

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UMA NOVA CLASSE DE DISTRIBUIÇÕES E UM NOVO MODELO PARA
TEMPO DE VIDA: PROPRIEDADES E APLICAÇÕES

RODRIGO GONÇALVES DOS SANTOS

DISSERTAÇÃO DE MESTRADO

ORIENTADOR: PROF. DR. GAUSS MOUTINHO CORDEIRO

Recife
2016

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

UMA NOVA CLASSE DE DISTRIBUIÇÕES E UM NOVO MODELO PARA
TEMPO DE VIDA: PROPRIEDADES E APLICAÇÕES

RODRIGO GONÇALVES DOS SANTOS

ORIENTADOR: PROF. DR. GAUSS MOUTINHO CORDEIRO

Área de Concentração: Estatística Aplicada

Dissertação apresentada ao programa de Pós-graduação em Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de **Mestre em Estatística**.

Recife, 2016

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

S237n Santos, Rodrigo Gonçalves dos.
Uma nova classe de distribuições e um novo modelo para tempo de vida:
propriedades e aplicações / Rodrigo Gonçalves dos Santos. – 2016.
61 f.: il., fig., tab.

Orientador: Gauss Moutinho Cordeiro.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN.
Estatística, Recife, 2016.
Inclui referências e apêndices.

1. Estatística aplicada. 2. Distribuição de probabilidades. 3. Momentos ordinários. 4. Máxima verossimilhança. I. Cordeiro, Gauss Moutinho (orientador). II. Título.

310

CDD (23. ed.)

UFPE- MEI 2016-042

RODRIGO GONÇALVES DOS SANTOS

**UMA NOVA CLASSE DE DISTRIBUIÇÕES E UM NOVO MODELO PARA TEMPO
DE VIDA: PROPRIEDADES E APLICAÇÕES**

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

Aprovada em: 23 de fevereiro de 2016.

BANCA EXAMINADORA

Prof. PhD. Gauss Moutinho Cordeiro
UFPE

Prof.^a Dra. Audrey Helen Mariz de Aquino Cysneiros (Examinador Interno)
UFPE

Prof. Dr. Rodrigo Bernardo da Silva (Examinador Externo)
UFPB

Agradecimentos

Agradeço à Deus por ter aberto diversas portas ao longo de toda a minha vida.

Aos meus pais, José Francisco e Maria Edna, pelo apoio e carinhos incondicionais desde sempre. Ao meu irmão e amigo Rodolfo pelo apoio e convivência até seguirmos nossos caminhos.

Pela paciência e compreensão da minha amada esposa Maria das Graças, minha melhor amiga, sempre com conselhos e palavras importantes durante toda nossa caminhada.

Ao professor Gauss Cordeiro, pela liberdade, paciência e conselhos sempre úteis, que sem dúvida serão lembrados ao longo de toda minha vida acadêmica.

Aos colegas de mestrado, pelos momentos de descontração e conhecimento por nós trocados. Em especial, duas amigas: Fernanda Clotilde e Terezinha Késsia, por compartilharem tantas ideias mirabolantes comigo.

Aos participantes da banca examinadora, pelas sugestões.

À Valéria e Elaine pela paciência e eficiência na secretaria de pós-graduação.

*Bem-aventurado o homem que
acha sabedoria, e o homem que
adquire conhecimento;
Provérbios 3:13*

Resumo

Nas últimas décadas, diversos autores têm introduzido novas distribuições e em especial, novas classes de distribuições. O objetivo é obter melhoria nos ajustes de modelos aos diferentes tipos de dados. Algumas áreas importantes com esse objetivo são confiabilidade e análise de sobrevivência. Além disso, novas técnicas de otimização, como a meta-heurística, implementadas em software gratuitos como o R, conjuntamente com o avanços computacionais permitiram maior velocidade no tratamento de modelos com muitos parâmetros. Neste contexto, esta dissertação apresenta uma nova família de distribuição e um novo modelo com cinco parâmetros. A nova família foi baseada no gerador de Alzaatreh, usando como distribuição base a distribuição Burr XII. Propriedades gerais para esta família foram obtidas, além de duas aplicações a dados reais de duas distribuições pertencentes à família. O novo modelo para tempo de vida foi desenvolvido a partir da composição da distribuição taxa de falha linear generalizada na família exponencializada generalizada. Algumas propriedades do modelo, como momentos ordinários, incompletos e ponderados por probabilidade e função geratriz foram obtidas. Duas aplicações a dados reais foram realizadas mostrando o potencial da nova distribuição.

Palavras-chave: Gerador de Alzaatreh. Distribuição Burr XII. Família exponencializada generalizada. Distribuição taxa de falha linear generalizada. Estimação por máxima verossimilhança. Momentos.

Abstract

In recent decades, various authors have introduced new distributions and especially, new classes of distributions. The objective is to obtain best fitting models to the different types of data. Important areas the same objective are reliability and survival analysis. In addition, new optimization techniques, as meta-heuristics, implemented in free software as R, together with computational advances have allowed increased speed in the treatment of the models with many parameters. In this context, this thesis presents a new family of distributions and a new five-parameter model. The new family was based on the Alzaatreh's generator, using as baseline the Burr XII distribution. General properties were obtained for this family, as well as two applications to real data of two distributions belonging to the family. The new lifetime model was developed from the composition of the generalized linear failure rate distribution in the exponentiated generalized family. Some properties of the new model as ordinary, incomplete and probability weighted moments and generating function were obtained. Two applications to real data were performed showing the potential of the new distribution.

Keywords: Alzaatreh's generator. Burr XII distribution. Exponentiated generalized family. Generalized linear failure rate distribution. Maximum likelihood estimation. Moments.

Lista de Figuras

2.1	Algumas formas da função densidade da distribuição BXII-N(a, b, μ, σ). . .	27
2.2	Algumas formas da função densidade da distribuição BXII-W(a, b, β, λ). . .	28
2.3	Algumas formas da função densidade da distribuição BXII-Kw(a, b, α, β). . .	29
2.4	Modelos ajustados e histograma do conjunto de dados fibra de carbono. . .	34
2.5	Modelos ajustados e histograma do conjunto de dados rolamentos de esferas. . .	36
3.1	Algumas formas da fdp da distribuição EGLFR($a, b, \alpha, \beta, \theta$).	40
3.2	Algumas formas da ftf da distribuição EGLFR($a, b, \alpha, \beta, \theta$).	41
3.3	Densidades ajustadas e histograma do conjuntos e dados Aarset.	45
3.4	Gráfico Quantil-Quantil para os dados aircond.	46
3.5	Média da distribuição EGLFR($a, b, \alpha, \beta, \theta$).	49
3.6	Variância da distribuição EGLFR($a, b, \alpha, \beta, \theta$).	50

Lista de Tabelas

2.1	Estatísticas descritivas.	33
2.2	Modelos ajustados aos dados de fibra de carbono	33
2.3	Modelos ajustados ao conjunto de dados rolamento de esferas.	35
3.1	Estatísticas descritivas.	43
3.2	Modelos ajustados ao conjunto de dados Aarset.	44
3.3	A hipótese nula, valores da estatística LR, graus de liberdade e p-valores para os melhores ajustes casos especiais da distribuição EGLFR para o conjunto de dados Aarset.	44
3.4	Modelos ajustados para o conjunto de dados aircond.	46
3.5	Estatísticas Anderson Darling e Cramér-von Mises para os modelos ajustados ao conjunto de dados aircond.	46

Sumário

1	Introdução	12
1.1	Revisão bibliográfica	13
1.1.1	Conjunto de dados em análise de sobrevivência e confiabilidade . .	13
1.1.2	Algumas famílias de distribuições contínuas	13
1.1.2.1	Família T-X de distribuições	14
1.1.2.2	Família das exponencializadas	14
1.1.2.3	Família de distribuições beta	15
1.1.2.4	Família de distribuições Kumaraswamy	15
1.1.2.5	Família de distribuições McDonald	16
1.1.2.6	Família de distribuições beta generalizada refletida	17
1.1.2.7	Família de distribuições Weibull	17
1.1.2.8	Família de distribuições composta	17
1.1.3	Inferência	19
1.1.4	Metodologia de análise e seleção de modelos	20
2	Família Burr-XII G	23
2.1	Introdução	23
2.2	Família Burr XII-G	25
2.3	Distribuições especiais da Burr XII-G	26
2.3.1	Distribuição Burr XII-normal	26
2.3.2	Distribuição Burr XII-Weibull	27
2.3.3	Distribuição Burr XII-Kumaraswamy	28
2.4	Propriedades	29
2.4.1	Função Quantílica e simulação	29
2.4.2	Momentos	29
2.4.3	Função geratriz	30

2.4.4	Desvios médios	30
2.4.5	Estatísticas de ordem	31
2.5	Estimação por máxima verossimilhança	31
2.6	Aplicações	32
2.7	Conclusões	35
3	Distribuição EGFLR	37
3.1	Introdução	38
3.2	A distribuição EGLFR	39
3.2.1	Casos especiais da distribuição EGLFR	39
3.3	Estimação por máxima verossimilhança	41
3.4	Análise de dados reais	42
3.5	Propriedades matemáticas	45
3.5.1	Momentos e função geratriz	48
3.5.2	Função quantílica	48
3.5.3	Momentos ponderados por probabilidade	49
3.5.4	Momentos incompletos	49
3.5.5	Estatísticas de ordem	50
3.6	Considerações finais	51
4	Considerações finais	52
	Referências	53
A	Apêndice	58
A.1	Reescrevendo a cdf da Burr XII-G	59
A.2	Estatísticas de ordem da Burr XII-G	61

Capítulo 1

Introdução

Esta dissertação apresenta uma nova família de distribuições e uma nova distribuição de probabilidade. Os capítulos foram desenvolvidos em forma de artigo e portanto podem ser lidos de forma independente.

Na Seção 1.1, apresenta-se uma breve revisão bibliográfica dos temas abordados nos demais capítulos.

A família de distribuições Burr XII - G (SANTOS et al., 2015) é apresentada no Capítulo 2. A partir do gerador chamado de família T-X (ALZAATREH et al., 2013), usou-se a função densidade de probabilidade (fdp) da distribuição Burr XII (BXII) (BURR, 1942) considerando os parâmetros a e b e tomando $s = 1$. Ainda no gerador, o integrando superior foi a razão entre a função de distribuição acumulada (fda) pela sua função de sobrevivência. Diversas propriedades dessa nova família foram apresentadas.

Por meio de duas aplicações a dados reais, foram introduzidas as distribuições BXII-Weibull e BXII-exponencial, com aplicações, mostrando seu potencial prático. O primeiro conjunto de dados utilizado foi referente à resistência de fibras de carbono de 10 centímetros obtido em Lawless (2011). O tempo de falha de 23 rolamentos de esferas (LIEBLEIN; ZELEN, 1956) foi o segundo conjunto de dados utilizado.

No Capítulo 3, a distribuição taxa de falha linear generalizada estendida (SANTOS; CORDEIRO, 2015) é introduzida. A nova distribuição de cinco parâmetros é desenvolvida a partir da família de distribuições exponencializada generalizada (CORDEIRO et al., 2013), tomando por distribuição base a distribuição taxa de falha linear generalizada (SARHAN; KUNDU, 2009).

Além de algumas propriedades matemáticas, duas aplicações a dados reais foram realizadas. A primeira aplicação consistiu em tempos de falhas de 50 dispositivos (AARSET, 1987) e o segundo conjunto de dados utilizado foi referente ao número de falhas de um sistema de ar-condicionado em uma frota de aviões a jato Boeing 720 (PROSCHAN, 1963).

Finalmente, no Capítulo 4, as considerações finais gerais sobre os temas apresentados e também trabalhos futuros.

1.1 Revisão bibliográfica

Nesta seção, os temas abordados são acréscimos à leitura dos artigos nos capítulos seguintes. Inicialmente, uma breve revisão sobre a origem dos dados em análise de sobrevivência e confiabilidade é abordada. Em seguida, mostram-se expressões da fda e fdp de algumas famílias de distribuições e distribuições contínuas relacionadas aos capítulos seguintes. A metodologia de estimação pontual de parâmetros é abordada na sequência. E por último, a metodologia de análise e escolha de modelos é brevemente comentada.

1.1.1 Conjunto de dados em análise de sobrevivência e confiabilidade

Os conjuntos de dados estudados no presente trabalho apresentam suporte nos reais positivos. Não é incomum dados que representem o tempo até o acontecimento de um evento específico ou até um ponto de corte (COLLETT, 2003), sendo esta uma característica possível nesse tipo de conjunto de dados: a censura.

De maneira geral, a análise de sobrevivência se interessa no tempo até a ocorrência de um ou mais eventos. Por exemplo, o tempo até a cura ou morte de um paciente depois de diagnosticado sua doença. Caso seja um componente eletrônico ou máquina, uma análise semelhante pode ser feita. Neste caso, estuda-se a confiabilidade do equipamento, sendo possível conhecer um pouco mais sobre sua durabilidade. Este cenário é mais comum na área industrial.

Seja T a variável aleatória que representa algum destes tempos citados anteriormente. Denota-se sua fdp por $f_T(t)$ e sua fda por $F_T(t)$. Uma definição muito importante é a função de sobrevivência dada por

$$\bar{F}_T(t) = P(T \geq t) = 1 - F_T(t),$$

que representa a probabilidade de que elemento em estudo sobreviva pelo menos até o tempo t , $t > 0$.

Outra definição básica é a função taxa de falha (ftf) definida por

$$h_T(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} = \frac{f_T(t)}{\bar{F}_T(t)},$$

que pode apresentar-se como uma função constante, decrescente, crescente, em forma de U (ou banheira) e banheira invertida.

1.1.2 Algumas famílias de distribuições contínuas

Nesta seção serão apresentadas algumas famílias de distribuições, relatando também algumas distribuições desenvolvidas dentro de cada família, assim como suas aplicações.

1.1.2.1 Família T-X de distribuições

A família de geradores T-X definido por Alzaatreh et al. (2013) é de extrema relevância por englobar diversas famílias de distribuições. A fda desse gerador é expressa por

$$F(x) = \int_a^{H[G(x;\boldsymbol{\eta})]} r(t) dt = R\{H[G(x;\boldsymbol{\eta})]\}, \quad x \in \mathcal{D} \subseteq \mathbb{R} \quad (1.1)$$

em que $r(t)$ e $R(t)$ são fdp e fda, respectivamente, de uma determinada variável aleatória T , e $H[G(x;\boldsymbol{\eta})]$ é função de uma fda $G(x;\boldsymbol{\eta})$, associada a fdp $g(x;\boldsymbol{\eta})$ de uma variável aleatória X , com vetor de parâmetros $\boldsymbol{\eta}$. Ainda tem-se que $H[G(x;\boldsymbol{\eta})]$ satisfaz algumas condições, tais como

1. $H[G(x;\boldsymbol{\eta})] \in [a, b]$.
2. É diferenciável e monotonicamente não-decrescente.
3. $\lim_{x \rightarrow -\infty} H[G(x;\boldsymbol{\eta})] = a$ e $\lim_{x \rightarrow \infty} H[G(x;\boldsymbol{\eta})] = b$.

Ao longo dos últimos anos, algumas famílias de distribuições têm sido desenvolvidas por este método. Diversos outros métodos para gerar distribuições foram desenvolvidos, como por exemplo: o método de equações diferenciais - dentre os quais destacam-se o sistema de equações de Pearson (PEARSON, 1894) e de Burr (BURR, 1942); e o método de adicionar parâmetros a uma distribuição existente tem sido largamente utilizada para generalizar distribuições existentes, sendo que a família exponencializada tem papel fundamental, por sua simplicidade, ao adicionar um único parâmetro de potência a uma fda. Uma revisão mais aprofundada sobre diferentes métodos de gerar distribuições pode ser vista em Lee et al. (2013).

1.1.2.2 Família das exponencializadas

No início do século XIX, alguns autores já se utilizavam da fda $G(x) = (1 - \alpha e^{-\lambda x})^\beta$, $x > \lambda^{-1} \log(\alpha)$, $\alpha > 0$, $\lambda > 0$, $\beta > 0$ para comparar a tabelas de mortalidade humana (GOMPERTZ, 1825) e também para representar o crescimento populacional (VERHULST, 1838). Gupta et al. (1998) introduziram a distribuição exponencial exponencializada. Esse conceito da família das exponencializadas (exp-G) é definido adicionando um parâmetro como expoente, sendo sua fda dada por

$$F_{\text{exp-G}}(x) = [G(x;\boldsymbol{\eta})]^a, \quad x \in \mathcal{D} \subseteq \mathbb{R},$$

e por consequência, sua fdp é

$$f_{\text{exp-G}}(x) = a g(x;\boldsymbol{\eta}) [G(x;\boldsymbol{\eta})]^{a-1}$$

em que $a > 0$. É importante ressaltar que Mudholkar e Srivastava (1993) já haviam definido a distribuição Weibull exponencializada. Mudholkar et al. (1995) mostraram que a função taxa de falha dessa distribuição apresenta as formas crescente, decrescente, unimodal e banheira. Muitas outras distribuições exponencializadas foram desenvolvidas como Fréchet (NADARAJAH; KOTZ, 2003), Gumbel e gama (NADARAJAH; KOTZ, 2006b), por exemplo.

Esta família pode ser obtida do gerador definido em 1.1, considerando, por exemplo, $r(t)$ como sendo a fdp da distribuição uniforme padrão e $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})^a$. Algumas outras famílias de distribuições possuem a família das exponencializadas como uma sub-família. Quando $a = 1$, a distribuição resume-se a $G(x)$.

Pode-se encontrar na literatura a transformação Lehmann tipo-I, que é um sinônimo para a família das exponencializadas. A transformação Lehmann tipo-II (LEHMANN, 1953) exponencializa $1 - G(x; \boldsymbol{\eta})$, sendo a fda relacionada a transformação dada por $F_{\text{LeII-G}} = 1 - [1 - G(x; \boldsymbol{\eta})]^a$.

1.1.2.3 Família de distribuições beta

Caso $r(t)$ seja a fdp da distribuição beta e $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})$ em (1.1), a fda e fdp do gerador beta (EUGENE et al., 2002) são, respectivamente,

$$F_{\text{B-G}}(x) = \frac{B_{G(x; \boldsymbol{\eta})}(a, b)}{B(a, b)}, \quad x \in \mathfrak{D} \subseteq \mathbb{R}$$

e

$$f_{\text{B-G}}(x) = \frac{1}{B(a, b)} g(x; \boldsymbol{\eta}) G(x; \boldsymbol{\eta})^{a-1} [1 - G(x; \boldsymbol{\eta})]^{b-1},$$

em que $a > 0$, $b > 0$, $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ é a função beta e $B_x(a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$ é a função beta incompleta (JEFFREY; ZWILLINGER, 2007).

Alguns trabalhos recentes referentes a essa família de distribuição podem ser citados. Nadarajah e Kotz (2006a) introduziram a distribuição beta-exponencial. Barreto-Souza et al. (2010) apresentaram a distribuição beta exponencial generalizada. Lee et al. (2007) definiram a distribuição beta-Weibull, que se mostrou versátil e possui quatro formas da ftf (banheira, unimodal, crescente e decrescente).

A distribuição beta Weibull modificada (SILVA et al., 2010), com cinco parâmetros, além de grande flexibilidade de sua fdp e ftf (apresentando quatro formas), possui dezesseis casos especiais, como as distribuições beta Weibull, beta exponencial, exponencial exponencializada, Weibull, Rayleigh e exponencial, por exemplo.

1.1.2.4 Família de distribuições Kumaraswamy

Considerando como T a variável aleatória Kumaraswamy (KUMARASWAMY, 1980) e $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})$, tem-se a fda e fdp da família Kumaraswamy-G (Kw-G) expressas

por

$$F_{\text{Kw-G}}(x) = 1 - [1 - G(x; \boldsymbol{\eta})^a]^b, \quad x \in \mathfrak{D} \subseteq \mathbb{R}$$

e

$$f_{\text{Kw-G}}(x) = a b g(x; \boldsymbol{\eta}) G(x; \boldsymbol{\eta})^{a-1} [1 - G(x; \boldsymbol{\eta})^a]^{b-1},$$

respectivamente, em que $a > 0$ e $b > 0$. Cordeiro e Castro (2011) introduziram essa nova família, também conhecida como Kumaraswamy tipo-I - G. Como algumas distribuições desenvolvidas podem-se ser citadas as distribuição Kw-Weibull e Kw-exponencial (CORDEIRO et al., 2010). A distribuição Kw-gama generalizada (PASCOA et al., 2011), com cinco parâmetros, generaliza mais de dez distribuições, dentre elas as distribuições gama, gama generalizada, Weibull, exponencial e qui-quadrado. Cordeiro et al. (2012) introduziram a distribuição Kw-Gumbel, para $x \in \mathbb{R}$, com quatro parâmetros.

1.1.2.5 Família de distribuições McDonald

Tomando $r(t)$ como a fdp da distribuição beta generalizada (MCDONALD; XU, 1995), com $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})$ em (1.1), a fda e fdp da família McDonald-G (Mc-G) (ALEXANDER et al., 2012) são, respectivamente,

$$F_{\text{Mc-G}}(x) = \frac{B_{G(x; \boldsymbol{\eta})^c}(a, b)}{B(a, b)}, \quad x \in \mathfrak{D} \subseteq \mathbb{R}$$

e

$$f_{\text{Mc-G}}(x) = \frac{c}{B(a, b)} g(x; \boldsymbol{\eta}) G(x; \boldsymbol{\eta})^{ac-1} [1 - G(x; \boldsymbol{\eta})^c]^{b-1},$$

em que $a > 0$, $b > 0$ e $c > 0$. Quando $c = 1$ obtém-se a família beta-G. A família Kw-G também é obtida a partir desta família, quando $a = 1$. As transformações Lehmann tipo-I e tipo-II são obtidas quando $b = c = 1$ e $a = c = 1$, respectivamente, além da própria distribuição $G(x; \boldsymbol{\eta})$, quando $a = b = c = 1$. Cordeiro et al. (2014) definiram a distribuição Mc-Weibull, além do modelo de regressão para dados censurado log-McDonald Weibull.

Outra distribuição com suporte nos reais positivos escolhida como distribuição de base foi a distribuição beta invertida, com fda dada por $G(x; \boldsymbol{\eta}) = B_{\frac{x}{1-x}}(\alpha, \beta)/B(\alpha, \beta)$. Cordeiro e Lemonte (2012) propuseram a distribuição Mc-Beta Invertida, com cinco parâmetros.

E, por fim, a distribuição McDonald taxa de falha linear generalizada (ELBATAL et al., 2014) apresenta seis parâmetros e se mostra como uma distribuição muito flexível, além de possuir diversas distribuições como casos especiais, embora tenha um problema de identificabilidade. Tomando a distribuição de base como a fda da distribuição taxa de falha linear, este problema é corrigido. Uma segunda opção, mantendo a distribuição de base taxa de falha linear generalizada, é utilizar o gerador beta-G.

1.1.2.6 Família de distribuições beta generalizada refletida

Recentemente, a família beta generalizada refletida - G (RGB-G) foi definida com a seguinte fda

$$F_{\text{RGB-G}}(x) = 1 - \frac{B_{[1-G(x;\boldsymbol{\eta})]^c}(a, b)}{B(a, b)}, \quad x \in \mathfrak{D} \subseteq \mathbb{R}$$

e fdp

$$f_{\text{RGB-G}}(x) = \frac{c}{B(a, b)} g(x; \boldsymbol{\eta}) [1 - G(x; \boldsymbol{\eta})]^{ac-1} \{1 - [1 - G(x; \boldsymbol{\eta})]^c\}^{b-1},$$

em que $a > 0$, $b > 0$ e $c > 0$. Esta classe geradora de distribuições também é chamada de família beta-exponencial-G (beta-exp-G) (LEE et al., 2013). A distribuição de base $G(x; \boldsymbol{\eta})$ é um caso especial quando $a = b = c = 1$; Se $c = 1$, então tem-se a família beta-G; A classe de distribuições exponencializada generalizada (CORDEIRO et al., 2013) é um caso especial quando $a = 1$. Tomando como base a distribuição Weibull inversa, com $G_{\text{IW}}(x) = e^{-\gamma x^{-\theta}}$, Elbatal et al. (2013) propuseram a distribuição RGB-Weibull inversa, com cinco parâmetros.

1.1.2.7 Família de distribuições Weibull

Bourguignon et al. (2014) definiram a família Weibull-G (W-G), usando $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})/\overline{G}(x; \boldsymbol{\eta})$, em que $\overline{G}(x; \boldsymbol{\eta})$ é a função de sobrevivência, a partir da fdp da distribuição Weibull como $r(t)$. Sua fda e fdp são expressas, respectivamente, por

$$F_{\text{W-G}}(x) = 1 - \exp \left\{ -a \left[\frac{G(x; \boldsymbol{\eta})}{\overline{G}(x; \boldsymbol{\eta})} \right]^b \right\}, \quad x \in \mathfrak{D} \subseteq \mathbb{R}$$

e

$$f_{\text{W-G}}(x) = a b g(x; \boldsymbol{\eta}) \frac{G(x; \boldsymbol{\eta})^{b-1}}{[\overline{G}(x; \boldsymbol{\eta})]^{b+1}} \exp \left\{ -a \left[\frac{G(x; \boldsymbol{\eta})}{\overline{G}(x; \boldsymbol{\eta})} \right]^b \right\}.$$

em que $a > 0$ e $b > 0$. Se $b = 1$, tem-se o gerador exponencial. Os próprios autores da família introduziram a distribuição W-Weibull, com quatro parâmetros. Esta distribuição generaliza as distribuições exponencial poder, Chen e Gompertz. Novas distribuições como a distribuição W-exponencial, com três parâmetros, também surgem dessa nova distribuição.

Outras famílias relacionadas com a distribuição Weibull foram desenvolvidas. Por exemplo, Alzaatreh et al. (2013) definiram a família Weibull-X, utilizando $H[G(x; \boldsymbol{\eta})] = -\log[\overline{G}(x; \boldsymbol{\eta})]$ em 1.1. Já Tahir et al. (2014) utilizaram $H[G(x; \boldsymbol{\eta})] = -\log[G(x; \boldsymbol{\eta})]$ e por isso, usaram o complementar de 1.1, ou seja, sua fda é definida por $1 - R\{-\log[G(x; \boldsymbol{\eta})]\}$.

1.1.2.8 Família de distribuições composta

Outra maneira de gerar novas distribuições contínuas é obtida da composição de modelos discretos, como por exemplo, binomial-negativa, geométrica e série de potência, todas

truncadas em zero, já que existe uma interpretação física.

Considere, então, um subsistema com N elementos (com tempos de falha X_1, X_2, \dots, X_N) funcionando independentemente, em que cada subsistema possui a unidades funcionando em paralelo, fazendo com que ocorra uma falha se e somente se todos as a unidades virem a falhar. Além disso, as unidades de cada subsistemas possuem tempo de falha denotadas por $U_{i1}, U_{i2}, \dots, U_{ia}$, para o i -ésimo subsistema, são independentes e identicamente distribuídos com uma função exponencializada de uma fda $T[G(x; \boldsymbol{\eta})]^a$ para $x > 0$ e parâmetro potência a e vetor de parâmetros $\boldsymbol{\eta}$.

Para $Y = \min(X_1, X_2, \dots, X_N)$, a fda de Y dado N é definida como

$$F(y|N) = 1 - \{1 - T[G(y; \boldsymbol{\eta})]^a\}^N$$

e para $Z = \max(X_1, X_2, \dots, X_N)$, a fda de Z dado N é

$$F(z|N) = \{T[G(z; \boldsymbol{\eta})]^a\}^N.$$

Como são consideradas apenas variáveis aleatórias truncadas em zero para N , a fda incondicional de Y é

$$F(y) = 1 - \sum_{i=1}^{\infty} \{1 - T[G(y; \boldsymbol{\eta})]^a\}^i P(N = i) \quad (1.8)$$

e da mesma forma, a fda incondicional de Z é definida como

$$F(z) = \sum_{i=1}^{\infty} \{T[G(z; \boldsymbol{\eta})]^a\}^i P(N = i). \quad (1.9)$$

As equações 1.8 e 1.9, para diferentes variáveis aleatórias discretas truncadas em zero N e variáveis aleatórias $G(x, \boldsymbol{\eta})$ definem diversas classes de modelos compostos. Alguns modelos compostos foram estudados e são mencionados a seguir.

Ramos et al. (2015) definiram um novo modelo composto tomando N como Poisson(λ) truncada em zero, com função de probabilidade (fp) $P(N = n) = \lambda^n / [n! (e^\lambda - 1)]$, para $n = 1, 2, \dots$ e o tempo de falha de cada subsistema, denotado por X_i , tendo distribuição Kw-G. Assim, a fda incondicional de $Y = \min(X_1, X_2, \dots, X_N)$, ou seja, o tempo do primeiro subsistema a falhar é definido por

$$F(y) = \frac{1 - e^{-\lambda F_{\text{Kw-G}}(y)}}{1 - e^{-\lambda}},$$

chamado de família de distribuições Kumaraswamy-G Poisson [Kw-GP($a, b, \boldsymbol{\eta}, \lambda$)].

No caso de $T[G(x; \boldsymbol{\eta})] = F_{\text{exp-G}}(x)$ e N tendo distribuição geométrica(p) truncada em zero, com fp $P[N = n] = (1 - p)p^n$, para $n = 1, 2, \dots$, Nadarajah et al. (2015) definiram a classe de distribuições chamada exponencializada G-geométrica tipo 1 [EGG1($a, \boldsymbol{\eta}, p$)], que tem fda incondicional de $Y = \min(X_1, X_2, \dots, X_N)$ definida como

$$F(y) = \frac{F_{\text{exp-G}}(y)}{1 - p + p F_{\text{exp-G}}(y)}.$$

Para $a = 1$, a classe de distribuições EGG1 se resume a classe de distribuição G-geométrica tipo 1 [GG1($\boldsymbol{\tau}, p$)] estudada por Alkarni (2013).

Gomes et al. (2015) definiram uma nova classe de distribuições composta tomando $T[G(x; \boldsymbol{\eta})] = F_{\text{exp-G}}(x)$ e N com distribuição Poisson(λ) truncada em zero. A fda da família de distribuições exponencializadas G-Poisson [EGP($a, \boldsymbol{\tau}, \lambda$)] é dada por

$$F_{EGP}(y) = \frac{1 - e^{-\lambda F_{\text{exp-G}}(y)}}{1 - e^{-\lambda}}.$$

Para $a = 1$, temos a família de distribuições G-Poisson (GP), estudada por Tahir et al. (2016).

1.1.3 Inferência

Na presente dissertação, o método de estimação dos parâmetros de todas as distribuições foi o método da máxima verossimilhança. Este método de estimação possui algumas propriedades muito úteis quando o tamanho da amostra cresce como consistência, eficiência, normalidade assintótica, se atendidas algumas condições de regularidade.

Considere que $\mathbf{x} = (x_1, \dots, x_n)$ é uma amostra aleatória observada de tamanho n de determinada população com fdp $f(x; \theta)$, sendo θ um vetor de parâmetros p -dimensional. Assim, a função de verossimilhança é definida como

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta).$$

Seja, para \mathbf{x} fixo, $\hat{\theta} = \hat{\theta}(\mathbf{x})$ o conjunto de valores dos parâmetros que faça com que $L(\theta; \mathbf{x})$ atinja o seu máximo. Então, $\hat{\theta}$ é conhecido como a estimativa de máxima verossimilhança (MLE). Em geral, para obter $\hat{\theta}$ é preferível maximizar a função de log-verossimilhança

$$l(\theta; \mathbf{x}) = \sum_{i=1}^n \log [f(x_i; \theta)], \quad (1.10)$$

considerando que também é possível diferenciá-la com relação a θ e $\hat{\theta}$ é tal que seja solução do sistema linear de p equações

$$U(\theta) = \frac{\partial}{\partial \theta} l(\theta; \mathbf{x}) = 0, \quad (1.11)$$

sendo $U(\theta)$ a função score. Como a função log é monótona crescente, maximizar $L(\theta; \mathbf{x})$ ou $l(\theta; \mathbf{x})$ com relação a θ é equivalente.

Nesse problema de maximização, o sistema linear apresentado quase sempre é não linear. Por esse motivo, buscam-se algoritmos de otimização. Por exemplo, generalized simulated Annealing (GenSA) (XIANG et al., 2013), genetic algorithms (GA) (SCRUCCA, 2013) e maximum likelihood estimation (maxLik) (HENNINGSEN; TOOMET, 2011) são

pacotes do software R (R CORE TEAM, 2015) para executar a maximização de funções.

As condições de regularidades usuais que supõe-se atendidas pela função de log-verossimilhança são: o espaço paramétrico ζ é aberto e $l(\zeta)$ tem máximo global em ζ ; Para todo x , as derivadas de quarta ordem de $l(\zeta)$ com relação aos parâmetros do modelo existem e são contínuas em um subconjunto aberto de ζ que contém o verdadeiro valor dos parâmetros; O suporte da variável aleatória X não depende de parâmetros desconhecidos; A matriz de informação esperada é finita e positiva definida. Essas condições são não restritivas e supõe-se para todas as distribuições citadas.

Intervalos de confiança e testes de hipóteses podem ser feitos levando-se em consideração a normalidade assintótica dos estimadores de máxima verossimilhança. Esses estimadores possuem normalidade com média θ e variância dada por K^{-1} , em que K é a matriz de informação de Fisher.

1.1.4 Metodologia de análise e seleção de modelos

Ao ajustar modelos ao mesmo conjunto de dados será necessário definir qual dos modelos está melhor ajustado. Em uma breve revisão da literatura, verificam-se diversas metodologias, cada qual com seus destaques e indicações de uso.

Seja ω_0 o processo desconhecido gerador das n observações disponíveis no vetor \mathbf{x} . Deseja-se saber, através de um modelo conhecido e ajustado às observações, $w(\mathbf{x}|\theta)$, qual a proximidade referente ao processo desconhecido ω_0 . Seja também $L = L(\hat{\theta}|\mathbf{x})$ em que $\hat{\theta}$ é a MLE de θ .

Akaike (1998) definiu um critério para escolha de modelos. Neste critério, definido por

$$AIC_w = 2 p - 2 \log(L),$$

AIC_w quantifica o quanto próximo do processo gerador real está o modelo $w(\mathbf{x}|\theta)$.

Portanto, entre dois modelos w_1 e w_2 , aquele que apresenta o menor valor do critério de informação de Akaike (AIC) é o modelo mais apropriado. Note que, em nenhum momento, existe a garantia de bom ajuste no modelo escolhido. Este critério é apenas uma ferramenta de escolha entre modelos.

Outro critério que pode auxiliar na escolha de um modelo é definido por Hannan e Quinn (1979). Embora tenha sido desenvolvido originalmente para seleção de modelos autorregressivos, tornou-se uma alternativa de seleção de modelos para modelos com outras estruturas de dados. Este critério é definido por

$$HQIC_w = 2 p \log[\log(n)] - 2 \log(L).$$

Da mesma forma que o critério anterior, o modelo que apresentar menor valor critério de informação de Hannan e Quinn (HQIC) deve ser escolhido e, novamente, assim como

o AIC, não existe garantia do bom ajuste do modelo. Este critério apresenta em sua expressão, diferentemente do AIC, n , o tamanho da amostra.

Para garantir a bondade do ajuste deve-se recorrer à métodos mais robustos, como os que serão apresentados a seguir. Nestes métodos estão sendo feitos testes estatísticos, baseados nas fdas escolhidas para a modelagem dos dados, diferentemente dos critérios anteriormente apresentados.

Esses métodos são baseados na função $F_n(x)$, que é a função de distribuição empírica (fde). Considere um amostra de tamanho n . A fde é definida como

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{i}{n}, & x_{(i)} \leq x < x_{(i+1)}, \quad i = 1, \dots, n-1 \\ 1, & x_{(n)} \geq x \end{cases} \quad (1.12)$$

em que $x_{(i)}$ é representa a i -ésima estatística de ordem. $F_n(x)$ é uma função escada.

D'Agostino (1986) cita duas classes de estatísticas baseadas na diferença vertical entre a fde e a fda. Na primeira delas, baseadas no supremo, ou seja, nos máximos de $|F_n(x) - F(x)|$, $|\cdot|$ indicando a função modular, tem-se, por exemplo, a estatística de Kolmogorov-Smirnov. As estatísticas na segunda classe são baseadas na seguinte forma

$$Q = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(x) dF(x),$$

em que $\psi(x)$ representa uma função de peso para a forma quadrática. Nesta classe leva-se em consideração todas as diferenças entre a fda e fde. Nos casos em que $\psi(x) = 1$ e $\psi(x) = \{F(x)[1 - F(x)]\}^{-1}$, tem-se as estatísticas de Cramér-von-Mises e Anderson-Darling, respectivamente. Ambas as estatísticas estão definidas em Chen e Balakrishnan (1995) como

$$W^* = \left\{ n \int_{-\infty}^{\infty} [F_n(x) - \hat{F}(x)]^2 d\hat{F}(x) \right\} \left(1 + \frac{0.5}{n} \right)$$

$$A^* = \left\{ n \int_{-\infty}^{\infty} \frac{[F_n(x) - \hat{F}(x)]^2}{\hat{F}(x)[1 - \hat{F}(x)]} d\hat{F}(x) \right\} \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2} \right)$$

em que $\hat{F}(x)$ é a fda avaliada na MLE $\hat{\phi}_n$ do vetor de parâmetros ϕ . O teste de hipótese formulado tem como hipótese nula a adequação dos dados à fda proposta. Para ambas as estatísticas, valores pequenos são evidências a favor da hipótese nula, o que representa contribuições pequenas das formas quadráticas nas integrais anteriores.

Os valores críticos para executar ambos os testes é de difícil obtenção analítica. Entretanto, podem ser obtidos através das funções qCvM e qAD do pacote goftest (FARAWAY et al., 2015) no software R. Pela construção das estatísticas, parece intuitivo que quanto menor o valor em ambas, melhor ajustado está o modelo.

Para comparar dois ou mais modelos que são ditos encaixados pode-se realizar o teste da razão de verossimilhanças (CASELLA; BERGER, 2002; MOOD et al., 1974). Considere um modelo com vetor de parâmetros $\Theta = (\theta_1, \dots, \theta_p)^T$, p -dimensional, e um sub-modelo com vetor de parâmetros Θ_0 , também p -dimensional, mas com k parâmetros conhecidos, $1 \leq k \leq p$. Esse teste tem por hipótese nula k igualdades propostas em Θ_0 , ou seja,

$$H_0 : \theta_1 = \theta_1^0, \dots, \theta_k = \theta_k^0, \theta_{k+1}, \dots, \theta_p,$$

em que $\theta_1^0, \dots, \theta_k^0$ são conhecidos e $\theta_{k+1}, \dots, \theta_p$ são desconhecidos. A estatística de teste é definida por $-2 \log(\lambda)$ sendo

$$\lambda = \frac{\sup_{\Theta_0} L(\Theta; \mathbf{x})}{\sup_{\Theta} L(\Theta; \mathbf{x})}.$$

Seja Λ a variável aleatória que assume valores λ . Então, $-2 \log(\Lambda)$ possui distribuição assintótica χ_k^2 . Adotado um nível de significância $100 \alpha\%$, rejeita-se a hipótese nula quando $-2 \log(\lambda) > q_{1-\alpha}$, sendo $q_{1-\alpha}$ quantil da distribuição χ_k^2 que possui área α após este.

Apesar de todas as ferramentas apresentadas, a análise de gráficos persiste importante para a escolha de modelos. Dentre os principais gráficos que podem ser apresentados para essa escolha, tem-se o histograma dos dados e os modelos ajustados e o gráfico quantil-quantil, por exemplo. A partir destes gráficos e com as estatísticas e critérios apresentados, será possível determinar qual dos modelos propostos está melhor ajustado aos dados.

Capítulo 2

Família de distribuições Burr XII - G: propriedades e aplicações

Resumo

Uma nova família de distribuições foi estudada usando o gerador de Alzaatreh (ALZAATREH et al., 2013) baseado na distribuição Burr XII (BURR, 1942), que estende algumas bem conhecidas classes de distribuições. Algumas propriedades da nova família são obtidas. Duas aplicações a dados reais comparam o desempenho da nova família com outras conhecidas classes (SANTOS et al., 2015).

Palavras-Chaves: Distribuição Burr XII. Família das exponencializadas. Momentos. Família T-X.

Abstract

We study a new family of distributions using the Alzaatreh's generator (ALZAATREH et al., 2013) based on the Burr XII distribution (BURR, 1942), which extends some well known classes of distributions. Some properties of the new family are obtained. Two applications to real data compare the performance of the new family with other known classes (SANTOS et al., 2015).

keywords: Burr XII distribution; Exponentialized family; Moments; T-X family.

2.1 Introdução

Nos últimos anos, diversos autores exploraram novos geradores de distribuições para analisar dados reais. Importantes famílias foram obtidas com aplicações a dados para tempo de vida. O gerador de novas distribuições é importante porque novos tipos de dados

de sobrevivência aparecem com função taxa de falha em que as distribuições conhecidas não se adequam.

Alzaatreh et al. (2013) propuseram uma família de distribuições contínuas, chamada de família T-X, que tornou-se muito útil. Tahir et al. (2015) forneceram alguns exemplos, tomando funções especiais de $H[G(x)]$. Dentre esses modelos pode-se citar: $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})$ e $T \in [0, 1]$ que refere-se a beta-G (EUGENE et al., 2002), Kumaraswamy tipo-I - G (CORDEIRO; CASTRO, 2011), McDonald-G (ALEXANDER et al., 2012) e exponencializada generalizada-G (EG-G ou Kumaraswamy tipo-II - G) (CORDEIRO et al., 2013). Uma nova família de distribuições contínuas é proposta considerando a função densidade de probabilidade (fdp) da Burr XII (BXII) e $H[G(x; \boldsymbol{\eta})] = G(x; \boldsymbol{\eta})/\bar{G}(x; \boldsymbol{\eta})$, sendo $\bar{G}(x; \boldsymbol{\eta}) = 1 - G(x; \boldsymbol{\eta})$.

O sistema Burr de distribuições foi proposto em 1942 por Irving W. Burr (BURR, 1942). A fdp da variável aleatória T seguindo a distribuição BXII é dada por

$$r(t; s, a, b) = \left(\frac{ab}{s^a}\right) t^{a-1} \left[1 + \left(\frac{t}{s}\right)^a\right]^{-b-1}, \quad t > 0,$$

em que $a > 0$ e $b > 0$ são parâmetros de forma e $s > 0$ é parâmetro de escala. A função de distribuição acumulada (fda) é dada por

$$R(t; s, a, b) = 1 - \left[1 + \left(\frac{t}{s}\right)^a\right]^{-b}, \quad t > 0. \quad (2.1)$$

Sem perda de generalidade, considerou-se o modelo BXII com $s = 1$. A distribuição BXII tem como casos especiais as distribuições log-logística e Weibull. Uma característica importante da distribuição BXII é que sua função taxa de falha (hrf) pode ser unimodal ou monótona, o que fornece uma alternativa para a distribuição log-logística. Adicionalmente, a distribuição BXII tem algumas vantagens por ter expressão em forma fechada para as funções de sobrevivência e taxa de falha. Paranaíba et al. (2011) construiu uma distribuição baseada no modelo BXII usando o gerador beta chamado distribuição beta Burr XII (BBXII). Paranaíba et al. (2013) estudou uma alternativa para o modelo BBXII chamado distribuição Kumaraswamy Burr XII (KwBXII).

Este capítulo propõe uma nova família de distribuições baseada na distribuição BXII chamada família Burr XII - G. O capítulo está delineado como segue. Na Seção 2.2, baseado no trabalho de Alzaatreh, define-se a família Burr XII - G e mostra-se a expressão da fdp da família como uma mistura de densidades exponencializada-G (exp-G). São apresentados na Seção 2.3 alguns casos especiais da família proposta, tomando as distribuições normal, Weibull e Kumaraswamy como modelos de base. Algumas propriedades matemáticas da família BXII-G são obtidas na Seção 2.4. Expressões em forma fechada foram derivadas para a função quantílica (fq), momentos, função geratriz de momentos (fgm) e desvios

médios, além de fornecer a fdp das estatísticas de ordem também como uma mistura de densidades exp-G. Na seção 2.5, foram determinados os componentes da função escore e a forma de obtenção das estimativas de máxima verossimilhança (EMVs) dos parâmetros do modelo. Finalmente, duas aplicações a conjuntos de dados reais são apresentados na seção 2.6, usando as distribuições Burr XII - exponencial e Burr XII - Weibull.

2.2 Família Burr XII-G

Como discutido na seção anterior, considerou-se a família T-X com $H[G(x)] = G(x)/\overline{G}(x)$. Assim, foi assumido que a variável aleatória T tem distribuição Burr XII e define a família Burr XII-G (BXII-G).

Esta família tem fda e fdp dada por

$$F(x; \theta) = \int_0^{\frac{G(x; \boldsymbol{\eta})}{\overline{G}(x; \boldsymbol{\eta})}} a b t^{a-1} (1 + t^a)^{-b-1} dt = 1 - \left\{ 1 + \left[\frac{G(x; \boldsymbol{\eta})}{\overline{G}(x; \boldsymbol{\eta})} \right]^a \right\}^{-b} \quad (2.2)$$

e

$$f(x; \theta) = a b g(x; \boldsymbol{\eta}) \frac{G(x; \boldsymbol{\eta})^{a-1}}{\overline{G}(x; \boldsymbol{\eta})^{a+1}} \left\{ 1 + \left[\frac{G(x; \boldsymbol{\eta})}{\overline{G}(x; \boldsymbol{\eta})} \right]^a \right\}^{-b-1}, \quad (2.3)$$

respectivamente, em que $\theta = (a, b, \boldsymbol{\eta})$, e $\boldsymbol{\eta}$ é o vetor de parâmetros relacionado à distribuição de base G.

Da equação (2.2), tomando $a = 1$, teremos

$$F(x; \theta) = 1 - [1 - G(x; \boldsymbol{\eta})]^b.$$

Esta equação revela que a família BXII-G(1, b, $\boldsymbol{\eta}$) é idêntica à família Lehmann tipo-II($\alpha, \boldsymbol{\eta}$) (LEHMANN, 1953).

Outra família obtida acontece quando $b = 1$ em (2.2) fornecendo

$$F(x; \theta) = \frac{G(x; \boldsymbol{\eta})^a}{\overline{G}(x; \boldsymbol{\eta})^a + G(x; \boldsymbol{\eta})^a}.$$

Esta fda é idêntica à definida em Cruz et al. (2013) chamada família odd log-logistic-G (OLL-G). Então, as famílias BXII($a, 1, \boldsymbol{\eta}$) e OLL-G($a, \boldsymbol{\eta}$) são idênticas.

E o último caso especial que pode ser derivado das duas últimas fdas se refere à distribuição de base $G(x) = G(x; \boldsymbol{\eta})$ que aparece quando $a = b = 1$ na família BXII-G.

A hrf correspondente às equações (2.2) e (2.3) é dada por

$$h(x; \theta) = a b \left[1 + \left(\frac{G(x; \boldsymbol{\eta})}{\overline{G}(x; \boldsymbol{\eta})} \right)^a \right]^{-1} \frac{G(x; \boldsymbol{\eta})^{a-1}}{\overline{G}(x; \boldsymbol{\eta})^{a+1}} g(x; \boldsymbol{\eta}).$$

Pode-se escrever a equação (2.2) como uma série de potências infinita de $G(x)$, reescrita como

$$F(x; \theta) = \sum_{i=0}^{\infty} \nu_i H_i(x), \quad (2.4)$$

em que $H_i(x) = H_i(x; \boldsymbol{\eta}) = G(x; \boldsymbol{\eta})^i$. No Apêndice A.1 estão descritos os passos para a obtenção dessa mistura e também a estrutura da constante $\nu_i = \nu_i(a, b)$. A fdp de X pode ser reescrita como uma mistura de funções de densidades exp-G

$$f(x; \theta) = \sum_{i=0}^{\infty} \nu_{i+1} h_{i+1}(x), \quad (2.5)$$

em que $h_{i+1}(x) = h_{i+1}(x; \eta) = (i + 1) g(x; \eta) G(x; \eta)^i$ é a função de densidade da exp-G com parâmetro potência $i + 1$. Algumas propriedades estruturais da distribuição exp-G foram estudadas por Gupta et al. (1998), Gupta e Kundu (2001), Mudholkar e Srivastava (1993), Mudholkar et al. (1995) e Nadarajah e Kotz (2006b), entre outros. Assim, algumas propriedades matemáticas da família proposta podem ser derivadas de (2.5) e estas propriedades da distribuição exp-G. Por exemplo, os momentos ordinários e incompletos, fgm de X serão determinadas por (2.5) e as quantidades correspondentes à exp-G.

2.3 Distribuições especiais da Burr XII-G

A família BXII-G pode ser aplicada em diversas áreas dependendo da escolha da distribuição de base $G(x)$. Se os dados apresentam suporte positivo, como tempos de falha, pode-se tomar $G(x)$ como a fda das distribuições Weibull ou exponencial. Se o interesse está em um modelo de proporção, as fdas das distribuições beta e Kumaraswamy poderiam ser boas escolhas para distribuições de base. Contudo, se os dados se apresentam na reta dos reais, as fdas das distribuições normal e logística contemplam esta variação.

Nesta seção, serão apresentados e estudados alguns casos especiais desta família. A densidade terá expressão de fácil manuseio quando $G(x)$ e $g(x)$ tiverem expressões analíticas simples.

2.3.1 Distribuição Burr XII-normal

A distribuição Burr XII - Normal (BXII-N) é definida de (2.2) tomando como distribuição de base $G(x)$ a distribuição normal com parâmetros $\mu \in \mathbb{R}$ e $\sigma > 0$. A fda da distribuição BXII-N(θ), em que $\theta = (a, b, \mu, \sigma^2)$, é definida por (para $x \in \mathbb{R}$)

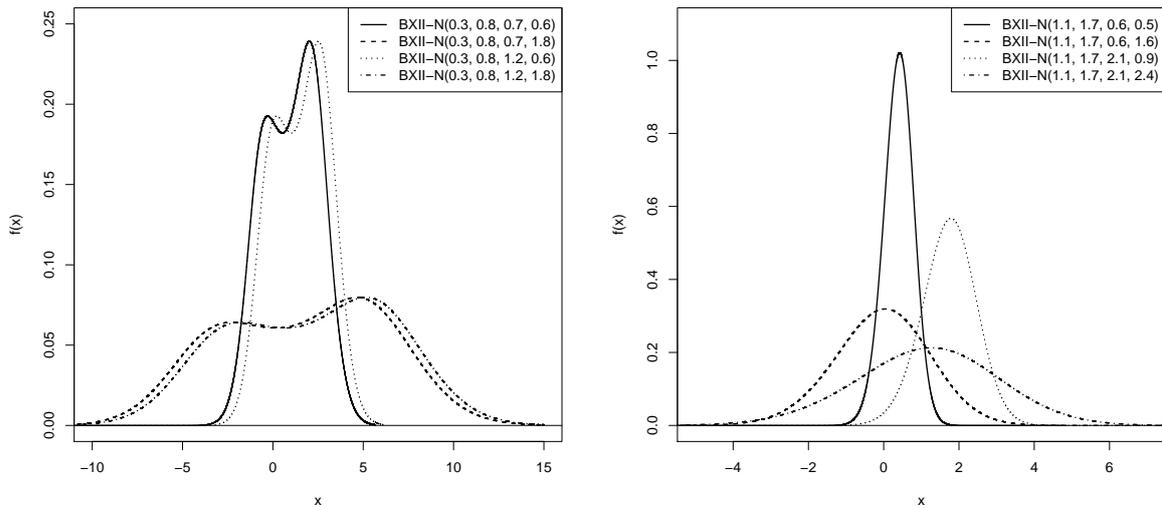
$$F(x; \theta) = 1 - \left[1 + \left(\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x-\mu}{\sigma}\right)} \right)^a \right]^{-b},$$

em que μ é um parâmetro de locação, σ é um parâmetro de escala, a e b são parâmetros de forma e $\Phi(\cdot)$ representa a fda da distribuição normal padrão. Seja $\phi(\cdot)$ a fdp da distribuição normal padrão. Baseado em (2.3), a fdp do modelo BXII-N segue como

$$f(x; \theta) = a b \phi\left(\frac{x - \mu}{\sigma}\right) \frac{\Phi\left(\frac{x-\mu}{\sigma}\right)^{a-1}}{\left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{a+1}} \left\{ 1 + \left[\frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x-\mu}{\sigma}\right)} \right]^a \right\}^{-b-1}. \quad (2.6)$$

Se $\mu = 0$ e $\sigma = 1$, o modelo BXII-N padrão $(a, b, 0, 1)$ é obtido. Gráficos de (2.6) para valores de parâmetros selecionados são mostrados na Figura 2.1, em que as formas bimodais se destacam.

Figura 2.1: Algumas formas da função densidade da distribuição BXII-N(a, b, μ, σ).



Fonte: Próprio autor.

2.3.2 Distribuição Burr XII-Weibull

O modelo Burr XII - Weibull (BXII-W) pode ser usado em análise de sobrevivência. Neste caso, tomando a fda da distribuição Weibull como $G(x)$, a fda e fdp (para $x > 0$) do modelo BXII-W(θ), em que $\theta = (a, b, \beta, \lambda)$, são dados por

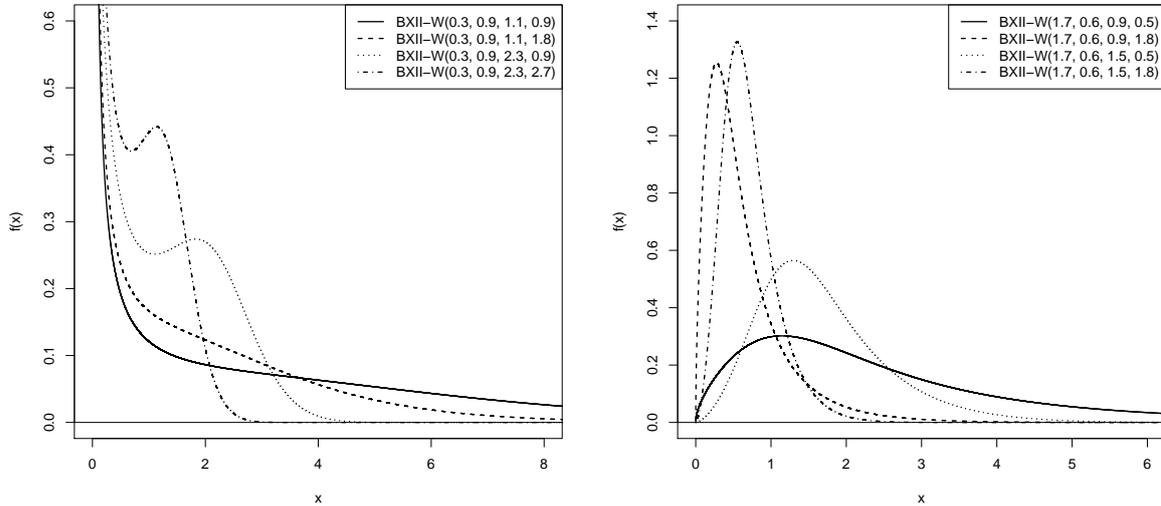
$$F(x; \theta) = 1 - \left[1 + \left(e^{-\lambda x^\beta} - 1 \right)^a \right]^{-b}$$

e

$$f(x; \theta) = a b \lambda \beta x^{\beta-1} e^{\lambda x^\beta} (e^{\lambda x^\beta} - 1)^{a-1} \left[1 + (e^{\lambda x^\beta} - 1)^a \right]^{-1-b}, \quad (2.7)$$

respectivamente. Um caso especial do modelo BXII-W é obtido quando $\beta = 1$ e então tem-se o modelo Burr XII - exponencial (BXII-E). Gráficos de (2.7) para alguns valores de parâmetros podem ser visualizados na Figura 2.2. Notam-se formas unimodais e decrescentes para a fdp, além da forma decrescente-crescente-decrescente.

Figura 2.2: Algumas formas da função densidade da distribuição $BXII-W(a, b, \beta, \lambda)$.



Fonte: Próprio autor.

2.3.3 Distribuição Burr XII-Kumaraswamy

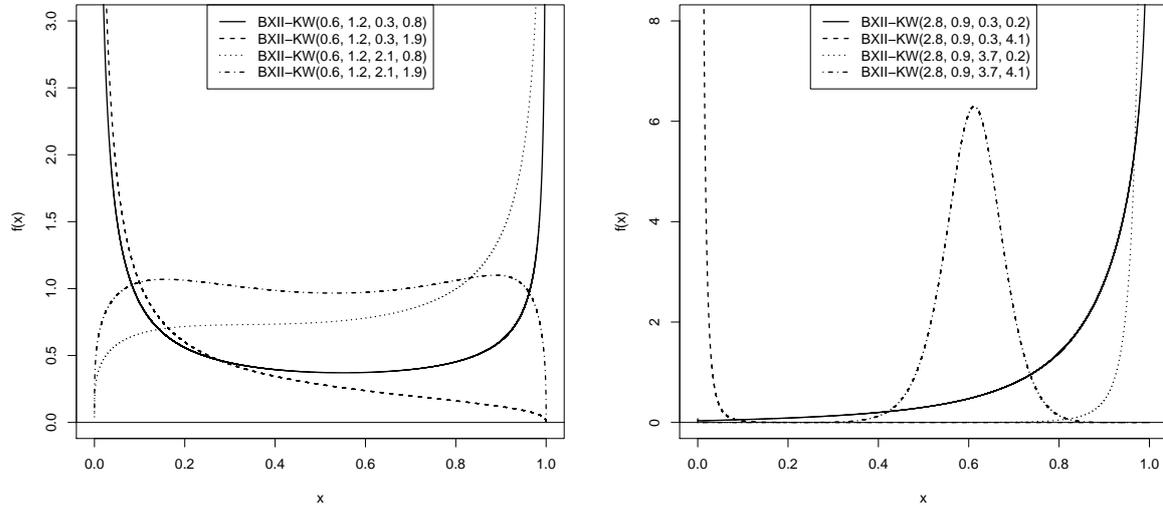
A distribuição Kumaraswamy é um bom modelo para ajustar dados no intervalo $(0, 1)$. O modelo Burr XII - Kumaraswamy (BXII-Kw) com vetor de parâmetros $\theta = (a, b, \alpha, \beta)$ tem fda dada por

$$F(x; \theta) = 1 - \{1 + [(1 - x^\alpha)^{-\beta} - 1]^a\}^{-b}, \quad 0 < x < 1.$$

Sua fdp se resume em

$$f(x; \theta) = \alpha \beta b x^{\alpha-1} (1 - x^\alpha)^{-a\beta-1} \{(1 - x^\alpha)^{-a\beta} [1 - (1 - x^\alpha)^\beta]\}^{-1-b} \times \{(1 - x^\alpha)^\beta + a [1 - (1 - x^\alpha)^\beta]\}, \quad 0 < x < 1. \quad (2.8)$$

Um caso especial de (2.8) segue se for definido $\alpha = \beta = 1$ produzindo o modelo $BXII-Kw(a, b, 1, 1)$, que se torna a distribuição Burr XII-uniforme(a, b). Gráficos de (2.8) para alguns valores de parâmetros são dados na Figura 2.3. Entre as diversas formas da fdp, tem-se as formas decrescente, crescente, em U, unimodal e bimodal.

Figura 2.3: Algumas formas da função densidade da distribuição BXII-Kw(a, b, α, β).

Fonte: Próprio autor.

2.4 Propriedades

2.4.1 Função Quantílica e simulação

A fq de X , sendo que $X \sim \text{BXII-G}(a, b, \eta)$ é obtida pela inversão de (2.2). Então, tem-se que

$$x = Q(u) = G^{-1} \left(\frac{z}{1-z} \right)$$

em que $z = z(a, b) = \left[\left(\frac{1}{1-u} \right)^{\frac{1}{b}} - 1 \right]^{\frac{1}{a}}$. Se U tem distribuição uniforme $(0, 1)$, $X = Q(U)$ segue a distribuição BXII-G(a, b, η). Note que se a distribuição de base $G(x)$ tem função inversa em forma fechada $G^{-1}(u)$, $X = Q(U)$ é um método direto para gerar da distribuição em questão. Entretanto, se não possuir função inversa, então $X = Q(U)$ pode ser obtido numericamente.

2.4.2 Momentos

Daqui em diante, seja Y_{i+1} a variável aleatória seguindo a distribuição exp-G com parâmetro de potência $i + 1$, i.e., tendo densidade $h_{i+1}(x; \eta)$. A primeira equação para o n th momento de X é derivada de (2.5) como

$$E(X^n) = \sum_{i=0}^{\infty} \nu_{i+1} E(Y_{i+1}^n). \quad (2.9)$$

Expressões em forma fechada para momentos de diversas distribuições exp-G são

dadas por Nadarajah e Kotz (2006b) podem ser usadas para obter $E(X^n)$ para modelos estendidos na família BXII-G.

Outra maneira de calcular $E(X^n)$ é determinado de (2.9) em termos da fq da distribuição de base

$$E(X^n) = \sum_{i=0}^{\infty} (i+1) \nu_{i+1} A(n, i),$$

em que $A(n, i) = \int_{-\infty}^{\infty} x^n G(x; \eta)^i g(x) dx = \int_0^1 Q_G(u)^n u^i du$.

Os momentos incompletos desempenham um papel importante para medir a desigualdade, por exemplo, pelas curvas de Lorenz e Bonferroni. O n -ésimo momento incompleto de X é dado por

$$m_n(y) = \int_{-\infty}^y x^n f(x) dx = \sum_{i=0}^{\infty} (i+1) \nu_{i+1} \int_0^{G(y)} Q_G(u)^n u^i du. \quad (2.10)$$

A última integral pode ser computada numericamente para a maioria das fqs das distribuições de base.

2.4.3 Função geratriz

Aqui, são fornecidas duas expressões para a fgm $M(t) = M(t; a, b, \eta) = E(e^{tX})$ de X . Claramente, a primeira é obtida de (2.5) como

$$M(t) = \sum_{i=0}^{\infty} \nu_{i+1} M_{i+1}(t), \quad (2.11)$$

em que $M_{k+1}(t)$ é a fgm de Y_{i+1} . Consequentemente, $M(t)$ pode ser determinada da função geratriz da distribuição exp-G.

Outra expressão para $M(t)$ pode ser derivada de (2.5) em termos da fq da distribuição de base

$$M(t) = \sum_{i=0}^{\infty} (i+1) \nu_{i+1} \rho(t, i), \quad (2.12)$$

em que $\rho(t, i) = \int_{-\infty}^{\infty} e^{tx} G(x; \eta)^i g(x; \eta) dx = \int_0^1 e^{tQ_G(u)} u^i du$.

As fgms de diversas distribuições na família BXII-G são obtidas das equações (2.11) e (2.12).

2.4.4 Desvios médios

Os desvios médios em torno da média ($\delta_1 = E(|X - \mu'_1|)$) e em torno da mediana ($\delta_2 = E(|X - M|)$) de X podem ser expressas como

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

respectivamente, em que $\mu'_1 = E(X)$, $M = Q(0.5)$ é a mediana, $F(\mu'_1)$ é facilmente calculada da fda (2.2) e $m_1(z) = \int_{-\infty}^z x f(x) dx$ é o primeiro momento incompleto obtido de (2.10) com $n = 1$.

Agora, serão fornecidas duas maneiras alternativas para computar δ_1 e δ_2 . Uma equação geral para $m_1(z)$ pode ser derivada de (2.5) como

$$m_1(z) = \sum_{i=0}^{\infty} \nu_{i+1} J_{i+1}(z), \quad (2.13)$$

em que $J_{i+1}(z) = \int_{-\infty}^z x h_{i+1}(x) dx$ é uma quantidade básica para computar os desvios médios de Y_{i+1} . Assim, δ_1 e δ_2 são funções do desvio médio da distribuição exp-G

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2 \sum_{i=0}^{\infty} \nu_{i+1} J_{i+1}(\mu'_1) \quad \text{e} \quad \delta_2 = \mu'_1 - 2 \sum_{i=0}^{\infty} \nu_{i+1} J_{i+1}(M).$$

2.4.5 Estatísticas de ordem

Estatísticas de ordem estão entre as mais fundamentais e úteis ferramentas na estatística não-paramétrica e inferência. A densidade da i -ésima estatística de ordem $X_{i:n}$, $f_{i:n}(x)$, em uma amostra aleatória de tamanho n de X é dada por

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) F(x)^{i-1} [1 - F(x)]^{n-i}. \quad (2.14)$$

É possível mostrar que a i -ésima estatística de ordem de X com distribuição BXII-G é dada por

$$f_{i:n}(x) = \sum_{j,l=0}^{\infty} s_{j,l} h_{l+j+1}(x). \quad (2.15)$$

No Apêndice A, Seção A.2 estão descritos os passos para a obtenção dessa mistura além da composição da constante $s_{j,l} = s_{j,l}(a, b)$.

A equação (2.15) revela que a fdp de $X_{i:n}$ é uma combinação linear dupla de funções de densidades da distribuição exp-G.

Importantes propriedades das estatísticas de ordem da distribuição BXII-G podem ser obtidas da mistura como os momentos ordinários e incompletos, fgm e desvios médios.

Claramente, a fda das estatísticas de ordem da distribuição BXII-G é dada por

$$F_{i:n}(x) = \sum_{j,l=0}^{\infty} s_{j,l} H_{l+j+1}(x).$$

2.5 Estimação por máxima verossimilhança

Diversas abordagens para estimação de parâmetros foram largamente usadas, mas o método de estimação por máxima verossimilhança é o mais comumente empregado.

Os EMVs possuem propriedades desejáveis e podem ser usadas quando intervalos de confiança e também estatísticas de teste são construídas. A aproximação normal para estes estimadores em amostras grandes é facilmente manipulado tanto analiticamente quanto numericamente. Seja x_1, \dots, x_n uma amostra observada da família BXII-G, com parâmetros a, b e $\boldsymbol{\eta}$, em que $\boldsymbol{\eta}$ é um vetor de parâmetros da distribuição de base. Então, seja $\theta = (a, b, \boldsymbol{\eta})$. A função de log-verossimilhança total para θ é dada por

$$\begin{aligned} \mathfrak{l}(\theta) = n \log(a b) + \sum_{i=1}^n \left[\log(g(x_i; \boldsymbol{\eta})) + (a-1) \log \left(\frac{G(x_i; \boldsymbol{\eta})}{\bar{G}(x_i; \boldsymbol{\eta})} \right) - \right. \\ \left. - (b+1) \log \left(1 + \left[\frac{G(x_i; \boldsymbol{\eta})}{\bar{G}(x_i; \boldsymbol{\eta})} \right]^a \right) \right], \end{aligned} \quad (2.16)$$

em que $g(x_i; \boldsymbol{\eta}) = d [G(x_i; \boldsymbol{\eta})]/dx_i$. Seja $V(x_i, \boldsymbol{\eta}) = G(x_i; \boldsymbol{\eta})/\bar{G}(x_i; \boldsymbol{\eta})$. A função escore $U(\theta) = (U_a, U_b, U_{\boldsymbol{\eta}})^T$ é obtida a partir das derivadas da função de log-verossimilhança. Seus componentes são dados por

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

$$U_b = \frac{n}{b} - \sum_{i=1}^n \log[1 + V(x_i; \boldsymbol{\eta})^a],$$

$$U_{\boldsymbol{\eta}_j} = \sum_{i=1}^n \left[-a(b+1) \frac{V(x_i; \boldsymbol{\eta})^{a-1} V_{\boldsymbol{\eta}_j}(x_i; \boldsymbol{\eta})}{1 + V(x_i; \boldsymbol{\eta})^a} + (a-1) \frac{V_{\boldsymbol{\eta}_j}(x_i; \boldsymbol{\eta})}{V(x_i; \boldsymbol{\eta})} + \frac{g_{\boldsymbol{\eta}_j}(x_i; \boldsymbol{\eta})}{g(x_i; \boldsymbol{\eta})} \right],$$

para $j = 1, 2, \dots, p$ na última equação, em que $V_{\boldsymbol{\eta}_j}(x_i; \boldsymbol{\eta}) = \partial V(x_i; \boldsymbol{\eta})/\partial \boldsymbol{\eta}_j$ e $g_{\boldsymbol{\eta}_j}(x_i; \boldsymbol{\eta}) = \partial g(x_i; \boldsymbol{\eta})/\partial \boldsymbol{\eta}_j$. As EMVs $(\hat{a}, \hat{b}, \hat{\boldsymbol{\eta}})$ são soluções simultâneas das equações $U_a = 0$, $U_b = 0$ e $U_{\boldsymbol{\eta}} = 0$.

Maximizações de (2.16) podem ser feitas usando rotinas disponíveis no software R (R CORE TEAM, 2015). Algumas estatísticas relacionadas ao ajuste podem ser obtidas no script AdequacyModel (MARINHO; BOURGUIGNON, 2012). Para estimação intervalar sobre os parâmetros θ requer a matriz de informação observada, que é uma matriz quadrada de ordem $(p+2)$ para θ que é obtida derivando a função escore com relação aos componentes de θ .

Para n grande, $(\hat{a} - a, \hat{b} - b, \hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ aproxima a distribuição normal $(p+2)$ -variada com vetor de médias zero e matrix de variância e covariância dada pelo inverso da matriz de informação observada. Essa aproximação da distribuição normal pode ser usada para derivar algumas propriedades dos componentes de θ .

2.6 Aplicações

Nesta seção, usou-se dois conjuntos de dados reais para ajustar alguns modelos da família proposta. Algumas estatísticas descritivas, como a média, mediana, variância,

assimetria e excesso de curtose, além do mínimo e máximo, dos dois conjuntos de dados utilizados estão disponíveis na Tabela 2.1.

Tabela 2.1: Estatísticas descritivas.

Dados	Média	Mediana	Variância	Assimetria	Curtose	Mínimo	Máximo
<i>carbono</i>	3.0498	2.9865	0.3851	0.6396	0.1913	1.901	5.02
<i>esferas</i>	72.21	67.8	1407.947	0.8749	0.1888	17.23	173.4

O primeiro conjunto de dados está disponível em Lawless (2011) e é referente à resistência de 64 fibras de carbono de tamanho 10cm. Através das estatísticas descritivas na Tabela 2.1, nota-se a cauda direita um pouco mais longa, fazendo com que a média seja maior que a mediana. Percebe-se também uma distribuição um pouco mais afunilada quando comparada ao achatamento da distribuição normal. Foram comparadas a distribuição BXII-W(a, b, λ, γ) e outras distribuições de 4 parâmetros de outras famílias: beta-Weibull, BW(a, b, λ, γ), Kumaraswamy-Weibull, KwW(a, b, λ, γ) e McDonald-Exponential, McE(a, b, c, λ).

A Tabela 2.2 lista as EMVs dos parâmetros, erros padrão entre parênteses e os valores das seguintes estatísticas: Critério de Informação de Akaike (AIC), Critério de Informação de Hannan-Quinn (HQIC), teste de Anderson-Darling (A^*) e Critério de Cramér-von Mises (W^*). Para todas as estatísticas, valores pequenos indicam um melhor ajuste. Pode-se notar que o modelo BXII-W fornece o melhor ajuste aos dados de fibra de carbono. Ainda é perceptível alguns erros-padrão superiores às estimativas nas distribuições BW e KwW, indicando parâmetros não significativos (iguais a zero) sob a suposição de normalidade dos estimadores de máxima verossimilhança. Entretanto, como o objetivo é a comparação de diferentes distribuições com quatro parâmetros, optou-se por deixá-los nos modelos.

A seguir, uma comparação visual, na Figura 2.4, das funções de densidade ajustadas e

Tabela 2.2: Modelos ajustados aos dados de fibra de carbono

Distribuição	Estimativas (Erros padrão)				A^*	W^*	AIC	HQIC
BXII-W	3.7325 (1.3448)	0.1029 (0.0573)	3.9496 (0.6796)	0.0271 (0.0139)	0.2357	0.0410	120.208	123.610
BW	22.8314 (2.7328)	0.1278 (0.0151)	0.6699 (3.5408)	3.1984 (0.0053)	2.0708	0.2665	120.652	124.054
KwW	41.1286 (1.0180)	0.6672 (0.0129)	0.8477 (2.4156)	1.6500 (0.0036)	0.3635	0.0692	121.723	125.125
McE	16.6732 (0.0816)	1.5457 (0.0127)	4.9129 (0.0893)	1.4193 (0.0016)	0.3695	0.0689	122.258	125.660

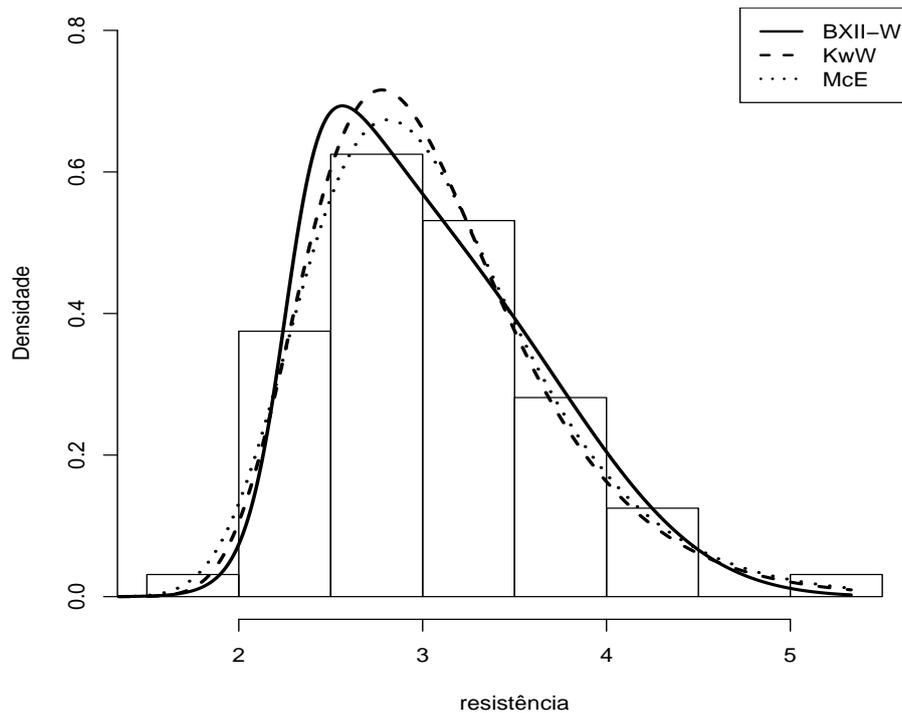
o histograma do conjunto de dados das fibras de carbono é feita. Apenas as três densidades com menores estatísticas A^* e W^* foram ajustadas, para melhor visualização das mesmas.

Nota-se que o modelo BXII-W difere ligeiramente dos outros modelos, com alguma

inclinação, aproximando-se do histograma.

O segundo conjunto de dados usado está disponível em Lieblein e Zelen (1956). Pelas

Figura 2.4: Modelos ajustados e histograma do conjunto de dados fibra de carbono.



Fonte: Próprio autor.

estatísticas descritivas, nota-se que a cauda a direita é bem mais mais longa e que possui também algum afunilamento na forma da distribuição, se comparada a distribuição normal. Estes dados são referentes ao tempo de falha de 23 rolamentos de esferas. A unidade de mensuração está em 10^6 rotações. Cinco modelos de três parâmetros foram ajustados aos dados: Burr XII-exponencial, $BXII-E(a, b, \lambda)$, Weibull-exponencial, $WE(a, b, \lambda)$, Weibull exponencializada, $EW(a, \lambda, \gamma)$, beta-exponencial, $BE(a, b, \lambda)$ e Kumaraswamy-exponencial, $KwE(a, b, \lambda)$.

A Tabela 2.3 lista as mesmas quantidades de interesse citadas anteriormente. Nota-se que as distribuições $BXII-E$, EW e BE apresentam parâmetros não significativos, mas como o intuito é comparar diferentes distribuições de três parâmetros, novamente, optou-se por deixá-los nos modelos. Analisando cada uma das quatro estatísticas, notou-se que o modelo $BXII-E$ fornece o melhor ajuste ao conjunto de dados referente aos rolamentos de esferas, se somente as estatísticas robustas de Anderson-Darling e Cramér-von Mises são consideradas. Outra comparação visual é realizada na Figura 2.5, considerando as densidades ajustadas

Tabela 2.3: Modelos ajustados ao conjunto de dados rolamento de esferas.

Distribuição	Estimativas (Erros Padrão)			A*	W*	AIC	HQIC
BXII-E	2.6546 (1.0141)	0.6441 (0.6915)	0.0129 (0.0060)	0.1917	0.0324	232.467	233.324
WE	3.0214e-05 (1.4750e-05)	4.1540 (0.6526)	0.2061 (0.0299)	0.4718	0.0821	235.073	235.930
EW	0.0278 (0.0319)	1.0804 (0.7003)	4.3001 (6.8453)	0.1947	0.0335	232.129	232.986
BE	4.8588 (3.1330)	1.2550 (2.6418)	0.0269 (0.0436)	0.1941	0.0332	232.136	232.993
KwE	4.6385 (2.4764)	1.1024 (0.9541)	0.0287 (0.0183)	0.1933	0.0330	232.174	233.031

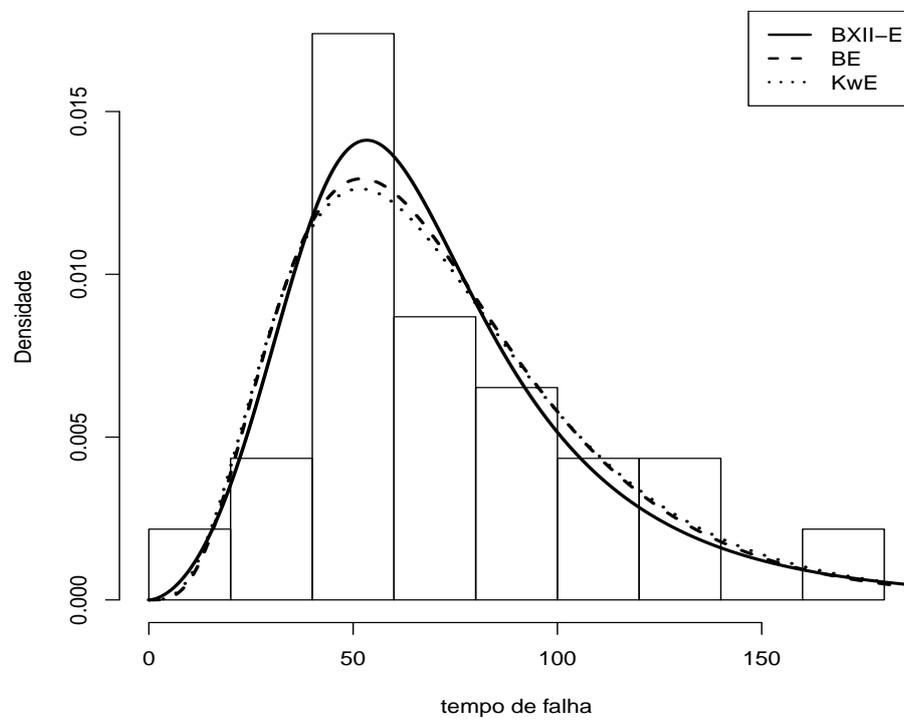
e o histograma dos dados. Novamente, optou-se pela melhor visualização das densidades, ajustando apenas as três que apresentaram menores estatísticas A^* e W^* . Pode-se notar que o modelo BXII-E difere do histograma dos outros modelos, sendo que a diferença mais notável ocorre próximo à moda do conjunto de dados, muito embora ainda haja uma diferença considerável entre os modelos ajustados e a moda do conjunto de dados.

2.7 Conclusões

Neste capítulo, foi apresentada uma nova família de distribuições que generaliza duas classes de modelos: as famílias Lehmann tipo-II e odd log-logistic. A distribuição de base é um exemplo básico desta família. Provou-se que a nova família é um modelo competitivo para outras famílias, quando comparadas com o mesmo número de parâmetros.

Além disso, foram investigadas algumas propriedades como os momentos ordinários e incompletos, funções quantil e geratriz, desvios médios, entre outros. Um dos principais resultados indica que a função densidade da BXII-G é uma mistura de funções densidades da distribuição exp-G.

Demonstrou-se que os modelos BXII-W e BXII-E podem fornecer melhores ajustes a dados reais se comparados com outras distribuições conhecidas, baseadas nos geradores beta, McDonald e Kumaraswamy.

Figura 2.5: Modelos ajustados e histograma do conjunto de dados rolamentos de esferas.

Fonte: Próprio autor.

Capítulo 3

Distribuição Taxa de falha linear generalizada estendida: propriedades e aplicações

Resumo

Inúmeras distribuições clássicas têm sido usadas extensivamente nas últimas décadas para modelagem de dados de tempo de vida. Um novo modelo de cinco parâmetros para tempo de vida foi definido e algumas propriedades matemáticas básicas foram derivadas baseadas em uma representação linear simples de sua função densidade de probabilidade (fdp). Existem, no mínimo, onze distribuições especiais do novo modelo. Sua função taxa de falha (ftf) pode apresentar as formas crescente, decrescente e banheira. Os parâmetros do modelo são estimados pelo método da máxima verossimilhança. A utilidade do modelo proposto foi mostrada por meio de duas aplicações a dados reais (SANTOS; CORDEIRO, 2015).

Palavras-chave: Família das exponencializadas generalizada. Distribuição taxa de falha linear generalizada. Estimação por máxima verossimilhança. Momentos. Função quantílica.

Abstract

Numerous classical distributions have been extensively used over the past decades for modeling lifetime data. We define a new five-parameter lifetime model and derive some basic mathematical properties based on a simple linear representation of its density function. There are at least eleven special distribution of the new model. Its failure rate function may present increasing, decreasing and bathtub shapes. The model parameters are estimated by maximum likelihood. We show the usefulness of the proposed model by

means of two real data sets (SANTOS; CORDEIRO, 2015).

keywords: Exponentiated generalized family. Generalized linear failure rate distribution. Maximum likelihood estimation. Moment. Quantile function.

3.1 Introdução

Bain (1974) foi pioneiro apresentando a distribuição taxa de falha linear (LFR), que apresenta função de distribuição acumulada (fda) dada por

$$H(x; \alpha, \beta) = 1 - e^{-\alpha x - \frac{\beta}{2} x^2}. \quad (3.1)$$

As distribuições exponencial (E) e Rayleigh (R) podem ser obtidos a partir da distribuição LFR impondo $\beta = 0$ e $\alpha = 0$, respectivamente.

Sarhan e Kundu (2009) introduziram a distribuição tri-paramétrica taxa de falha linear generalizada (GLFR) cuja fda é

$$G(x; \xi) = (1 - e^{-\alpha x - \frac{\beta}{2} x^2})^\theta, \quad x > 0. \quad (3.2)$$

Sua fdp é dada por

$$g(x, \xi) = \theta (\alpha + \beta x) e^{-\alpha x - \frac{\beta}{2} x^2} (1 - e^{-\alpha x - \frac{\beta}{2} x^2})^{\theta-1}, \quad x > 0,$$

em que $\xi = (\alpha, \beta, \theta)$, for $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta > 0$ e $\theta > 0$. Esta distribuição consiste na aplicação da distribuição LFR na família das exponencializadas (exp-G) pela adição de um parâmetro de potência. Note que, se $\theta = 1$, a distribuição LFR é obtida como caso especial.

Cordeiro et al. (2013) propuseram a família de distribuições exponencializadas generalizada (EG-G). A fda dessa família é dada por

$$F(x; a, b) = \{1 - [1 - G(x)]^a\}^b, \quad x \in \mathbb{R}, \quad (3.3)$$

e sua fdp é

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

para $a > 0$, $b > 0$, em que $g(x) = dG(x)/dx$. Esta família engloba as classes Lehmann tipo-I ($a = 1$) e Lehmann tipo-II ($b = 1$).

Neste capítulo, uma nova distribuição com cinco parâmetros é proposta e consiste em inserir a fda da distribuição GLFR (3.2) na equação (3.3), sendo referida como modelo taxa de falha linear generalizada estendida (EGLFR). Esta nova distribuição generaliza onze distribuições, como as distribuições exponencial (E), Rayleigh (R), exponencial

generalizada (GE), Rayleigh generalizada (GR), entre outras.

O capítulo está organizado como segue. Na Seção 3.2, a distribuição EGLFR é definida e seus casos especiais apresentados. Na Seção 3.3, discute-se a estimação por máxima verossimilhança e os componentes da função escore são apresentados. Na Seção 3.4, duas aplicações são fornecidas como prova empírica da flexibilidade do novo modelo. Algumas expressões explícitas são apresentadas para os momentos ordinários e incompletos, funções quantílica e geratriz, momentos ponderados por probabilidade (MPPs) e estatísticas de ordem na Seção 3.5. Na Seção 3.6, algumas considerações finais.

3.2 A distribuição EGLFR

A variável aleatória X tendo distribuição EGLFR é denotada por $X \sim \text{EGLFR}(a, b, \alpha, \beta, \theta)$. A fda de X é dada por

$$F(x, \zeta) = \left\{ 1 - \left[1 - \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^\theta \right]^a \right\}^b, \quad x > 0, \quad (3.4)$$

em que $\zeta = (a, b, \alpha, \beta, \theta)$. Sua fdp é dada por

$$f(x; \zeta) = a b \theta (\alpha + \beta x) e^{-\alpha x - \frac{\beta}{2} x^2} \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^{\theta-1} \left[1 - \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^\theta \right]^{a-1} \\ \times \left\{ 1 - \left[1 - \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^\theta \right]^a \right\}^{b-1}.$$

A ftf da variável aleatória X reduz-se a

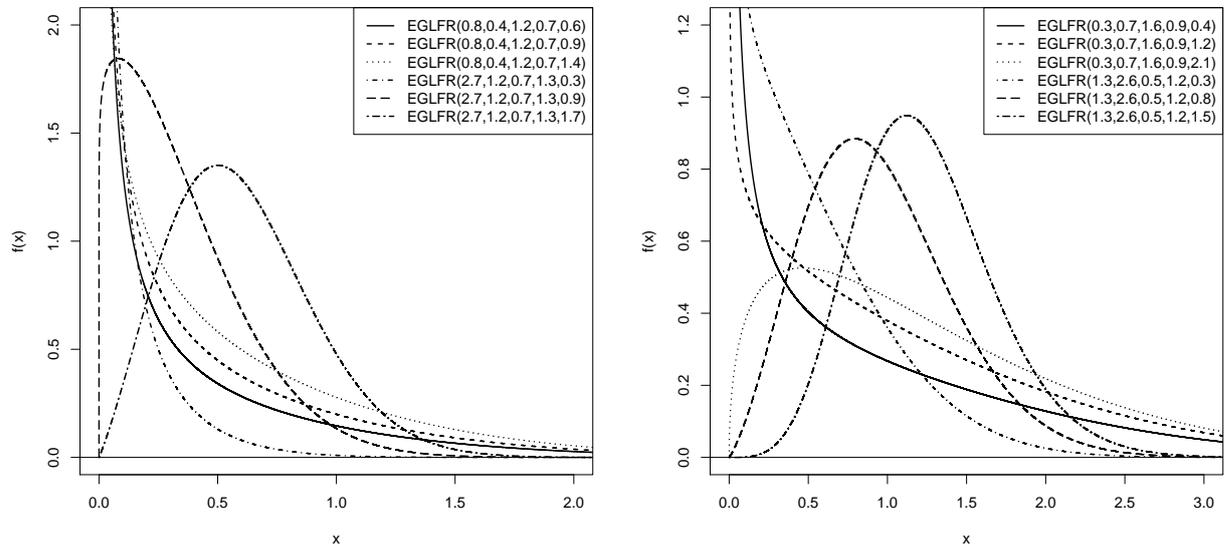
$$r(x; \zeta) = a b \theta (\alpha + \beta x) e^{-\alpha x - \frac{\beta}{2} x^2} \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^{\theta-1} \left[1 - \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^\theta \right]^{a-1} \times \\ \left\{ 1 - \left[1 - \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^\theta \right]^a \right\}^{b-1} / \left(1 - \left\{ 1 - \left[1 - \left(1 - e^{-\alpha x - \frac{\beta}{2} x^2} \right)^\theta \right]^a \right\}^b \right).$$

Na Figura 3.1 e 3.2 mostram-se algumas formas da fdp e ftf de X . Claramente, a ftf apresenta as formas crescente, decrescente e banheira.

3.2.1 Casos especiais da distribuição EGLFR

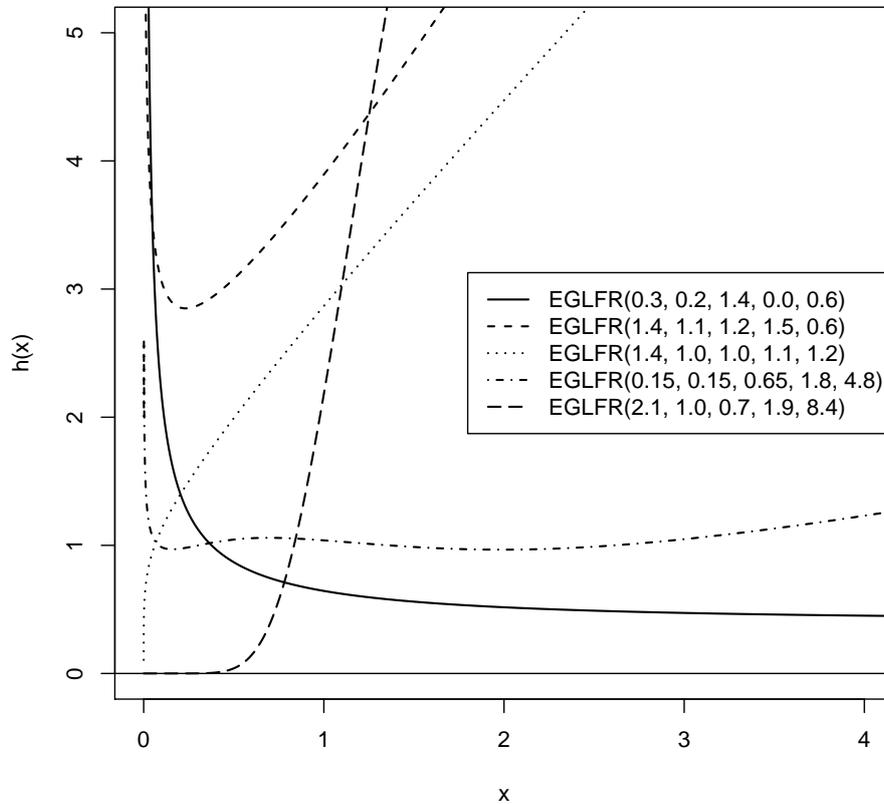
Nesta seção, serão apresentados onze modelos especiais que são derivados da distribuição EGLFR na equação (3.4). Todos os modelos especiais foram obtidos por alguns valores inteiros dos parâmetros. É possível chegar a mesma distribuição através da reparametrização.

1. Para $b = 1$, tem-se o modelo Lehmann tipo-II - taxa de falha linear generalizada, LeII - GLFR $(a, \alpha, \beta, \theta)$. Não foram encontrados registros desta distribuição na literatura.

Figura 3.1: Algumas formas da fdp da distribuição $EGLFR(a, b, \alpha, \beta, \theta)$.

Fonte: Próprio autor.

2. Para $\beta = 0$, o modelo exponencializada generalizada - exponencial generalizada é obtido, $EG-GE(a, b, \alpha, \theta)$. Também não foram encontrados registros desta distribuição na literatura.
3. Para $\alpha = 0$, segue o modelo exponencializada generalizada - Rayleigh generalizada, $EG - GR(a, b, \beta, \theta)$. Não foram encontrados registros desta distribuição na literatura.
4. Para $a = b = 1$, resume-se na distribuição GLFR.
5. Para $b = 1$ e $\beta = 0$, segue o modelo Lehmann tipo-II - exponencial generalizada, $LeII-GE(a, \alpha, \theta)$. Não foram encontrados registros desta distribuição na literatura.
6. Para $b = 1$ e $\alpha = 0$, obtem-se o modelo Lehmann tipo II - Rayleigh generalizada, $LeII-GR(a, \beta, \theta)$. Não foram encontrados registros desta distribuição na literatura.
7. Para $a = b = 1$ e $\beta = 0$, tem-se o modelo exponencial generalizada, $GE(\alpha, \theta)$, definido por Mudholkar e Srivastava (1993), mas já utilizada para comparar tabelas de mortalidade humana e representar crescimento da população no século XIX por Gompertz e Verhulst (GUPTA; KUNDU, 2001).
8. Para $a = b = 1$ e $\alpha = 0$, o modelo Rayleigh generalizado é obtido, $GR(\beta, \theta)$, introduzido por Surlles e Padgett (2001) como distribuição Burr tipo X e corretamente

Figura 3.2: Algumas formas da ftf da distribuição $EGLFR(a, b, \alpha, \beta, \theta)$.

Fonte: Próprio autor.

nomeado como o modelo Rayleigh generalizado (KUNDU; RAQAB, 2005).

9. Para $a = b = 1$ e $\theta = 1$, tem-se a distribuição LFR.
10. Para $a = b = 1$, $\beta = 0$ e $\theta = 1$, obtem-se a distribuição exponencial.
11. Para $a = b = 1$, $\alpha = 0$ e $\theta = 1$, segue a distribuição Rayleigh.

3.3 Estimação por máxima verossimilhança

Nesta seção, é abordada a metodologia de estimação por máxima verossimilhança (EMVs) dos parâmetros desconhecidos em $\zeta = (a, b, \alpha, \beta, \theta)$ da distribuição EGLFR. Considere que x_1, \dots, x_n é uma amostra observada desta distribuição. A função log-

verossimilhança total para ζ é dada por

$$l(\zeta) = n \log(ab\theta) + \sum_{i=1}^n \left\{ \log(\alpha + \beta x_i) - \alpha x_i - \frac{\beta}{2} x_i^2 + \right. \\ \left. + (\theta - 1) \log(u_i) + (a - 1) \log(1 - u_i^\theta) + (b - 1) \log[1 - (1 - u_i^\theta)^a] \right\},$$

em que $u_i = 1 - e^{-\alpha x_i - \frac{\beta}{2} x_i^2}$.

A função escore $U(\zeta) = (U_a, U_b, U_\alpha, U_\beta, U_\theta)^T$ é dada pela derivada da função log-verossimilhança com relação aos parâmetros. Seus componentes são

$$U_a = \frac{n}{a} + \sum_{i=1}^n \log(1 - u_i^\theta) \left[\frac{(b-1)(1 - u_i^\theta)^a}{(1 - u_i^\theta)^a - 1} + 1 \right], \\ U_b = \frac{n}{b} + \sum_{i=1}^n \log[1 - (1 - u_i^\theta)^a], \\ U_\alpha = \sum_{i=1}^n (\alpha + \beta x_i)^{-1} - x_i + x_i(1 - u_i) \left[\frac{\theta - 1}{u_i} - \frac{\theta(a-1)u_i^{\theta-1}}{1 - u_i^\theta} + \right. \\ \left. + \frac{a\theta(b-1)u_i^{\theta-1}(1 - u_i^\theta)^{a-1}}{1 - (1 - u_i^\theta)^a} \right], \\ U_\beta = \sum_{i=1}^n \frac{x_i}{\alpha + \beta x_i} - \frac{x_i^2}{2} + \frac{x_i^2(1 - u_i)}{2} \left[\frac{\theta - 1}{u_i} - \frac{\theta(a-1)u_i^{\theta-1}}{1 - u_i^\theta} + \right. \\ \left. + \frac{a\theta(b-1)u_i^{\theta-1}(1 - u_i^\theta)^{a-1}}{1 - (1 - u_i^\theta)^a} \right], \\ U_\theta = \frac{n}{\theta} + \sum_{i=1}^n \log(u_i) \left[1 - \frac{(a-1)u_i^\theta}{1 - u_i^\theta} + \frac{a(b-1)u_i^\theta(1 - u_i^\theta)^{a-1}}{1 - (1 - u_i^\theta)^a} \right].$$

A EMV $\hat{\zeta}$ de ζ pode ser obtido resolvendo simultaneamente as equações $U(\hat{\zeta}) = 0$. Não existem expressões em forma fechada para a EMV $\hat{\zeta}$ e deve ser computado através de métodos numéricos usando um algoritmo de otimização não linear.

Para estimação intervalar e testes de hipóteses sobre os parâmetros do modelo requer-se a matriz de informação observada, que é obtido diferenciando a função escore com relação a cada componente de ζ . Sob certas condições, a distribuição assintótica de $(\hat{\zeta} - \zeta)$ é $N_5(0, K(\zeta)^{-1})$, em que $K(\zeta)$ é a matriz de informação observada. Portanto, a aproximação normal pode ser usada para derivar propriedades de $\hat{\zeta}$.

3.4 Análise de dados reais

Nesta seção, são apresentadas duas análises de dados reais para ilustrar a performance do novo modelo na prática. O primeiro conjunto de dados é dado por Aarset (1987) e

consiste em tempo até a primeira falha de cinquenta dispositivos. O segundo conjunto de dados é dado por Proschan (1963) e consiste no número de falhas sucessivas no sistema de ar-condicionado da frota de aviões a jato Boeing 720. A Tabela 3.1 fornece uma sumarização descritiva de ambos os conjuntos de dados. Analisando esta tabela para o primeiro conjunto de dados (*aarset*), nota-se a assimetria negativa, indicando uma leve concentração à esquerda da massa de probabilidade, além de possuir distribuição mais dispersa. Já para o segundo conjunto de dados (*aircond*), indica assimetria positiva (cauda a direita mais pesada) e uma distribuição mais afunilada.

Para o primeiro conjunto de dados, comparou-se os modelos especiais da distribuição

Tabela 3.1: Estatísticas descritivas.

Dados	Média	Mediana	Variância	Assimetria	Curtose	Mínimo	Máximo
<i>aarset</i>	49.53	65	1163.654	-0.2586	-1.7061	0.1	86
<i>aircond</i>	92.07	54	11645.93	2.1221	4.9379	1	603

EGLFR usando o teste da razão de verossimilhança (LR) (CASELLA; BERGER, 2002). Na Tabela 3.2, as EMVs dos parâmetros e seus erros padrão são apresentadas. Na Tabela 3.3, são fornecidos os valores da estatística LR, graus de liberdade correspondente e p-valor para alguns casos especiais da distribuição EGLFR. O modelo EGLFR caracteriza-se por pequenos erros padrão, enquanto o modelo EG-GR apresenta dois parâmetros com alta variabilidade. Além disso, o modelo EGLFR possui o maior valor de log-verossimilhança.

Executando o teste de hipótese apresentado na Tabela 3.3, todos os casos que são rejeitados ao nível de significância de 5% foram apresentadas. Neste mesmo nível, para $\alpha = 0$, cujo p-valor é 0.091, a distribuição EG-GR não é rejeitada.

Adicionalmente, uma comparação visual é fornecida entre quatro densidades ajustadas e o histograma dos dados na Figura 3.3. Notou-se que as densidades EGLFR e EG-GR têm uma vantagem sobre as outras densidades para tempos altos. O modelo EG-GR é penalizado pelo ajuste ruim para pequenos valores do tempo.

Para o segundo conjunto de dados, outras distribuições também com cinco parâmetros foram usadas: Kumaraswamy exponencializada - Dagum, $EK-D(a, b, \lambda, \delta, \beta)$, introduzido por Huang e Oluyede (2014) (2014) (que pode ser entendida como exponencializada generalizada - Dagum, EG-D); McDonald - taxa de falha linear, $Mc-LFR(a, b, c, \alpha, \beta)$, definido por Elbatal et al. (2014); McDonald - beta invertida, $Mc-IB(a, b, c, \alpha, \beta)$, proposto por Cordeiro e Lemonte (2012); e beta generalizada refletida - Weibull inversa, $RGB-IW(a, b, c, \lambda, \theta)$, introduzida por Elbatal et al. (2013).

A Tabela 3.4 lista a EMVs dos parâmetros do modelo, erros padrão entre parênteses e valores das estatísticas: Critério de Informação de Akaike (AIC) e Critério de Informação de Hannan-Quinn (HQIC). A Tabela 3.5 lista os valores das estatísticas Anderson-Darling

Tabela 3.2: Modelos ajustados ao conjunto de dados Aarset.

Distribuições	EMV (Erro padrão)					log-verossimilhança
	a	b	α	β	θ	
EGLFR	38.4999 (18.0535)	0.1373 (0.0230)	0.0035 (0.0001)	6.4983e-05 (9.9354e-06)	4.5731 (0.6179)	-228.1943
EG-GR	0.3359 (0.0354)	3.2114 (14.5997)	0 (-)	0.0016 (0.0002)	0.0286 (0.0668)	-229.6161
EG-GE	395.4852 (47.8087)	0.1571 (0.0298)	0.0033 (0.0007)	0 (-)	4.5484 (0.8648)	-230.8900
GLFR	1 (-)	1 (-)	0.0038 (0.0009)	0.0003 (3.9181e-05)	0.5327 (0.0935)	-233.1447
LeII-GLFR	0.0013 (0.0001)	1 (-)	0.0038 (0.0009)	0.0003 (4.1643e-05)	383.7255 (38.7859)	-233.1447
LeII-GR	199.7615 (30.7621)	1 (-)	0 (-)	0.0003 (3.6147e-05)	0.0017 (0.0001)	-234.5654
GR	1 (-)	1 (-)	0 (-)	0.0003 (3.9972e-05)	0.3519 (0.0698)	-234.5654
LFR	1 (-)	1 (-)	0.0136 (0.0045)	0.0002 (6.9670e-05)	1 (-)	-238.0636
LeII-GE	35.2089 (7.8541)	1 (-)	0.0187 (0.0021)	0 (-)	0.0221 (0.0056)	-239.9951
GE	1 (-)	1 (-)	0.0187 (0.0019)	0 (-)	0.7798 (0.1634)	-239.9951
E	1 (-)	1 (-)	0.0218 (0.0022)	0 (-)	1 (-)	-241.0895
R	1 (-)	1 (-)	0 (-)	0.0006 (8.4133e-05)	1 (-)	-264.0528

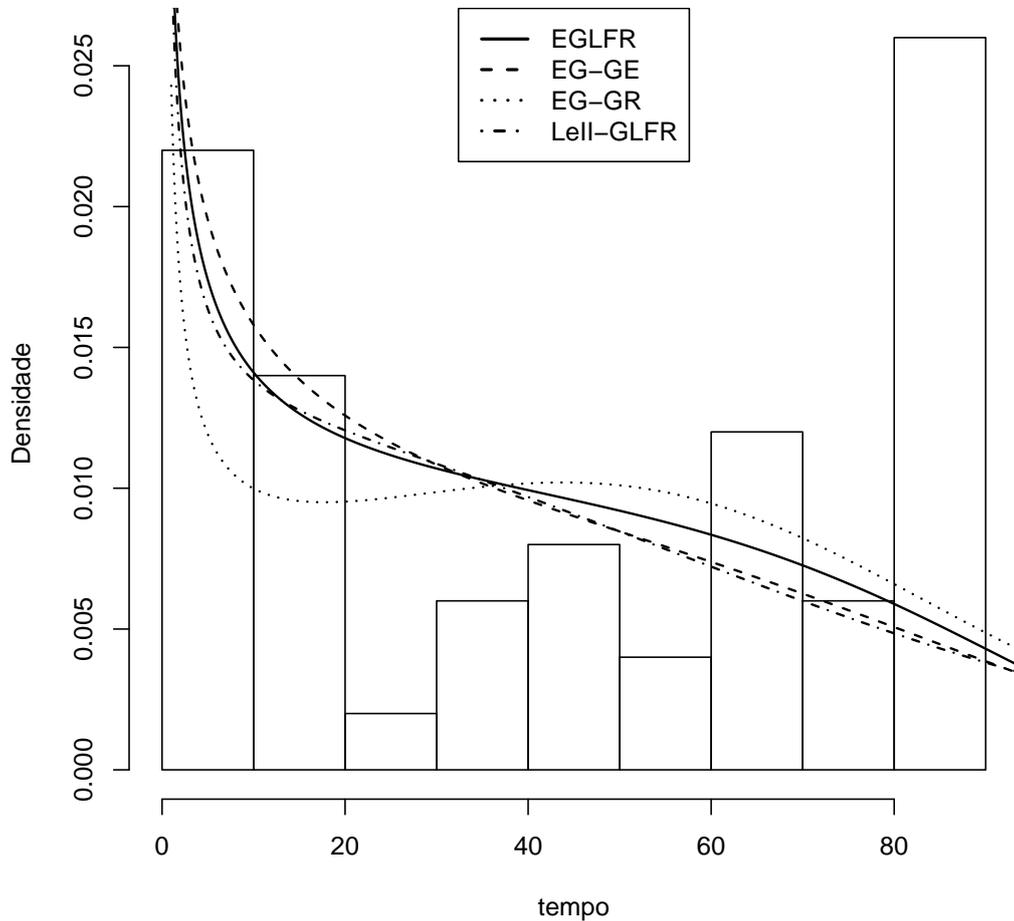
Tabela 3.3: A hipótese nula, valores da estatística LR, graus de liberdade e p-valores para os melhores ajustes casos especiais da distribuição EGLFR para o conjunto de dados Aarset.

Distribuição	H_0	Estatística de teste	g.l.	p-valor
EG-GE	$\beta = 0$	5.3914	1	0.020
GLFR	$a = 1, b = 1$	9.9008	2	0.007
LeII-GLFR	$b = 1$	9.9008	1	0.001
LeII-GR	$b = 1, \beta = 0, \alpha = 0$	12.7422	3	0.005

(A*) e Cramér-von Mises (W*) como descrito por Chen e Balakrishnan (1995). Para todas as estatísticas, valores pequenos são desejados.

Na Tabela 3.4, notou-se que as estatísticas AIC e HQIC têm seus menores valores para a distribuição EGLFR, que indica o melhor ajuste. Adicionalmente, o modelo EGLFR apresenta os menores erros padrão para todas as estimativas. Outras distribuições apresentam grande variabilidade em algumas estimativas.

Na Tabela 3.5, as estatísticas A* e W* apresentam os menores valores para a distribuição EGLFR, indicando assim o melhor ajuste. Estas duas últimas estatísticas são

Figura 3.3: Densidades ajustadas e histograma do conjuntos e dados Aarset.

Fonte: Próprio autor.

mais robustas que as estatísticas AIC e HQIC.

Uma comparação visual é fornecida na Figura 3.4 através do gráfico quantil-quantil, escolhido devido aos ajustes semelhantes dos modelos. Além disso, apenas os três modelos com menores valores para as estatísticas A^* e W^* foram estudados visualmente. Os três modelos têm ajustes semelhantes para pequenos valores, enquanto o modelo EGLFR tem alguma vantagem sobre Mc-LFR e Mc-IB para grandes valores.

3.5 Propriedades matemáticas

Nesta seção serão apresentadas algumas propriedades matemáticas da distribuição EGLFR. Inicialmente, alguns resultados úteis serão apresentados.

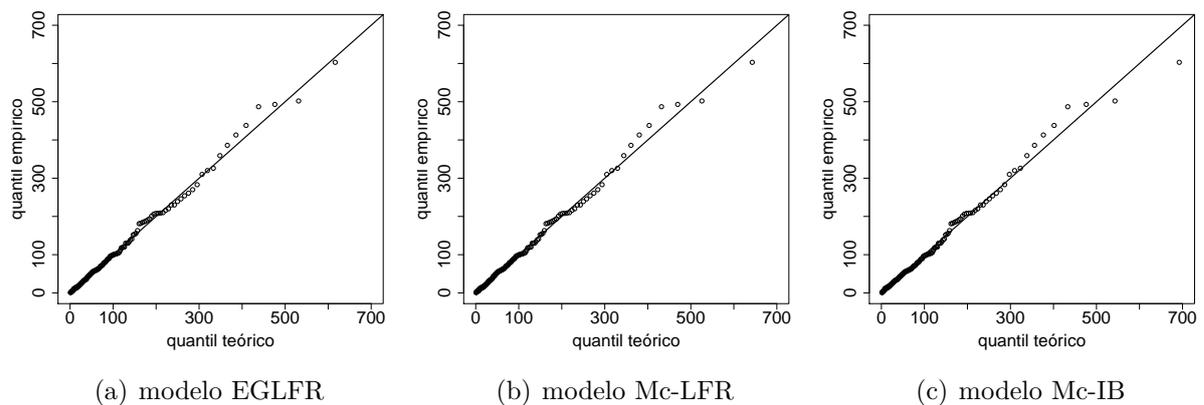
Tabela 3.4: Modelos ajustados para o conjunto de dados aircond.

Distribuição	EMV (Erros padrão)					AIC	HQIC
	a	b	α	β	θ		
EGLFR	0.1223 (0.0296)	0.2725 (0.0595)	0.0389 (0.0048)	7.0746e-05 (3.3342e-05)	5.4996 (1.9824)	2072.1978	2078.7542
EK-D	11.4997 (7.5376)	0.1584 (0.0158)	11.8155 (2.2)	0.7401 (0.0451)	14.2439 (7.3512)	2073.8833	2080.4397
Mc-LFR	0.1167 (0.0592)	0.2465 (0.0619)	12.1994 (9.7997)	0.0298 (0.0097)	1.3987e-05 (5.0201e-05)	2073.6893	2080.2457
Mc-IB	144.7402 (44.9590)	61.0816 (30.6679)	976.3466 (297.4680)	0.0240 (0.0390)	0.4339 (0.1213)	2075.2300	2081.7864
RGB-IW	43.3687 (20.1383)	0.2020 (0.0282)	1.4997 (7.9185)	36.3314 (7.2991)	0.3789 (0.0452)	2073.0023	2079.5587

Tabela 3.5: Estatísticas Anderson Darling e Cramér-von Mises para os modelos ajustados ao conjunto de dados aircond.

Distribuição	A*	W*
EGLFR	0.1602	0.0207
EK-D	0.2814	0.0396
Mc-LFR	0.1767	0.0244
Mc-IB	0.2632	0.0351
RGB-IW	0.2839	0.0377

Figura 3.4: Gráfico Quantil-Quantil para os dados aircond.



Fonte: Próprio autor.

Uma representação para a seguinte integral

$$\eta_{k,n}(y) = \eta_{k,n}(y; \alpha, \beta) = \int_0^y x^k e^{-(n+1)(\alpha x + \frac{\beta}{2}x^2)} dx, \quad (3.5)$$

nos casos em que $y \rightarrow \infty$ (denotado $\eta_{k,n}$) e $0 < y < \infty$ será necessária. Para o último caso, a equação (3.5), usando Prudnikov et al. (1986, p. 321), tem forma fechada dada por (CORDEIRO et al., 2014)

$$\eta_{k,n}(z) = \sum_{j=0}^1 \frac{(-1)^j 2^{(k-1+j)/2} \Gamma\left(\frac{k+1+j}{2}\right)}{[(n+1)\beta]^{(k+1+j)/2} j!} {}_2F_2\left(1, \Delta(1, [k+1+j]/2); \Delta(2, 1+j); z^{-1}\right), \quad (3.6)$$

em que $\Delta(r_1, r_2) = (r_2/r_1, (r_2+1)/r_1, (r_2+2)/r_1, \dots, (r_2+r_1-1)/r_1)$, $z = 2\beta/[(n+1)\alpha^2]$, ${}_pF_q(t_1, \dots, t_p; l_1, \dots, l_q; x)$ é a função hipergeométrica generalizada definida por

$${}_pF_q(t_1, \dots, t_p; l_1, \dots, l_q; x) = \sum_{k=0}^{\infty} \frac{(t_1)_k (t_2)_k \dots (t_p)_k x^k}{(l_1)_k (l_2)_k \dots (l_q)_k k!},$$

e $(t)_k = \Gamma(t+k)/\Gamma(t)$ é chamado de fatorial ascendente.

Seja $J_{k,n}(y) = \int_0^y x^k e^{-(n+1)\frac{\beta}{2}x^2} dx$. Expandindo o termo exponencial $e^{-(n+1)\alpha x}$ e usando a identidade de Prudnikov et al. (1986, p. 348), para o primeiro caso, obtém-se (CORDEIRO et al., 2014), para $0 < y < \infty$,

$$J_{k,n}(y) = \frac{y^{2(k+1)}}{2} G_{2,3}^{1,2} \left((n+1) \frac{\beta}{2} y^2 \left| \begin{matrix} \frac{-k}{2}, \frac{1-k}{2} \\ 0, \frac{-k-1}{2}, \frac{k}{2} \end{matrix} \right. \right), \quad (3.7)$$

em que $G_{p,q}^{m,n}$ é chamada de função Meijer G definida como segue

$$G_{p,q}^{m,n} \left(x \left| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right. \right) = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(b_j + t) \prod_{j=1}^n \Gamma(1 - a_j - t)}{\prod_{j=n+1}^p \Gamma(a_j + t) \prod_{j=m+1}^q \Gamma(1 - b_j - t)} x^{-t} dt.$$

Então, equação (3.5) reduz-se à

$$\eta_{k,n}(y) = \sum_{j=0}^{\infty} \frac{(-1)^j [(n+1)\alpha]^j}{j!} J_{k+j,n}(y). \quad (3.8)$$

Cordeiro et al. (2013) provaram para qualquer fda G da família EG-G o seguinte resultado

$$F(x) = \sum_{j=0}^{\infty} w_j G(x)^j \quad (3.9)$$

em que $w_j = w_j(a, b) = \sum_{i=0}^{\infty} \frac{(-1)^{i+j} \Gamma(b+1) \Gamma(ia+1)}{\Gamma(b-i) \Gamma(ai-j+1) i! j!}$. Equação (3.9) representa uma mistura de fdas exp-G. Tomando a fda da distribuição GLFR na equação (3.2) como $G(x)$ e usando a expansão binomial, equação (3.9) pode ser reescrita como

$$F(x) = \sum_{l=0}^{\infty} \kappa_l [1 - H(x; l\alpha, l\beta)], \quad (3.10)$$

em que $\kappa_l = \kappa_l(a, b, \theta) = \frac{(-1)^l \Gamma(\theta l + 1)}{\Gamma(\theta l - l + 1) l!} \sum_{j=0}^{\infty} w_j$.

Finalmente, a fdp de X pode ser expressa como

$$f(x) = \sum_{n=0}^{\infty} v_n h(x; (n+1)\alpha, (n+1)\beta), \quad (3.11)$$

em que $v_n = v_n(a, b, \theta) = -\kappa_{n+1}$ e $h(x; \alpha, \beta) = dH(x; \alpha, \beta)/dx = (\alpha + \beta x) e^{-\alpha x - \frac{\beta}{2} x^2}$ denota fdp da distribuição LFR. A representação na forma de mistura em (3.11) é muito útil para derivar algumas propriedades matemáticas da distribuição EGLFR. Várias das expressões seguintes foram obtidas usando o software MATHEMATICA (RESEARCH, 2012).

3.5.1 Momentos e função geratriz

Uma representação simples do k -ésimo momento de X é dado por

$$\mu'_k = E(X^k) = \sum_{n=0}^{\infty} v_n (n+1) (\alpha \eta_{k,n} + \beta \eta_{k+1,n}), \quad (3.12)$$

em que $\eta_{k,n}$ é definido pela equação (3.6).

A função geratriz de momentos (fgm) de X , $M(t) = E(e^{tX})$, é definida por

$$M(t) = \sum_{n=0}^{\infty} v_n \left\{ \sqrt{\iota_2} + \sqrt{\pi} e^{\frac{(\iota_1-t)^2}{\iota_2}} \left[t + \iota_1 \operatorname{erf} \left(\frac{(t-\iota_1)^2}{\sqrt{\iota_2}} \right) + |t-\iota_1| \operatorname{erf} \left(\frac{|\iota_1-t|}{\sqrt{\iota_2}} \right) \right] \right\} / \sqrt{\iota_2},$$

em que $\iota_1 = \alpha(n+1)$, $\iota_2 = 2\beta(n+1)$ e $\operatorname{erf}(x) = 2\pi^{-1/2} \int_0^x e^{-t^2} dt$ é a função erro. As Figuras 3.5 e 3.6 mostram gráficos da média e variância, respectivamente, como função dos parâmetros a e b , fixando os outros parâmetros.

Analisando a média, verifica-se, para valores pequenos e fixos do parâmetro b , uma função crescente da média conforme o parâmetro a cresce. Já com o parâmetro a fixo, tem-se uma função decrescente da média conforme o parâmetro b cresce, em ambos os gráficos da Figura 3.5. Foi quase imperceptível o impacto da variação do parâmetro θ de 0.8 para 1.3, ocasionando um pequeno aumento da média. Analisando a variância na Figura 3.6, para valores pequenos e fixos de b , nota-se uma função crescente da variância para pequenos valores do parâmetro a e decrescente, conforme o parâmetro a cresce. Fixando o parâmetro a , tem-se uma função decrescente da variância conforme o parâmetro b cresce. Novamente, foi muito pequeno o impacto da variação do parâmetro θ de 0.8 para 1.3, ocasionando uma leve diminuição na variância.

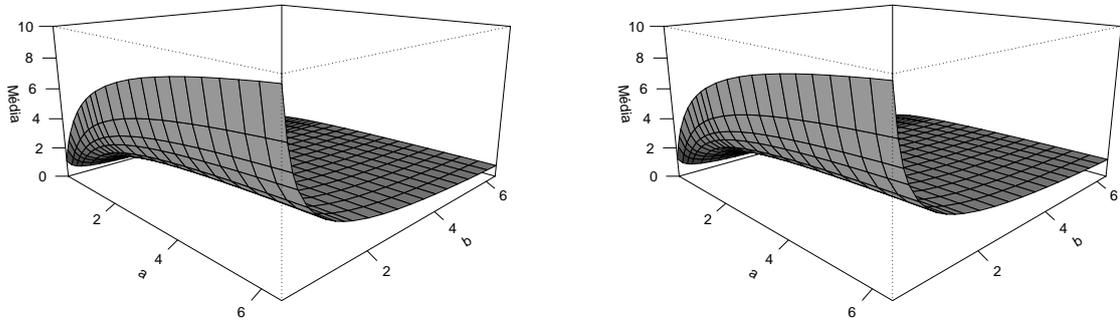
3.5.2 Função quantílica

A função quantílica (fq) de X , $Q(u) = F^{-1}(u)$, pode ser determinada como a solução positiva da equação de segundo grau $\frac{\beta}{2}x^2 + \alpha x + c = 0$, em que $c = \log \left[1 - \sqrt[{\theta}]{1 - \sqrt[{\alpha}]{1 - \sqrt[{\beta}]{u}}} \right]$. Então, a fq de X é dada por

$$Q(u) = \beta^{-1} \left[\sqrt{\alpha^2 - 2\beta \log \left(1 - \sqrt[{\theta}]{1 - \sqrt[{\alpha}]{1 - \sqrt[{\beta}]{u}}} \right)} - \alpha \right]. \quad (3.13)$$

Usando a equação (3.13), a mediana da distribuição EGLFR segue fazendo $u = 1/2$.

Figura 3.5: Média da distribuição EGLFR($a, b, \alpha, \beta, \theta$).



(a) $\alpha = 0.2$, $\beta = 0.4$ e $\theta = 0.8$

(b) $\alpha = 0.2$, $\beta = 0.4$ e $\theta = 1.3$

Fonte: Próprio autor.

3.5.3 Momentos ponderados por probabilidade

O MPP foi originalmente proposto por Greenwood et al. (1979) e usado como uma alternativa ao método da máxima verossimilhança para estimação de parâmetros. O (s, r) -ésimo MPP de X é definido por $\delta_{s,r} = E[X^s F(X)^r] = \int_0^\infty x^s [F(x)]^r f(x) dx$. Claramente, os momentos ordinários são obtidos quando $r = 0$.

Por outro lado, usando a identidade de uma série de potência elevado a um número inteiro para $F(x)^r$, em que $F(x)$ e $f(x)$ são dados pelas equações (3.10) e (3.11), respectivamente, o PMW de X se reduz à

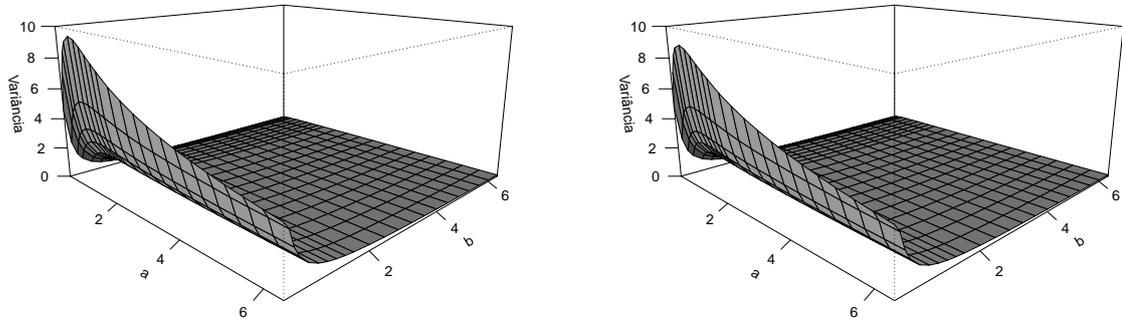
$$\delta_{s,r} = \sum_{l,n=0}^{\infty} \tau_{r,l} v_n (n+1) (\alpha \eta_{s,l+n} + \beta \eta_{s+1,l+n}), \quad (3.14)$$

em que $\tau_{r,0} = \kappa_0^r$ e $\tau_{r,l} = (l\kappa_0)^{-1} \sum_{k=1}^l (kr - l + k) \kappa_k \tau_{r,l-k}$, para $l \geq 1$, $\eta_{s,l+n}$, κ_l e v_n são definidos pelas equações (3.6), (3.10) e (3.11), respectivamente.

3.5.4 Momentos incompletos

O s -ésimo momento incompleto de X é definido por $m_s(y) = E(X^s | X < y)$. Usando as equações (3.8) e (3.11), escreve-se como

$$m_s(y) = \sum_{n=0}^{\infty} v_n (n+1) [\alpha \eta_{s,n}(y) + \beta \eta_{s+1,n}(y)]. \quad (3.15)$$

Figura 3.6: Variância da distribuição EGLFR($a, b, \alpha, \beta, \theta$).(a) $\alpha = 0.2$, $\beta = 0.4$ e $\theta = 0.8$ (b) $\alpha = 0.2$, $\beta = 0.4$ e $\theta = 1.3$ **Fonte:** Próprio autor.

Os momentos incompletos são muito úteis para a construção das curvas de Bonferroni e Lorenz. Para uma dada probabilidade $p > 0$, as curvas de Bonferroni e Lorenz são dadas por

$$B(p) = \frac{m_1[Q(p)]}{p \mu'_1} \quad \text{e} \quad L(p) = \frac{m_1[Q(p)]}{\mu'_1}.$$

Estas curvas são também úteis para obter os índices de concentração de Bonferroni (B) e de Gini (G), que são medidas que quantificam a desigualdade de renda. Os estudos sobre a curva e o índice de Bonferroni tem sido estendidos também para diferentes campos como o bem-estar, pobreza e confiabilidade (GIORGI; NADARAJAH, 2010).

Esses índices são dados por $B = 1 - \int_0^1 B(p) dp$ e $G = 1 - 2 \int_0^1 L(p) dp$. Este último tem uma fórmula simples $G = (2\delta_{1,1} - \mu'_1)/\mu'_1$, que pode ser determinado usando as equações (3.12) e (3.14). O índice B segue das equações (3.12), (3.13) e (3.15).

3.5.5 Estatísticas de ordem

Seja x_1, \dots, x_n uma amostra aleatória da distribuição EGLFR. A fdp da i -ésima estatística de ordem $X_{i:n}$ é $f_{i:n}(x) = [B(i, n - i + 1)]^{-1} f(x) F(x)^{i-1} [1 - F(x)]^{n-i}$, em que $B(\cdot, \cdot)$ denota a função beta. Tomando a fda e fdp nas equações (3.10) e (3.11), respectivamente, $f_{i:n}(x)$ é escrito como

$$f_{i:n}(x) = \sum_{m,j,l=0}^{\infty} \gamma_{m,j,l} h(x; (j+l+m+1)\alpha, (j+l+m+1)\beta), \quad i = 1, 2, \dots, n,$$

em que $\gamma_{m,j,l} = \frac{(j+1)v_j\tau_{i-1,l}\psi_{m,i}}{(j+l+m+1)B(i,n-i+1)}$ e v_j e $\tau_{i-1,l}$ são dados por (3.11) e (3.14), respectivamente, e $\psi_{m,i} = \sum_{k=0}^{n-i} \binom{n}{k} \tau_{k,m}$.

Estatísticas de ordem são muito úteis em estatística não-paramétrica e inferência. Por exemplo, a duração de um teste pode ser extremamente longo, então se torna necessário censurar o teste após a i -ésima falha (LOVRIC, 2011).

3.6 Considerações finais

Neste capítulo, introduz-se um novo modelo de cinco parâmetros, chamado de distribuição taxa de falha linear generalizado estendido, que generaliza onze distribuições. A nova distribuição tem suporte nos reais positivos e é um modelo bem competitivo, não só por generalizar distribuições assim como quando comparadas a outras distribuições também de cinco parâmetros, como provado empiricamente por meio de duas aplicações. Estudou-se algumas de suas propriedades gerais e foram fornecidas expressões explícitas para os momentos ordinários, incompletos e ponderados por probabilidade, as funções geratriz e quantil, além das estatísticas de ordem. Espera-se que o modelo proposto possa atrair aplicações mais amplas, principalmente pelo elenco de onze distribuições que são generalizadas.

Capítulo 4

Considerações finais

Na presente dissertação, foram introduzidas a nova família Burr XII-G (BXII-G) e a nova distribuição para tempo de vida, taxa de falha linear generalizada estendida (EGLFR). Diversas propriedades de ambas foram apresentadas, como momentos ordinários e incompletos, funções geratriz e quantílica, fornecendo diversas expressões explícitas. O método da máxima verossimilhança foi utilizado para a estimação dos parâmetros. Nas duas aplicações mostrou-se a utilidade da família BXII-G por meio das distribuições BXII-Weibull e seu caso particular BXII-exponencial, comparadas a outras distribuições também de quatro e três parâmetros, respectivamente. A nova distribuição EGLFR também mostrou-se útil por meio de duas aplicações, uma delas realizada apenas com a nova distribuição e seus onze casos especiais.

O teste da razão de verossimilhança, além dos critérios de informação de Akaike e Hannan-Quinn foram utilizados para seleção de modelos. As estatísticas mais robustas de Anderson-Darling e Cramér-von Mises também foram utilizadas.

Para trabalho futuros, planeja-se desenvolver distribuições contínuas utilizando a família BXII-G, assim como modelos de regressão log-linear.

Referências

AARSET, M. V. How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, v. 36, n. 1, p. 106–108, 1987.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: *Selected Papers of Hirotugu Akaike*. [S.l.]: Springer, 1998. p. 199–213.

ALEXANDER, C. et al. Generalized beta-generated distributions. *Computational Statistics & Data Analysis*, Elsevier, v. 56, n. 6, p. 1880–1897, 2012.

ALKARNI, S. H. A class of truncated binomial lifetime distributions. *Open Journal of Statistics*, Scientific Research Publishing, v. 3, n. 05, p. 305, 2013.

ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. *Metron*, Springer, v. 71, n. 1, p. 63–79, 2013.

BAIN, L. J. Analysis for the linear failure-rate life-testing distribution. *Technometrics*, Taylor & Francis, v. 16, n. 4, p. 551–559, 1974.

BARRETO-SOUZA, W.; SANTOS, A. H.; CORDEIRO, G. M. The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 80, n. 2, p. 159–172, 2010.

BOURGUIGNON, M.; SILVA, R. B.; CORDEIRO, G. M. The Weibull–G family of probability distributions. *Journal of Data Science*, China Data Mine Association, v. 12, n. 1, p. 53–68, 2014.

BURR, I. W. Cumulative frequency functions. *The Annals of Mathematical Statistics*, JSTOR, v. 13, n. 2, p. 215–232, 1942.

CASELLA, G.; BERGER, R. L. *Statistical inference*. [S.l.]: Duxbury Pacific Grove, CA, 2002.

CHEN, G.; BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, American Society for Quality, v. 27, n. 2, p. 154–161, 1995.

COLLETT, D. *Modelling survival data in medical research*. [S.l.]: CRC press, 2003.

CORDEIRO, G. M.; CASTRO, M. de. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 81, n. 7, p. 883–898, 2011.

- CORDEIRO, G. M.; HASHIMOTO, E. M.; ORTEGA, E. M. The McDonald Weibull model. *Statistics*, Taylor & Francis, v. 48, n. 2, p. 256–278, 2014.
- CORDEIRO, G. M.; LEMONTE, A. J. The McDonald inverted beta distribution. *Journal of the Franklin Institute*, Elsevier, v. 349, n. 3, p. 1174–1197, 2012.
- CORDEIRO, G. M.; NADARAJAH, S.; ORTEGA, E. M. The Kumaraswamy Gumbel distribution. *Statistical Methods & Applications*, Springer, v. 21, n. 2, p. 139–168, 2012.
- CORDEIRO, G. M.; ORTEGA, E. M.; CUNHA, D. C. da. The exponentiated generalized class of distributions. *Journal of Data Science*, China Data Mining Association, v. 11, n. 1, p. 1–27, 2013.
- CORDEIRO, G. M.; ORTEGA, E. M.; LEMONTE, A. J. The exponential-Weibull lifetime distribution. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 84, n. 12, p. 2592–2606, 2014.
- CORDEIRO, G. M.; ORTEGA, E. M.; NADARAJAH, S. The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, Elsevier, v. 347, n. 8, p. 1399–1429, 2010.
- CRUZ, J. N. da et al. A new family of distributions: Theory and applications in survival analysis. 2013. Submitted.
- D'AGOSTINO, R. B. *Goodness-of-fit-techniques*. [S.l.]: CRC press, 1986.
- ELBATAL, I.; CONDINO, F.; DOMMA, F. Reflected generalized beta inverse Weibull distribution: definition and properties. *arXiv preprint arXiv:1309.6108*, 2013.
- ELBATAL, I.; MEROVCI, F.; MARZOUK, W. McDonald generalized linear failure rate distribution. *Pakistan Journal of Statistics and Operation Research*, v. 10, n. 3, p. 267–288, 2014.
- EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 31, n. 4, p. 497–512, 2002.
- FARAWAY, J. et al. *gofest: Classical Goodness-of-Fit Tests for Univariate Distributions*. [S.l.], 2015. R package version 1.0-3. Disponível em: <<http://CRAN.R-project.org/package=gofest>>.
- GIORGI, G. M.; NADARAJAH, S. Bonferroni and Gini indices for various parametric families of distributions. *Metron*, Springer, v. 68, n. 1, p. 23–46, 2010.
- GOMES, A. E.; DA-SILVA, C. Q.; CORDEIRO, G. M. The exponentiated G Poisson model. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 44, n. 20, p. 4217–4240, 2015.
- GOMPERTZ, B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, JSTOR, p. 513–583, 1825.

GREENWOOD, J. A. et al. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, v. 15, n. 5, p. 1049–1054, 1979.

GUPTA, R.; KUNDU, D. Generalized exponential distribution: An alternative to gamma and Weibull distributions. *Biom. J*, v. 43, p. 117–130, 2001.

GUPTA, R. C.; GUPTA, P. L.; GUPTA, R. D. Modeling failure time data by Lehman alternatives. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 27, n. 4, p. 887–904, 1998.

HANNAN, E. J.; QUINN, B. G. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 190–195, 1979.

HENNINGSEN, A.; TOOMET, O. maxlik: A package for maximum likelihood estimation in R. *Computational Statistics*, v. 26, n. 3, p. 443–458, 2011. Disponível em: <<http://dx.doi.org/10.1007/s00180-010-0217-1>>.

HUANG, S.; OLUYEDE, B. O. Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data. *Journal of Statistical Distributions and Applications*, Springer, v. 1, n. 1, p. 1–20, 2014.

JEFFREY, A.; ZWILLINGER, D. *Table of integrals, series, and products*. [S.l.]: Academic Press, 2007.

KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, Elsevier, v. 46, n. 1, p. 79–88, 1980.

KUNDU, D.; RAQAB, M. Z. Generalized Rayleigh distribution: different methods of estimations. *Computational statistics & data analysis*, Elsevier, v. 49, n. 1, p. 187–200, 2005.

LAWLESS, J. F. *Statistical models and methods for lifetime data*. [S.l.]: John Wiley & Sons, 2011.

LEE, C.; FAMOYE, F.; ALZAATREH, A. Y. Methods for generating families of univariate continuous distributions in the recent decades. *Wiley Interdisciplinary Reviews: Computational Statistics*, Wiley Online Library, v. 5, n. 3, p. 219–238, 2013.

LEE, C.; FAMOYE, F.; OLUMOLADE, O. Beta-Weibull distribution: some properties and applications to censored data. *Journal of modern applied statistical methods*, v. 6, n. 1, p. 17, 2007.

LEHMANN, E. L. The power of rank tests. *The Annals of Mathematical Statistics*, JSTOR, p. 23–43, 1953.

LIEBLEIN, J.; ZELEN, M. Statistical investigation of the fatigue life of deep-groove ball bearings. *Journal of Research of the National Bureau of Standards*, Citeseer, v. 57, n. 5, p. 273–316, 1956.

LOVRIC, M. *International encyclopedia of statistical science*. [S.l.]: Springer London, 2011.

- MARINHO, P. R. D.; BOURGUIGNON, M. Package adequacy model. 2012.
- MCDONALD, J. B.; XU, Y. J. A generalization of the beta distribution with applications. *Journal of Econometrics*, Elsevier, v. 66, n. 1, p. 133–152, 1995.
- MOOD, A.; GRAYBILL, F.; BOES, D. *Introduction to the Theory of Statistics*. [S.l.]: McGraw-Hill, New York, NY, 1974.
- MUDHOLKAR, G. S.; SRIVASTAVA, D. K. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, Institute of Electrical and Electronics Engineers, v. 42, n. 2, 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*, Taylor & Francis, v. 37, n. 4, p. 436–445, 1995.
- NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. The exponentiated G geometric family of distributions. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 85, n. 8, p. 1634–1650, 2015.
- NADARAJAH, S.; KOTZ, S. The exponentiated Fréchet distribution. *Interstat Electronic Journal*, v. 312001, 2003.
- NADARAJAH, S.; KOTZ, S. The beta exponential distribution. *Reliability engineering & system safety*, Elsevier, v. 91, n. 6, p. 689–697, 2006.
- NADARAJAH, S.; KOTZ, S. The exponentiated type distributions. *Acta Applicandae Mathematica*, Springer, v. 92, n. 2, p. 97–111, 2006.
- PARANAÍBA, P. F. et al. The beta Burr XII distribution with application to lifetime data. *Computational Statistics & Data Analysis*, Elsevier, v. 55, n. 2, p. 1118–1136, 2011.
- PARANAÍBA, P. F. et al. The Kumaraswamy Burr XII distribution: theory and practice. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 83, n. 11, p. 2117–2143, 2013.
- PASCOA, M. A. de; ORTEGA, E. M.; CORDEIRO, G. M. The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical methodology*, Elsevier, v. 8, n. 5, p. 411–433, 2011.
- PEARSON, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, JSTOR, p. 71–110, 1894.
- PROSCHAN, F. Theoretical explanation of observed decreasing failure rate. *Technometrics*, Taylor & Francis Group, v. 5, n. 3, p. 375–383, 1963.
- PRUDNIKOV, A.; BRYCHKOV, Y. A.; MARICHEV, O. Integrals and series, 1986. *Gordon and Breach Science Publishers, Amsterdam*, 1986.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2015. Disponível em: <<https://www.R-project.org/>>.
- RAMOS, M. W. A. et al. The Kumaraswamy G Poisson family of distributions. *Journal of Statistical Theory and Applications*, v. 14, n. 3, p. 222–239, 2015.

- RESEARCH, I. W. *Mathematica*. 2012. Wolfram Research, Inc.
- SANTOS, R. G.; CORDEIRO, G. M. Extended generalized linear failure rate distribution: properties and applications. *Journal of Applied Statistics*, p. 14, 2015. Submitted.
- SANTOS, R. G.; CUNHA, E. T. da; CORDEIRO, G. M. The Burr XII-G family of distribution: properties and applications. *Journal of Data Science*, p. 18, 2015. Submitted.
- SARHAN, A. M.; KUNDU, D. Generalized linear failure rate distribution. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 38, n. 5, p. 642–660, 2009.
- SCRUCCA, L. GA: A package for genetic algorithms in R. *Journal of Statistical Software*, v. 53, n. 4, p. 1–37, 2013. Disponível em: <<http://www.jstatsoft.org/v53/i04/>>.
- SILVA, G. O.; ORTEGA, E. M.; CORDEIRO, G. M. The beta modified Weibull distribution. *Lifetime Data Analysis*, Springer, v. 16, n. 3, p. 409–430, 2010.
- SURLES, J.; PADGETT, W. Inference for reliability and stress-strength for a scaled Burr type X distribution. *Lifetime Data Analysis*, Springer, v. 7, n. 2, p. 187–200, 2001.
- TAHIR, M. et al. The logistic-X family of distributions and its applications. *Commun. Stat. Theory Methods (2015a)*. *forthcoming*, 2015.
- TAHIR, M. et al. The Poisson-X family of distributions. *Journal of Statistical Computation and Simulation*, Taylor & Francis, p. 1–21, 2016.
- TAHIR, M. et al. A new Weibull-G family of distributions. *Hacet. J. Math. Stat.(2015b)*. *forthcoming*, 2014.
- VERHULST, P.-F. Notice sur la loi que la population suit dans son accroissement. correspondance mathématique et physique publiée par a. *Quetelet*, v. 10, p. 113–121, 1838.
- XIANG, Y. et al. Generalized simulated annealing for global optimization: the GenSA package. *The R Journal Volume 5(1):13-29, June 2013*, 2013. Disponível em: <<http://journal.r-project.org/archive/2013-1/xiang-gubian-suomela-et-al.pdf>>.

Apêndice A

Apêndice

Neste apêndice serão fornecidas as demonstrações para a obtenção de misturas, ou seja, combinações lineares de funções de distribuição acumulada (cdfs) ou de funções densidade de probabilidade (pdfs) da distribuição exponencializadas-G (exp-G). A cdf da distribuição exp-G, com parâmetro de potência i , é definida por $H_i(x) = G(x)^i$ e sua pdf é $h_i(x) = i g(x) G(x)^{i-1}$. Inicialmente, algumas expansões úteis no desenvolvimento das misturas deverão ser apresentadas.

A expansão binomial generalizada é definida como

$$(1+x)^q = \sum_{k=0}^{\infty} \frac{\Gamma(q+1)}{\Gamma(q-k+1) k!} x^k, \quad |x| < 1 \text{ e } q \in \{\mathbb{R} - \mathbb{N}\}, \quad (\text{A.1})$$

sendo que para $x = 1$ a série converge para $q > -1$. Para $q \in \mathbb{N}$, utiliza-se a expansão binomial

$$(1+x)^q = \sum_{k=0}^q \frac{q!}{(q-k)! k!} x^k, \quad (\text{A.2})$$

Outro resultado útil é a identidade da série de potência elevada a um número natural n

$$\left(\sum_{k=0}^{\infty} a_k x^k \right)^n = \sum_{k=0}^{\infty} c_{n,k} x^k, \quad (\text{A.3})$$

em que $c_{n,0} = a_0^n$ e $c_{n,k} = (k a_0)^{-1} \sum_{j=1}^k (j n - k + j) a_j c_{n,k-j}$ (para $k \geq 1$).

A identidade a seguir resume a razão de duas séries de potências em uma única série

$$\frac{\sum_{i=0}^{\infty} b_i (z - z_0)^i}{\sum_{i=0}^{\infty} c_i (z - z_0)^i} = \sum_{i=0}^{\infty} d_i (z - z_0)^i, \quad (\text{A.4})$$

em que d_i é obtido recursivamente $\sum_{i=0}^k c_i d_{k-i} = b_k$, $k = 0, 1, 2, \dots$, cuja convergência ocorre para $|z - z_0| < r$, $r > 0$.

A.1 Reescrevendo a cdf da Burr XII-G

Nesta seção apresentaremos os passos para a obtenção mistura de exp-G para a cdf da família de distribuição Burr XII-G (BXII-G).

Podemos reescrever a equação (2.2) como

$$F(x) = 1 - \left[\frac{\overline{G}(x)^a}{\overline{G}(x)^a + G(x)^a} \right]^b.$$

Usando (A.1) nos termos da última equação, tem-se

$$[1 - G(x)]^a = \sum_{i=0}^{\infty} w_i(a) G(x)^i$$

e

$$G(x)^a = \sum_{n=0}^{\infty} u_n(a) G(x)^n,$$

em que $w_i(a) = (-1)^i \binom{a}{i}$ e $u_n(a) = \sum_{j=n}^{\infty} (-1)^{n+j} \binom{a}{j} \binom{j}{n}$. Portanto, o denominador é reescrito como

$$[1 - G(x)]^a + G(x)^a = \sum_{n=0}^{\infty} q_n(a) G(x)^n,$$

em que $q_n(a) = w_n(a) + u_n(a)$. Logo,

$$F(x) = 1 - \left[\frac{\sum_{n=0}^{\infty} w_n(a) G(x)^n}{\sum_{n=0}^{\infty} q_n(a) G(x)^n} \right]^b.$$

Usando (A.4), reescrevemos a última equação como

$$F(x) = 1 - \left[\sum_{i=0}^{\infty} d_i(a) G(x)^i \right]^b,$$

em que $d_i(a)$ é obtido de $\sum_{i=0}^k q_i(a) d_{k-i}(a) = w_i(a)$, $k = 0, 1, 2, \dots$

Utilizando novamente (A.1) na série de potência acima, considerando $x = G(x) - 1$, tem-se

$$\left[\sum_{i=0}^{\infty} d_i(a) G(x)^i \right]^b = 1 + \sum_{n=1}^{\infty} \frac{\tau_n}{n!} \left[\sum_{i=0}^{\infty} d_i(a) G(x)^i - 1 \right]^n.$$

em que $\tau_n = \tau_n(b) = \prod_{j=0}^{n-1} (b - j)$. Reescrevendo $l_i(a) = d_i(a) - 1$, para $i = 0$ e $l_i(a) = d_i(a)$, $\forall i \geq 1$ e usando a identidade (A.3)

$$F(x) = 1 - \left[1 + \sum_{i=0}^{\infty} \sum_{n=1}^{\infty} \frac{\tau_n}{n!} \mu_{i,n}(a) G(x)^i \right] = \sum_{i=0}^{\infty} \nu_i(a, b) G(x)^i,$$

em que $\nu_i = \nu_i(a, b) = -\sum_{n=1}^{\infty} \frac{\tau_n}{n!} \mu_{i,n}(a)$, $\mu_{i,n}(a) = (n l_0)^{-1} \sum_{p=1}^n [p(i+1) - n] l_p \mu_{i,n-p}$, para $i \geq 0$ e $\mu_{i,0} = \mu_{i,0}(a) = l_0^i$.

Logo, obtemos a cdf da família BXII-G como uma mistura de cdfs exp-G. De fato, podemos expressar

$$F(x) = \sum_{i=0}^{\infty} \nu_i H_i(x), \quad (\text{A.5})$$

em que $H_i(x) = H_i(x; \eta) = G(x)^i$.

A.2 Estatísticas de ordem da Burr XII-G

Podemos demonstrar que a pdf das estatísticas de ordem da distribuição BXII-G é uma mistura dupla de pdfs da distribuição exponencializadas-G. A densidade das estatísticas de ordem segue de (2.14)

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

Usando (A.2) no termo $[1 - F(x)]^{n-i}$

$$[1 - F(x)]^{n-i} = \sum_{k=0}^{n-i} w_k F(x)^k$$

em que $w_k = (-1)^k \binom{n-i}{k}$. A cdf $F(x)$ foi definida em (A.5). Usando (A.3), podemos escrever

$$\left[\sum_{j=0}^{\infty} w_j H_j(x) \right]^{k+i-1} = \sum_{j=0}^{\infty} c_{k+i-1,j} H_j(x),$$

em que $c_{k+i-1,j} = (j w_0)^{-1} \sum_{p=1}^j [p(k+i) - j] w_p c_{k+i-1,j-p}$ e $c_{k+i-1,0} = w_0^i$. Então, equação (2.14) se reduz a

$$f_{i:n}(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{n-i} \frac{n! c_{k+i-1,j}}{(i-1)!(n-i)!} f(x) H_j(x).$$

Como $f(x)$ também é definida como uma mistura a partir da derivação de (A.5), nós temos

$$f_{i:n}(x) = \sum_{j,l=0}^{\infty} \sum_{k=0}^{n-i} \frac{n! c_{k+i-1,j}}{(i-1)!(n-i)!} (l+1) g(x) G(x)^{l+j}.$$

Finalmente, equação (2.14) é reescrito como

$$f_{i:n}(x) = \sum_{j,l=0}^{\infty} s_{j,l} h_{l+j+1}(x),$$

em que $s_{l,j} = \left(\frac{l+1}{l+j+1} \right) \frac{n! w_{l+1}}{(i-1)!(n-i)!} \sum_{k=0}^{n-i} c_{k+i-1,j}$.