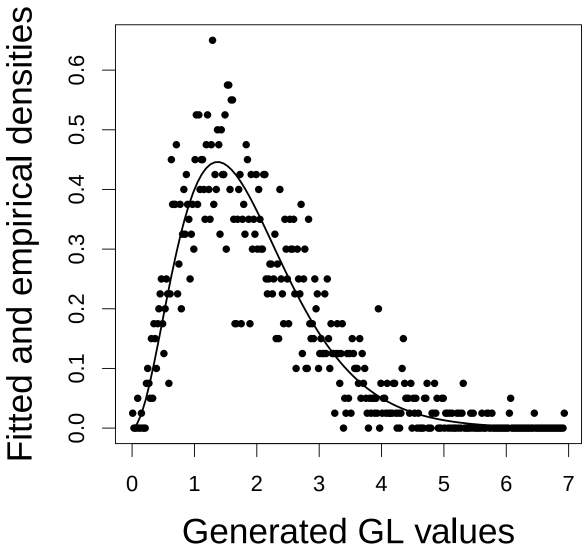


Figure 5.5: Plots of the theoretical and empirical densities for the GL(3,2) distribution.

From Section 5.8, taking $\boldsymbol{U_\theta} = \mathbf{0}$, the MLEs are given by the solutions of the system of non-linear equations:

$$\begin{cases} n\,\frac{2+\lambda}{(\lambda+1)\lambda} - \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \left[\frac{a-1}{\lambda\, x_i - \log\left(1+\frac{\lambda\, x_i}{1+\lambda}\right)}\right]\left[1 - \frac{1}{(1+\lambda)(\lambda+1+\lambda\, x_i)}\right] = 0, \\ \sum_{i=1}^{n} \log\left[\lambda\, x_i - \log\left(1+\frac{\lambda\, x_i}{1+\lambda}\right)\right] - n\,\Psi(a) = 0. \end{cases}$$

The MLEs do not have closed-form expressions and we use the numerical BFGS procedure to obtain them, which is reportedly fast and accurate. The simulation process will be conducted following the steps:

1. Simulated GL distributed data of $N \in \{50, 100, 150\}$ are obtained by means of the GL RNG.

2. Two scenarios are considered: (a) $a = 2$ and $\lambda \in \{1,2,3,4,5\}$ and (b) $a = 5$ and $\lambda \in \{1,2,3,4,5\}$.

3. Generated data is submitted to ML estimation to obtain parameter estimates and sample biases and MSEs.

The initial parameter values chosen in the iterative process are $(\lambda, a) = (1, 1)$. Table 5.1 gives the mean estimates for (λ, a) and their associated MSEs for all scenarios. The MSE values decrease as the sample size increases as expected. Further, smaller values of the MSE at a specified pair are associated to the smaller parameter values. In order to assess separately the influence of increasing both λ and a on the MLE of a, we display plots of $\text{Bias}(\widehat{a})$ as functions of $\lambda \in \{1,2,3,4\}$ for $a \in \{2,5\}$ in Figure 5.6. These plots indicate that the bias decreases when λ increases and it is slower for large values of a.

Table 5.1: Average of MLEs and their corresponding estimates for the MSE

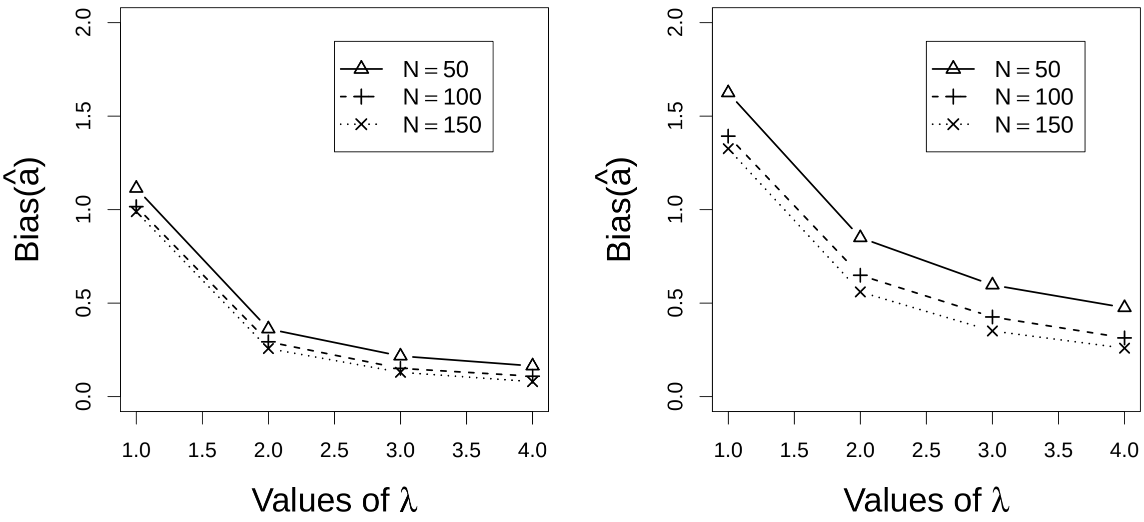
(a, λ)	N	$\hat{a}$	MSE($\hat{a}$)	$\hat{\lambda}$	MSE($\hat{\lambda}$)
(2,1)	50	3.114	1.724	1.213	0.105
	100	3.016	1.252	1.179	0.060
	150	2.989	1.114	1.170	0.050
(2,2)	·	2.362	0.397	2.138	0.203
	·	2.293	0.201	2.076	0.084
	·	2.257	0.142	2.048	0.054
(2,3)	·	2.217	0.264	3.152	0.420
	·	2.151	0.116	3.052	0.173
	·	2.130	0.080	3.027	0.115
(2,4)	·	2.164	0.225	4.185	0.717
	·	2.109	0.104	4.078	0.322
	·	2.080	0.061	4.024	0.195
(5,1)	50	6.625	5.087	1.160	0.085
	100	6.393	3.064	1.125	0.043
	150	6.326	2.469	1.115	0.031
(5,2)	·	5.850	2.566	2.189	0.248
	·	5.649	1.224	2.120	0.107
	·	5.560	0.799	2.090	0.064
(5,3)	·	5.597	1.904	3.225	0.483
	·	5.426	0.867	3.131	0.210
	·	5.351	0.563	3.090	0.132
(5,4)	·	5.476	1.676	4.258	0.831
	·	5.314	0.747	4.139	0.357
	·	5.259	0.481	4.099	0.229

5.9.2 Applications to real data

Here, we perform two applications to real data to show that the proposed model is more adequate than other more common lifetime distributions. We consider a study in the survival context and an application to the SAR image data.

First application: Modeling reliability data

We consider the analysis of the time between failures for repairable item in the form of a data set discussed by Murthy *et al.* (2004) [pp. 278, Data Set 15.1]. Such quantity is a widely used variable in reliability and, assuming the condition of a constant failure rate, it can be defined as the time between two consecutive failures. In terms of an initial analysis, Table 5.2 provides a descriptive discussion of the current data set. The descriptive statistics indicate that the empirical distribution of the data is right skewed (skeness $= 1.295462 > 0$) and leptokurtic



(a) For $a = 2$ (b) For $a = 5$

Figure 5.6: Plots of the bias estimates for a in terms of λ in the GL distribution.

(kurtosis $= 4.31917 > 3$). Further, the sample mean and median are not much different and the relationship between mean and dispersion is very expressive (CV = 73.08%).

Table 5.2: Descriptive statistics from the real data set

Size	kurtosis	skewness	Mean	Median	CV %	Min	Max
30	4.31917	1.295462	1.5430	1.2350	73.08	0.11	4.73

We compare the proposed GL distribution with three other lifetime models: (a) the Lindley, (b) the Weibull and (c) the complementary exponential geometric (Louzada, 2011) distributions. As an initial inferential discussion, Table 5.3 provides the MLEs and their corresponding standard errors. The estimates present low standard errors what makes a comparative study of fitness in terms of the estimates acceptable. Figure 5.7 displays the histogram of the data and the fitted densities.

Table 5.3: MLEs and their standard errors based on the first real data set

Models	α_1	α_2
GL	1.702903 (0.006475723)	1.496357 (0.004269048)
L	0.9762392 (0.0006029884)	•
W	1.709983 (0.0016919445)	1.463319 (0.0013725423)
CEG	1.2440647 (0.003156975)	0.2369999 (0.0006839305)

In order to quantify the performance of the current distributions, we adopt three goodness-of-fit measures for discriminating both the empirical (f_n) and fitted ($\widehat{f}$) densities: (a) Symmetrized Kullback-Leibler divergence (Eguchi and Copas, 2006; Seghouane and Amari, 2007),

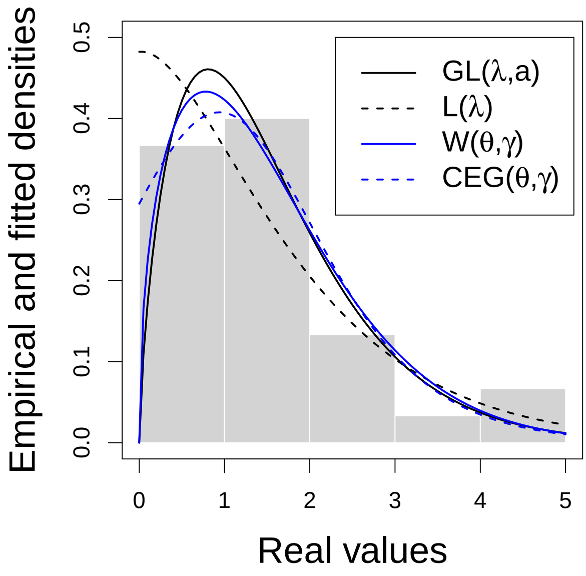


Figure 5.7: Application to real data.

$\mathrm{KL}(f_n, \widehat{f})$, (b) Symmetrized chi-square divergence (Taneja, 2006), $\chi^2(f_n, \widehat{f})$, and (c) Kolmogorov-Smirnov measures, $\mathrm{KS}(f_n, \widehat{f})$. Table 5.4 lists the values of these measures based on the current data set. The results indicate that the GL model outperforms the other distributions according to these criteria.

Table 5.4: Godness-of-fit measures based on a real data set

Measures	GL	Lindley	Weibull	CEG
$\chi^2(f_n, \widehat{f})$	0.04216311	0.04746565	0.04549579	0.04558648
$\mathrm{KL}(f_n, \widehat{f})$	0.01834463	0.02254057	0.02000603	0.01910870
$\mathrm{KS}(f_n, \widehat{f})$	0.16983330	0.19526670	0.17020000	0.17070000
p-value$_{\mathrm{KS}}$	0.74514940	0.62345490	0.74416100	0.74262650

Second application: Modeling intensity data from SAR images

SAR images have been indicated as remote sensing important tools. Among several advantages, the analysis of SAR images can be justified by the ability of producing high spacial resolution images and of operating in all-weather and all-day. However, such images are strongly contaminated by an interference pattern, nominated by *speckle noise* (Oliver and Quegan, 1998). This phenomenon requires a specialized model to be used in the processing of SAR images.

In general (or *polarimetric* terms), these images are obtained obeying the following methodology: orthogonally polarized pulses (in vertical, 'V', or horizontal, 'H', orientations) are sent towards a target, and the returned echo is captured with respect to each polarization: HH, HV, VH (in practice, HV $\approx$ VH) and VV. As a consequence, resulting images are understood as an outcome from a sequence either (i) of complex random vectors (called *single look*) or (ii) of

Hermitian positive definite random matrices (called multilook) (Martinez and Pottier, 2007). This application consists in an advance in case (ii) on which each pixel is represented by 3×3 Hermitian positive definite matrices, whose diagonal elements are positive real intensities:

$$\boldsymbol{Z} = \begin{bmatrix} |Z_{\text{HH}}| & Z_{\text{HH-HV}} & Z_{\text{HH-VV}} \\ Z^*_{\text{HH-HV}} & |Z_{\text{HV}}| & Z_{\text{HV-VV}} \\ Z^*_{\text{HH-VV}} & Z^*_{\text{HV-VV}} & |Z_{\text{VV}}| \end{bmatrix},$$

where $\{|Z_{\text{HH}}|, |Z_{\text{HV}}|, |Z_{\text{VV}}|\}$ represents the set of intensities from Z_{HH}, Z_{HV}, and Z_{VV} polarization channels (complex random vectors) and $\{Z_{\text{HH-HV}}, Z_{\text{HH-VV}}, Z_{\text{HV-VV}}\}$ indicates the set of possible products between two different polarization channels such that $Z_{\text{A-B}} = Z_{\text{A}} Z^*_{\text{B}}$, for A,B $\in \{\text{HH}, \text{HV}, \text{VV}\}$, and * denotes the conjugate of a complex number. In particular, we propose the GL model for describing the terms $|Z_{\text{VV}}|$ in one sample extracted from SAR images.

According to the survey proposed by Gao (2010), *empirical distributions* for describing intensity SAR data consist of models which have no sound deduction in theory, coming from the experience of analyzing real data. Some works (Oliver and Quegan, 1998; Fernandes, 1998; de Fatima and Fernandes, 2000; Gao, 2010) have indicated that the Weibull distribution is one of the best bi-parametric models for describing single-look images for intensity. However, the performance of such distribution is affected by multi-look effect (Ulaby *et al.*, 1986). The proposed distribution can be also used as an alternative model for describing intensities in polarization channels extracted from SAR images. In the subsequent discussion, we present evidence in favor of the GL distribution to be used as pre-processing step of SAR images.

Table 5.5: Descriptive statistics from the second real data set

Window	kurtosis	skewness	Mean	Median	CV %	Min	Max
41×41	5.885843	1.386258	0.0341400	0.0302800	59.80	8.33×10^{-4}	0.1417

Table 5.6: MLEs and their standard errors based on the second real data set

Models	α_1	α_2
GL	2.945013 (0.04340277)	87.226878 (0.6839667)
L	30.22958 (0.6631777)	•
W	0.03852693 (0.0005583914)	1.78024116 (0.032154107)
CEG	77.35430015 (0.9419955)	0.09774584 (0.005164741)

To that end, we use an image of San Francisco obtained by the AIRSAR sensor – an airborne mission with PolSAR capabilities designed by the Jet Propulsion Laboratory. This figure was obtained at `http://earth.eo.esa.int/polsarpro/datasets.html` by means of the `polSARpro` software. Figure 5.8(a) displays a 41×41 pixels image (HH channel) of San Francisco recorded by this sensor, acquired with four nominal looks. Table 5.5 presents a fast descriptive analysis from extracted data. This empirical distribution provides more evidence to be right skewed and to have a leptokurtic form than that furnished by the data in the first application.

Table 5.7: Godness-of-fit measures based on a real data set

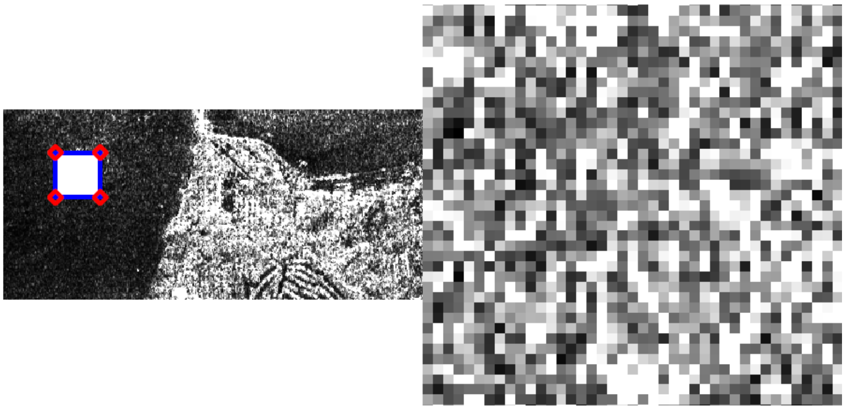
Measures	GL	Lindley	Weibull	CEG
$\chi^2(f_n, \widehat{f})$	190.70514	2891.6875	460.0431	484.9014
$\mathrm{KL}(f_n, \widehat{f})$	72.99628	552.5212	116.2774	148.6457
$\mathrm{KS}(f_n, \widehat{f})$	0.02866805	0.2211755	0.0517323	0.05193516
p-value$_{\mathrm{KS}}$	0.53451750	$< 10^{-16}$	0.04779928	0.03515184

Based on these data and in order to advance in inferential process, the MLEs are listed in Table 5.6. The fitted densities for the intensity SAR returns are plotted in Figure 5.8. Note that larger returns values are well described by the GL distribution.

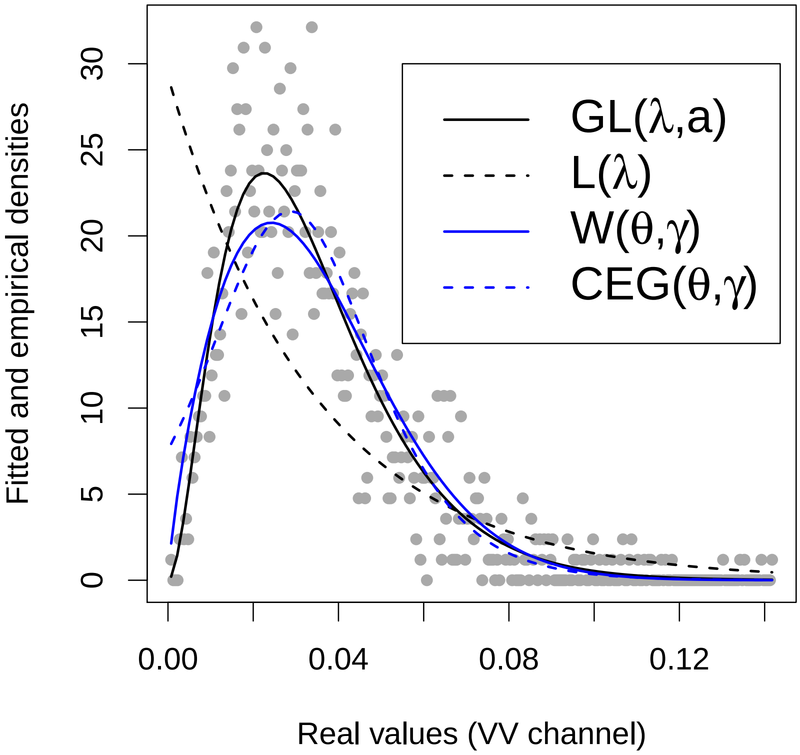
Table 5.7 gives the goodness-of-fit measures for the second application. Under the three criteria – χ^2, KL and KS – the GL model is meaningfully better than the other models. Additionally, assuming to a decision error (nominal level) of 5%, the p-value$_{\mathrm{KS}}$ indicates that only for the GL distribution the null hypothesis that SAR intensity data come from $X \sim \mathrm{GL}(\lambda, \mathrm{a})$ is not rejected.

5.10 Concluding remarks

In this chapter, we apply the gamma-G class of distributions pioneered by Zografos and Balakrishnan (2009) and Ristić and Balakrishnan (2011) to define a new distribution named as gamma Lindley (GL) distribution. The structural properties of this distribution, which include the moments, quantile and generating functions, order statistics, is studied in details. Additionally, we derive a power series for the qf which is useful to obtain some mathematical measures. The proposed distribution has been shown to be very versatile to fit skew real data. Maximum likelihood method is used to estimate the unknown parameters in the proposed model. A simulation study was performed to evaluate the performance of the MLEs. Finally, we consider two applications from survival data and synthetic aperture radar (SAR) images under which the proposed distribution is compared with three other lifetime models. According to three goodness-of-fit criteria, the GL distribution outperforms other distributions in terms of model fitting.



(a) Application to AIRSAR image.



(b) Fits on SAR data.

Figure 5.8: Application to AIRSAR image.

5.11 Appendix

Appendix A:Quantile function

We derive a power series for $Q_{GL}(u)$ following the steps. First, we use a power series for $Q^{-1}(a, 1-u)$. Second, we obtain a power series for the argument $1 - \exp[-Q^{-1}(a, 1-u)]$.

Third, we derive a power series for the Lindley qf using the Lagrange theorem to obtain a power series for $Q_{GL}(u)$.

We introduce the following quantities defined by Cordeiro and Lemonte (2011). Let $Q^{-1}(a,z)$ be the inverse function of

$$Q(a,z) = 1 - \frac{\gamma(a,z)}{\Gamma(a)} = \frac{\Gamma(a,z)}{\Gamma(a)} = u.$$

The inverse quantile function $Q^{-1}(a,1-u)$ is determined in the Wolfram website [1] as

$$\begin{aligned} Q^{-1}(a,1-u) &= w + \frac{w^2}{a+1} + \frac{(3a+5)w^3}{2(a+1)^2(a+2)} + \frac{[a(8a+33)+31]w^4}{3(a+1)^3(a+2)(a+3)} \\ &+ \frac{\{a(a[a(125a+1179)+3971]+5661)+2888\}\, w^5}{24(a+1)^4(a+2)^2(a+3)(a+4)} + O(w^6), \end{aligned}$$

where $w = [u\Gamma(a+1)]^{1/a}$. We can write the last equation as

$$z = Q^{-1}(a,1-u) = \sum_{r=0}^{\infty} \delta_r\, u^{r/a}, \tag{5.28}$$

where $\delta_i = \overline{b}_i\,\Gamma(a+1)^{i/a}$. Here, $\overline{b}_0 = 0$, $\overline{b}_1 = 1$ and any coefficient $\overline{b}_{i+1}$ (for $i \geq 1$) can be obtained from the cubic recurrence equation

$$\begin{aligned} \overline{b}_{i+1} =& \frac{1}{i(a+i)} \Bigg\{ \sum_{r=1}^{i} \sum_{s=1}^{i-s+1} \overline{b}_r \overline{b}_s \overline{b}_{i-r-s+2}\, s\,(i-r-s+2) \\ & \times \sum_{r=2}^{i} \overline{b}_r \overline{b}_{i-r+2}\, r\,[r-a-(1-a)(i+2-r)] \Bigg\}. \end{aligned}$$

The first coefficients are $\overline{b}_2 = 1/(a+1)$, $\overline{b}_3 = (3a+5)/[2(a+1)^2(a+2)]$, ... Now, we present some algebraic details for the GL qf, say $Q_{GL}(u)$. The cdf of X is given by (5.4). By inverting $F(x) = u$, we obtain (5.7).

By replacing (5.28) in equation (5.7), we can write

$$Q_{GL}(u) = Q_L\left(1 - \exp\left[-\sum_{r=0}^{\infty} \delta_r\, u^{r/a}\right]\right).$$

By expanding the exponential function and using (2.7), we have

$$\begin{aligned} 1 - \exp\left(-\sum_{r=0}^{\infty} \delta_r\, u^{r/a}\right) &= 1 - \sum_{l=0}^{\infty} \frac{(-1)^l\, \left(\sum_{r=0}^{\infty} \delta_r\, u^{r/a}\right)^l}{l!} \\ &= 1 - \sum_{l=0}^{\infty} \frac{(-1)^l\, \sum_{r=0}^{\infty} f_{l,r}\, u^{r/a}}{l!} = 1 - \sum_{r=0}^{\infty} p_r\, u^{r/a}, \end{aligned} \tag{5.29}$$

where $p_r = \sum_{l=0}^{\infty} \frac{(-1)^l f_{l,r}}{l!}$, $f_{l,r} = (r\,\delta_0)^{-1} \sum_{q=1}^{r} [q(l+1) - r]\, \delta_q\, f_{l,r-q}$ for $r \geq 1$ and $f_{l,0} = \delta_0^l$. Combining (5.7) and (7.28), we obtain

$$Q_{GL}(u) = Q_L\left(1 - \sum_{r=0}^{\infty} p_r\, u^{r/a}\right).$$

[1] http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/06/01/03/

Using the power series for $Q_L(u)$ in the last equation and expanding the binomial term gives

$$Q_{GL}(u) \;=\; \left\{\sum_{n=0}^{\infty} t_n \left(1-\sum_{r=0}^{\infty} p_r\, u^{r/a}\right)^n\right\} = \left\{\sum_{n=0}^{\infty} t_n \sum_{j=0}^{n}(-1)^j \binom{n}{j} \left(\sum_{r=0}^{\infty} p_r u^{r/a}\right)^j\right\}.$$

Now, using (2.7), we can write

$$Q_{GL}(u) = \left\{\sum_{n=0}^{\infty}\sum_{j=0}^{n}(-1)^j \binom{n}{j}\, t_n \sum_{r=0}^{\infty} h_{j,r}\, u^{r/a}\right\} = \left\{\sum_{n,r=0}^{\infty}\sum_{j=0}^{n}(-1)^j\, t_n\, h_{j,r} \binom{n}{j}\, u^{r/a}\right\},$$

where $h_{j,r} = (r\,p_0)^{-1}\sum_{m=0}^{r}[m(j+1)-r]\,p_m\,h_{j,r-m}$. Finally,

$$Q_{GL}(u) = \sum_{r=0}^{\infty} \tau_r\, u^{r/a},$$

where $\tau_r = \overline{p}_j\, h_{j,r}$, $\overline{p}_j = \sum_{n=0}^{\infty}\sum_{j=0}^{n}(-1)^j\, t_n \binom{n}{j}$.

Appendix B: Moments

Here, we demonstrate the results obtained in Section 5.5. The nth ordinary moment of $X \sim$GL(λ, a) can be expressed in terms of $Q_L(u)$ as follows

$$\mu'_n \;=\; \mathrm{E}(X^n) \;=\; \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^1 Q_L(u)^n\, u^{a+k-1}\,\mathrm{d}u.$$

Inserting (5.8) in the last equation and using (2.7), we have

$$\begin{aligned}
\mu'_n \;&=\; \sum_{k=0}^{\infty} b_k\,(a+k)\int_0^1 \left(\sum_{i=0}^{\infty} t_i\, u^i\right)^n u^{a+k-1}\,\mathrm{d}u \\
&=\; \sum_{i,k=0}^{\infty} b_k\,(a+k)\,e_{n,i}\int_0^1 u^{a+i+k-1}du = \sum_{i,k=0}^{\infty}\frac{b_k\,(a+k)\,e_{n,i}}{a+i+k},
\end{aligned}$$

where $e_{n,i} = (i t_0^{-1})\sum_{m=1}^{i}[m(n+1)-i]\,t_m\,e_{n,i-m}$ and $e_{n,0} = t_0^n$.

Now, we use equation (2.7) and the power series $Q_L(u) = \sum_{i=0}^{\infty} t_i\, x^i$ to obtain the nth incomplete moment given in Section 5.3. We obtain

$$m_n(y) \;=\; E(X^n|X<y) = \sum_{k=0}^{\infty}(a+k)b_k \int_0^{1-\mathrm{e}^{-\lambda y}\left(1+\frac{\lambda y}{1+\lambda}\right)} Q_L(u)^n u^{a+k-1}\,\mathrm{d}u.$$

So, using the Lindley qf and equation (2.7), we have

$$\begin{aligned}
m_n(y) \;&=\; \sum_{i,k=0}^{\infty}(a+k)\,b_k\,v_{n,i}\int_0^{1-\mathrm{e}^{-\lambda y}\left(1+\frac{\lambda y}{1+\lambda}\right)} u^{a+i+k-1}\,\mathrm{d}u \\
&=\; \sum_{i,k=0}^{\infty}\frac{(a+k)\,b_k\,v_{n,i}}{a+i+k}\left\{1-\mathrm{e}^{-\breve{}\mathrm{y}}\left(1+\frac{\breve{}\mathrm{y}}{1+\breve{}}\right)\right\}^{a+i+k},
\end{aligned}$$

where $v_{n,i} = (it_0)^{-1} \sum_{m=1}^{i} [m(n+1) - i]\, t_m\, v_{n,i-m}$ and the coefficients $t_i{}'$ are given in Appendix A.

Expanding the binomial term and the exponential function in the last equation, we obtain

$$
\begin{aligned}
m_n(y) &= \sum_{i,j,k,l=0}^{\infty} \frac{(-1)^j\, b_k\, v_{n,i}\, u_{j,l}\, (a+k)}{a+i+k} \left(1 + \frac{\lambda y}{1+\lambda}\right)^j \binom{a+i+k}{j} y^l \\
&= \sum_{i,j,k,l,r=0}^{\infty} \frac{(-1)^j\, b_k\, v_{n,i}\, u_{j,l}\, (a+k)\, \lambda^r}{(a+i+k)(1+\lambda)^r} \binom{a+i+k}{j} \binom{j}{r} y^{l+r}
\end{aligned}
$$

where $u_{j,l} = (la_0)^{-1} \sum_{m=1}^{l} [m(l+1) - j]\, a_m\, u_{j,l-m}$ and $a_l = -\lambda/l!$.

Let $J_s = \{(l,r) \in \mathbf{N} \times \mathbf{N}; \mathrm{l} + \mathrm{r} = \mathrm{s}\}$. We have

$$
m_n(y) = \sum_{l,r=0}^{\infty} \zeta_{l,r}\, y^{l+r} = \sum_{s=0}^{\infty} q_s\, y^s,
$$

where $\zeta_{l,r} = \sum_{i,j,k=0}^{\infty} \frac{(-1)^j\, b_k\, v_{n,i}\, u_{j,l}\, (a+k)\, \lambda^r}{(a+i+k)(1+\lambda)^r} \binom{a+i+k}{j} \binom{j}{r}$ and $q_s = \sum_{(l,r) \in J_s} \zeta_{l,r}$, for $s \geq 0$.

Further, we obtain the $(s,p)th$ PWM of X using equations (2.7) and (5.22) as follows

$$
\begin{aligned}
\xi_{s,\,p} &= E[X^s\, F(X)^p] = \int_0^\infty x^s\, F(x)^p\, f(x)\, \mathrm{d}x \\
&= \int_0^\infty x^s \left(\sum_{r=0}^{\infty} v_r G(x)^r \right)^p g(x) \sum_{l=0}^{\infty} d_l\, G(x)^l\, \mathrm{d}x \\
&= \sum_{l,r=0}^{\infty} \frac{\alpha_{p,r}\, d_l\, \lambda^2}{1+\lambda} \int_0^\infty (1+x)\, x^s\, \mathrm{e}^{-\breve{}\mathrm{x}} \left\{ 1 - \mathrm{e}^{\breve{}\mathrm{x}} \left(1 + \frac{\breve{}\mathrm{x}}{1+\breve{}} \right) \right\}^{\mathrm{l+r}} \mathrm{dx},
\end{aligned}
$$

where $\alpha_{p,r} = (r\, v_0)^{-1} \sum_{m=1}^{r} [m(p+1) - r]\, v_r\, \alpha_{p,r-m}$, $v_r = \sum_{k=0}^{\infty} b_k s_r(a+k)$ and $s_r(a)$ is given by (5.21)(Section 5.7). Further, $d_l = \sum_{k=0}^{\infty} b_k\, s_r(a+k-1)$.

Thus, expanding the binomial term and the exponential function in the last equation, we have

$$
\xi_{s,\,p} = \sum_{l,k,r,t=0}^{\infty} \frac{(-1)^t\, \alpha_{p,r}\, d_l\, \lambda^{k+2}}{(1+\lambda)^{k+1}} \binom{t}{k} \binom{l+r}{t} \int_0^\infty (1+x)\, x^{s+k}\, \mathrm{e}^{-(\breve{}+\breve{}\mathrm{t})\mathrm{x}}\, \mathrm{dx}
$$

The integral in the last equation comes from the gamma function. Thus, we obtain

$$
\xi_{s,p} = \sum_{k,r,t}^{\infty} \frac{\alpha_{p,r}\, \omega_{k,r,t}}{\lambda^s\, (1+t)^s} \left[\Gamma(s+k+1) + \frac{\Gamma(s+k+2)}{\lambda\, (1+t)} \right],
$$

where $\omega_{k,r,t} = \sum_{l=0}^{\infty} \frac{(-1)^t\, d_l\, \lambda^{k+2}}{(1+\lambda)^{k+1}\, (\lambda + t\lambda)^k} \binom{t}{k} \binom{l+r}{t}$ and d_l is given above.

Appendix C: Generating function

Here, we present the algebraic details of the second representation for $M(t)$ based on the quantile power series of X. Using (5.11) with $\mu = 0$ and $\sigma = 1$, we obtain

$$M(t) \;=\; \int_0^1 \exp\left[t\, Q_{GL}(u)\right]\, du = \int_0^1 \exp\left[t\left(\sum_{r=0}^{\infty} \tau_r\, u^{r/a}\right)\right]\,\mathrm{d}u,$$

where $\tau_r \;=\; \overline{p}_j\, h_{j,r}\; \overline{p}_j \;=\; \sum_{s=0}^{\infty}\sum_{j=0}^{s}(-1)^j\binom{s}{j}\, t_s,\; t_s \;=\; \sum_{k=s+1}^{\infty}(-1)^{k-s}\binom{k}{s}\,\pi_k,\; \pi_k$ is given in Appendix A. and $h_{j,i} = (i\,p_0)^{-1}\sum_{m=0}^{i}[m(j+1)-i]\; p_m\, h_{j,i-m}$. Other quantities are well-defined in Section 5.4.

Expanding the exponential function, we have

$$M(t) = \int_0^1 \sum_{k=0}^{\infty} \frac{t^k\left(\sum_{r=0}^{\infty}\tau_r\, u^{r/a}\right)^k}{k!}\, du = \sum_{k,r=0}^{\infty} \frac{d_{k,r}}{\left(\frac{r}{a}+1\right)}\,\frac{t^k}{k!},$$

where $d_{k,r} = (r\,\tau_0)^{-1}\sum_{m=1}^{r}[m(k+1)-r]\,\tau_m\, d_{k,r-m}$ (for $r \geq 1$), $d_{k,0} = \tau_0^r$, $d_{0,0} = 1$, the quantities τ_j's are given by $\tau_j = \overline{p}_j\, h_{j,r}$ and the other quantities $\overline{p}_j$ and $h_{j,r}$ are defined before.

References

CORDEIRO, G. M.; LEMONTE, A. J. The β-Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics & Data Analysis*, v. 55, p. 1445–1461, 2011.

DE FÁTIMA CARVALHO FERREIRA, A.; FERNANDES, D. Speckle filter for weibull-distributed sar images. *IEEE 2000 international geoscience and remote sensing symposium (igarss 2000)*, v. 2, p. 642–644, 2000.

EGUCHI, S.; COPAS, J. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, v. 97, p. 2034–2040, 2006.

FERNANDES, D. Segmentation of SAR images with Weibull distribution. *IEEE international geoscience and remote sensing symposium proceedings*, v. 1, p. 24–26, 1998.

GAO, G. Statistical modeling of SAR images: A survey. *Sensors*, v. 10, p. 775–795, 2010.

GENTLE, J. Randomnumber generation andmonte carlomethods. *Statistics and Computing. Springer*, 2003.

GHITANY, M.; ATIEH, B.; NADARAJAH, S. Lindley distribution and its applications. *Mathematics and Computers in Simulation*, v. 78, p. 493–506, 2008.

LINDLEY, D. Fiducial distributions and Bayes's theorem. *Journal of the Royal Statistical Society Series B (Methodological)*, v. 20, p. 102–107, 1958.

LOPEZ, C.; POTTIER, E. On the extension of multidimensional speckle noise model from single-look to multilook SAR imagery. *IEEE Transactions on Geoscience and Re-mote Sensing*, v. 45(2), p. 305–320, 2007.

LOUZADA, F.; ROMAN, M.; CANCHO, V. The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart. *Computational Statistics & Data Analysis*, v. 55, p. 2516–2524, 2011.

MURTHY, D.; XIE, M.; JIANG, R. Weibull models. *Wiley Series in Probability and Statistics. Wiley*, 2004.

NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. The gamma-G family of distributions: Mathematical properties and applications. *To appear in Communications in Statistics - Theory and Methods*, 2013.

NADARAJAH, S.; BAKOUCH, H. S.; TAHMASBI, R. A generalized lindley distribution. *Sankhya B*, v. 73, p. 331–359, 2012.

OLIVER, C.; QUEGAN, S. Understanding synthetic aperture radar images. The SciTech radar and defense series. *SciTech Publishing*, 1998.

RACHED, Z.; ALALAJI, F.; LORNE CAMPBELL, L. Renyi's divergence and entropy rates for finite al-phabet Markov sources. *IEEE Transactions on Information Theory*, v. 47, p. 1553-1561, 2001.

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

SANKARAN, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics*, v. 26, p. 145–149, 1970.

SEGHOUANE, A. K.; AMARI, S. I. The AIC criterion and symmetrizing the Kullback-Leibler divergence. *IEEE Transactions on Neural Networks*, v. 18, p. 97–106, 2007.

SHANNON, C. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, p. 379–423, 1948.

TANEJA, I. Bounds on triangular discrimination, harmonic mean and symmetric chi-square divergences. *Journal of Concrete and Applicable Mathematics*, v. 4, p. 91–111, 2006.

ULABY, F.; KOUYATE, F.; BRISCO, B.; WILLIAMS, T. Textural infornation in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, v. 24, p. 235–245, 1986.

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

CHAPTER 6

A new generalized gamma distribution

Resumo

A modelagem e análise de tempo de vida é um aspecto importante do trabalho estatístico em uma ampla variedade de informação científica e campos tecnológicos. Neste capítulo, introduzimos e estudamos a distribuição gama-Nadarajah-Haghighi, que pode ser interpretada como uma distribuição gama generalizada truncada(Stacy, 1962). Esse modelo pode apresentar função taxa de falha nas formas constante, decrescente, crescente, em forma de um tipo especial de banheira e um tipo especial de banheira invertida, dependendo dos valores dos seus parâmetros. Demonstramos que a nova função de densidade pode ser expressa como uma combinação linear de funções densidade de exponecializadas Nadarajah-Haghighi (Lemonte, 2013). Várias de suas propriedades estruturais são derivadas, incluindo algumas expressões explícitas para o momentos, funções quantílica e geradora, assimetria, curtose, desvios, curvas de Bonferroni e Lorenz, probabilidade ponderada de momentos e dois tipos de entropia. Também obtemos as estatísticas de ordem. O método de máxima verosimilhança é usado para estimar os parâmetros do modelo e da matriz de informação observada é derivada. Ilustramos o potencial da nova distribuição por meio de duas aplicações para conjuntos de dados reais.

Palavras-chave: Distribuição gama generalizada. Distribuição Nadarajah-Haghighi. Estimação de máxima verossimilhança. Função taxa de falha.

Abstract

The modeling and analysis of lifetimes is an important aspect of statistical work in a wide variety of scientific and technological fields. We introduce and study the gamma-Nadarajah-

Haghighi distribution which can be interpreted as a truncated generalized gamma distribution (Stacy, 1962). It can have a constant, decreasing, increasing, upside-down bathtub or bathtub-shaped hazard rate function depending on the values of its parameters. We demonstrate that the new density function can be expressed as a mixture of exponentiated Nadarajah-Haghighi density functions (Lemonte, 2013). Various of its structural properties are derived, including some explicit expressions for the moments, quantile and generating functions, skewness, kurtosis, mean deviations, Bonferroni and Lorenz curves, probability weighted moments and two types of entropy. We also obtain the order statistics. The method of maximum likelihood is used for estimating the model parameters and the observed information matrix is derived. We illustrate the potentiality of the new distribution by means of two applications to real data sets.

Keywords: Generalized gamma distribution; Hazard rate function; Nadarajah-Haghighi distribution; Maximum likelihood estimation.

6.1 Introduction

In the last decades, several distributions have been proposed based on different modifications of the beta, gamma and Weibull distributions, among others, to provide bathtub hrfs. We cite the exponentiated Weibull distribution pioneered by Mudholkar and Srivastava (1993). Cordeiro *et al.* (2010) defined the Kumaraswamy Weibull distribution, Kong and Sepanski (2007) studied the beta gamma distribution and Cordeiro *et al.* (2011) proposed the exponentiated generalized gamma distribution.

If $G(x;\alpha,\lambda) = 1 - \exp[-(\lambda x)^{\alpha}]$ is the Weibull cumulative distribution with parameters $\alpha > 0$ (shape parameter) and $\lambda > 0$ (scale parameter), then equation (2.2) yields the *generalized gamma* ("GeGa") cumulative distribution (Stacy, 1962)

$$F(x;a,\alpha,\lambda) = \frac{\gamma(a,(\lambda x)^{\alpha})}{\Gamma(a)},$$

and the corresponding pdf given by

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

The GeGa distribution having the Weibull, gamma, exponential and Rayleigh as special models is widely used for modelling lifetime data. Ortega *et al.* (2003) discussed influence diagnostics in GeGa regression models, Nadarajah and Gupta (2007) applied the GeGa distribution to drought data, Huang and Hwang (2006) presented a simple method for estimating the model parameters using its characterization and moment estimation and Cox *et al.* (2007) developed a parametric survival analysis and taxonomy for its hrf. More recently, Ortega *et al.* (2008) compared three types of residuals based on the deviance component in GeGa regression models under censored observations and Ortega *et al.* (2009) proposed a modification of these regression models to allow the possibility that long-term survivors may be presented in the data.

Nadarajah and Haghighi (2011) proposed a new 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

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

and

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

respectively, where $\alpha > 0$ (shape parameter) and $\lambda > 0$ (scale parameter). If Y follows the Nadarajah-Haghighi (NH) distribution, we write $Y \sim NH(\alpha,\lambda)$. The generalization always has its mode at zero and yet allows for increasing, decreasing and constant hrfs. Lemonte (2013) studied the exponentiated NH (exp-NH) distribution.

Here, we combine the works of Zografos and Balakrishnan (2009) and Nadarajah and Haghighi (2011) to derive some mathematical properties of a new three-parameter distribution called the *gamma Nadarajah-Haghighi* (GNH) distribution, with the hope that it may give a "better fit" compared to the GeGa distribution in certain practical situations. Additionally, we study some of its mathematical properties and discuss maximum likelihood estimation of the model parameters. This distribution is expected to have immediate application for modeling failure times due to fatigue and lifetime data in fields such as engineering, finance, economics and insurance, among others.

The rest of the chapter is organized as follows. In Section 6.2, we present the density function and hrf and provide plots of such functions for selected parameter values. In Section 6.3, we derive useful expansions for its cumulative and density functions. In Section 6.4, 6.5 and 6.6 we provide general properties of the GNH distribution including the moments, quantile and generating functions, skewness, kurtosis, mean deviations, Bonferroni and Lorenz curves and probability weighted moments. The Rényi and Shannon entropies are derived in Section 6.8. The estimation of the model parameters is performed by maximum likelihood in Section 6.9 and two applications to real data are given in Section 6.10. Concluding remarks are addressed in Section 6.11.

6.2 The new distribution

In this section, we introduce a new generalization of the gamma distribution. If $G(x;\alpha,\lambda) = 1 - \exp[1-(1+\lambda x)^{\alpha}]$ is the NH cumulative distribution with parameters α and λ, then equation (2.2) yields the GNH cumulative distribution (for $x > 0$)

$$F(x;a,\alpha,\lambda) = \frac{\gamma(a,(1+\lambda x)^{\alpha}-1)}{\Gamma(a)}, \tag{6.1}$$

where $\lambda > 0$ is a scale parameter and the other positive parameters α and a are shape parameters.

The corresponding density function becomes

$$f(x;a,\alpha,\lambda) = \frac{\alpha\,\lambda}{\Gamma(a)}(1+\lambda x)^{\alpha-1}\,[(1+\lambda x)^{\alpha}-1]^{a-1}\,\exp\{-[(1+\lambda x)^{\alpha}-1]\}. \tag{6.2}$$

In this chapter, a random variable X following (6.2) is denoted by $X \sim \mathrm{GNH}(a, \alpha, \lambda)$. Evidently, the density function (6.2) does not involve any complicated function. It is a positive point of the current generalization. It can be proved that the GNH density function is log-convex if $\alpha < 1$ and $a < 1$, and it is log-concave if $\alpha > 1$ and $a > 1$. Furthermore,

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

The study of the GHN distribution seems important since it extends some useful distributions previously studied in the literature. In fact, the NH distribution is obtained by taking $a = 1$. The gamma distribution corresponds to $\alpha = 1$, whereas the exponential distribution is obtained by taking $\alpha = 1$ and $a = 1$. For $\alpha = 1, a = \eta/2$ and $\lambda = 2$, equation (6.2) reduces to the chi-squared distribution, where η denotes the degrees of freedom. Figure 6.1 displays some plots of the density function (6.2) for some parameter values. It is evident that the new distribution is much more flexible than the NH distribution.

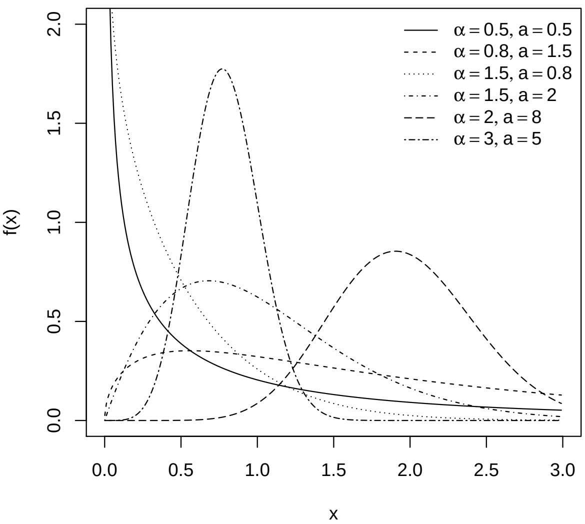


Figure 6.1: Plots for the GNH pdf for some parameter values; $\lambda = 1$.

We note five motivations for the proposed distribution:

• Ability (or the inability) of the GNH distribution to model data that have their mode fixed at zero;

• If Y is a Gamma-G (GG) random variable with shape parameters α and a, and scale parameter λ, then the density in (6.2) is the same as that of the random variable $Z = Y - \lambda^{-1}$ truncated at zero; that is, the GNH distribution can be interpreted as a truncated GG distribution;

• As we shall see later, the GNH hrf can be constant, decreasing, increasing, upside-down bathtub (special case, without the constant part) or bathtub-shaped (special case, without the constant part);

• Some distributions commonly used for parametric models in survival analysis are special cases of the GNH distribution, such as the NH, gamma, chi-squared and exponential distributions;

• It can be applied in some interesting situations as follows: biological and reliability studies, see Cordeiro *et al.* (2011a); failure times of fatiguing materials (see Section 6.10), among others.

The GHN hrf is given by

$$\tau(x;a,\alpha,\lambda) = \frac{\alpha\,\lambda(1+\lambda x)^{\alpha-1}\,[(1+\lambda x)^{\alpha}-1]^{a-1}\,\exp\{-[(1+\lambda x)^{\alpha}-1]\}}{\Gamma(a,(1+\lambda x)^{\alpha}-1)}, \tag{6.3}$$

where $\Gamma(a,z) = \Gamma(a) - \gamma(a,z) = \int_z^{\infty} t^{a-1}\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$. Figure 6.2 displays plots of the GNH hrf for some parameter values.

The new distribution is easily simulated as follows: if V is a gamma random variable with shape parameter $a > 0$, then

$$X = \lambda^{-1}\left\{(1+V)^{1/\alpha} - 1\right\}$$

has the GNH(a,α,λ) distribution.

6.3 Useful expansions

Expansions for equations (6.1) and (6.2) can be derived using the concept of exponentiated distributions. Consider the exp-NH distribution with power parameter $a > 0$ defined by $Y \sim$ exp-NH(a,α,λ) with cdf and pdf given by

$$H_a(x) = \{1 - \exp\{1-(1+\lambda x)^{\alpha}\}\}^a$$

and

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

respectively. Then, equation (6.2) can be expressed as

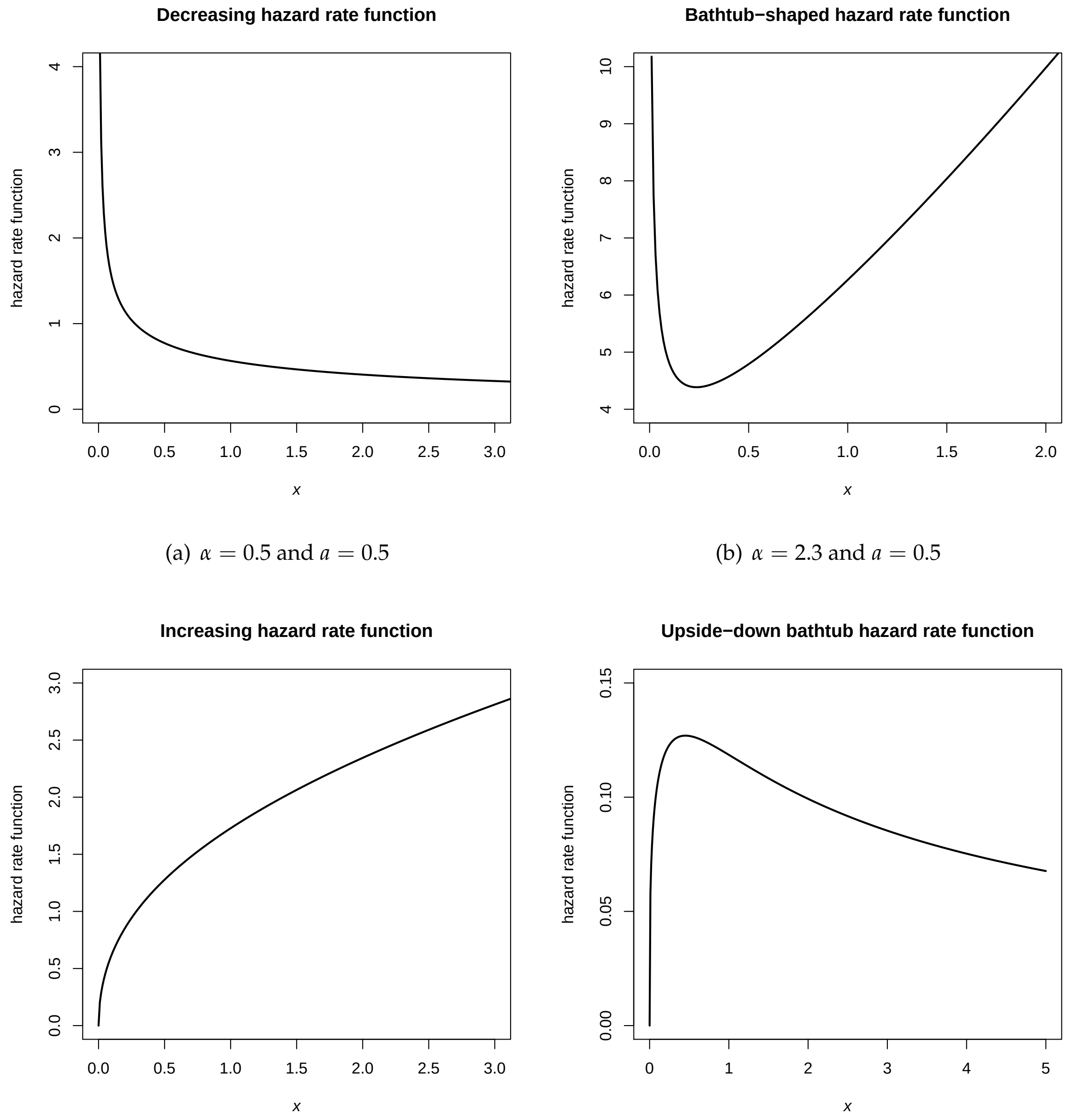
$$f(x) = \sum_{k=0}^{\infty} b_k\,h_{a+k}(x) = g(x)\sum_{k=0}^{\infty}(a+k)\,b_k\,G(x)^{a+k-1}, \tag{6.4}$$

where

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)}$$

and

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



(a) $\alpha = 0.5$ and $a = 0.5$

(b) $\alpha = 2.3$ and $a = 0.5$

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

(d) $\alpha = 0.3$ and $a = 1.3$

Figure 6.2: The GNH hrf for some parameter values; $\lambda = 1$.

denotes the exp-NH$(a+k, \alpha, \lambda)$ density function. Equation (6.4) is the main result of this section. It reveals that the GNH density function is a mixture of exp-NH densities. So, several GNH properties can be obtained by knowing those properties of the exp-NH distribution.

The cdf corresponding to (6.4) becomes

$$F(x) = \sum_{k=0}^{\infty} b_k \, H_{a+k}(x) = \sum_{k=0}^{\infty} b_k \, \{1 - \exp\left[1 - (1+\lambda x)^{\alpha}\right]\}^{a+k}, \tag{6.5}$$

where $H_{a+k}(x) = \{1 - \exp\left[1 - (1+\lambda x)^{\alpha}\right]\}^{a+k}$ denotes the exp-NH cdf with parameters $a+k$,

α and λ.

For $z \in (0,1)$ and any real non-integer α, we have

$$z^{\alpha} = \sum_{r=0}^{\infty} s_r(\alpha)\, z^r, \tag{6.6}$$

where

$$s_r(\alpha) = \sum_{l=r}^{\infty} (-1)^{r+l} \binom{\alpha}{l} \binom{l}{r}.$$

Combining (6.4) and (6.6), we obtain

$$f(x) = g(x) \sum_{r=0}^{\infty} d_r\, G(x)^r, \tag{6.7}$$

where $d_r = (r+1)^{-1} \sum_{k=0}^{\infty} b_k\, (a+k)\, s_r(a+k-1)$.

6.4 Quantile Function

The GNH qf, say $Q(u) = F^{-1}(u)$, can be expressed in terms of the NH qf ($Q_{NH}(\cdot)$). Inverting equation (6.1), it follows the qf of X as

$$F^{-1}(u) = Q_{GNH}(u) = Q_{NH}\left\{1 - \exp[-Q^{-1}(a, 1-u)]\right\}, \tag{6.8}$$

for $0 < u < 1$, where $Q^{-1}(a,u)$ is the inverse function of $Q(a,z) = 1 - \gamma(a,z)/\Gamma(a)$. Quantities of interest can be obtained from (6.8) by substituting appropriate values for u. Further, after some algebra, the NH qf can expressed as (Nadarajah and Haghighi, 2011)

$$Q_{NH}(u) = \lambda^{-1}\left\{[1 - \log(1-u)]^{1/\alpha} - 1\right\}. \tag{6.9}$$

By replacing (6.9) in equation (6.8), we obtain

$$Q_{GNH}(u) = \lambda^{-1}\left\{\left[1 - \log[1 - (1 - \exp[-Q^{-1}(a, 1-u)])]\right]^{1/\alpha} - 1\right\}. \tag{6.10}$$

The inverse function $Q^{-1}(a,u)$ follows from the Wolfram website as

$$z = Q^{-1}(a, 1-u) = \sum_{i=0}^{\infty} a_i\, u^{i/a}, \tag{6.11}$$

where $a_0 = 0$, $a_1 = \Gamma(a+1)^{1/a}$, $a_2 = \Gamma(a+1)^{2/a}/(a+1)$, $a_3 = (3a+5)\Gamma(a+1)^{3/a}/[2(a+1)^2(a+2)]$, etc.

Then, after some algebra using (2.7), (6.9) and (6.11), we obtain

$$Q_{GNH}(u) = \sum_{i=0}^{\infty} \overline{q}_i\, u^{i/a}, \tag{6.12}$$

where $\overline{q}_0 = (q_0 - 1)\lambda^{-1}$, $\overline{q}_i = q_i \lambda^{-1}$ $(i \geq 1)$. The quantity q_i and some other quantities of interest and algebraic details are given in Appendix A. Equations (6.10)-(6.12) are the main results of this section.

6.5 Moments

The ordinary and incomplete moments of X can be immediately obtained from the moments of Y having the exp-NH(a, α, λ) distribution. Thus, we can write from (6.4)

$$\mu'_n = E(X^n) = \sum_{k=0}^{\infty} b_k\, E(Y_k^n).$$

Using the moments of $Y_k \sim \exp - \mathrm{NH(a+k)}$ (Lemonte, 2013), we have

$$\mu'_n = E(X^n) = \lambda^{-n} \sum_{k=0}^{\infty} b_k\, R_n(\alpha, a+k), \tag{6.13}$$

where $R_n(\alpha, a+k) = \int_0^1 \left\{[1-\log(1-u)]^{1/\alpha} - 1\right\}^n u^{a+k-1}\,\mathrm{du}$ is an integral to be computed numerically.

Alternatively, we can write μ'_n in closed-form, based on the quantity $E(Y_k^n)$ (Lemonte, 2013), as

$$\mu'_n = \lambda^{-n}(a+k) \sum_{j,k=0}^{\infty} \sum_{i=0}^{n} \frac{(-1)^{n+j-i}\,(a+k)\,\mathrm{e}^{\mathrm{j}+1}\,\mathrm{b_k}}{(j+1)^{i/\alpha+1}} \binom{a+k-1}{j} \binom{n}{i} \Gamma\left(\frac{i}{\alpha}+1, j+1\right), \tag{6.14}$$

where $\Gamma(a,x) = \int_x^{\infty} z^{a-1}\,\mathrm{e}^{-\mathrm{z}}\,\mathrm{dz}$.

Let $T_n(y)$ denote the nth incomplete moment of X. That is, $T_n(y) = \int_0^y x^n f(x)\,dx$. From equation (6.4), we can write

$$T_n(y) = \sum_{k=0}^{\infty} b_k\, T_n^{\mathrm{NH}}(y, a+k), \tag{6.15}$$

where $T_n^{\mathrm{NH}}(y, a+k)$ denotes the incomplete moment of Y_k. In Lemonte (2013), two expressions for $T_n^{\mathrm{NH}}(y, a+k)$ are given. The first one is

$$T_n^{\mathrm{NH}}(y, a+k) = (a+k)\lambda^{-n} \int_{1-\exp^{1-(1+\lambda y)^{\alpha}}}^{\infty} \left\{\left[1-\log(1-u)\right]^{1/\alpha} - 1\right\}^n u^{a+k-1}\,\mathrm{du},$$

which involves numerical integration. The second one is given in closed-formed as

$$\begin{aligned} T_n^{\mathrm{NH}}(y, a+k) &= \lambda^{-n} \sum_{j,k=0}^{\infty} \sum_{i=0}^{n} \frac{(-1)^{n+j-i}\,(a+k)\,\mathrm{e}^{\mathrm{j}+1}\,\mathrm{b_k}}{(j+1)^{i/\alpha+1}} \binom{a+k-1}{j} \binom{n}{i} \times \\ &\times \Gamma\left(\frac{i}{\alpha}+1, (j+1)(1+\lambda y)^{\alpha}\right). \end{aligned}$$

Using the incomplete first moment, we can derive the mean deviations of X about the mean μ'_1 and about the median M as, respectively,

$$\delta_1 = 2\left[\mu'_1\, F(\mu'_1) - T_1(\mu'_1)\right] \qquad \text{and} \qquad \delta_2 = \mu'_1 - 2T_1(M). \tag{6.16}$$

For a positive random variable X, the Bonferroni and Lorenz curves are defined by $B(\pi) = T_1(q)/(\pi\mu'_1)$ and $L(\pi) = T_1(q)/\mu'_1$, respectively, where $q = F^{-1}(\pi) = Q_{GNH}(\pi)$ comes from the qf (6.10) for a given probability π.

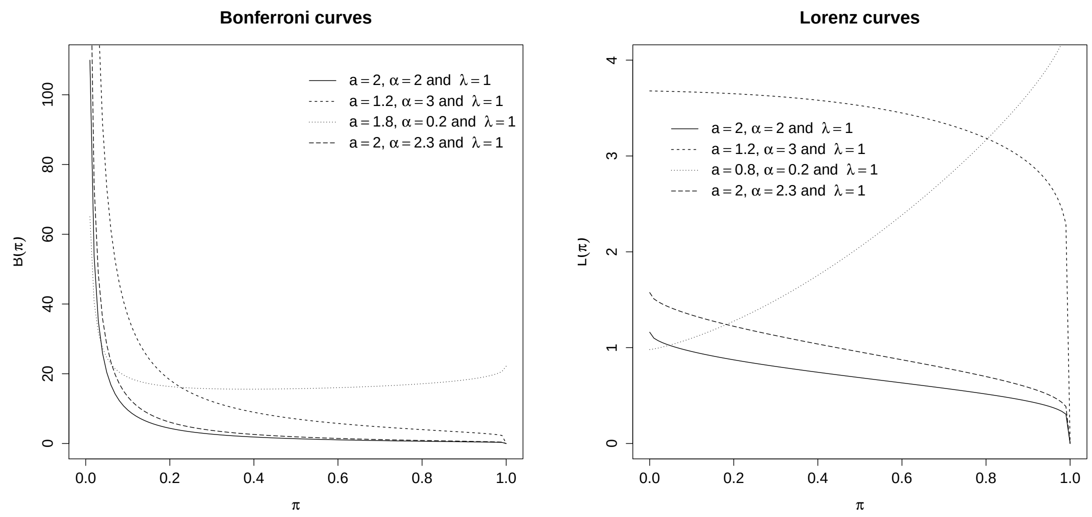


Figure 6.3: Bonferroni and Lorenz curves for some parameter values.

Next, we obtain the probability weighted moments (PWMs) of X. The primary use of these moments is to estimate the parameters of a distribution whose inverse can not be expressed explicitly. The $(s,p)th$ PWM of X is formally defined as

$$\xi_{s,\,p} = E[X^s\,F(X)^p] = \int_0^\infty x^s\,F(x)^p\,f(x)\,\mathrm{d}x.$$

Using equations (2.7), (6.4) and (6.5), we have

$$\xi_{s,p} = \sum_{r,s,m,i=0}^{\infty} \varsigma_{p,r}\,\omega_{r,s,m,i}\,\Gamma\left(\frac{m}{\alpha}+1,1\right), \tag{6.17}$$

where

$$\omega_{r,s,m,i} = \sum_{j,k,l,n=0}^{\infty} \frac{(-1)^{i+j+n+m+s-1}\,\alpha\,d_l}{\lambda^s\,k!\,(\alpha i+1)}\binom{k}{i}\binom{l+r}{j}\binom{\alpha-1}{n}\binom{s+n}{m},$$

d_l is defined in Section 6.3, $v_m = \sum_{k=0}^{\infty} b_k\,s_m(a+k)$, $s_m(a+k)$ is given in Section 6.3, $\varsigma_{p,r} = (rv_0)^{-1}\sum_{m=1}^{r}[m(p+1)-r]\,v_m\,\varsigma_{p,r-m}$ (for $r \geq 1$) and $\varsigma_{p,0} = v_0^p$.

Equations (6.13), (6.14), (6.15) and (6.17) are the main results of this section.

To illustrate the behaviour of the skewness and kurtosis as functions of the parameters, Figure 6.5 displays the Galton's skewness (Johnson et al, 1994, p. 40) and Moors' kurtosis (Moors, 1988) for a selected values of a and α, with $\lambda = 1$. These measures are considered more robust than those usual skewness and kurtosis measures and they exist even for distributions without moments. The Galton's skewness is given by

$$G = \frac{Q\left(\frac{3}{4}\right)+Q\left(\frac{1}{4}\right)-2Q\left(\frac{1}{2}\right)}{Q\left(\frac{3}{4}\right)-Q\left(\frac{1}{4}\right)},$$

whereas the Moors kurtosis is given by

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

Figure 6.5 suggests that both Galton's skewness and Moors' kurtosis increase as α increases, whereas the behaviour for increasing a seems to be different. Also, the parameter a seems to have a much smaller effect in these quantities than α.

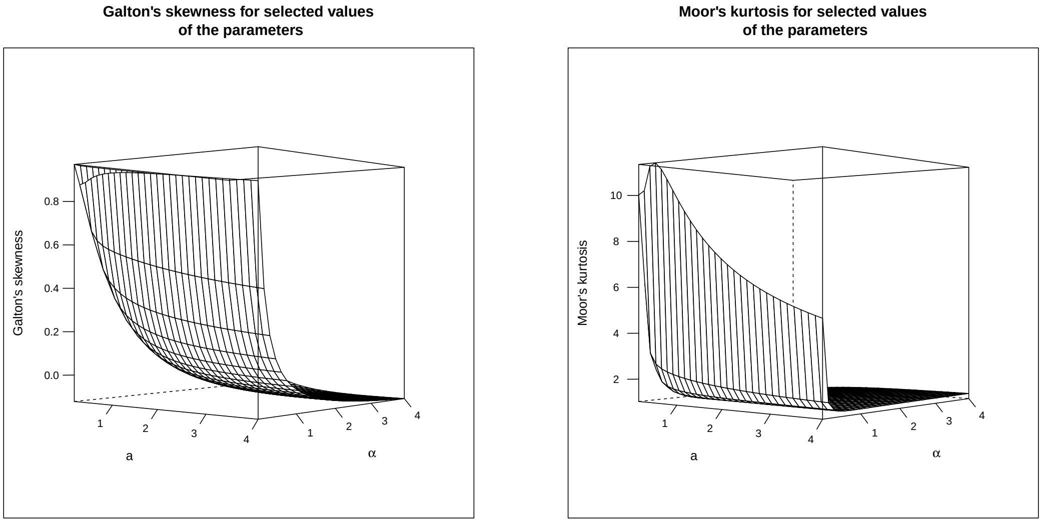


Figure 6.4: Galton's skewness and Moor's kurtosis for the GNH distribution.

6.6 Generating function

A first representation for the mgf $M(t)$ of X can be based on the qf. We have

$$M(t) = \int_0^1 \exp\left[t\, Q_{GNH}(u)\right]\, du.$$

Expanding the exponential function, using (6.12) and after some algebra, we obtain

$$M(t) = \sum_{i,k=0}^{\infty} \frac{d_{k,i}}{\left(\frac{i}{a}+1\right)} \frac{t^k}{k!}, \tag{6.18}$$

where $d_{k,i} = (i\,\overline{q}_0)^{-1}\sum_{m=1}^{i}[m(k+1)-i]\,\overline{q}_m\, d_{k,i-m}$ for $k \geq 1$, $d_{k,0} = \overline{q}_0^k$, $d_{0,i} = 1$, $\overline{q}_i = q_i\lambda^{-1}$ (for $i \geq 1$), $\overline{q}_0 = (q_0 - 1)\lambda^{-1}$ and q_i and other quantities are defined in Appendix A.

A second representation for $M(t)$ is determined from the exp-NH generating function. We ca write $M(t) = \sum_{k=0}^{\infty} b_k\, M_k(t)$, where b_k is given in Section 6.3 and $M_k(t)$ is the mgf of $Y_k \sim \exp - \mathrm{NH(a+k)}$ given by

$$M_k(t) = \sum_{i,r=0}^{\infty} \frac{\eta_i\, g_{i,r}\, t^k}{r/\beta + 1}, \tag{6.19}$$

where $\eta_i = \sum_{k=0}^{\infty} \frac{(-1)^i}{\lambda^k\, k!}\binom{k}{i}$, $g_{i,r} = (r\zeta_0)^{-1}\sum_{n=1}^{r}[n(i+1)-r]\,\zeta_n\, g_{i,r-n}$, $\zeta_r = \sum_{m=0}^{\infty} f_m d_{m,r}$, $d_{m,r} = (ra_0)^{-1}\sum_{v=0}^{r}[v(m+1)-r]\, a_v\, d_{m,r-v}$, $f_m = \sum_{j=m}^{\infty}(-1)^{j-m}\binom{j}{m}\left(\alpha^{-1}\right)_j/j!$, where $\left(\alpha^{-1}\right)_j = \left(\alpha^{-1}\right)\left(\alpha^{-1}-1\right)\ldots\left(\alpha^{-1}-j+1\right)$ is the descending factorial. Other quantities and details about (6.19) are provided in Appendix B.

Equations (6.18) and (6.19) are the main results of this section.

6.7 Order statistics

Suppose $X_1, \ldots, X_n$ is a random sample from the standard GNH distribution and let $X_{1:n} < \cdots < X_{n:n}$ denote the corresponding order statistics. Using (6.4) and (6.5), the pdf of $X_{i:n}$ can be expressed as

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

where the coefficients b_k's are given in Section 6.3. Based on equations (2.7) and (2.8), we obtain

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

where $\eta_{i+j-1,0} = \kappa_0^{i+j-1}$ and $\eta_{i+j-1,k} = (kb_0)^{-1} \sum_{m=1}^{k} \left[m(i+j) - k \right] b_m\, \eta_{i+j-1,k-m}$. Thus, the pdf of $X_{i:n}$ reduces to

$$f_{i:n}(z) = g(z) \sum_{k,r=0}^{\infty} m_{k,r}\, G(z)^{2a+k+r-1}, \tag{6.20}$$

where

$$m_{k,r} = \sum_{j=0}^{n-i} \frac{(-1)^j\, n!\, b_r\, \eta_{i+j-1,k}}{(i-1)!\,(n-i-j)!\,j!}.$$

Equation (6.20) can be expressed as

$$f_{i:n}(z) = \sum_{k,r=0}^{\infty} f_{k,r}\, h_{2a+k+r}(x), \tag{6.21}$$

where

$$f_{k,r} = \frac{m_{k,r}}{2a+k+r}.$$

Equation (6.21) is the main result of this section. It reveals that the pdf of the GNH order statistics is a double linear combination of exp-NH densities with parameters $2a+k+r$, α and λ. So, several mathematical quantities of the GNH order statistics such as ordinary and incomplete moments, mgf and mean deviations can be obtained from those quantities of the exp-NH distribution.

6.8 Entropies

Entropy can be understood as a measure of variation or uncertainty of a random variable X. The Rényi and Shannon entropies are the two more common measures (Shannon, 1948; Rényi, 1961). The Rényi entropy of a random variable with pdf $f(x;\boldsymbol{\theta})$ is defined as

$$H_{\mathrm{R,c}}(\boldsymbol{\theta}) = \frac{1}{1-c} \log \left\{ \mathrm{E}\left[f^{c-1}(X;\boldsymbol{\theta}) \right] \right\} = \frac{1}{1-c} \log \left(\int_0^{\infty} f^{c}(x;\boldsymbol{\theta})\, \mathrm{d}x \right), \tag{6.22}$$

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

The Shannon entropy of a random variable X is defined by $H_S(\boldsymbol{\theta}) = E\{-\log[f(X;\boldsymbol{\theta})]\}$. It is the special case of the Rényi entropy when $c \uparrow 1$.

Direct calculation yields that the Shannon entropy of the random variable X as

$$H_S(\boldsymbol{\theta}) = -\{\log(\alpha) + \log(\lambda) - \log[\Gamma(a)]\} - (\alpha - 1)\,\mathrm{E}\,[\log(1+\lambda\,X)] - 1 + \mathrm{E}\,(1+\lambda\,X)^{\alpha} \\ -(a-1)\,\mathrm{E}\,\log[(1+\lambda X)^{\alpha} - 1].$$

Next, we obtain an expansion for the GNH Rényi entropy. From a result by Lemonte and Cordeiro (2011)

$$x^{\lambda} = \sum_{j=0}^{\infty} f_j\, x^j,$$

where

$$f_j = f_j(\lambda) = \sum_{k=j}^{\infty} (-1)^{k-j} \binom{k}{j} \frac{(\lambda)_k}{k!},$$

and from equation (6.7), we can write

$$f^c(x) = g^c(x) \left(\sum_{r=0}^{\infty} d_r\, G^r(x)\right)^c = g^c(x) \sum_{j=0}^{\infty} f_j \left(\sum_{r=0}^{\infty} d_r\, G^r(x)\right)^j.$$

By applying equation (2.7), we have

$$f^c(x) = g^c(x) \sum_{j,r=0}^{\infty} \underbrace{f_j\, c_{j,r}}_{w_{j,r}}\, G^r(x) = \sum_{j,r=0}^{\infty} w_{j,r}\, g^c(x)\, G^r(x).$$

Finally, using the above result in (6.22), the Rényi entropy can be reduced to

$$H_{\mathrm{R,c}}(\boldsymbol{\theta}) = \frac{1}{1-c} \log\left(\sum_{j,r=0}^{\infty} w_{j,r} \int_0^{\infty} g^c(x)\, G^r(x)\, \mathrm{d}x\right) \\ = \frac{1}{1-c} \log\left\{\sum_{j,r=0}^{\infty} w_{j,r}\, \mathrm{E}_Y\left[g^{c-1}(Y)\, G^r(Y)\right]\right\},$$

where E_Y denotes the expected value based on the exp-NH random variable Y defined at the beginning of Section 6.3.

6.9 Maximum likelihood estimation

Here, we propose a method for obtaining the maximum likelihood estimates (MLEs) of the GNH model parameters. Let $x_1, \ldots, x_n$ be a sample of size n from $X \sim \mathrm{GHN}(a, \alpha, \lambda)$. The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (a, \alpha, \lambda)^{\top}$ can be expressed as

$$\ell(\boldsymbol{\theta}) = n\{\log(\alpha) + \log(\lambda) - \log[\Gamma(a)]\} + (\alpha - 1)\sum_{i=1}^{n} \log(1+\lambda\, x_i) + n - \sum_{i=1}^{n} (1+\lambda\, x_i)^{\alpha} \\ + (a-1)\sum_{i=1}^{n} \log[(1+\lambda x_i)^{\alpha} - 1].$$

The components of the score vector are given by

$$\boldsymbol{U}_\theta = (U_a, U_\alpha, U_\lambda)^\top = \left(\frac{\mathrm{d}\ell(\boldsymbol{\theta})}{\mathrm{d}a}, \frac{\mathrm{d}\ell(\boldsymbol{\theta})}{\mathrm{d}\alpha}, \frac{\mathrm{d}\ell(\boldsymbol{\theta})}{\mathrm{d}\lambda}\right)^\top,$$

$$U_a = -n\,\Psi^{(0)}(a) + \sum_{i=1}^{n} \log[(1+\lambda x_i)^\alpha - 1],$$

$$\begin{aligned} U_\alpha &= \frac{n}{\alpha} + \sum_{i=1}^{n} \log(1+\lambda x_i) - \sum_{i=1} (1+\lambda x_i)^\alpha \log(1+\lambda x_i) + \\ &+ (a-1) \sum_{i=1}^{n} \frac{(1+\lambda x_i)^\alpha \log(1+\lambda x_i)}{(1+\lambda x_i)^\alpha - 1} \end{aligned}$$

and

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

where $\Psi(\cdot)$ is the digamma function.

Setting these equations to zero, $U(\boldsymbol{\theta}) = 0$, and solving them simultaneously yields the MLE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, under some regularity conditions. These equations can not be solved analytically but statistical software to compute them numerically using iterative techniques such as the Newton-Raphson algorithm are available. For interval estimation of the model parameters, we use the 3×3 total observed information matrix $J(\boldsymbol{\theta})$ given by

$$J(\boldsymbol{\theta}) = \begin{pmatrix} J_{aa} & J_{a\alpha} & J_{a\lambda} \\ \bullet & J_{\alpha\alpha} & J_{\alpha\lambda} \\ \bullet & \bullet & J_{\lambda\lambda} \end{pmatrix} = \begin{pmatrix} \frac{\mathrm{d}^2\ell(\boldsymbol{\theta})}{\mathrm{d}a\,\mathrm{d}a} & \frac{\mathrm{d}^2\ell(\boldsymbol{\theta})}{\mathrm{d}a\,\mathrm{d}\alpha} & \frac{\mathrm{d}^2\ell(\boldsymbol{\theta})}{\mathrm{d}a\,\mathrm{d}\lambda} \\ \bullet & \frac{\mathrm{d}^2\ell(\boldsymbol{\theta})}{\mathrm{d}\alpha\,\mathrm{d}\alpha} & \frac{\mathrm{d}^2\ell(\boldsymbol{\theta})}{\mathrm{d}\alpha\,\mathrm{d}\lambda} \\ \bullet & \bullet & \frac{\mathrm{d}^2\ell(\boldsymbol{\theta})}{\mathrm{d}\lambda\,\mathrm{d}\lambda} \end{pmatrix},$$

whose elements are

$$J_{aa} = -n\,\Psi^{(1)}(a), \quad J_{a\alpha} = \sum_{i=1}^{n} \frac{(1+\lambda x_i)^\alpha \log(1+\lambda x_i)}{(1+\lambda x_i)^\alpha - 1}, \quad J_{a\lambda} = \alpha \sum_{i=1}^{n} \frac{x_i (1+\lambda x_i)^{\alpha-1}}{(1+\lambda x_i)^\alpha - 1},$$

$$\begin{aligned} J_{\alpha\alpha} &= -\frac{n}{\alpha^2} - \sum_{i=1} (1+\lambda x_i)^\alpha \log^2(1+\lambda x_i) \\ &+ (a-1)\sum_{i=1}^{n}\left\{\frac{(1+\lambda x_i)^\alpha \log^2(1+\lambda x_i)\,[(1+\lambda x_i)^\alpha - 1] - (1+\lambda x_i)^{2\alpha}\log^2(1+\lambda x_i)}{[(1+\lambda x_i)^\alpha - 1]^2}\right\}, \end{aligned}$$

$$\begin{aligned} J_{\alpha\lambda} &= \sum_{i=1}^{n}\left(\frac{x_i}{1+\lambda x_i}\right) - \alpha\sum_{i=1}^{n} x_i(1+\lambda x_i)^{\alpha-1}\log(1+\lambda x_i) - \sum_{i=1}^{n} x_i(1+\lambda x_i)^{\alpha-1} \\ &+ (a-1)\sum_{i=1}^{n}\left\{\frac{\alpha x_i (1+\lambda x_i)^{\alpha-1}\log(1+\lambda x_i) + x_i(1+\lambda x_i)^{\alpha-1}}{(1+\lambda x_i)^\alpha - 1}\right. \\ &\qquad \left. - \frac{\alpha x_i (1+\lambda x_i)^{2\alpha-1}\log(1+\lambda x_i)}{[(1+\lambda x_i)^\alpha - 1]^2}\right\} \end{aligned}$$

and

$$J_{\lambda\lambda} = -\frac{n}{\lambda^2} - (\alpha - 1)\sum_{i=1}^{n}\left(\frac{x_i}{1 + \lambda\, x_i}\right)^2 - \alpha\,(\alpha - 1)\sum_{i=1}^{n} x_i^2\,(1 + \lambda\, x_i)^{\alpha-2}$$
$$+ \alpha\,(a-1)\sum_{i=1}^{n} x_i \left\{\frac{(\alpha-1)\, x_i\,(1 + \lambda\, x_i)^{\alpha-2}\,[(1 + \lambda\, x_i)^{\alpha} - 1] - \alpha\, x_i\,(1 + \lambda\, x_i)^{2\alpha-2}}{[\,(1 + \lambda\, x_i)^{\alpha} - 1\,]^2}\right\}.$$

Under standard regularity conditions (Cox and Hinkley, 1974) that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the multivariate normal $N_3(0, J(\widehat{\boldsymbol{\theta}})^{-1})$ distribution can be used to construct approximate confidence intervals for the parameters.

6.10 Applications to real data

We conduct two applications of the GNH 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 6.9. All computations were done using the `Ox` matrix programming language (for more details see Doornik, 2006). The first example is a data set from Nichols and Padgett (2006) consisting of 100 observations on breaking stress of carbon fibres (in Gba). For the second example, we consider the data set consisting of the number of successive failures for the air conditioning system of each member in a fleet of 13 Boeing 720 jet airplanes (Proschan, 1963). Obviously, due to the genesis of the GNH distribution, the positive data are by excellence ideally modelled by this distribution. Thus, the use of the GNH distribution for fitting these two data sets is well justified.

Table 6.1 provides some descriptive measures for the two data sets, which include central tendency statistics, standard deviation (SD), skewness (CS) and kurtosis (CK), among others.

Table 6.1: Descriptives statistics.

Statistic	Real data sets	
	stress carbon fibres	number sucessive of failures
Mean	2.621	92.704
Median	2.700	54.000
Mode	2.170	14.000
SD	1.014	107.916
CS	0.374	2.122
CK	0.173	4.938
Minimum	0.390	1.000
Maximum	5.710	603.000

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

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

where $i = 1, \ldots, n$ and $X_{j:n}$ is the j-*th* order statistics of the sample (Mudholkar *et al.*, 1996).

The TTT plots for the fibres data and for the number of successive failure data are presented in Figure 6.5. The TTT plot for the fibres data in Figure 6.5(a) indicates a decreasing hrf, whereas the TTT plot for the number of sucessive failure data in Figure 6.5(b) reveals an upside-down bathtub hrf. Therefore, these plots indicate the appropriateness of the GNH distribution to fit these data, since the new model can present both forms of the hrf.

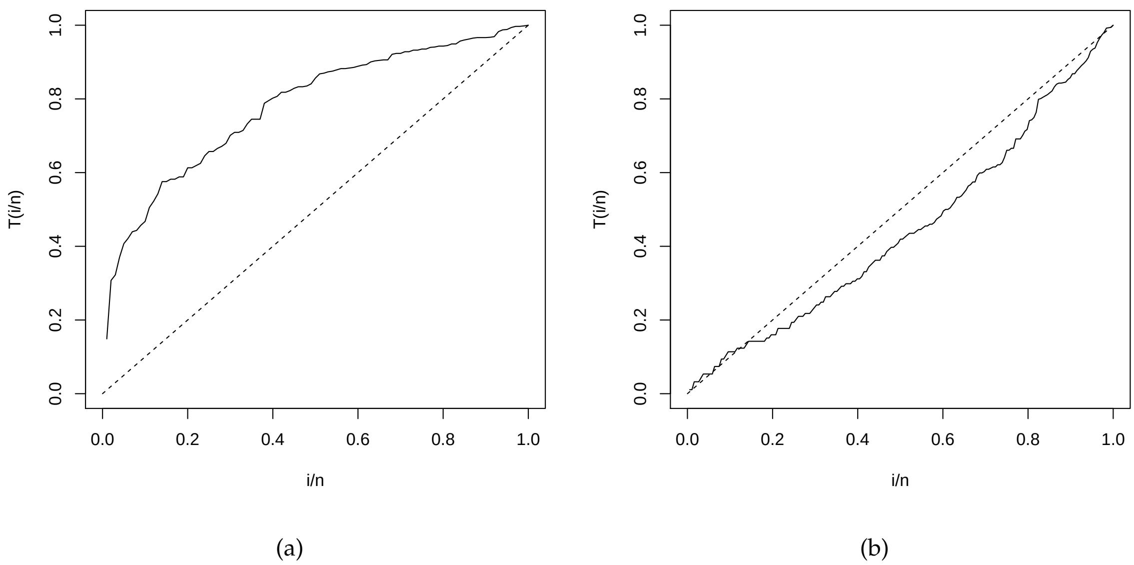


Figure 6.5: TTT plots – (a) stress carbon fibres data; (b) number of sucessive failures air conditioning system data.

The three-parameter GeGa, (Stacy, 1962), exp-NH (Lemonte, 2013) and gamma exponentiated exponential (GEE) (Ristić and Balakrishnan, 2012) distributions are also fitted to these data. Their densities are given by

$$\begin{aligned} f_{GeGa}(x) &= \frac{\alpha\,\lambda^{\alpha a}}{\Gamma(a)} x^{\alpha a-1}\exp[-(\lambda x)^{\alpha}],\; x>0,\\ f_{exp-NH}(x) &= a\,\alpha\,\lambda\,\frac{(1+\lambda x)^{\alpha-1}\exp\{1-(1+\lambda x)^{\alpha}\}}{[1-\exp\{1-(1+\lambda x)^{\alpha}\}]^{1-a}},\; x>0,\\ f_{GEE}(x) &= \frac{\lambda\,\alpha^{a}}{\Gamma(a)} \mathrm{e}^{-\lambda x}(1-\mathrm{e}^{-\lambda x})^{\alpha-1}[-\log(1-\mathrm{e}^{-\lambda x})]^{a-1},\; x>0, \end{aligned}$$

respectively, where $\lambda > 0, \alpha > 0$ and $a > 0$ are parameters.

Table 6.2 lists the MLEs (and the corresponding standard errors in parentheses) of the parameters of the four fitted distributions to both data sets. From the figures of Table 6.2, for the stress carbon fibres data set, we note that $\widehat{\alpha} > 1$ and $\widehat{a} > 1$ for the GNH and exp-NH models, which implies that the hrf for these distributions are decreasing in accordance with Figure 6.5(a). Further, for the number of successive failure data, we have $\widehat{\alpha} < 1$ and $\widehat{a} > 1$ for the GNH and exp-NH models, implying that the hrf is upside-down bathtub in accordance with Figure 6.5(b).

Table 6.2: MLEs (standard erros in parenthesis).

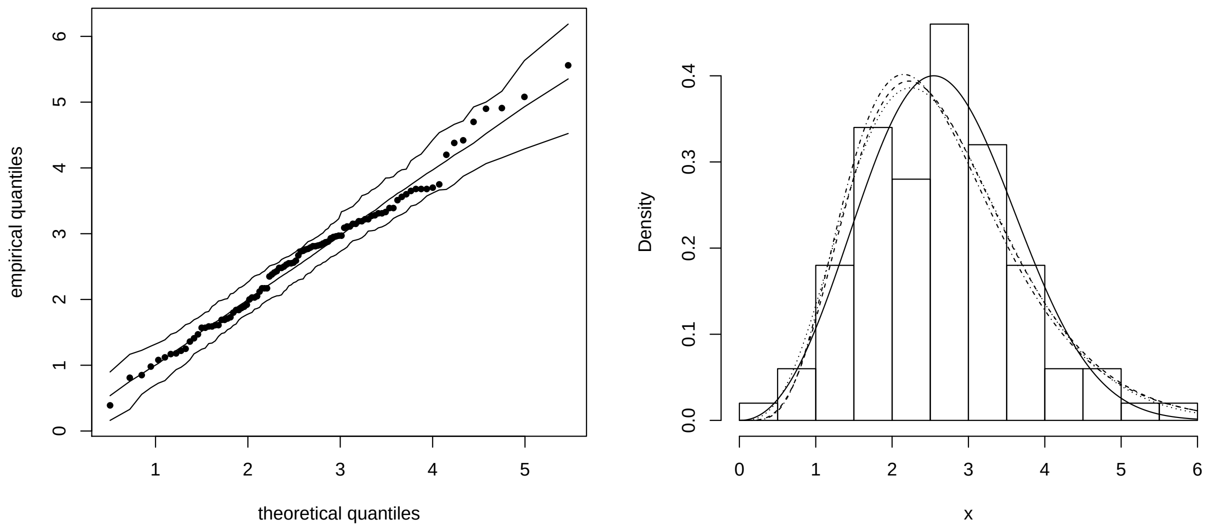
Distribution	Estimates					
	stress carbon fibres			number of sucessive failures		
GNH(α, λ, a)	2.9102	0.2273	3.1231	0.4906	0.0838	1.6009
	(2.2944)	(0.2677)	(0.7864)	(0.0955)	(0.0795)	(0.4855)
GG(α, λ, a)	1.1528	1.7814	2.9543	1.2642	0.3735	5.1353
	(0.1335)	(0.7722)	(0.7690)	(5.7261)	(0.1747)	(4.5431)
exp-NH(α, λ, a)	2.6211	0.1981	4.1170	0.6275	0.0273	1.1007
	(0.0632)	(0.0365)	(0.9971)	(0.0080)	(0.0262)	(0.1291)
GEE(λ, α, a)	0.2697	8.0755	6.1597	0.0001	3.3052	16.6892
	(0.0243)	(1.1726)	(1.0632)	(10^{-6})	(0.3448)	(1.7474)

Now, we shall apply formal goodness-of-fit tests to verify which distribution fits better to these real data sets. We consider the Cramér-von Mises (W^*) and Anderson-Darling (A^*), which are described in details in Chen and Balakrishnan (1995), and Kolmogorov-Smirnov (KS) statistics. Table 6.3 gives the values of the KS, W^* and A^* statistics (and the p-values of the tests in parentheses) for the data sets. Thus, according to these formal tests, the GNH model fits the current data better than the other models, i.e., these values indicate that the null hypothesis is strongly not rejected for the GNH distribution. Thus, according to these goodness-of-fit tests, the GNH model fits the current data better than the other models. These results illustrate the potentiality of the GNH distribution and the importance of the additional shape parameter.

Figure 6.6 and Figure 6.7 display the QQ plot with envelope, which allows us to compare the empirical distribution with the fitted GNH distribution and plots of the estimated pdf's of the fitted GNH, GG, exp-NH and GEE models to these data. They indicate that the GNH distribution is superior to the other distributions in terms of model fitting. These QQ-plots support the result obtained by the KS, W^* and A^* tests. From these plots, we conclude that the proposed distribution provides a better fit to these data than the GG, exp-NH and GEE models.

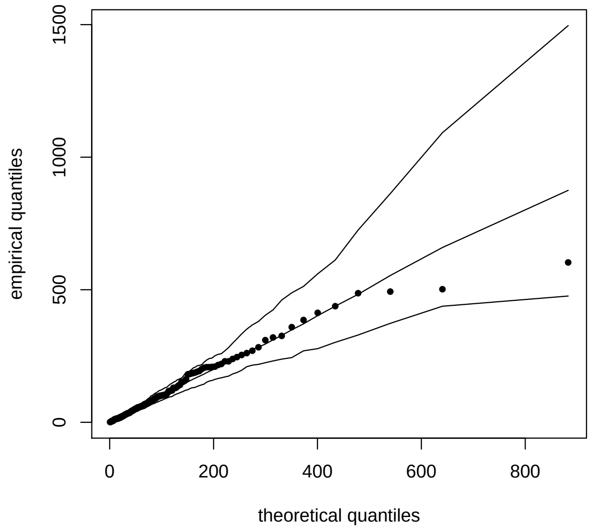
Table 6.3: Goodness-of-fit tests.

Distribution	stress carbon fibres			number of sucessive failures		
	KS	W*	A*	KS	W*	A*
GNH	**0.0645**	**0.0706**	**0.4132**	**0.0442**	**0.0311**	**0.2401**
	(0.7998)	(0.2720)	(0.3317)	(0.8554)	(0.8316)	(0.7735)
GG	0.0793	0.1443	0.7307	0.0443	0.0460	0.3236
	(0.5548)	(0.0286)	(0.0550)	(0.8543)	(0.5714)	(0.5231)
exp-NH	0.0859	0.1100	0.5653	0.0462	0.0671	0.4493
	(0.4514)	(0.0801)	(0.1397)	(0.8175)	(0.3039)	(0.2744)
GEE	0.9997	0.1615	0.8265	0.9959	0.0345	0.2563
	(<0.001)	(0.0165)	(0.0318)	(<0.001)	(0.7783)	(0.7204)

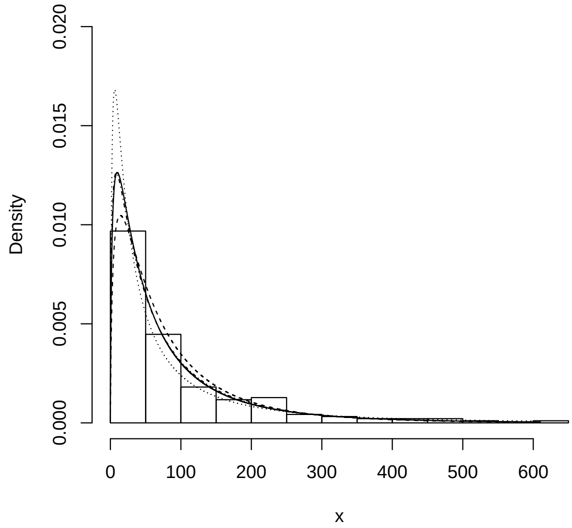


(a) (b)

Figure 6.6: (a) QQ plot with envelope for the GNH distribution and (b) fitted densities of the GNH (solid line), GG (dashed line), exp-NH (dotted line) and GEE (dotdash line) distributions for fibre data.



(a)



(b)

Figure 6.7: (a) QQ plot with envelope for the GNH distribution and (b) fitted densities of the GNH (solid line), GG (dashed line) exp-NH (dotted line) and GEE (dotdash line) distributions for number of successive failures for the air conditioning system.

6.11 Concluding remarks

In this chapter, we propose a new generalized gamma distribution called the gamma-Nadarajah-Haghighi (GNH) distribution. We demonstrate that the hazard rate function of the GNH distribution can be increasing, decreasing, bathtub-shaped and upside-down bathtub shaped. A detailed study on some mathematical properties of the new distribution is presented. The model parameters are estimated by maximum likelihood and the observed information matrix is determined. The potentiality of the new model is demonstrated by means of two real data sets. In fact, the GNH distribution model fits the two data sets well. We hope that the proposed model may attract wider applications in statistics.

6.12 Appendix

Appendix A: Quantile function

We derive a power series for $Q_{GNH}(u)$ in the following way. First, we use a known power series for $Q^{-1}(a, 1-u)$. Second, we obtain a power series for the argument $1-\exp[-Q^{-1}(a, 1-u)]$. Third, we consider the NH qf given in Nadarajah and Haghighi (2011).

We introduce the following quantities defined by Cordeiro and Lemonte (2011). Let $Q^{-1}(a,z)$ be the inverse function of

$$Q(a,z) = 1 - \frac{\gamma(a,z)}{\Gamma(a)} = \frac{\Gamma(a,z)}{\Gamma(a)} = u.$$

The inverse function $Q^{-1}(a,1-u)$ is determined in the Wolfram website [1] as

$$\begin{aligned} Q^{-1}(a,1-u) &= w + \frac{w^2}{a+1} + \frac{(3a+5)w^3}{2(a+1)^2(a+2)} + \frac{[a(8a+33)+31]w^4}{3(a+1)^3(a+2)(a+3)} \\ &+ \frac{\{a(a[a(125a+1179)+3971]+5661)+2888\}\, w^5}{24(a+1)^4(a+2)^2(a+3)(a+4)} + O(w^6), \end{aligned}$$

where $w = [u\Gamma(a+1)]^{1/a}$. We can write the last equation as in (6.10), where the δ_i''s are given by $\delta_i = \overline{b}_i\,\Gamma(a+1)^{i/a}$. Here, $\overline{b}_0 = 0$, $\overline{b}_1 = 1$ and any coefficient $\overline{b}_{i+1}$ (for $i \geq 1$) can be obtained from the cubic recurrence equation

$$\overline{b}_{i+1} = \frac{1}{i(a+i)}\left\{ \sum_{r=1}^{i}\sum_{s=1}^{i-s+1} \overline{b}_r\overline{b}_s\overline{b}_{i-r-s+2}\, s\,(i-r-s+2) \sum_{r=2}^{i} \overline{b}_r\overline{b}_{i-r+2}\, r\,[r-a-(1-a)(i+2-r)] \right\}.$$

The first coefficients are $\overline{b}_2 = 1/(a+1)$, $\overline{b}_3 = (3a+5)/[2(a+1)^2(a+2)], \ldots$. Now, we present some algebraic details for the GNH qf, say $Q_{GNH}(u)$. The cdf of X is given by (6.1). By inverting $F(x) = u$, we obtain (6.10). The NH qf is given by (6.9).

So, using (6.11), we have

$$1 + Q^{-1}(a,1-u) = \sum_{i=0}^{\infty} r_i u^{i/a},$$

where $r_0 = 1$ and $r_i = a_i$ $(i \geq 1)$.

Now, replacing the last result in (6.10), we obtain

$$Q_{GNH}(u) = \lambda^{-1}\left\{ \left(\sum_{i=0}^{\infty} r_i\, u^{i/a}\right)^{1/\alpha} - 1 \right\}$$

By expanding $\left(\sum_{i=0}^{\infty} r_i\, u^{i/a}\right)^{1/\alpha}$ and using (2.7) and (2.8), we have

$$\left(\sum_{i=0}^{\infty} r_i\, u^{i/a}\right)^{1/\alpha} = \sum_{j=0}^{\infty} f_i(\alpha^{-1}) \left(\sum_{i=0}^{\infty} r_i\, u^{i/a}\right)^{j} = \sum_{i,j=0}^{\infty} f_i \epsilon_{j,i} u^{i/a},$$

where $f_j(\alpha^{-1}) = \sum_{k=j}^{\infty}(-1)^{k-j}\binom{k}{j}(\alpha^{-1})_k/k!$, $(\alpha^{-1})_k = \alpha^{-1}(\alpha^{-1}-1)\ldots(\alpha^{-1}-k+1)$ is the descending factorial, $\epsilon_{j,i} = (ir_0)^{-1}\sum_{m=1}^{i}[m(j+1)-i]\, r_m\, \epsilon_{j,i-m}$ (for $i \geq 1$) and $\epsilon_{j,i} = r_0^j$.

Using the last result, we obtain

$$Q_{GNH}(u) = \sum_{i=0}^{\infty} \overline{q}_i\; u^{i/a},$$

where $\overline{q}_0 = (q_0 - 1)\lambda^{-1}$, $\overline{q}_i = q_i\lambda^{-1}$ $(i \geq 1)$ and $q_i = f_i(\alpha^{-1})\sum_{j=0}^{\infty}\epsilon_{j,i}$.

[1] http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/06/01/03/

Appendix B: Generating function

Here, we present the algebraic details of the second representation for $M(t)$ based on the quantile power series of X. From equation (6.12), we can write

$$M(t) = \int_0^1 \exp\left[t\, Q_{GNH}(u)\right]\, du = \int_0^1 \exp\left[t\left(\sum_{i=0}^{\infty} \overline{q}_i\, u^{i/a}\right)\right] du,$$

$\overline{q}_0 = (q_0 - 1)\lambda^{-1}$, $\overline{q}_i = q_i\lambda^{-1}$ $(i \geq 1)$, $q_i = f_i(\alpha^{-1})\sum_{j=0}^{\infty} \epsilon_{j,i}$, $f_j = \sum_{k=j}^{\infty}(-1)^{k-j}\binom{k}{j}(\beta)_k/k!$, $(\alpha^{-1})_k = \alpha^{-1}(\alpha^{-1}-1)\ldots(\alpha^{-1}-k+1)$ is the descending factorial, $\epsilon_{j,i} = (ir_0)^{-1}\sum_{m=1}^{i}[m(j+1)-i]\, r_m\, \epsilon_{j,i-m}$ and $\beta = 1/\alpha$.

Expanding the exponential function, we have

$$M(t) = \int_0^1 \sum_{k=0}^{\infty} \frac{t^k\left(\sum_{i=0}^{\infty} \overline{q}_i\, u^{i/a}\right)^k}{k!}\, du = \sum_{i,k=0}^{\infty} \frac{d_{k,i}}{\left(\frac{i}{a}+1\right)}\, \frac{t^k}{k!},$$

where $d_{k,i} = (i\,\overline{q}_0)^{-1}\sum_{m=1}^{i}[m(k+1)-i]\,\overline{q}_m\, d_{k,i-m}$ (for $i \geq 1$), $d_{k,0} = \overline{q}_0^{\,i}$, $d_{0,0} = 1$.

Next, we obtain the exp-NH generating function using the exp-NH qf as follows:

$$M_{exp-NH}(t) = \int_0^1 \exp\left[tQ_{exp-NH}(u)\right]\, du = \int_0^1 \exp\{t\lambda^{-1}[1-\log(1-u^{1/\beta})]^{1/\alpha} - 1\}\, du.$$

Expanding the exponential function, we obtain

$$M_{exp-NH}(t) = \sum_{k=0}^{\infty} \frac{t^k}{\lambda^k\, k!} \int_0^1 \left\{[1-\log(1-u^{1/\beta})]^{1/\alpha} - 1\right\}^k du, \tag{6.23}$$

where $1 - \log(1-u^{1/\beta})$ can be expressed as

$$1 - \log(1-u^{1/\beta}) = \sum_{r=0}^{\infty} v_r\, u^{r/\beta},$$

where $v_r =$. Using (2.7) and (2.8), we have

$$\begin{aligned}\left(\sum_{r=0}^{\infty} v_r\, u^{r/\beta}\right)^{1/\alpha} &= \sum_{m=0}^{\infty} f_m \left(\sum_{r=0}^{\infty} v_r\, u^{r/\beta}\right)^m \\ &= \sum_{m=0}^{\infty} f_m \sum_{r=0}^{\infty} \nu_{m,r}\, u^{r/\beta} = \sum_{m,r=0}^{\infty} f_m\, \nu_{m,r}\, u^{r/\beta}, \end{aligned} \tag{6.24}$$

where $f_m = \sum_{j=m}^{\infty}(-1)^{j-m}\binom{j}{m}\left(\alpha^{-1}\right)_j/j!$, $\left(\alpha^{-1}\right)_j = (\alpha^{-1})(\alpha^{-1}-1)\ldots(\alpha^{-1}-j+1)$ is the decending factorial and $\nu_{m,r} = (rv_0)^{-1}\sum_{n=1}^{r}[n(m+1)-r]\, v_n \nu_{m,r-n}$.

Further, using (6.24) and the binomial expansion, we can write

$$\begin{aligned}\left\{[1-\log(1-u^{1/\beta})]^{1/\alpha} - 1\right\}^k &= \left(\sum_{m,r=0}^{\infty} f_m\, \nu_{m,r}\, u^{r/\beta} - 1\right)^k \\ &= \sum_{i=0}^{\infty} (-1)^i \binom{k}{i} \left(\sum_{r=0}^{\infty} \gamma_r\, u^{r/\beta}\right)^i \\ &= \sum_{i,r=0}^{\infty} (-1)^i \binom{k}{i}\, g_{i,r}\, u^{r/\beta}, \end{aligned} \tag{6.25}$$

where $\gamma_r = \sum_{m=0}^{\infty} f_m \, \nu_{m,r}$ and $g_{i,r} = (r\gamma_0)^{-1} \sum_{s=1}^{r} [s(i+1) - r] \, \gamma_s \, g_{i,r-s}$ (for $r \geq 1$) and $g_{i,0} = \gamma_0^i$.

Thus, replacing (6.25) in (6.23), we obtain

$$\begin{aligned} M(t) &= \sum_{k=0}^{\infty} \frac{t^k}{\lambda^k \, k!} \sum_{i,r=0}^{\infty} (-1)^i \binom{k}{i} g_{i,r} \int_0^1 u^{r/\beta} \, du \\ &= \sum_{i,r=0}^{\infty} \frac{\eta_i \, g_{i,r} \, t^k}{r/\beta + 1}, \end{aligned}$$

where $\eta_i = \sum_{k=0}^{\infty} \frac{(-1)^i}{\lambda^k \, k!} \binom{k}{i}$.

References

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

CHEN, G.; BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, v. 27, p. 154–161, 1995.

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. 1399–1429, 2010.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, v. 81, p. 827–842, 2011.

CORDEIRO, G. M.; LEMONTE, A. J. The β-Birnbaum-Saunders distribution: An improved distribution for fatigue life modeling.*Computational Statistics and Data Analysis*, v. 55, p. 1445–1461, 2011.

COX, D. R.; HINKLEY, D. V. *Theoretical Statistics*. London: Chapman and Hall, 1974.

COX, C.; CHU, H.; SCHNEIDER, M. F.; A. MUÑOZ, A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat. Med.*, v. 26, p. 4352–4374, 2007.

DOORNIK, J.A. *An Object-Oriented Matrix Language*. 5. ed., London, UK: Timberlake Consultants Press, 2007.

GRADSHTEYN, I. S.; RYZHIK, I. M. *Table of integrals*, series, and products. New York: Academic Press, 2007.

GUPTA, R. D.; KUNDU, D. Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, v. 43, 117–130, 2001.

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

HUANG, P. H.; HWANG, T. Y. On new moment estimation of parameters of the generalized gamma distribution using it's characterization. *Taiwanese J. Math.*, v. 10, p. 1083–1093, 2006.

KONG, L.; LEE, C.; SEPANSKI, J. H. On the properties of beta-gamma distribution. *J. Mod. Appl. Stat. Methods*, v. 6, p. 187–211, 2007.

LEMONTE, A. J. 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 (Series D)*, v. 37, p. 25–32, 1988.

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.

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

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; KOLLIA, G. D. A Generalization of the Weibull Distribution with Application to the Analysis of Survival Data. *Journal of the American Statistical Association*, v. 91, p. 1575–1583, 1993.

NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. *Mathematics and Computers in Simulation*, v. 74, p. 1–7, 1997.

NADARAJAH, S. Explicit expressions for moments of order statistics. *Statistics and Probability Letters*, v. 78, p. 196–205, 2008.

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

NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. The gamma-G family of distributions. Mathematical properties and applications. To appear in *Comm. Statist. Theory - Methods*, 2013.

NICHOLS, M. D.; PADGETT, W. J. A Bootstrap control chart for Weibull percentiles. *Quality and Reliability Engineering International*, p. 22, p. 141–151, 2006.

ORTEGA, E. M. M.; BOLFARINE, H.; PAULA, G. A. Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis*, v. 42, p. 165–186, 2003.

ORTEGA, E. M. M.; PAULA, G. A.; BOLFARINE, H. Deviance residuals in generalized log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation*, v. 78, p. 747–764, 2008.

ORTEGA, E. M. M.; CANCHO, V.G.; PAULA, G. A. Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*, v. 15, p. 79–106, 2009.

PROSCHAN, F. Theoretical explanation of observed decreasing failure rate. *Technometrics*, v. 5, p. 375–383, 1963.

RÉNYI, A. On measures of information and entropy Proceedings of the 4th Berkeley Symposium on Mathematics, *Statistics and Probability 1960*, p. 547–561, 1961.

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

SHANNON, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, v. 27, p. 379–423, 1948.

STACY, E. W. A generalization of the gamma distribution. *Annals of Mathematical Statistics*, v. 33, p. 1187–1192, 1962.

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

CHAPTER 7

The Gamma Extended Weibull Distribution

Resumo

Neste capítulo, estudamos a distribuição tempo de vida de quatro parâmetros denominada gama Weibull estendida, que generaliza as distribuições de Weibull e Weibull estendida, entre vários outros. Obtemos expressões explícitas para os momentos incompletos, funções quantílica e geradora, desvios médios, entropias e confiabilidade. O método de máxima verossimilhanccã é usado para estimar os parâmetros do modelo. A aplicabilidade do novo modelo é ilustrada por meio de um conjunto de reais.

Palavras-chave: Desvios médios. Distribuição gama Weibull estendida. Distribuição Weibull estendida. Estimação de máxima verossimilhança. Função geradora. Função quantílica.

Abstract

We study a four-parameter lifetime distribution named the gamma extended Weibull model, which generalizes the Weibull and extended Weibull distributions, among several others. We obtain explicit expressions for the raw and incomplete moments, generating and quantile functions, mean deviations, entropies and reliability. The method of maximum likelihood is used for estimating the model parameters. The applicability of the new model is illustrated by means of a real data set.

Keywords: Extended Weibull distribution; Gamma extended Weibull distribution; Generating function; Maximum likelihood estimation; Mean deviation; Quantile function.

7.1 Introduction

There are hundreds of continuous univariate distributions and recent developments focus on constructing wider distributions from classic ones. The two-parameter Weibull has been the most popular distribution for modeling lifetimes. However, its major weakness is its inability to accommodate non-monotone hazard rates. This has led to new generalizations of this distribution. One of the first extensions allowing for non-monotone hazard rates, including the bathtub shaped hazard rate function (hrf), is the *exponentiated Weibull* (Exp-W) distribution studied by Mudholkar and Srivastava (1993), Mudholkar *et al.* (1995) and Mudholkar *et al.* (1996). It has been well established in the literature that the Exp-W distribution provides significantly better fits than traditional models based on the exponential, gamma, Weibull and log-normal distributions. In the last paper, the authors presented a three-parameter *extended Weibull* (EW) model to yield a more flexible distribution. Further, Shao *et al.* (2004) used this distribution to study flood frequency and Hao and Singh (2008) described some of its applications in hydrology. We take the EW distribution as the baseline distribution for a further generalization introduced here.

The three-parameter EW distribution is defined by the pdf and cdf (Mudholkar *et al.*, 1996)

$$g_{\lambda,\alpha,\beta}(x) = \lambda\,\beta\,x^{\beta-1}\,(1+\alpha\lambda\,x^{\beta})^{-\frac{1}{\alpha}-1},\ \alpha>0 \quad \text{and} \quad g_{\lambda,\beta}(x) = \lambda\,\beta\,x^{\beta-1}\,\mathrm{e}^{-\lambda x^{\beta}},\ \alpha=0, \tag{7.1}$$

$$G_{\lambda,\alpha,\beta}(x) = 1-(1+\alpha\lambda x^{\beta})^{-\frac{1}{\alpha}},\ \alpha>0 \quad \text{and} \quad G_{\lambda,\beta}(x) = 1-\mathrm{e}^{-\lambda x^{\beta}},\ \alpha=0, \tag{7.2}$$

respectively, where $\lambda > 0$ is a scale parameter and $\alpha \geq 0$ and $\beta > 0$ are shape parameters. The support of the EW distribution is $(0,\infty)$. The forms of the pdf and cdf when α goes to zero tend to those ones of the case $\alpha = 0$. Clearly, the cdf (7.2) extends the Weibull cdf and this fact justifies the name EW model. Due to the shape parameter α, more flexibility can be incorporated in model (7.1), which is useful for lifetime data. The survival function associated to (7.1) is $S_{\lambda,\alpha,\beta}(x) = 1 - G_{\lambda,\alpha,\beta}(x)$ for $\alpha > 0$ and $S_{\lambda,\beta}(x) = 1 - G_{\lambda,\beta}(x)$ for $\alpha = 0$.

A family of univariate distributions generated by gamma random variables was proposed by Zografos and Balakrishnan (2009) and Ristić and Balakrishnan (2012). For any baseline cdf G(x), $x \in \mathbb{R}$, they defined the *gamma-G* (GG for short) model with an extra parameter $a > 0$ by the pdf and cdf given by

$$f(x) = \frac{g(x)}{\Gamma(a)}\,\{-\log[1-G(x)]\}^{a-1} \tag{7.3}$$

and

$$F(x) = \frac{1}{\Gamma(a)}\int_0^{-\log[1-G(x)]} t^{a-1}\,\mathrm{e}^{-t}\,dt = \gamma_1(a,-\log[1-G(x)]), \tag{7.4}$$

respectively, where $g(x) = dG(x)/dx$, $\Gamma(a) = \int_0^\infty t^{a-1}\,\mathrm{e}^{-t}dt$ is the gamma function, $\gamma(a,z) = \int_0^z t^{a-1}\mathrm{e}^{-t}dt$ denotes the incomplete gamma function and $\gamma_1(a,z) = \gamma(a,z)/\Gamma(a)$ is the incomplete gamma function ratio.

The GG distribution has the same parameters of the parent G distribution plus one extra shape parameter $a > 0$. Each new GG distribution can be determined from a specified G distribution. For $a = 1$, the G distribution is a basic exemplar with a continuous crossover towards cases with different shapes (for example, a particular combination of skewness and kurtosis).

In this paper, we introduce a new four-parameter model called the *"gamma extended Weibull"* (denoted with the prefix "GEW" for short) distribution which contains several distributions as special models including the EW distribution. This distribution represents only a basic exemplar of the GEW distribution. We study some of its mathematical properties. The paper is outlined as follows. In Section 7.1, we define the GEW distribution and provide some of its special models. Further, two useful expansions for its density and cumulative distributions are given in Section 7.2. In Section 7.3, we obtain its quantile function (qf). The generating function, moments and mean deviations are presented in Sections 7.4, 7.5 and 7.6, respectively. Closed-form expressions for the Rényi and Shannon entropies are derived in Section 7.7. The reliability is investigated in Section 7.8. Maximum likelihood estimation of the model parameters and some inferential tools are discussed in Section 7.9. The usefulness of the new model is provided by means of an application to a real data set in Section 7.10. Some conclusions are offered in Section 7.11.

7.2 The GEW Distribution

By taking the pdf (7.1) and cdf (7.2) of the EW distribution with scale parameter $\lambda > 0$ and shape parameters $\alpha \geq 0$ and $\beta > 0$, the pdf and cdf of the GEW distribution are obtained from equations (7.3) and (7.4) (for $x > 0$) as

$$f(x;\boldsymbol{\tau},a) = \begin{cases} \frac{\lambda\,\beta\,x^{\beta-1}\,(1+\alpha\lambda\,x^{\beta})^{-\frac{1}{\alpha}-1}}{\alpha^{a-1}\,\Gamma(a)} \left\{\log[(1+\alpha\lambda\,x^{\beta})]\right\}^{a-1}, & \alpha > 0, \\ \frac{\lambda^{a}\,\beta\,x^{\beta a-1}\,\mathrm{e}^{-\lambda x^{\beta}}}{\Gamma(a)}, & \alpha = 0, \end{cases} \tag{7.5}$$

$$F(x;\boldsymbol{\tau},a) = \begin{cases} \gamma_1\left(a, \frac{1}{\alpha}\log[1+\alpha\lambda\,x^{\beta}]\right) & \alpha > 0, \\ \gamma_1(a, \lambda x^{\beta}), & \alpha = 0, \end{cases} \tag{7.6}$$

where $\tau = (\alpha, \beta, \lambda)$. Clearly, when $\alpha \to 0$, the first expressions in (7.5) and (7.6) tend to the second ones in these equations. For $\alpha = 0$, we note that the GEW distribution is identical to the *generalized gamma* (GG) distribution pioneered by Stacy (1962). Hereafter, a random variable X having pdf (7.5) is denoted by $X \sim$GEW$(\boldsymbol{\tau}, a)$. Evidently, the density function (7.5) does not involve any complicated function and the EW distribution arises as the basic exemplar for $a = 1$. It is a positive point for the current generalization. The GEW model has several submodels: the extended Weibull (EW) when $a = 1$, gamma Weibull (GW) when $\alpha = 0$, gamma extended exponential (GEE) when $\beta = 1$, extended exponential (EE) when $\beta = a = 1$, gamma exponential (GE) when $\alpha = 0$ and $\beta = 1$, Weibull (W) when $\alpha = 0$ and $a = 1$, and exponential when $\alpha = 0$ and $a = \beta = 1$. We motivate the paper by comparing the performance of the GEW distribution and some of its sub-models applied to a real data set in Section 7.10. In Figure 7.1, we display shapes of the pdf (7.5) for some parameter values. Plots of the density function (7.5) and its hrf for selected parameter values are displayed in Figures 7.1 and 7.2, respectively.

The new distribution is easily simulated as follows: if V is a gamma random variable with parameter a, then

$$X = \left(\alpha^{-1}\,\lambda^{-1}\,\left\{[\exp(-V)]^{-\alpha} - 1\right\}\right)^{1/\beta}$$

has the GEW$(\boldsymbol{\tau}, a)$ distribution. This generate scheme is straightforward because of the existence of fast generators for gamma random variables.

7.3 Useful expansions

Expansions for equations (7.5) and (7.6) can be derived using the concept of exponentiated distributions. The *exponentiated extended Weibull* (exp-EW) distribution follows by raising the cdf (7.2) to a power $a > 0$. Let $Y \sim$ exp-EW$(\boldsymbol{\tau}, a)$ for $a > 0$ be a random variable having this distribution. The cdf and pdf of Y are given by

$$H_a(x;\boldsymbol{\tau}) = \begin{cases} \left[1 - (1+\alpha\lambda\, x^{\beta})^{-\frac{1}{\alpha}}\right]^a, & \alpha > 0, \\ \left[1 - \exp(-\lambda\, x^{\beta})\right]^a, & \alpha = 0, \end{cases}$$

and

$$h_a(x;\boldsymbol{\tau}) = \begin{cases} a\,\lambda\,\beta\, x^{\beta-1}\,(1+\alpha\lambda\, x^{\beta})^{-\frac{1}{\alpha}-1}[1-(1+\alpha\lambda\, x^{\beta})^{-\frac{1}{\alpha}}]^{a-1}, & \alpha > 0, \\ a\,\lambda\,\beta\, x^{\beta-1}\,\mathrm{e}^{-\lambda x^{\beta}}[1-\exp(-\lambda\, x^{\beta})]^{a-1}, & \alpha = 0, \end{cases} \tag{7.7}$$

respectively.

Using http://functions.wolfram.com/ GammaBetaErf/ InverseGammaRegularized/ 06/ 01/ 03/ 0001/, we can write

$$\begin{aligned} &\left\{\frac{1}{\alpha}\log\left[1-\alpha\lambda\, x^{\beta}\right]\right\}^{a-1} \\ &= \quad (a-1)\sum_{k=0}^{\infty}\binom{k+1-a}{k}\sum_{j=0}^{k}\frac{(-1)^{j+k}\binom{k}{j}\,p_{j,k}}{(a-1-j)}\,[1-(1-\alpha\lambda\, x^{\beta})^{-\frac{1}{\alpha}}]^{a+k-1}, \end{aligned}$$

where $a > 0$ is a real parameter and the constants $p_{j,k}$ are given recursively by

$$p_{j,k} = k^{-1}\sum_{m=1}^{k}\frac{(-1)^m\,[m(j+1)-k]}{(m+1)}\,p_{j,k-m},$$

for $k = 1, 2, \ldots$ and $p_{j,0} = 1$.

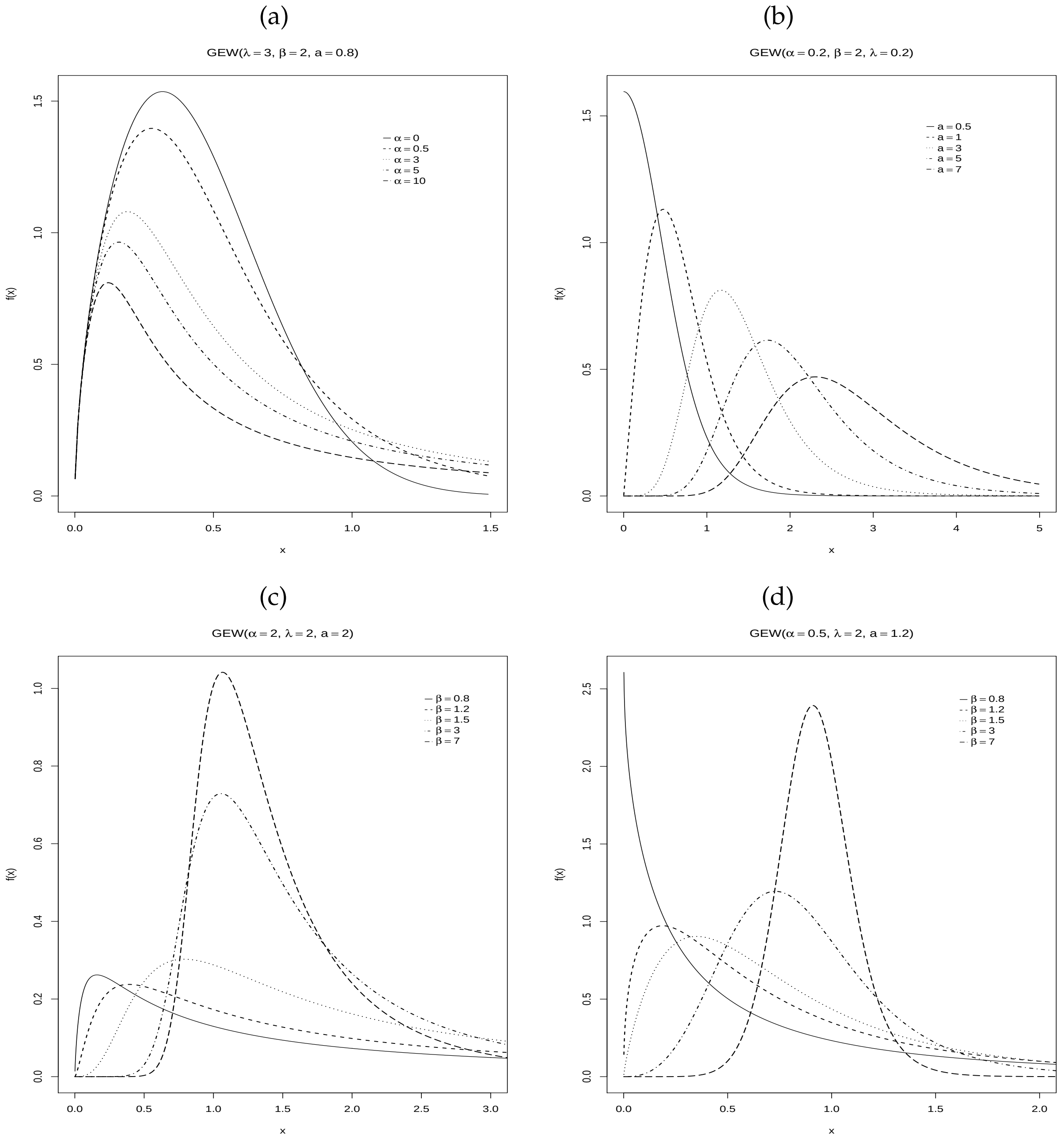


Figure 7.1: Plots of the GEW density.

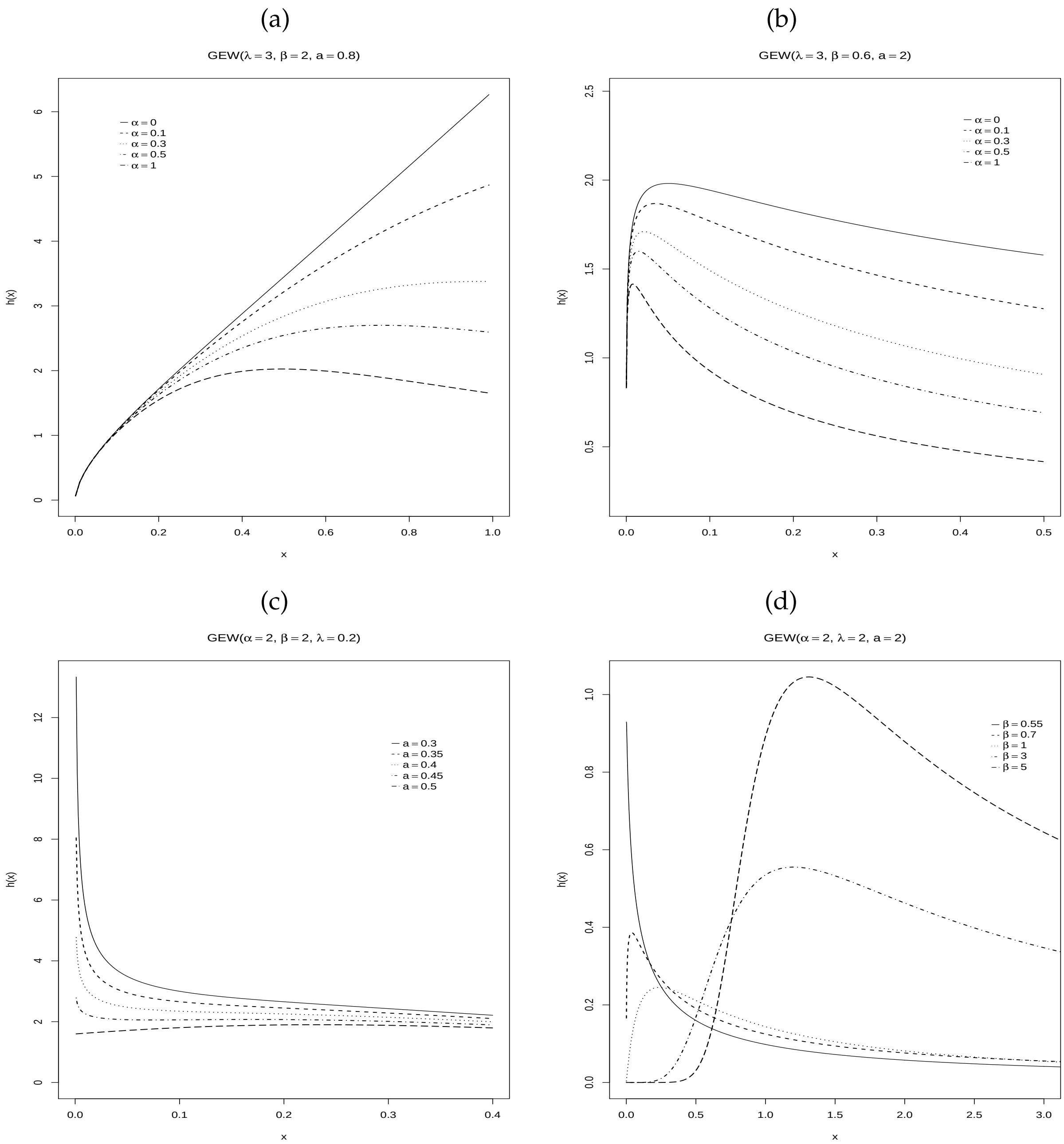


Figure 7.2: Plots of the GEW hazard rate function.

Further, for any real parameter $a > 0$, we define

$$b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)} \sum_{j=0}^{k} \frac{(-1)^{j+k}\binom{k}{j}\, p_{j,k}}{(a-1-j)}$$

and then the first equation in (7.5) can be expressed as

$$f(x) = \sum_{k=0}^{\infty} b_k\, h_{a+k}(x), \tag{7.8}$$

where $h_{a+k}(x) = (a+k)\,\lambda\,\beta\,x^{\beta-1}\,(1+\alpha\lambda\,x^{\beta})^{-\frac{1}{\alpha}-1}\,[1-(1+\alpha\lambda\,x^{-\frac{1}{\alpha}})]^{a+k-1}$ denotes the exp-EW$(\boldsymbol{\tau}, a+k)$ density function. The cdf corresponding to (7.8) is given by

$$F(x) = \sum_{k=0}^{\infty} b_k\, H_{a+k}(x),$$

where $H_{a+k}(x) = \left[1-(1+\alpha\lambda\,x^{\beta})^{-\frac{1}{\alpha}}\right]^{a+k}$ represents the exp-EW cdf with parameters α, λ, β and $a+k$.

Similarly, we can derive an expansion for the GW density (when $\alpha = 0$). Using the second equation in (7.5), we obtain the same expression in (7.8), but the function $h_{a+k}(x)$ denotes, in the case $\alpha = 0$, the GEW$(\boldsymbol{\tau}, a+k)$ density function given by the second equation of (7.7).

After some algebra using (7.7) in (7.8), we obtain the representation

$$f_{\text{GEW}}(x;\boldsymbol{\tau},a) = \begin{cases} \sum\limits_{j=0}^{\infty} e_j\; g_{\lambda^\star,\alpha^\star,\beta}(x) & \text{if} \quad \alpha > 0, \\ \sum\limits_{j=0}^{\infty} e_j\; g_{\lambda^\star,\beta}(x) & \text{if} \quad \alpha = 0, \end{cases} \tag{7.9}$$

where $e_j = \sum_{k=0}^{\infty}(-1)^j\,(a+k)\,b_k\,\binom{a+k-1}{j}$, $g_{\lambda^\star,\alpha^\star,\beta}(x)$ denotes the EW pdf with parameters $\lambda^\star = (j+1)\lambda$, $\alpha^* = \alpha/(j+1)$ and β and $g_{\lambda^\star,\beta}(x)$ denotes the Weibull pdf with parameters $\lambda^\star$ and β.

By integrating (7.9), we obtain

$$F_{\text{GEW}}(x;\boldsymbol{\tau},a) = \begin{cases} \sum\limits_{j=0}^{\infty} e_j\, G_{\lambda^\star,\alpha^\star,\beta}(x) & \text{if} \quad \alpha > 0, \\ \sum\limits_{j=0}^{\infty} e_j\, G_{\lambda^\star,\beta}(x) & \text{if} \quad \alpha = 0. \end{cases} \tag{7.10}$$

Equations (7.9) and (7.10) are the main formulae of this section. They indicate that the GEW density function is a linear combination of EW (when $\alpha > 0$) and Weibull (when $\alpha = 0$) densities. So, several GEW structural properties can be obtained from those properties of the EW and Weibull distributions.

7.4 Quantile Function

First, we consider the general case $\alpha > 0$. The GEW qf, say $Q(u) = F^{-1}(u)$, can be expressed in terms of the EW qf ($Q_{EW}(\cdot)$). Inverting $F(x) = u$ given by (7.6), we obtain the qf of

X (for $0 < u < 1$) as

$$F^{-1}(u) = Q_{GEW}(u) = Q_{EW}\left\{1 - \exp[-Q^{-1}(a, 1-u)]\right\}, \tag{7.11}$$

where $Q^{-1}(a, u)$ is the inverse function of $Q(a, z) = 1 - \gamma_1(a, z)$. Quantities of interest can be obtained from (7.11) by substituting appropriate values for u. Further, the EW qf is given by

$$Q_{EW}(u) = \left\{\frac{1-(1-u)^{\alpha}}{\alpha\lambda(1-u)^{\alpha}}\right\}^{1/\beta} \tag{7.12}$$

We can obtain the inverse function $Q^{-1}(a, u)$ in the Wolfram website as

$$z = Q^{-1}(a, 1-u) = \sum_{i=0}^{\infty} a_i\, u^{i/a},$$

where $a_0 = 0$, $a_1 = \Gamma(a+1)^{1/a}$, $a_2 = \Gamma(a+1)^{2/a}/(a+1)$, $a_3 = (3a+5)\Gamma(a+1)^{3/a}/[2(a+1)^2(a+2)]$, etc.

By expanding the exponential function and using (2.7), we obtain

$$\exp\left(-\sum_{r=0}^{\infty} a_r u^{r/a}\right) = \sum_{r=0}^{\infty} p_r\, u^{r/a},$$

where the p_r's are defined in Appendix A. We can write

$$Q_{GEW}(u) = \left(\frac{1}{\alpha\lambda}\right)^{1/\beta}\left[\frac{\left(\sum_{r=0}^{\infty} p_r\, u^{r/a}\right)^{\alpha}}{1-\left(\sum_{r=0}^{\infty} p_r\, u^{r/a}\right)^{\alpha}}\right]^{-\beta^{-1}}.$$

By expanding $\left(\sum_{r=0}^{\infty} p_r\, u^{r/a}\right)^{\alpha}$, we can write $Q_{GEW}(u)$ as follows

$$Q_{GEW}(u) = \left(\frac{1}{\alpha\lambda}\right)^{1/\beta}\left(\frac{\sum_{i=0}^{\infty}\tau_i\, u^{i/a}}{\sum_{i=0}^{\infty}\eta_i\, u^{i/a}}\right)^{-1/\beta},$$

where $\tau_i = \sum_{s=0}^{\infty} f_s\, \zeta_{s,i}$, $\zeta_{s,i} = (i\, p_0)^{-1}\sum_{m=0}^{i}[m(s+1)-i]p_m\, \zeta_{s,i-m}$, $f_s = \sum_{k=s}^{\infty}\frac{(-1)^{k-s}\,(\alpha)_k}{k!}\binom{k}{s}$, $(\alpha)_k$ is the descending factorial, $\eta_0 = 1-\tau_0$ and $\eta_r = -\tau_r$ (r≥ 1).

Now, the ratio between the two power series reduces to

$$Q_{GEW}(u) = \left(\frac{\eta_0}{\alpha\lambda}\right)^{1/\beta}\sum_{i=0}^{\infty}\gamma_i\, u^{i/a} = \sum_{i=0}^{\infty}\gamma_i^{\star}\, u^{i/a}. \tag{7.13}$$

More details about (7.13) and other quantities are given in Appendix A.

Secondly, for the case $\alpha = 0$, the algebraic calculations are much simpler. The Weibull qf is given by

$$Q_W(u) = \left[-\lambda^{-1}\log(1-u)\right]^{1/\beta}. \tag{7.14}$$

Thus, substituting this result in (7.12) and after some algebra, we obtain

$$Q_{GW}(u) = \sum_{j=0}^{\infty}\wp_j^{\star}\, u^j, \tag{7.15}$$

where $\wp_j^{\star} = \left(\frac{1}{\lambda}\right)^{\beta^{-1}}\wp_j$, $\wp_j = \sum_{k=j}^{\infty}(-1)^{k-j}\,(\beta^{-1})_k\binom{k}{j}/k!$ and $(a)_k = a(a-1)\ldots(a-k+1)$ denotes the descending factorial.

Equations (7.13) and (7.15) are the basic results of this section, since we can obtain from them several mathematical quantities for the proposed model.

7.5 Generating Function

We now provide formulae for the moment generating function (mgf) $M(t) = E(\mathrm{e}^{\mathrm{tX}})$ of X, using the qf of X obtained in the last section. We consider two different cases. First, for $\alpha > 0$, by expanding the exponential function, we have

$$\begin{aligned} M(t) &= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^{\infty} \mathrm{e}^{\mathrm{tx}}\,\mathrm{g(x)G(x)^{a+k-1}\,dx} \\ &= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^{1} \mathrm{e}^{\mathrm{t\,Q_{EW}(u)}}\,\mathrm{u^{a+k-1}\,du}. \end{aligned} \tag{7.16}$$

But, expanding the binomial term in (7.12), we have

$$\begin{aligned} Q_{EW}(u) &= \left\{\frac{1-(1-u)^{\alpha}}{\alpha\lambda(1-u)^{\alpha}}\right\}^{1/\beta} \\ &= \left(\frac{1}{\alpha\lambda}\right)^{1/\beta} \left\{\frac{1-\sum_{j=0}^{\infty}(-u)^j\binom{\alpha}{j}}{\sum_{j=0}^{\infty}(-u)^j\binom{\alpha}{j}}\right\}^{1/\beta} \\ &= \left(\frac{1}{\alpha\lambda}\right)^{1/\beta} \left\{\frac{\sum_{j=0}^{\infty}\theta_j\,u^j}{\sum_{j=0}^{\infty}\overline{\theta}_j\,u^j}\right\}^{-1/\beta}, \end{aligned}$$

where $\overline{\theta}_0 = 1-\theta_0$, $\overline{\theta}_j = \theta_j$ $(j \geq 1)$ and $\theta_j = (-1)^j\binom{\alpha}{j}$.

Thus, the ratio between the two power series, we obtain

$$\begin{aligned} Q_{EW}(u) &= \left(\frac{q_0}{\alpha\lambda}\right)^{1/\beta} \left[\sum_{j=0}^{\infty}\kappa_j\,u^j\right]^{1/\beta} \\ &= \sum_{j=0}^{\infty}\overline{\nu}_j u^j, \end{aligned} \tag{7.17}$$

where

$$\kappa_j + \frac{1}{q_0}\sum_{j=1}^{r}\kappa_{r-j}\,q_j - \theta_r = 0,$$

$\overline{\nu} = \left(\frac{q_0}{\alpha\lambda}\right)^{1/\beta}\nu_j$, $\nu_j = \sum_{n=0}^{\infty} g_n\eta_{n,j}$, $\eta_{n,j} = (j\kappa_0)^{-1}\sum_{m=1}^{j}[m(n+1)-j]\,\kappa_m\,\eta_{j-m}$, $g_n = \sum_{i=n}^{\infty}\frac{(-1)^{i-n}\binom{i}{n}(-1/\beta)_i}{i!}$ and $(-1/\beta)_i$ is descending factorial.

Thus, replacing the result (7.17) in (7.16), we obtain

$$\begin{aligned}
M(t) &= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^1 \exp\left\{t \sum_{j=0}^{\infty} \overline{v}_j u^j\right\} u^{a+k-1} du \\
&= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^1 \sum_{r=0}^{\infty} \frac{t^r}{r!} \left(\sum_{j=0}^{\infty} \overline{v}_j u^j\right)^r u^{a+k-1} du \\
&= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^1 \left(1 + \sum_{r=1}^{\infty}\sum_{j=0}^{\infty} \frac{t^r}{r!}\, d_{r,j}\, u^j\right) u^{a+k+j-1} du \\
&= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^1 \left(1 + \sum_{j=0}^{\infty} \xi_j\, u^j\right) u^{a+k+j-1} du \\
&= \sum_{j,k=0}^{\infty} \frac{b_k\,(a+k)\,\xi_j^{\star}}{a+j+k}, \qquad (7.18)
\end{aligned}$$

where $\xi_0^{\star} = 1 + \xi_0$, $\xi_j^{\star} = \xi_j\ (j \geq 1)$, $\xi_j = \frac{t^r}{r!}\, d_{r,j}$, $d_{r,j} = (j\overline{v}_0)^{-1} \sum_{i=0}^{j} [i(r+1) - j]\, \overline{v}_r\, d_{r,j-i}$, for $r \geq 1$ and $j \geq 1$ and $d_{r,0} = \overline{v}_0^r$.

Secondly, for $\alpha = 0$, in a similar manner we can obtain the mgf. From (7.14), we obtain

$$\begin{aligned}
Q_W(u) &= \left[-\lambda^{-1} \log(1-u)\right]^{1/\beta} = \lambda^{-1/\beta} \left(\sum_{i=0}^{\infty} \frac{(-1)^{i+2}\, u^{i+1}}{i+1}\right)^{1/\beta} \\
&= u^{1/\beta} \left(\sum_{i=0}^{\infty} \psi_i\, u^i\right)^{1/\beta},
\end{aligned}$$

where $\psi_i = \frac{(-1)^{i+2}}{(i+1)\lambda}$.

Thus, expanding the last parenthesis, we have

$$Q_W(u) = u^{1/\beta} \sum_{i=0}^{\infty} \theta_i\, u^i,$$

where $\theta_i = \sum_{n=0}^{\infty} f_n\, \varsigma_{n,i}$, $\varsigma_{n,i} = (i\psi_0)^{-1} \sum_{m=0}^{i} [m(n+1) - i]\, \psi_i\, \varsigma_{n,i-m}$, $f_n = \sum_{j=n}^{\infty} \frac{(-1)^{j-n}\,(1/\beta)_j}{j!} \binom{j}{n}$ and $(1/\beta)_j$ is the descending factorial.

Using the last equation, the mgf for $\alpha = 0$ is given by

$$\begin{aligned}
M(t) &= \sum_{k=0}^{\infty} b_k\,(a+k) \int_0^1 \mathrm{e}^{\mathrm{t\,Q_W(u)}}\, \mathrm{u}^{\mathrm{a+k-1}}\, \mathrm{du} \\
&= \sum_{i,k=0}^{\infty} \frac{b_k\,(a+k)\,\varphi_i^{\star}}{a+k+i+1/\beta}, \qquad (7.19)
\end{aligned}$$

where $\varphi_0^{\star} = 1 + \varphi_0$, $\varphi_i^{\star} = \varphi_i\ (i \geq 1)$, $\varphi_i = \sum_{l=1}^{\infty} \tau_i\, t_{l,i}$, $\tau_l = \frac{t^l}{l!}$, $t_{l,i} = (i\theta_0)^{-1} \sum_{m=1}^{i} [m(l+1) - i]\, \theta_m\, t_{l,i-m}$ for $l \geq 1$ and $i \geq 1$ and $d_{l,0} = \theta_0^l$.

By substituting known parameters in equations (7.18) and (7.19), we can obtain specific formulae for GEW special models.

7.6 Moments

Some of the most important features and characteristics of a distribution can be studied through moments. Consequently, we can obtain from (7.8)

$$\mu'_r = E(X^r) = \begin{cases} \alpha^{-(\frac{r}{\beta}+1)}\,\lambda^{-\frac{r}{\beta}} \sum_{j=0}^{\infty} e_j^{\star}\, B\left(\frac{r}{\beta}+1, \frac{j+1}{\alpha} - \frac{r}{\beta}\right) & \text{if } \alpha > 0 \text{ and } r < \beta/\alpha, \\ \lambda^{-\frac{r}{\beta}}\,\Gamma\left(\frac{r}{\beta}+1\right) \sum_{j=0}^{\infty} \frac{e_j}{(j+1)^{r/\beta}} & \text{if } \alpha = 0, \end{cases} \tag{7.20}$$

where $e_j^{\star} = e_j\,(j+1)$.

Established algebraic expansions to determine $E(X^r)$ can be more efficient than computing these moments directly by numerical integration of (7.5), which can be prone to rounding errors among others. Further, the central moments (μ_r) and cumulants (κ_r) of X are determined from (7.20) by using the well-known relationships

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

respectively, where $\kappa_1 = \mu'_1$. Then, $\kappa_2 = \mu'_2 - \mu_1'^2$, $\kappa_3 = \mu'_3 - 3\mu'_2\,\mu'_1 + 2\mu_1'^3$, $\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu_2'^2 + 12\mu'_2\mu_1'^2 - 6\mu_1'^4$, etc. The skewness $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ and kurtosis $\gamma_2 = \kappa_4/\kappa_2^2$ coefficients can be obtained readily from the second, third and fourth cumulants. Figure 7.3 and 7.4 displays some plots of the skewness and kurtosis of the GEW model.

For lifetime models, it is of interest to know the rth incomplete moment of X defined by $T_r(y) = \int_0^y x^r\, f(x)dx$. Moreover, it is simple to verify from (7.9) that $T_r(y)$ (when $\alpha > 0$) can be expressed as

$$T_r(y) = \sum_{j=0}^{\infty} e_j\, \lambda^{\star}\, \beta\, \rho(y; r, \alpha^{\star}\lambda^{\star}, \alpha^{\star^{-1}}), \tag{7.21}$$

where

$$\rho(y; r, p, q) = \int_0^y x^r\, (1 + p\,x)^{-q-1} dx$$

for $r, p, q > 0$.

Using Maple, this integral can be determined as

$$\rho(y; r, p, q) = A(r, p, q) \left\{ 2\,y^{r-q}\; {}_2F_1\left(q^{-r}, q+1; q+1-r, -\frac{1}{py}\right) \left[B(r, p, q) + C(r, p, q)\right] \right\}, \tag{7.22}$$

where ${}_2F_1$ is the hypergeometric function defined by

$${}_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{x^j}{j!},$$

$$A(r, p, q) = \left\{ p\, \sin[\pi(q-r)]\, (q-r)\, \Gamma(q+1)\, \Gamma(r+1-q) \right\}^{-1},$$

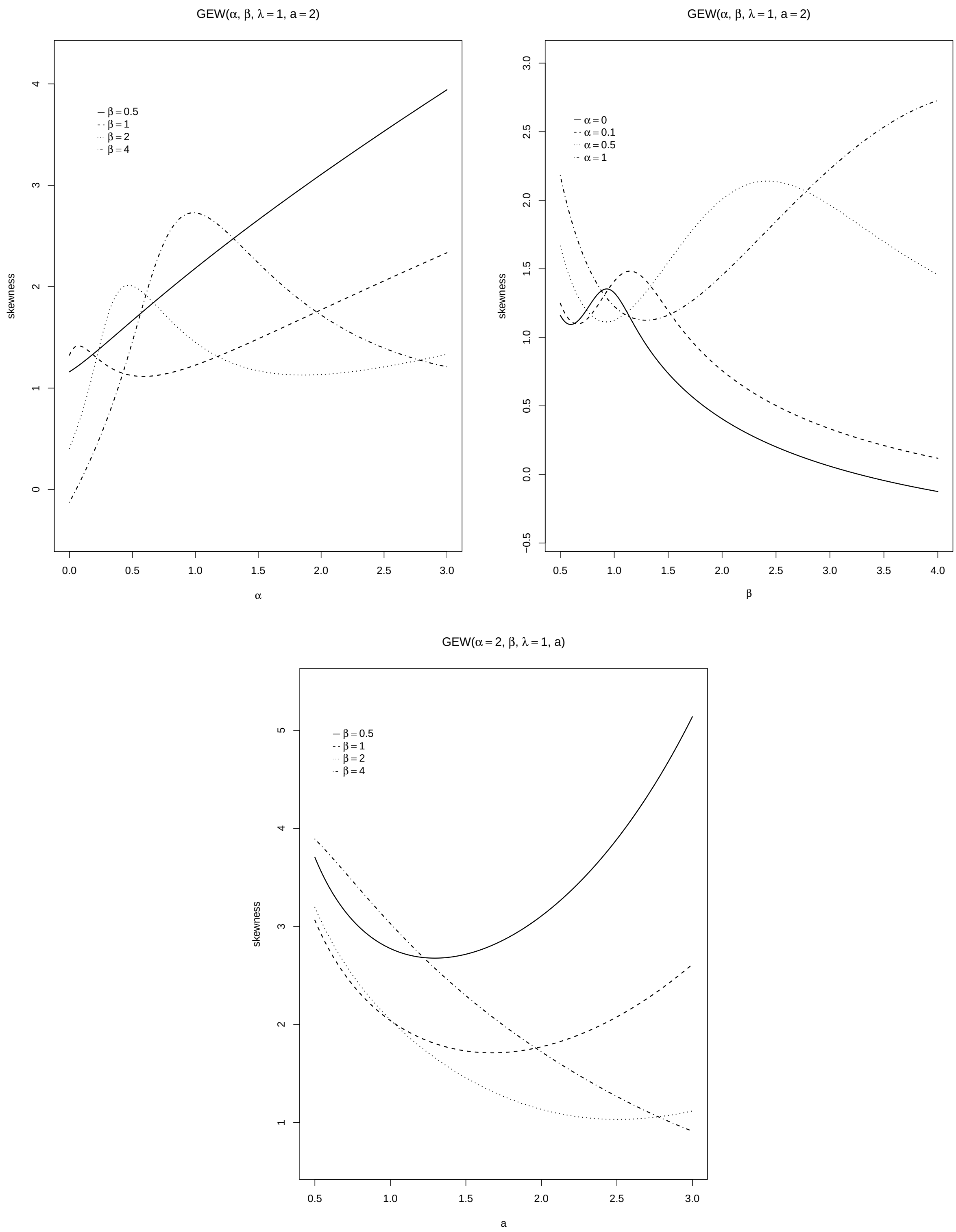


Figure 7.3: Skewness of the GEW distribution for several choices of the parameters.

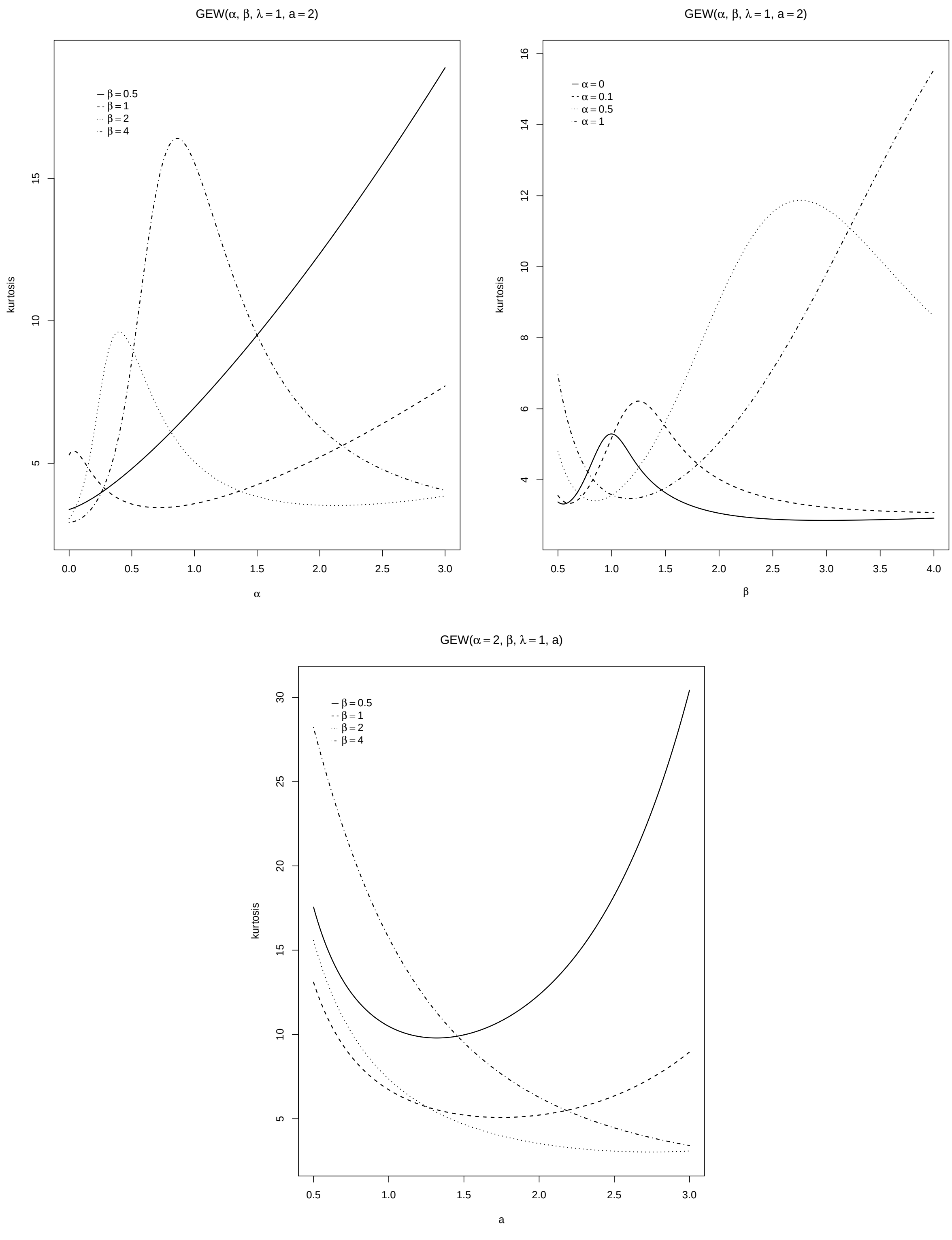


Figure 7.4: Kurtosis of the GEW distribution for several choices of the parameters.

$$B(r,p,q) = p^{-q}\,\Gamma(q+1)\,\Gamma(r+1-q)\,\big[\cos(q\pi)\,\sin(\pi r) - \sin(q\pi)\,\cos(\pi r)\big],$$

and

$$C(r,p,q) = \pi\,p^{-r}\,\Gamma(r+1)(q-r).$$

Combining equations (7.21) and (7.22), we obtain the incomplete moments. For $\alpha = 0$, $T_r(y)$ can be expressed as

$$T_r(y) = \lambda^{-r}\sum_{j=0}^{\infty}\frac{e_j}{(j+1)^r}\,\gamma(r+1,(j+1)\lambda y).$$

7.7 Mean Deviations

The mean deviations about the mean and about the median can be used as measures of spread in a population. They are given by $\delta_1 = E(|X-\mu_1'|) = 2\mu_1'\,F(\mu_1') - 2\,T_1(\mu_1')$ and $\delta_2 = E(|X-m|) = \mu_1' - 2\,T_1(m)$, respectively, where the mean μ_1' is determined from (7.20) and $T_1(\cdot)$ follows from (7.21) with $r = 1$ as

$$T_1(y) = \lambda\sum_{j=0}^{\infty} e_j\,\rho(y;1,\alpha^{\star}\lambda^{\star},\alpha^{\star^{-1}}).$$

Here, $\rho(y;1,p,q)$ can be reduced to

$$\rho(y;1,p,q) = -[q(q-1)p^2]^{-1}\left[\frac{(pqy+1)}{(1+py)^q} - 1\right].$$

Further, for $\alpha = 0$, we have

$$T_1(q) = \sum_{j=0}^{\infty} e_j\left(\frac{1}{\lambda^{\star}}\right)^{1/\beta}\int_0^{\lambda^{\star}q^{\beta}} v^{1/\beta}\,\mathrm{e}^{-v}dv$$

and then $T_1(q)$ becomes

$$T_1(q) = \sum_{j=0}^{\infty} e_j\left(\frac{1}{(j+1)\lambda}\right)^{1/\beta}\left[\Gamma\left(1+\frac{1}{\beta}\right) - \gamma\left(1+\frac{1}{\beta},(j+1)\lambda q^{\beta}\right)\right].$$

Both equations for $T_1(\cdot)$ can be used to determine Lorenz and Bonferroni curves defined by $L(\pi) = T_1(q)/\mu_1'$ and $B(\pi) = T_1(q)/(\pi\,\mu_1')$, respectively, where $q = Q(\pi)$ is the qf (7.12) at a given probability π.

7.8 Entropy

The Rényi entropy is defined by

$$I_R(\rho) = \frac{1}{1-\rho}\log\left\{\int f(x)^{\rho}dx\right\},$$

where $\rho > 0$ and $\rho \neq 1$. For the GEW distribution, the integral in $I_R(\rho)$ reduces to

$$\frac{1}{1-\rho}\left[\frac{\lambda\beta}{\alpha^{a-1}\Gamma(a)}\right]^{\rho}\int_0^\infty x^{\rho(\beta-1)}(1+\alpha\lambda x^\beta)^{-\rho(1+1/\alpha)}[\log(1+\alpha\lambda x^\beta)]^{\rho(a-1)}dx\,.$$

Setting $y = x^{\rho(\beta-1)}$ and $\gamma = \beta/[\rho(\beta-1)]$, $b = \rho(a-1)+1$, $\xi = \alpha/[\alpha(\rho-1)+\rho]$ and $\delta = \lambda[\alpha(\rho-1)+\rho]$, we have

$$\begin{aligned} I_R(\rho) &= C\int_0^\infty y^{1/[\rho(\beta-1)]}(1+\xi\delta y^\gamma)^{-(1+1/\xi)}[\log(1+\xi\delta y^\gamma)]^{b-1}dy \\ &= C\,E\{Y^{1/[\rho(\beta-1)]}\}\,, \end{aligned}$$

where $Y \sim GEW(\xi,\gamma,\delta,b)$,

$$C = \frac{1}{\rho\,(1-\rho)(\beta-1)}\left[\frac{\lambda\beta}{\alpha^{a-1}\Gamma(a)}\right]^{\rho}\,,$$

and $E\{Y^{1/[\rho(\beta-1)]}\}$ is given by (7.20) for $\rho > \alpha/(\beta-1)$.

For $\alpha > 0$, the Shannon entropy is given by

$$\begin{aligned} -E\{\log[f(X)]\} &= -\log(\lambda)-\log(\beta)+\log[\Gamma(a)]+(a-1)\log(\alpha)-(\beta-1)E[\log(X)] \\ &\quad +(1+1/\alpha)E[\log(1+\alpha\lambda X^\beta)]-(a-1)E\{\log[\log(1+\alpha\lambda X^\beta)]\}. \end{aligned} \tag{7.23}$$

We now obtain the expectations in the right side of (7.23). Setting $y = \log(1+\alpha\lambda X^\beta)$, we have

$$\begin{aligned} E[\log(X)] &= \int_0^\infty \log(x)\,\frac{\lambda\beta x^{\beta-1}}{\alpha^{a-1}\Gamma(a)}(1+\alpha\lambda x^\beta)^{-(1+1/\alpha)}[\log(1+\alpha\lambda x^\beta)]^{a-1}dx \\ &= \frac{1}{\beta\alpha^a\,\Gamma(a)}\Big[\int_0^\infty \log(e^y-1)e^{-y/\alpha}y^{a-1}dy \\ &- \log(\alpha\lambda)\int_0^\infty e^{-y/\alpha}y^{a-1}dy\Big]. \end{aligned} \tag{7.24}$$

The second integral in (7.24) is equal to $\alpha^a\Gamma(a)$. For the first integral, using the power series

$$\log(t) = 2\sum_{j=0}^{\infty}\frac{1}{2j+1}\left(\frac{t-1}{t+1}\right)^{2j+1}\,,\ \text{for}\quad t>0\,,$$

we have

$$\begin{aligned} \int_0^\infty \log(e^y-1)\,e^{-y/\alpha}y^{a-1}dy &= 2\sum_{j=0}^{\infty}\frac{1}{2j+1}\int_0^\infty (1-2e^{-y})^{2j+1}\,e^{-y/\alpha}\,y^{a-1}dy \\ &= 2\sum_{j=0}^{\infty}\frac{1}{2j+1}\sum_{k=0}^{2j+1}\binom{2j+1}{k}(-2)^k\int_0^\infty e^{-y(k+1/\alpha)}\,y^{a-1}dy \\ &= 2\sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}\frac{1}{2j+1}\binom{2j+1}{k}(-2)^k\left(\frac{\alpha}{k\alpha+1}\right)^a\Gamma(a)\,. \end{aligned}$$

Thus,

$$E[\log(X)] = \frac{1}{\beta}\left[\sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}\frac{2}{2j+1}\binom{2j+1}{k}\frac{(-2)^k}{(k\alpha+1)^a}-\log(\alpha\lambda)\right]\,.$$

Also,

$$
\begin{aligned}
E[\log(1+\alpha\lambda X^{\beta})] &= \frac{\lambda\beta}{\alpha^{a-1}\Gamma(a)}\int_0^{\infty} x^{\beta-1}(1+\alpha\lambda x^{\beta})^{-(1+1/\alpha)}[\log(1+\alpha\lambda x^{\beta})]^{a}dx \\
&= \frac{\lambda\beta}{\alpha^{a}\,\Gamma(a)}\int_0^{\infty} y^{a}\,\mathrm{e}^{-y/\alpha}dy = a\,\lambda\,\beta\,\alpha\,,
\end{aligned}
$$

and

$$
\begin{aligned}
& E\{\log[\log(1+\alpha\lambda X^{\beta})]\} \\
= & \frac{\lambda\beta}{\alpha^{a-1}\Gamma(a)}\int_0^{\infty} \log[\log(1+\alpha\lambda x^{\beta})]\,x^{\beta-1}\,(1+\alpha\lambda x^{\beta})^{-(1+1/\alpha)}[\log(1+\alpha\lambda x^{\beta})]^{a-1}dx \\
= & \frac{1}{\alpha^{a}\,\Gamma(a)}\int_0^{\infty} y^{a-1}\,\log(y)\mathrm{e}^{-y/\alpha}dy = \log(\alpha)+\psi(a)\,,
\end{aligned}
$$

where $\psi(.)$ is the digamma function.

Thus, the Shannon entropy for $\alpha > 0$ reduces to

$$
\begin{aligned}
-E\{\log[f(x)]\} &= -\log(\lambda)-\log(\beta)+\log[\Gamma(a)]+(\alpha+1)\,a\,\lambda\,\beta-(a-1)\,\psi(a) \\
&\quad +\frac{\beta-1}{\beta}\left[\log(\alpha\lambda)-\sum_{j=0}^{\infty}\sum_{k=0}^{2j+1}\frac{2}{2j+1}\binom{2j+1}{k}\frac{(-2)^{k}}{(k\alpha+1)^{a}}\right].
\end{aligned}
$$

For $\alpha = 0$, we have

$$
-E\{\log[f(x)]\} = -\,a\,\log(\lambda)-\log(\beta)+\log[\Gamma(a)]+(a\,\beta-1)E[\log(X)]-\lambda E(X^{\beta})\,.
$$

By setting $y = \lambda x^{\beta}$, we can write

$$
\begin{aligned}
E[\log(X)] &= \frac{\lambda^{a}\beta}{\Gamma(a)}\int_0^{\infty}\log(x)\,x^{a\beta-1}\exp(-\lambda x^{\beta})dx \\
&= \frac{1}{\beta\Gamma(a)}\int_0^{\infty}[\log(y)-\log(\lambda)]\;y^{a-1}\,\mathrm{e}^{-y}dy = \frac{1}{\beta}\,[\psi(a)-\,\log(\lambda)]
\end{aligned}
$$

and

$$
E(X^{\beta}) \;=\; \frac{\lambda^{a}\beta}{\Gamma(a)}\int_0^{\infty} x^{\beta(a+1)-1}\exp(-\lambda x^{\beta})dx = \frac{a}{\lambda}\,.
$$

Thus, the Shannon entropy for $\alpha = 0$ is given by

$$
-E\{\log[f(x)]\} = -\,a\,[1+\log(\lambda)]-\log(\beta)+\log[\Gamma(a)]+\frac{a\,\beta-1}{\beta}\,[\psi(a)-\log(\lambda)]\;.
$$

7.9 Estimation

The maximum likelihood method is used for estimating the parameters of the GEW model. We determine the maximum likelihood estimates (MLEs) from complete samples only. Let $x_1,\ldots,x_n$ be a sample of size n from the GEW(α,β,λ,a) distribution. The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (\alpha,\beta,\lambda,a)^{T}$ is given by $l(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log[f(x_i)] = \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$ where

$$
\begin{aligned}
l_i(\boldsymbol{\theta}) &= \log(\lambda)+\log(\beta)-\log[\Gamma(a)]-(a-1)\log(\alpha)+(\beta-1)\log(x_i) \\
&\quad -(1+1/\alpha)\log(1+\alpha\lambda x_i^{\beta})+(a-1)\log[\log(1+\alpha\lambda x_i^{\beta})]. \qquad (7.25)
\end{aligned}
$$

The components of the score vector $U(\boldsymbol{\theta})$ are given by $\sum_{i=1}^{n} \partial l_i(\boldsymbol{\theta})/\partial\theta_j$, $j = 1, \ldots, 4$, where θ_j is substituted by the parameters α, β, λ or a. The elements $\partial l_i(\boldsymbol{\theta})/\partial\theta_j$, $j = 1, \ldots, 4$, are given below:

$$\begin{aligned}
\partial l_i(\boldsymbol{\theta})/\partial\alpha &= -\frac{a-1}{\alpha} + \frac{\log(1+\alpha\,\lambda\,x_i^{\beta})}{\alpha^2} - \frac{(\frac{1}{\alpha}+1)\,\lambda\,x_i^{\beta}}{1+\alpha\,\lambda\,x_i^{\beta}} + \frac{(a-1)\,\lambda\,x_i^{\beta}}{1+\alpha\,\lambda\,x_i^{\beta}\,\log(1+\alpha\,\lambda\,x_i^{\beta})} \\
\partial l_i(\boldsymbol{\theta})/\partial\beta &= \frac{1}{\beta} + \log(x_i) - \frac{(\frac{1}{\alpha}+1)\,\alpha\,\lambda\,x_i^{\beta}\,\log(x_i)}{1+\alpha\,\lambda\,x_i^{\beta}} + \frac{(a-1)\,\alpha\,\lambda\,x_i^{\beta}\,\log(x_i)}{(1+\alpha\,\lambda\,x_i^{\beta})\,\log(1+\alpha\,\lambda\,x_i^{\beta})}, \\
\partial l_i(\boldsymbol{\theta})/\partial\lambda &= \frac{1}{\lambda} - \frac{(\frac{1}{\alpha}+1)\,\alpha\,x_i^{\beta}}{1+\alpha\,\lambda\,x_i^{\beta}} + \frac{(a-1)\,\alpha\,x_i^{\beta}}{(1+\alpha\,\lambda\,x_i^{\beta})\,\log(1+\alpha\,\lambda\,x_i^{\beta})}, \\
\partial l_i(\boldsymbol{\theta})/\partial a &= -\psi(a) - \log(\alpha) + \log[\log(1+\alpha\,\lambda\,x_i^{\beta})]\,.
\end{aligned}$$

Setting these equations to zero and solving them simultaneously yield the MLEs of the four parameters. For interval estimation on the model parameters, we require the expected information matrix. The elements of the 4×4 total observed information matrix $J(\theta) = \{J_{rs}\}$, where $r, s \in \{\alpha, \beta, \lambda, a\}$, are given in Appendix B.

The multivariate normal $N_4(0, J(\widehat{\boldsymbol{\theta}})^{-1})$ distribution, where $J(\widehat{\boldsymbol{\theta}})^{-1}$ is the observed information matrix evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, can be used to construct approximate confidence regions for the parameters.

The likelihood ratio (LR) statistic can be used for comparing the GEW distribution with some of its special models. We can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain LR statistics for testing some of its sub-models. In any case, hypothesis tests of the type $H_0 :=_0$ versus $H :\neq_0$, where is a vector formed with some components of $\boldsymbol{\theta}$ and $_0$ is a specified vector, can be performed using LR statistics.

7.10 Application

In the application, we use Tippett's (1950) warp break data for six types of weaving warps. We describe Tippett's experiment from Tippett (1950, p.105): "The results of a weaving experiment was conducted in a factory. There were 6 lots of warp yarn labelled respectively *AL*, *AM*, etc. They were spun from two growths of cotton, *A* and *B*, and each cotton was spun to three twists (i.e., the number of turns in the yarn per inch): low (*L*), medium (*M*), and high (*H*). The combination of these three factors give 6 kinds of yarn, which are the experimental treatments. From each yarn were prepared 9 warps (a warp is a quantity of warp yarn that goes into one loom as a unit), and, as a loom came available in the course of events, a warp chosen random from the 54 was assigned to it, until ultimately all 54 were disposed of. More than one warp was woven in some looms, but that did not upset the randomness of the distribution. The number of warp threads that broke during the waeving of each warp was counted and expressed as a rate of so many breaks per unit of warp."

We analyse the warp breakage rates for individual warps disregarding the factors. We fit the GEW model and other sub-models to these data by the method of maximum likelihood.

The MLEs of the parameters and the AIC (Akaike Information Criterion) measure for the models are listed in Table 7.1.

Table 7.1: MLEs of the model parameters for the warp breakage rates data (Tippett, 1950), the corresponding SEs (given in parentheses) and the AIC measure.

Model	α	λ	β	a	AIC
GEW	0.895137	0.000026	9.792369	23.609396	420.257020
	(0.042691)	(0.000011)	(0.714861)	(2.615331)	
GW	0	0.008219	1.622223	2.061746	425.080529
	(-)	(0.000971)	(0.025583)	(0.071981)	
EW	0.104458	0.000192	2.499730	1	427.270840
	(0.025452)	(0.000032)	(0.052776)	(-)	
W	0	0.000436	2.236797	1	427.436886
	(-)	(0.000047)	(0.029037)	(-)	
GEE	0.000961	0.190944	1	5.364537	421.924911
	(0.013883)	(0.026556)	(-)	(0.502296)	
EE	0.000100	0.035669	1	1	472.453098
	(0.035378)	(0.000717)	(-)	(-)	
GE	0	0.250448	1	6.896945	422.114535
	(-)	(0.006646)	(-)	(0.176429)	
E	0	0.035666	1	1	470.448845
	(-)	(0.00066)	(-)	(-)	

The plots of the fitted densities of all models are given in Figure 7.5. They indicate that the new distribution provides a better fit than the other sub-models. The required numerical evaluations were implemented by using a R script (sub-routine `nlminb` that can be found at `http://cran.r-project.org`). The data set `warpbreaks` is available as an R data frame.

A comparison of the new distribution with four of its sub-models using LR statistics is performed in Table 7.2. These statistics indicate that the new distribution is the most adequate model to explain the data.

Table 7.2: LR tests for the warp breakage rates data (Tippett, 1950).

Model	Hypotheses	Statistic LR	p-value
GEW vs GW	$H_0 : \alpha = 0$ vs $H_1 : H_0$ is false	6.823509	0.008996563
GEW vs EW	$H_0 : a = 1$ vs $H_1 : H_0$ is false	9.013820	0.002679457
GEW vs GEE	$H_0 : \beta = 1$ vs $H_1 : H_0$ is false	3.667891	0.055470343
GEW vs GE	$H_0 : \alpha = 0, \beta = 1$ vs $H_1 : H_0$ is false	5.857515	0.053463433

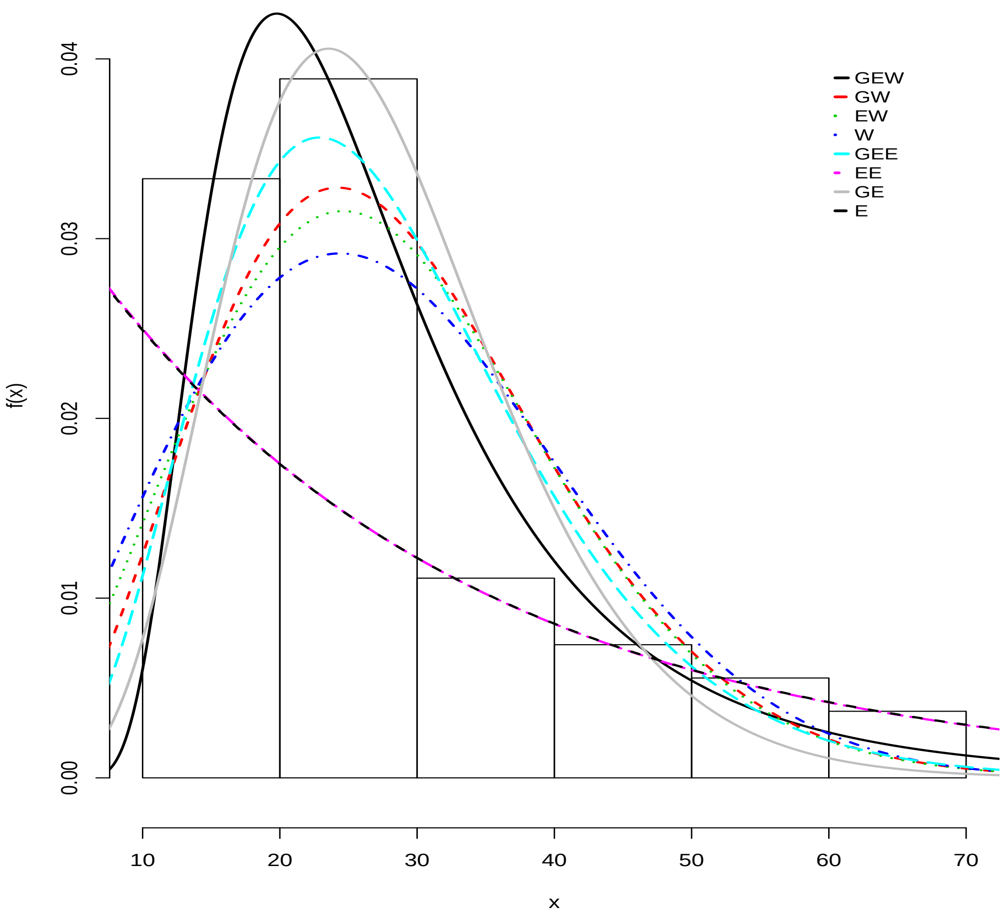


Figure 7.5: Plots of the GEW density and sub-models for the warp breakage rates data (Tippett, 1950).

7.11 Concluding remarks

We introduce a new model named the gamma extended Weibull (GEW) distribution and study some of its structural properties. It generalizes some important distributions in the literature and provides means of its continuous extension to still more complex situations. The new model contains several distributions as special models including the extended Weibull (Mudholkar *et al.*, 1996), gamma Weibull (Zografos and Balakrishnan, 2009) and generalized gamma (Stacy, 1962. We provide explicit expressions for the density function, ordinary and incomplete moments, generating and quantile functions, mean deviations, entropies and reliability. The model parameters are estimated by maximum likelihood. The usefulness of the new model is illustrated by means of an application to real data, where the GEW model provides a better fit than some of its submodels.

7.12 Appendix

Appendix A

We derive a power series for $Q_{GEW}(u)$ in the following way. First, we use a known power series for $Q^{-1}(a,1-u)$. Second, we obtain a power series for the argument $1-\exp[-Q^{-1}(a,1-u)]$. Third, we consider the EW qf to obtain a power series for $Q_{GEW}(u)$.

Let $Q^{-1}(a,z)$ be the inverse function of

$$Q(a,z) = 1 - \frac{\gamma(a,z)}{\Gamma(a)} = \frac{\Gamma(a,z)}{\Gamma(a)} = u.$$

The inverse function $Q^{-1}(a,1-u)$ is determined in the Wolfram website [1] as

$$\begin{aligned} Q^{-1}(a,1-u) &= w + \frac{w^2}{a+1} + \frac{(3a+5)w^3}{2(a+1)^2(a+2)} + \frac{[a(8a+33)+31]w^4}{3(a+1)^3(a+2)(a+3)} \\ &+ \frac{\{a(a[a(125a+1179)+3971]+5661)+2888\}\, w^5}{24(a+1)^4(a+2)^2(a+3)(a+4)} + O(w^6), \end{aligned}$$

where $w = [u\Gamma(a+1)]^{1/a}$. We can write the last equation as

$$z = Q^{-1}(a,1-u) = \sum_{i=0}^{\infty} a_i\, u^{i/a}, \tag{7.26}$$

where the a_i's are given by $a_i = \overline{b}_i\, \Gamma(a+1)^{i/a}$. Here, $\overline{b}_0 = 0$, $\overline{b}_1 = 1$ and any coefficient $\overline{b}_{i+1}$ (for $i \geq 1$) can be obtained from the cubic recurrence equation

$$\begin{aligned} \overline{b}_{i+1} &= \frac{1}{i(a+i)}\left\{ \sum_{r=1}^{i} \sum_{s=1}^{i-s+1} \overline{b}_r \overline{b}_s \overline{b}_{i-r-s+2}\, s\,(i-r-s+2) \times \right. \\ &\times \left. \sum_{r=2}^{i} \overline{b}_r \overline{b}_{i-r+2}\, r\,[r-a-(1-a)(i+2-r)] \right\}. \end{aligned}$$

The first coefficients are $\overline{b}_2 = 1/(a+1)$, $\overline{b}_3 = (3a+5)/[2(a+1)^2(a+2)]$, ... Now, we present some algebraic details to derive the GEW qf, say $Q_{GEW}(u)$. The EW qf is given by

$$Q_{EW}(u) = \left\{ \frac{1-(1-u)^{\alpha}}{\alpha\lambda(1-u)^{\alpha}} \right\}^{1/\beta}.$$

By replacing (7.26) in equation (7.11), we can write

$$Q_{GEW}(u) = \left\{ \frac{1-\{\exp[-Q^{-1}(a,1-u)]\}^{\alpha}}{\alpha\lambda\,\{\exp[-Q^{-1}(a,1-u)]\}^{\alpha}} \right\}^{1/\beta}. \tag{7.27}$$

[1] http://functions.wolfram.com/GammaBetaErf/InverseGammaRegularized/06/01/03/

By expanding the exponential function and using (2.7) and (7.26), we have

$$\begin{aligned}
\exp[-Q^{-1}(a,1-u)] &= \exp\left(-\sum_{i=0}^{\infty} a_i\, u^{i/a}\right) \\
&= \sum_{l=0}^{\infty}(-1)^{l+1}\,\frac{\left(\sum_{i=0}^{\infty} a_i\, u^{i/a}\right)^l}{l!} \\
&= -1+\sum_{i=0}^{\infty} h_i\, u^{i/a} = \sum_{i=0}^{\infty} p_i\, u^{i/a}, \qquad (7.28)
\end{aligned}$$

where $p_0 = -1 + h_0$, $p_i = h_i$ (i ≥ 1),h$_i = \sum_{l=1}^{\infty} \frac{(-1)^{l+1} f_{l,i}}{l!}$, $f_{l,i} = (i\, a_0)^{-1} \sum_{q=1}^{i} [q(l+1) - i]\, a_q\, f_{l,i-q}$ for $i \geq 1$ and $f_{l,0} = a_0^l$. Combining (7.27) and (7.28), we obtain

$$\begin{aligned}
Q_{GEW}(u) &= \left(\frac{1}{\alpha\lambda}\right)^{1/\beta}\left[\frac{\left(\sum_{i=0}^{\infty} p_i\, u^{i/a}\right)^{\alpha}}{1-\left(\sum_{i=0}^{\infty} p_i\, u^{i/a}\right)^{\alpha}}\right]^{-1/\beta} \\
&= \left(\frac{1}{\alpha\lambda}\right)^{1/\beta}\left(\frac{\sum_{i=0}^{\infty} \tau_i\, u^{i/a}}{1-\sum_{i=0}^{\infty} \tau_i\, u^{i/a}}\right)^{-1/\beta} \\
&= \left(\frac{1}{\alpha\lambda}\right)^{1/\beta}\left(\frac{\sum_{i=0}^{\infty} \tau_i\, u^{i/a}}{\sum_{i=0}^{\infty} \eta_i\, u^{i/a}}\right)^{-1/\beta},
\end{aligned}$$

where $\tau_i = \sum_{s=0}^{\infty} f_s\, \zeta_{s,i}$, $\zeta_{s,i} = (i\, p_0)^{-1} \sum_{m=0}^{i} [m(s+1) - i] p_m\, \zeta_{s,i-m}$, $f_s = \sum_{k=s}^{\infty} \frac{(-1)^{k-s} \binom{k}{s} (\alpha)_k}{k!}$, $(\alpha)_k$ is the descending factorial, $\eta_0 = 1 - \tau_0$ and $\eta_r = -\tau_r$ (r≥ 1).

From the quotient of the two power series, we have

$$\begin{aligned}
Q_{GEW}(u) &= \left(\frac{\eta_0}{\alpha\lambda}\right)^{1/\beta}\left(\sum_{i=0}^{\infty} \rho_i\, u^{i/a}\right)^{-1/\beta} \\
&= \left(\frac{\eta_0}{\alpha\lambda}\right)^{1/\beta}\sum_{i=0}^{\infty} \gamma_i\, u^{i/a} = \sum_{i=0}^{\infty} \gamma_i^{\star}\, u^{i/a}
\end{aligned}$$

where

$$\rho_i + \frac{1}{\eta_0}\sum_{i=1}^{r} \rho_{r-i}\, \eta_i - \tau_r = 0,$$

$\gamma_i^{\star} = \left(\frac{\eta_0}{\alpha\lambda}\right)^{1/\beta} \gamma_i$, $\gamma_i = \sum_{n=0}^{\infty} g_n \epsilon_{n,i}$, $\epsilon_{n,i} = (i\, \rho_0)^{-1} \sum_{m=0}^{i} [m(n+1) - i] \rho_m\, \epsilon_{n,i-m}$,
$g_n = \sum_{j=n}^{\infty} \frac{(-1)^{j-n} \binom{j}{n} (-1/\beta)_j}{j!}$ and $(-1/\beta)_j$ is descending factorial.

Appendix B

Let $J_{\theta_j\theta_k} = \partial^2 l(\boldsymbol{\theta})/\partial\theta_j\partial\theta_k = \sum_{i=1}^{n} \partial^2 \log[f(x_i)]/\partial\theta_j\partial\theta_k$, for $j,k = 1,\ldots,4$, be the elements of the information matrix $J(\boldsymbol{\theta})$, where θ_j and θ_k are substituted by the parameters α, β, λ or a. Let $l_i(\boldsymbol{\theta}) = \log[f(x_i)]$ and $q(x_i) = 1 + \alpha\lambda x_i^{\beta}$. The quantities $\partial^2 l_i(\boldsymbol{\theta})/\partial\theta_j\partial\theta_k$ are given by

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \alpha^2} = \frac{a-1}{\alpha^2} - \frac{2\log[q(x_i)]}{\alpha^3} + \frac{2\,\lambda\,x_i^{\beta}}{\alpha^2\,\log[q(x_i)]} + \frac{(\frac{1}{\alpha}+1)\,\lambda^2\,x_i^{2\beta}}{\{\log[q(x_i)]\}^2}$$
$$- \frac{(a-1)\,\lambda^2\,x_i^{2\beta}}{\{\log[q(x_i)]\}^2\,\log\{\log[q(x_i)]\}} - \frac{(a-1)\,\lambda^2\,x_i^{2\beta}}{\{\log[q(x_i)]\}^2\,\{\log\{\log[q(x_i)]\}\}^2},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \alpha \partial \beta} = \frac{\lambda\,x_i^{\beta}\,\log(x_i)}{\alpha\,q(x_i)} - \frac{(\frac{1}{\alpha}+1)\,\lambda\,x_i^{\beta}\,\log(x_i)}{q(x_i)} + \frac{(\frac{1}{\alpha}+1)\,\lambda^2\,x_i^{2\beta}\,\alpha\,\log(x_i)}{[q(x_i)]^2}$$
$$+ \frac{(a-1)\,\lambda\,x_i^{\beta}\,\log(x_i)}{q(x_i)\,\log[q(x_i)]} - \frac{(a-1)\,\lambda^2\,x_i^{2\beta}\,\alpha\,\log(x_i)}{[q(x_i)]^2\,\log[q(x_i)]}$$
$$- \frac{(a-1)\,\lambda^2\,x_i^{2\beta}\,\alpha\,\log(x_i)}{[q(x_i)]^2\,\{\log[q(x_i)]\}^2},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \alpha \partial \lambda} = \frac{x_i^{\beta}}{\alpha\,q(x_i)} - \frac{(\frac{1}{\alpha}+1)\,x_i^{\beta}}{q(x_i)} + \frac{(\frac{1}{\alpha}+1)\,\lambda\,x_i^{2\beta}\,\alpha}{[q(x_i)]^2} + \frac{(a-1)\,x_i^{\beta}}{q(x_i)\,\log[q(x_i)]}$$
$$- \frac{(a-1)\,\lambda\,x_i^{2\beta}\,\alpha}{[q(x_i)]^2\,\log[q(x_i)]} - \frac{(a-1)\,\lambda\,x_i^{2\beta}\,\alpha}{[q(x_i)]^2\,\{\log[q(x_i)]\}^2},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \alpha \partial a} = -\frac{1}{\alpha} + \frac{\lambda\,x_i^{\beta}}{q(x_i)\,\log[q(x_i)]},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \beta^2} = -\frac{1}{\beta^2} - \frac{(\frac{1}{\alpha}+1)\,\alpha\,\lambda\,x_i^{\beta}\,[\log(x_i)]^2}{q(x_i)} + \frac{(\frac{1}{\alpha}+1)\,\alpha^2\,\lambda^2\,x_i^{2\beta}\,[\log(x_i)]^2}{[q(x_i)]^2}$$
$$+ \frac{(a-1)\,\alpha\,\lambda\,x_i^{\beta}\,[\log(x_i)]^2}{q(x_i)\,\log[q(x_i)]} - \frac{(a-1)\,\alpha^2\,\lambda^2\,x_i^{2\beta}\,[\log(x_i)]^2}{[q(x_i)]^2\,\log[q(x_i)]}$$
$$- \frac{(a-1)\,\alpha^2\,\lambda^2\,x_i^{2\beta}\,[\log(x_i)]^2}{[q(x_i)]^2\,\{\log[q(x_i)]\}^2},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \beta \partial \lambda} = -\frac{(\frac{1}{\alpha}+1)\,\alpha\,x_i^{\beta}\,\log(x_i)}{q(x_i)} + \frac{(\frac{1}{\alpha}+1)\,\alpha^2\,\lambda\,x_i^{2\beta}\,\log(x_i)}{[q(x_i)]^2} + \frac{(a-1)\,\alpha\,x_i^{\beta}\,\log(x_i)}{q(x_i)\,\log[q(x_i)]}$$
$$- \frac{(a-1)\,\alpha^2\,\lambda\,x_i^{2\beta}\,\log(x_i)}{[q(x_i)]^2\,\log[q(x_i)]} - \frac{(a-1)\,\alpha^2\,\lambda\,x_i^{2\beta}\,\log(x_i)}{[q(x_i)]^2\,\{\log[q(x_i)]\}^2},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \beta \partial a} = \frac{\alpha\,\lambda\,x_i^{\beta}\,\log(x_i)}{q(x_i)\,\log[q(x_i)]},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \lambda^2} = -\frac{1}{\lambda^2} + \frac{(\frac{1}{\alpha}+1)\,\alpha^2\,x_i^{2\beta}}{[q(x_i)]^2} - \frac{(a-1)\,\alpha^2\,x_i^{2\beta}}{[q(x_i)]^2\,\log[q(x_i)]} - \frac{(a-1)\,\alpha^2\,x_i^{2\beta}}{[q(x_i)]^2\,\{\log[q(x_i)]\}^2},$$

$$\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \lambda \partial a} = \frac{\alpha\,x_i^{\beta}}{q(x_i)\,\log[q(x_i)]},\ \frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial a^2} = -\psi_1(a),$$

where $\psi_1(a) = d^2\log[\Gamma(a)]/da^2$ is the trigamma function.

References

ABRAMOWITZ, M.; STEGUN, I. A. *Handbook of Mathematical Funtions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, 1972.

CORDEIRO, G. M.; ORTEGA, E. M. M.; SILVA, G. O. The exponentiated generalized gamma distribution with application to lifetime data. *Journal of Statistical Computation and Simulation*, v. 81,p. 827–842, 2011.

GRADSHTEYN, I. S.; RYZHIK, I. M. *Table of Integrals, Series, and Products*. 6. ed. San Diego: Academic Press, 2000.

GUPTA, R.D.; KUNDU, D. Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, v. 41, p. 173–188, 1999.

GUPTA, R.D.; KUNDU, D. Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biometrical Journal*, v. 43, p. 117–130, 2001.

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

HAO, Z.; SINGH, V. P. Entropy-based parameter estimation for extended Burr XII distribution. *Stochastic Environmental Research and Risk Assessment*, v. 23, p. 1113–1122, 2008.

KOTZ, S.; LUMELSKII, Y.; PENSKY, M. *The Stress-Strength Model and its Generalizations: Theory and Applications*. New Jersey: World Scientific, 2003.

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.

MUDHOLKAR, 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.

MUDHOLKAR, G. S.; SRIVASTAVA, D. K.; Kollia, G. D. A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, v. 91, p. 1575–1583, 1996.

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

RISTIĆ, M. M.; BALAKRISHNAN, N. The gamma exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, v. 82, p. 1191–1206, 2012.

SHAO, Q.; WONG, H.; XIA, J.; WAI-CHEUNG, I. Models for extremes using the extended three-parameter Burr XII system with application to flood frequency analysis. *Hydrological Sciences Journal*, v. 49, p. 685–702, 2004.

STACY, E. W. A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, v. 33, p. 1187–1192, 1962.

TIPPETT, L. H. C. *Technological applications of statistics*. New York : John Wiley & Sons, 1950. p. 105–106

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