



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

Faculdade de Ciências e Tecnologia
Câmpus de Presidente Prudente

Nathan Cordeiro Ginak

Família de distribuições de probabilidade Gama Exponenciada- X e suas propriedades

Presidente Prudente
2025

Nathan Cordeiro Ginak

Família de distribuições de probabilidade Gama Exponenciada-X e suas propriedades

Relatório Final para Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística da FCT/Unesp para aproveitamento na disciplina TCC 2

Orientador:

Prof. Dr. Guilherme Aparecido Santos Aguilar

**Presidente Prudente
2025**

G492f Ginak, Nathan Cordeiro

Família de distribuições de probabilidade Gama Exponenciada-X e suas propriedades / Nathan Cordeiro Ginak. -- Presidente Prudente, 2025

60 p. : tabs.

Trabalho de conclusão de curso (Bacharelado - Estatística) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Guilherme Aparecido Santos Aguilar

1. Distribuições de Probabilidade. 2. Análise de Sobrevida. 3. Estimativa de Máxima Verossimilhança. I. Título.

TERMO DE APROVAÇÃO

Nathan Cordeiro Ginak

Família de distribuições de probabilidade Gama Exponenciada-X e suas propriedades

Relatório de Final de Trabalho de Conclusão de Curso aprovado como requisito para obtenção de créditos na disciplina Trabalho de Conclusão do curso de graduação em Estatística da Faculdade de Ciências e Tecnologia da Unesp, pela seguinte banca examinadora:

 Documento assinado digitalmente
GUILHERME APARECIDO SANTOS AGUILAR
Data: 02/12/2025 09:16:37-0300
Verifique em <https://validar.iti.gov.br>

Orientador: _____

Prof. Dr. Guilherme Aparecido Santos Aguilar
Departamento de Estatística

 Documento assinado digitalmente
FERNANDO ANTONIO MOALA
Data: 02/12/2025 22:34:35-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Fernando Antônio Moala
Departamento de Estatística

Presidente Prudente, 02 de dezembro de 2025.

Agradecimentos

Dedico esse trabalho à minha mãe Maria Cristina Cordeiro e minha avó Carmelita Maria Cordeiro, que sempre me incentivaram nos estudos, me ensinaram a ser uma pessoa honesta e ética e foram o meu porto seguro nos tempos de dificuldade. De onde estão, espero que estejam orgulhosas de mim.

Agradeço ao meu orientador professor Guilherme Aparecido Santos Aguilar, por toda sua ajuda durante a realização deste trabalho. Sem sua ajuda, esse trabalho não existiria.

À minha namorada Fernanda dos Santos Bregolim, por ser minha parceira em todos os momentos desde que estamos juntos.

Aos meus colegas de faculdade, pelo companheirismo ao longo desses quatro anos.

Resumo

Este trabalho tem como objetivo apresentar a família de distribuições de probabilidade Gama Exponenciada-X, gerada através do método T-X, com o intuito de proporcionar distribuições de probabilidade mais flexíveis para a modelagem estatística. Primeiramente serão apresentadas as principais características da família Gama Exponenciada-X, como as funções acumulada, densidade, função de sobrevivência e de risco, além da expansão em séries da distribuição, com o intuito de facilitar a dedução de propriedades importantes. Em seguida, conceitos estatísticos importantes serão apresentados, como função quantil, momentos, função geratriz de momentos, estatística de ordem, Entropia de Rényi e estimadores de máxima verossimilhança para a família. Posteriormente será analisado um caso particular desta família de distribuições, onde será feita a estimação dos parâmetros por meio do método de máxima verossimilhança, e também ajustado a dois conjunto de dados distintos, com o intuito de verificar sua performance comparado a outros modelos já existentes.

Palavras-chave: Distribuições de Probabilidade, Método T-X, Gama Exponenciada, Propriedades estatísticas

Abstract

This work aims to present the Exponentiated Gamma-X family of probability distributions, generated through the T-X method, with the goal of providing more flexible probability distributions for statistical modeling. First, the main characteristics of the Exponentiated Gamma-X family will be presented, such as the cumulative distribution function, density function, survival and hazard functions, as well as the series expansion of the distribution, in order to facilitate the derivation of important properties. Next, key statistical concepts will be introduced, including the quantile function, moments, moment-generating function, order statistics, Rényi entropy, and maximum likelihood estimators for the family. Subsequently, a particular case of this family of distributions will be analyzed, in which parameter estimation will be carried out using the maximum likelihood method, and the model will be fitted to two distinct datasets to assess its performance compared to other existing models.

Keywords: Probability distributions, T-X method, Exponentiated gamma, statistical properties

Lista de Figuras

4.1	Gráficos da função densidade da GE-W com diferentes parâmetros	22
4.2	Gráficos da função de risco da GE-W com diferentes parâmetros.	22
4.3	Gráficos da função densidade da GE-Lom com diferentes parâmetros. . .	23
4.4	Gráficos da função de risco da GE-Lom com diferentes parâmetros. . . .	24
5.1	Gráficos da função densidade da GE-C com diferentes parâmetros	26
5.2	Função de risco da distribuição GE-C para diferentes combinações de parâmetros com comportamento crescente.	27
5.3	Função de risco da distribuição GE-C para diferentes combinações de parâmetros com comportamento decrescente.	27
5.4	Função de risco da distribuição GE-C para diferentes combinações de parâmetros com formato unimodal.	28
5.5	Função de risco da distribuição GE-C para diferentes combinações de parâmetros com formato em curva da banheira.	28
5.6	Superfície da assimetria da distribuição GE-C em função de θ e β , com $\lambda = 1.2$ fixo.	30
5.7	Superfície da curtose da distribuição GE-C em função de θ e β , com $\lambda = 1.2$ fixo.	31
5.8	Superfície da assimetria da distribuição GE-C em função de θ e λ , com $\beta = 0.8$ fixo.	31
5.9	Superfície da curtose da distribuição GE-C em função de θ e λ , com $\beta = 0.8$ fixo.	32
5.10	Superfície da assimetria da distribuição GE-C em função de β e λ , com $\theta = 1.5$ fixo.	32
5.11	Superfície da curtose da distribuição GE-C em função de β e λ , com $\theta = 1.5$ fixo.	33

5.12 Curva de sobrevivência empírica (Kaplan–Meier) e funções de sobrevivência ajustadas das distribuições testadas para o primeiro conjunto de dados.	39
5.13 Curva de sobrevivência empírica (Kaplan–Meier) e funções de sobrevivência ajustadas das distribuições testadas para o segundo conjunto de dados.	40

Lista de Tabelas

3.1	Tabela de funções que podem ser utilizadas no método T-X	5
5.1	Momentos principais da GE-C para diferentes valores de θ , β e λ	30
5.2	Médias, vieses, variâncias, MSE, MAE e RMSE dos EMV de θ , β e λ da distribuição GE-C para $(\theta, \beta, \lambda) = (1.5, 0.6, 2.3)$ em diferentes tamanhos amostrais.	36
5.3	Médias, vieses, variâncias, MSE, MAE e RMSE dos EMV de θ , β e λ da distribuição GE-C para $(\theta, \beta, \lambda) = (0.8, 1.6, 0.7)$ em diferentes tamanhos amostrais (segunda simulação).	37
5.4	Critérios de ajuste para o conjunto de dados 1	39
5.5	Critérios de ajuste para o conjunto de dados 2	40

Lista de Abreviações

AIC *Akaike Information Criterion*

AICc *Akaike Information Criterion Corrected*

BIC *Bayesian Information Criterion*

EMV Estimador de Máxima Verossimilhança

GE-C Gama Exponenciada-Chen

GE-Lom Gama Exponenciada-Lomax

GE-W Gama Exponenciada-Weibull

GE-X Gama Exponenciada-X

Sumário

Lista de Abreviações	xi
1 Introdução	1
2 Referencial Teórico	4
3 Metodologia	5
4 Família Gama Exponenciada-X	7
4.1 Resultados Iniciais	7
4.2 Expansão em Séries	9
4.3 Propriedades Estatísticas	10
4.3.1 Quantis	11
4.3.2 Momentos e função geratriz de momentos	11
4.3.3 Desvios Médios	14
4.3.4 Entropia de Rényi	15
4.3.5 Estatística de ordem	16
4.3.6 Estimação dos Parâmetros	17
4.4 Exemplos de distribuições	20
4.4.1 Gama Exponenciada-Weibull	20
4.4.2 Gama Exponenciada-Lomax	22
5 Gama Exponenciada-Chen	25
5.1 Principais funções	25
5.2 Momentos	29
5.3 Simulação	33
5.4 Aplicação	35

5.4.1	Primeiro <i>Dataset</i>	39
5.4.2	Segundo <i>Dataset</i>	40
6	Conclusão	42
6.1	Discussão dos resultados	42
6.2	Considerações finais	43
	Referências Bibliográficas	45

1 Introdução

Uma distribuição de probabilidade é definida como uma função que descreve o comportamento de uma variável aleatória, atribuindo probabilidades aos possíveis resultados de um experimento aleatório. Para o caso de variáveis aleatórias contínuas, essa distribuição por uma função densidade de probabilidade (fdp), cuja integral em um intervalo fornece a probabilidade de a variável assumir valores que estão contidos dentro deste intervalo. Em decorrência disso, distribuições de probabilidades, em especial as contínuas, sempre foram essenciais para a modelagem estatística de eventos do mundo real. Contudo, nem todas as distribuições mais conhecidas na literatura são suficientes para este fim, por não serem flexíveis o suficiente para se ajustarem aos padrões complexos de determinados eventos.

Para suprir essa necessidade, diversos métodos para a criação de novas distribuições de probabilidade surgiram, muitos deles se baseando na adição de mais parâmetros a uma distribuição já conhecida, como proposto por Marshall e Olkin (1997), que formulou o seguinte modelo para adição de novos parâmetros a uma distribuição de probabilidade

$$\overline{F(x)} = \frac{\alpha G(x)}{G(x) + \alpha \overline{G(x)}} \quad \alpha > 0$$

na qual $G(x)$ é uma função densidade acumulada (fda) de uma variável aleatória X .

De maneira, semelhante, Gupta et. al. (1998) propuseram a família de distribuições exponenciadas, definida por

$$F(x) = [G(x)]^\theta \quad \theta > 0$$

Outro método utilizado para criação de novas distribuições de probabilidade são aqueles que consistem em combinar distribuições já existentes. O exemplo mais notável e um dos pioneiros desta tática foi o modelo beta-G, proposto por Eugene et.al (2002),

que é dada por

$$F(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^{G(x)} t^{\alpha-1} (1-t)^{\beta-1} dt \quad \alpha, \beta > 0$$

A partir disso, novos métodos semelhantes surgiram, como o Gamma-G tipo 1 por Zografos e Balakrishnan (2009), na qual teve suas principais propriedades estudadas por Nadarajah et al. (2015), a Kumaraswamy-G por Cordeiro e Castro (2011), a nova Gamma-G por Ristić e Balakrishnan(2012), a classe exponencial generalizada por Cordeiro et al.(2013), o modelo Weibull-G proposto por Bourguignon et al. (2014), a Lomax-G por Cordeiro et al. (2014), Gamma-X por Alzaatreh et al. (2014), Weibull-X por Alzaatreh e Ghosh (2015), o modelo transmutado exponenciado generalizado-G por Yousof et al. (2015), a nova Weibull-G por Tahir et. al (2016), a Gama-Weibull-G por Oluyede et al. (2018), a nova Weibull-X por Ahmad et al. (2018), a Kumaraswamy-G generalizada por Nofal et al. (2018), a Chen-G por Anzagra et al. (2020), a Zubair-G por Ahmad (2020), a odd Weibull Inversa-G por Thomas et al. (2020), a nova família de distribuições generalizadas de Salahuddin et al. (2021), a Gamma-X por Pourezza et al.(2021), a nova Lomax-G por Sapkota et al. (2023) e Lomax-X deslocada por Atchadé et al. (2024).

Com base nessas referências, o intuito deste trabalho é propor a nova família de distribuições de probabilidade denominada Gama Exponenciada-X (GE-X), que será obtida utilizando a distribuição Gama Exponenciada proposta por Gupta et.al no método T-X, proposto por Alzaatreh, Lee e Famoye (2013).

A partir deste método, será abordada algumas propriedades da nova família de distribuições, como sua função densidade, acumulada, de sobrevivência e de risco, além de propriedades estatísticas importantes como momentos, função geratriz, estatística de ordem e entropia de Rényi, além de algumas distribuições que pertencem a família criada. Posteriormente, uma distribuição desta família será apresentada, a Gama Exponenciada-Chen, na qual veremos algumas de suas características, além de simulações de Monte Carlo para verificar as propriedades assintóticas de seus estimadores de máxima verossimilhança. Após isso, será feito o ajuste desta distribuição a dois

conjuntos de dados, de maneira a avaliar seu desempenho comparado a outros modelos existentes. Espera-se que a distribuição Gama Exponenciada-Chen apresente maior flexibilidade e métricas superiores, possibilitando mais uma alternativa a modelagem estatística.

2 Referencial Teórico

Os materiais estudados para a formulação deste trabalho englobam basicamente o campo de distribuições de probabilidade e métodos de estimação necessários para a aplicação.

Para o estudo da distribuição Gama Exponenciada, foi analisado o artigo “Exponentiated Gamma Distribution: Different Methods of Estimation”, de Ibrahim e Bakoban, (2012), já que os autores apresentam a distribuição proposta de maneira detalhada, apresentando suas propriedades estatísticas mais detalhadamente e propondo diferentes métodos para a estimação de seus parâmetros. Também foi analisada a dissertação "Família Distribuição Gama Exponenciada" de Aguilar (2017) e do artigo "Zero-Truncated Poisson Exponentiated Gamma Distribution: Application and Estimation Methods" por Aguilar et al. (2019), para visualização do comportamento da Gama Exponenciada na criação de novas distribuições de probabilidade, além de melhor entendimento do processo de simulação estatística. Outra referência utilizada foi o livro "Statistical Inference", de Casella e Berger (2002, para revisão de conceitos importantes de inferência estatística, principalmente no que tange o processo de estimação de parâmetros

Para melhor entendimento do método T-X, proposto para a criação da GE-X, foi analisado o artigo “A new method for generating families of continuous distribution”, por Alzaatreh, Lee e Famoye (2013), onde é dada toda a fundamentação teórica do método. Por fim, para verificar uma aplicação do método T-X, foi feita a leitura do artigo “On the Weibull-X family of distributions” (Alzaatreh e Ghosh, 2015), onde também são abordadas mais detalhadamente propriedades como entropia, função geratriz de momentos e estatísticas de ordem e o processo para simulação de variáveis aleatórias e estimação de parâmetros.

3 Metodologia

A família de distribuições será criada através do método T-X, proposto por Alzaatreh, Lee e Famoye (2013).

Seja $r(t)$ a função densidade de probabilidade (fdp) da variável aleatória T , onde $T \in [m, n]$, e $W(F(x; \boldsymbol{\xi}))$ uma função da fda $F(x; \boldsymbol{\xi})$ de uma variável aleatória qualquer X com vetor de parâmetros $\boldsymbol{\xi}$, de maneira que $W(F(x; \boldsymbol{\xi}))$ satisfaça as seguintes condições:

$$\begin{cases} W(F(x; \boldsymbol{\xi})) \in [m, n] \\ W(F(x; \boldsymbol{\xi})) \text{ é diferenciável e monótona não decrescente} \\ W(F(x; \boldsymbol{\xi})) \rightarrow m \text{ quando } x \rightarrow -\infty \text{ e } W(F(x; \boldsymbol{\xi})) \rightarrow n \text{ quando } x \rightarrow \infty. \end{cases}$$

Essa função pode variar, a depender do domínio da variável aleatória T . Abaixo está listado alguns exemplos de funções que podem ser utilizadas.

Domínio de T	$W(F(x))$
$[0, 1]$	$F(x)$
$[0, 1]$	$F^\alpha(x)$
$[m, \infty), \quad m \geq 0$	$-\ln [1 - F(x)]$
$[m, \infty), \quad m \geq 0$	$\frac{F(x)}{1-F(x)}$
$[m, \infty), \quad m \geq 0$	$-\ln [1 - F^\alpha(x)]$
$[m, \infty), \quad m \geq 0$	$\frac{F^\alpha(x)}{1-F^\alpha(x)}$
$(-\infty, \infty)$	$\ln [-\ln (1 - F(x))]$
$(-\infty, \infty)$	$\ln \left[\frac{F(x)}{1-F(x)} \right]$
$(-\infty, \infty)$	$\ln [-\ln (1 - F^\alpha(x))]$
$(-\infty, \infty)$	$\ln \left[\frac{F^\alpha(x)}{1-F^\alpha(x)} \right]$

Tabela 3.1: Tabela de funções que podem ser utilizadas no método T-X

Dessa forma, a fda do modelo T-X é definida como

$$G(x) = \int_m^{W(F(x;\boldsymbol{\xi}))} r(t) dt = R\{W(F(x;\boldsymbol{\xi}))\} \quad (3.1)$$

onde $R(t)$ representa a fda da variável aleatória T .

Consequentemente, pela regra da cadeia, temos que a fdp pode ser obtida por

$$g(x) = \left\{ \frac{d}{dx} W(F(x;\boldsymbol{\xi})) \right\} r\{W(F(x;\boldsymbol{\xi}))\} \quad (3.2)$$

Para a criação da família Gama Exponenciada-X, adotaremos que T segue distribuição Gama Exponenciada, além de $W(F(x;\boldsymbol{\xi})) = -\ln[1 - F(x;\boldsymbol{\xi})]$, que se trata da função $H(x)$, ou seja, a função risco acumulada da variável aleatória X com fda $F(x)$, definida como $H(x;\boldsymbol{\xi}) = \int h(x;\boldsymbol{\xi}) dx$, onde $h(x;\boldsymbol{\xi})$ é a função risco.

Posteriormente, será analisada a distribuição Gama Exponenciada-Chen, gerada pelo método proposto, tomando $F(x)$ como a função acumulada da distribuição Chen, proposta inicialmente por Chen (2000). A análise será composta da geração de amostras aleatórias para a distribuição, simulações de Monte Carlo para testar as propriedades dos estimadores de máxima verossimilhança da distribuição e ajuste a dois conjuntos de dados, onde serão avaliadas métricas como AIC, AICc, BIC e p-valor do teste de razão de máxima verossimilhança. O ajuste será realizado através do método de máxima verossimilhança.

A geração de amostras aleatórias desta distribuição, o processo de estimação dos parâmetros e o ajuste desta distribuição a um conjunto de dados será realizada através de algoritmos computacionais, feitos no software R com auxílio dos pacotes *Newdistns*, desenvolvido por Nadarajah e Rocha (2016) para simulação de variáveis e *fitdistrplus*, desenvolvido por Delignette-Muller e Dutang (2015) para o processo de estimação dos parâmetros e ajuste da distribuição.

4 Família Gama Exponenciada-X

4.1 Resultados Iniciais

Em análise de sobrevivência, o principal objetivo para novas criações de distribuições de probabilidade é a criação de funções de risco bastante flexíveis, ou seja, que podem apresentar diferentes formas a depender do valor de seus parâmetros. Baseando-se nisso, Gupta et al. (1998) propuseram a distribuição Gama Exponenciada, que apresentou maior maleabilidade para ajustar-se a diferentes formatos de dados, notadamente por permitir funções de risco com comportamento tanto monótono quanto não monótono.

Sejam $V(x) = 1 - e^{-x}(x + 1)$ e $v(x) = xe^{-x}$, respectivamente, as funções densidade acumulada e densidade de probabilidade da distribuição Gama com parâmetros $\alpha = 2$ e $\beta = 1$. A distribuição Gama Exponenciada é definida pela fda:

$$R(x; \theta) = [V(x)]^\theta = [1 - e^{-x}(x + 1)]^\theta \quad \theta, x > 0 \quad (4.1)$$

e função densidade de probabilidade

$$r(x; \theta) = \theta xe^{-x}[1 - e^{-x}(x + 1)]^{\theta-1} \quad \theta, x > 0 \quad (4.2)$$

Para obter a família Gama Exponenciada-X, aplicaremos a equação (4.2) no método T-X, definido pela equação (3.1), com função $W(F(x; \xi)) = -\ln[1 - F(x; \xi)] = -\ln \overline{F(x; \xi)} = H(x; \xi)$.

Desta maneira, temos que a fda e a fdp da família GE-X são dadas, respectivamente, por

$$G(x; \theta, \xi) = [1 - (1 - F(x; \xi))(1 - \ln \overline{F(x; \xi)})]^\theta \quad (4.3)$$

e

$$g(x; \theta, \boldsymbol{\xi}) = \theta f(x; \boldsymbol{\xi}) \{-\ln \overline{F(x; \boldsymbol{\xi})}\} [1 - (1 - F(x; \boldsymbol{\xi}))(1 - \ln \overline{F(x; \boldsymbol{\xi})})]^{\theta-1} \quad (4.4)$$

Com isso, podemos também chegar nas expressões da função de sobrevivência $S(x)$ e da função de risco $h(x)$, definidas abaixo.

$$S(x; \theta, \boldsymbol{\xi}) = 1 - [1 - (1 - F(x; \boldsymbol{\xi}))(1 - \ln \overline{F(x; \boldsymbol{\xi})})]^\theta \quad (4.5)$$

e

$$h(x; \theta, \boldsymbol{\xi}) = \frac{\theta f(x; \boldsymbol{\xi}) \{-\ln \overline{F(x; \boldsymbol{\xi})}\} [1 - (1 - F(x; \boldsymbol{\xi}))(1 - \ln \overline{F(x; \boldsymbol{\xi})})]^{\theta-1}}{1 - [1 - (1 - F(x; \boldsymbol{\xi}))(1 - \ln \overline{F(x; \boldsymbol{\xi})})]^\theta} \quad (4.6)$$

Podemos também escrever (4.3) e (4.4) em termos das funções risco e risco acumulada, como

$$G(x; \theta, \boldsymbol{\xi}) = [1 - e^{-H(x; \boldsymbol{\xi})}(H(x; \boldsymbol{\xi}) + 1)]^\theta \quad (4.7)$$

e

$$g(x; \theta, \boldsymbol{\xi}) = \theta h(x; \boldsymbol{\xi}) H(x; \boldsymbol{\xi}) e^{-H(x; \boldsymbol{\xi})} [1 - e^{-H(x; \boldsymbol{\xi})}(H(x; \boldsymbol{\xi}) + 1)]^{\theta-1} \quad (4.8)$$

respectivamente.

Da mesma forma, podemos escrever a função de risco da GE-X como

$$h(x; \theta, \boldsymbol{\xi}) = h_F(x; \boldsymbol{\xi}) h_R(H(x; \boldsymbol{\xi})) \quad (4.9)$$

onde h_F se trata da função de risco da variável aleatória X com fda $F(x; \boldsymbol{\xi})$ e h_R é a função de risco da distribuição Gama Exponenciada.

Se $\theta = 1$, chegamos a família de distribuições Gamma-X, conforme proposto por Pourezza (2021), com parâmetros $\alpha = 2$ e $\beta = 1$.

Proposição 1. *Se uma variável aleatória T tem distribuição Gama Exponenciada com parâmetro θ , então a variável aleatória $X = F^{-1}(1 - e^T)$ segue a distribuição Gama Exponenciada-X.*

Demonstração. Conforme proposto por Alzaatreh, Lee e Famoye (2013), como temos

que $G(x) = R\{-\ln(1 - F(x))\}$, tomando a sua função inversa, chegamos na relação entre as variáveis aleatórias X e T apresentadas. $\square$

Esta proposição nos dá uma maneira mais simples de simular uma amostra aleatória da distribuição Gama Exponenciada- X , onde simulamos primeiramente uma amostra aleatória T de uma variável aleatória com distribuição Gama Exponenciada, e então calculando $X = F^{-1}(1 - e^{-T})$, que segue distribuição Gama Exponenciada- X . Essa estratégia será utilizada posteriormente para a simulação de uma distribuição de probabilidade da família proposta.

4.2 Expansão em Séries

Para derivar determinadas propriedades estatísticas, a expansão em séries da função densidade será importante, proporcionando expressões mais simples a serem trabalhadas.

Teorema 1. *Seja a variável aleatória X com distribuição Gama Exponenciada- X . Podemos reescrever sua função densidade (4.4) como uma combinação linear de funções densidade da família exponenciada, de maneira que*

$$g(x; \theta, \xi) = \theta(-\ln \overline{F(x; \xi)}) \sum_{i,j \geq 0} \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} (1 - \ln \overline{F(x; \xi)})^i \pi(x) \quad (4.10)$$

onde $\pi(x) = (j+1)f(x; \xi)(F(x; \xi))^j$ representa a função densidade da distribuição Exponenciada- G , com função geradora $F(x; \xi)$, e parâmetro $\alpha = j+1$.

Demonstração. A expansão binomial generalizada é definida por

$$(1-z)^a = \sum_{m=0}^{\infty} (-1)^m \binom{a}{m} z^m$$

de maneira que $a > 0$ e $|z| < 1$. Dessa forma, aplicando-a em (4.4), de forma que $z = [e^{\ln \overline{F(x; \xi)}}(1 - \ln \overline{F(x; \xi)})] = [(1 - F(x; \xi))(1 - \ln \overline{F(x; \xi)})]$ e $a = \theta - 1$, chegamos em

$$g(x; \theta, \xi) = \theta f(x; \xi) (-\ln \overline{F(x; \xi)}) \sum_{i=0}^{\infty} (-1)^i \binom{\theta-1}{i} (1 - F(x; \xi))^i (1 - \ln \overline{F(x; \xi)})^i$$

Aplicando a expansão novamente, agora para $z = F(x)$ e $a = i$, reduzimos a série para

$$g(x; \theta, \xi) = \theta f(x; \xi) (-\ln \overline{F(x; \xi)}) \sum_{i,j \geq 0} (-1)^{i+j} \binom{i}{j} \binom{\theta-1}{i} (F(x; \xi))^j (1 - \ln \overline{F(x; \xi)})^i$$

Conforme proposto por Gupta et al. (1998), a função densidade da família de distribuições Exponenciada-G é definida por

$$f(x) = \alpha g(x) G(x)^{\alpha-1}$$

Dessa forma, reagrupando alguns termos, temos

$$g(x; \theta, \xi) = \theta (-\ln \overline{F(x; \xi)}) \sum_{i,j \geq 0} \binom{i}{j} \binom{\theta-1}{i} (-1)^{i+j} (1 - \ln \overline{F(x; \xi)})^i f(x; \xi) F(x; \xi)^j$$

Por fim, conforme apresentando anteriormente, fazendo $f(x; \xi) F(x; \xi)^j = \pi(x)/j + 1$, chegamos na expressão apresentada. $\square$

4.3 Propriedades Estatísticas

Nesta seção será apresentada propriedades estatísticas essenciais para a família GE-X.

4.3.1 Quantis

A função quantil de distribuições de probabilidade são importantes para a realização de diversas tarefas, como para encontrar medidas de posição e simulações de Monte Carlo.

Pelo fato da distribuição Gama Exponenciada não apresentar uma forma fechada para sua função quantil, os quantis da Gama Exponenciada-X só podem ser obtidos por simulações numéricas.

Teorema 2. *Seja x_p o p -ésimo quantil da Gama Exponenciada-X. Temos que x_p pode ser obtido resolvendo numericamente a expressão*

$$(1 - F(x_p))(1 - \ln \overline{F}(x_p)) = 1 - p^{\frac{1}{\theta}} \quad (4.11)$$

Demonstração. Pela propriedade da função acumulada, temos que $G(x_p) = p$. Assim, temos que

$$p = [1 - e^{\ln \overline{F}(x_p)}(1 - \ln \overline{F}(x_p))]^{\theta}$$

Elevando ambos os lados ambos os lados por $\frac{1}{\theta}$, chegamos em

$$p^{\frac{1}{\theta}} = 1 - e^{\ln \overline{F}(x_p)}(1 - \ln \overline{F}(x_p))$$

Rearranjando os termos, chegamos na equação apresentada. □

4.3.2 Momentos e função geratriz de momentos

Para melhor entendimento das características de uma distribuição de probabilidade, é imprescindível saber quais são os seus momentos, pois são através deles que podemos tirar informações como medidas centrais, dispersão, assimetria e curtose.

Teorema 3. *Seja X uma variável aleatória com distribuição Gama Exponenciada- X . O seu r -ésimo momento é definido como*

$$E(X^r) = \theta \sum_{i,j \geq 0} \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} w_{i,j,r} \quad (4.12)$$

onde

$$\begin{aligned} w_{i,j,r} &= \int_0^\infty x^r (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx \\ &= E_\pi \left(x^r (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \right) \end{aligned}$$

na qual $E_\pi(Y)$ representa a esperança da variável aleatória Y , cuja distribuição de probabilidade é dada por $\pi(x)$

Demonstração. O r -ésimo momento da variável aleatória X é obtido pela expressão.

$$E(X^r) = \int_0^\infty x^r g(x; \theta, \boldsymbol{\xi}) dx$$

Assim, da equação (4.10), temos que o r -ésimo momento de X se torna

$$E(X^r) = \int_0^\infty x^r \theta (-\ln \overline{F(x; \boldsymbol{\xi})}) \sum_{i,j \geq 0} \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx$$

Pelo teorema de Fubini-Tonelli, temos que os sinais de integração e de somatório podem ser alternados, já que $x^r \theta (-\ln \overline{F(x; \boldsymbol{\xi})}) \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x)$ é absolutamente somável (ou seja, para uma função qualquer $f_n(x)$, $\sum |f_n(x)| < \infty$). Assim, reajustando os termos, chegamos em

$$E(X^r) = \theta \sum_{i,j \geq 0} \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} \int_0^\infty x^r (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx$$

Fazendo a substituição $\int_0^\infty x^r (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx = w_{ij,r}(x)$, chegamos ao resultado. $\square$

Em teoria de probabilidade, a função geratriz de momentos (ou FGM) é uma representação alternativa da função densidade de probabilidade, de maneira a encapsular informações sobre seus momentos, sendo assim, uma maneira alternativa de se encontrar os momentos de uma distribuição de probabilidade. A função geradora de momentos é particularmente útil porque, em muitos casos, permite a obtenção de resultados analíticos de forma mais simples e direta, através de sucessivas derivadas da função no ponto 0. Além disso, é uma ferramenta importante para o estudo de somas de variáveis aleatórias.

Teorema 4. *Seja a variável aleatória X com distribuição Gama Exponenciada- X . Sua função geratriz de momentos $M_X(t)$ é dada por*

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k) \quad (4.13)$$

Demonstração. A função geratriz geratriz de momentos de uma variável aleatória X é definida através da expressão.

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} g(x; \theta, \boldsymbol{\xi}) dx$$

Desta maneira, utilizando a equação (4.10), temos que

$$M_X(t) = \int_0^\infty e^{tx} \theta (-\ln \overline{F(x; \boldsymbol{\xi})}) \sum_{i,j \geq 0} \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx$$

Aplicando a expansão da função exponencial $e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!}$, chegamos em

$$M_X(t) = \int_0^\infty \sum_{k=0}^\infty \sum_{i,j \geq 0} \theta \frac{x^k t^k}{k!} \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx$$

Conforme o teorema de Fubini-Tonelli, em argumento análogo ao apresentado no Teorema 3 e reagrupando alguns termos, chegamos em.

$$M_X(t) = \sum_{k=0}^\infty \frac{t^k}{k!} \sum_{i,j \geq 0} \theta \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} \int_0^\infty x^k (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx$$

Mas pelo Teorema 3

$$E(X^k) = \sum_{i,j \geq 0} \theta \binom{i}{j} \binom{\theta-1}{i} \frac{(-1)^{i+j}}{j+1} \int_0^\infty x^k (-\ln \overline{F(x; \boldsymbol{\xi})}) (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x) dx$$

Assim, chegamos ao resultado apresentado.

□

4.3.3 Desvios Médios

A dispersão de uma distribuição pode ser medida através da soma dos valores absolutos dos desvios em relação a média, ou então em relação a mediana, para o caso de distribuições assimétricas.

Seja X tem distribuição Gama Exponenciada-X, com média $\mu = E(X)$ e mediana m , os desvios em relação a média e em relação são definidos por

$$\delta_1 = \int_0^\infty |x - \mu| g(x; \theta, \boldsymbol{\xi}) dx \quad \text{e} \quad \delta_2 = \int_0^\infty |x - m| g(x; \theta, \boldsymbol{\xi}) dx.$$

Essas equações podem simplificadas, conforme Nadarajah e Kotz (2006), chegando em

$$\delta_1 = 2\mu F(\mu) - 2I(\mu) \quad \text{e} \quad \delta_2 = \mu + 2mF(m) - m - I(m)$$

onde a função $I(m)$ é obtida por

$$I(m) = \int_0^m xg(x; \theta, \xi) dx$$

Conforme a equação (4.10), chegamos que

$$I(m) = \sum_{i,j \geq 0} (-1)^{i+j} \theta \binom{i}{j} \binom{\theta-1}{i} \int_0^m x (-\ln \overline{F(x; \xi)}) (1 - \ln \overline{F(x; \xi)})^i \pi(x) dx \quad (4.14)$$

4.3.4 Entropia de Rényi

A entropia é uma medida utilizada na teoria de probabilidades que mede o grau de incerteza de uma variável aleatória. Quanto maior seu valor, maior o seu grau de incerteza. Existem diversas maneiras para o cálculo da entropia, como a entropia de Rényi, proposta por Rényi (1961), que se trata da generalização de outras entropias, como da entropia de Shannon.

Dado uma variável aleatória X , com função densidade $f(x)$, a entropia de Rényi é definida como

$$I_\alpha(X) = (\alpha - 1)^{-1} \ln \left[\int_0^\infty f^\alpha(x) dx \right] \quad \alpha > 0, \alpha \neq 1$$

Teorema 5. *Seja X uma variável aleatória com distribuição Gama Exponenciada- X . A entropia de Rényi para essa variável é definida por*

$$I_\alpha(X) = (1 - \alpha)^{-1} \ln \left[\theta^\alpha \sum_{i,j \geq 0} \binom{i}{j} \binom{\alpha(\theta - 1)}{i} \frac{(-1)^{i+j}}{j+1} \int_0^\infty a_{i,j}(x) \pi(x) dx \right] \quad (4.15)$$

onde

$$a_{i,j}(x) = f(x; \boldsymbol{\xi})^{\alpha-1} (-\ln \overline{F(x; \boldsymbol{\xi})})^\alpha (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i$$

Demonstração. A função $g^\alpha(x)$ é definida por

$$g^\alpha(x; \theta, \boldsymbol{\xi}) = \theta^\alpha f^\alpha(x; \boldsymbol{\xi}) \{(-\ln \overline{F(x; \boldsymbol{\xi})})^\alpha\} [1 - (1 - F(x; \boldsymbol{\xi}))(1 - \ln F(x; \boldsymbol{\xi}))]^{\alpha(\theta-1)}$$

Fazendo um processo semelhante ao feito no Teorema 1 , chegamos que a expansão em séries para $g^\alpha(x)$ é dada por

$$g^\alpha(x; \theta, \boldsymbol{\xi}) = \theta^\alpha f^{\alpha-1}(x; \boldsymbol{\xi}) (-\ln \overline{F(x; \boldsymbol{\xi})})^\alpha \sum_{i,j \geq 0} \binom{i}{j} \binom{\alpha(\theta - 1)}{i} \frac{(-1)^{i+j}}{j+1} (1 - \ln \overline{F(x; \boldsymbol{\xi})})^i \pi(x)$$

Inserindo esta expansão na expressão da entropia, chegamos ao resultado. $\square$

4.3.5 Estatística de ordem

As estatísticas de ordem são cruciais para o estudo de distribuições, por proporcionarem informações importantes como o mínimo e o máximo, importantes para sabermos mais sobre a distribuição e a estrutura de um determinado conjunto de dados.

Considere $(X_{(1)}, \dots, X_{(n)})$ uma amostra aleatória ordenada da variável aleatória X com função densidade $f(x)$ e função acumulada $F(x)$. A função densidade de probabilidade de $X_{(i)}$ (ou seja, a estatística de i -ésima ordem), denotada por $f_{X_{(i)}}(x)$ é dada por

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} [F(x; \boldsymbol{\xi})]^{i-1} [1 - F(x; \boldsymbol{\xi})]^{n-i} f(x; \boldsymbol{\xi})$$

Aplicando a expansão binomial generalizada, a equação acima pode ser simplificada para

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (1 - F(x; \boldsymbol{\xi}))^{n-i+j} f(x; \boldsymbol{\xi})$$

Seja X uma variável aleatória com distribuição Gama Exponenciada-X. A função densidade de $X_{(i)}$ pode ser escrita como

$$\begin{aligned} g_{X_{(i)}}(x; \theta, \boldsymbol{\xi}) &= \frac{n!}{(i-1)!(n-i)!} \theta \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \left(1 - \left[1 - (1 - F(x; \boldsymbol{\xi})) (1 - \ln \overline{F(x; \boldsymbol{\xi})}) \right]^\theta \right)^{n-i+j} \\ &\quad \times f(x; \boldsymbol{\xi}) (-\ln \overline{F(x; \boldsymbol{\xi})}) \left[1 - (1 - F(x; \boldsymbol{\xi})) (1 - \ln \overline{F(x; \boldsymbol{\xi})}) \right]^{\theta-1} \end{aligned} \quad (4.16)$$

4.3.6 Estimação dos Parâmetros

A estimação de parâmetros é um dos pilares no processo de inferência estatística, consistindo em métodos para determinar os valores desconhecidos que caracterizam um modelo probabilístico através dos dados observados. Um dos meios mais utilizados para este fim é o método da Máxima Verossimilhança, que se baseia na maximização da função de verossimilhança, que mede a probabilidade dos dados observados condicionada aos parâmetros.

Os estimadores são muito utilizados em na inferência frequentista por possuir excelentes propriedades assintóticas, como consistência (a medida que o tamanho amostral cresce, a estimativa se aproxima cada vez ao valor do parâmetro verdadeiro) e eficiência (com o aumento do tamanho amostral, o estimador de máxima verossimilhança apresenta a menor variância entre todos os estimadores possíveis).

A função de verossimilhança para uma amostra aleatória de tamanho n de uma variável aleatória Y com vetor paramétrico $\underline{\theta}$ e função densidade $s(y; \underline{\theta})$ é definida como

$$L(\underline{\theta}) = \prod_{i=1}^n s(y; \underline{\theta})$$

Com o intuito de facilitar os cálculos, acaba se utilizando o logaritmo da função de verossimilhança, a log-verossimilhança

$$\mathcal{L}(\underline{\theta}) = \sum_{i=1}^n \ln s(y; \underline{\theta})$$

Seja $X_1, \dots, X_n$ uma amostra aleatória da variável aleatória X com distribuição Gama Exponenciada-X. Sendo $\Theta = (\theta, \boldsymbol{\xi})^\top$, com base em (4.8), a função log-verossimilhança de Θ é dada por

$$\begin{aligned} \mathcal{L}(\Theta) = & n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln [1 - e^{-H(x_i; \boldsymbol{\xi})} (H(x_i; \boldsymbol{\xi}) + 1)] \\ & + \sum_{i=1}^n \ln h(x_i; \boldsymbol{\xi}) + \sum_{i=1}^n \ln H(x_i; \boldsymbol{\xi}) - \sum_{i=1}^n H(x_i; \boldsymbol{\xi}) \end{aligned} \quad (4.17)$$

Derivando (4.17) em relação as componentes de Θ , chegamos a função score $U(\Theta) = (U_\theta, U_{\boldsymbol{\xi}})^\top$

na qual suas componentes são dadas por

$$U_\theta = \frac{\partial \mathcal{L}(\Theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n \ln [1 - e^{-H(x_i; \boldsymbol{\xi})} (H(x_i; \boldsymbol{\xi}) + 1)] \quad (4.18)$$

e

$$\begin{aligned}
U_{\xi} = \frac{\partial \mathcal{L}(\Theta)}{\partial \xi} &= \sum_{i=1}^n \frac{1}{h(x_i; \xi)} \frac{\partial h(x_i; \xi)}{\partial \xi} + \sum_{i=1}^n \frac{1}{H(x_i; \xi)} \frac{\partial H(x_i; \xi)}{\partial \xi} - \sum_{i=1}^n \frac{\partial H(x_i; \xi)}{\partial \xi} \\
&+ (\theta - 1) \sum_{i=1}^n \frac{H(x_i; \xi) e^{-H(x_i; \xi)}}{1 - e^{H(x_i; \xi)}(H(x_i; \xi) + 1)} \frac{\partial H(x_i; \xi)}{\partial \xi}
\end{aligned} \quad (4.19)$$

Os estimadores de Máxima Verossimilhança (EMV) $\hat{\theta}_{EMV}$ e $\hat{\xi}_{EMV}$ para θ e ξ serão obtidos resolvendo $U_{\theta} = 0$ e $U_{\xi} = 0$ simultaneamente. Vale ressaltar que essas equações não apresentam soluções analíticas, devendo serem resolvidas numericamente.

Como temos a função score para cada um dos parâmetros, podemos montar a matriz de informação de Fisher, que é dada por

$$\mathcal{I}(\Theta) = \begin{pmatrix} U_{\theta\theta} & U_{\theta\xi} \\ U_{\xi\theta} & U_{\xi\xi} \end{pmatrix}$$

onde cada componente é dada por

$$U_{\theta\theta} = \frac{\partial^2 \mathcal{L}(\Theta)}{\partial \theta^2} = -\frac{n}{\theta^2} \quad (4.20)$$

$$U_{\theta\xi} = U_{\xi\theta} = \frac{\partial^2 \mathcal{L}(\Theta)}{\partial \theta \partial \xi} = \sum_{i=1}^n \frac{H(x_i; \xi) e^{-H(x_i; \xi)}}{1 - e^{H(x_i; \xi)}(H(x_i; \xi) + 1)} \frac{\partial H(x_i; \xi)}{\partial \xi} \quad (4.21)$$

e

$$\begin{aligned}
U_{\xi\xi} = \frac{\partial^2 \mathcal{L}}{\partial \xi \partial \xi^{\top}} &= \sum_{i=1}^n \left[\left(\frac{(\theta - 1) e^{-H(x_i; \xi)} (1 - H(x_i; \xi) - e^{-H(x_i; \xi)})}{[1 - e^{H(x_i; \xi)}(H(x_i; \xi) + 1)]^2} - \frac{1}{H(x_i; \xi)^2} \right) \right. \\
&\times \frac{\partial H(x_i; \xi)}{\partial \xi} \frac{\partial H(x_i; \xi)}{\partial \xi^{\top}} + \left(\frac{(\theta - 1) H(x_i; \xi) e^{-H(x_i; \xi)}}{1 - e^{H(x_i; \xi)}(H(x_i; \xi) + 1)} + \frac{1}{H(x_i; \xi)} - 1 \right) \\
&\times \left. \frac{\partial^2 H(x_i; \xi)}{\partial \xi \partial \xi^{\top}} - \frac{1}{h(x_i; \xi)^2} \frac{\partial h(x_i; \xi)}{\partial \xi} \frac{\partial h(x_i; \xi)}{\partial \xi^{\top}} + \frac{1}{h_i} \frac{\partial^2 h(x_i; \xi)}{\partial \xi \partial \xi^{\top}} \right]
\end{aligned} \quad (4.22)$$

A inversa da matriz de Fisher nos fornece a matriz de variância e covariância dos estimadores de máxima verossimilhança de nossa distribuição, ou seja.

$$\mathcal{I}^{-1}(\Theta) = \begin{pmatrix} Var(\hat{\theta}_{EMV}) & Cov(\hat{\theta}_{EMV}, \hat{\xi}_{EMV}) \\ Cov(\hat{\xi}_{EMV}, \hat{\theta}_{EMV}) & Var(\hat{\xi}_{EMV}) \end{pmatrix} \quad (4.23)$$

A matriz acima é importante para encontrar os $100(1 - \alpha)\%$ intervalos de confiança para os estimadores de máxima de verossimilhança das seguintes formas

$$\hat{\theta}_{EMV} \pm z_{\frac{\alpha}{2}} Var(\hat{\theta}_{EMV}) \quad ; \quad \hat{\xi}_{EMV} \pm z_{\frac{\alpha}{2}} Var(\hat{\xi}_{EMV}) \quad (4.24)$$

Onde $z_{\frac{\alpha}{2}}$ é o quantil da distribuição Normal padrão, de maneira que $P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

4.4 Exemplos de distribuições

Nessa seção, algumas distribuições que pertecem a família Gama Exponencia-X serão apresentadas, com suas respectivas funções acumulada, densidade, de sobrevivência e de risco. Também será mostrado o comportamento dessas distribuições para um determinado conjunto de parâmetros.

4.4.1 Gama Exponenciada-Weibull

A distribuição Weibull é uma das mais aplicadas em análise de sobrevivência, sendo aplicada na modelagem de tempo de vida, tempo de falha e resistência de materiais. Um dos principais motivos de sua utilização é sua função de risco em formato de banheira.

Sendo a função acumulada da distribuição Weibull dada por

$$F(x; k, \lambda) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k} \quad x \geq 0, \quad k, \lambda > 0$$

onde k representa o parâmetro de forma e λ representa o parâmetro de escala. A

distribuição Gama Exponenciada-Weibull (GE-W) é obtida por

$$G(x; \theta, k, \lambda) = \left[1 - e^{-\left(\frac{x}{\lambda}\right)^k} \left(1 + \left(\frac{x}{\lambda}\right)^k \right) \right]^\theta \quad x \geq 0, \quad \theta, k, \lambda > 0 \quad (4.25)$$

com respectiva função densidade

$$g(x; \theta, k, \lambda) = \theta \frac{k}{\lambda^k} x^{k-1} \left(\frac{x}{\lambda}\right)^k e^{-\left(\frac{x}{\lambda}\right)^k} \left[1 - e^{-\left(\frac{x}{\lambda}\right)^k} \left(1 + \left(\frac{x}{\lambda}\right)^k \right) \right]^{\theta-1} \quad x \geq 0, \quad \theta, k, \lambda > 0 \quad (4.26)$$

Com essas equações também podemos definir a função de sobrevivência

$$S(x; \theta, k, \lambda) = 1 - \left[1 - e^{-\left(\frac{x}{\lambda}\right)^k} \left(1 + \left(\frac{x}{\lambda}\right)^k \right) \right]^\theta \quad x \geq 0, \quad \theta, k, \lambda > 0 \quad (4.27)$$

e também a função de risco

$$h(x; \theta, k, \lambda) = \frac{\theta \frac{k}{\lambda^k} x^{k-1} \left(\frac{x}{\lambda}\right)^k e^{-\left(\frac{x}{\lambda}\right)^k} \left[1 - e^{-\left(\frac{x}{\lambda}\right)^k} \left(1 + \left(\frac{x}{\lambda}\right)^k \right) \right]^{\theta-1}}{1 - \left[1 - e^{-\left(\frac{x}{\lambda}\right)^k} \left(1 + \left(\frac{x}{\lambda}\right)^k \right) \right]^\theta} \quad x \geq 0, \quad \theta, k, \lambda > 0 \quad (4.28)$$

A figura 4.1 mostra alguns exemplos de formas que a densidade da GE-W pode assumir, verificando como parâmetro influência no formato do seu gráfico. Já a figura 4.2 mostra a função de risco da GE-W para um um outro conjunto de parâmetros. Nota-se que para $k < 1$, a função de risco assume uma forma unimodal, enquanto que para $k > 1$, temos a função de risco em formato crescente.

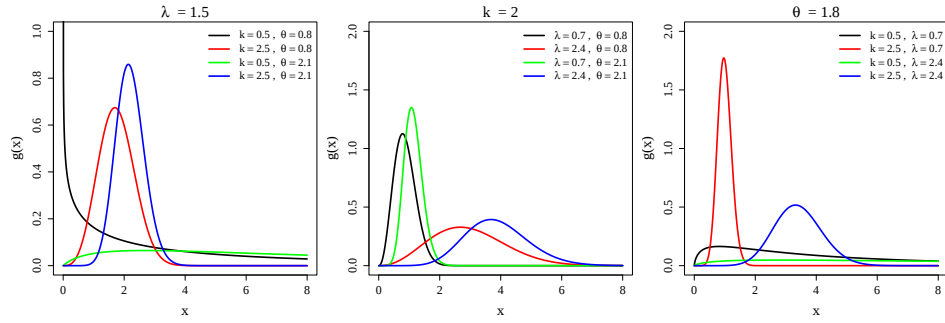


Figura 4.1: Gráficos da função densidade da GE-W com diferentes parâmetros

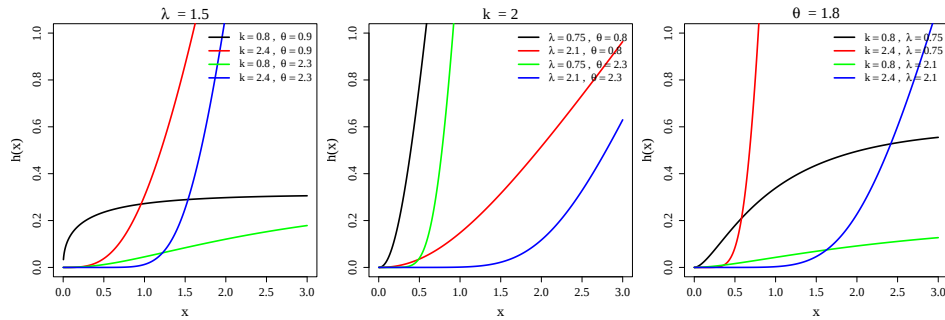


Figura 4.2: Gráficos da função de risco da GE-W com diferentes parâmetros.

4.4.2 Gama Exponenciada-Lomax

A distribuição de Lomax, também chamada de Pareto tipo II, é uma distribuição de probabilidade com a particularidade de apresentar um calda pesada. É muito utilizado em negócios, economia e ciências atuárias.

Sendo a função acumulada da distribuição de Lomax dada por

$$F(x; \alpha, \lambda) = 1 - \left(1 + \frac{x}{\lambda}\right)^{-\alpha} \quad x \geq 0, \quad \alpha, \lambda > 0$$

onde α é o parâmetro de forma e λ o parâmetro de escala. A distribuição Gama Exponenciada-Lomax (GE-Lom) é obtida por

$$G(x; \theta, \alpha, \lambda) = \left[1 - \left(1 + \frac{x}{\lambda}\right)^{-\alpha} \left(\alpha \ln \left\{1 + \frac{x}{\lambda}\right\} + 1\right)\right]^{\theta} \quad x \geq 0, \quad \theta, \alpha, \lambda > 0 \quad (4.29)$$

com respectiva função densidade

$$g(x; \theta, \alpha, \lambda) = \theta \frac{\alpha^2}{\lambda} \ln \left\{ 1 + \frac{x}{\lambda} \right\} \left(1 + \frac{x}{\lambda} \right)^{-(\alpha+1)} \times \left[1 - \left(1 + \frac{x}{\lambda} \right)^{-\alpha} \left(\alpha \ln \left\{ 1 + \frac{x}{\lambda} \right\} + 1 \right) \right]^{\theta-1} \quad x \geq 0, \quad \theta, \alpha, \lambda > 0 \quad (4.30)$$

Com essas duas equações, podemos encontrar a função de sobrevivência e a função de risco, respectivamente, por

$$S(x; \theta, \alpha, \lambda) = 1 - \left[1 - \left(1 + \frac{x}{\lambda} \right)^{-\alpha} \left(\alpha \ln \left\{ 1 + \frac{x}{\lambda} \right\} + 1 \right) \right]^{\theta} \quad x \geq 0, \quad \theta, \alpha, \lambda > 0 \quad (4.31)$$

e

$$h(x; \theta, \alpha, \lambda) = \frac{\theta \frac{\alpha^2}{\lambda} \ln \left\{ 1 + \frac{x}{\lambda} \right\} \left(1 + \frac{x}{\lambda} \right)^{-(\alpha+1)} \left[1 - \left(1 + \frac{x}{\lambda} \right)^{-\alpha} \left(\alpha \ln \left\{ 1 + \frac{x}{\lambda} \right\} + 1 \right) \right]^{\theta-1}}{1 - \left[1 - \left(1 + \frac{x}{\lambda} \right)^{-\alpha} \left(\alpha \ln \left\{ 1 + \frac{x}{\lambda} \right\} + 1 \right) \right]^{\theta}} \quad (4.32)$$

A figura 4.3 mostra alguns exemplos de gráficos que a GE-Lom pode assumir, verificando como parâmetro influência no formato de sua densidade. Já a figura 4.4 mostra diferentes formas para para sua função de risco. Podemos perceber que para $\theta < 1$, a função de risco tem formato decrescente, enquanto que $\theta > 1$, o risco apresenta forma unimodal.

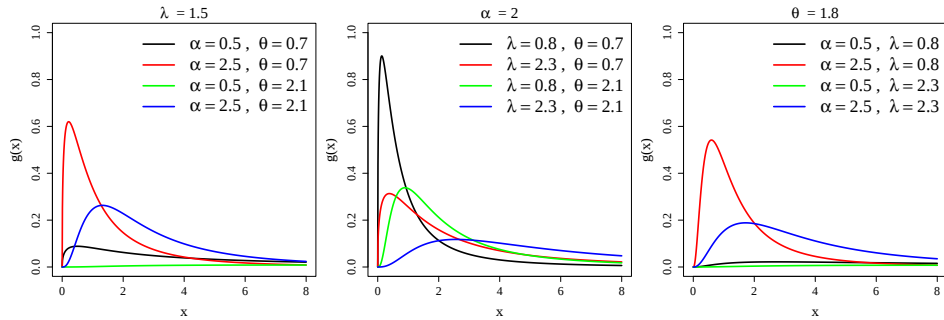


Figura 4.3: Gráficos da função densidade da GE-Lom com diferentes parâmetros.

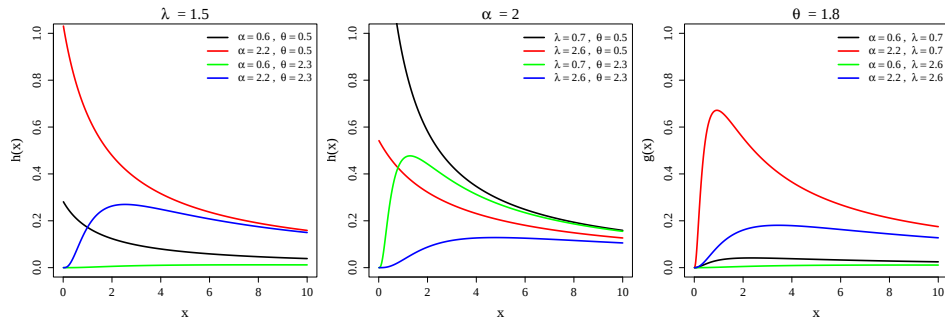


Figura 4.4: Gráficos da função de risco da GE-Lom com diferentes parâmetros.

À primeira vista, observa-se que o modelo Gama-Exponenciada-X apresenta grande flexibilidade, permitindo modificar distribuições de probabilidade já conhecidas e, com isso, gerar novos formatos para a função densidade e, conseqüentemente, para suas funções de risco. Essa característica amplia as possibilidades de modelagem estatística.

Para atestar a flexibilidade da família GE-X, no próximo capítulo estudaremos mais a fundo uma distribuição da família, a distribuição Gama Exponenciada-Chen.

5 Gama Exponenciada-Chen

5.1 Principais funções

Chen (2000) propôs a distribuição Chen, um distribuição com dois parâmetros que pode apresentar sua função de risco crescente ou então com o formato de banheira. Dado o padrão flexível de sua função de risco, a distribuição Chen é muito utilizada nos campos de análise de sobrevivência e confiabilidade, como em engenharia industrial, testes acelerados e em medicina. Seja X uma variável aleatória com distribuição Chen. Sua fda e fdp são dadas, respectivamente, por

$$F(x; \lambda, \beta) = 1 - e^{\lambda(1-e^{x^\beta})} \quad (5.1)$$

e

$$f(x; \lambda, \beta) = \lambda \beta x^{\beta-1} e^{x^\beta + \lambda(1-e^{x^\beta})} \quad (5.2)$$

onde λ é o parâmetro de escala e β é o parâmetro de forma.

Inserindo 5.1 em 4.3, chegamos na distribuição Gama Exponenciada-Chen (GE-C), definida por

$$G(x; \theta, \beta, \lambda) = [1 - e^{\lambda(1-e^{x^\beta})}(1 - \lambda(1 - e^{x^\beta}))]^\theta \quad x \geq 0, \quad \theta, \beta, \lambda > 0 \quad (5.3)$$

com respectiva função densidade

$$g(x; \theta, \beta, \lambda) = \theta \beta \lambda^2 x^{\beta-1} (e^{x^\beta} - 1) \exp \{x^\beta + \lambda(1 - e^{x^\beta})\} \quad (5.4)$$

$$\times [1 - e^{\lambda(1-e^{x^\beta})} (1 - \lambda(1 - e^{x^\beta}))]^{\theta-1} \quad x \geq 0, \quad \theta, \beta, \lambda > 0$$

Com isso, conseguimos também suas funções de sobrevivência e de risco, definidas respectivamente como:

$$S(x; \theta, \beta, \lambda) = 1 - [1 - e^{\lambda(1-e^{x^\beta})} (1 - \lambda(1 - e^{x^\beta}))]^\theta \quad x \geq 0, \quad \theta, \beta, \lambda > 0 \quad (5.5)$$

e

$$h(x; \theta, \beta, \lambda) = \frac{\theta \beta \lambda^2 x^{\beta-1} (e^{x^\beta} - 1) \exp \{x^\beta + \lambda(1 - e^{x^\beta})\} [1 - e^{\lambda(1-e^{x^\beta})} (1 - \lambda(1 - e^{x^\beta}))]^{\theta-1}}{1 - [1 - e^{\lambda(1-e^{x^\beta})} (1 - \lambda(1 - e^{x^\beta}))]^\theta} \quad (5.6)$$

A figura 5.1 mostra como a densidade da distribuição Gama Exponenciada-Chen muda com a variação de seus parâmetros.

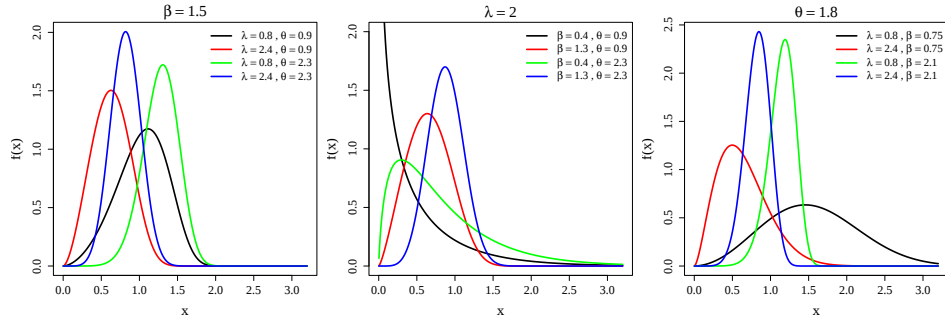


Figura 5.1: Gráficos da função densidade da GE-C com diferentes parâmetros

Uma característica muito notável da GE-C é que ela pode apresentar as quatro diferentes formas para a sua função de risco mais frequentemente vistas na literatura, ou seja, pode apresentar risco crescente, decrescente, unimodal ou com *bathtub shape* (curva da banheira).

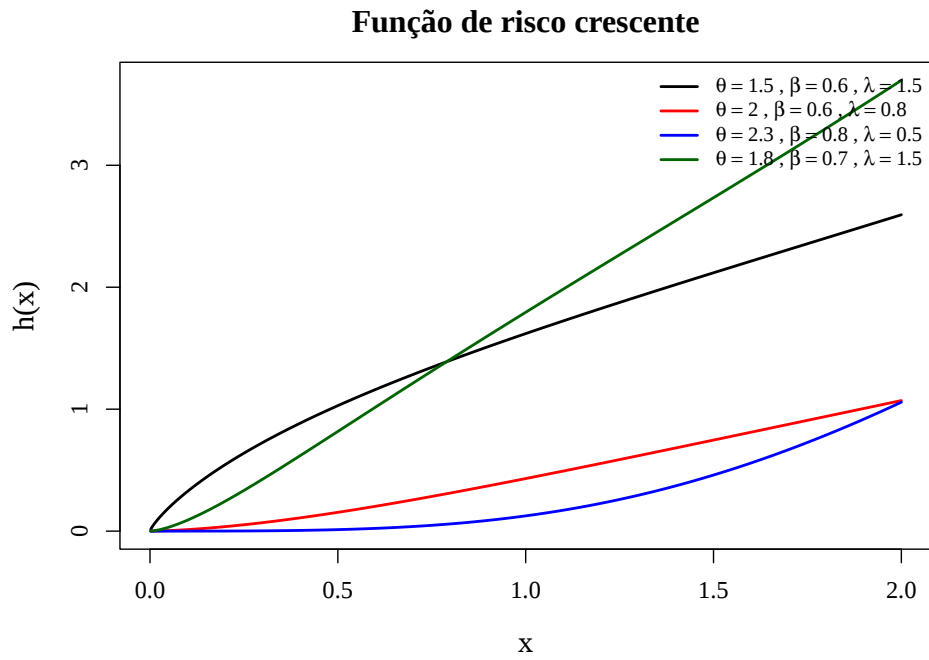


Figura 5.2: Função de risco da distribuição GE-C para diferentes combinações de parâmetros com comportamento crescente.

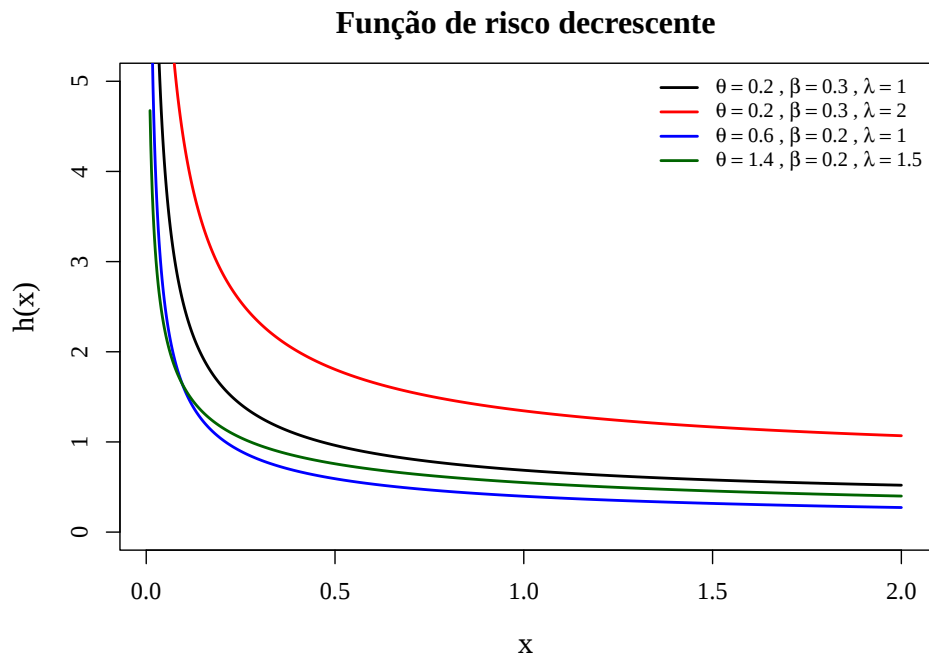


Figura 5.3: Função de risco da distribuição GE-C para diferentes combinações de parâmetros com comportamento decrescente.

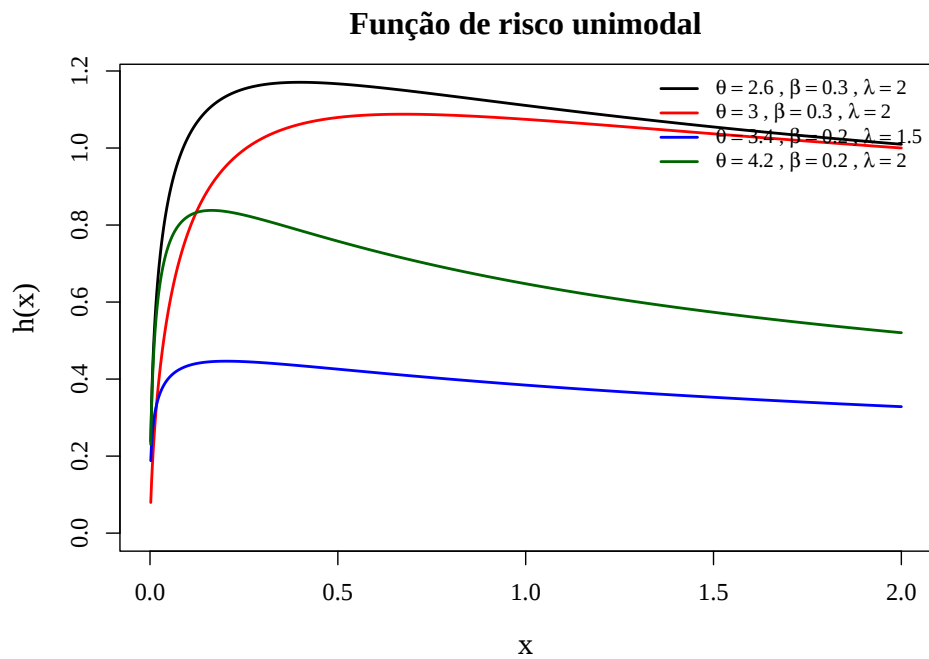


Figura 5.4: Função de risco da distribuição GE-C para diferentes combinações de parâmetros com formato unimodal.

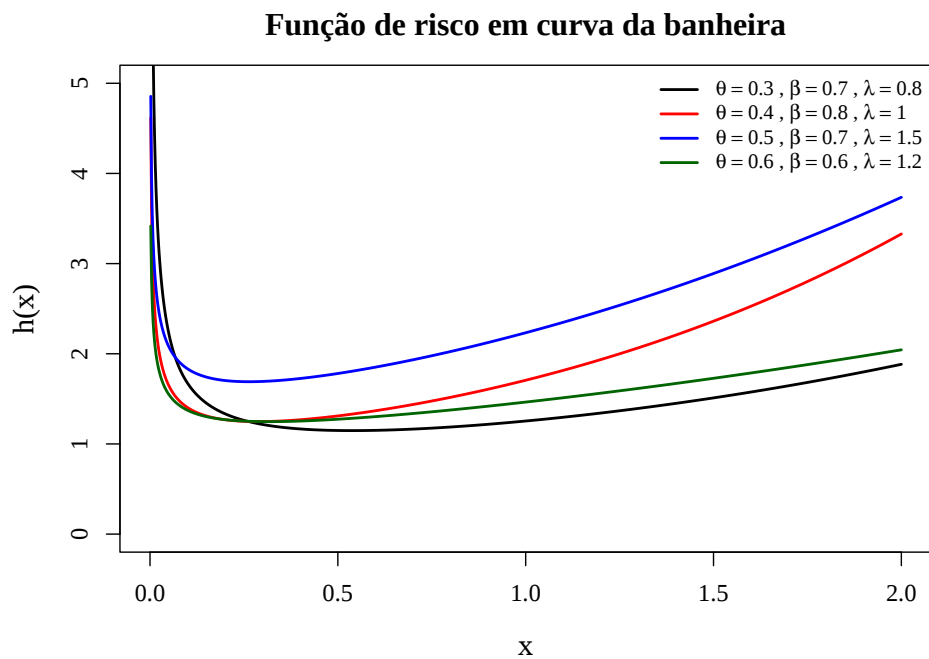


Figura 5.5: Função de risco da distribuição GE-C para diferentes combinações de parâmetros com formato em curva da banheira.

5.2 Momentos

A análise dos momentos de ordem superior da distribuição GE-C mostra que o modelo é bastante flexível em termos de forma, principalmente ao se analisar sua assimetria e curtose.

A assimetria (denotada por γ é um valor que mede quanto uma distribuição ou conjunto de dados é assimétrico, ou seja, quanto o lado esquerdo e direito do ponto central da distribuição são distintos. É também conhecido como o coeficiente de assimetria de Fisher-Pearson. Sua expressão é dada por:

$$\gamma = E\left(\frac{(x - \mu)^3}{\sigma^3}\right) = \frac{E(X^3) - 3E(X)E(X^2) + 2[E(X)]^3}{Var^{\frac{3}{2}}(X)}$$

Por sua vez, a curtose (representada por κ) é uma medida de dispersão que caracteriza o pico de uma distribuição de probabilidade. Uma alta curtose representa que a distribuição apresenta caudas mais pesadas, mostrando uma grande quantidade de *outliers*. É obtida pela expressão:

$$\kappa = E\left(\frac{(x - \mu)^4}{\sigma^4}\right) = \frac{E(X^4) - 4E(X)E(X^3) + 6E(X^2)[E(X)]^2 - 3[E(X)]^4}{[Var(X)]^2}$$

Observa-se, pelas curvas de assimetria, que a GE-C gera, na maior parte dos casos considerados, distribuições assimétricas à direita (coeficiente de assimetria positivo), com cauda mais alongada no lado dos valores maiores de X . Dependendo da combinação dos parâmetros, essa assimetria pode ser mais moderada, aproximando a distribuição de um formato quase simétrico, ou bastante acentuada, indicando forte concentração de massa próxima da origem e uma cauda direita pesada. Isso evidencia que o modelo consegue representar desde situações com leve desbalanceamento até cenários com clara predominância de valores baixos e poucos valores extremos elevados.

Os gráficos de curtose, por sua vez, indicam que a GE-C consegue produzir tanto

distribuições com caudas relativamente leves quanto distribuições marcadamente leptocúrticas, com caudas pesadas e maior probabilidade de observações extremas. Em alguns conjuntos de parâmetros, os valores de curtose aproximam-se dos da distribuição normal (mesocurtose), enquanto em outros o coeficiente é substancialmente maior, caracterizando um pico mais acentuado no centro e maior espessura das caudas.

θ	β	λ	$E(X)$	$E(X^2)$	$Var(X)$	$E(X^3)$	$E(X^4)$	Assimetria	Curtose
1.5	0.8	1.2	1.07	1.36	0.227	1.98	3.17	0.415	2.88
1.4	0.6	0.7	1.79	4.11	0.886	11.0	33.4	0.575	3.08
2.0	0.3	1.1	2.43	11.4	5.49	80.3	754.0	2.00	9.05

Tabela 5.1: Momentos principais da GE-C para diferentes valores de θ , β e λ

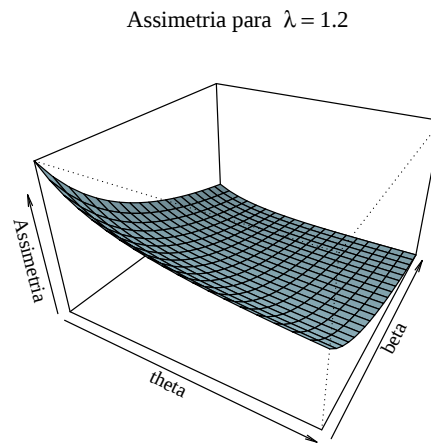


Figura 5.6: Superfície da assimetria da distribuição GE-C em função de θ e β , com $\lambda = 1.2$ fixo.

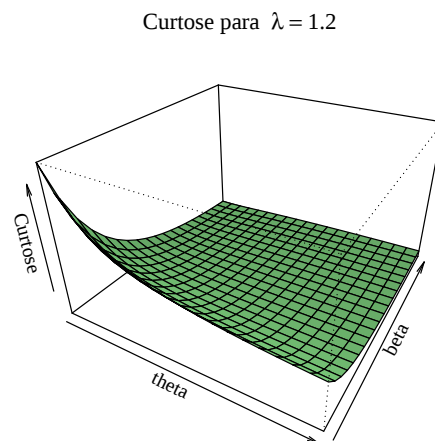


Figura 5.7: Superfície da curtose da distribuição GE-C em função de θ e β , com $\lambda = 1.2$ fixo.

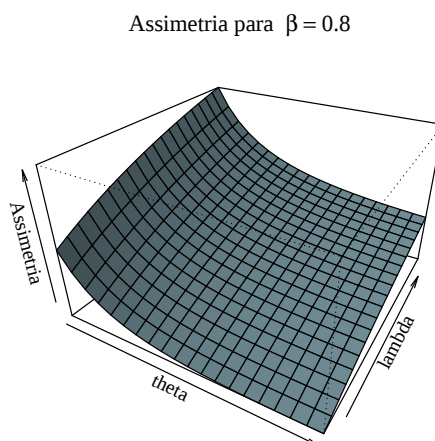


Figura 5.8: Superfície da assimetria da distribuição GE-C em função de θ e λ , com $\beta = 0.8$ fixo.

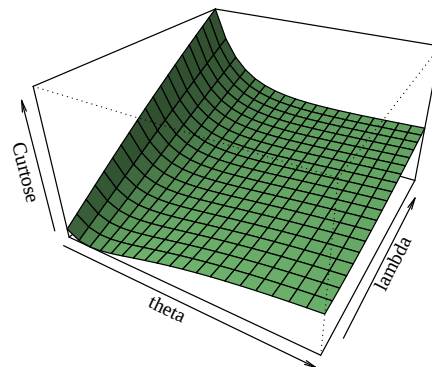
Curtose para $\beta = 0.8$ 

Figura 5.9: Superfície da curtose da distribuição GE-C em função de θ e λ , com $\beta = 0.8$ fixo.

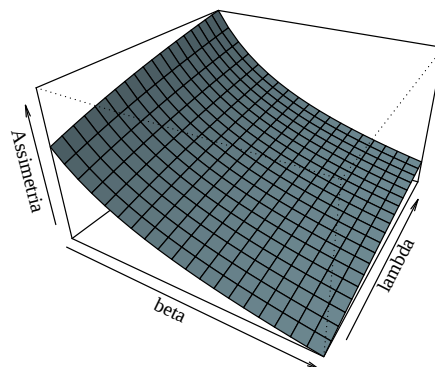
Assimetria para $\theta = 1.5$ 

Figura 5.10: Superfície da assimetria da distribuição GE-C em função de β e λ , com $\theta = 1.5$ fixo.

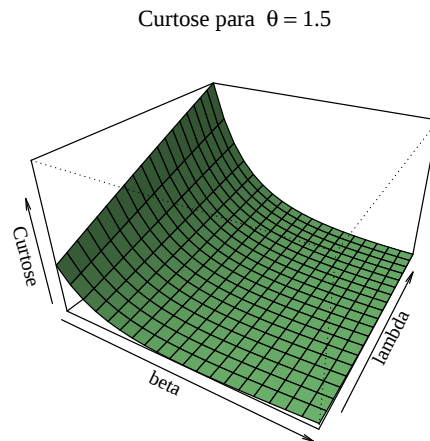


Figura 5.11: Superfície da curtose da distribuição GE-C em função de β e λ , com $\theta = 1.5$ fixo.

5.3 Simulação

Para verificar as propriedades assintóticas dos estimadores de máxima verossimilhança da distribuição GE-C, foram conduzidas duas simulações de Monte Carlo com $B = 1000$ amostras de tamanhos $n = 20, 50, 100, 200$ para diferentes valores dos parâmetros da distribuição, a fim de analisar as métricas de vício (BIAS), erro quadrático médio (MSE), erro absoluto médio (MAE) e erro padrão médio (RMSE), com suas expressões definidas para uma estimativa $\hat{\alpha}$ de um parâmetro α qualquer definidas por:

$$\text{BIAS}(\hat{\alpha}) = \frac{1}{B} \sum_{i=1}^B (\hat{\alpha}_i - \alpha),$$

$$\text{MSE}(\hat{\alpha}) = \frac{1}{B} \sum_{i=1}^B (\hat{\alpha}_i - \alpha)^2,$$

$$\text{MAE}(\hat{\alpha}) = \frac{1}{B} \sum_{i=1}^B |\hat{\alpha}_i - \alpha|,$$

$$\text{RMSE}(\hat{\alpha}) = \sqrt{\frac{1}{B} \sum_{i=1}^B (\hat{\alpha}_i - \alpha)^2}.$$

As amostras para a simulação foram obtidas através da Proposição 1, presente na seção 4.1 do capítulo 4. Sendo a função quantil da distribuição Chen dada por

$$Q(u) = \left(\ln \left[1 - \frac{1}{\lambda} \ln(1 - u) \right] \right)^{\frac{1}{\beta}} \quad 0 < u < 1$$

Tomando $u = 1 - e^{-T}$, com T uma variável aleatória com distribuição Gama Exponenciada com parâmetro θ , chegamos na variável aleatória abaixo.

$$X = \left(\ln \left[1 + \frac{T}{\lambda} \right] \right)^{\frac{1}{\beta}} \sim GE - C(\theta, \beta, \lambda)$$

Dessa forma, foram simuladas 200000 amostras de X , com o auxílio do pacote *Newdistns*, com cada observação sendo armazenada em uma matriz de 200 linhas (numero máximo de tamanho da amostra) e 1000 colunas (número de amostras). Sendo assim, as primeiras 20 linhas da primeira coluna representam a primeira amostra de tamanho $n = 20$, as 50 primeiras linhas, ou seja, a primeira amostra adicionando-se as 30 próximas linhas da primeira coluna representa a primeira amostra de tamanho $n = 50$, considerando as 100 primeiras linhas da mesma coluna obtêm-se a primeira amostra de tamanho $n = 100$ e, finalmente, considerando-se todas as linhas da primeira co-

luna, tem-se a primeira amostra de tamanho $n = 200$. Repetindo esse processo nas 1000 colunas, obtemos as 1000 amostras de cada um dos diferentes tamanhos amostrais considerados no estudo.

Os resultados das simulações de Monte Carlo indicam que os estimadores da distribuição Gama Exponenciada–Chen apresentam o comportamento esperado do ponto de vista assintótico. De modo geral, observou-se que, à medida que o tamanho amostral aumenta, há redução da variabilidade das estimativas e do erro quadrático médio em torno dos valores verdadeiros, o que é compatível com a consistência e a normalidade assintótica previstas pela teoria.

Por outro lado, verificou-se a presença de um viés moderado nas estimativas de θ e β nos cenários considerados. Embora esse viés não seja desprezível para amostras menores, sua magnitude não compromete o desempenho global do estimador e tende a se reduzir com o aumento de n , indicando que, na prática, o método mantém boa qualidade inferencial para tamanhos amostrais moderados e grandes.

5.4 Aplicação

Para verificar a aplicabilidade da distribuição Gama Exponenciada–Chen, foram considerados dois *datasets*. O primeiro é referente a voltagem necessária para a falha de 128 isolantes elétricos, visto primeiramente Nelson (2004), já o segundo é referente a resistência de fibras de vidro de 1.5 cm, medidos no *National Physical Laboratory*, Inglaterra, utilizado pela primeira vez por Smith e Naylor (1987). A distribuição GE-C foi comparada com as distribuições Gama, Weibull e Chen, todas com 2 parâmetros, e também com a distribuição Gama generalizada, com 3 parâmetros. Para avaliá-las, foram analisados o Critério de Informação de Akaike (AIC), proposta por Akaike (1974), o Critério de Informação Bayesiano (BIC), também chamado de Critério de Schwarz, por Schwarz (1978) e o Critério de Informação de Akaike Corrigido (AICc), proposto inicialmente por Liang e Zou (2008). Quanto menor o valor dessas métricas, melhor é o ajuste do modelo aos dados testados. Cada uma dessas métricas são dadas, respectivamente, por:

n	θ					β					λ							
	Mean	Bias	Variance	MSE	MAE	RMSE	Mean	Bias	Variance	MSE	MAE	RMSE	Mean	Bias	Variance	MSE	MAE	RMSE
20	2.120	0.617	3.340	3.720	1.500	1.930	1.040	0.444	1.080	1.270	0.580	1.130	2.350	0.050	0.791	0.793	0.731	0.890
50	1.860	0.359	1.820	1.950	0.987	1.400	0.711	0.111	0.113	0.125	0.224	0.353	2.310	0.013	0.317	0.317	0.457	0.563
100	1.710	0.213	0.917	0.961	0.677	0.981	0.641	0.041	0.030	0.032	0.131	0.178	2.330	0.028	0.152	0.153	0.311	0.391
200	1.600	0.105	0.355	0.365	0.431	0.605	0.617	0.017	0.011	0.012	0.084	0.108	2.320	0.019	0.073	0.074	0.215	0.271

Tabela 5.2: Médias, vieses, variâncias, MSE, MAE e RMSE dos EMV de θ , β e λ da distribuição GE-C para $(\theta, \beta, \lambda) = (1.5, 0.6, 2.3)$ em diferentes tamanhos amostrais.

n	θ					β					λ							
	Mean	Bias	Variance	MSE	MAE	RMSE	Mean	Bias	Variance	MSE	MAE	RMSE	Mean	Bias	Variance	MSE	MAE	RMSE
20	1.300	0.499	1.840	2.090	0.843	1.450	1.840	0.242	0.524	0.582	0.763	0.784	0.784	0.084	0.326	0.333	0.449	0.577
50	0.961	0.161	0.489	0.515	0.399	0.717	1.690	0.094	0.166	0.175	0.327	0.418	0.728	0.028	0.128	0.128	0.271	0.358
100	0.859	0.059	0.116	0.119	0.235	0.345	1.640	0.043	0.071	0.073	0.212	0.270	0.715	0.015	0.053	0.053	0.179	0.231
200	0.827	0.028	0.040	0.040	0.151	0.201	1.620	0.019	0.031	0.031	0.140	0.177	0.708	0.008	0.024	0.024	0.121	0.154

Tabela 5.3: Médias, vieses, variâncias, MSE, MAE e RMSE dos EMV de θ , β e λ da distribuição GE-C para $(\theta, \beta, \lambda) = (0.8, 1.6, 0.7)$ em diferentes tamanhos amostrais (segunda simulação).

$$\begin{aligned}
AIC &= -2 \mathcal{L}(\Theta) + 2d, \\
AICc &= AIC + \frac{2(d+2)(d+3)}{n-d-3}, \\
BIC &= -2 \mathcal{L}(\Theta) + \log(n).
\end{aligned}$$

onde n é o número de observações do conjunto de dados e d o número de parâmetros do modelo.

Além disso, será conduzido o teste de razão de verossimilhança entre a distribuição GE-C com a distribuição Chen, para comparar ambas diretamente, considerando o nível de significância $\alpha = 0.05$. Neste teste, são serão testadas duas hipóteses:

- Modelo restrito (H_0): distribuição de Chen, com os parâmetros β e λ ;
- Modelo completo (H_1): distribuição GE-C, com o parâmetro adicional θ .

Sendo $\mathcal{L}(\hat{\Theta}_0)$ e $\mathcal{L}(\hat{\Theta}_1)$ a log-verossimilhança da distribuição Chen e Gama Exponenciada-Chen, respectivamente. A estatística do teste Λ é definida por

$$\Lambda_{\text{obs}} = 2[\mathcal{L}(\hat{\Theta}_1) - \mathcal{L}(\hat{\Theta}_0)] \sim \chi^2_{k_1 - k_0}$$

Onde k_0 e k_1 representam, respectivamente, o número de parâmetros da distribuição Chen e Gama Exponenciada-Chen. O p -valor do teste então é definida por:

$$p\text{-valor} = \mathbb{P}(\chi^2_{k_1 - k_0} \geq \Lambda_{\text{obs}}),$$

em que Λ_{obs} é o valor da estatística calculado a partir dos dados. Se $p\text{-valor} < 0.05$, temos evidência estatística que o ajuste proporcionado pela Gama Exponenciada-Chen é superior.

5.4.1 Primeiro *Dataset*

O primeiro *dataset* contém dados referentes a degradação destrutiva obtida em um teste acelerado de um material de isolamento elétrico, mais especificamente a voltagem necessária para que o material falhe. Para as distribuições utilizadas e seus respectivos parâmetros testados foram a distribuição Gama com $\alpha = 3.7638425$ e $\beta = 0.3349427$, a distribuição Chen com $\beta = 0.667719569$ e $\lambda = 0.004255702$, a distribuição Gama generalizada com $a = 16.827239$, $d = 1.942226$ e $p = 12.820222$ e a distribuição Gama Exponenciada-Chen, com $\theta = 0.421851141$, $\beta = 0.683044833$ e $\lambda = 0.003940258$.

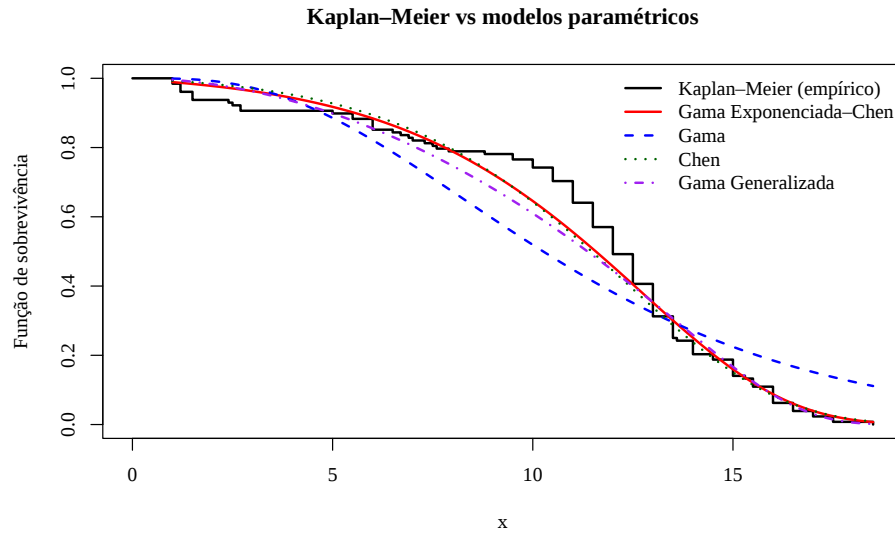


Figura 5.12: Curva de sobrevivência empírica (Kaplan–Meier) e funções de sobrevivência ajustadas das distribuições testadas para o primeiro conjunto de dados.

Tabela 5.4: Critérios de ajuste para o conjunto de dados 1

Modelo	$-\mathcal{L}(\hat{\Theta})$	AIC	AICc	BIC	p -valor
GE-C	-349.1120	704.2239	704.4175	712.7800	–
Gama	-394.3587	792.7173	792.8133	798.4214	–
Chen	-351.1609	706.3218	706.4178	712.0259	0.0429
Gama Generalizada	-352.2363	710.4727	710.6662	719.0288	–

Analisando a tabela 5.4 e a figura 5.12, podemos verificar que a GE-C apresentou bom ajuste aos dados, além de apresentar os menores valores para o AIC e AICc. Apesar

do BIC da GE-C ter sido o 2º melhor, sendo superado pela distribuição Chen, o teste de razão de verossimilhança indica que a distribuição GE-C proporciona um melhor ajuste.

5.4.2 Segundo *Dataset*

O segundo *dataset* apresenta a pressão máxima suportada por 63 fibras de vidro de 1,5 cm até que elas se rompam. É um conjunto de dados muito utilizado no campo de análise de sobrevivência, principalmente em estudos de novas distribuições de probabilidade. Foram utilizadas as distribuições distribuição Gama com $\alpha = 17.43906$ e $\beta = 11.57322$, a distribuição Chen com $\beta = 1.96040288$ e $\lambda = 0.07203649$, a distribuição Gama generalizada com $a = 1.769247$, $d = 4.810269$ e $p = 7.762520$ e a distribuição Gama Exponenciada-Chen, com $\theta = 1.045792$, $\beta = 1.602082$ e $\lambda = 0.269332$.

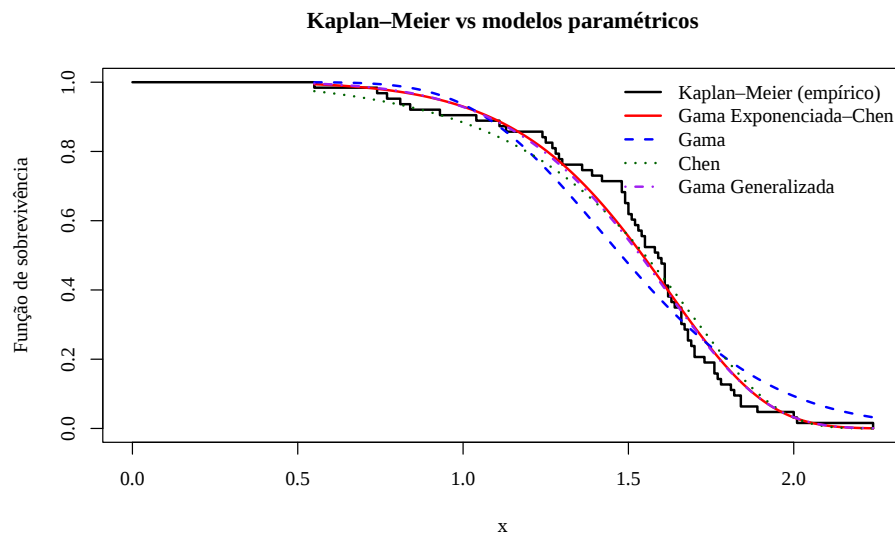


Figura 5.13: Curva de sobrevivência empírica (Kaplan-Meier) e funções de sobrevivência ajustadas das distribuições testadas para o segundo conjunto de dados.

Tabela 5.5: Critérios de ajuste para o conjunto de dados 2

Modelo	$-\mathcal{L}(\hat{\Theta})$	AIC	AICc	BIC	p -valor
GE-C	-14.36054	34.72108	35.12786	41.15048	—
Gama	-23.95154	51.90308	52.10308	56.18935	—
Chen	-16.46133	36.92267	37.12267	41.20894	0.0403
Gama Generalizada	-14.58761	35.17522	35.58200	41.60462	—

Analisando as tabela 5.5 e a figura 5.13, podemos verificar que a GE-C apresentou o melhor ajuste comparado a todas as outras, mostrando mais uma vez sua flexibilidade, mesmo entre em distribuições com o mesmo número de parâmetros.

6 Conclusão

6.1 Discussão dos resultados

Neste trabalho foi proposta e estudada a família de distribuições de probabilidade Gama Exponenciada-X (GE-X), construída a partir do método T-X com geradora Gama Exponenciada. Foram obtidas expressões gerais para as funções de distribuição acumulada, densidade, sobrevivência e risco da família, bem como sua representação em série, o que permitiu deduzir propriedades importantes como momentos, função geratriz de momentos, estatísticas de ordem e entropia de Rényi. Esse arcabouço teórico mostra que a família GE-X constitui uma classe ampla e flexível para modelagem de dados de sobrevivência e confiabilidade.

Dentre os casos particulares, foi dada ênfase à distribuição Gama Exponenciada-Chen (GE-C). A análise dos gráficos da função densidade e, principalmente, das funções de risco evidenciou que a GE-C é capaz de reproduzir diferentes formatos de risco ao longo do tempo, com comportamentos crescente, unimodal e em forma de curva da banheira, característica desejável em aplicações de confiabilidade.

Os resultados para os momentos e para as superfícies de assimetria e curtose mostram que, ao variar seus parâmetros, a GE-C consegue gerar distribuições desde moderadamente assimétricas até fortemente assimétricas, com níveis de curtose que vão de valores próximos aos de distribuições clássicas a casos mais leptocúrticos. Essa variedade de combinações reforça a capacidade da GE-C de ajustar conjuntos de dados com diferentes graus de assimetria e caudas mais pesadas.

Os estudos de simulação de Monte Carlo realizados em dois cenários paramétricos distintos indicaram que os estimadores de máxima verossimilhança para os parâmetros θ , β e λ apresentam o comportamento esperado do ponto de vista assintótico. À medida que o tamanho amostral aumenta, observa-se, em geral, redução do vício (BIAS), do erro

quadrático médio (MSE), do erro absoluto médio (MAE) e da raiz do erro quadrático médio (RMSE), o que é consistente com as propriedades de consistência e eficiência assintótica dos EMV. Em amostras menores, nota-se um viés moderado sobretudo nas estimativas de θ e β , mas esse efeito tende a se reduzir para tamanhos amostrais maiores, sugerindo que o método de estimação é mais estável em situações com maior quantidade de informação.

Na etapa de aplicação a dados reais, a distribuição GE-C foi ajustada a dois conjuntos de dados clássicos de análise de sobrevivência, sendo comparada com modelos concorrentes de dois e três parâmetros, como as distribuições Gama, Chen e Gama Generalizada. No primeiro conjunto de dados, a GE-C apresentou os menores valores de AIC e AICc, além de bom alinhamento gráfico com a curva de Kaplan–Meier. Embora o BIC tenha favorecido ligeiramente a distribuição de Chen, o teste de razão de verossimilhança indicou, ao nível de 5% de significância, evidência estatística em favor da GE-C frente ao modelo de Chen, mostrando que o parâmetro adicional é justificável em termos de ajuste. No segundo conjunto de dados, a GE-C voltou a apresentar o melhor desempenho global entre os modelos avaliados, com menores valores de AIC, AICc e BIC e boa aderência visual à função de sobrevivência empírica, reforçando a competitividade da distribuição proposta em relação a modelos já consolidados na literatura.

6.2 Considerações finais

De modo geral, os resultados teóricos, de simulação e de aplicação indicam que a família de distribuições Gama Exponenciada-X se mostra como uma alternativa promissora para modelagem de dados, proporcionando grande mobilidade e alterações na forma da função de risco de distribuições já existentes, além de boa capacidade de ajuste quando comparada a modelos tradicionais. Por outro lado, o estudo também evidencia algumas limitações práticas, como a maior complexidade das expressões analíticas e a sensibilidade numérica dos procedimentos de estimação em algumas configurações de parâmetros e tamanhos amostrais mais reduzidos, o que demanda maior cuidado na escolha de valores iniciais e na implementação computacional.

Como perspectivas futuras, destacam-se a extensão da análise para cenários com censura (à direita e intervalar), a investigação de métodos alternativos de estimação, como abordagens bayesianas e métodos robustos, bem como a inclusão de covariáveis por meio de modelos de regressão baseados na GE-C. Outra linha natural de continuidade é explorar outros casos particulares da família GE-X em diferentes áreas de aplicação, comparando-os com modelos consagrados em bases de dados de diferentes naturezas. Tais desenvolvimentos podem consolidar ainda mais a família Gama Exponenciada-X como uma ferramenta versátil para a modelagem estatística em análise de sobrevivência e confiabilidade.

Referências Bibliográficas

AGUILAR, G. A. S. *Família Distribuição Gama Exponenciada*. 98 p. Dissertação (Dissertação (Mestrado em Matemática Aplicada e Computacional)) — Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Programa de Pós-Graduação em Matemática Aplicada e Computacional, Presidente Prudente, SP, 2017. Disponível em: <<http://hdl.handle.net/11449/149970>>.

AGUILAR, G. A. S.; MOALA, F. A.; CORDEIRO, G. M. Zero-truncated poisson exponentiated gamma distribution: Application and estimation methods. *Journal of Statistical Theory and Practice*, v. 13, p. 57, 2019. Disponível em: <<https://doi.org/10.1007/s42519-019-0059-2>>.

AHMAD, Z. The zubair-g family of distributions: Properties and applications. *Annals of Data Science*, v. 7, 06 2020.

AHMAD, Z.; ELGARHY, M.; HAMEDANI, G. G. A new Weibull-X family of distributions: properties, characterizations and applications. *Journal of Statistical Distributions and Applications*, v. 5, n. 1, p. 1–18, December 2018. Disponível em: <https://ideas.repec.org/a/spr/jstada/v5y2018i1d10.1186_s40488-018-0087-6.html>.

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, 1974.

ALZAATREH, A.; GHOSH, I. On the weibull-x family of distributions. *Journal of Statistical Theory and Applications*, v. 14, n. 2, 2015. Disponível em: <<https://doi.org/10.2991/jsta.2015.14.2.5>>.

ALZAATREH, A.; LEE, C.; FAMOYE, F. A new method for generating families of continuous distributions. *METRION*, v. 71, 2013. Disponível em: <<https://doi.org/10.1007/s40300-013-0007-y>>.

ALZAATREH, A.; LEE, C.; FAMOYE, F. The gamma-normal distribution: Properties and applications. *Computational Statistics & Data Analysis*, v. 69, p. 67–80, 2014.

ANZAGRA, L.; SARPONG, S.; NASIRU, S. Chen-g class of distributions. *Cogent Mathematics & Statistics*, v. 7, n. 1, p. 1721401, 2020.

ATCHADÉ, M. N. et al. A new shifted lomax-x family of distributions : Properties and applications to actuarial and financial data. *Computational Journal of Mathematical and Statistical Sciences*, v. 4, p. 41–71, 11 2024.

BOURGUINON, M.; SILVA, R. B.; CORDEIRO, G. M. The weibull-g family of probability distributions. *Journal of Data Science*, v. 12, n. 1, p. 53–68, 2014.

CASELLA, G.; BERGER, R. *Statistical Inference*. Duxbury Thomson Learning, 2002. (Duxbury advanced series). ISBN 9780495391876. Disponível em: <<https://books.google.com.br/books?id=ZpkPPwAACAAJ>>.

- CHEN, Z. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters*, v. 49, n. 2, p. 155–161, 2000.
- CORDEIRO, G. M.; AND, M. de C. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 81, n. 7, p. 883–898, 2011. Disponível em: <<https://doi.org/10.1080/00949650903530745>>.
- CORDEIRO, G. M. et al. The lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, v. 247, p. 465–486, 2014. ISSN 0096-3003. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0096300314>>.
- CORDEIRO, G. M.; ORTEGA, E. M. M.; CUNHA, D. C. C. The exponentiated generalized class of distributions. *Journal of Data Science*, v. 11, p. 1–27, 2013.
- DELIGNETTE-MULLER, M. L.; DUTANG, C. fitdistrplus: An r package for fitting distributions. *Journal of Statistical Software*, v. 64, n. 4, p. 1–34, 2015. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v064i04>>.
- EUGENE, N.; LEE, C.; FAMOYE, F. Beta-normal distribution and its applications. *Communications in Statistics—Theory and Methods*, v. 31, p. 497–512, 2002.
- GUPTA, R. C.; GUPTA, P. L.; GUPTA, R. D. Modeling failure time data by lehman alternatives. *Communications in Statistics—Theory and Methods*, v. 27, n. 4, p. 887–904, 1998.
- IBRAHIM, S.; BAKOBAN, R. Exponentiated gamma distribution: Different methods of estimations. *Journal of Applied Mathematics*, 2012. Disponível em: <<https://doi.org/10.1155/2012/284296>>.
- LIANG, H.; ZOU, G. Improved aic selection strategy for survival analysis. *Computational Statistics & Data Analysis*, v. 52, n. 5, p. 2538–2548, 2008. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/19158943/>>.
- MARSHALL, A. W.; OLKIN, I. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, v. 84, n. 3, p. 641–652, 1997.
- NADARAJAH, S.; CORDEIRO, G. M.; ORTEGA, E. M. M. The zografos–balakrishnan-g family of distributions: Mathematical properties and applications. *Communications in Statistics—Theory and Methods*, v. 44, 2015. Disponível em: <<https://doi.org/10.1080/03610926.2012.740127>>.
- NADARAJAH, S.; KOTZ, S. The beta exponential distribution. *Reliability Engineering and System Safety*, v. 91, n. 6, p. 689–697, 2006. Disponível em: <<https://ideas.repec.org/a/eee/reensy/v91y2006i6p689-697.html>>.
- NADARAJAH, S.; ROCHA, R. Newdistns: An r package for new families of distributions. *Journal of Statistical Software*, v. 69, n. 10, p. 1–32, 2016. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v069i10>>.

NELSON, W. *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*. Wiley, 2004. (Wiley series in probability and statistics). ISBN 9780470317471. Disponível em: <<https://www.wiley.com/en-us/Accelerated+Testing%3A+Statistical+Models%2C+Test+Plans%2C+and+Data+Analysis+-p-9780470317471>>.

NOFAL, Z. et al. The generalized kumaraswamy-g family of distributions. *Journal of Statistical Theory and Applications*, 09 2018.

OLUYEDE, B. et al. The gamma-weibull-g family of distributions with applications. *Austrian Journal of Statistics*, v. 47, n. 1, p. 45–76, Jan. 2018. Disponível em: <<https://www.ajs.or.at/index.php/ajs/article/view/vol47-1-3>>.

POURREZA, H.; JAMKHANEH, E. B.; DEIRI, E. A family of gamma-generated distributions: Statistical properties and applications. *Statistical Methods in Medical Research*, v. 30, n. 8, p. 1850–1873, 2021.

RÉNYI, A. On measures of entropy and information. In: . [s.n.], 1961. Disponível em: <<https://api.semanticscholar.org/CorpusID:123056571>>.

RISTIC, M.; BALAKRISHNAN, N. The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, v. 82, p. 1191–1206, 08 2012.

SALAHUDDIN, N. et al. A novel generalized family of distributions for engineering and life sciences data applications. *Mathematical Problems in Engineering*, v. 2021, 05 2021.

SAPKOTA, L. et al. New lomax-g family of distributions: Statistical properties and applications. *AIP Advances*, v. 13, 09 2023.

SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v. 6, n. 2, p. 461–464, 1978. Disponível em: <<https://www.jstor.org/stable/2958889>>.

SMITH, R.; NAYLOR, J. A comparison of maximum likelihood and bayesian estimators for the three-parameter weibull distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 36, n. 3, p. 358–369, 1987. Disponível em: <<https://www.jstor.org/stable/2347795>>.

TAHIR, M. et al. A new weibull-g family of distributions. *Hacettepe Journal of Mathematics and Statistics*, v. 45, n. 2, p. 629–647, 2016.

THOMAS, E.; DAVID, I.; ADUBISI, O. Inverse odd weibull generated family of distribution. *Pakistan Journal of Statistics and Operation Research*, p. 617–633, 08 2020.

YOUSOF, H. M. et al. The transmuted exponentiated generalized-g family of distributions. *Pakistan Journal of Statistics and Operation Research*, v. 11, p. 441–464, 12 2015.

ZOGRAFOS, K.; BALAKRISHNAN, N. On families of beta- and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, v. 6, n. 4, p. 344–362, 2009.