

Universidade de Brasília
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

**A Distribuição Gama Bimodal: Propriedades,
Modelo de Regressão e Aplicações**

Letícia Ferreira Reis

Projeto apresentado para obtenção do título
de Bacharel em Estatística

Letícia Ferreira Reis

A Distribuição Gama Bimodal: Propriedades, Modelo de Regressão e Aplicações

Orientador:
Prof. Dr. Roberto Vila Gabriel

Projeto apresentado para obtenção do título
de Bacharel em Estatística

Brasília
2020

Agradecimentos

À minha mãe, Elaine, por ter sempre me apoiado e me ajudado, por ter se esforçado para me agradar quando os planos do TCC fugiam do planejado, por cozinhar minhas comidas favoritas, por me ensinar novos *hobbies*, por me ouvir e por seu senso de humor.

Ao meu pai, Saulo, por ter me estimulado a estudar desde que era pequena, por ter se esforçado para me dar uma boa educação, por ter me dado condições para que eu pudesse me dedicar aos estudos e por todo o carinho.

À minha vó, Maria José, e aos demais familiares por todo o carinho que sempre me dedicaram e por terem sido compreensivos nos momentos em que estive ausente devido aos afazeres acadêmicos.

Aos orientadores Roberto Vila e Helton Saulo por sempre responderem minhas dúvidas prontamente, por apresentarem um projeto bem estruturado para meu TCC e por me ensinarem como produzir um artigo científico em uma revista internacional de estatística.

Ao Prof. Edwin Ortega e Fábio Prativiera, por colaborarem com a construção do artigo que originou esse trabalho e pela solicitude que demonstraram.

A Rebeca e às Professoras Ana Maria e Maria Tereza, por me orientarem em dois PIBIC, por me ensinarem o passo a passo de um trabalho científico, por me ensinarem as melhores formas de escrever e de me comunicar, pela empatia que demonstram e por serem mulheres inspiradoras.

Às minhas amigas Emanuele, Emilly, Ana Caroline, Clara e ao meu amigo Gabriel Oliveira por me proporcionarem momentos incríveis, por me ajudarem a relaxar, pelas risadas, por apoiarem meus planos e por todos os conselhos.

Ao meu namorado, Rodrigo, por todo o companheirismo, por estudar comigo, por me ajudar no TCC e por me lembrar de cuidar da saúde.

Aos amigos que a Estatística proporcionou: João Gabriel, Gauthier, Luiz Eduardo, João Marcos, Lucas Alfaia e Wallace por me ajudarem nas atividades do curso, por terem me motivado e pelos momentos de relaxamento.

À equipe do Dimensionamento da Força de Trabalho no Serviço Público por me mostrarem como funciona um projeto de pesquisa multidisciplinar, por mostrarem como modelos matemáticos impactam na tomada de decisões.

Às equipes GESIR e GERIV da Caixa Econômica Federal, em especial a Carlos, Luiz Enciso, Mauro, Cecília, Fábio e Tiago por me introduzirem ao mundo da ciência de dados, por me apresentarem problemas estatísticos do mundo real, por tudo de programação que me ensinaram e pela forma que me tratavam.

À equipe Predictors do Banco do Brasil, em especial, Filipe, Leandro, Rafael, Giovanni e Slompo por toda a paciência que tiveram no meu primeiro emprego, pelas conversas, pelas dicas e por tudo que me ensinaram sobre modelagem.

Resumo

Este trabalho baseia-se no artigo: *A Bimodal Gamma Distribution: Properties, Regression Model and Applications*, escrito por R. Vila, L. Ferreira, H. Saulo, F. Prata e E. M. M. Ortega, publicado no periódico *Statistics* em 2020 [30]. Um dos objetivos principais deste trabalho é propor uma distribuição Gama bimodal utilizando uma transformação quadrática baseada no modelo Alpha-skew-normal. Ao longo dos capítulos foram abordadas várias propriedades dessa distribuição como média, variância, momentos, taxa de risco e medidas de entropia. Além disso, estimou-se os parâmetros da distribuição pelo método de máxima verossimilhança e verificou-se a consistência desses estimadores por simulações de Monte Carlo. Também foi apresentado um modelo de regressão para dados censurados baseado na distribuição Gama bimodal. A aplicação deste modelo em dois bancos de dados resultou em uma melhora na modelagem quando comparado a outros modelos conhecidos.

Palavras-chave: Distribuição bimodal. Distribuição Gama. Simulação de Monte Carlo. Modelo de regressão. GAMLSS.

Abstract

This work is based on the article: *A Bimodal Gamma Distribution: Properties, Regression Model and Applications*, authored by R. Vila, L. Ferreira, H. Saulo, F. PrataViera and E. M. M. Ortega, published in the journal *Statistics* in 2020 [30]. One of the main objectives of this work is to propose a bimodal Gamma distribution using a quadratic transformation based on the alpha-skew-normal model. In the following chapters, several properties of this distribution such as mean, variance, moments, hazard rate and entropy measures were discussed. Also, Monte Carlo simulations were performed to check the bias in the maximum likelihood estimation. Further, a new regression model with censored data based on the bimodal gamma distribution was proposed. This regression model was applied to two real datasets found in the literature and compared to other regression models. The bimodal gama regression model gave the better fits in the two datasets.

Keywords: Bimodal distribution. Gamma distribution. Monte Carlo simulation. Regression model. GAMLSS.

Conteúdo

Introdução	9
1 Algumas Definições Básicas	12
1.1 Variáveis Aleatórias	12
1.2 Esperança Matemática	14
1.3 Momentos reais	14
1.4 Estimação de Parâmetros e Informação de Fisher	16
1.4.1 Simulações de Monte Carlo	17
1.5 Medidas de Sobrevivência	18
1.6 Medidas de Entropia	19
1.7 Métricas de Qualidade de Ajuste	20
2 Distribuição Gama	23
2.1 Parâmetros	23
2.2 Moda	24
2.3 Momentos, Média, Variância e Função Geradora de Momentos	25
2.4 Medidas de Confiabilidade	25
2.5 Entropia de Shannon	27
2.6 Estimadores de Máxima Verossimilhança e Matriz de Informação	27
3 A Distribuição Gama Bimodal: BGama(θ_δ)	29
3.1 A Função de Densidade Gama bimodal	29
3.2 Caracterização de Bimodalidade e Unimodalidade	31
3.3 Momentos reais, Variância e Função Geradora de Momentos	35
3.4 Medidas de Entropia	41
3.5 Confiabilidade, taxa de risco e vida média residual	44
4 Estimação por Máxima Verossimilhança na BGama(θ_δ)	48
4.1 Simulações de Monte Carlo	53

5	O Modelo de Regressão BGama(θ_δ) com Censura	54
6	Aplicações	56
6.1	Aplicação 1. Dados do Rio Wheaton	56
6.2	Aplicação 2. Dados de câncer gástrico	59
	Conclusões	62
	Bibliografia	63
	Apêndice	66

Lista de Figuras

2.1	FDP para diferentes valores de parâmetros na distribuição Gama.	24
2.2	Função de risco da distribuição Gama para diferentes valores de α	26
3.1	A FDP da distribuição BGama para alguns valores dos parâmetros, com $\beta = 0,7$	30
3.2	Função de Risco da BGama para alguns valores dos parâmetro com $\beta = 0,7$	45
6.1	(a) PDF estimada do modelo BGama(θ_δ). (b) CDF empírica e estimada do modelo BGama(θ_δ). (c) QQ plot dos resíduos quantílicos do ajuste.	58
6.2	Gráficos de diagnóstico para o modelo de regressão BGama em dados de câncer gástrico. (a) Gráfico de índices \hat{q}_i (b) Função de densidade estimada para \hat{q}_i . (c) <i>QQ-plot</i> para \hat{q}_i . (d) <i>Worm plot</i> para \hat{q}_i	60
6.3	Função de Sobrevivência estimada do modelo de regressão BGama(θ_δ) e a função empírica de sobrevivência. (a) Ajuste considerando apenas o parâmetro $\alpha_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)$. (b) Ajuste considerando ambos os parâmetros $\alpha_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)$ e $\beta_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)$, para os dados de câncer gástrico.	61

Lista de Tabelas

4.1	Viés dos valores simulados (EQM nos parênteses) dos estimadores de Máxima Verossimilhança do Modelo BGama	53
6.1	Medidas KS e p -valores associados aos dados do Rio Wheaton	56
6.2	MLEs dos parâmetros do modelo para os dados do Rio Wheaton	57
6.3	Medidas Estatísticas	58
6.4	MLE, SE e p -valores dos parâmetros do modelo de regressão BGama(θ_δ) nos dados de câncer gástrico	59
6.5	Medidas GD, AIC e BIC para os modelos de regressão BGama, Gama e Weibull para os dados de câncer gástrico	60

Introdução

Nas últimas décadas, a quantidade de dados cresceu de forma exponencial. Diversas áreas do saber passaram a monitorar suas informações e a tentar modelar eventos a fim de obter um panorama mais claro de seus estudos. Com esse aumento no número de bancos de dados, percebeu-se que tanto os fenômenos humanos como os naturais possuem certos aspectos que não podem ser representados por distribuições clássicas sem perda de informação. Nesses casos, é necessário buscar novas distribuições para melhorar a precisão das inferências e das previsões.

Uma situação comum em que as distribuições clássicas falham é quando os dados apresentam dois intervalos com frequências maiores que os demais, ou seja, quando existem duas modas. Esse evento é conhecido como bimodalidade, e pode ser encontrado, por exemplo, em pixels de imagens de placas de aterosclerose (Shankar, 2003 [28]), em processos para controle de qualidade (Hsu, 2008 [18]), no tamanho de peixes (Rao et. al, 2008 [26]), no diâmetro de ovos (Famoye, 2004 [10]) e no tempo em que determinado vulcão entra e sai de erupção (Weisberg, 2005 [33]).

A bimodalidade pode aparecer sob diferentes aspectos: os picos de frequência podem estar mais próximos ou mais separados, podem possuir graus diferentes de assimetria, podem exibir magnitudes distintas. A curtose também varia nessas situações, de modo que um dos picos pode ser mais achatado, enquanto outro pode possuir aspecto afunilado. Há uma gama de combinações possíveis, por essa razão é interessante que a modelagem nesses cenários seja feita por uma distribuição cuja função de densidade de probabilidade (FDP) além de bimodal, também seja flexível.

A mistura de distribuições é a alternativa mais praticada nesses casos. Essa técnica consiste na soma ponderada de k distribuições $\{g(x; \Theta_1), \dots, g(x; \Theta_k)\}$ (Ghojogh, 2019 [11]), isto é,

$$f(x; \Theta_1, \dots, \Theta_k) = \sum_{i=1}^k w_i g(x; \Theta_i),$$

onde $\sum_{i=1}^k w_i = 1$.

O uso de misturas, no entanto, pode levar a problemas de identificabilidade (Gómez e

Gallardo, 2019 [13]). A identificabilidade é a propriedade que garante que os parâmetros do modelo, quando estimados, possuam solução única. Quando essa propriedade não é satisfeita, um mesmo conjunto de dados pode resultar em mais de um modelo probabilístico, de forma que as inferências nesse conjunto perdem sua precisão (Godfrey, 1985 [12]).

Além do mais, a literatura foca em misturas de distribuições Normal, deixando de lado distribuições mais flexíveis como a distribuição Gama (Ahmed e Hussein, 2008 [1]). Além de possuírem menor flexibilidade, sabe-se que os principais algoritmos para a construção de uma mistura Normal com determinada matriz de covariância não são eficientes e nem acurados (Wang e Taaffe, 2015 [32]). Também existe a dificuldade de determinar o número k de distribuições normais necessárias para que a mistura possua um bom ajuste.

Esse processo também pode exigir que um número elevado de parâmetros seja estimado. Um mistura bimodal de normais possui pelo menos cinco parâmetros: as médias (μ_1, μ_2) , as variâncias (σ_1^2, σ_2^2) e o parâmetro de mistura (π) . A medida que se aumenta o número k de distribuições, o número de parâmetros cresce ainda mais. Como a estimação é feita por algoritmos iterativos, esse processo é demorado e não possui garantias de que irá convergir.

Pode-se também alcançar a bimodalidade por meio da adição de parâmetros à distribuições existentes. Alzaatreh (2015) [3], por exemplo, introduziu uma nova família de distribuições, a Gama generalizada ¹ ao adicionar dois parâmetros a distribuição Gama. Hassan (2010) [17] também chegou a esse resultado ao adicionar um parâmetro de bimodalidade à distribuição exponencial, de modo que sua distribuição apresentava: um parâmetro de locação, dois parâmetros de escala e um parâmetro de bimodalidade, totalizando quatro parâmetros.

Outra forma de encontrar distribuições bimodais é por meio da multiplicação de uma distribuição unimodal por um fator quadrático. Elal-Olivero (2010) [8] realizou essa operação em uma densidade Normal e assim introduziu assimetria e bimodalidade à função, criando a Alpha-Skew Normal (ASN). Uma variável aleatória Z possui distribuição ASN caso sua função densidade de probabilidade (FDP) e sua função de distribuição acumulada (FDA), respectivamente, sejam dadas por

$$g(z) = \frac{1 + (1 - \delta z)^2}{2 + \delta^2} \phi(z) \quad \text{e} \quad G(z) = \Phi(z) + \delta \left(\frac{2 - \delta z}{2 + \delta^2} \right) \phi(z), \quad z, \delta \in \mathbb{R},$$

onde δ controla a bimodalidade, Φ e ϕ representam a Função de Distribuição de Probabilidade (FDP) e Função de Distribuição Acumulada da distribuição Normal, respectivamente.

Esse tipo de procedimento permite propor distribuições bimodais que possuem um me-

¹A FDP da Gama generalizada é dada por: $f(x; \theta, \beta, \alpha) = \begin{cases} \frac{\beta}{\Gamma(k)\theta} \left(\frac{x}{\theta}\right)^{k\beta-1} e^{-\left(\frac{x}{\theta}\right)^\beta}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

nor número de parâmetros e logo exigem menor esforço computacional durante o processo de estimação. Nessas distribuições, também, é mais fácil encontrar fórmulas fechadas para as propriedades de interesse.

Assim, o presente estudo busca investigar o comportamento de uma distribuição bimodal originada de uma distribuição Gama multiplicada pelo mesmo fator quadrático proposto por Elal-Olivero (2010) [8]. Uma vez que a distribuição Gama é conhecida por sua flexibilidade, pretende-se propor uma nova distribuição, que além de bimodal, seja flexível. Portanto, esse é um trabalho inédito com o intuito de apresentar essa distribuição e suas características.

Esta dissertação está organizada em 6 capítulos. Nos dois primeiros capítulos deste trabalho foram recordadas algumas noções fundamentais de Teoria de Probabilidades e Estatística, que serão utilizadas à medida que forem necessárias na exposição desta dissertação. Já nos outros capítulos o foco principal é o estudo do comportamento da distribuição Gama bimodal. Mais especificamente, no Capítulo 1 é feito um breve resumo sobre algumas noções básicas de probabilidade e estatística; no Capítulo 2 são apresentadas algumas das propriedades da distribuição Gama; no Capítulo 3 introduz-se a distribuição Gama bimodal e suas respectivas propriedades matemáticas; no Capítulo 4 é feita a estimação de parâmetros por máxima verossimilhança e é verificada a performance dessa estimação por simulações de Monte Carlo; no Capítulo 5 é proposto um modelo de regressão baseado na distribuição Gama bimodal; e no Capítulo 6 é feita a aplicação desse modelo em dois bancos de dados encontrados na literatura.

Algumas Definições Básicas

Antes de apresentar a distribuição Gama bimodal, é importante abordar alguns conceitos teóricos de Probabilidade e Estatística, que serão necessários para a demonstração de algumas das propriedades principais da distribuição proposta neste trabalho e para a construção e avaliação do modelo de regressão associado. Sendo assim, nesta seção serão abordadas noções teóricas de variáveis aleatórias, esperança matemática, momentos reais, estimação de parâmetros, matriz Informação de Fisher, medidas de sobrevivência, medidas de Entropia e Métricas de Qualidade de Ajuste.

1.1 Variáveis Aleatórias

Em situações da vida real o estatístico realiza experimentos, determinísticos ou aleatórios, e em seguida observa os resultados. Uma vez que os resultados associados a experimentos determinísticos podem ser preditos com certeza, estes não são de interesse na Teoria da Probabilidade. Uma experiência simples mostra que é possível descrever o conjunto de todos os resultados possíveis ao repetir um experimento aleatório um grande número de vezes. Este conjunto é conhecido como espaço amostral, e frequentemente é representado pela letra grega Ω . De forma não rigorosa, um evento, denotado por letras maiúsculas (A, B, C, \dots) , pode ser definido como um elemento da classe de subconjuntos do espaço amostral com as seguintes restrições: que a classe, chamada de σ -álgebra, seja fechada por complementares e por uniões enumeráveis. Uma função $\mathbb{P}(\cdot)$ que associa a cada evento A um número no intervalo $[0, 1]$ é conhecida como medida de probabilidade sempre que $\mathbb{P}(\Omega) = 1$ e a função $\mathbb{P}(\cdot)$ seja σ -aditiva. Em muitas situações, é comum ter conjuntos numéricos associados ao resultado de um experimento aleatório. Na teoria de probabilidade, estes conjuntos numéricos, sobre restrições de mensurabilidade com respeito à σ -álgebra, são conhecidos como variáveis aleatórias e denotados por X, Y, Z , etc. Um exemplo simples de variável aleatória é descrito, de maneira informal, pelo número de vezes que um colega frequenta o bar em uma semana.

Dependendo da classe do conjunto de valores (contradomínio) que uma variável aleatória assume, ela é classificada como discreta, contínua ou mista. Dentro das classes de variáveis aleatórias contínuas tem-se as variáveis absolutamente contínuas. Doravante, neste trabalho, por simplicidade escrevemos a expressão variável aleatória contínua para referirmos a uma variável absolutamente contínua. Se X é uma variável aleatória contínua, a probabilidade de se obter um resultado menor ou igual a um número real x é dada pela função de distribuição acumulada (FDA).

Definição 1.1.1. *Se X uma variável aleatória contínua, existe uma função F_X não negativa de $x \in \mathbb{R}$, tal que (Magalhães, 2015 [21])*

$$F_X(x) = \int_{-\infty}^x f(x) dx = \mathbb{P}(X \leq x),$$

onde $f(\cdot)$ é a função densidade de probabilidade (FDP) associada à X .

A Função Densidade de Probabilidade (FDP) representa a taxa de variação instantânea da FDA naquele valor x . Se $f(x)$ é uma FDP então deve cumprir certos pressupostos.

Proposição 1.1.1. *Se $f(x)$ é uma FDP, então (Magalhães, 2015 [21])*

$$(1) f(x) \geq 0, \forall x \in \mathbb{R};$$

$$(2) \int_{-\infty}^{\infty} f(x) dx = 1.$$

A moda da distribuição (contínua) de uma variável aleatória pode ser interpretada como o(s) valor(es) mais prováveis de serem amostrados. Note que a partir da FDP sempre é possível determinar a moda.

Definição 1.1.2. *A moda de uma distribuição contínua é o valor no qual a FDP atinge seu máximo local ou global.*

Por meio da FDP, também, é possível saber se determinada distribuição pertence à família exponencial. Essa família possui características que podem facilitar os cálculos de certas medidas relacionadas à distribuição.

Definição 1.1.3. *Uma família de FDPs é chamada de família exponencial se pode ser escrita da seguinte forma (Casella, 2010 [5])*

$$f(x; \theta) = h(x)c(\theta) \exp \left[\sum_{i=1}^k w_i(\theta)t_i(x) \right],$$

onde $h(x) \geq 0$ e $t_i(x)$ são funções das observações amostrais que não dependem do vetor de parâmetros θ . Enquanto $c(\theta) > 0$ e w_1, \dots, w_k são funções do vetor de parâmetros e independentes dos valores amostrais.

1.2 Esperança Matemática

A esperança matemática é uma medida muito importante para a modelagem e inferência estatística, visto que na abordagem frequentista, supõe-se que em um número infinito de repetições independentes de um experimento, o valor médio dos resultados converge para seu valor médio (Ross, 2010 [27]).

Definição 1.2.1. *A esperança matemática ou média de uma variável contínua X é definida por Magalhães, 2015 [21],*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

sempre que a integral esteja bem definida.

Na seguinte proposição serão enunciadas as propriedades básicas da esperança matemática.

Proposição 1.2.1. *Se X e Y são duas variáveis aleatórias definidas sobre o mesmo espaço amostral Ω , não necessariamente contínuas, então, para todo $a, b \in \mathbb{R}$ (Magalhães, 2015 [21]),*

(a) $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b;$

(b) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y);$

(c) *Se $X(\omega) \leq Y(\omega) \forall \omega \in \Omega$, então $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

Como a distribuição Gama bimodal é construída em função da distribuição Gama usual através de uma transformação quadrática, o conceito de esperança de funções de variáveis aleatórias, também, foi fundamental nesse estudo.

Teorema 1.2.1. *Seja g uma função (mensurável) de valor real. Se X é uma variável aleatória contínua, então a função composição $g(X)$ é uma variável aleatória contínua, logo tem sentido definir sua esperança, sempre que esta exista, como*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

1.3 Momentos reais

Os momentos de uma distribuição são características fundamentais para entender como a função se comporta. Sabendo-se os momentos pode-se derivar diversas medidas como média, variância, assimetria e curtose.

Definição 1.3.1. Dado $k \in \mathbb{N}$, o momento de ordem k de uma variável X é definido por $\mathbb{E}(X^k)$, sempre que essa quantidade exista. No caso que $k \in \mathbb{R}$, $\mathbb{E}(X^k)$ é conhecido como k -ésimo momento real. O momento central de ordem k , com $k \in \mathbb{N}$, é definido como

$$\mathbb{E}[(X - \mu)^k],$$

onde $E(X) = \mu$; veja por exemplo Magalhães, 2015 [21].

Note que o momento central de ordem 1 é zero. O momento central de ordem 2 é conhecido como variância. Essa medida indica o quanto a distribuição se dispersa.

Definição 1.3.2. Se o momento de ordem 2 é finito, a variância é definida como o momento central de ordem 2 (Magalhães, 2015 [21]), isto é,

$$\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mu)^2],$$

ou alternativamente, $\sigma^2 = \mathbb{E}(X^2) - \mu^2$, onde σ denota o desvio-padrão.

Já o momento central de ordem 3 dividido pelo cubo do desvio-padrão resulta no coeficiente de assimetria. A partir desse coeficiente pode-se saber se a distribuição é simétrica, assimétrica à direita ou assimétrica à esquerda. Valores próximos a 0 indicam simetria, valores negativos indicam assimetria à direita e valores positivos indicam assimetria à esquerda.

Definição 1.3.3. Se X é uma variável aleatória com desvio-padrão $\sigma > 0$ e média μ , seu coeficiente de assimetria, denotado por α_3 , é definido por (Magalhães, 2015 [21])

$$\alpha_3 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right].$$

O momento central de ordem 4 dividido pelo desvio-padrão elevado a quarta potência, por sua vez, é conhecido como coeficiente de curtose. Se esse coeficiente possui valor 3, a distribuição possui o mesmo achatamento da distribuição normal. Se a curtose é maior que 3 a distribuição é leptocúrtica, ou seja, é menos achatada que a distribuição normal, mas caso seja menor que 3, a função é platicúrtica e, portanto, mais achatada que a distribuição normal.

Definição 1.3.4. Se X é uma variável aleatória com desvio-padrão $\sigma > 0$ e média μ , seu coeficiente de curtose, denotado por α_4 , é definido por (Magalhães, 2015 [21]):

$$\alpha_4 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right].$$

Sobre certas condições, o cálculo dos momentos pode ser obtido a partir da função geradora de momentos.

Definição 1.3.5. *A função geradora de momentos de uma variável aleatória X é definida por*

$$M_X(t) = \mathbb{E}[\exp(tx)],$$

sempre que esta esperança seja finita em algum intervalo real $-t_0 < t < t_0$ para $t > 0$.

Teorema 1.3.1. *Se a função geradora de momentos de uma variável aleatória X está definida no intervalo $-t_0 < t < t_0$, então $\mathbb{E}(X^n)$ existe para $n = 1, 2, 3, \dots$; veja Magalhães, (2015) [21]. Em consequência, os momentos podem ser encontrados como segue*

$$\mathbb{E}(X^n) = M_X^{(n)}(t) \Big|_{t=0} = \frac{\partial^n}{\partial t^n} M_X(t) \Big|_{t=0}.$$

Sobre as condições do teorema acima, tem-se que a função geradora de momentos de X admite a seguinte expansão: $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r \mathbb{E}X^r}{r!}$; veja Johnson e Kotz, (1993) [19].

1.4 Estimação de Parâmetros e Informação de Fisher

Para estimar os parâmetros da distribuição Gama bimodal foi utilizado o método da máxima verossimilhança. Esse método consiste em escolher um valor do parâmetro para o qual a amostra observada é mais provável, buscando o ponto de máximo da função de verossimilhança.

Definição 1.4.1. *Seja X uma variável aleatória com distribuição $G(\theta)$ caracterizada pelo vetor de parâmetros populacionais $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Seja (X_1, \dots, X_n) uma amostra aleatória, de tamanho n , da variável aleatória X . Se $\mathbf{x} = (x_1, \dots, x_n)$ denota o valor observado fixo da amostra, a **função de máxima verossimilhança** para o vetor de parâmetros $\boldsymbol{\theta}$ é definida por*

$$L(\boldsymbol{\theta}; \mathbf{x}) = L(\theta_1, \dots, \theta_k; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_k).$$

Definição 1.4.2. *Se $\hat{\boldsymbol{\theta}}$ é um valor do vetor de parâmetros no qual a função $L(\boldsymbol{\theta}; \mathbf{x})$ atinge seu máximo como função de $\boldsymbol{\theta}$, onde $\mathbf{x} = (x_1, \dots, x_n)$ é um valor observado fixo da amostra, então $\hat{\boldsymbol{\theta}}$ é um **estimador de máxima verossimilhança** de $\boldsymbol{\theta}$ com base em uma amostra aleatória de X (Casella, 2010 [5]).*

Outra medida importante na estimação de parâmetros é conhecida como Informação de Fisher. Chama-se informação por possuir relação inversamente proporcional ao limite de variância de um estimador não viciado de variância uniformemente mínima (ENNVUM). Quanto maior a informação mais se sabe sobre o parâmetros e mais se limita a variância.

Definição 1.4.3. Denomina-se **Informação de Fisher** a quantidade definida por

$$I_F(\boldsymbol{\theta}) = \mathbb{E} \left(\left[\frac{\partial \log[f(x; \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right]^2 \right).$$

Sobre hipóteses padrões de regularidade, a Informação de Fisher pode ser escrita da seguinte maneira

$$I_F(\boldsymbol{\theta}) = -\mathbb{E} \left(\frac{\partial^2 \log[f(x; \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}^2} \right).$$

1.4.1 Simulações de Monte Carlo

Sabe-se que a estimação por máxima verossimilhança pode apresentar sensibilidade numérica, isto é, grandes mudanças nas estimativas podem ocorrer quando há pequenas mudanças nos dados. Por isso, o estimador de Monte Carlo é empregado para analisar o quanto variam os parâmetros estimados por máxima verossimilhança.

No nosso caso, considera-se uma variável aleatória X , com distribuição Gama Bimodal e vetor de parâmetros $\boldsymbol{\theta}$. A partir dessa variável são geradas r amostras aleatórias independentes de tamanho n e com a mesma distribuição de X . São então calculados os estimadores de máxima verossimilhança dos parâmetros da Gama Bimodal para cada uma dessas amostras. A partir desses valores pode-se obter um estimador dos parâmetros de X .

Suponha que X possui um parâmetro α . Um estimador para esse parâmetro baseado nos parâmetros estimados por máxima verossimilhança em cada amostra $(\hat{\alpha}_1, \dots, \hat{\alpha}_r)$ é dada por

$$\hat{\alpha} = \frac{1}{r} \sum_{i=1}^r \hat{\alpha}_i.$$

O viés da estimação é dada por

$$\text{Viés} = \hat{\alpha} - \alpha$$

e o erro quadrático médio é

$$EQM = \frac{1}{r} \sum_{i=1}^r (\hat{\alpha}_i - \hat{\alpha})^2 + \text{Viés}^2,$$

onde r é o número de amostras geradas.

Para gerar as amostras da variável aleatória X , utilizou-se o método da Transformada Inversa. Esse método consiste em simular amostras de uma variável aleatória U com distribuição Uniforme(0,1) e realizar a transformação dessa variável utilizando-se a inversa

da função de distribuição acumulada de X

$$X = F^{-1}(U),$$

onde F é a distribuição acumulada de X . Nos próximos capítulos é visto que é possível encontrar uma fórmula para a distribuição acumulada da Gama Bimodal.

1.5 Medidas de Sobrevivência

A teoria de análise de sobrevivência consiste no uso de técnicas estatísticas no estudo do tempo que um evento de interesse leva para ocorrer. Este estudo, além de permitir estimar o tempo de espera até o evento, também permite entender o efeito de outras variáveis sobre esse tempo. Por exemplo, na economia estas técnicas podem ser utilizadas para estimar o tempo até o pagamento de um empréstimo e na saúde podem ser utilizadas para estimar o efeito de um medicamento no tempo de sobrevivência de um paciente.

Assim, nesta seção serão apresentados alguns dos principais conceitos nesta área: censura, função de sobrevivência, função de risco, função de risco acumulada e vida média residual.

Uma **censura** pode ser interpretada como a ausência do evento de interesse durante o tempo de estudo. Existem três tipos principais de censura, os quais serão definidos e exemplificados abaixo.

- **Censura do Tipo I.** Este tipo de censura ocorre quando o estudo possui tempo pré-definido e o objeto de interesse não falha neste intervalo, então a informação sobre o tempo de falha dessa unidade não é completa. Neste caso a censura é uma variável aleatória. Por exemplo, o paciente sobreviveu além do tempo de estudo.
- **Censura do Tipo II.** Este segundo tipo de censura ocorre quando o pesquisador define que o experimento deve continuar até o decorrer de um determinado número de falhas. Ao escolher o número de falhas, escolhe-se também a quantidade de censuras, assim a censura deixa de ser uma variável aleatória. Por exemplo, define-se que um estudo com oito pacientes termina ao ocorrer de três óbitos, promovendo a censura dos outros cinco pacientes que não morreram.
- **Censura aleatória.** Esta classe de censura acontece quando não é possível acompanhar o objeto de estudo até o fim, ou a unidade falha devido a circunstâncias externas ao experimento. Esta censura define uma variável aleatória. Por exemplo, um paciente mudou de cidade e por isso deixou de ser acompanhado pelo estudo.

Definição 1.5.1. A função de sobrevivência, denotada por $S(t)$, é definida como a

probabilidade de uma observação não falhar até o tempo t . Em outras palavras,

$$S(t) = \mathbb{P}(T > t) = \int_t^{\infty} f(t) dt, \quad t > 0,$$

onde $f(t)$ é a FDP associada à variável tempo de vida T .

A função de sobrevivência é contínua e não crescente, de modo que $\lim_{t \rightarrow 0} S(t) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$.

Definição 1.5.2. A função de risco para variáveis contínuas, denotada por $h(t)$, é definida como a probabilidade instantânea de que um equipamento ou um ser vivo experimente um evento de interesse em um intervalo de tempo $[t, t + \Delta t)$, dado que esse evento não ocorreu antes do tempo t . Em outras palavras,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T > t)}{\Delta t}.$$

Como a função de sobrevivência é sempre monótona e não decrescente, é preferível analisar a função de risco para determinar se determinada distribuição é adequada a modelagem da sobrevivência dos dados.

Definição 1.5.3. A vida média residual, denotada por $MRL(t)$, é definida como o tempo de médio de espera até que determinado evento de interesse ocorra, dado que ele não ocorreu até o momento t . Em outras palavras,

$$MRL(t) = \frac{1}{S(t)} \int_t^{\infty} R(u) du, \quad t \geq 0,$$

onde $S(t)$ denota a função de sobrevivência.

1.6 Medidas de Entropia

A entropia representa a quantidade de incerteza contida em determinada distribuição, logo quanto maior o valor da entropia menos informação se tem sobre a variável. Existem várias maneiras de se quantificar essa medida e cada maneira é apropriada para uma forma de aplicação. No que segue, serão definidas três das principais entropias: a entropia de Shannon, a entropia Quadrática e a entropia de Renyi.

Definição 1.6.1. A entropia de Shannon associada a uma variável aleatória X é definida por

$$H_1(X) = -\mathbb{E}\{\log[f(X; \boldsymbol{\theta})]\} = -\int_0^{\infty} f(x; \boldsymbol{\theta}) \log[f(x; \boldsymbol{\theta})] dx.$$

A entropia de Shannon é amplamente utilizada dentro da área de teoria da informação, onde é aplicada em problemas de criptografia, transmissão de dados, codificação, etc.

Definição 1.6.2. A entropia Quadrática é mensurada pela seguinte fórmula:

$$H_2(X) = -\log \left[\int_0^\infty f^2(x; \boldsymbol{\theta}) dx \right].$$

Salienta-se que a entropia Quadrática possui aplicações na área de Ecologia e em algoritmos de *Machine Learning*.

Definição 1.6.3. A entropia de Renyi é definida por

$$H_\gamma(X) = -\frac{1}{1-\gamma} \log \left[\int_0^\infty f^\gamma(x; \boldsymbol{\theta}) dx \right], \quad \gamma \geq 0 \text{ e } \gamma \neq 1.$$

A entropia de Renyi constitui uma família de medidas de informação que generaliza a entropia de Shannon (Muller-Lennert, 2013 [23]).

1.7 Métricas de Qualidade de Ajuste

Ao propor uma nova distribuição, é de interesse saber se ela pode trazer ganhos na modelagem de situações reais. Por isso, é fundamental o uso de métricas de qualidade de ajuste para que seja possível comparar os modelos baseados na nova distribuição a modelos já conhecidos na literatura. Assim, abaixo estão descritas as métricas que foram utilizadas nesse estudo.

- **AIC (Akaike information criterion).** O critério de Akaike definido pela fórmula

$$AIC = 2k - 2 \log(\hat{L}).$$

Esse critério permite medir o quanto de informação os dados perdem pela modelagem, logo, quanto menor o valor do AIC, melhor o modelo. Aqui, k denota o número de parâmetros e \hat{L} é a medida de verossimilhança do modelo. Note que essa métrica penaliza modelos com maior número de parâmetros.

- **BIC (Bayesian information criterion).** O critério de informação Bayesiana nos fornece uma medida que serve para avaliar a performance do modelo. Sua fórmula, dada por

$$BIC = k \log(n) - \log(\hat{L}),$$

se assemelha à do AIC, entretanto há uma maior penalização a parâmetros adicionais. Aqui, k denota o número de parâmetros, \hat{L} é a medida de verossimilhança do

modelo e n é o tamanho da amostra. Assim como no AIC, os modelos com menor valor de BIC tendem a ser melhores.

- **W* (Cramér–von Mises criterion).** O critério de Cramér–von Mises é utilizado para avaliar o quão bem uma distribuição se ajusta aos dados. Para isso se compara a FDA dessa distribuição com a distribuição acumulada empírica das observações, através da seguinte fórmula:

$$T = \frac{1}{12n} \sum_{i=1}^n \left[\frac{2i-1}{2n} - F(x_i) \right]^2,$$

onde $F(\cdot)$ é a função de distribuição acumulada da distribuição de interesse e n é o tamanho da amostra. A partir da estatística T realiza-se um teste de hipóteses, sob a hipótese nula de que os dados seguem a distribuição de interesse.

- **KS (Kolmogorov-Smirnov).** Assim como W^* também é uma medida que avalia a distância entre a distribuição avaliada e a distribuição acumulada empírica dos dados. A estatística deste teste é dada por

$$D_n = \sup_x |F_n(x) - F(x)|,$$

onde $F(\cdot)$ é a FDA da distribuição avaliada e $F_n(\cdot)$ é a distribuição empírica das observações, dada pela seguinte fórmula.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[-\infty, x]}(X_i),$$

onde $I_{[-\infty, x]}(X_i)$ é uma função indicadora que possui valor 1 caso $X_i \leq x$

Nesse caso, a hipótese nula é de que a distribuição avaliada ajusta os dados de maneira satisfatória. Se a distribuição realmente é apropriada aos dados, a estatística D_n deve estar próxima a 0, mas se o ajuste é falho, a D_n assume valores altos.

- **Resíduos Quantílicos.** O Resíduo Quantílico é um método que possui o intuito de contornar a dificuldade em analisar os resíduos tradicionais. Quando a variável resposta é contínua, realiza-se a transformação *probit* da FDA F . No caso de variáveis discretas, antes de realizar a transformação é feita a randomização a fim de tornar a FDA contínua. Na equação a seguir pode-se ver como é obtida a função F^* , na qual será aplicada a transformação *probit*:

$$F^*_{*i} = \begin{cases} F(y_i, \hat{\mu}_i, \hat{\phi}), & \text{se } F \text{ é contínua,} \\ F(y_i^-, \hat{\mu}_i, \hat{\phi}) + u_i p(y_i, \hat{\mu}_i, \hat{\phi}), & \text{se } F \text{ é discreta,} \end{cases}$$

onde $\hat{\mu}_i$ e $\hat{\phi}$ denotam a média e o parâmetro de dispersão estimados, respectivamente. Aqui, u_i é uma variável aleatória com distribuição uniforme no intervalo $(0, 1]$ e $F(y_i^-, \hat{\mu}_i, \hat{\phi})$ é o limite inferior de F em y_i .

Assim, os resíduos quartílicos são encontrados pela transformação

$$r_{q,i} = \Phi^{-1}(F*),$$

onde $\Phi(\cdot)$ denota a FDA da distribuição normal padrão.

Quando o modelo ajustado é adequado, os resíduos $r_{q,i}$ possuem distribuição normal padrão.

Distribuição Gama

Como a distribuição Gama bimodal deste trabalho será definida em função do parâmetro $\delta \in \mathbb{R}$, que controla a bimodalidade da distribuição, e em função da distribuição Gama (clássica), através da aplicação de uma transformação quadrática; vários resultados da distribuição Gama desta seção serão úteis no desenvolvimento das demonstrações de propriedades da nova distribuição bimodal. Por esta razão, este capítulo tem como objetivo relembrar alguns elementos característicos da distribuição Gama.

Diz-se que uma variável aleatória Y possui distribuição Gama com vetor de parâmetros $\boldsymbol{\theta}_0 = (\alpha, \beta)$, $\alpha > 0, \beta > 0$, denotado por $Y \sim \text{Gama}(\boldsymbol{\theta}_0)$, quando sua FDP é descrita pela função

$$f(y; \boldsymbol{\theta}_0) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y), & y > 0, \\ 0, & y \leq 0, \end{cases}$$

onde $\Gamma(\alpha)$ denota a função Gama, a qual é definida pela seguinte integral imprópria

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \exp(-y) dy. \quad (2.1)$$

A partir de manipulações algébricas da Equação (2.1) é possível encontrar as seguintes propriedades da função Gama:

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, se $\alpha > 1$;
- $\Gamma(n) = (n - 1)!$, se n é inteiro e positivo.

2.1 Parâmetros

Por definição, a distribuição Gama depende dos parâmetros α e β . O parâmetro α está associado à forma da distribuição, isto é,

1. Se $\alpha < 1$, a distribuição é exponencial com caráter assintótico no eixo vertical e horizontal;
2. Se $\alpha = 1$, a distribuição é exponencial com média $\frac{1}{\beta}$;
3. Se $\alpha > 1$, a distribuição é unimodal e assimétrica à esquerda.

Na medida em que os valores de α aumentam, a distribuição tende a se tornar simétrica, e quando $\alpha \rightarrow \infty$ a densidade converge para a distribuição Normal.

O parâmetro β , por sua vez, determina a taxa de decaimento. Quanto maior o β , mais leves são as caudas da distribuição.

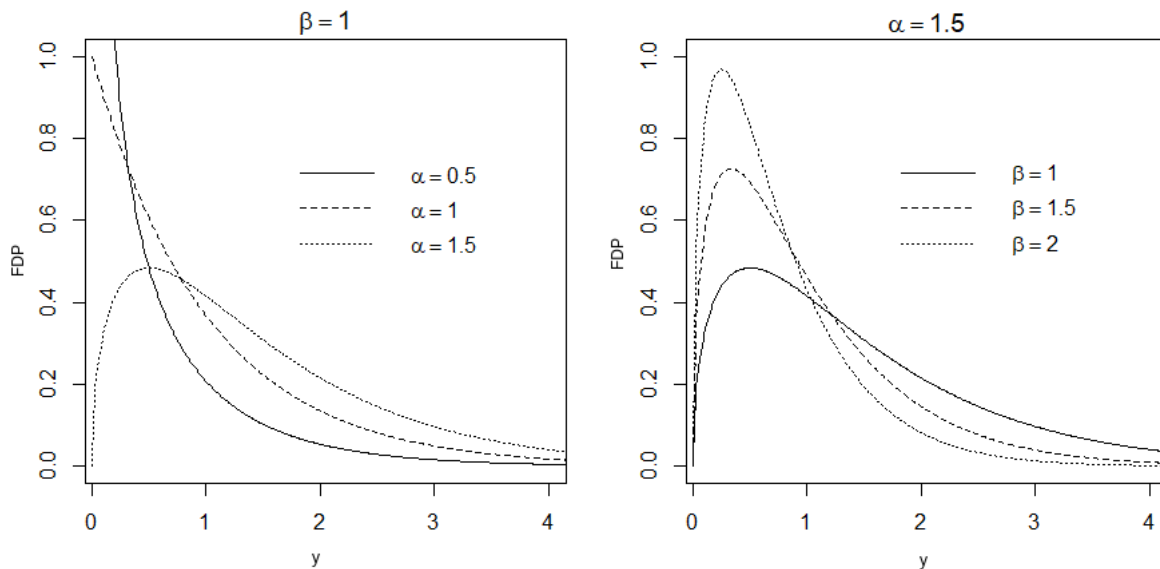


Figura 2.1: FDP para diferentes valores de parâmetros na distribuição Gama.

Dependendo dos valores de parâmetros escolhidos, podem-se obter outras densidades de probabilidades, também, muito conhecidas dentro da estatística. Por exemplo, se $\alpha = 1, \beta > 0$ obtém-se o modelo exponencial, se $\alpha = n/2, n > 0$ inteiro, $\beta > 1/2$ obtém-se a distribuição Qui-Quadrado com n graus de liberdade, e se $\alpha = k, k > 0$ inteiro, $\beta > 0$ deriva-se a distribuição Erlang de ordem k .

2.2 Moda

Uma moda, denotada por y_0 , da distribuição Gama satisfaz a equação $\frac{d \log[f(y; \theta_0)]}{dy} = 0$ ou, equivalentemente,

$$\frac{\alpha - 1}{y} - \beta y = 0.$$

Resolvendo esta equação, obtém-se que o ponto

$$y_0 = \frac{\alpha - 1}{\beta}$$

representa um ponto crítico da função $\log[f(y; \boldsymbol{\theta}_0)]$ sempre que $\alpha > 1$. Uma vez que $\frac{d^2 \log[f(y; \boldsymbol{\theta}_0)]}{dy^2} < 0$, para todo $y > 0$, e a função logaritmo é estritamente monótona, conclui-se que y_0 é um ponto de máximo da função $f(y; \boldsymbol{\theta}_0)$. Assim, a moda da distribuição Gama é dada por $y_0 = (\alpha - 1)/\beta$, quando $\alpha > 1$. Quando $\alpha \leq 1$, a distribuição não possui moda, visto que apresenta caráter exponencial, como foi visto na Seção 2.1.

2.3 Momentos, Média, Variância e Função Geradora de Momentos

Seja $Y \sim \text{Gama}(\boldsymbol{\theta}_0)$ uma variável aleatória seguindo uma distribuição Gama. O ν -ésimo **momento real** de Y é dado por

$$\mathbb{E}(Y^\nu) = \frac{1}{\beta^\nu} \frac{\Gamma(\alpha + \nu)}{\Gamma(\alpha)}, \quad \nu > -\alpha.$$

A partir dessa equação, note que a **média** μ e a **variância** σ^2 de Y são dadas por

$$\mu = \frac{\alpha}{\beta}; \quad \sigma^2 = \frac{\alpha}{\beta^2}.$$

Já a **função geradora de momentos** pode ser expressa como segue

$$M_Y(t) = \mathbb{E}[\exp(tY)] = \int_0^\infty e^{tY} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta - t)^\alpha}.$$

Portanto,

$$M_Y(t) = \left(\frac{\beta}{\beta - t} \right)^\alpha.$$

2.4 Medidas de Confiabilidade

A **função de sobrevivência** da distribuição Gama é dada por

$$S(t; \boldsymbol{\theta}_0) = 1 - F(t) = 1 - \beta^\alpha \frac{\gamma(\alpha, t)}{\Gamma(\alpha)},$$

onde a integral definida

$$\gamma(\alpha, t) = \int_0^t t^{\alpha-1} e^{-t} dt$$

é conhecida como a função gama incompleta inferior.

Salienta-se que não existe fórmula fechada para a função de sobrevivência, o que também acontece no caso das outras medidas de sobrevivência, pois todas elas são expressas em função da integral definida acima, a qual não tem uma expressão fechada.

Já a fórmula da **função de risco**, é dada por

$$h(t; \theta_0) = \frac{f(t; \theta_0)}{S(t; \theta_0)} = \frac{\exp(-\beta t) t^{\alpha-1}}{\gamma(\alpha, t)}.$$

O comportamento da função de risco é alterado conforme o valor de α , como segue (veja Figura 2.2)

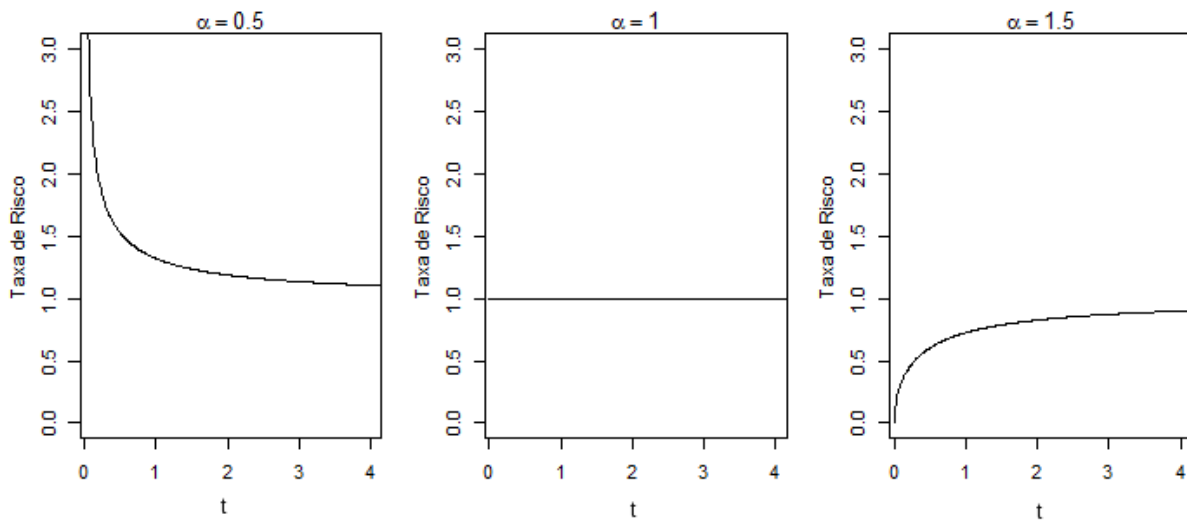


Figura 2.2: Função de risco da distribuição Gama para diferentes valores de α .

- Se $\alpha > 1$, a função cresce monotonicamente de 0 até determinado ponto, onde se estabiliza;
- Se $\alpha = 1$, a função é constante;
- Se $\alpha < 1$, a função decresce monotonicamente de ∞ para um ponto assintótico.

2.5 Entropia de Shannon

A **entropia de Shannon** associada a uma variável aleatória $Y \sim \text{Gama}(\boldsymbol{\theta}_0)$ é dada por

$$H_1(Y) = -\mathbb{E}\{\log[f(Y; \boldsymbol{\theta}_0)]\} = \log[\Gamma(\alpha)] - \alpha \log(\beta) + \alpha - (\alpha - 1)\mathbb{E}[\log(Y)].$$

No que segue será encontrada uma expressão para $\mathbb{E}[\log(Y)]$. De fato, por definição de esperança tem-se que

$$\mathbb{E}[\log(Y)] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \log(y) y^{\alpha-1} e^{-\beta y} dy. \quad (2.2)$$

Utilizando a identidade

$$\frac{dy^{\alpha-1}}{d\alpha} = \log(y) y^{\alpha-1}$$

em (2.2), segue que

$$\mathbb{E}[\log(Y)] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{dy^{\alpha-1}}{d\alpha} e^{-\beta y} dy.$$

Desde que $y^{\alpha-1}$ e $\frac{dy^{\alpha-1}}{d\alpha}$ sejam funções contínuas em α e y , a regra da integral de Leibniz nos permite trocar a integral com a derivada. Logo, a esperança $\mathbb{E}[\log(Y)]$ pode ser reescrita como segue

$$\begin{aligned} \mathbb{E}[\log(Y)] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{d}{d\alpha} \left[\int_0^\infty y^{\alpha-1} e^{-\beta y} dy \right] \\ &= \Psi^{(0)}(\alpha) - \log(\beta), \end{aligned}$$

onde $\Psi^{(0)}(\alpha) = \frac{1}{\Gamma(\alpha)} \frac{d\Gamma(\alpha)}{d\alpha}$ é conhecida como a função poligama de ordem 0. Em consequência, conclui-se que a fórmula da entropia associada a Y é expressada como

$$H_1(Y) = \alpha - \log \beta + \log \Gamma(\alpha) + (1 - \alpha)\Psi^{(0)}(\alpha).$$

2.6 Estimadores de Máxima Verossimilhança e Matriz de Informação

Seja $Y \sim \text{Gama}(\boldsymbol{\theta}_0)$ uma variável aleatória seguindo uma distribuição Gama com vetor de parâmetros $\boldsymbol{\theta}_0 = (\alpha, \beta)$. Se (Y_1, \dots, Y_n) é uma amostra aleatória, de tamanho n , tomada de Y , com valores observados denotados por $\mathbf{y} = (y_1, \dots, y_n)$, então a função de

log-verossimilhança para o vetor de parâmetros $\boldsymbol{\theta}_0$ é dada por

$$l(\boldsymbol{\theta}_0; \mathbf{y}) = n\alpha \log(\beta) - n\beta\bar{y} + (\alpha - 1) \sum_{i=1}^n \log(y_i) - n \log[\Gamma(\alpha)], \quad (2.3)$$

onde $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ denota a média dos elementos do valor observado.

Um cálculo simples mostra que as derivadas parciais, com respeito aos parâmetros, de primeira ordem da função log-verossimilhança são expressas como

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}_0; \mathbf{y})}{\partial \alpha} &= n \log(\beta) + \sum_{i=1}^n \log y_i - n\Psi^{(0)}; \\ \frac{\partial l(\boldsymbol{\theta}_0; \mathbf{y})}{\partial \beta} &= \frac{n\alpha}{\beta} - n\bar{y}. \end{aligned}$$

Já as derivadas parciais de segundo ordem são dadas por

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}_0; \mathbf{y})}{\partial \alpha^2} &= -n\Psi^{(1)}; \\ \frac{\partial^2 l(\boldsymbol{\theta}_0; \mathbf{y})}{\partial \beta^2} &= -\frac{n\alpha}{\beta^2}; \\ \frac{\partial^2 l(\boldsymbol{\theta}_0; \mathbf{y})}{\partial \beta \partial \alpha} &= \frac{\partial^2 l(\boldsymbol{\theta}_0; \mathbf{y})}{\partial \alpha \partial \beta} = \frac{n}{\beta}; \end{aligned}$$

onde $\Psi^{(1)}(\alpha) = \frac{d^2 \log[\Gamma(\alpha)]}{d\alpha^2}$ denota a função poligama de ordem 1.

Logo, a **matriz de informação de Fisher** é

$$I_Y(\boldsymbol{\theta}_0) = - \begin{bmatrix} -n\Psi^{(1)} & \frac{n}{\beta} \\ \frac{n}{\beta} & -\frac{n\alpha}{\beta^2} \end{bmatrix}.$$

Os estimadores de máxima verossimilhança, por sua vez, podem ser obtidos resolvendo-se as seguintes equações com relação a α e β

$$\frac{\partial l(\hat{\boldsymbol{\theta}}; \mathbf{y})}{\partial \alpha} = 0; \quad \frac{\partial l(\hat{\boldsymbol{\theta}}; \mathbf{y})}{\partial \beta} = 0;$$

respectivamente. Como estas equações não constituem um sistema linear, métodos numéricos têm que ser empregados para encontrar as respectivas soluções.

A Distribuição Gama Bimodal: BGama($\boldsymbol{\theta}_\delta$)

Neste capítulo, baseados na referência Vila, Ferreira, Saulo, Prativiera e Ortega, (2020) [30], serão estabelecidas algumas propriedades matemáticas básicas da distribuição Gama bimodal (BGama) e suas respectivas demonstrações. Primeiramente será apresentada a função densidade de probabilidade da distribuição BGama e em seguida foram caracterizadas a bimodalidade e unimodalidade, os momentos reais, as medidas de confiabilidade e a entropia da distribuição.

3.1 A Função de Densidade Gama bimodal

A distribuição Gama bimodal é fruto da multiplicação de uma função de densidade Gama (clássica) pelo fator quadrático $1 + (1 - \delta x)^2$ proposto por Elal-Olivero (2010) [8]. Logo, a densidade da Gama bimodal seria proporcional à seguinte expressão

$$[1 + (1 - \delta x)^2] f(x; \boldsymbol{\theta}_0),$$

onde $f(\cdot; \boldsymbol{\theta}_0)$ denota a função densidade de probabilidade (FDP) da distribuição Gama apresentada na Seção 2 e δ é um parâmetro real. Formalmente, definimos a distribuição Gama bimodal como segue.

Definição 3.1.1 (O modelo da Gama bimodal). *Dizemos que uma variável aleatória positiva X possui distribuição Gama bimodal (BGama) com vetor de parâmetros $\boldsymbol{\theta}_\delta = (\alpha, \beta, \delta)$, $\alpha > 0, \beta > 0$ e $\delta \in \mathbb{R}$, se sua FDP é dada por*

$$f(x; \boldsymbol{\theta}_\delta) = \begin{cases} \frac{1}{Z(\boldsymbol{\theta}_\delta)} [1 + (1 - \delta x)^2] \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (3.1)$$

onde α é parâmetro de forma, β é o parâmetro de escala, δ é o parâmetro responsável

pela bimodalidade da distribuição, $\Gamma(\alpha)$ denota a função Gama e a função

$$Z(\boldsymbol{\theta}_\delta) = 2 + \frac{\alpha\delta}{\beta} \left[(1 + \alpha) \frac{\delta}{\beta} - 2 \right]$$

é conhecida a constante de normalização da distribuição.

Salienta-se que, no caso particular $\delta = 0$, a distribuição Gama bimodal reduz-se à conhecida distribuição Gama (clássica) com parâmetros $\boldsymbol{\theta}_0 = (\alpha, \beta, 0) = (\alpha, \beta)$, ou seja, $\text{BGama}(\boldsymbol{\theta}_0)$. Na Figura 3.1 pode-se observar como a distribuição proposta (3.1) se comporta com algumas combinações de valores dos parâmetros distribucionais.

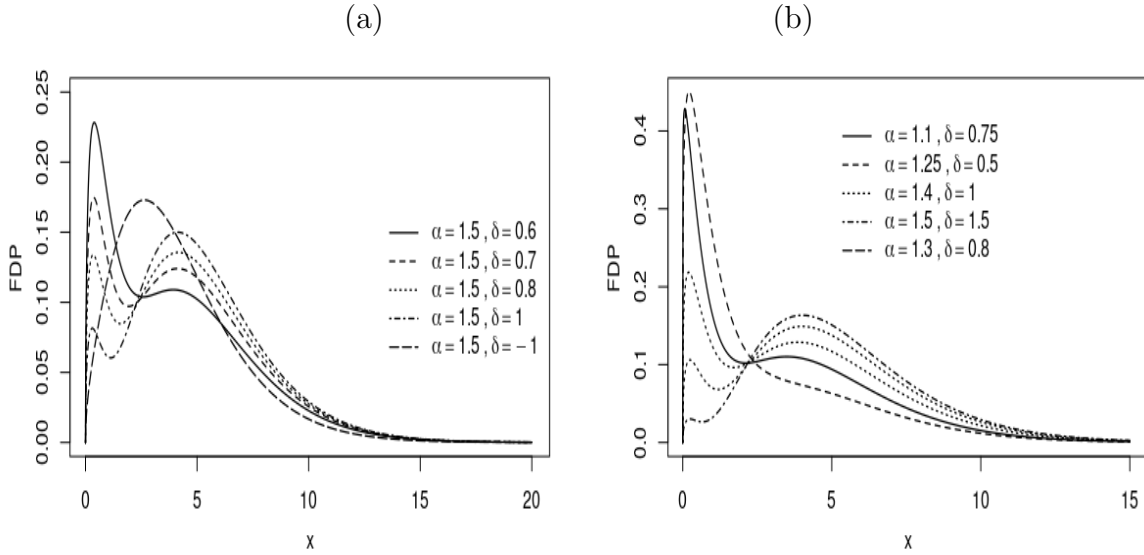


Figura 3.1: A FDP da distribuição BGama para alguns valores dos parâmetros, com $\beta = 0,7$.

Uma vez que a função $f(\cdot; \boldsymbol{\theta}_\delta)$ é proporcional ao produto do fator quadrático e a densidade Gama $f(\cdot; \boldsymbol{\theta}_0)$, ela é não negativa. Além do mais, a integral de $f(x; \boldsymbol{\theta}_\delta)$ sobre seu domínio $(0, +\infty)$ é igual a 1. De fato, considere a variável aleatória $Y \sim \text{BGama}(\boldsymbol{\theta}_0)$ com distribuição Gama, então

$$\begin{aligned} \int_{-\infty}^{\infty} f(x; \boldsymbol{\theta}_\delta) dx &= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \int_0^{\infty} [1 + (1 - \delta x)^2] \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \{1 + \mathbb{E}[(1 - \delta Y)^2]\} \\ &= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \left[1 + \frac{\alpha\delta^2}{\beta^2} + \left(1 - \frac{\alpha\delta}{\beta}\right)^2 \right]. \end{aligned} \quad (3.2)$$

Visto que

$$\begin{aligned} Z(\boldsymbol{\theta}_\delta) &= 2 + \frac{\alpha\delta}{\beta} \left[(1 + \alpha) \frac{\delta}{\beta} - 2 \right] \\ &= 1 + \frac{\alpha\delta^2}{\beta^2} + \left(1 - \frac{\alpha\delta}{\beta} \right)^2, \end{aligned}$$

a afirmação segue.

Em outras palavras, foi verificado que a função $f(\cdot; \boldsymbol{\theta}_\delta)$, definida em (3.1), é uma FDP.

Proposição 3.1.1 (Monotonicidade da densidade bimodal Gama). *A FDP do modelo da Gama bimodal (3.1) é decrescente sempre que $\alpha \leq 1$, $\delta > 0$ e $x < \frac{1}{\delta}$.*

Demonstração. Se $\delta > 0$ e $x < \frac{1}{\delta}$, a função quadrática que multiplica a densidade Gama (clássica) para gerar a Gama bimodal, $g(x) = 1 + (1 - \delta x)^2$ é decrescente e não-negativa pois sua derivada $g'(x) = -2(1 - \delta x)$ é menor que 0. Também foi visto na Seção 2.1 que quando $\alpha < 1$, a densidade da distribuição Gama (clássica) é decrescente. Assim, a função $f(x; \boldsymbol{\theta}_\delta)$ é decomposta como o produto de duas funções não-negativas decrescentes, sendo portanto decrescente. \square

3.2 Caracterização de Bimodalidade e Unimodalidade

Esta seção é direcionada a caracterizar rigorosamente o comportamento das modas da distribuição BGama. Mais precisamente, nesta seção, serão estabelecidas condições suficientes, em função dos parâmetros, para garantir a existência de unimodalidade e bimodalidade na distribuição BGama.

Proposição 3.2.1 (Modas). *O ponto x é uma moda da FDP Gama bimodal, se e somente se for solução da seguinte equação polinomial cúbica*

$$[\beta\delta^2x^2 - 2\delta(\delta + \beta)x + 2(\delta + \beta)]x - [1 + (1 - \delta x)^2](\alpha - 1) = 0,$$

ou equivalentemente,

$$\beta\delta^2x^3 - \delta[2(\delta + \beta) + \delta(\alpha - 1)]x^2 + 2[\delta + \beta + 2(\alpha - 1)]x - 2(\alpha - 1) = 0.$$

Demonstração. Sabe-se, pela Definição 1.1.2, que as possíveis modas de uma distribuição são os pontos de máximo da função de densidade associada, ou seja, estes pontos são os zeros da função derivada da FDP. Uma vez que a função logaritmo é estritamente monótona, para encontrar a equação cúbica acima, é suficiente derivar o logaritmo da densidade BGama com respeito à variável independente x e em seguida igualar a zero, de modo que as modas são soluções dessas equações. \square

Teorema 3.2.1 (Unimodalidade). *A FDP da distribuição BGama é unimodal nos seguintes casos:*

- 1) para $\delta = 0$ e $\alpha > 1$;
- 2) para $\delta \geq \beta$ e $\alpha = 1$.

Demonstração. 1) Uma vez que no caso $\delta = 0$, a distribuição Gama bimodal reduz-se à distribuição Gama (clássica) e esta distribuição é unimodal quando $\alpha > 1$ (veja Seção 2.2), a prova do primeiro item segue.

2) Se $\alpha = 1$ e x é uma moda da distribuição, então, pela Proposição 3.2.1, x é uma solução da seguinte equação quadrática

$$p_2(x) = \beta\delta^2x^2 - 2\delta(\delta + \beta)x + 2(\delta + \beta) = 0,$$

cujos discriminante é dado por

$$\Delta_2 = 4\delta^2(\delta + \beta)(\delta - \beta).$$

No que segue serão destacados dois casos: quando $\delta = \beta$ e quando $\delta > \beta$.

Se $\delta = \beta$, então $\Delta_2 = 0$, o que implica na existência de uma única raiz, $x_0 = 2/\beta$. Como $f(x; \boldsymbol{\theta}_\delta) \rightarrow 1$ na medida em que $x \rightarrow 0^+$ e $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ quando $x \rightarrow +\infty$, tem-se que a função densidade $f(\cdot; \boldsymbol{\theta}_\delta)$ é uma função crescente no intervalo $(0, x_0)$ e é decrescente no intervalo (x_0, ∞) . Isto é, o valor x_0 é o único ponto de máximo da distribuição Gama bimodal.

Já quando $\delta > \beta$ tem-se que $\Delta_2 > 0$. Assim, a equação $p_2(x) = 0$ possui dois zeros racionais e distintos,

$$x_1 = \frac{1}{\beta\delta} (\delta + \beta - \sqrt{\delta^2 - \beta^2}) > 0; \quad x_2 = \frac{1}{\beta\delta} (\delta + \beta + \sqrt{\delta^2 - \beta^2}) > 0;$$

onde $x_1 < x_2$. Como neste caso, $f(x; \boldsymbol{\theta}_\delta) \rightarrow \beta^3/[\beta^2 + \delta(\delta - \beta)]$ quando $x \rightarrow 0^+$, e $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ quando $x \rightarrow \infty$, a densidade $f(\cdot; \boldsymbol{\theta}_\delta)$ decresce no intervalo $(0, x_1)$, cresce no intervalo (x_1, x_2) e decresce novamente no intervalo de (x_2, ∞) . Depreende-se então que x_1 e x_2 são, respectivamente, pontos de mínimo e máximo. Logo a prova da afirmação no segundo item segue. \square

Para estabelecer as próximas propriedades é preciso definir as seguintes equações

$$a_{\delta,\beta} = \delta(4 + \delta)(\delta + \beta) + \beta(3\delta - 4)(3\delta + 4); \quad (3.3)$$

$$b_{\delta,\beta} = 16(1 + \delta)(\delta + \beta)^2 + (\delta^2 + 18\beta\delta(4 + \delta) - 96\beta)(\delta + \beta) - 2\delta^2; \quad (3.4)$$

$$c_{\delta,\beta} = 4(4 + \delta)(\delta + \beta)^3 + 12\beta(3\delta - 4)(\delta + \beta)^2 - 4\delta(\delta + \beta) - 27\beta. \quad (3.5)$$

Teorema 3.2.2 (Bimodalidade e Unimodalidade). *A FDP da distribuição Gama bimodal, com $\alpha > 1$, possui as seguintes configurações*

- 1) *É bimodal se $\delta > \beta$, $a_{\delta,\beta} > 0$, $b_{\delta,\beta} > 0$ e $c_{\delta,\beta} > 0$;*
- 2) *É unimodal se $0 < \delta < \beta$, $a_{\delta,\beta} < 0$, $b_{\delta,\beta} < 0$ e $c_{\delta,\beta} < 0$;*
- 3) *É bimodal se $\delta = \beta > \frac{8\sqrt{3}}{11} - \frac{4}{11}$;*
- 4) *É unimodal se $0 < \delta = \beta < \frac{\sqrt{1745}}{12} - \frac{35}{12}$;*

onde $a_{\delta,\beta}$, $b_{\delta,\beta}$ e $c_{\delta,\beta}$ são os mesmos das equações (3.3), (3.4) e (3.5), respectivamente.

Demonstração. Se x é uma moda da densidade BGama (3.1), pela Proposição 3.2.1, o ponto x deve ser a solução da seguinte equação cúbica

$$p_3(x) = \beta\delta^2x^3 - \delta[2(\delta + \beta) + \delta(\alpha - 1)]x^2 + 2[\delta + \beta + 2(\alpha - 1)]x - 2(\alpha - 1) = 0.$$

Pela regra dos signos de Descartes (ver Griffiths, 1947 [15] ou Xue, 2012 [34]) essa equação possui uma ou três raízes positivas. Essa regra afirma que quando o polinômio é ordenado de forma decrescente pelo grau de seus termos, o número de raízes positivas é no máximo o número de permutações de sinal, ou menor por uma diferença par. Sabe-se, também, que o discriminante de uma equação polinomial cúbica $ax^3 + bx^2 + cx + d$ é dado por

$$\Delta_3 = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd.$$

Neste caso, em específico, é dado por

$$\begin{aligned} \Delta_3 = \Delta_3(\alpha) = & 16\delta^4(\alpha - 1)^4 + 16\delta^2a_{\beta,\delta}(\alpha - 1)^3 \\ & + 4\delta^2b_{\beta,\delta}(\alpha - 1)^2 + 4\delta^2c_{\beta,\delta}(\alpha - 1) + 16\delta^2(\delta - \beta)(\delta + \beta)^3. \end{aligned}$$

Se $\delta > \beta$, $a_{\delta,\beta}$, $b_{\delta,\beta}$, $c_{\delta,\beta}$ são positivos e $\alpha > 1$ então $\Delta_3(\alpha) > 0$. Conseqüentemente a equação $p_3(x) = 0$ possui três raízes positivas diferentes, denotadas por x_1, x_2 e x_3 . Sem perda de generalidade, assuma que $x_1 < x_2 < x_3$. Como $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ na medida em que $x \rightarrow 0^+$ e $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ quando $x \rightarrow \infty$, tem-se que a densidade BGama cresce nos intervalos $(0, x_1)$ e (x_2, x_3) , e decresce nos intervalos (x_1, x_2) e (x_3, ∞) . Isto é, x_1 e x_3 são dois pontos de máximo e x_2 é um ponto de mínimo da função $f(x; \boldsymbol{\theta}_\delta)$, de onde a propriedade de bimodalidade do primeiro item segue.

Se $0 < \delta < \beta$ e $a_{\delta,\beta}$, $b_{\delta,\beta}$, $c_{\delta,\beta}$ são negativos, então $\Delta_3(\alpha) < 0$, quando $\alpha > 1$. Portanto, a equação polinomial $p_3(x) = 0$ possui uma única raiz positiva, denotada por x_0 , e duas raízes imaginárias conjugadas complexas. Como $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ quando $x \rightarrow 0^+$ e $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ quando $x \rightarrow +\infty$, o valor x_0 tem que ser um ponto de máximo de $f(x; \boldsymbol{\theta}_\delta)$. Isto prova a afirmação de unimodalidade do segundo item.

No que segue serão verificadas as afirmações dos Itens 3) e 4). Para isso, observe que, quando $\delta = \beta$, os parâmetros $a_{\delta,\beta}$, $b_{\delta,\beta}$ e $c_{\delta,\beta}$ podem ser reescritos como

$$\begin{aligned} a_{\delta,\beta} &= \beta(11\beta^2 + 8\beta - 16); \\ b_{\delta,\beta} &= \beta^2(18\beta^2 + 105\beta - 65); \\ c_{\delta,\beta} &= \beta(32\beta^3 + 272\beta^2 - 200\beta - 27). \end{aligned}$$

Com o intuito de verificar o Item 3), assumamos que $\delta = \beta > \frac{8\sqrt{3}}{11} - \frac{4}{11}$. Neste caso, note que os valores de $a_{\delta,\beta}$, $b_{\delta,\beta}$ e $c_{\delta,\beta}$ são positivos, o que implica na positividade do discriminante $\Delta_3(\alpha)$. Então, como na prova do Item 1), pode-se comprovar que a equação $p_3(x) = 0$ possui 3 raízes diferentes, denotadas por x_1, x_2, x_3 , sendo x_1 e x_3 pontos de máximo e x_2 ponto de mínimo da função $f(x; \theta_\delta)$, verificando-se a bimodalidade enunciada no terceiro item.

Finalmente, provaremos o Item 4). A partir das equações acima envolvendo $a_{\delta,\beta}$, $b_{\delta,\beta}$ e $c_{\delta,\beta}$, note que, quando $0 < \delta = \beta < \frac{\sqrt{1745}}{12} - \frac{35}{12}$, os valores de $a_{\delta,\beta}$, $b_{\delta,\beta}$ e $c_{\delta,\beta}$ são negativos, e portanto $\Delta_3(\alpha) < 0$. Logo, como na prova do Item 2), a equação polinomial $p_3(x) = 0$ possui uma única raiz positiva, denotada por x_0 , e duas raízes complexas conjugadas. Portanto, x_0 é o único ponto de máximo da função $f(x; \theta_\delta)$. Isto garante a afirmação de unimodalidade do quarto item. \square

Observação 3.2.1. *Na prova do Item 1) do Teorema 3.2.2, outra forma de verificar que o polinômio $p_3(x)$ possui três raízes positivas é através do uso da fórmula de Vieta (Vinberg, 2003 [31]). Esta fórmula afirma que em um polinômio de grau n , $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$ com raízes x_1, \dots, x_n , as seguintes relações são válidas:*

$$\begin{aligned} x_1 + x_2 + \dots + x_{n-1} + x_n &= -\frac{a_{n-1}}{a_n}; \\ (x_1 x_2 + x_1 x_3 + \dots + x_1 x_n) + (x_2 x_3 + x_2 x_4 + \dots + x_2 x_n) + \dots + x_{n-1} x_n &= -\frac{a_{n-2}}{a_n}; \\ x_1 x_2 \dots x_{n-1} x_n &= (-1)^n \frac{a_0}{a_n}. \end{aligned}$$

Aplicando estas relações ao polinômio $p_3(x)$, tem-se

$$\begin{aligned} x_1 + x_2 + x_3 &= \frac{2(\delta + \beta) + \delta(\alpha - 1)}{\beta\delta}; \\ x_1 x_2 + x_1 x_3 + x_2 x_3 &= \frac{2[\delta + \beta + 2(\alpha - 1)]}{\beta\delta^2}; \\ x_1 x_2 x_3 &= \frac{2(\alpha - 1)}{\beta\delta^2}. \end{aligned}$$

No Item 1), como $\alpha > 1$ e $0 < \beta < \delta$, todas as raízes têm que ser positivas e diferentes de zero.

3.3 Momentos reais, Variância e Função Geradora de Momentos

Assim como a distribuição Gama, a distribuição Gama bimodal apresenta fórmulas fechadas para os momentos, e também, para a função geradora de momentos. Além disso, sobre determinada restrição nos parâmetros, todos os momentos desta distribuição são finitos.

Sendo assim, os resultados desta seção destinam-se a apresentar as fórmulas de medidas relacionadas aos momentos da BGama.

Proposição 3.3.1 (Momentos reais). *Se $X \sim BGama(\boldsymbol{\theta}_\delta)$, para cada número real fixado ν tal que $\nu > -\alpha$, tem-se*

$$\mathbb{E}(X^\nu) = \frac{1}{\beta^\nu Z(\boldsymbol{\theta}_\delta)} \left[2 - \frac{2\delta}{\beta}(\nu + \alpha) + \frac{\delta^2}{\beta^2}(\nu + \alpha)(\nu + \alpha + 1) \right] \frac{\Gamma(\nu + \alpha)}{\Gamma(\alpha)}. \quad (3.6)$$

Demonstração. Se $X \sim BGama(\boldsymbol{\theta}_\delta)$, pela definição de momentos reais de uma distribuição, tem-se

$$\begin{aligned} \mathbb{E}(X^\nu) &= \int_{-\infty}^{\infty} x^\nu f(x; \boldsymbol{\theta}_\delta) dx \\ &= \int_0^{\infty} x^\nu \frac{1 + (1 - \delta x)^2}{Z(\boldsymbol{\theta}_\delta)} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) dx. \end{aligned}$$

Desenvolvendo o quadrado da soma de dois termos e aplicando a linearidade da integral, a expressão acima é

$$\begin{aligned} &= \frac{2}{Z(\boldsymbol{\theta}_\delta)} \int_0^{\infty} x^\nu \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) dx \\ &\quad - \frac{2\delta}{Z(\boldsymbol{\theta}_\delta)} \int_0^{\infty} x^{\nu+1} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) dx \\ &\quad + \frac{\delta^2}{Z(\boldsymbol{\theta}_\delta)} \int_0^{\infty} x^{\nu+2} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) dx. \end{aligned}$$

Considere uma variável aleatória $Y \sim BGama(\boldsymbol{\theta}_0)$ com distribuição Gama (clássica). Uma simples observação mostra que a última expressão pode ser escrita da seguinte forma

$$= \frac{1}{Z(\boldsymbol{\theta}_\delta)} [2 \mathbb{E}(Y^\nu) - 2\delta \mathbb{E}(Y^{\nu+1}) + \delta^2 \mathbb{E}(Y^{\nu+2})].$$

Portanto, o ν -ésimo momento real $\mathbb{E}(X^\nu)$ pode ser expressado em função dos ν -ésimo momentos reais $\mathbb{E}(Y^\nu)$ de uma distribuição Gama como segue

$$\mathbb{E}(X^\nu) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} [2\mathbb{E}(Y^\nu) - 2\delta\mathbb{E}(Y^{\nu+1}) + \delta^2\mathbb{E}(Y^{\nu+2})].$$

Mas como $\mathbb{E}(Y^\nu) = \frac{\Gamma(\nu+\alpha)}{\beta^\nu\Gamma(\alpha)}$, $\nu > -\alpha$, (veja Seção 2.3), tem-se que

$$\mathbb{E}(X^\nu) = \frac{1}{\beta^\nu\Gamma(\alpha)Z(\boldsymbol{\theta}_\delta)} \left[2\Gamma(\nu+\alpha) - \frac{2\delta}{\beta}\Gamma(\nu+1+\alpha) + \frac{\delta^2}{\beta^2}\Gamma(\nu+2+\alpha) \right]. \quad (3.7)$$

Finalmente, ao aplicar a propriedade da função Gama, $\Gamma(x+1) = x\Gamma(x)$ (veja Seção 2), no Item (3.7), a expressão (3.6) para $\mathbb{E}(X^\nu)$ segue. \square

Pela Proposição 3.3.1 tem-se todos os momentos da distribuição Gama bimodal, em particular os dois primeiros momentos existem, e em consequência, através de simples manipulações algébricas, obtém-se o seguinte resultado.

Corolário 3.3.1 (Média e Variância). *Se $X \sim BGama(\boldsymbol{\theta}_\delta)$ com $\mu = \mathbb{E}(X)$ e $\sigma^2 = \text{Var}(X)$, então*

$$\mu = \frac{\alpha(\alpha^2\delta^2 - 2\alpha\beta\delta + 3\alpha\delta^2 + 2\beta^2 - 2\beta\delta + 2\delta^2)}{\beta(\alpha^2\delta^2 - 2\alpha\beta\delta + \alpha\delta^2 + 2\beta^2)};$$

$$\sigma^2 = \frac{\alpha[-8(\alpha+1)\beta^3\delta + 4(\alpha+1)(2\alpha+3)\beta^2\delta^2 - 4\alpha(\alpha+1)(\alpha+2)\beta\delta^3 + \alpha(\alpha+1)^2(\alpha+2)\delta^4 + 4\beta^4]}{\beta^2[-2\alpha\beta\delta + \alpha(\alpha+1)\delta^2 + 2\beta^2]^2}.$$

O seguinte resultado fornece expressões fechadas para os n -ésimos momentos padrão da distribuição Gama bimodal. No caso particular que $n = 3$ ou $n = 4$ obtém-se a expressão fechada para a assimetria e a curtose, respectivamente.

Proposição 3.3.2 (Momentos Padrão). *Se $X \sim BGama(\boldsymbol{\theta}_\delta)$, para cada número natural fixado n , tem-se*

$$\mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^n \right] = C \sum_{k=0}^n \binom{n}{k} (-\mu)^{n-k} \left[2 - 2\frac{\delta}{\beta}(k+\alpha) + \frac{\delta^2}{\beta^2}(k+\alpha+1) \right] \prod_{i=0}^{k-1} (\alpha+i),$$

onde $C = [\sigma^n \beta^\nu Z(\boldsymbol{\theta}_\delta)]^{-1}$ e os parâmetros μ e σ são fornecidos no Corolário 3.3.1. Além do mais, na expressão acima assume-se que $\prod_{i=0}^{-1} (\alpha+i) = 1$.

Demonstração. Para encontrar os momentos padrões a partir de σ , μ e da existência dos outros momentos (veja Proposição 3.3.1), será utilizada a seguinte expansão binomial

$$(p+q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}. \quad (3.8)$$

Tomando $p = -\mu/\sigma$ e $q = X/\sigma$ na identidade acima, e em seguida tomando a esperança em ambos os lados da equação resultante, pela propriedade da linearidade do operador esperança, segue que

$$\begin{aligned} \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^n \right] &= \frac{1}{\sigma^n} \sum_{k=0}^n \binom{n}{k} (-\mu)^{n-k} \mathbb{E}(X^k) \\ &= \frac{1}{\sigma^n \beta^\nu Z(\boldsymbol{\theta}_\delta)} \sum_{k=0}^n \binom{n}{k} (-\mu)^{n-k} \left[2 - 2\frac{\delta}{\beta}(k + \alpha) + \frac{\delta^2}{\beta^2}(k + \alpha + 1) \right] \frac{\Gamma(k + \alpha)}{\Gamma(\alpha)}, \end{aligned}$$

onde na segunda igualdade é utilizada a Proposição 3.3.1. Em seguida, utilizando a identidade

$$\Gamma(\alpha + k) = \prod_{i=0}^{k-1} (\alpha + i)$$

e assumindo que $\prod_{i=0}^{-1} (\alpha + i) = 1$, a prova da proposição segue imediatamente. \square

Proposição 3.3.3. *Se $X \sim BGama(\boldsymbol{\theta}_\delta)$, para cada número natural n fixado tem-se que*

$$\begin{aligned} 1) \quad &\mathbb{E}[\log(X^n)] = n\delta\tilde{C}[(2\alpha + 1)\delta - 2\beta] + n\tilde{C}[2\beta^2 - 2n\alpha\beta\delta + \alpha(\alpha + 1)\delta^2] [\Psi^{(0)}(\alpha) - \log(\beta)]; \\ 2) \quad &\mathbb{E}\{[\log(X)]^n\} \\ &= \hat{C} \left[\frac{n(n-1)\delta^2}{\beta} - 2n\delta + \alpha(\alpha + 1)\delta^2 \right] \sum_{k=0}^{n-2} \binom{n-2}{k} (-1)^{n-2-k} [\log(\beta)]^{n-2-k} \Psi^{(k)}(\alpha) \\ &+ \hat{C} \{ 2\beta + [\alpha(\alpha + 1)\delta^2] \} [\Psi^{(n)}(\alpha) - (n-1)\log(\beta) \Psi^{(n-1)}(\alpha)] - \hat{C}n\delta \left[2 + \frac{(2\alpha + 1)\delta}{\beta} \right] \Psi^{(n-1)}(\alpha); \end{aligned}$$

onde $\tilde{C} = [\beta^2 Z(\boldsymbol{\theta}_\delta)]^{-1}$, $\hat{C} = [\beta\Gamma(\alpha)Z(\boldsymbol{\theta}_\delta)]^{-1}$ e $\Psi^{(m)}(z)$ é a função poligama de ordem m definida por $\Psi^{(m)}(z) = \frac{d^{m+1} \log[\Gamma(z)]}{dz^{m+1}}$.

Demonstração. Seja $Y \sim BGama(\boldsymbol{\theta}_0)$ uma variável aleatória com distribuição Gama clássica. Utilizando integração por partes tem-se

$$\begin{aligned} \mathbb{E}[Y \log(Y^n)] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^\alpha \log(y^n) \exp(-\beta y) dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \left\{ -y^\alpha \log(y^n) \frac{\exp(-\beta y)}{\beta} \Big|_0^\infty - \int_0^\infty \frac{\exp(-\beta y)}{-\beta} \left[\alpha y^{\alpha-1} \log(y^n) + y^\alpha \frac{n}{y} \right] dy \right\} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\exp(-\beta y)}{\beta} \left[\alpha y^{\alpha-1} \log(y^n) + y^\alpha \frac{n}{y} \right] dy, \end{aligned}$$

onde na última igualdade são utilizados os seguintes limites

$$\lim_{y \rightarrow +\infty} y^\alpha \log(y^n) \exp(-\beta y) = 0 \quad \text{e} \quad \lim_{y \rightarrow 0} y^\alpha \log(y^n) \exp(-\beta y) = 0.$$

Uma observação simples mostra que a última integral acima é

$$= \frac{n}{\beta} + \frac{\alpha}{\beta} \mathbb{E}[\log(Y^n)].$$

Portanto, obtém-se a seguinte igualdade

$$\mathbb{E}[Y \log(Y^n)] = \frac{n}{\beta} + \frac{\alpha}{\beta} \mathbb{E}[\log(Y^n)]. \quad (3.9)$$

Analogamente ao feito acima, utilizando duas vezes integração por partes, verifica-se que

$$\mathbb{E}[Y^2 \log(Y^n)] = \frac{n(\alpha + 1)}{\beta^2} + \frac{n}{\beta} \mathbb{E}(Y) + \frac{\alpha(\alpha + 1)}{\beta^2} \mathbb{E}[\log(Y^n)]. \quad (3.10)$$

Desde que

$$\mathbb{E}[\log(X^n)] = \frac{2 \mathbb{E}[\log(Y^n)] - 2\delta \mathbb{E}[Y \log(Y^n)] + \delta^2 \mathbb{E}[Y^2 \log(Y^n)]}{Z(\boldsymbol{\theta}_\delta)} \quad (3.11)$$

por combinar a seguinte identidade

$$\mathbb{E}[\log(Y^n)] = n[\Psi^{(0)}(\alpha) - \log(\beta)],$$

com as identidades nos Itens (3.9), (3.10) e (3.11), pela Proposição 3.3.1, a prova do primeiro item segue.

No que resta da prova, será verificado o Item 2). Desde que $Y \sim \text{BGama}(\boldsymbol{\theta}_0)$, utilizando integração por partes, tem-se que

$$\begin{aligned} \mathbb{E}\{Y[\log(Y)]^n\} &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^\alpha [\log(y)]^n \exp(-\beta y) dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \left\{ -y^\alpha [\log(y)]^n \frac{\exp(-\beta y)}{\beta} \Big|_0^\infty + \int_0^\infty \frac{\exp(-\beta y)}{\beta} \left\{ \alpha y^{\alpha-1} [\log(y)]^n + y^\alpha [\log(y)]^{n-1} \frac{n}{y} \right\} dy \right\}. \end{aligned}$$

Uma vez que $\lim_{y \rightarrow 0} y^\alpha [\log(y)]^n e^{-\beta y} = \lim_{y \rightarrow +\infty} y^\alpha [\log(y)]^n e^{-\beta y} = 0$, a integral acima é

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\exp(-\beta y)}{\beta} \left\{ \alpha y^{\alpha-1} [\log(y)]^n + y^\alpha [\log(y)]^{n-1} \frac{n}{y} \right\} dy.$$

Novamente, desde que $Y \sim \text{BGama}(\boldsymbol{\theta}_0)$, note que esta integral pode ser expressa como

$$= \frac{n}{\beta} \mathbb{E}\{[\log(Y)]^{n-1}\} + \frac{\alpha}{\beta} \mathbb{E}\{[\log(Y)]^n\}.$$

Logo, tem-se que

$$\mathbb{E}\{Y[\log(Y)]^n\} = \frac{n}{\beta} \mathbb{E}\{[\log(Y)]^{n-1}\} + \frac{\alpha}{\beta} \mathbb{E}\{[\log(Y)]^n\}. \quad (3.12)$$

Analogamente ao feito acima, utilizando duas vezes integração por partes, verifica-se que

$$\begin{aligned} \mathbb{E}\{Y^2[\log(Y)]^n\} &= \frac{n(n-1)}{\beta^2} \mathbb{E}\{[\log(Y)]^{n-2}\} \\ &+ \frac{n(2\alpha+1)}{\beta^2} \mathbb{E}\{[\log(Y)]^{n-1}\} + \frac{\alpha(\alpha+1)}{\beta^2} \mathbb{E}\{[\log(Y)]^n\}. \end{aligned} \quad (3.13)$$

Uma vez que

$$\mathbb{E}\{[\log(X)]^n\} = \frac{2\mathbb{E}\{[\log(Y)]^n\} - 2\delta\mathbb{E}\{Y[\log(Y)]^n\} + \delta^2\mathbb{E}\{Y^2[\log(Y)]^n\}}{Z(\boldsymbol{\theta}_\delta)}, \quad (3.14)$$

$$\mathbb{E}\{[\log(Y)]^n\} = \frac{1}{\Gamma(\alpha)} \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} [\log(\beta)]^{n-k} \Psi^{(k)}(\alpha); \quad (3.15)$$

e

$$\mathbb{E}\{[\log(Y)]^{n-1}\} = \frac{\Psi^{(n-1)}(\alpha)}{\Gamma(\alpha)} + \mathbb{E}\{[\log(Y)]^{n-2}\}; \quad (3.16)$$

$$\mathbb{E}\{[\log(Y)]^n\} = \frac{\Psi^{(n)}(\alpha)}{\Gamma(\alpha)} - (n-1) \frac{\log \beta \Psi^{(n-1)}(\alpha)}{\Gamma(\alpha)} + \mathbb{E}\{[\log(Y)]^{n-2}\}; \quad (3.17)$$

por combinar as identidades (3.12)-(3.17), a prova do segundo item segue. \square

Proposição 3.3.4. *Se $X \sim BGama(\boldsymbol{\theta}_\delta)$ então*

$$M_X(t) = \frac{1}{\beta^{-\alpha} Z(\boldsymbol{\theta}_\delta)} [2 + \delta^2 \alpha(\alpha+1) - 2\delta\alpha(\beta-t)] (\beta-t)^{-(\alpha+2)}, \quad t < \beta.$$

Demonstração. Seja $Y \sim BGama(\boldsymbol{\theta}_0)$. Aplicando a formula de integração por partes tem-se

$$\mathbb{E}[Y \exp(tY)] = \frac{\alpha}{\beta-t} M_Y(t); \quad \mathbb{E}[Y^2 \exp(tY)] = \frac{\alpha(\alpha+1)}{(\beta-t)^2} M_Y(t). \quad (3.18)$$

utilizando a identidade

$$\begin{aligned} &\int_0^\infty e^{tx} [1 + (1-\delta x)^2] x^{\alpha-1} \exp(-\beta x) dx \\ &= 2 \int_0^\infty x^{\alpha-1} \exp[-(\beta-t)x] dx + \delta^2 \int_0^\infty x^{\alpha+1} \exp[-(\beta-t)x] dx - 2\delta \int_0^\infty x^\alpha \exp[-(\beta-t)x] dx \end{aligned}$$

note que a função geradora de momentos $M_X(t)$ pode ser expressada da seguinte forma

$$M_X(t) = 2 M_Y(t) - 2\delta \mathbb{E}[Y \exp(tY)] + \delta^2 \mathbb{E}[Y^2 \exp(tY)]. \quad (3.19)$$

Combinando as identidades no Item (3.18) com a relação (3.19), obtém-se

$$M_X(t) = \frac{1}{Z(\boldsymbol{\theta}_\delta)(\beta - t)^2} [2 + \delta^2 \alpha(\alpha + 1) - 2\delta \alpha(\beta - t)] M_Y(t), \quad t < \beta.$$

Como $M_Y(t) = (1 - \frac{t}{\beta})^{-\alpha}$ para $t < \beta$ (ver Seção 2.3), a prova da proposição segue da identidade acima. \square

Observação 3.3.1. *A função característica de $X \sim \text{BGama}(\boldsymbol{\theta})$, definida por $\phi_X(t)$, pode ser obtida a partir da função de momentos por meio da seguinte identidade*

$$M_X(t) = \phi_X(-it).$$

O próximo resultado mostra o decaimento exponencial das caudas da distribuição BGama (3.1) com respeito ao valor da variável aleatória.

Corolário 3.3.2 (Distribuição de Caudas Leves). *Se $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$, então existe $t > 0$ tal que*

$$\mathbb{P}(X > x) \leq \exp(-tx),$$

para x grande o suficiente.

Demonstração. Uma vez que, para $t < \beta$, a função geradora de momentos $M_X(t)$ da distribuição Gama bimodal é finita (veja Proposição 3.3.4), a prova segue. \square

Observação 3.3.2. *Outra forma de calcular a taxa de decaimento da distribuição associada a uma variável aleatória X é através da seguinte quantidade*

$$\tau_X = - \lim_{x \rightarrow \infty} \frac{d \log[f_X(x)]}{dx}.$$

Aplicando esta fórmula à distribuição Gama bimodal, tem-se

$$\begin{aligned} \tau_{\text{BGama}(\boldsymbol{\theta}_\delta)} &= \lim_{x \rightarrow \infty} \left[\frac{2\delta(1 - \delta x)}{1 + (1 - \delta x)^2} - (\alpha - 1) \frac{1}{x} + \beta \right] = \beta \\ &= \tau_{\text{BGama}(\boldsymbol{\theta}_0)} = \tau_{\text{BGama}(\alpha=1, \beta, \delta=0)} = \tau_{\text{exp}(\beta)}. \end{aligned}$$

Percebe-se que a taxa de decaimento da distribuição Gama bimodal, assim como as distribuições Gama e Exponencial, depende apenas de seu parâmetro de escala β . Em outras palavras, perto das caudas a distribuição se assemelha a uma Exponencial.

Pode-se verificar também que

$$0 = \tau_{\text{InvGamma}}(\boldsymbol{\theta}_0) = \tau_{\text{LogNorm}}(\mu, \sigma^2) = \tau_{\text{GenPareto}}(\boldsymbol{\theta}_0, \xi) < \tau_{\text{BGama}}(\boldsymbol{\theta}_\delta) < \tau_{\text{Normal}}(\mu, \sigma^2) = \infty.$$

Portanto, a cauda da distribuição Normal é mais leve que a cauda da distribuição BGama, visto que a distribuição Normal decai com taxa infinita, enquanto a distribuição BGama possui taxa igual a β . Porém a distribuição estudada possui caudas mais leves que a distribuição Pareto-generalizada, Lognormal, e Gamma-inversa.

3.4 Medidas de Entropia

Como foi abordada na Seção 1.6, a entropia é uma medida fundamental que informa o quanto de incerteza uma distribuição possui. Assim, esta seção dedica-se a apresentar expressões matemáticas para a entropia Quadrática e para a entropia de Shannon associadas a uma distribuição Gama bimodal.

Proposição 3.4.1 (Entropia Quadrática). *Se $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$ e $\alpha > 1$, então*

$$\begin{aligned} H_2(X) = & \log(2) + \frac{1}{2} \log(\pi) + 2 \log[\Gamma(\alpha)] + \log[Z(\boldsymbol{\theta}_\delta)] - \log [1 + \delta^2 \sigma^2 + (1 - \delta\mu)^2] \\ & - \log(\beta) - \log(\alpha - 1) - \log[\Gamma(\alpha - 1)] - \log \left[\Gamma\left(\alpha - \frac{1}{2}\right) \right], \end{aligned} \quad (3.20)$$

onde μ e σ^2 denotam a média e variância de X , respectivamente; veja Corolário 3.3.1.

Demonstração. A entropia quadrática é definida por; veja Definição 1.6.2,

$$H_2(X) = -\log \left[\int_0^\infty f^2(x; \boldsymbol{\theta}) dx \right].$$

Se $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$, uma simples observação mostra que

$$\begin{aligned} & \int_0^\infty f^2(x; \boldsymbol{\theta}_\delta) dx \\ &= \int_0^\infty \frac{[1 + (1 - \delta x)^2]^2}{[Z(\boldsymbol{\theta}_\delta)]^2} \frac{\beta^{2\alpha}}{\Gamma(\alpha)^2} x^{2\alpha-2} e^{-2\beta x} dx \\ &= \frac{\beta \Gamma(2\alpha - 1)}{2^{2\alpha-1} Z(\boldsymbol{\theta}_\delta) \Gamma^2(\alpha)} \int_0^\infty [1 + (1 - \delta x)^2] \frac{1 + (1 - \delta x)^2}{Z(\boldsymbol{\theta}_\delta)} \frac{\beta^{2\alpha}}{\Gamma(2\alpha - 1)} x^{(2\alpha-1)-1} e^{-2\beta x} dx. \end{aligned}$$

Note que a expressão acima é

$$= \frac{\beta \Gamma(2\alpha - 1)}{2^{2\alpha-1} Z(\boldsymbol{\theta}_\delta) \Gamma^2(\alpha)} \{1 + \mathbb{E}[(1 - \delta X)^2]\},$$

onde $X \sim \text{BGama}(2\alpha - 1, 2\beta, \delta)$ e $\mathbb{E}[(1 - \delta X)^2] = \delta^2 \sigma^2 + (1 - \delta \mu)^2$. Portanto,

$$\int_0^\infty f^2(x; \boldsymbol{\theta}_\delta) dx = \frac{\beta \Gamma(2\alpha - 1)}{2^{2\alpha-1} Z(\boldsymbol{\theta}_\delta) \Gamma^2(\alpha)} \{1 + \delta^2 \sigma^2 + (1 - \delta \mu)^2\}.$$

Combinando esta última identidade com as relações

$$\Gamma(2\alpha + 1) = 2\alpha \Gamma(2\alpha); \quad \Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}} \Gamma(z) \Gamma\left(z + \frac{1}{2}\right);$$

tem-se que

$$\int_0^\infty f^2(x; \boldsymbol{\theta}_\delta) dx = \frac{\beta(\alpha - 1) \Gamma(\alpha - 1) \Gamma(\alpha - \frac{1}{2})}{2Z(\boldsymbol{\theta}_\delta) \sqrt{\pi} \Gamma^2(\alpha)} [1 + \delta^2 \sigma^2 + (1 - \delta \mu)^2].$$

Finalmente, tomando o logaritmo e em seguida multiplicando por -1 ambos os lados da equação acima, a fórmula (3.20) da entropia Quadrática segue. \square

Proposição 3.4.2 (Entropia de Shannon). *Se $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$ então a entropia de Shannon pode ser escrita como*

$$\begin{aligned} H_1(X) &= \log[Z(\boldsymbol{\theta}_\delta)] + \log[\Gamma(\alpha)] - \alpha \log(\beta) + \beta \mu - \log [1 + (1 - \delta \mu)^2 + \delta^2 \sigma^2] \\ &+ \sum_{n=1}^{\infty} \frac{1}{n} \sum_{k=0}^n \binom{n}{k} (-1)^{-k} [1 + (1 - \delta \mu)^2 + \delta^2 \sigma^2]^{-k} \mathbb{E}\{[g(X)]^k\} \\ &- (\alpha - 1) \delta \tilde{C} [(2\alpha + 1)\delta - 2\beta] + \tilde{C} [2\beta^2 - 2\alpha\beta\delta + \alpha(\alpha + 1)\delta^2] [\Psi^{(0)}(\alpha) - \log(\beta)], \end{aligned}$$

sempre que a série acima convirja. Aqui, $g(x)$ é a função definida por $g(x) = 1 + (1 - \delta x)^2$, $\tilde{C} = [\beta^2 Z(\boldsymbol{\theta}_\delta)]^{-1}$, os parâmetros μ, σ^2 denotam a média e a variância de X (veja Corolário 3.3.1), respectivamente, e $\Psi^{(m)}(z)$ é a função poligama de ordem m definida por

$$\Psi^{(m)}(z) = \frac{d^{m+1} \log[\Gamma(z)]}{dz^{m+1}}.$$

Demonstração. Seja $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$. A fórmula da entropia de Shannon, conforme a Definição 1.6.1, é dada por

$$H_1(X) = -\mathbb{E}\{\log[f(X; \boldsymbol{\theta})]\}.$$

Considere a função $g(x) = 1 + (1 - \delta x)^2$. Desenvolvendo o logaritmo $\log[f(X; \boldsymbol{\theta})]$ na identidade acima, tem-se

$$H_1(X) = \mathbb{E}\{\log[Z(\boldsymbol{\theta}_\delta)] + \log[\Gamma(\alpha)] - \alpha \log(\beta) - \log[g(X)] - (\alpha - 1) \log(X) + \beta X\}.$$

utilizando a linearidade da esperança, segue que

$$\begin{aligned} H_1(X) &= \log[Z(\boldsymbol{\theta}_\delta)] + \log[\Gamma(\alpha)] - \alpha \log(\beta) \\ &\quad - \mathbb{E}\{\log[g(X)]\} - (\alpha - 1)\mathbb{E}[\log(X)] + \beta\mu. \end{aligned} \quad (3.21)$$

Pela Proposição 3.3.3, com $n = 1$, é conhecido que

$$\begin{aligned} \mathbb{E}[\log(X)] &= \delta\tilde{C}[(2\alpha + 1)\delta - 2\beta] \\ &\quad + \tilde{C}[2\beta^2 - 2\alpha\beta\delta + \alpha(\alpha + 1)\delta^2][\Psi^{(0)}(\alpha) - \log(\beta)], \end{aligned} \quad (3.22)$$

onde $\tilde{C} = [\beta^2 Z(\boldsymbol{\theta}_\delta)]^{-1}$.

Por outro lado, para obter uma expressão para a esperança $\mathbb{E}\{\log[g(X)]\}$, considera-se a expansão de Taylor da função $\log[g(X)]$ ao redor do ponto $\mathbb{E}[g(X)] = 1 + (1 - \delta\mu)^2 + \delta^2\sigma^2$. Seja $\mathbb{E}[g(X)] = \mu_g$. Logo,

$$\log[g(X)] = \log(\mu_g) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n\mu_g^n} [g(X) - \mu_g]^n, \quad (3.23)$$

onde, segundo a fórmula de Hadamard, o raio de convergência R da série é dada por

$$R^{-1} = \limsup_{n \rightarrow \infty} \left| \frac{(-1)^{n+1}}{n\mu_g^n} \right|^{1/n} = \frac{1}{\mu_g} = \frac{1}{1 + (1 - \delta\mu)^2 + \delta^2\sigma^2}.$$

Tomando a esperança em ambos os lados de (3.23) e assumindo que é possível trocar a esperança com a série, tem-se que

$$\begin{aligned} \mathbb{E}\{\log[g(X)]\} &= \log(\mu_g) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n\mu_g^n} \mathbb{E}\{[g(X) - \mu_g]^n\} \\ &= \log(\mu_g) - \sum_{n=1}^{\infty} \frac{1}{n} \sum_{k=0}^n \binom{n}{k} (-\mu_g)^{-k} \mathbb{E}\{[g(X)]^k\}, \end{aligned} \quad (3.24)$$

onde na segunda igualdade utiliza-se a expansão binomial (3.8). Desde que $\mathbb{E}(X^\nu) < \infty$, para todo $\nu > -\alpha$ (veja Proposição 3.3.1), tem-se que $\mathbb{E}\{[g(X)]^k\} < \infty$.

Finalmente, substituindo as esperanças (3.22) e (3.24) na identidade (3.21), a prova da proposição segue. \square

3.5 Confiabilidade, taxa de risco e vida média residual

O objetivo desta seção é fornecer expressões matemáticas para a função de sobrevivência, a função de risco e a vida média residual

$$S(t; \boldsymbol{\theta}_\delta) = \int_t^\infty f(x; \boldsymbol{\theta}_\delta) dx;$$

$$h(t; \boldsymbol{\theta}_\delta) = \frac{f(t; \boldsymbol{\theta}_\delta)}{S(t; \boldsymbol{\theta}_\delta)};$$

$$\text{MRL}(t; \boldsymbol{\theta}_\delta) = \frac{1}{S(t; \boldsymbol{\theta}_\delta)} \int_t^\infty R(x; \boldsymbol{\theta}_\delta) dx;$$

da distribuição Gama bimodal, respectivamente.

Seja $Y \sim \text{BGama}(\boldsymbol{\theta}_0)$ uma variável aleatória com distribuição Gama bimodal. Por usar a fórmula de integração por partes, tem-se que

$$\mathbb{E}(\mathbf{1}_{\{Y \geq t\}} Y) = \frac{\exp(-\beta t)}{\beta} t^\alpha + \frac{\alpha}{\beta} \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}), \quad (3.25)$$

$$\mathbb{E}(\mathbf{1}_{\{Y \geq t\}} Y^2) = \frac{\exp(-\beta t)}{\beta} t^\alpha \left(t + \frac{\alpha + 1}{\beta} \right) + \frac{\alpha(\alpha + 1)}{\beta^2} \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}), \quad (3.26)$$

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\{Y \geq t\}} Y^3) &= \frac{\exp(-\beta t)}{\beta} t^\alpha \left[t^2 + \frac{\alpha + 2}{\beta} t + \frac{(\alpha + 1)(\alpha + 2)}{\beta^2} \right] \\ &+ \frac{\alpha(\alpha + 1)(\alpha + 2)}{\beta^3} \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}). \end{aligned} \quad (3.27)$$

Proposição 3.5.1. *Se $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$ então, a função de sobrevivência, a função de distribuição acumulada e a taxa de risco, respectivamente, podem ser expressados como segue*

$$1) S(t; \boldsymbol{\theta}_\delta) = \frac{\delta t}{\beta Z(\boldsymbol{\theta}_\delta)} \left[\delta \left(t + \frac{\alpha + 1}{\beta} \right) - 2 \right] f(t; \boldsymbol{\theta}_0) + S(t; \boldsymbol{\theta}_0);$$

$$2) F(t; \boldsymbol{\theta}_\delta) = -\frac{\delta t}{\beta Z(\boldsymbol{\theta}_\delta)} \left[\delta \left(t + \frac{\alpha + 1}{\beta} \right) - 2 \right] f(t; \boldsymbol{\theta}_0) + F(t; \boldsymbol{\theta}_0);$$

$$3) h(t; \boldsymbol{\theta}_\delta) = \frac{[1 + (1 - \delta t)^2] h(t; \boldsymbol{\theta}_0)}{\frac{\delta t}{\beta} [\delta (t + \frac{\alpha + 1}{\beta}) - 2] h(t; \boldsymbol{\theta}_0) + Z(\boldsymbol{\theta}_\delta)};$$

onde $S(t; \boldsymbol{\theta}_0) = 1 - F(t; \boldsymbol{\theta}_0) = \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_t^\infty y^{\alpha-1} e^{-\beta y} dy$.

Demonstração. Uma vez que a integral da função de sobrevivência pode ser expressada

em função de uma variável Y com distribuição Gama (clássica), isto é,

$$S(t; \boldsymbol{\theta}_\delta) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} [2 \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}) - 2\delta \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}Y) + \delta^2 \mathbb{E}(\mathbf{1}_{\{Y \geq t\}}Y^2)], \quad Y \sim \text{BGama}(\boldsymbol{\theta}_0),$$

ao combinar as identidades (3.25) e (3.26), a prova do item 1) segue. Já as demonstrações dos itens 2) e 3) seguem diretamente por combinar as equações $F(t; \boldsymbol{\theta}_\delta)$ e $h(t; \boldsymbol{\theta}_\delta)$ com o primeiro item, respectivamente. \square

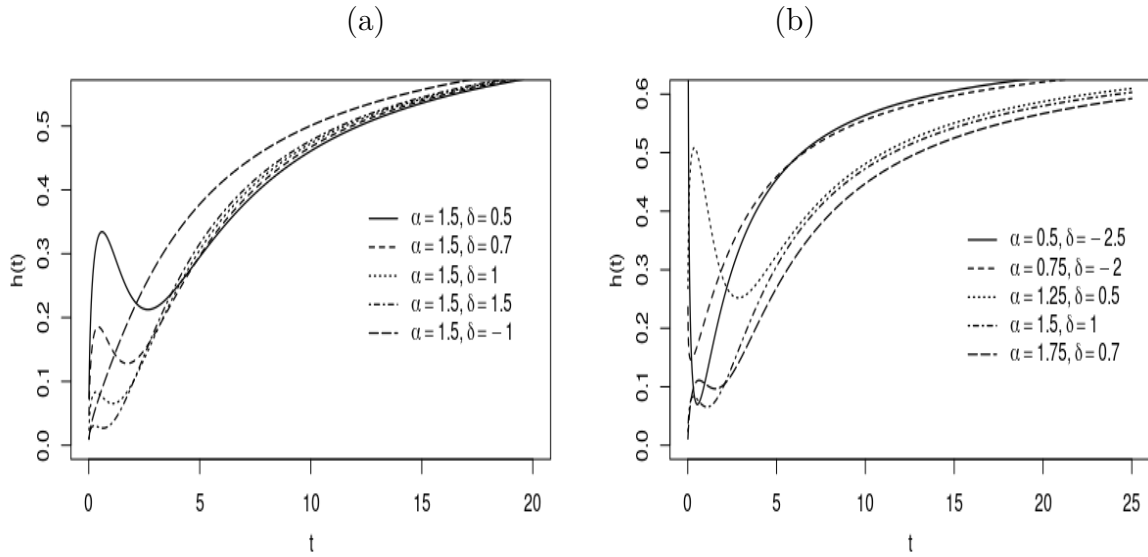


Figura 3.2: Função de Risco da BGama para alguns valores dos parâmetro com $\beta = 0,7$.

Observação 3.5.1 (Monotonicidade da função de risco quando $\delta = 0$). *Na Seção 2.4, foi visto que quando $\alpha > 1$, a função de risco $h(t; \boldsymbol{\theta}_0)$ é côncava e crescente; quando $\alpha < 1$, a função de risco é convexa e decrescente; e quando $\alpha = 1$ a função de risco é constante. Lembrando que, quando $\delta = 0$, tem-se a distribuição Gama (clássica).*

Proposição 3.5.2 (Monotonicidade decrescente da função de risco). *A função de risco $H(x; \boldsymbol{\theta}_\delta)$ da distribuição BGama (3.1) é decrescente quando $\alpha \leq 1$, $\delta > 0$ e $x \in (\frac{1}{\delta} - \frac{\alpha+1}{2\beta}, \frac{1}{\delta})$.*

Demonstração. Pela Proposição 3.5.1,

$$H(x; \boldsymbol{\theta}_\delta) = \frac{1 + (1 - \delta x)^2}{\frac{\delta x}{\beta} [\delta(x + \frac{\alpha+1}{\beta}) - 2] + \frac{Z(\boldsymbol{\theta}_\delta)}{H(x; \boldsymbol{\theta}_0)}}. \quad (3.28)$$

Através das derivadas da função $x \mapsto 1 + (1 - \delta x)^2$, sabe-se que esta decresce quando $x < 1/\delta$ e que a função $x \mapsto \frac{\delta x}{\beta} [\delta(x + \frac{\alpha+1}{\beta}) - 2]$ cresce quando $x > \frac{1}{\delta} - \frac{\alpha+1}{2\beta}$. Então, de acordo com essas informações e com o comportamento de $H(x; \boldsymbol{\theta}_0)$ apresentado na Observação 3.5.1, $x \mapsto \frac{\delta x}{\beta} [\delta(x + \frac{\alpha+1}{\beta}) - 2] + \frac{Z(\boldsymbol{\theta}_\delta)}{H(x; \boldsymbol{\theta}_0)}$ é uma função crescente quando $\alpha \leq 1$

e $x > \frac{1}{\delta} - \frac{\alpha+1}{2\beta}$. Logo, a taxa de risco, representada pela expressão (3.28), é o produto de duas funções decrescentes não negativas quando $\alpha \leq 1$, $\delta > 0$ e $x \in (\frac{1}{\delta} - \frac{\alpha+1}{2\beta}, \frac{1}{\delta})$, sendo assim decrescente. \square

Proposição 3.5.3 (Monotonicidade crescente da taxa de risco). *A taxa de risco $H(x; \boldsymbol{\theta}_\delta)$ da distribuição BGama (3.1) é crescente nas seguintes situações*

- 1) se $\delta = \beta$, $\alpha = 1$ e $x \in (0, 2/\beta)$;
- 2) se $\delta > \beta$, $\alpha = 1$ e $x \in (\delta + \beta - \sqrt{\delta^2 - \beta^2}/\beta\delta, \delta + \beta + \sqrt{\delta^2 - \beta^2}/\beta\delta)$;
- 3) se $\alpha > 1$, $\delta > \beta$, $a_{\delta,\beta} > 0$, $b_{\delta,\beta} > 0$ e $c_{\delta,\beta} > 0$, caso $x \in (0, x_1)$ ou $x \in (x_2, x_3)$, onde x_1, x_2, x_3 são as três raízes positivas distintas da equação polinomial $p_3(x) = 0$;
- 4) Se $\alpha > 1$, $0 < \delta < \beta$, $a_{\delta,\beta} < 0$, $b_{\delta,\beta} < 0$ e $c_{\delta,\beta} < 0$ para $x \in (0, x_0)$, onde x_0 é a única raiz positiva de $p_3(x) = 0$;

onde a equação polinomial $p_3(x)$ é definida por

$$p_3(x) = \beta\delta^2 x^3 - \delta[2(\delta + \beta) + \delta(\alpha - 1)]x^2 + 2[\delta + \beta + 2(\alpha - 1)]x - 2(\alpha - 1),$$

e os parâmetros $a_{\delta,\beta}$, $b_{\delta,\beta}$ e $c_{\delta,\beta}$ são definidos em (3.3), (3.4) e (3.5), respectivamente.

Demonstração. Nos Teoremas 3.2.1 e 3.2.2, foi visto que a densidade $f(x; \boldsymbol{\theta}_\delta)$ é crescente nos intervalos mencionados. Como $R(x; \boldsymbol{\theta}_\delta)$ é decrescente e por definição, $H(x; \boldsymbol{\theta}_\delta) = \frac{f(x; \boldsymbol{\theta}_\delta)}{R(x; \boldsymbol{\theta}_\delta)}$, a função de risco é o produto de duas funções crescentes e não negativas. Logo, as provas dos itens 1)-4) seguem. \square

Proposição 3.5.4. *Se $X \sim BGama(\boldsymbol{\theta}_\delta)$ então*

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\{X \geq t\}} X) &= \frac{1}{Z(\boldsymbol{\theta}_\delta)\Gamma(\alpha)} \left\{ 2 + \delta^2 \left[t^2 + \frac{\alpha+2}{\beta}t + \frac{(\alpha+1)(\alpha+2)}{\beta^2} \right] \right\} \beta^{\alpha-1} t^\alpha \exp(-\beta t) \\ &\quad - \frac{2\delta}{Z(\boldsymbol{\theta}_\delta)\Gamma(\alpha)} \left(t + \frac{\alpha+1}{\beta} \right) \beta^{\alpha-1} t^\alpha \exp(-\beta t) \\ &\quad + \frac{\alpha}{\beta Z(\boldsymbol{\theta}_\delta)} \left[2 + \frac{\delta^2}{\beta^2}(\alpha+1)(\alpha+2) - 2\frac{\delta}{\beta}(\alpha+1) \right] S(t; \boldsymbol{\theta}_0). \end{aligned}$$

Demonstração. Sabe-se que

$$\mathbb{E}(\mathbf{1}_{\{X \geq t\}} X) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} [2 \mathbb{E}(\mathbf{1}_{\{Y \geq t\}} Y) - 2\delta \mathbb{E}(\mathbf{1}_{\{Y \geq t\}} Y^2) + \delta^2 \mathbb{E}(\mathbf{1}_{\{Y \geq t\}} Y^3)], \quad Y \sim BGama(\boldsymbol{\theta}_0),$$

então se forem combinadas as identidades (3.25), (3.26) e (3.27), a prova segue. \square

Observação 3.5.2 (Função de vida média residual). *Ao usar a fórmula de integração por partes, note que a esperança $\mathbb{E}(\mathbf{1}_{\{X \geq t\}}X)$ pode ser escrita da seguinte maneira*

$$\mathbb{E}(\mathbf{1}_{\{X \geq t\}}X) = tS(t; \boldsymbol{\theta}_\delta) + \int_t^\infty R(x; \boldsymbol{\theta}_\delta) dx.$$

Uma vez que $xR(x; \boldsymbol{\theta}_\delta) \rightarrow 0$, quando $x \rightarrow \infty$, tem-se que

$$MRL(t; \boldsymbol{\theta}_\delta) = \left[\frac{1}{S(t; \boldsymbol{\theta}_\delta)} \mathbb{E}(\mathbf{1}_{\{X \geq t\}}X) \right] - t,$$

onde $S(t; \boldsymbol{\theta}_\delta)$ e $\mathbb{E}(\mathbf{1}_{\{X \geq t\}}X)$ são as características distribucionais fornecidas nas Proposições 3.5.1 e 3.5.4, respectivamente.

Observação 3.5.3. *No caso particular em que $\delta = 0$, note que*

$$\mathbb{E}(\mathbf{1}_{\{Y \geq t\}}Y) = \frac{\beta^{\alpha-1}}{\Gamma(\alpha)} t^\alpha e^{-\beta t} + \left(\frac{\alpha}{\beta} \right) S(t; \boldsymbol{\theta}_0), \quad Y \sim BGama(\boldsymbol{\theta}_0).$$

Então, pela Observação 3.5.2,

$$MRL(t; \boldsymbol{\theta}_0) = \frac{\beta^{\alpha-1} t^\alpha e^{-\beta t}}{\Gamma(\alpha) S(t; \boldsymbol{\theta}_0)} + \left(\frac{\alpha}{\beta} \right) - t.$$

A identidade acima, também, foi verificada por Govil (1983) [14], Equação (10).

Estimação por Máxima Verossimilhança na BGama($\boldsymbol{\theta}_\delta$)

Seja $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$ uma variável aleatória seguindo uma distribuição Gama bimodal com vetor de parâmetros $\boldsymbol{\theta}_\delta = (\alpha, \beta, \delta)$. Em seguida considere uma amostra aleatória, de tamanho n , (X_1, \dots, X_n) , independente e identicamente distribuída com distribuição BGama e com vetor de parâmetros $\boldsymbol{\theta}_\delta$ em um conjunto aberto $\Theta \in \mathbb{R}^3$, chamado de espaço paramétrico, onde valores distintos de $\boldsymbol{\theta}_\delta$ implicam em distribuições distintas para X_1 . Se $\mathbf{x} = (x_1, \dots, x_n)$ denota o valor observado da amostra aleatória (X_1, \dots, X_n) , então a função de log-verossimilhança de $\boldsymbol{\theta}_\delta$ é dada por

$$l(\boldsymbol{\theta}_\delta; \mathbf{x}) = -\log[Z(\boldsymbol{\theta}_\delta)] + \sum_{i=1}^n \log[1 + (1 - \delta x_i)^2] \quad (4.1)$$

$$+ \alpha \log(\beta) - \log[\Gamma(\alpha)] + (\alpha - 1) \sum_{i=1}^n \log(x_i) - n\beta\bar{x},$$

onde

$$Z(\boldsymbol{\theta}_\delta) = 2 + \frac{\alpha\delta}{\beta} \left[(1 + \alpha) \frac{\delta}{\beta} - 2 \right] \quad \text{e} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

As derivadas parciais de primeira e segunda ordem (e mistas) de $Z(\boldsymbol{\theta}_\delta)$ são dadas por

$$\begin{aligned} \frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial \alpha} &= \frac{\delta}{\beta} \left[(1 + 2\alpha) \frac{\delta}{\beta} - 2 \right], & \frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial \beta} &= -\frac{2\alpha\delta}{\beta^2} \left[(1 + \alpha) \frac{\delta}{\beta} - 1 \right], \\ \frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial \delta} &= \frac{2\alpha}{\beta} \left[(1 + \alpha) \frac{\delta}{\beta} - 1 \right], & \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial \alpha^2} &= \frac{2\delta^2}{\beta^2}, \\ \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial \beta^2} &= \frac{2\alpha\delta}{\beta^3} \left[(1 + \alpha) \frac{3\delta}{\beta} - 2 \right], & \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial \delta^2} &= \frac{2\alpha(1 + \alpha)}{\beta^2}, \end{aligned} \quad (4.2)$$

e

$$\begin{aligned}\frac{\partial^2 Z(\theta_\delta)}{\partial \alpha \partial \beta} &= \frac{\partial^2 Z(\theta_\delta)}{\partial \beta \partial \alpha} = -\frac{2\delta}{\beta^2} \left[(1 + 2\alpha) \frac{\delta}{\beta} - 1 \right], \\ \frac{\partial^2 Z(\theta_\delta)}{\partial \alpha \partial \delta} &= \frac{\partial^2 Z(\theta_\delta)}{\partial \delta \partial \alpha} = \frac{2}{\beta} \left[(1 + 2\alpha) \frac{\delta}{\beta} - 1 \right], \\ \frac{\partial^2 Z(\theta_\delta)}{\partial \beta \partial \delta} &= \frac{\partial^2 Z(\theta_\delta)}{\partial \delta \partial \beta} = -\frac{2\alpha}{\beta^2} \left[(1 + \alpha) \frac{2\delta}{\beta} - 1 \right].\end{aligned}$$

Note que $f(x; \theta_\delta)$ é uma função positiva e diferenciável de $\theta_\delta = (\alpha, \beta, \delta)$. Se o supremo $\hat{\theta}$ existe, ele deve satisfazer as seguintes equações de verossimilhança:

$$\frac{\partial l(\hat{\theta}; \mathbf{x})}{\partial \alpha} = 0, \quad \frac{\partial l(\hat{\theta}; \mathbf{x})}{\partial \beta} = 0, \quad \frac{\partial l(\hat{\theta}; \mathbf{x})}{\partial \delta} = 0. \quad (4.3)$$

Qualquer raiz não trivial das equações de verossimilhança (4.3) pode ser um MLE. No entanto, se o valor do parâmetro é um máximo global da função $l(\theta_\delta; \mathbf{x})$, este é chamado de estimador de máxima verossimilhança no senso estrito.

Ao derivar pela primeira vez a função $l(\theta_\delta; \mathbf{x})$, com respeito aos parâmetros, encontra-se

$$\begin{aligned}\frac{\partial l(\theta_\delta; \mathbf{x})}{\partial \alpha} &= -\frac{1}{Z(\theta_\delta)} \frac{\partial Z(\theta_\delta)}{\partial \alpha} + \log(\beta) - \Psi^{(0)}(\alpha) + \sum_{i=1}^n \log(x_i) \\ &= -\frac{\delta[(1 + 2\alpha)\delta - 2\beta]}{2\beta^2 + \alpha\delta[(1 + \alpha)\delta - 2\beta]} + \log(\beta) - \Psi^{(0)}(\alpha) + \sum_{i=1}^n \log(x_i), \\ \frac{\partial l(\theta_\delta; \mathbf{x})}{\partial \beta} &= -\frac{1}{Z(\theta_\delta)} \frac{\partial Z(\theta_\delta)}{\partial \beta} + \frac{\alpha}{\beta} - n\bar{x} \\ &= \frac{2\alpha\delta[(1 + \alpha)\delta - \beta]}{2\beta^3 + \alpha\delta\beta[(1 + \alpha) - 2\beta]} + \frac{\alpha}{\beta} - n\bar{x}, \\ \frac{\partial l(\theta_\delta; \mathbf{x})}{\partial \delta} &= -\frac{1}{Z(\theta_\delta)} \frac{\partial Z(\theta_\delta)}{\partial \delta} - 2 \sum_{i=1}^n \frac{1 - \delta x_i}{1 + (1 - \delta x_i)^2} \\ &= -\frac{2\alpha[(1 + \alpha)\delta - \beta]}{2\beta^2 + \alpha\delta[(1 + \alpha)\delta - 2\beta]} - 2 \sum_{i=1}^n \frac{1 - \delta x_i}{1 + (1 - \delta x_i)^2}.\end{aligned} \quad (4.4)$$

Percebe-se que as equações em (4.3) não são lineares, por isso são utilizados métodos numéricos para a estimação dos parâmetros da distribuição Gama bimodal.

Já as derivadas de segunda ordem de $l(\boldsymbol{\theta}_\delta; \mathbf{x})$ podem ser escritas como

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \alpha^2} &= D_{\boldsymbol{\theta}_\delta}(\alpha, \alpha) - \Psi^{(1)}(\alpha), \\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \beta^2} &= D_{\boldsymbol{\theta}_\delta}(\beta, \beta) - \frac{\alpha}{\beta^2}, \\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \delta^2} &= D_{\boldsymbol{\theta}_\delta}(\delta, \delta) + 2 \sum_{i=1}^n \frac{x_i [1 - (1 - \delta x_i)^2]}{[1 + (1 - \delta x_i)^2]^2}.\end{aligned}\tag{4.5}$$

Note que as derivadas mistas de segunda ordem de $l(\boldsymbol{\theta}_\delta; \mathbf{x})$ são dadas por

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \alpha \partial \beta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \beta \partial \alpha} = D_{\boldsymbol{\theta}_\delta}(\alpha, \beta) + \frac{1}{\beta}, \\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \alpha \partial \delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \delta \partial \alpha} = D_{\boldsymbol{\theta}_\delta}(\alpha, \delta), \\ \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \beta \partial \delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; \mathbf{x})}{\partial \delta \partial \beta} = D_{\boldsymbol{\theta}_\delta}(\beta, \delta),\end{aligned}$$

onde

$$D_{\boldsymbol{\theta}_\delta}(u, v) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \left[\frac{1}{Z(\boldsymbol{\theta}_\delta)} \frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial u} \frac{\partial Z(\boldsymbol{\theta}_\delta)}{\partial v} - \frac{\partial^2 Z(\boldsymbol{\theta}_\delta)}{\partial u \partial v} \right], \quad u, v \in \{\alpha, \beta, \delta\}.$$

Sabe-se que, pelo Teorema de Schwarz's, as variáveis parciais mistas são comutativas em $\boldsymbol{\theta}_\delta \in \mathbb{R}^3$ visto que essas funções possuem segundas derivadas contínuas nesse conjunto.

Se $X \sim \text{BGama}(\boldsymbol{\theta}_\delta)$, sobre clássicas condições de regularidade, a matriz de informação de Fisher associada a X , denotada por $I_X(\boldsymbol{\theta}_\delta)$, é dada por

$$I_X(\boldsymbol{\theta}_\delta) = - \begin{bmatrix} D_{\boldsymbol{\theta}_\delta}(\alpha, \alpha) - \Psi^{(1)}(\alpha) & D_{\boldsymbol{\theta}_\delta}(\alpha, \beta) + \frac{1}{\beta} & D_{\boldsymbol{\theta}_\delta}(\alpha, \delta) \\ D_{\boldsymbol{\theta}_\delta}(\alpha, \beta) + \frac{1}{\beta} & D_{\boldsymbol{\theta}_\delta}(\beta, \beta) - \frac{\alpha}{\beta^2} & D_{\boldsymbol{\theta}_\delta}(\beta, \delta) \\ D_{\boldsymbol{\theta}_\delta}(\alpha, \delta) & D_{\boldsymbol{\theta}_\delta}(\beta, \delta) & D_{\boldsymbol{\theta}_\delta}(\delta, \delta) + 2 \mathbb{E} \left\{ \frac{X[1 - (1 - \delta X)^2]}{[1 + (1 - \delta X)^2]^2} \right\} \end{bmatrix},$$

onde, pela Proposição 3.3.1, tem-se a finitude da esperança $\mathbb{E} \left\{ \frac{X[1 - (1 - \delta X)^2]}{[1 + (1 - \delta X)^2]^2} \right\}$. Mais precisamente,

$$\mathbb{E} \left\{ \left| \frac{X[1 - (1 - \delta X)^2]}{[1 + (1 - \delta X)^2]^2} \right| \right\} \leq 2 \mathbb{E}X - 2\delta \mathbb{E}X^2 + \delta^2 \mathbb{E}X^3 < \infty.$$

Teorema 4.0.1. *Se $\Theta = \{\alpha \in \mathbb{R}^+ : \varepsilon_0 < \alpha < \alpha_0\}$ é um espaço paramétrico onde $0 < \varepsilon_0 = \varepsilon_0(\beta, \delta) < \alpha_0$ é um valor fixo e*

$$\alpha_0 = \alpha_0(\beta, \delta) = \frac{1}{2\delta} \left[(2\beta - \delta) + \sqrt{(2\beta - \delta)^2 + 2\delta(4\beta - \delta)} \right],$$

com β, δ conhecidos de tal modo que $0 < \delta < 2\beta$. Então, com probabilidade 1, quando $n \rightarrow \infty$, a função de verossimilhança $\frac{d l(\alpha; \mathbf{x})}{d\alpha} = 0$ possui uma solução consistente, denotada por $\hat{\alpha}$.

Demonstração. Se β e δ são conhecidos, pode-se simplificar a notação escrevendo-se BGama(α), $f(x; \alpha)$, $Z(\alpha)$ e $D_\alpha(\alpha, \alpha)$ ao invés de BGama(θ_δ), $f(x; \theta_\delta)$, $Z(\theta_\delta)$ e $D_{\theta_\delta}(\alpha, \alpha)$, respectivamente.

Se $X \sim \text{BGama}(\alpha)$, segundo Cramér (1946) [7], para provar a tese deste teorema é suficiente verificar as seguintes condições:

- ① $\mathbb{E}\left\{\frac{d \log[f(X; \alpha)]}{d\alpha}\right\} = 0$, para todo $\alpha \in \Theta$;
- ② $-\infty < \mathbb{E}\left\{\frac{d^2 \log[f(X; \alpha)]}{d\alpha^2}\right\} < 0$, para todo $\alpha \in \Theta$;
- ③ Existe uma função $H(x)$, independente do parâmetro desconhecido α , tal que para todo $\alpha \in \Theta$,

$$\left| \frac{d^3 \log[f(x; \alpha)]}{d\alpha^3} \right| < H(x) \quad \text{e} \quad \mathbb{E}[H(X)] = M(\alpha) < \infty.$$

De fato, ao tomar $n = 1$, a equação (4.4) se torna

$$\frac{d \log[f(x; \alpha)]}{d\alpha} = -\frac{1}{Z(\alpha)} \frac{dZ(\alpha)}{d\alpha} + \log(\beta) - \Psi^{(0)}(\alpha) + \log(x).$$

Então,

$$\mathbb{E}\left\{\frac{d \log[f(X; \alpha)]}{d\alpha}\right\} = \mathbb{E}[\log(X)] - \frac{1}{Z(\alpha)} \frac{dZ(\alpha)}{d\alpha} + \log(\beta) - \Psi^{(0)}(\alpha). \quad (4.6)$$

Combinando o Item 1) da Proposição 3.3.3 com as identidades apresentadas em (4.2), chega-se na seguinte igualdade

$$\mathbb{E}[\log(X)] - \frac{1}{Z(\alpha)} \frac{dZ(\alpha)}{d\alpha} = \Psi^{(0)}(\alpha) - \log(\beta).$$

Substituindo esta identidade em (4.6), tem-se que $\mathbb{E}\left\{\frac{d \log[f(X; \alpha)]}{d\alpha}\right\} = 0$ para todo $\alpha \in \Theta$, verificando-se assim o Item ①.

Tomando $n = 1$ em (4.5), utilizando a definição de $D_\alpha(\alpha, \alpha)$ e utilizando as identidades em (4.2), segue-se que

$$\begin{aligned} \frac{d^2 \log[f(x; \alpha)]}{d\alpha^2} &= D_\alpha(\alpha, \alpha) - \Psi^{(1)}(\alpha) \\ &= \frac{\delta^3 / \beta^4}{Z^2(\alpha)} [2\delta\alpha^2 - 2(2\beta - \delta)\alpha - (4\beta - \delta)] - \Psi^{(1)}(\alpha). \end{aligned} \quad (4.7)$$

Visto que $0 < \delta < 2\beta$, a constante α_0 é bem definida. Note que, para $0 < \alpha < \alpha_0$, a seguinte desigualdade é válida:

$$2\delta\alpha^2 - 2(2\beta - \delta)\alpha - (4\beta - \delta) < 0.$$

Também sabe-se que (veja Guo e Qi (2010) [16], Corolário 1.2.),

$$\Psi^{(1)}(\alpha) > \frac{e^{1/\alpha}}{[\exp(1/\alpha) - 1]\alpha^2} > 0.$$

Portanto, $\frac{d^2 \log[f(x; \alpha)]}{d\alpha^2} < 0$, para todo $x > 0$. Consequentemente, a afirmação no Item ② é satisfeita.

Para verificar o terceiro item, primeiro deriva-se, com respeito a α , ambos lados da equação (4.7) para obter o seguinte

$$\begin{aligned} \frac{d^3 \log[f(x; \alpha)]}{d\alpha^3} &= \frac{\delta^3/\beta^6}{Z^3(\alpha)} \left(-2\delta[(1+2\alpha)\delta - 2\beta] [2\delta\alpha^2 - 2(2\beta - \delta)\alpha - (4\beta - \delta)] \right. \\ &\quad \left. + [4\alpha\delta - 2(2\beta - \delta)] \{2\beta^2 + \alpha\delta[(1+\alpha)\delta - 2\beta]\} \right) - \Psi^{(2)}(\alpha), \end{aligned}$$

onde

$$\begin{aligned} G(\alpha) &= \frac{\delta^3/\beta^6}{Z^3(\alpha)} \left\{ 2\delta[(1+2\alpha)\delta + 2\beta] [2\delta\alpha^2 + 2(2\beta + \delta)\alpha + (4\beta + \delta)] \right. \\ &\quad \left. + [4\alpha\delta + 2(2\beta + \delta)] \{2\beta^2 + \alpha\delta[(1+\alpha)\delta + 2\beta]\} \right\}. \end{aligned}$$

Então, para todo $x > 0$ e $\alpha \in \Theta$,

$$\left| \frac{d^3 \log[f(x; \alpha)]}{d\alpha^3} \right| \leq G(\alpha) + |\Psi^{(2)}(\alpha)|. \quad (4.8)$$

Como $G(\alpha)$ é uma função crescente em α e $\alpha < \alpha_0$, tem-se que

$$G(\alpha) \leq G(\alpha_0), \quad \text{para todo } \alpha \in \Theta. \quad (4.9)$$

Combinando as inequações $\Psi^{(n)}(\alpha) > -(n-1)! \exp[-n\Psi^{(0)}(\alpha)]$, para n ímpar (Batir, 2007 [4]), e $\Psi^{(0)}(\alpha) > \log(\alpha + \frac{1}{2}) - \frac{1}{\alpha}$ (Elezovic, 2000 [9]), chega-se à seguinte desigualdade

$$-\Psi^{(2)}(\alpha) < \exp \left\{ 2 \left[\frac{1}{\alpha} - \log \left(\alpha + \frac{1}{2} \right) \right] \right\} < \exp \left\{ 2 \left[\frac{1}{\varepsilon_0} - \log \left(\varepsilon_0 + \frac{1}{2} \right) \right] \right\}, \quad \text{para todo } \alpha \in \Theta.$$

Por outro lado, pelo Corolário 1.2 de Guo e Qi (2010) [16],

$$\Psi^{(2)}(\alpha) < \frac{\exp(1/\alpha) \{1 - 2\alpha[\exp(1/\alpha) - 1]\}}{[\exp(1/\alpha) - 1]^2 \alpha^4} < 0.$$

Logo,

$$|\Psi^{(2)}(\alpha)| = -\Psi^{(2)}(\alpha) < \exp\left\{2\left[\frac{1}{\varepsilon_0} - \log\left(\varepsilon_0 + \frac{1}{2}\right)\right]\right\}. \quad (4.10)$$

Combinando os Itens (4.8), (4.9) e (4.10), tem-se que

$$\left|\frac{d^3 \log[f(x; \alpha)]}{d\alpha^3}\right| < G(\alpha_0) + \exp\left\{2\left[\frac{1}{\varepsilon_0} - \log\left(\varepsilon_0 + \frac{1}{2}\right)\right]\right\}, \quad \text{para todo } \alpha \in \Theta,$$

onde $H(x) = G(\alpha_0) + \exp\{2[\frac{1}{\varepsilon_0} - \log(\varepsilon_0 + \frac{1}{2})]\} = \text{constante}$. Logo, a prova do Item ③ está completa, e assim, a prova do teorema segue por Cramér (1946) [7]. \square

4.1 Simulações de Monte Carlo

Nesta seção foram realizadas simulações de Monte Carlo para avaliar o comportamento dos estimadores de verossimilhança de α e β . Dessa forma, o parâmetro δ foi considerado como conhecido e apenas α e β foram estimados.

Os estudos de simulação geraram amostras de tamanho $n \in \{10, 60, 120\}$ e com as seguintes variações de parâmetro: $\alpha \in \{0.50, 1.00, 1.50\}$, $\beta \in \{1.00\}$, e $\delta \in \{-10, -5, -1, 1, 5, 10\}$. Para cada um desses cenários foram simuladas 1000 amostras.

Para cada valor de parâmetro δ e tamanho de amostra, os valores empíricos para o viés e Erro Quadrático Médio (EQM) dos estimadores por ML são apresentados na Tabela 4.1. Nesta tabela é possível notar que à medida que o tamanho da amostra cresce, o viés e o EQM diminuem, indicando que o estimador é assintoticamente não-viesado. Em outras palavras, os parâmetros estimados por Máxima Verossimilhança são consistentes.

Tabela 4.1: Viés dos valores simulados (EQM nos parênteses) dos estimadores de Máxima Verossimilhança do Modelo BGama

n	δ	BG($\alpha = 0, 50, \beta = 1, 00, \delta$)		BG($\alpha = 1, 00, \beta = 1, 00, \delta$)		BG($\alpha = 1, 50, \beta = 1, 00, \delta$)	
		Viés($\hat{\alpha}$)	Viés($\hat{\beta}$)	Viés($\hat{\alpha}$)	Viés($\hat{\beta}$)	Viés($\hat{\alpha}$)	Viés($\hat{\beta}$)
20	-10	0,4053 (0,7019)	0,2028 (0,1850)	0,5157 (1,3893)	0,2018 (0,2165)	0,5574 (2,2031)	0,1770 (0,2427)
	-5	0,2785 (0,3735)	0,1617 (0,1328)	0,4580 (1,0668)	0,1895 (0,1878)	0,5453 (1,8772)	0,1795 (0,2199)
	1	0,0336 (0,0225)	0,1196 (0,1426)	0,1014 (0,0965)	0,0810 (0,0592)	0,2176 (0,3481)	0,0931 (0,0685)
	5	0,3599 (0,6604)	0,1661 (0,1435)	0,6729 (1,8993)	0,2345 (0,2417)	0,7525 (2,7966)	0,2253 (0,2561)
	10	0,5653 (1,2008)	0,2435 (0,2370)	0,4242 (2,0143)	0,1161 (0,3131)	0,6814 (2,7151)	0,2084 (0,2587)
60	-10	0,1294 (0,1080)	0,0632 (0,0302)	0,1567 (0,2546)	0,0601 (0,0391)	0,0321 (0,5791)	0,0021 (0,0697)
	-5	0,0774 (0,0503)	0,0449 (0,0215)	0,1392 (0,1843)	0,0565 (0,0330)	0,0835 (0,4332)	0,0207 (0,0561)
	1	-0,0033 (0,0059)	0,0134 (0,0191)	0,0287 (0,0241)	0,0225 (0,0142)	0,0623 (0,0699)	0,0268 (0,0152)
	5	0,0766 (0,0516)	0,0373 (0,0172)	0,2218 (0,3057)	0,0753 (0,0391)	0,2363 (0,5169)	0,0692 (0,0459)
	10	0,1912 (0,1766)	0,0798 (0,0358)	-0,2430 (0,5223)	-0,1394 (0,1098)	0,1682 (0,5757)	0,0482 (0,0551)
120	-10	0,0626 (0,0458)	0,0286 (0,0124)	0,0767 (0,1162)	0,0275 (0,0169)	-0,1228 (0,3672)	-0,0502 (0,0448)
	-5	0,0327 (0,0206)	0,0173 (0,0089)	0,0677 (0,0821)	0,0257 (0,0141)	-0,0376 (0,2493)	-0,0217 (0,0328)
	1	-0,0107 (0,0031)	-0,0054 (0,0085)	0,0133 (0,0118)	0,0090 (0,0065)	0,0298 (0,0321)	0,0114 (0,0068)
	5	0,0273 (0,0169)	0,0118 (0,0068)	0,1101 (0,1292)	0,0354 (0,0159)	0,1176 (0,2350)	0,0326 (0,0197)
	10	0,0944 (0,0706)	0,0373 (0,0140)	-0,4445 (0,3810)	-0,2162 (0,0848)	0,0386 (0,3181)	0,0073 (0,0300)

O Modelo de Regressão BGama($\boldsymbol{\theta}_\delta$) com Censura

O objetivo desta seção é propor um modelo de regressão paramétrico baseado na distribuição Gama bimodal que pode ser utilizado na área de Análise de Sobrevivência. Esse modelo é uma alternativa aos modelos de regressão de locação-escala e pode ser implementado computacionalmente utilizando-se de funções prontas no software R [www.r-project.org]. Por ser um modelo paramétrico, é possível calcular a função de verossimilhança e compará-lo a outros modelos.

No Modelo de Regressão Gama bimodal a relação entre um vetor de covariáveis $\mathbf{v} = (v_1, \dots, v_p)^T$ e os parâmetros α e β da distribuição BGama($\boldsymbol{\theta}_\delta$) é expressa pelas seguintes equações

$$\alpha_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1); \quad \beta_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2);$$

onde $i = 1, \dots, n$, além disso $\boldsymbol{\tau}_1 = (\tau_{11}, \dots, \tau_{1p})^T$ e $\boldsymbol{\tau}_2 = (\tau_{21}, \dots, \tau_{2p})^T$ denotam os coeficientes de regressão, e $\mathbf{v}_i^T = (v_{i1}, \dots, v_{ip})$.

A partir dessas especificações e da Proposição 3.5.1, tem-se que a função de sobrevivência do modelo BGama é expressa por

$$R(x|\mathbf{v}) = 1 + \frac{\delta x^{\exp(\mathbf{v}^T \boldsymbol{\tau}_1)} \exp(\mathbf{v}^T \boldsymbol{\tau}_2)^{\exp(\mathbf{v}^T \boldsymbol{\tau}_1)-1} \exp[-\exp(\mathbf{v}^T \boldsymbol{\tau}_2) x]}{\Gamma[\exp(\mathbf{v}^T \boldsymbol{\tau}_1)] Z(\boldsymbol{\theta}_\delta)} \times \left\{ \delta \left[x + \frac{\exp(\mathbf{v}^T \boldsymbol{\tau}_1) - 1}{\exp(\mathbf{v}^T \boldsymbol{\tau}_2)} \right] - 2 \right\} - I(\exp(\mathbf{v}^T \boldsymbol{\tau}_1), \exp(\mathbf{v}^T \boldsymbol{\tau}_2) x), \quad (5.1)$$

onde $I(k, y) = \gamma(k, y)/\Gamma(k)$ é a função razão-Gama incompleta; $\gamma(k, y) = \int_0^y w^{k-1} e^{-w} dw$ é a função Gama incompleta e $\Gamma(\cdot)$ é a função Gama.

No caso de censuras, considere uma amostra $(x_1, \mathbf{v}_1), \dots, (x_n, \mathbf{v}_n)$ de n observações independentes. Nessa amostra cada indivíduo i possui certo tempo de vida X_i e tempo

de censura C_i , onde X_i e C_i são variáveis aleatórias independentes e $x_i = \min(X_i, C_i)$, para $i = 1, \dots, n$. A censura, nesse caso, é não-informativa, por isso o tempo de vida observado e os tempos de censura são independentes. Se F é o tempo de vida e C é o tempo de censura de indivíduos para os quais x_i representa o tempo de vida ou censura, então a função de log-verossimilhança total do vetor de parâmetros $\boldsymbol{\psi} = (\delta, \boldsymbol{\tau}_1^T, \boldsymbol{\tau}_2^T)^T$, do modelo (5.1), é dada por

$$\begin{aligned} l(\boldsymbol{\psi}) &= \sum_{i \in F} \log[1 + (1 - \delta x_i)^2] \sum_{i \in F} \log[Z_i(\boldsymbol{\theta}_\delta)] \sum_{i \in F} \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1) (\mathbf{v}_i^T \boldsymbol{\tau}_2) \\ &\quad - \sum_{i \in F} \log\{\Gamma[\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)]\} + \sum_{i \in F} (\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1) - 1) \log(x_i) \\ &\quad - \sum_{i \in F} \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2) x_i + \sum_{i \in C} l_i^{(c)}(\boldsymbol{\psi}), \end{aligned} \quad (5.2)$$

onde

$$\begin{aligned} l_i^{(c)}(\boldsymbol{\psi}) &= \log \left\{ 1 + \frac{\delta x_i^{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)} \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)^{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1) - 1} \exp[-\exp(\mathbf{v}_i^T \boldsymbol{\tau}_2) x_i]}{\Gamma[\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)] Z_i(\boldsymbol{\theta}_\delta)} \right. \\ &\quad \left. \times \left\{ \delta \left[x_i + \frac{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1) - 1}{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)} \right] - 2 \right\} - I(\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1), \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2) x_i) \right\} \end{aligned}$$

e

$$Z_i(\boldsymbol{\theta}_\delta) = 2 + \frac{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_1) \delta}{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)} \left\{ [1 + \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)] \left[\frac{\delta}{\exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)} \right] - 2 \right\}.$$

O MLE, denotado por $\widehat{\boldsymbol{\psi}}$, do vetor de parâmetros desconhecidos pode ser determinado pela maximização da função log-verossimilhança (5.2). Os valores iniciais para os parâmetros $\boldsymbol{\tau}_1$ e $\boldsymbol{\tau}_2$ podem ser obtidos pelo ajuste do modelo BGama($\boldsymbol{\theta}_0$), ou seja, pelo ajuste do modelo Gama clássico.

A construção de intervalos de confiança aproximados para os parâmetros deste modelo pode ser realizada a partir de uma distribuição Normal multivariada, $N_{2p+1}(0, J(\widehat{\boldsymbol{\psi}})^{-1})$. O método Delta é um método simples e útil para realizar essas aproximações.

Aplicações

Nessa seção, foram apresentadas duas aplicações da distribuição $BGama(\boldsymbol{\theta}_\delta)$. Na primeira aplicação, foram modelados dados bimodais sobre inundações de um rio e, na segunda, aplicou-se o Modelo de Regressão $BGama(\boldsymbol{\theta}_\delta)$ em dados com censura de pacientes com câncer gástrico. A qualidade do ajuste desses modelos foi comparada a de outros modelos encontrados na literatura por meio das métricas abordadas na Seção 1.8.

6.1 Aplicação 1. Dados do Rio Wheaton

Os dados utilizados nesta aplicação se tratam de excedências no volume (em m^3/s) do Rio Wheaton próximo a Carcross no território Yukon, Canadá. São informações sobre 72 inundações que ocorreram no período de 1958 a 1984, arredondados em um dígito decimal. Em estudos posteriores, Choulakian e Stephens (2001) [6] e Akinsete et al. (2008) [2] analisaram esses dados, e Akinsete et al. (2008) [2] modelou as excedências com as distribuições: Pareto, Weibull com três parâmetros, Pareto generalizada e Beta-Pareto. O ajuste destas distribuições foi comparado pela medida KS, conforme está na Tabela 6.1.

Tabela 6.1: Medidas KS e p -valores associados aos dados do Rio Wheaton

Modelo	KS	p -valor
Pareto	2,7029	<0,000
Weibull com três parâmetros	1,6734	0,0074
Pareto Generalizada	1,205	0,1094
Beta Pareto	1,2534	0,0864

Além das distribuições ajustadas neste trabalho, utilizou-se também a distribuição Gama Generalizada Kumaraswamy (KumGG) (para $x > 0$) definida por Pascoa et al. (2011) [25] nas comparações. Essa comparação é relevante visto que pode-se obter a maioria das distribuições clássicas utilizadas na análise de sobrevivência a partir de casos

particulares dessa distribuição, como por exemplo, a distribuição Weibull Exponenciada ($\lambda = 1$ e $\varphi = 1$), Gamma ($\lambda = 1$, $\varphi = 1$ e $\tau = 1$) e Weibull ($\lambda = 1$, $\varphi = 1$ e $k = 1$). A densidade da KumGG é dada por

$$f(x; \alpha, \tau, k, \lambda, \varphi) = \frac{\lambda \varphi \tau}{\alpha \Gamma(k)} \left(\frac{x}{\alpha}\right)^{\tau k - 1} \exp\left[-\left(\frac{x}{\alpha}\right)^\tau\right] \gamma_1^{\lambda - 1}\left[k, \left(\frac{x}{\alpha}\right)^\tau\right] \left(1 - \gamma_1^\lambda\left[k, \left(\frac{x}{\alpha}\right)^\tau\right]\right)^{\varphi - 1},$$

onde todos os parâmetros são positivos: $\gamma_1(k, y) = \gamma(k, y)/\Gamma(k)$ é a função razão-Gamma incompleta; α é o parâmetro de escala e os outros parâmetros positivos τ , k , φ e λ são parâmetros de forma.

Também foi feito o ajuste com a distribuição Weibull Modificada (MW) (para $x \geq 0$), definida por Lai et al. (2003) [20], com os parâmetros $\alpha > 0$, $\tau \geq 0$ e $k \geq 0$, com a densidade:

$$f(x; \alpha, \tau, k) = \alpha x^{(\tau - 1)}(\tau + kx) \exp[kx - \alpha x^\tau \exp(kx)].$$

Na Tabela 6.2 podem-se ver os MLEs estimados em cada modelo, e na Tabela 6.3 são apresentadas medidas estatísticas da qualidade de ajuste do Modelo. Na Tabela 6.3 nota-se que a distribuição BGama(θ_δ) obteve as menores estatísticas de AIC, BIC e W*. Ao comparar a medida KS com os resultados da Tabela 6.1, nota-se que também obteve o maior p -valor. Portanto a distribuição BGama(θ_δ) é, dentre essas, a que possui o melhor ajuste para esses dados.

Tabela 6.2: MLEs dos parâmetros do modelo para os dados do Rio Wheaton

Modelo	α	β	δ		
BGama(θ_δ)	1,054	0,176	0,177		
	(0,145)	(0,111)	(0,032)		
	α	τ	k	λ	φ
Kw-GG	548,542	0,103	0,098	158,570	869,87
	(252,1)	(0,082)	(0,007)	(70,450)	(196,8)
Gamma	14,558	1	0,838	1	1
	(2,816)	(-)	(0,121)	(-)	(-)
EW	11,278	1,380	0,591	1	1
	(1,506)	(0,284)	(0,149)	(-)	(-)
Weibull	11,632	0,901	1	1	1
	(1,601)	(0,085)	(-)	(-)	(-)
MW	0,124	0,775	0,010		
	(0,034)	(0,124)	(0,007)		

Tabela 6.3: Medidas Estatísticas

Model	AIC	BIC	W^*	KS
BGama(θ_δ)	501,51	508,34	0,038	0,065 (0,918)
Kw-GG	514,01	525,39	0,159	0,099 (0,473)
gamma	506,68	511,24	0,130	0,102 (0,433)
EW	505,85	512,68	0,074	0,096 (0,516)
Weibull	506,99	511,55	0,137	0,105 (0,402)
MW	507,34	514,17	0,097	0,100 (0,466)

Na Figura 6.1, são apresentadas a FDP e a FDA do ajuste pela BGama(θ_δ) e também o gráfico *QQ-PLLOT* dos resíduos quantílicos. Na Figura 6.1-(c), onde é apresentado um *QQ-plot* para a distribuição BGama(θ_δ), nota-se que os resíduos quantílicos seguem aproximadamente uma distribuição Normal, o que indica que o modelo possui ajuste satisfatório. Nas outras imagens, também, percebe-se que o ajuste é coerente com os dados visto que não apresenta pontos muito distantes do que foi estimado.

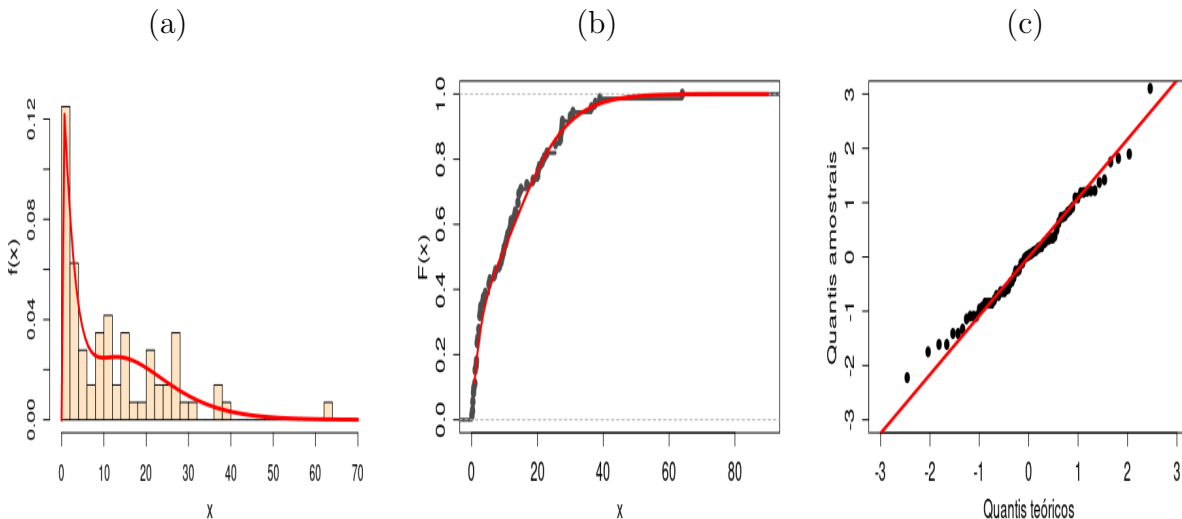


Figura 6.1: (a) PDF estimada do modelo BGama(θ_δ). (b) CDF empírica e estimada do modelo BGama(θ_δ). (c) QQ plot dos resíduos quantílicos do ajuste.

6.2 Aplicação 2. Dados de câncer gástrico

Nesta segunda aplicação, foram considerados os dados analisados por Martinez et al. (2013) [22] e Ortega et al. (2017) [24]. Estas duas últimas pesquisas usaram o modelo de fração de cura para ajustar as informações sobre câncer gástrico. O *dataset* consiste em uma amostra com 201 pacientes em diferentes estados clínicos, em que 76 desses pacientes receberam quimiorradioterapia adjuvante e foram operados e os outros 125 pacientes passaram apenas pela cirurgia. A variável resposta é o tempo de morte em meses desde a cirurgia. A censura desses dados é de 53.2% e as variáveis utilizadas foram:

- x_i : Tempo até a morte, em meses após a cirurgia;
- v_{i1} : Tipo de terapia (0=quimiorradioterapia adjuvante; 1=apenas cirurgia) para $i = 1, \dots, 201$.

O Modelo de Regressão BGama(θ_δ) possui os seguintes coeficientes para essas variáveis

$$\alpha_i = \exp(\tau_{10} + v_{i1}\tau_{11}) \quad \text{e} \quad \beta_i = \exp(\tau_{20} + v_{i1}\tau_{21}), \quad i = 1, \dots, 201,$$

onde τ_{10} e τ_{20} são os interceptos dos parâmetros α e β , respectivamente. E τ_{11} e τ_{21} são os coeficientes relacionados a covariável para os parâmetros α e β , respectivamente.

Tabela 6.4: MLE, SE e p -valores dos parâmetros do modelo de regressão BGama(θ_δ) nos dados de câncer gástrico

Parâmetro	Estimativa	SE	p -Valor
τ_{10}	-0,221	0,073	0,003
τ_{11}	1,306	0,100	<0,001
τ_{20}	-3,506	0,077	<0,001
τ_{21}	1,077	0,102	<0,001
δ	0,032	0,002	<0,001

Na Tabela 6.4 observa-se pela medida do p -valor que existem diferenças significativas do efeito dos tratamentos em ambos os parâmetros, ou seja, ao nível de 0,05 de significância a covariável é significativa. Quando apenas a cirurgia é realizada aumenta-se o valor dos parâmetros α e β visto que τ_{11} e τ_{21} são positivos e significantes. O parâmetro δ também foi considerado significativo no modelo, ou seja, o ajuste proporcionado por uma Gama Bimodal e o proporcionado por uma Gama (clássica) é estatisticamente diferente.

Tabela 6.5: Medidas GD, AIC e BIC para os modelos de regressão BGama, Gama e Weibull para os dados de câncer gástrico

Modelo	GD	AIC	BIC
BGama(θ_δ)	866,53	876,53	893,04
BGama(θ_0)	871,28	879,28	892,49
Weibull	872,34	880,34	893,55

Este modelo parece ser o mais razoável quando comparado a outros, visto que performou melhor na maioria das métricas de ajuste. Na Tabela 6.5 verifica-se que o modelo de regressão BGama(θ_δ) possui o menor Desvio Global (GD) e os menores valores de AIC, entretanto seu BIC empatou com o da distribuição Gama.

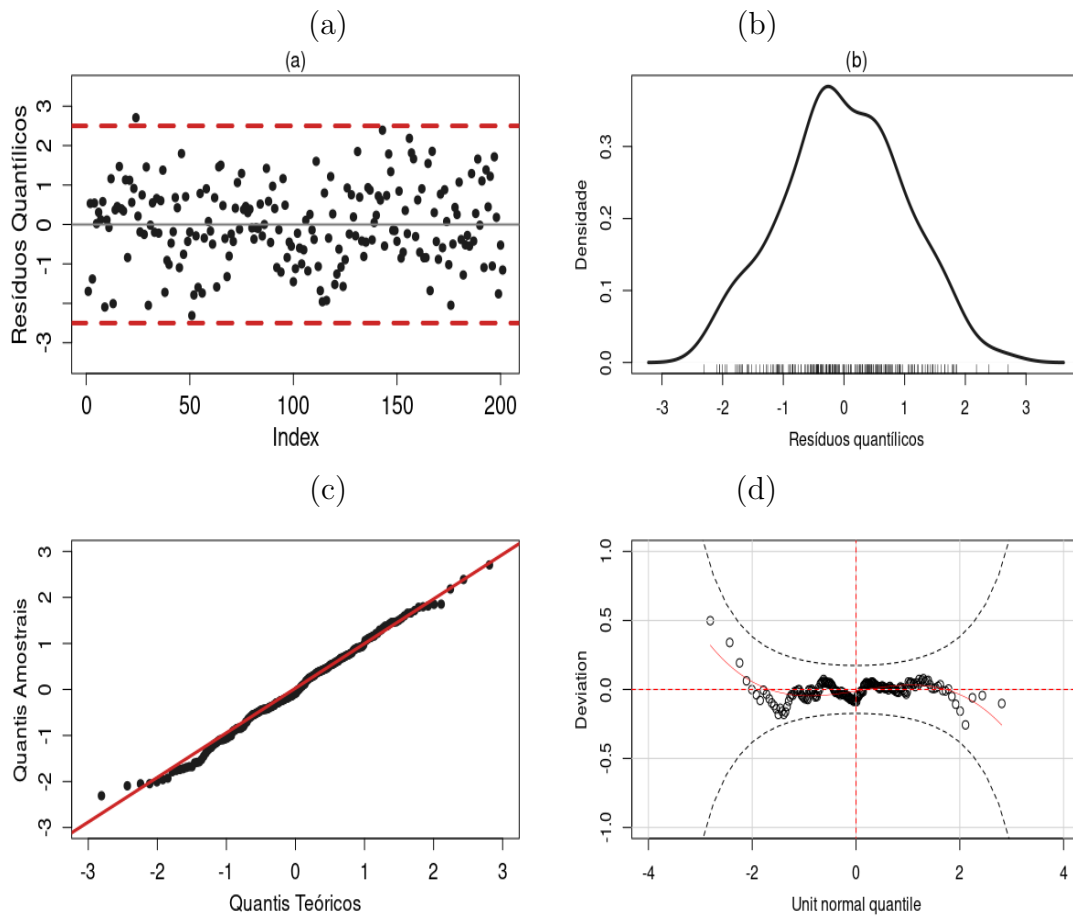


Figura 6.2: Gráficos de diagnóstico para o modelo de regressão BGama em dados de câncer gástrico. (a) Gráfico de índices \hat{q}_i (b) Função de densidade estimada para \hat{q}_i . (c) QQ -plot para \hat{q}_i . (d) $Worm$ plot para \hat{q}_i .

Os gráficos de densidade, QQ -plot e $worm$ plot, para os quantis dos resíduos confirmam que o modelo providência um bom ajuste. Na Figura 6.2 pode-se ver que os resíduos parecem seguir distribuição normal, e que além disso, não possuem pontos discrepantes. Note que resíduos quantílicos confirmam que o modelo se ajusta bem aos dados visto que:

se mantiveram próximos as bandas de confiança em (a); aproximaram-se a uma curva gaussiana em (b); se mantiveram próximos a linha do QQ -plot em (c); e conservaram-se próximos à linha horizontal no *worm plot* em (d).

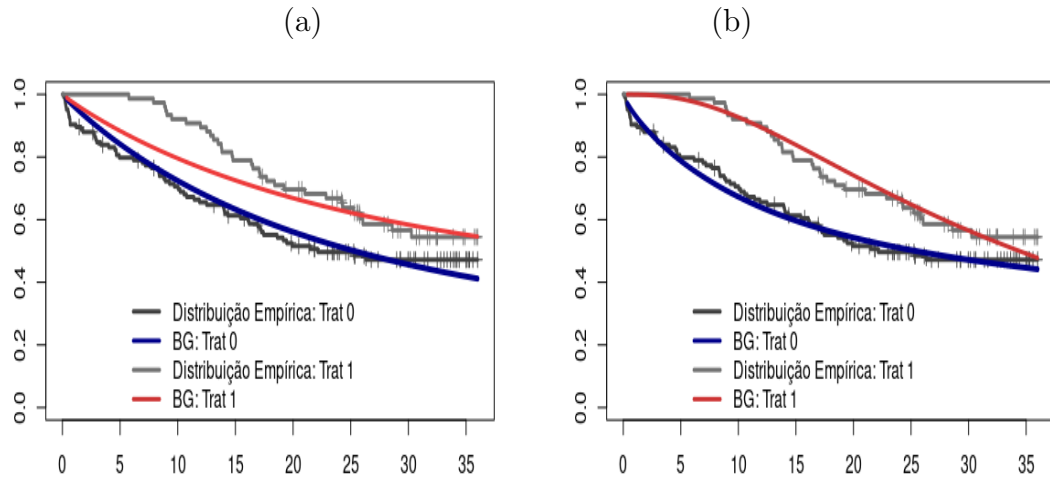


Figura 6.3: Função de Sobrevivência estimada do modelo de regressão BGama(θ_δ) e a função empírica de sobrevivência. (a) Ajuste considerando apenas o parâmetro $\alpha_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)$. (b) Ajuste considerando ambos os parâmetros $\alpha_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_1)$ e $\beta_i = \exp(\mathbf{v}_i^T \boldsymbol{\tau}_2)$, para os dados de câncer gástrico.

Por fim, para avaliar a qualidade do ajuste do Modelo de Regressão Gama bimodal, comparou-se as estimativas do modelo com a distribuição empírica dos dados. Para essa análise foram utilizadas duas variações do modelo: na primeira variação apenas o parâmetro α foi modelado em função das covariáveis; e, no segundo modelo, os dois parâmetros, α e β , foram ajustados em função do tipo de tratamento. Na figura 6.3 nota-se que o segundo modelo descreve os dados de forma satisfatória.

Conclusões

Ao longo deste trabalho foram apresentadas diversas propriedades da distribuição Gama bimodal. Foi matematicamente descrito em que situações a distribuição assume a uni-bimodalidade e em quais intervalos a função cresce e decresce. Além disso, foi demonstrado que esta distribuição possui caudas leves, mas não tão leves como as da distribuição normal.

Foi visto que é possível obter fórmulas fechadas para os momentos e para a função geradora de momentos, facilitando o cálculo da média, desvio padrão, assimetria e curtose. Em relação as medidas de entropia, também, foi possível encontrar fórmulas fechadas para a entropia de Shannon e a entropia Quadrática. Já em relação as medidas de sobrevivência, foi possível encontrar expressões para a função de sobrevivência, função de distribuição acumulada, taxa de risco e vida média residual, o que contribuiu para a caracterização da forma da taxa de risco.

Os Estimadores de Máxima Verossimilhança, por sua vez, precisam ser estimados computacionalmente, uma vez que não existe solução linear para o ponto máximo da função de verossimilhança. Também foi verificado através de simulações computacionais que esses estimadores são consistentes, visto que seu viés e variância diminuem na medida em que o número de amostras aumenta.

Por fim, foi apresentado um modelo de regressão baseado na distribuição Gama bimodal. A partir das medidas de ajuste dos modelos, percebeu-se que o modelo proposto tem melhor desempenho em comparação a outros encontrados na literatura. Logo, possui aplicabilidade e pode trazer ganhos a modelagem de situações reais.

Portanto, a distribuição Gama bimodal mostrou ser uma distribuição bimodal flexível que pode ser útil na modelagem de dados reais. Por ter menos parâmetros e por permitir a obtenção de fórmulas fechadas para grande parte de suas medidas, pode ser uma boa alternativa as misturas. Em trabalhos futuros pode-se buscar novas aplicações para esta distribuição. Também, é possível criar novos modelos baseados na Gama Bimodal.

Bibliografia

- [1] AHMED, S. E.; HUSSEIN, A.: Gamma Mixture: Bimodality, Inflexions, and L-Moments. *Communications in Statistics*, **37**, p. 1147–1161, (2008)
- [2] AKINSETE, A.; FAMOYE, F.; LEE, C.: The beta-Pareto distribution. *Statistics*. **42**, p. 547–563, (2008)
- [3] ALZAATREH, A.: Family of generalized gamma distributions: Properties and applications. *Hacettepe Journal of Mathematics and Statistics*, **45**, (2015)
- [4] BATIR, N.: On some properties of digamma and polygamma functions. *Journal of Mathematical Analysis and Applications*, **328** (1), p. 452–465, (2007)
- [5] CASELLA, G.; BERGER, R. L.: *Inferência Estatística*, 2.ed. São Paulo: Cengage Learning, (2010)
- [6] CHOULAKIAN, V.; STEPHENS, M. A.: Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, **43**, p. 478–484, (2001)
- [7] CRAMER, H.: *Mathematical methods of statistics. Mathematical Inequalities & Applications*, New Jersey: Princeton, (1946)
- [8] ELAL-OLIVERO, D.: Alpha-skew-normal distribution. *Proyecciones Journal of Mathematics*, **29**, p. 224–240, (2010)
- [9] ELEZOVIC, N.; GIORDANO, C.; PECARIC, J.: The best bounds in Gautschi’s inequality. *Mathematical Inequalities & Applications*, **3**, p. 239–252, (2000)
- [10] FAMOYE, F.; LEE, C.; EUGENE, N.: Beta-normal distribution: Bimodality properties and application, *Statistical Methods and Applications*, **3**, p. 85–103, (2004)
- [11] GHOJOGH, B., GHOJOGH, A.; CROWLEY, M.; KARRAY, F.: Fitting A Mixture Distribution to Data: Tutorial. ArXiv preprint. ArXiv:1901.06708, (2019)

- [12] GODFREY, K. R.; DISTEFANO, J. J.: Identifiability of Model Parameter. IFAC Proceedings Volumes, **18** (5), p. 89–114, (1985)
- [13] GÓMEZ, Y. M.; GÓMEZ-DÉNIS, E.; VENEGAS, O.; GALLARDO, D. I.; GÓMEZ, H. W.: An Asymmetric Bimodal Distribution with Application to Quantile Regression. *Symmetry*, **11** (7), 899, (2019)
- [14] GOVIL, K. K.; AGGARWAL, K. K.: Mean residual life function for normal, gamma and lognormal densities. *Reliability Engineering*, **5** (1), p. 47–51, (1983)
- [15] GRIFFITHS, L.: *Introduction to the Theory of Equations*. J. Wiley, (1947)
- [16] GUO, B.; QI, F.: Some properties of the psi and polygamma functions. *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics*, **39**, (2010)
- [17] HASSAN, M.; HIJAZI, R.: A bimodal exponential power distribution. *Pakistan Journal of Statistics*, **26** (2), (2010)
- [18] HSU, Y.; PEARN, W.; WU, P.: Capability adjustment for gamma processes with mean shift consideration in implementing six sigma program. *European Journal of Operational Research*, **119**, p. 517–529, (2008)
- [19] JOHNSON, N. L.; KOTZ, S.; KEMP, A.: *Univariate Discrete Distributions*. New York: Wiley, (1993)
- [20] LAI, C. D.; XIE, M.; MURTHY D. N. P.: A modified Weibull distribution. *IEEE Trans Reliab.* **52**, p. 33–37, (2003)
- [21] MAGALHÃES, M. N.: *Probabilidade e Variáveis Aleatórias*, 3. ed. São Paulo: Editora da Universidade de São Paulo, (2015)
- [22] MARTINEZ, E. Z.; ACHCAR, J. A.; JÁCOME, A. A. A.; SANTOS, J. S.: Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput Methods Programs Biomed.* **112**, p. 343–355, (2013)
- [23] MÜLLER-LENNERT, M.; DUPUIS, F.; SZEHR, O.; FEHR, S.; TOMAMICHEL, M.: On quantum Renyi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, (2013)
- [24] ORTEGA, E. M. M.; CORDEIRO, G. M.; HASHIMOTO, E. M.; SUZUKI, A. K.: Regression models generated by gamma random variables with long-term survivors. *Communications for Statistical Applications and Methods*, **24**, p. 43–65, (2017)

- [25] PASCOA, M. A. R.; ORTEGA E. M. M.; CORDEIRO, G. M.: The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, **8**, p. 411–433, (2011)
- [26] RAO, K. S.: A bimodal distribution. *Bulletin of Calcutta Mathematical Society*, **80**, p. 238–240, (1988)
- [27] ROSS, S.: *Probabilidade: um curso moderno com aplicações*. 8.ed. Porto Alegre: Bookman, (2010)
- [28] SHANKAR, P. M.; FORSBERG, F.; LOWN, L.: Statistical modeling of atherosclerotic plaque in carotid B mode images-a feasibility study. *Ultrasound in Medicine and Biology*, **29**, p. 1305–1309, (2003)
- [29] STASINOPOULOS, D. M.; RIGBY, R. A.: Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, **23** (7), (2007)
- [30] VILA, R.; FERREIRA, L.; SAULO, H.; PRATAVIERA, F.; ORTEGA, E. M. M.: A Bimodal Gamma Distribution: Properties, Regression Model and Applications. *Statistics*, **54**, Issue 3, p. 469–493, (2020)
- [31] VINBERG, É.: *A Course in Algebra*. Graduate studies in mathematics. American Mathematical Society, (2003)
- [32] WANG, J.; Taaffe, M.: Multivariate Mixtures of Normal Distributions: Properties, Random Vector Generation, Fitting, and as Models of Market Daily Changes. *Inform Journal on Computing*, Monterey, CA, p. 193–203, (2015)
- [33] WEISBERG, S.: *Applied Linear Regression*. 1. ed. New York: John Wiley & Sons, (2005)
- [34] XUE, J.: *Loop Tiling for Parallelism*. The Springer International Series in Engineering and Computer Science. Springer US, (2012)

Apêndice

```
#####  
## Bimodal Gamma Distribution #  
#####  
  
#####  
## Packages  
#####  
  
rm(list=ls(all=TRUE))  
  
require(gbs)  
require(datasets)  
require(MASS)  
require(numDeriv)  
require(fitdistrplus)  
require(GoFKernel)  
  
#####  
## CDFs  
#####  
  
## bimodal-Gamma distribution: proposal  
pbimodalGamma <- function(x,  
                           alpha = 1.0,  
                           beta = 1.0,  
                           delta = 1.0){
```

```

A1 <- delta*(2-delta*(x+(alpha+1)/beta))*beta^(alpha-1)*x^(alpha) *
exp(-beta*x)
A2 <- (2+delta^2 * alpha*(alpha+1)/(beta^2) - 2 * delta * alpha / beta)
* pgamma(x,alpha,beta) * gamma(alpha)
A3 <- (2 + (alpha*delta)/beta * ((1+alpha)*delta/beta -
2))*gamma(alpha)

cdf <- (A1+A2)/A3

return(cdf)
}

#####
## Random number generation
#####

rbimodalGamma <- function(n, ALPHA = 0.5, BETA = 1.0, DELTA = 1.0){

  cdfbimodalGamma <- function(x,
                                alpha = ALPHA,
                                beta = BETA,
                                delta = DELTA){

    A1 <- delta * (2-delta*(x+(alpha+1)/beta)) * beta^(alpha-1) *
x^(alpha) * exp(-beta*x)
    A2 <- (2 + delta^2 * alpha * (alpha+1)/(beta^2) - 2
*delta*alpha/beta) * pgamma(x,alpha,beta) * gamma(alpha)
    A3 <- (2 + (alpha*delta)/beta* ((1+alpha)*delta/beta -
2))*gamma(alpha)
    cdf <- (A1+A2)/A3
    return(cdf)
  }

  X <- numeric(n)
  j <- 0
  while(j < n){

```

```

j <- j+1

f.inv <- inverse(cdfbimodalGamma ,lower=0.000001,upper=Inf)

X[j] <- f.inv(runif(1))

}

return(X)

}

#rbimodalGamma(n=20, ALPHA = 0.5, BETA = 1.0, DELTA = 1.0)

#####
## PDFs
#####

## bimodal Gamma distribution:

dbimodalGamma <- function(x,
                           alpha      = 1.0,
                           beta       = 1.0,
                           delta      = 1.0,
                           log        = FALSE){

  b1 <- 1
  b2 <- ( 1 + (1 - delta * x )^2 ) / ( 2 +(alpha*delta/beta)*
    ( (1+alpha)*(delta/beta) -2) )
  b3 <- (beta^alpha / gamma(alpha))*x^(alpha-1)*exp(-beta*x)
  a1 <- b1*b2 * b3

  pdf <- a1
  if(log == TRUE){
    pdf <- log(pdf)
  }
}

```

```

}
return(pdf)
}

```

```

#####
## Hazard rate
#####

```

```

hrbimodalGamma <- function(x,
                             alpha = 1.0,
                             beta = 1.0,
                             delta = 1.0,
                             log = FALSE){

  b1 <- 1
  b2 <- ( 1 + (1 - delta * x )^2 ) / ( 2 +(alpha*delta/beta)*
    ( (1+alpha)*(delta/beta) -2) )
  b3 <- (beta^alpha / gamma(alpha))*x^(alpha-1)*exp(-beta*x)
  a1 <- b1*b2 * b3

  pdf <- a1

  A1 <- delta*(2-delta*(x+(alpha+1)/beta))*beta^(alpha-1)*x^(alpha) *
    exp(-beta*x)
  A2 <- (2+delta^2 * alpha*(alpha+1)/(beta^2) - 2
    *delta*alpha/beta)*pgamma(x,alpha,beta)*gamma(alpha)
  A3 <- (2 + (alpha*delta)/beta* ((1+alpha)*delta/beta - 2))*gamma(alpha)

  cdf <- (A1+A2)/A3

  hr <- pdf/(1-cdf)

  return(hr)
}

```

```
}

```

```
#####
## Log-likelihood - Bimoda Gamma distribution
#####

```

```
loglik_dbimodalGamma <- function(theta, x, delta){

```

```
  value <- sum(dbimodalGamma(x,
                             alpha = theta[1],
                             beta  = theta[2],
                             delta  = delta,
                             log    = TRUE))

```

```
  return(-value)
}

```

```
loglik_dbimodalGammaTotal <- function(theta, x, delta){

```

```
  value <- sum(dbimodalGamma(x,
                             alpha = theta[1],
                             beta  = theta[2],
                             delta  = theta[3],
                             log    = TRUE))

```

```
  return(-value)
}

```

```
loglik_dbimodalGamma_SE <- function(par, x){

```

```
  value <- sum(dbimodalGamma(x,
                             alpha = par[1],

```



```

        beta  = par[2],
        delta = par[3],
        log   = TRUE))

    return(value)
}

loglik_dbimodalGamma_SE_HE <- function(par, t, delta){

    value <- sum(dbimodalGamma(x=t,
                               alpha = par[1],
                               beta  = par[2],
                               log   = TRUE))

    return(value)
}

#####
## MLE estimation - Bimodal Gamma - choose delta - Proposal
#####

mle_dbimodalGamma <- function(x){

    deltas      <- seq(-50,50,by=1)
    alphas      <- numeric()
    betas       <- numeric()
    lambdas     <- numeric()
    logLikelihood <- numeric()

    obj <- function(lambda, xbar, logx.bar) {
        digamma(lambda * xbar) - logx.bar - log(lambda)
    }

    xbar <- mean(x)
    logx.bar <- mean(log(x))
    u <- uniroot(obj, lower = 0.00001, upper =10e5, xbar = xbar,
                logx.bar = logx.bar)
    lambda.hat <- u$root
    alfa.hat   <- xbar * lambda.hat

```

```

#f1 <- fitdist(x, "gamma")

for(i in 1:length(deltas)){

  estimation <- optim(par=c(alfa.hat,lambda.hat),
    fn=loglik_dbimodalGamma,x=x,method="BFGS",delta=deltas[i])
  a <- estimation$par[1]
  b <- estimation$par[2]
  alphas[i] <- a
  betas[i] <- b
  deltas[i] <- deltas[i]
  parameters <- c(a, b, deltas[i])
  logLikelihood[i] <- loglik_dbimodalGamma_SE(par = parameters, x = x)
}

maximum <- max(logLikelihood)
best <- which(logLikelihood == maximum)

results <- list(alphaEstimate = alphas[best],
  betaEstimate = betas[best],
  deltaEstimate = deltas[best],
  logLikelihood = logLikelihood[best])
return(results)
}

mle_dbimodalGamma_fixed_delta <- function(x,delta){

  obj <- function(lambda, xbar, logx.bar) {
    digamma(lambda * xbar) - logx.bar - log(lambda)
  }

  xbar <- mean(x)

```

```

logx.bar <- mean(log(x))
u <- uniroot(obj, lower = 0.00001, upper =10e5, xbar = xbar,
logx.bar = logx.bar)
lambda.hat <- u$root
alfa.hat <- xbar * lambda.hat

#f1 <- fitdist(x, "gamma")

estimation <- optim(par=c(alfa.hat,lambda.hat),
fn=loglik_dbimodalGamma,x=x,method="BFGS",delta=delta,hessian=TRUE)

SHess      = solve(estimation$hessian)
SE         = sqrt(diag(SHess))
se.coef   = SE
tval      = estimation$par/se.coef
matcoef   = cbind(c(estimation$par), se.coef, tval, 2*(1-pnorm(abs(tval))))

results <- list( estimates = estimation$par,
                SEs = SE
                )
return(results)
}

#####
## Monte Carlo Study
#####

## Global variables

NREP <- 1000 # Monte Carlo trials
N <- 50 # sample size

beta <- 1.0
alpha <- 0.5
DELTA <- seq(-5,5,by=0.1)
deltapar <- c(-10,-5,1,5,10)

```

```

alpha_MLE <- matrix(NA,NREP,length(deltapar))
beta_MLE <- matrix(NA,NREP,length(deltapar))
delta_MLE <- matrix(NA,NREP,length(deltapar))

## Sample loop

for(j in 1:length(deltapar)){

  delta <- deltapar[j]

  #### Monte Carlo loop
  set.seed(201410)

  for(i in 1:NREP) {

    dados <- rbimodalGamma(n=N, ALPHA = alpha, BETA =beta, DELTA =
    delta)

    estimates <- mle_dbimodalGamma_fixed_delta(x=dados, delta = delta )

    alpha_MLE[i,j] <- estimates$estimates[1]
    beta_MLE[i,j] <- estimates$estimates[2]

  }

}

## MLE where random numbers from BG distribution

betahat_MLE <- colMeans(beta_MLE); betahat_MLE
alphahat_MLE <- colMeans(alpha_MLE); alphahat_MLE

bias_beta_MLE <- colMeans(beta_MLE) - beta; bias_beta_MLE
bias_alpha_MLE <- colMeans(alpha_MLE) - alpha; bias_alpha_MLE

```

```

var_beta_MLE    <- colVars(beta_MLE)
var_alpha_MLE   <- colVars(alpha_MLE)

mse_beta_MLE    <- var_beta_MLE + (bias_beta_MLE)^2
mse_alpha_MLE   <- var_alpha_MLE + (bias_alpha_MLE)^2

#####
## Graphs
#####

## Density
xx <- seq(0.0,20,by=0.01)

par(mar=c(3.9,4.2,1.4,0.5),mgp=c(2.4, 1, 0))
plot(xx, dbimodalGamma(xx,alpha = 1.5, beta = 0.7, delta =
0.6),type="l",col=1,lty=1,lwd=1.3,xlim=c(0.1,20),ylim=c(0,0.25),xlab="x",
ylab="FDP",main=' ', font.main = 1)
lines(xx, dbimodalGamma(xx,alpha = 1.5, beta = 0.7, delta =
0.7),type="l",col=1,lty=2,lwd=1.3)
lines(xx, dbimodalGamma(xx,alpha = 1.5, beta = 0.7, delta =
0.8),type="l",col=1,lty=3,lwd=1.3)
lines(xx, dbimodalGamma(xx,alpha = 1.5, beta = 0.7, delta =
1),type="l",col=1,lty=4,lwd=1.3)
lines(xx, dbimodalGamma(xx,alpha = 1.5, beta = 0.7, delta =
2),type="l",col=1,lty=5,lwd=1.3)
legend(13, 0.17, c(expression(alpha==1.5 ~ ", "
~delta==0.6),expression(alpha==1.5 ~ ", "
~delta==0.7),expression(alpha==1.5 ~ ", "
~delta==0.8),expression(alpha==1.5 ~ ", " ~delta==1),
expression(alpha==1.5 ~ ", " ~delta==2)) ,
      lty = c(1, 2, 3, 4, 5), col= c (1, 1, 1, 1, 1),lwd= c (1.3, 1.3,
1.3, 1.3, 1.3), box.col = 0)

xx <- seq(0.0,15,by=0.01)

par(mar=c(3.9,4.2,1.4,0.5),mgp=c(2.4, 1, 0))
plot(xx, dbimodalGamma(xx,alpha = 1.1, beta = 0.7, delta=0.75)

```

```

,type="l",col=1,lty=1,lwd=1.3,xlim=c(0.1,15),ylim=c(0,0.45),xlab="x",
ylab="FDP",main=' ', font.main = 1)
lines(xx, dbimodalGamma(xx,alpha = 1.25, beta = 0.7, delta =0.5)
,type="l",col=1,lty=2,lwd=1.3)
lines(xx, dbimodalGamma(xx,alpha = 1.4, beta = 0.7, delta =1)
,type="l",col=1,lty=3,lwd=1.3)
lines(xx, dbimodalGamma(xx,alpha = 1.5, beta = 0.7, delta =1.5)
,type="l",col=1,lty=4,lwd=1.3)
lines(xx, dbimodalGamma(xx,alpha = 1.3, beta = 0.7, delta =0.8)
,type="l",col=1,lty=3,lwd=1.3)
legend(4.9, 0.43, c(expression(alpha==1.1 ~ "," ~ delta==0.75)
,expression(alpha==1.25 ~ "," ~ delta==0.5),expression(alpha==1.4 ~ ","
~ delta==1),
                expression(alpha==1.5 ~ "," ~
                delta==1.5),expression(alpha==1.3 ~ "," ~
                delta==0.8)) ,
#
lty = c(1, 2, 3, 4, 5), col= c (1, 1, 1, 1, 1),lwd= c (1.3, 1.3,
1.3, 1.3, 1.3), box.col = 0)

## Hazard Rate

xx <- seq(0.01,25,by=0.01)

par(mar=c(3.9,4.2,1.4,0.5),mgp=c(2.4, 1, 0))
plot(xx, hrbimodalGamma(xx,alpha = 1.5, beta = 0.7 , delta = 0.5)
,type="l",col=1,lty=1,lwd=1.3,xlim=c(0,20),ylim=c(0,0.55),xlab="x",ylab="
H(x)",main=' ', font.main = 1)
lines(xx, hrbimodalGamma(xx,alpha = 1.5, beta = 0.7 , delta = 0.7)
,type="l",col=1,lty=2,lwd=1.3)
lines(xx, hrbimodalGamma(xx,alpha = 1.5, beta = 0.7 , delta = 1.0)
,type="l",col=1,lty=3,lwd=1.3)
lines(xx, hrbimodalGamma(xx,alpha = 1.5, beta = 0.7 , delta = 1.5)
,type="l",col=1,lty=4,lwd=1.3)
lines(xx, hrbimodalGamma(xx,alpha = 1.5, beta = 0.7 , delta = 2.0)
,type="l",col=1,lty=5,lwd=1.3)
legend(12, 0.4, c(expression(paste(alpha==1.5," ",",",delta==0.5)),
                expression(paste(alpha==1.5," ",",",delta==0.7)),

```

```

        expression(paste(alpha==1.5," ",delta==1.0)),
        expression(paste(alpha==1.5," ",delta==1.5)),
        expression(paste(alpha==1.5," ",delta==2.0))),
    lty = c(1, 2, 3, 4, 5), col= c (1, 1, 1, 1, 1),lwd= c (1.3, 1.3,
1.3, 1.3, 1.3), box.col = 0)

xx <- seq(0.01,25,by=0.01)

par(mar=c(3.9,4.2,1.4,0.5),mgp=c(2.4, 1, 0))
plot(xx, hrbimodalGamma(xx,alpha = 0.50, beta = 0.7 , delta =
-2.5),type="l",col=1,lty=1,lwd=1.3,xlim=c(0,25),ylim=c(0,0.60),xlab="x",y
lab="H(x)",main='(b)', font.main = 1)
lines(xx, hrbimodalGamma(xx,alpha = 0.75, beta = 0.7 , delta =
-2.0),type="l",col=1,lty=2,lwd=1.3)
lines(xx, hrbimodalGamma(xx,alpha = 1.25, beta = 0.7 , delta =
0.5),type="l",col=1,lty=3,lwd=1.3)
lines(xx, hrbimodalGamma(xx,alpha = 1.50, beta = 0.7 , delta =
1.0),type="l",col=1,lty=4,lwd=1.3)
lines(xx, hrbimodalGamma(xx,alpha = 1.75, beta = 0.7 , delta =
0.7),type="l",col=1,lty=5,lwd=1.3)
legend(15, 0.4, c(expression(paste(alpha==0.50," ",delta==--2.5)),
expression(paste(alpha==0.75,"
",delta==--2.0)),
expression(paste(alpha==1.25," ",delta==0.5)),
expression(paste(alpha==1.50," ",delta==1.0)),
expression(paste(alpha==1.75,"
",delta==0.7))),
lty = c(1, 2, 3, 4, 5), col= c (1, 1, 1, 1, 1),lwd= c (1.3, 1.3,
1.3, 1.3, 1.3), box.col = 0)

#####
## BGama Regression Model
#####

#####
## Auxiliary codes

```

```
#####

require(numDeriv)
require(flexsurv)
require(gamlss)
BGA <- function (mu.link = "log", sigma.link="log", nu.link = "identity")
{
  mstats <- checklink(  "mu.link", "Bimodal Gamma", substitute(mu.link),
                        c("1/mu^2","sqrt", "log", "identity","own"))
  dstats <- checklink("sigma.link", "Bimodal Gamma", substitute(sigma.link),
                      c("sqrt","log", "identity", "own"))
  vstats <- checklink(  "nu.link", "Bimodal Gamma", substitute(nu.link),
                        c("sqrt", "log", "identity", "own"))

  structure(
    list(family = c("BGA", "Bimodal Gamma"),
         parameters = list(mu=TRUE, sigma=TRUE, nu=TRUE),
         nopar = 3,
         type = "Continuous",
         mu.link = as.character(substitute(mu.link)),
         sigma.link = as.character(substitute(sigma.link)),
         nu.link = as.character(substitute(nu.link)),
         mu.linkfun = mstats$linkfun,
         sigma.linkfun = dstats$linkfun,
         nu.linkfun = vstats$linkfun,
         mu.linkinv = mstats$linkinv,
         sigma.linkinv = dstats$linkinv,
         nu.linkinv = vstats$linkinv,
         mu.dr = mstats$mu.eta,
         sigma.dr = dstats$mu.eta,
         nu.dr = vstats$mu.eta,

         dldm = function(y,mu,sigma,nu){ #----- ok
           lpdf<-function(t,x,sigma,nu){log(dauxiBGA(t,x,sigma,nu))}
           dldm<-grad(func=lpdf,t=y,x=mu,sigma=sigma,nu=nu,method='simple')
           dldm
         },
         d2ldm2 = function(y,mu,sigma,nu){#----- ok
           lpdf<-function(t,x,sigma,nu){log(dauxiBGA(t,x,sigma,nu))}
           dldm<-grad(func=lpdf,t=y,x=mu,sigma=sigma,nu=nu,method='simple')

```



```

d2ldm2 <- -dldm * dldm
d2ldm2 <- ifelse(d2ldm2 < -1e-15, d2ldm2,-1e-15)
d2ldm2
},
dldd = function(y,mu,sigma,nu){#-----ok
  lpdf<-function(t,mu,x,nu){log(dauxiBGA(t,mu,x,nu))}
  dldd<-grad(func=lpdf,t=y,mu=mu,x=sigma,nu=nu,method='simple')
  dldd
} ,
d2ldd2 = function(y,mu,sigma,nu){#-----ok
  lpdf<-function(t,mu,x,nu){log(dauxiBGA(t,mu,x,nu))}
  dldd<-grad(func=lpdf,t=y,mu=mu,x=sigma,nu=nu,method='simple')
  d2ldd2 <- -dldd*dldd
  d2ldd2 <- ifelse(d2ldd2 < -1e-15, d2ldd2,-1e-15)
  d2ldd2
},
dldv = function(y,mu,sigma,nu){#----- ok
  lpdf<-function(t,mu,sigma,x){log(dauxiBGA(t,mu,sigma,x))}
  dldv<-grad(func=lpdf,t=y,mu=mu,sigma=sigma,x=nu,method='simple')
  dldv
},
d2ldv2 = function(y,mu,sigma,nu){#----- ok
  lpdf<-function(t,mu,sigma,x){log(dauxiBGA(t,mu,sigma,x))}
  dldv<-grad(func=lpdf,t=y,mu=mu,sigma=sigma,x=nu,method='simple')
  d2ldv2<- -dldv * dldv
  d2ldv2 <- ifelse(d2ldv2 < -1e-15, d2ldv2,-1e-15)
  d2ldv2
},
d2ldmdd = function(y,mu,sigma,nu){#----- ok
  lpdf<-function(t,x,sigma,nu){log(dauxiBGA(t,x,sigma,nu))}
  dldm<-grad(func=lpdf,t=y,x=mu,sigma=sigma,nu=nu,method='simple')
  lpdf<-function(t,mu,x,nu){log(dauxiBGA(t,mu,x,nu))}
  dldd<-grad(func=lpdf,t=y,mu=mu,x=sigma,nu=nu)
  d2ldmdd = -(dldm * dldd)
  d2ldmdd<-ifelse(is.na(d2ldmdd)==TRUE,0,d2ldmdd)
  d2ldmdd
},
d2ldmdv = function(y,mu,sigma,nu){#----- ok
  lpdf<-function(t,x,sigma,nu){log(dauxiBGA(t,x,sigma,nu))}

```

```

dldm<-grad(func=lpdf,t=y,x=mu,sigma=sigma,nu=nu,method='simple')
lpdf<-function(t,mu,sigma,x){log(dauxiBGA(t,mu,sigma,x))}
dldv<-grad(func=lpdf,t=y,mu=mu,sigma=sigma,x=nu)
d2ldmdv = -(dldm * dldv)
d2ldmdv
},
d2ldddv = function(y,mu,sigma,nu){#----- ok
  lpdf<-function(t,mu,x,nu){log(dauxiBGA(t,mu,x,nu))}
  dlld<-grad(func=lpdf,t=y,mu=mu,x=sigma,nu=nu,method='simple')
  lpdf<-function(t,mu,sigma,x){log(dauxiBGA(t,mu,sigma,x))}
  dldv<-grad(func=lpdf,t=y,mu=mu,sigma=sigma,x=nu)
  d2ldddv = -(dlld * dldv)
  d2ldddv
},
#----- ok
G.dev.incr = function(y,mu,sigma,nu,...)
{
  -2*dBGA(y,mu,sigma,nu,log=TRUE)
} ,
rqres = expression(rqres(pfun="pBGA", type="Continuous", y=y,
mu=mu, sigma=sigma, nu=nu)) ,
#mu.initial = expression(mu <- (y+mean(y))/2),
mu.initial = expression(mu <- (y+mean(y))/2),
#mu.initial = expression(mu <- rep(mean(y), length(y))),
#mu.initial = expression(mu <- rep(sd(y)*2.2+sqrt(pi),
length(y))),
sigma.initial = expression(sigma <- rep(1, length(y))), #OK
nu.initial = expression(nu <- rep(1.45,length(y))), #)k
mu.valid = function(mu) all(mu > 0),
sigma.valid = function(sigma) all(sigma > 0),
nu.valid = function(nu) TRUE,
y.valid = function(y) all(y > 0)
),
class = c("gamlss.family","family"))
}
#-----
dBGA <- function(x, mu = 2, sigma = 2, nu = 0.5, log = FALSE){
  if (any(mu < 0)) stop(paste("mu must be positive", "\n", ""))
  if (any(sigma < 0)) stop(paste("sigma must be positive", "\n", ""))

```

```

zd1 <- 2 + ((mu*nu^2)/sigma^2) + (((mu^2)*(nu^2))/sigma^2) -
((2*mu*nu)/sigma)
zd2 <- 1 + (1-nu*x)^2
f <- (zd2/zd1)*dgamma(x, shape=mu, rate =sigma)

if(log==FALSE) fy<-f else fy<-log(f)
fy
}
#-----
pBGA <- function(q, mu = 2, sigma = 2, nu = 0.5, lower.tail = TRUE,
log.p = FALSE){
  if (any(mu < 0)) stop(paste("mu must be positive", "\n", ""))
  if (any(sigma < 0)) stop(paste("sigma must be positive", "\n", ""))
  delta <- nu
  alpha <- mu
  beta <- sigma
  A1 <- delta*(2-delta*(q+(alpha+1)/beta))*beta^(alpha-1)*q^(alpha) *
exp(-beta*q)
  A2 <- (2+delta^2 * alpha*(alpha+1)/(beta^2) - 2
*delta*alpha/beta)*pgamma(q,alpha,beta)*gamma(alpha)
  A3 <- (2 + (alpha*delta)/beta* ((1+alpha)*delta/beta - 2))*gamma(alpha)

  FF1 <- (A1+A2)/A3

  if(lower.tail==TRUE) cdf<-FF1 else cdf<- 1-FF1
  if(log.p==FALSE) cdf<- cdf else cdf<- log(cdf)
  cdf
}
#-----
hBGA<-function(x, mu = 2, sigma = 2, nu = 0.5){
  delta <- nu
  alpha <- mu
  beta <- sigma
  b1 <- 1
  b2 <- ( 1 + (1 - delta * x )^2 ) / ( 2 +(alpha*delta/beta)*(
(1+alpha)*(delta/beta) -2) )
  b3 <- (beta^alpha / gamma(alpha))*x^(alpha-1)*exp(-beta*x)
  a1 <- b1*b2 * b3

```

```

pdf <- a1

A1 <- delta*(2-delta*(x+(alpha+1)/beta))*beta^(alpha-1)*x^(alpha) *
exp(-beta*x)
A2 <- (2+delta^2 * alpha*(alpha+1)/(beta^2) - 2
*delta*alpha/beta)*pgamma(x,alpha,beta)*gamma(alpha)
A3 <- (2 + (alpha*delta)/beta* ((1+alpha)*delta/beta - 2))*gamma(alpha)

cdf <- (A1+A2)/A3

hr <- pdf/(1-cdf)
}

#####
dauxiBGA <- function(t,mu,sigma,nu){
  zd1 <- 2 + ((mu*nu^2)/sigma^2) + (((mu^2)*(nu^2))/sigma^2) -
  ((2*mu*nu)/sigma)
  zd2 <- 1 + (1-nu*t)^2
  f <- (zd2/zd1)*dgamma(t, shape=mu, rate =sigma)
  f }

#####
## Aplication 1: Wheaton River
#####
##The Wheaton River data Exceedances of Wheaton River flood data.
data1<-c(1.7,2.20,14.4, 1.1, 0.4, 20.6, 5.3, 0.7, 1.9, 13.0, 12.0, 9.3,
  1.4,18.7,8.50, 25.5, 11.6, 14.1,22.1,1.1,2.5,14.4,1.7,37.6,
  0.6, 2.2 ,39.0, 0.3 ,15.0 ,11.0, 7.3, 22.9, 1.7, 0.1 ,1.1 ,0.6,
  9.0, 1.7, 7.0, 20.1 ,0.4 ,2.8 ,14.1 ,9.9 ,10.4 ,10.7 ,30.0, 3.6,
  5.6, 30.8, 13.3, 4.2 ,25.5 ,3.4 ,11.9 ,21.5 ,27.6 ,36.4, 2.7, 64.0,
  1.5, 2.5, 27.4, 1.0, 27.1, 20.2, 16.8, 5.3, 9.7, 27.5, 2.5, 27.0)

# mu= alpha
#sigma = beta

```

```

#nu = delta
#Todos os parâmetros com ligação identidade
#fit1 <- gamlss(data1~1, family = BGA(mu.link = "identity",sigma.link =
"identity"),n.cyc=200)

#plot(fit1)
#wp(fit1)
#summary(fit1)

#model with alpha and beta
fit <- gamlss(data1~1, family = BGA,n.cyc=200)

plot(fit)
wp(fit)
summary(fit)

#Cramer Von Mises:
n<-length(data1)
x<-1:n # para calcular o CVM

(cvm<-1/(12*n)+sum(((2*x-1)/(2*n)-pBGA(sort(data1), mu=1.054,
sigma=0.176,nu=0.177))^2))

#KS
ks.test(data1,"pBGA",mu=1.054, sigma=0.176,nu=0.177)

#####
# Estimates of Kum-GG,
# gamma, exponential Weibull, Weibull and modified Weibull

library("AdequacyModel")
require(GenSA)

#####
#kumGamma

```

```
#####
fit.sa5<- function(data,density) {
  minusllike<-function(x)
  -sum(log(density(data,x[1],x[2],x[3],x[4],x[5]))) #minus the loglik
  lower <- c(0.011,0.011,0.011,0.011,0.011) #may need some changes here
  upper <- c(1000,1000,1000,1000,1000)
  out <- GenSA(lower = lower, upper = upper, fn = minusllike,
  control=list(verbose=TRUE,maxit=7000,max.time=3))
  return(out[c("value","par","counts")])
}
dkum.GG <- function(x,alpha,tau,k,lambda,phi){
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(f)
}
fit.sa5(data1,dkum.GG)
#Densidade
dkum.GG<- function(par,x){
  k=par[3]
  phi=par[5]
  alpha=par[1]
  lambda=par[4]
  tau=par[2]
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(f)
}
#Acumulada
pkum.GG <- function(par,x){
  k=par[3]
  phi=par[5]
  alpha=par[1]
  lambda=par[4]
  tau=par[2]
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
```

```

GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
F <- 1-((1-GG1^lambda)^phi)
return(F)
}

gkum.GG <- goodness.fit(pdf=dkum.GG, cdf=pkum.GG,
                        starts = c(548.54266113, 0.10309103, 0.09875128
                                   ,158.57014699,
                                   869.87847130 ), data = data1,method="BFGS",
                        domain=c(0,Inf),mle=NULL);gkum.GG

#####
#Gamma
#####
fit.sa2<- function(data,density) {
  minusllike<-function(x) -sum(log(density(data,x[1],x[2]))) #minus the loglik
  lower <- c(0.011,0.011) #may need some changes here
  upper <- c(1000,1000)
  out <- GenSA(lower = lower, upper = upper, fn = minusllike,
               control=list(verbose=TRUE,maxit=7000,max.time=3))
  return(out[c("value","par","counts")])
}

dG <- function(x,alpha,k){
  phi=1
  lambda=1
  g1 <- dgengamma.orig(x, shape=1,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=1,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(f)
}

fit.sa2(data1,dG)
#Densidade
dG <- function(par,x){
  k=par[2]
  phi=1
  alpha=par[1]

```

```

lambda=1
tau=1
g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
F <- 1-((1-GG1^lambda)^phi)
return(f)
}
#Acumulada
pG <- function(par,x){
  k=par[2]
  phi=1
  alpha=par[1]
  lambda=1
  tau=1
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(F)
}

gG <- goodness.fit(pdf=dG, cdf=pG,
                   starts = c(14.5587845, 0.8382682), data = data1
                   ,method="BFGS", domain=c(0,Inf),mle=NULL);gG

#####
#Weibull
#####
fit.sa2<- function(data,density) {
  minusllike<-function(x) -sum(log(density(data,x[1],x[2])))
  lower <- c(0.011,0.011) #may need some changes here
  upper <- c(1000,1000)
  out <- GenSA(lower = lower, upper = upper, fn = minusllike,
              control=list(verbose=TRUE,maxit=7000,max.time=3))
  return(out[c("value","par","counts")])
}
dW <- function(x,alpha,tau){
  k=1

```



```

phi=1
lambda=1
g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
F <- 1-((1-GG1^lambda)^phi)
return(f)
}
fit.sa2(data1,dW)
#Densidade
dW <- function(par,x){
  k=1
  phi=1
  alpha=par[1]
  lambda=1
  tau=par[2]
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(f)
}
#Acumulada
pW <- function(par,x){
  k=1
  phi=1
  alpha=par[1]
  lambda=1
  tau=par[2]
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(F)
}

gW <- goodness.fit(pdf=dW, cdf=pW,
                   starts = c(11.6321918, 0.9011661), data =
                   data1,method="BFGS", domain=c(0,Inf),mle=NULL);gW

```

```
#####
#Exp Weibull
#####
fit.sa3<- function(data,density) {
  minusllike<-function(x) -sum(log(density(data,x[1],x[2],x[3]))) #minus
  the loglik
  lower <- c(0.011,0.011,0.011) #may need some changes here
  upper <- c(1000,1000,1000)
  out <- GenSA(lower = lower, upper = upper, fn = minusllike,
  control=list(verbose=TRUE,maxit=7000,max.time=3))
  return(out[c("value","par","counts")])
}
d.EW <- function(x,alpha,tau,lambda){
  phi=1
  k=1
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(f)
}
fit.sa3(data1,d.EW)
#Densidade
dEW<- function(par,x){
  k=1
  phi=1
  alpha=par[1]
  lambda=par[3]
  tau=par[2]
  g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
  GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
  f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
  F <- 1-((1-GG1^lambda)^phi)
  return(f)
}
#Acumulada
pEW <- function(par,x){
  k=1
```

```

phi=1
alpha=par[1]
lambda=par[3]
tau=par[2]
g1 <- dgengamma.orig(x, shape=tau,scale=alpha,k)
GG1 <- pgengamma.orig(x,shape=tau,scale=alpha,k)
f <- phi*lambda*g1*(GG1^(lambda-1))*(1-GG1^(lambda))^(phi-1)
F <- 1-((1-GG1^lambda)^phi)
return(F)
}

```

```

gEW <- goodness.fit(pdf=dEW, cdf=pEW,
                    starts = c(10.0000000, 0.8315737, 1.1283636),
                    data = data1,method="BFGS",
                    domain=c(0,Inf),mle=NULL);gEW

```

```
#####
```

```
#Modified Weibull
```

```
#####
```

```

fit.sa3<- function(data,density) {
  minusllike<-function(x) -sum(log(density(data,x[1],x[2],x[3]))) #minus the loglik
  lower <- c(0.000001,0.000001,0.000001) #may need some changes here
  upper <- c(10,8,10)
  out <- GenSA(lower = lower, upper = upper, fn = minusllike,
              control=list(verbose=TRUE,max.time=2))
  return(out[c("value","par","counts")])
}

```

```

dMW <- function(x,a,b,c){
  f <- a*(x^(b-1))*(b+c*x)*exp((c*x)-a*(x^b)*exp(x*c))
  return(f)
}

```

```
fit.sa3(data1,dMW)
```

```

dMW <- function(par,x){
  a<-par[1]
  b<-par[2]
  c<-par[3]
  f <- a*(x^(b-1))*(b+c*x)*exp(c*x-a*(x^b)*exp(x*c))
}

```

```

    return(f)
}
pMW <- function(par,x){
  a<-par[1]
  b<-par[2]
  c<-par[3]
  FF <- 1-exp(-a*(x^b)*exp(c*x))
  return(FF)
}

gMW<-goodness.fit(pdf=dMW, cdf=pMW,
                  starts = c(0.1241774 ,0.7757742, 0.0100739), data =
                  data1,method="BFGS",
                  domain=c(0.000001,Inf),mle=NULL);gMW

#chutes<-fit.sa3(data1,dMW)$par
#chutes<-c(0.0000326729,0.1889702,0.410336)
#require(MASS)
#fit.MW<- fitdistr(data1,dMW,start=list(a=chutes[1],b=chutes[2],c=chutes[3]),
#                  control=list(ndeps=c(1e-5,1e-5,1e-5),maxit=100))
#fit.MW

#####
#density and ecdf
x11()
hist(data1,25,freq=F,xlab="x",ylab="f(x)",main="",
      cex.lab=1.2,ylim=c(0,0.13),xlim=c(0,70),col="bisque")
curve(dBGA(x,mu=exp(fit$mu.coefficients),sigma=exp(fit$sigma.coefficients
),nu=fit$nu.coefficients),
      add=T,lwd=3,col="red")
box()

x11()
plot(ecdf(data1),lwd=3,xlab='x',
ylab='F(x)',cex.lab=1.2,cex.axis=1.2,xlim=c(0,90),main="", col =
'gray30')
curve(pBGA(x,mu=exp(fit$mu.coefficients),sigma=exp(fit$sigma.coefficients
),nu=fit$nu.coefficients),

```

```

add=T,lwd=3,col="red")

Res.q0 <- fit$residuals
x11()
qqnorm(Res.q0,pch=19,ylim = c(-3,3),xlim=c(-3,3),col="black",
       ylab="Quantis amostrais",xlab="Quantis teóricos",main =
       "",cex.lab=1.2,cex.axis=1.2)
qqline(Res.q0, datax = F, distribution = qnorm,
       col="red",lwd=3)

#####
## Aplicacion 2: Gastric Cancer
#####
require(gamlss)
require(gamlss.cens)
require(survival)

#Dados
# tempo
t<-as.numeric(c(0.63, 35, 10.72, 30.69, 24.97, 19.84, 32.5, 31.38, 0.2, 29.08,
                22.14, 31.64, 0.26, 25.76, 23.85, 24.21, 36, 34.74, 33.36, 4.67,
                35.2, 33.65, 22.99, 34.21, 28.22, 17.24, 26.78, 23.39, 25.33,
                5.76, 25.3, 33.91, 16.94, 34.9, 16.18, 30.63, 33.09, 7.89, 4.01,
                14.77, 9.7, 17.43, 27.89, 23.55, 13.78, 20.49, 18.39, 25.36,
                20.49, 17.14, 0.1, 0.49, 14.01, 8.85, 36, 0.56, 26.05, 36, 36,
                30.26, 22.86, 36, 8.95, 36, 36, 36, 25.79, 1.61, 17.7, 15.43,
                24.21, 35.72, 35.59, 19.93, 36, 32.53, 28.98, 23.52, 24.57, 21.48,
                11.88, 32.4, 31.84, 34.41, 13.95, 26.32, 6.55, 33.88, 10.03,
                28.65, 35.03, 13.82, 19.44, 12.5, 30.16, 23.39, 4.67, 10.2, 8.32,
                1.18, 13.49, 16.09, 21.02, 3.32, 6.91, 29.31, 12.83, 25.26, 19.14,
                23.82, 29.28, 2.89, 0.66, 0.3, 21.38, 0.33, 12.5, 27.37, 24.01,
                9.47, 11.97, 7.17, 2.8, 9.05, 16.38, 24.8, 25.56, 26.32, 25.33,
                14.05, 7.93, 21.84, 17.34, 4.93, 20.76, 24.28, 26.15, 11.41,
                25.79, 16.94, 25.59, 21.88, 19.38, 8.39, 36, 3.49, 36, 1.45,
                18.62, 24.31, 13.13, 4.54, 6.15, 36, 28.59, 36, 36, 36, 2.63,
                30.95, 19.21, 36, 17.07, 17.14, 36, 0.66, 36, 36, 10.53, 16.51,
                36, 7.7, 36, 35.89, 14.7, 0.23, 9.28, 36, 32.8, 36, 25.23, 1.78,

```

8.91, 14.34, 8.36, 15.76, 10.76, 14.38, 35.3, 32.47, 34.08, 2.73,
29.97, 36, 30.49, 2.96, 31.68, 29.31, 0.53, 8.78, 13.09))

Censura

```
d<-as.integer(c(1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1,
                0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1,
                1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0,
                1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0,
                0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1,
                1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1,
                0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
                0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0,
                0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0,
                0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1))
```

Tratamento

```
trat<-as.factor(c(0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0,
                  0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0,
                  0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0,
                  1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1,
                  0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0,
                  0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1,
                  0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0,
                  1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1,
                  1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
                  0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1))
```

```
dados <- data.frame(t,d,trat)
```

```
str(dados)
```

```
### Ajuste KM
```

```
fit <- survfit(Surv(t,d)~trat)
```

```
plot(fit,mark.time = T,lwd=c(3,3,3),lty=c(1,2,3))
```

```
##### Gamma
```

```
fit11 <- gamlss(Surv(t,d)~trat,
               sigma.formula = ~trat,
               family=cens(GA),c.crit=0.01,n.cyc=2000,data=dados)
plot(fit11)
wp(fit11)
summary(fit11)

#####

##### BGA considerando sigma = beta
fit22sigma <- gamlss(Surv(t,d)~1,
                    sigma.formula = ~trat,
                    nu.formula = ~1,
                    family=cens(BGA),c.crit=0.01,n.cyc=2000,data=dados)
plot(fit22sigma)
wp(fit22sigma,ylim.all = 1,cex.lab = 1.35)
summary(fit22sigma)

##### BGA considerando mu = alpha
fit22mu <- gamlss(Surv(t,d)~trat,
                  sigma.formula = ~1,
                  nu.formula = ~1,
                  family=cens(BGA),c.crit=0.01,n.cyc=2000,data=dados)
plot(fit22mu)
wp(fit22mu,ylim.all = 1,cex.lab = 1.35)
summary(fit22mu)

##### BGA considerando mu e sigma
set.seed(98)
fit22 <- gamlss(Surv(t,d)~trat,
                sigma.formula = ~trat,
                nu.formula = ~1,
                family=cens(BGA),c.crit=0.01,n.cyc=2000,data=dados)
plot(fit22)
wp(fit22,ylim.all = 1,cex.lab = 1.35)
summary(fit22)

#####
```

```

#### WEibull
fit33 <- gamlss(Surv(t,d)~trat,
               sigma.formula = ~trat,
               family=cens(WEI),c.crit=0.01,n.cyc=200,data=dados)

plot(fit33)
wp(fit33)
summary(fit33)

AIC(fit11,fit33,fit22)
BIC(fit11,fit33,fit22)

km.1 <- survfit(Surv(t,d)~trat)

plot(km.1,mark.time = T,lwd=c(3,3),col=c("gray25","gray45"))
curve(1-pBGA(x,mu=exp(-0.062),sigma=exp(-3.699),
            nu=0.019),add=T,lwd=4,col="darkblue" )
curve(1-pBGA(x,mu=exp(-0.062),sigma=exp(-3.699-0.207),
            nu=0.019),add=T,lwd=3,col="brown3" )
legend(5,0.4,cex=c("Distribuição Empírica: Trat 0","BG: Trat
0","Distribuição Empírica: Trat 1","BG: Trat 1"),
      ,lwd=c(3,3,3,3),lty=c(1,1,1,1),col=c("gray25","darkblue","gray45",
      "brown3"),bty="n")

x11()
plot(km.1,mark.time = T,lwd=c(3,3),col=c("gray25","gray45"))
curve(1-pBGA(x,mu=exp(-0.221),sigma=exp(-3.506),
            nu=0.032),add=T,lwd=4,col="darkblue" )
curve(1-pBGA(x,mu=exp(-0.221+1.306),sigma=exp(-3.506+1.077),
            nu=0.032),add=T,lwd=3,col="brown3" )
legend(5,0.4,cex=1,c("Distribuição Empírica: Trat 0","BG: Trat
0","Distribuição Empírica: Trat 1","BG: Trat 1"),
      ,lwd=c(3,3,3,3),lty=c(1,1,1,1),col=c("gray25","darkblue","gray45",
      "brown3"),bty="n")

n <- length(dados$t)

```



```
index <- 1:n
```

```
Res.q1 <- fit22$residuals
```

```
plot(index,Res.q1,col="gray11",pch=19,  
      ylab="Resíduos Quantílicos",xlab="Index",main =  
      "(a)",cex.lab=1.35,cex.axis=1.35, font.main = 1,  
      ylim=c(-3.5,3.5))  
abline(h=-2.5,lwd=4,lty=2,col="firebrick3")  
abline(h=0,lwd=2,lty=1,col="gray50")  
abline(h=2.5,lwd=4,lty=2,col="firebrick3")
```

```
x11()  
plot(density(Res.q1),lwd=3,xlab="Resíduos  
quantílicos",ylab="Densidade",main="(b)",col="gray11", font.main =1)  
rug(jitter(Res.q1))
```

```
x11()  
qqnorm(Res.q1,pch=19,ylim = c(-3,3),xlim=c(-3,3),col="gray11",  
       ylab="Quantis Amostrais",xlab="Quantis Teóricos",main =  
       "",cex.lab=1.2,cex.axis=1.2, font.main = 1)  
qqline(Res.q1, datax = F, distribution = qnorm, lwd = 3,  
       col="firebrick3")
```

```
x11()  
wp(fit22,ylim.all = 1, col = 'red', pch = 1)
```