



Universidade de Brasília
Departamento de Estatística

**A Distribuição Beta Bimodal:
Propriedades e Aplicações**

Lucas Melo Alfaia

Projeto apresentado para obtenção do
título de Bacharel em Estatística

**Brasília
2021**

Lucas Melo Alfaia

**A Distribuição Beta Bimodal:
Propriedades e Aplicações**

Orientador: Roberto Vila

Projeto apresentado para obtenção do
título de Bacharel em Estatística

**Brasília
2021**

Este trabalho é todo dedicado as minhas avós Lurdes e Ninfa. Obrigado por tudo, estou com saudades ...

Agradecimentos

Aos familiares, por todo o apoio, que muito contribuíram para a realização deste trabalho. Aos meus pais e irmão, que me incentivaram nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava ao curso e à realização deste trabalho.

Aos amigos do curso: Álvaro Eduardo, David Dakan, Luiz Eduardo, Fernando Vitor, Matheus Pereira, Paulo Octávio, Victor de Souza Barros e Letícia Ferreira, seja pela ajuda nas dúvidas, nos momentos de entretenimento, nos conselhos quando me desestabilizava emocionalmente.

Aos amigos da vida: Pedro Viana, Fábio Junior, Felipe Rodrigues, Eduardo Horowitz dentre outros pelo companheirismo.

Ao Orientador Roberto Vila por sempre responder minhas dúvidas prontamente, por me ensinar como produzir um artigo científico.

A todos os professores do Departamento de Estatística no geral que tiveram paciência comigo e que em algum momento contribuíram para a minha vida acadêmica.

Resumo

O presente trabalho baseia-se na Pré-publicação: “A Model for Bimodal Rates and Proportions”, escrito por R. Vila, L. Alfaia, A. F. B. Menezes, M. N. Çankaya e M. Bourguignon, e disponibilizado no link [Arxiv.org/2108.07934](https://arxiv.org/abs/2108.07934). O modelo Beta é considerado como a distribuição mais importante para ajustar dados dentro em um intervalo $(0, 1)$. No entanto, a distribuição beta não é adequada para modelar dados nesse intervalo que apresentam comportamento bimodal. Sendo isso, o objetivo deste trabalho é propor uma distribuição Beta bimodal utilizando uma transformação quadrática baseada no modelo Alpha-skew-normal. Ao longo dos capítulos foram abordados propriedades básicas dessa distribuição como média, variância, momentos, medidas de entropia e taxas de risco. Além disso, foi apresentado um modelo de regressão baseado nesta distribuição e averiguamos os resíduos. Dentro da estrutura desse modelo estimou-se os parâmetros da distribuição pelo método de máxima verossimilhança e verificou-se a consistência desses estimadores por simulações de Monte Carlo. Na aplicação, o modelo se ajustou bem ao banco de dados analisado.

Palavras Chaves Distribuição Bimodal, Distribuição Beta. Simulações de Monte Carlo. Modelo de regressão.

Abstract

The present work is based on the Pre-publication: “A Model for Bimodal Rates and Proportions”, written by R. Vila, L. Alfaia, AFB Menezes, MN Cankaya and M. Bourguignon, and available at [Arxiv.org/2108.07934](https://arxiv.org/abs/2108.07934). The Beta model is considered as the most important distribution to fit data within a range (0,1). Therefore, the objective of this work is to propose a bimodal Beta distribution using a quadratic transformation based on the Alpha-skew-normal model. Throughout the chapters, basic properties of this distribution, such as mean, variance, moments, entropy measures and hazard rates, were addressed. Furthermore, a regression model based on this distribution was presented and we investigated residuals. Within the structure of this model, the parameters of the distribution were estimated by the maximum likelihood method and the consistency of these estimators was verified by Monte Carlo simulations. In the application, the model fits well to the analyzed database.

Keywords Bimodal Distribution, Beta Distribution. Monte Carlo simulations. Regression model.

Conteúdo

Introdução	7
1 Algumas definições básicas	9
1.1 Variáveis Aleatórias	9
1.2 Esperança matemática	10
1.3 Momentos reais	11
1.4 Estimação de Parâmetros e Informação de Fisher	12
1.5 Simulação de Monte Carlo	13
1.6 Medidas de Sobrevivência	14
1.7 Medidas de Entropia	15
1.8 Métricas de qualidade de ajuste	16
1.9 Processos Estocásticos	17
2 Distribuição Beta	19
2.1 Momentos, Média, Moda, Variância, Função Geradora de Momentos e Entropia de Shannon	20
2.2 Estimadores de Máxima Verossimilhança e matriz de Informação	21
3 Distribuição Beta Bimodal	23
3.1 A função densidade da Beta bimodal	23
3.2 Formas da distribuição BBeta	24
3.3 Medidas de risco	27
3.4 Momentos reais	30
3.5 Medidas de entropia	33
3.6 Representação estocástica	38
3.7 Identificabilidade da distribuição Beta bimodal	39
4 Modelo de regressão, estimativa e análise diagnóstica	41
5 Estudo de simulação	44
5.1 Estimativas dos parâmetros	44
5.2 Resíduos	45

6	Aplicação de dados reais	48
	Conclusão	52
	Bibliografia	53
	Appendices	55
	.1 Misturas finitas de distribuições Beta é identificável	56
	.2 Estimativa de máxima verossimilhança	57
	.3 Matriz de informação	57

Lista de Figuras

2.1	Função densidade de probabilidade para diferentes valores dos parâmetros da distribuição Beta.	20
3.1	Função densidade de probabilidade para diferentes valores de parâmetros da distribuição Beta bimodal.	24
3.2	Função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Beta com α e $\beta = 2$	28
3.3	Função de Sobrevivência BBeta para diferentes valores dos parâmetros da distribuição com α e $\beta = 2$	29
3.4	Função de risco BBeta para alguns valores dos parâmetros e α e $\beta = 2$. . .	30
3.5	Assimetria para alguns valores de ρ , α e β são valores fixos iguais a 2. . . .	31
3.6	Curtose para alguns valores de ρ , α e β são valores fixos iguais a 2.	32
5.1	Boxplots dos parâmetros estimados obtidos em experimentos de Monte Carlo para diferentes tamanhos de amostra	45
5.2	Gráficos de probabilidade normal das estatísticas de ordem média.	46
6.1	Gráficos empíricos de dados	49
6.2	Gráfico Half-normal dos resíduos de quantis randomizados com envelope simulado.	51

Lista de Tabelas

5.1	Viés estimado e erro quadrático médio.	45
5.2	Medidas descritivas dos resíduos de quantis randomizados.	47
6.1	Estimativas de ML, erros padrão e intervalo de confiança de 95 %.	50
6.2	Medidas de adequação.	50

Introdução

Em diversas áreas do conhecimento surgem situações em que os dados coletados possuem um comportamento com presença de mais de uma moda. Fenômeno conhecido como multimodalidade (Ahmed, 2008). Por exemplo, na medicina (encontrado em pixels de imagens de placas de aterosclerose (Shankar, 2003)); em processos para controle de qualidade (Hsu, 2008), na geociências (no tempo em que determinado vulcão entra e sai de erupção (Weisberg, 2005)); em análises de tamanho de peixes (Rao, 1988); entre outras situações, o comportamento dos conjuntos de dados encontrados têm características multimodais.

Nesse contexto é necessário propor novos modelos probabilísticos que consigam descrever esse comportamento. Neste trabalho estamos interessados nas distribuições bimodais que possuam alta flexibilidade em seu comportamento para realizar inferência no conjunto de dados. Distribuições bimodais são caracterizada pela presença de no máximo dois pontos de máximo no gráfico da função densidade de probabilidade (f.d.p.).

Geralmente, quando se deseja descrever um conjunto de dados com características bimodais, é comum usar misturas de distribuições. Entretanto, essa prática pode gerar problemas sérios de identificabilidade na estimação dos parâmetros, isto é, parâmetros diferentes podem gerar a mesma distribuição de probabilidade (Gomez, 2019). Além do mais, a literatura foca em misturas de distribuições simétricas como as Gaussianas e deixa de lado distribuições mais flexíveis como distribuições que podem ser tanto assimétricas como simétricas, como o caso das distribuições Gama ou Beta (Ahmed, 2008). No caso da distribuição Beta encontramos poucos estudos na literatura que implementem inferência de casos bimodais.

Um método bastante eficiente de encontrar distribuições bimodais é a multiplicação de qualquer distribuição unimodal por um fator quadrático adequado, para maiores detalhes veja, por exemplo, (Elal-Olivero, 2010), (Vila et al., 2020a), (Vila et al., 2020b) e mais recentemente (Vila and Çankaya, 2021a). No artigo (Elal-Olivero, 2010), o autor realizou essa operação pela primeira vez em uma distribuição Gaussiana e assim introduziu assimetria e bimodalidade à sua f.d.p.. Este modelo é conhecido como Alpha-Skew Normal (ASN). Uma variável aleatória (v.a.) Z tem distribuição ASN se sua f.d.p. e sua

função de distribuição acumulada (f.d.a.), respectivamente, são dadas por

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

em que o parâmetro δ controla a bimodalidade e, Φ e ϕ representam a f.d.p. e f.d.a. da distribuição Normal padrão, respectivamente.

O presente estudo pretende analisar o comportamento de uma nova distribuição originada pela distribuição Beta, isto é, propor uma distribuição Beta bimodal usando transformações na f.d.p. original, sem desrespeitar as restrições que define uma função de densidade de probabilidade. O fator quadrático proposto por (Elal-Olivero, 2010) será essencial para o processo da proposta da nova distribuição bimodal.

Esta dissertação está organizada em 6 capítulos. 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 Beta; no Capítulo 3 introduz-se a distribuição Beta bimodal e suas respectivas propriedades matemáticas; no Capítulo 4 um modelo de regressão da distribuição Beta bimodal é proposto; no Capítulo 5 é feita a estimação das variáveis regressoras do modelo proposto pelo método de máxima verossimilhança e é verificada a performance dessa estimação por simulações de Monte Carlo; e no Capítulo 6 é feita a aplicação desse modelo em um banco de dados reais.

Capítulo 1

Algumas definições básicas

Antes de apresentar a distribuição Beta bimodal, é importante abordar alguns conceitos teóricos de Probabilidade e Estatística. Sendo assim, neste Capítulo 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 de processos estocásticos.

1.1 Variáveis Aleatórias

Uma variável aleatória (v.a.) corresponde a um resultado numérico de um experimento, cujo resultado é aleatório, isto é, não-determinístico. Dado isto, pode-se utilizar toda a Teoria da Probabilidade para se chegar em conclusões acerca do comportamento do resultado do experimento. A Teoria da Probabilidade é o ramo da matemática relacionado a estes fenômenos aleatórios (ou casuais). Muitas pessoas têm se dedicado ao seu estudo, devido ao seu interesse intrínseco, bem como as aplicações bem sucedidas em muitas áreas das ciências físicas, biológicas e sociais, na engenharia e no mundo dos negócios. A experiência mostra que muitos fenômenos não-determinísticos exibem uma regularidade estatística, pois, dado um experimento simples, é possível descrever o conjunto de todos os resultados ao repetir o experimento aleatório um grande número de vezes. Este conjunto é conhecido como espaço amostral, e frequentemente é representado pela letra grega Ω . A grosso modo, 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, isto é, uma função aditiva. Valores numéricos associados a um experimento aleatório, dado as restrições de mensurabilidade com respeito à σ -álgebra, podem ser classificados como v.a.'s, e frequentemente denotadas por X, Y, Z, W etc. Se o conjunto de valores possíveis de uma v.a. X for finito ou infinito enumerável, chamamos

X de v.a. discreta, isto é, os valores possíveis de X podem ser dispostos em ordem crescente e serem associados cada um a um número inteiro positivo em ordem crescente. Se o contradomínio de X for um intervalo ou uma coleção de intervalos, então diremos que X é uma v.a. (absolutamente) contínua, isto é, dependendo da classe do conjunto dos valores (contradomínio) da variável, ela pode ser classificada como discreta, contínua ou mista. Nesta dissertação estaremos interessados nas variáveis aleatórias contínuas.

Definição 1 (Função de distribuição acumulada, (Magalhães, 2015)). *Se X é uma v.a. contínua, então existe uma função não negativa $f(\cdot)$, chamada função densidade de probabilidade (f.d.p) associada a X , tal que a função de distribuição acumulada (f.d.a) é representada da seguinte forma*

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

A f.d.p. $f(\cdot)$ representa a taxa de variação instantânea da f.d.a. naquele valor x . Se $f(\cdot)$ é uma f.d.p. então deve cumprir certos pressupostos.

Proposição 1.1.1. *Se $f(\cdot)$ é uma f.d.p., então as seguintes propriedades decorrem imediatamente da Definição 1:*

1. $f(x) \geq 0, \forall x \in \mathbb{R};$
2. $\int_{-\infty}^{+\infty} f(x) dx = 1.$

1.2 Esperança matemática

Esperança Matemática ou valor esperado de uma v.a., um dos conceitos mais relevantes da Teoria das Probabilidades. A esperança matemática é a média aritmética de uma v.a. - fenômeno ou experimento aleatório, conceito amplamente conhecido na análise exploratória de dados, somente que este é calculado para os dados da amostra, enquanto a esperança matemática é calculada nos modelos probabilísticos que se ajustam a esses conjuntos de dados. Esse estudo é de extrema importância pois é uma forma de caracterizar o comportamento da v.a..

Definição 2 (Esperança, (Magalhães, 2015)). *A esperança matemática de uma v.a. contínua X é definida por*

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

sempre que a integral acima esteja bem definida.

Na seguinte proposição serão enunciadas algumas das propriedades básicas da esperança matemática.

Proposição 1.2.1 ((Magalhães, 2015)). *Se X e Y são duas v.a.'s definidas sobre o mesmo espaço amostral Ω , então para todo $a, b \in \mathbb{R}$,*

1. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$;
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
3. *Se $X(\omega) \leq Y(\omega), \forall \omega \in \Omega$, então $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

1.3 Momentos reais

Vimos que podemos definir uma v.a. a partir de sua f.d.p. ou por sua f.d.a.. Outra forma de caracterizar uma v.a. e pelo seus momentos reais. Sabendo-se os momentos pode-se derivar diversas medidas, como pro exemplo, média, variância, assimetria e curtose.

Definição 3 (Momentos, (Magalhães, 2015)). *Dado $k \in \mathbb{N}$, o momento de ordem k de uma v.a. X é definido por $\mathbb{E}(X^k)$, sempre que essa quantidade exista. No caso que $k \in \mathbb{R}$, a quantidade $\mathbb{E}(X^k)$ é conhecida 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 $\mathbb{E}(X) = \mu$ é a esperança matemática de X .

O momento central de ordem 1 é zero. Com os momentos centrais de ordem 2, 3 e 4 e possível obter expressões fechadas para a variância, coeficiente de assimetria e coeficiente de curtose, respectivamente.

Definição 4 (Variância, (Magalhães, 2015)). *Dado que o momento de ordem 2 seja finito, a variância é representada pelo momento central de ordem 2, isto é,*

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

Definição 5 (Coeficiente de assimetria, (Magalhães, 2015)). *Seja X uma v.a. com média μ e desvio padrão $\sigma > 0$. O momento central de ordem 3 dividido pelo desvio-padrão ao cubo é conhecido como coeficiente de assimetria e é denotado por α_3 . Em outras palavras, o coeficiente de assimetria de X é definido por*

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

Definição 6 (Coeficiente de curtose, (Magalhães, 2015)). *Seja X uma v.a. com média μ e desvio padrão $\sigma > 0$. O momento central de ordem 4 dividido pelo seu desvio-padrão a quarta potencia é chamado de coeficiente de curtose de X e é denotado por α_4 .*

Precisamente, o coeficiente de curtose de X é dado por

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

Para essas medidas, é necessário a distribuição ter seus momentos bem formulados. Para este propósito, a definição da função geradora de momentos é essencial, pois sobre certas condições, o cálculo dos momentos pode ser obtido através dela.

Definição 7 (Função geradora de momentos). *A função geradora de momentos (FGM) de uma v.a. 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 FGM de uma v.a. X esta bem definida em algum vizinhança da origem, então $\mathbb{E}(X^n)$ existe para todo n inteiro e positivo. Consequentemente,*

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

Sobre as condições do teorema acima, tem-se que a FGM de X admite a expansão (veja, por exemplo, (Jonhson et al., 1993))

$$M_X(t) = \sum_{r=0}^{\infty} \mathbb{E}(X^r) \frac{t^r}{r!}. \quad (1.1)$$

1.4 Estimação de Parâmetros e Informação de Fisher

O método denotado máxima verossimilhança é um dos procedimentos mais usados para se obter estimadores. Ele trata o problema de estimação baseado nos resultados obtidos pela amostra e devemos determinar qual valor do parâmetro que maximiza a probabilidade de ter gerado tal amostra. Isso é feito buscando o ponto de máximo da função de verossimilhança.

Definição 8 (Função de máxima verossimilhança). *Seja X uma v.a. com distribuição $G(\boldsymbol{\theta})$ caracterizada pelo vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$. Seja (X_1, \dots, X_n) uma amostra aleatória simples, de tamanho n , da v.a. 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 9 (Estimador de máxima verossimilhança, (Casella and Berger, 2010)). *Se $\hat{\theta}$ é um valor no domínio do vetor de parâmetros θ na qual a função $L(\theta; \mathbf{x})$ atinge seu valor de máximo, então $\hat{\theta}$ é um estimador de máxima verossimilhança de θ com base em uma amostra aleatória de X .*

Uma medida de grande importância no estudo de estimação de parâmetros é conhecida como Informação de Fisher (IF). A IF é definida através de uma média na qual fornece um limitante para a variância do estimador não viciado de variância uniformemente mínima (ENNVUM) de θ . O conceito dessa informação é associado a uma espécie de curvatura média da função de verossimilhança no sentido de que quanto maior a curvatura mais precisa é a informação contida na verossimilhança. Quanto maior a informação mais se sabe sobre os parâmetros e mais se limita a variância.

Definição 10. *A medida de informação esperada de Fisher de θ através de X é definida pelos seguintes elementos*

$$[I_F(\theta)]_{ij} = \mathbb{E} \left[\left(\frac{\partial \log f(X; \theta)}{\partial \theta_i} \right) \left(\frac{\partial \log f(X; \theta)}{\partial \theta_j} \right) \right], \quad i, j = 1, \dots, n.$$

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

$$[I_F(\theta)]_{ij} = \mathbb{E} \left[- \frac{\partial^2 \log f(X; \theta)}{\partial \theta_i \partial \theta_j} \right], \quad i, j = 1, \dots, n.$$

1.5 Simulação de Monte Carlo

Dentro da abordagem das simulações, o método de Monte Carlo é um método computacional que utiliza números aleatórios e estatísticas para resolver diversos problemas de física e de matemática onde são muito difíceis de serem resolvidos com outros métodos. O método baseia-se em gerar r amostras aleatórias de tamanho n e com a mesma distribuição de uma v.a. X . São calculados os estimadores de máxima verossimilhança dos parâmetros da distribuição analisada. A partir desses valores pode-se obter um estimador dos parâmetros de X .

Dado um parâmetro α de X , gera-se estimadores de máxima verossimilhança para cada amostra $(\hat{\alpha}_1, \dots, \hat{\alpha}_r)$. Um estimador então quantificado por esses valores é dado por $\hat{\alpha} = \sum_{i=1}^r \hat{\alpha}_i / r$

O viés da estimação é dada por

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

e o erro quadrático médio é definido por

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

Um método para gerar amostras de uma v.a. X e o método da Transformada Inversa. Esse método baseia-se em simular amostras de uma v.a. distribuída uniformemente no intervalo $(0, 1)$, e realizar a transformação dessa variável utilizando a inversa da f.d.a. de X . Isto é, $X = F^{-1}(U)$ sendo U a v.a. com distribuição uniforme e F a distribuição acumulada de X .

1.6 Medidas de Sobrevida

Em problemas práticos, geralmente não é possível coletar todos os dados ao mesmo tempo. Dessa forma, faz-se necessária a definição clara e objetiva do tempo inicial do estudo. A partir da definição do tempo inicial, o estudo da análise de sobrevivência consiste em denominar o tempo de falha acerca do tempo inicial como uma variável resposta a ser estudada, isto é, o tempo até a ocorrência de um determinado evento. A unidade de medida do tempo de falha também é uma característica importante no planejamento do estudo, podendo ser medida em horas, dias, meses, anos, etc. A presença de censura caracteriza os modelos de análise de sobrevivência. Censura pode ser definida como a ausência da ocorrência do evento no tempo de análise. Ou seja, por motivos que não envolvem o objetivo central do estudo, o elemento amostral não apresenta tempo de falha.

Definição 11 (Função de sobrevivência). *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 $f(\cdot)$ é a f.d.p associada á variável tempo de vida T .

Definição 12 (Função de risco). *A função de risco para v.a.'s contínuas é definida como a probabilidade instantânea de um evento de interesse em um estudo ocorrer num intervalo de tempo $[t, t + \Delta t]$, dado que esse evento não ocorreu antes do tempo t . Em outras palavras,*

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

Definição 13 (Vida média residual). *A vida média residual é definida como o tempo médio de espera até que um determinado evento de interesse ocorra, dado que ele não*

ocorreu até o momento t . Em outras palavras,

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

1.7 Medidas de Entropia

As medidas de entropia medem a quantidade de incerteza ou informação que existe na amostra. A intuição por trás da quantificação da informação é a ideia de medir quanta surpresa existe em um evento. Os eventos raros (aqueles de baixa probabilidade) são mais surpreendentes e, portanto, têm mais informações do que os eventos comuns (de alta probabilidade). Calcular as informações de uma v.a. é equivalente a calcular as informações para a distribuição de probabilidade dos eventos dessa v.a.. Dentro das ciências da computação a intuição da entropia seria o valor médio de bits necessários para representar ou transmitir um evento retirado de uma distribuição de probabilidade da v.a.. Três entropias, as quais serão analisadas no decorrer do texto, serão definidas a seguir

Definição 14 (Entropia de Shannon, (Shannon, 1948)). *A Entropia de Shannon de uma v.a. X , sendo $f(x; \boldsymbol{\theta})$ a f.d.p. de X com seu vetor de parâmetros $\boldsymbol{\theta}$, é definida por*

$$H_1(X) = - \int_{-\infty}^{+\infty} f(x; \boldsymbol{\theta}) \log f(x; \boldsymbol{\theta}) \, dx.$$

Definição 15 (Entropia Quadrática, (Rényi, 1961)). *A Entropia Quadrática de uma v.a. X , sendo $f(x; \boldsymbol{\theta})$ a f.d.p. de X com seu vetor de parâmetros $\boldsymbol{\theta}$, é definida por*

$$H_2(X) = - \log \int_{-\infty}^{+\infty} f^2(x; \boldsymbol{\theta}) \, dx.$$

Definição 16 (Entropia de Tsallis, (Tsallis, 2009)). *A Entropia de Tsallis associada a uma v.a. não negativa X é definida por*

$$S_q(X) = \frac{1}{q-1} \left[1 - \int_{-\infty}^{\infty} f^q(x; \boldsymbol{\theta}) \, dx \right], \quad q \neq 1.$$

Usando a regra de L'Hôpital para o cálculo de limites, note que, quando $q \rightarrow 1$, $S_q(X)$ se aproxima da Entropia de Shannon. Isto é, a Entropia de Tsallis pode ser vista como uma extensão da entropia de Shannon.

1.8 Métricas de qualidade de ajuste

Ao propor uma nova distribuição, é fundamental o estudo de uma modelagem estatística baseado nessa nova distribuição, com o fim de averiguar se há ganhos de eficiência para a análise de dados reais em relação a outros modelos já conhecidos na literatura. Assim, abaixo estão descritas as métricas que foram utilizadas nesse estudo.

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

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

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 \widehat{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

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

se assemelha a do AIC, entretanto há uma maior penalização a parâmetros adicionais. Aqui, k denota o número de parâmetros adicionais, \widehat{L} é a medida de verossimilhança do modelo e n é o tamanho da amostra. Assim como 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 f.d.a. dessa distribuição com a f.d.a. 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 f.d.a. da distribuição de interesse e n é o tamanho da amostra. A partir da estatística T realiza-se um teste de hipóteses sob hipótese nula de que os dados seguem a distribuição de interesse.

- **KS (Kolmogorov-Smirnov).** Assim como critério de Cramér-von Mises, o critério de Kolmogorov-Smirnov, 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 f.d.a. da distribuição avaliada e $F_n(\cdot)$ denota 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 $X_i \leq x$ e o valor de 0 em outro caso.

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 f.d.a. F . No caso de v.a.'s discretas, antes de realizar a transformação é feita a randomização a fim de tornar a f.d.a. 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 v.a. com distribuição uniforme no intervalo $(0, 1]$ e $F(y_i^-; \hat{\mu}_i; \hat{\phi})$ denota o limite pela esquerda da função F em y_i .

Assim, os resíduos quantílicos são encontrados pela transformação $r_{q,i} = \Phi^{-1}(F^*)$, onde $\Phi(\cdot)$ denota a f.d.a. da distribuição normal padrão.

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

1.9 Processos Estocásticos

Ter abordagens diferentes para representar uma f.d.p. pode ser de grande auxílio para o estudo da mesma. Para esse fim, vamos introduzir o conceito de processos estocásticos. Um processo estocástico é uma coleção de variáveis aleatórias $\{X_t : t \in T\}$ indexadas em t , onde t é um índice definido num conjunto T chamado espaço tempo.

Assim, um processo estocástico é a descrição de um fenômeno aleatório que varia no tempo.

O processo estocástico pode representar a coleção das quantidades de carros que passam por um determinado ponto de uma rodovia, a evolução dos níveis de estoque semanais de uma firma, o comportamento de uma partícula de gás, variações nas qualidades dos produtos, variações nos preços de ações, vendas numa determinada loja, evolução do número de desempregados num determinado país, etc.

Os valores assumidos por um processo estocástico são denominados de estados e o conjunto de todos os estados possíveis é dito espaço de estados, frequentemente denotado por \mathcal{S} .

Definição 17. *Um processo estocástico é uma coleção $\{X_t : t \in T\}$ de v.a.'s definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathcal{P})$, onde T é um conjunto de índices.*

- Em virtude de suas inúmeras aplicações o parâmetro $t \in T$ é usualmente interpretado como tempo.
 - (a) Caso T seja finito ou infinito enumerável o processo é dito a tempo ou parâmetro discreto, onde $T = \{0, 1, 2, \dots\}$.
 - (b) Caso T contenha um intervalo da reta então processo é dito a tempo ou parâmetro contínuo, onde $T = [0, +\infty)$.
- O processo estocástico $\{X_t : t \in T\}$ pode ser classificado em função do seu espaço de estados \mathcal{S} como segue
 - (a) Caso \mathcal{S} seja finito ou infinito enumerável o estado é dito discreto.
 - (b) Caso \mathcal{S} contenha um intervalo da reta \mathbb{R} o processo é dito a valores contínuos.

Capítulo 2

Distribuição Beta

A distribuição Beta unimodal é frequentemente usada para modelagem de variáveis aleatórias tomando valores no intervalo $(0, 1)$. Devido a grande versatilidade no comportamento de uma v.a. com distribuição Beta para modelar f.d.p.'s com suporte $(0, 1)$, o modelo Beta tem inúmeras aplicações para representar quantidades físicas cujos valores estejam restritos a um intervalo identificável. Nesse contexto há inúmeros casos de aplicações como taxas, frações e proporções.

Uma v.a. X segue uma distribuição Beta com parâmetros de forma $\alpha > 0$ e $\beta > 0$, denotado por $X \sim \text{Beta}(\alpha, \beta)$, se sua f.d.a., denotada por $F(x; \alpha, \beta)$, é dada por

$$F(x; \alpha, \beta) = I_x(\alpha, \beta), \quad 0 < x < 1, \quad (2.1)$$

em que $I_x(\alpha, \beta) = B_x(\alpha, \beta)/B(\alpha, \beta)$ é a proporção da função Beta incompleta, $B_x(\alpha, \beta) = \int_0^x \omega^{\alpha-1}(1-\omega)^{\beta-1} d\omega$ é a função Beta incompleta, $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ é a função Beta e $\Gamma(\alpha) = \int_0^\infty \omega^{\alpha-1}e^{-\omega} d\omega$ denota a função Gama.

A f.d.p., denotada por $f(x; \alpha, \beta)$, associada à distribuição Beta (2.1) é

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1. \quad (2.2)$$

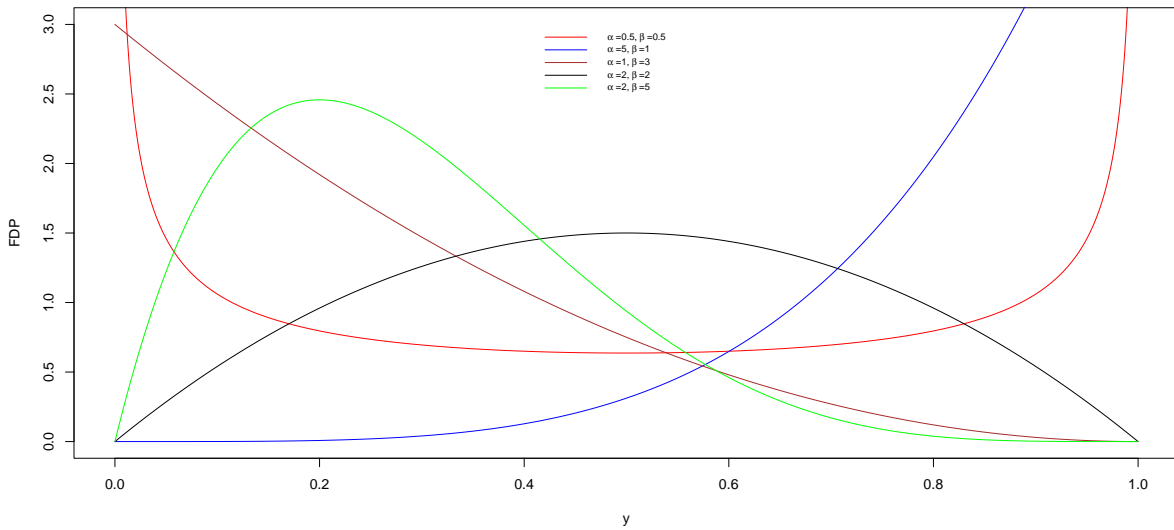


Figura 2.1: Função densidade de probabilidade para diferentes valores dos parâmetros da distribuição Beta.

Nas seções seguintes serão relembradas as principais características da distribuição Beta.

2.1 Momentos, Média, Moda, Variância, Função Geradora de Momentos e Entropia de Shannon

O k -ésimo momento real de $X \sim \text{Beta}(\alpha, \beta)$ é dado por

$$\mathbb{E}(X^k) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+k-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + k)}{\Gamma(\alpha + \beta + k)\Gamma(\alpha)}. \quad (2.3)$$

Sendo assim, a média e variância são apresentadas pelos cálculos a seguir.

A média e a variância associadas à distribuição (2.2) são escritas da seguinte forma

$$\mathbb{E}(X) = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)}{\Gamma(\alpha + \beta + 1)\Gamma(\alpha)} = \frac{\Gamma(\alpha + \beta)\alpha\Gamma(\alpha)}{(\alpha + \beta)\Gamma(\alpha + \beta)\Gamma(\alpha)} = \frac{\alpha}{\alpha + \beta}$$

Sendo assim, mostrando o cálculo do segundo momento,

$$\mathbb{E}(X^2) = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 2)}{\Gamma(\alpha + \beta + 2)\Gamma(\alpha)} = \frac{\Gamma(\alpha + \beta)(\alpha + 1)\alpha\Gamma(\alpha)}{(\alpha + \beta + 1)(\alpha + \beta)\Gamma(\alpha + \beta)\Gamma(\alpha)} = \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)}.$$

E então, podemos calcular a variância de X como segue

$$\text{Var}(X) = \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)} - \frac{\alpha^2}{(\alpha + \beta)^2} = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

Outra medida de bastante interesse é o cálculo da moda da distribuição Beta. A moda da distribuição Beta é obtida por derivar a f.d.p., igualar a 0 e em seguida analisar os pontos de máximo. A moda da distribuição Beta, denotada por $\text{Mo}(X)$, é dada por

$$\text{Mo}(X) = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

Por usar a expansão em (1.1) e a fórmula dos momentos (2.3), a FGM da v.a. X pode ser escrita como segue

$$M_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{t^k}{k!}.$$

A Entropia de Shannon associada a uma v.a. $X \sim \text{Beta}(\alpha, \beta)$ é dada por

$$\begin{aligned} h(X) &= \mathbb{E}[-\log f(x; \alpha, \beta)] \\ &= \int_0^1 -f(x; \alpha, \beta) \log(f; \alpha, \beta) dx \\ &= \log B(\alpha, \beta) - (\alpha - 1)\psi(\alpha) - (\beta - 1)\psi(\beta) + (\alpha + \beta - 2)\psi(\alpha + \beta). \end{aligned}$$

2.2 Estimadores de Máxima Verossimilhança e matriz de Informação

Seja X uma v.a. não negativa seguindo uma distribuição Beta (2.2) com vetor de parâmetros (α, β) . Seja $\mathbf{X} = (X_1, \dots, X_n)$ uma amostra aleatória, de tamanho n , de X . Se os valores observados correspondentes à amostra \mathbf{X} amostra são denotados pelo vetor $\mathbf{x} = (x_1, \dots, x_n)$, então a função log-verossimilhança para (α, β) é escrita da seguinte forma:

$$\begin{aligned} \log L(\alpha, \beta; \mathbf{x}) &= \sum_{i=1}^n \log f(x_i; \alpha, \beta) \\ &= (\alpha - 1) \sum_{i=1}^n \log(x_i) + (\beta - 1) \sum_{i=1}^n \log(1 - x_i) - n \log B(\alpha, \beta). \end{aligned}$$

As derivadas parciais de primeira ordem da função log-verossimilhança são, em relação a α e β , dadas por

$$\frac{\partial \log L(\alpha, \beta; \mathbf{x})}{\partial \alpha} = \sum_{i=1}^n \log(x_i) - n \frac{\partial \log B(\alpha, \beta)}{\partial \alpha}; \quad (2.4)$$

$$\frac{\partial \log L(\alpha, \beta; \mathbf{x})}{\partial \beta} = \sum_{i=1}^n \log(1 - x_i) - n \frac{\partial \log B(\alpha, \beta)}{\partial \beta}; \quad (2.5)$$

respectivamente. Já as derivadas parciais de segunda ordem, em relação a α e β , são dadas por

$$\begin{aligned} \frac{\partial^2 \log L(\alpha, \beta; \mathbf{x})}{\partial \alpha^2} &= -n \frac{\partial^2 \log B(\alpha, \beta)}{\partial \alpha^2}; \\ &= \frac{\partial^2 \log L(\alpha, \beta; \mathbf{x})}{\partial \beta^2} = -n \frac{\partial^2 \log B(\alpha, \beta)}{\partial \beta^2}; \end{aligned}$$

respectivamente. As expressões acima podem ser rescritas da forma a seguir, isto é, as derivadas parciais de segunda ordem, incluindo as mistas, são

$$\begin{aligned} \frac{\partial^2 \log L(\alpha, \beta; \mathbf{x})}{\partial \alpha^2} &= \psi_1(\alpha) - \psi_1(\alpha + \beta); \\ \frac{\partial^2 \log L(\alpha, \beta; \mathbf{x})}{\partial \beta^2} &= \psi_1(\beta) - \psi_1(\alpha + \beta); \\ \frac{\partial^2 \log L(\alpha, \beta; \mathbf{x})}{\partial \alpha \partial \beta} &= -\psi_1(\alpha + \beta); \end{aligned}$$

onde $\psi(\alpha)$ denota o logaritmo da derivada da função Gama, conhecida com função Digama, e $\psi_1(\alpha)$ é, então, a derivada da função Digama.

Logo, a matriz da Informação de Fisher é

$$I_F(\alpha, \beta) = \begin{bmatrix} \psi_1(\alpha) - \psi_1(\alpha + \beta) & -\psi_1(\alpha + \beta) \\ -\psi_1(\alpha + \beta) & \psi_1(\beta) - \psi_1(\alpha + \beta) \end{bmatrix}.$$

As estimativas de máxima verossimilhança (EMV) podem ser obtidas igualando as derivadas parciais de primeira ordem (2.4) e (2.5) a zero, porém, observe que estas equações não constituem um sistema linear, logo métodos numéricos têm que ser empregados para encontrar as respectivas raízes.

Capítulo 3

Distribuição Beta Bimodal

Neste capítulo, baseado na referência (Vila et al., 2021b), foram estabelecidos as primeiras propriedades matemáticas básicas da distribuição Beta bimodal (BBeta) e suas respectivas demonstrações. Serão apresentadas, a respeito da distribuição BBeta, a f.d.p., caracterizações da forma de unimodalidade e bimodalidade da distribuição, medidas de risco, momentos reais, medidas de entropia, representação estocástica da distribuição, identificabilidade da distribuição, estimativas de máxima verossimilhança e a matriz de informação.

3.1 A função densidade da Beta bimodal

A função densidade da Beta bimodal será construída a partir de uma ligeira modificação do método de (Elal-Olivero, 2010). Isto é, a distribuição Beta clássica será multiplicada pelo fator quadrático $\rho + (1 - \delta x)^2$ devidamente ponderado. Sendo assim, foi encontrada a distribuição Beta bimodal que será proposta neste trabalho.

Definição 18. Dizemos que uma v.a. não-negativa X segue a distribuição BBeta com vetor de parâmetros $\boldsymbol{\theta}_\delta = (\alpha, \beta, \rho, \delta)$, $\alpha > 0, \beta > 0, \rho \geq 0$ e $\delta \in \mathbb{R}$, denotado por $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$, se sua p.d.f. é dada por

$$f(x; \boldsymbol{\theta}_\delta) = \begin{cases} \frac{\rho + (1 - \delta x)^2}{Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)} x^{\alpha-1} (1 - x)^{\beta-1}, & 0 \leq x \leq 1, \\ 0, & \text{caso contrário,} \end{cases} \quad (3.1)$$

em que $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ é a função Beta, $\Gamma(\alpha)$ é a função Gama, e

$$Z(\boldsymbol{\theta}_\delta) = 1 + \rho - 2\delta \frac{\alpha}{\alpha + \beta} + \delta^2 \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} \quad (3.2)$$

denota a constante de normalização. Em outras palavras, a constante $Z(\boldsymbol{\theta}_\delta)$ é definida de

tal forma que

$$\int_0^1 f(x; \boldsymbol{\theta}_\delta) dx = 1.$$

Observe que no caso particular $\delta = 0$, a função em (3.1) se simplifica, pois a função se torna uma distribuição beta multiplicado por $(\rho + 1)/(\rho + 1) = 1$ e obtemos a distribuição Beta (estudada no Capítulo 2) com vetor de parâmetros $\boldsymbol{\theta}_0 = (\alpha, \beta, \rho, 0) = (\alpha, \beta)$. Os parâmetros α , β e ρ , controlam a forma da distribuição. A uni- ou bimodalidade da distribuição é controlada pelo parâmetro δ . Note que, para α , β e $\delta \neq 0$ fixos, o parâmetro ρ também controla a uni- ou bimodalidade da distribuição.

A figura 3.1 a seguir mostra o comportamento dessa nova distribuição com valores dos parâmetros α e β fixo:

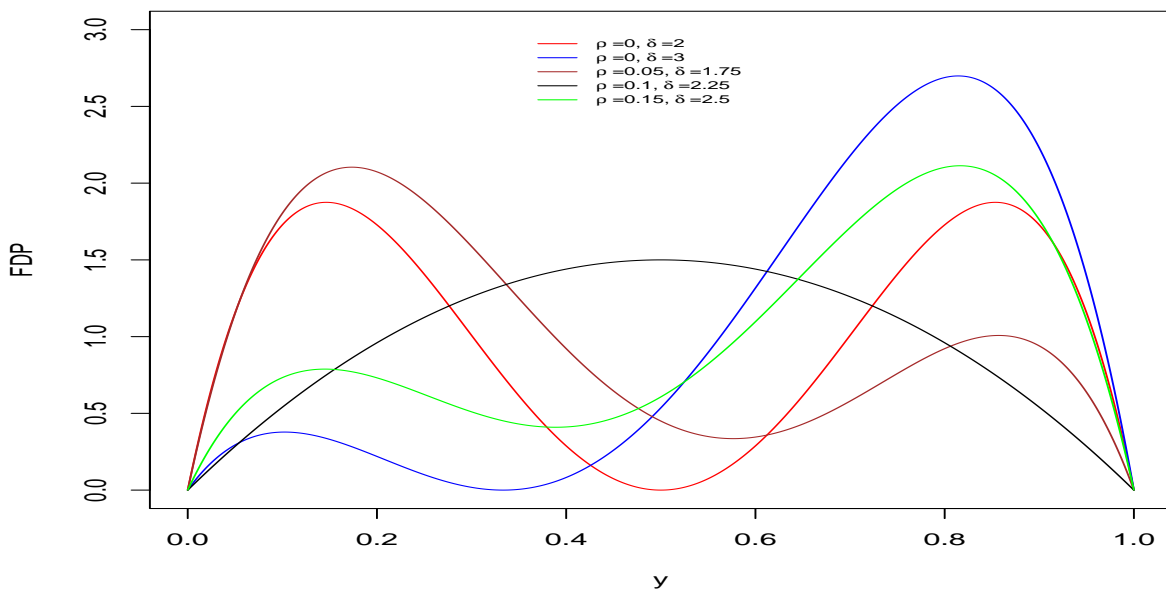


Figura 3.1: Função densidade de probabilidade para diferentes valores de parâmetros da distribuição Beta bimodal.

Uma vantagem desta distribuição em relação a outros modelos bimodais é a falta de problema de identificabilidade que ocorre nos métodos de misturas de distribuições, por exemplo.

3.2 Formas da distribuição BBeta

Esta seção tem como objetivo caracterizar rigorosamente o comportamento das modas da distribuição BBeta. Mais precisamente, nesta seção serão estabelecidas condições

suficientes, em função dos parâmetros, para garantir a existência de unimodalidade e bimodalidade da distribuição BBeta.

Sabe-se que o comportamento das modas, isto é, os pontos de máximo de uma função, podem ser obtidos pela derivada da mesma e igualando a zero, ou seja, os zeros (chamados de pontos críticos da função) da derivada da função. Aplicando o método na distribuição BBeta estudaremos o comportamento da função em relação a caracterização da unimodalidade e bimodalidade, mas antes apresentaremos as seguintes restrições para o auxílio em garantir a bimodalidade da BBeta.

Definimos o conjunto \mathcal{A} formado por todos os $\boldsymbol{\theta}_\delta = (\alpha, \beta, \rho, \delta) \in (0, +\infty)^2 \times [0, +\infty) \times \mathbb{R}$ de modo que

$$\alpha > 1, \beta > 1, \delta > 1, \rho \neq 0; \quad (3.3)$$

$$\delta(\alpha - 3) > -2(\alpha + \beta - 2); \quad (3.4)$$

$$2\delta(2 + \delta - \alpha) < (\rho + 1)(\alpha + \beta - 2); \quad (3.5)$$

$$(\rho + 1)(\alpha - 1) > 2\delta. \quad (3.6)$$

Estas condições serão necessárias para todas as provas dos teoremas a seguir. Além disso, observe que o conjunto \mathcal{A} é não vazio porque o ponto $\boldsymbol{\theta}_\delta = (6, 6, 0.1, 2) \in \mathcal{A}$, por exemplo.

Teorema 3.2.1 (Bimodalidade no caso $\rho \neq 0$). *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ tal que $\boldsymbol{\theta}_\delta \in \mathcal{A}$, então a distribuição BBeta é bimodal.*

Demonstração. Sendo a derivada da função de densidade BBeta a seguir

$$f'(x; \boldsymbol{\theta}_\delta) = \frac{x^{\alpha-2}(1-x)^{\beta-2}}{Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)} p_3(x), \quad (3.7)$$

em que $p_3(x)$ denota o seguinte polinômio de ordem 3:

$$p_3(x) = [\rho + (1 - \delta x)^2][(\alpha - 1)(1 - x) - (\beta - 1)x] - 2\delta(1 - \delta x)x(1 - x). \quad (3.8)$$

Excluindo os pontos $x = 0$ e $x = 1$, pois, por definição, estes não são pontos críticos, vemos que $f'(x; \boldsymbol{\theta}_\delta) = 0$ se e somente se

$$\begin{aligned} p_3(x) &= -x^3\delta^2(\alpha + \beta - 2) + x^2\delta[\delta(\alpha - 3) + 2(\alpha + \beta - 2)] \\ &+ x[2\delta(2 + \delta - \alpha) - (\rho + 1)(\alpha + \beta - 2)] - 2\delta + (\rho + 1)(\alpha - 1) = 0. \end{aligned}$$

Usando (3.3) em (3.7)-(3.8) temos que $f'(x; \boldsymbol{\theta}_\delta) \neq 0$ para todos $x > 1$. Isso garante que as raízes de $p_3(x)$ ocorrem dentro do intervalo $(0, 1)$.

Afirmamos então que, sob as condições (3.3), (3.4), (3.5) e (3.6), o polinômio $p_3(x)$

tem exatamente três raízes diferentes dentro do intervalo $(0, 1)$.

De fato, sob as condições (3.3), (3.4), (3.5) e (3.6), pela regra de signos de Descartes (veja, por exemplo, (Griffiths, 1947) e (Xue, 2012)), $p_3(x)$ possui uma ou três raízes positivas. Mas pelas condições (3.3) - (3.6), e pela fórmula de Vieta (veja, (Vinberg, 2003)),

$$\begin{aligned} x_1 + x_2 + x_3 &= \frac{\delta[\delta(\alpha - 3) + 2(\alpha + \beta - 2)]}{\delta^2(\alpha + \beta - 2)}; \\ x_1x_2 + x_2x_3 + x_1x_3 &= -\frac{[2\delta(2 + \delta - \alpha) - (\rho + 1)(\alpha + \beta - 2)]}{\delta^2(\alpha + \beta - 2)}; \\ x_1x_2x_3 &= \frac{-2\delta + (\rho + 1)(\alpha - 1)}{\delta^2(\alpha + \beta - 2)}; \end{aligned}$$

obtemos que a equação polinomial $p_3(x) = 0$ tem exatamente três raízes positivas em $(0, 1)$, denotados por x_1, x_2 e x_3 , e a afirmação segue.

Sem perda de generalidade vamos supor que $x_1 < x_2 < x_3$. Visto que, para $\alpha > 1, \beta > 1$, $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ então $x \rightarrow 0^+$ e $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ então $x \rightarrow 1^-$, segue-se que o gráfico da densidade BBeta (3.1) cresce nos intervalos $(0, x_1)$ e (x_2, x_3) , e diminui em (x_1, x_2) e $(x_3, 1)$. Ou seja, x_1 e x_3 são dois pontos de máximos e x_2 é o único ponto de mínimo. Assim, completamos a prova do teorema. \square

Teorema 3.2.2 (Bimodalidade no caso $\rho = 0$). *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$, $\rho = 0$, $\alpha > 1, \beta > 1$, $\delta > 1$ e*

$$[\delta(\alpha + 1) + \alpha + \beta - 2]^2 > 4\delta(\alpha + \beta)(\alpha - 1), \quad (3.9)$$

então a distribuição BBeta é bimodal.

Demonstração. Quando $\rho = 0$ em (3.7), temos o resultado da derivada a seguir

$$f'(x; \boldsymbol{\theta}_\delta) = \frac{x^{\alpha-2}(1-x)^{\beta-2}(1-\delta x)}{Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)} \left\{ (1-\delta x)[(\alpha-1)(1-x) - (\beta-1)x] - 2\delta x(1-x) \right\}, \quad (3.10)$$

a qual pode ser rescrita como

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

Note que $f'(x; \boldsymbol{\theta}_\delta) = 0$ se e somente se (excluindo os pontos de fronteira) $x = 1/\delta$ e, pela conhecida fórmula de Bhaskara,

$$x_{\pm} = \frac{\delta(\alpha + 1) + \alpha + \beta - 2 \pm \sqrt{[\delta(\alpha + 1) + \alpha + \beta - 2]^2 - 4\delta(\alpha + \beta)(\alpha - 1)}}{2\delta(\alpha + \beta)}.$$

Observe que, pelas condições $\alpha > 1, \beta > 1, \delta > 1$, em (3.10) temos que $f'(x; \boldsymbol{\theta}_\delta) \neq 0$ para todo $x > 1$. Portanto, sob a hipótese (3.9), segue-se que a equação $f'(x; \boldsymbol{\theta}_\delta) = 0$ têm as seguintes três raízes positivas $x = 1/\delta, x_-$ e x_+ no intervalo $(0, 1)$, onde $x_- < x = 1/\delta < x_+$.

Visto que, para $\alpha > 1, \beta > 1, f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ a medida que $x \rightarrow 0^+$ e $f(x; \boldsymbol{\theta}_\delta) \rightarrow 0$ quando $x \rightarrow 1^-$, a bimodalidade da distribuição BBeta é garantida, onde x_- e x_+ são os dois pontos de máximos e $x = 1/\delta$ é o único ponto de mínimo. \square

3.3 Medidas de risco

Nesta seção forneceremos fórmulas fechadas para o f.d.a., função de sobrevivência e função de risco.

Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$, a f.d.a., a função de sobrevivência e a função de risco de X são, respectivamente, dadas por

$$F(x; \boldsymbol{\theta}_\delta) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \left[(1 + \rho) I_x(\alpha, \beta) - 2\delta \frac{B_x(\alpha + 1, \beta)}{B(\alpha, \beta)} + \delta^2 \frac{B_x(\alpha + 2, \beta)}{B(\alpha, \beta)} \right], \quad (3.11)$$

$$S(x; \boldsymbol{\theta}_\delta) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \left[\frac{B(\alpha + i, \beta)}{B(\alpha, \beta)} - \frac{B_x(\alpha + i, \beta)}{B(\alpha, \beta)} \right],$$

$$H(x; \boldsymbol{\theta}_\delta) = \frac{[\rho + (1 - \delta x)^2] x^{\alpha-1} (1 - x)^{\beta-1}}{\sum_{i=0}^2 c_i [B(\alpha + i, \beta) - B_x(\alpha + i, \beta)]}, \quad (3.12)$$

em que $I_x(\alpha, \beta) = B_x(\alpha, \beta)/B(\alpha, \beta)$ é a função Beta incompleta regularizada, $B_x(\alpha, \beta)$ é a função Beta incompleta, $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$.

As devidas demonstrações serão apresentas, mas antes disso, precisamos de estabelecer o seguinte teorema chave.

Teorema 3.3.1. *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ então, para $0 \leq a < b \leq 1$ e $t > -\alpha$,*

$$\mathbb{E}(X^t \mathbf{1}_{\{a \leq X \leq b\}}) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \left[\frac{B_b(\alpha + t + i, \beta)}{B(\alpha, \beta)} - \frac{B_a(\alpha + t + i, \beta)}{B(\alpha, \beta)} \right],$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$.

Demonstração. Ao usar a definição de valor esperado e definição de densidade BBeta,

temos

$$\mathbb{E}(X^t \mathbf{1}_{\{a \leq X \leq b\}}) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \mathbb{E}(Y^{t+i} \mathbf{1}_{\{a \leq Y \leq b\}}),$$

em que $Y \sim \text{BBeta}(\boldsymbol{\theta}_0) = \text{Beta}(\alpha, \beta)$. Uma vez que os momentos de Y são conhecidos e dados por

$$\mathbb{E}(Y^{t+i} \mathbf{1}_{\{a \leq Y \leq b\}}) = \frac{B_b(\alpha + t + i, \beta)}{B(\alpha, \beta)} - \frac{B_a(\alpha + t + i, \beta)}{B(\alpha, \beta)},$$

a prova segue. \square

Tomando $t = 0$, $b = x$ e $a = 0$ no Teorema 3.3.1, obtemos a seguinte expressão fechada para o f.d.a. da distribuição BBeta.

Corolário 3.3.2. *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ então*

$$F(x; \boldsymbol{\theta}_\delta) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \frac{B_x(\alpha + i, \beta)}{B(\alpha, \beta)},$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$.

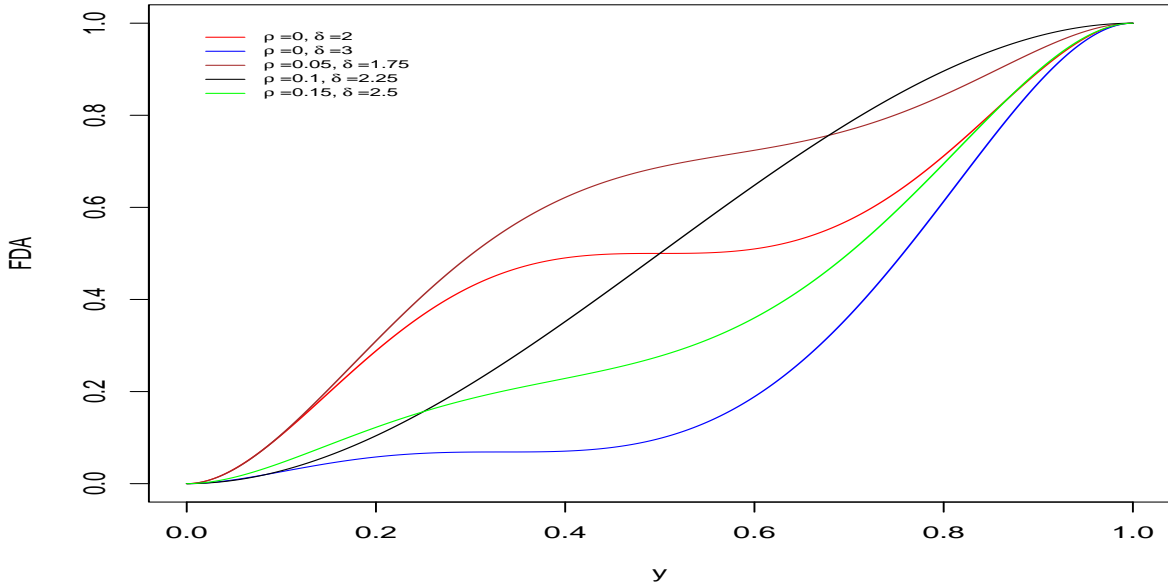


Figura 3.2: Função de distribuição acumulada para diferentes valores dos parâmetros da distribuição Beta com α e $\beta = 2$.

Considerando $t = 0$, $b = 1$ e $a = x$ no Teorema 3.3.1, obtemos a seguinte expressão fechada para a função de sobrevivência da distribuição BBeta.

Corolário 3.3.3. *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ então a função de sobrevivência de X , definida por*

$S(x; \boldsymbol{\theta}_\delta) = 1 - F(x; \boldsymbol{\theta}_\delta)$, é escrita como

$$S(x; \boldsymbol{\theta}_\delta) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \left[\frac{B(\alpha + i, \beta)}{B(\alpha, \beta)} - \frac{B_x(\alpha + i, \beta)}{B(\alpha, \beta)} \right],$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$.

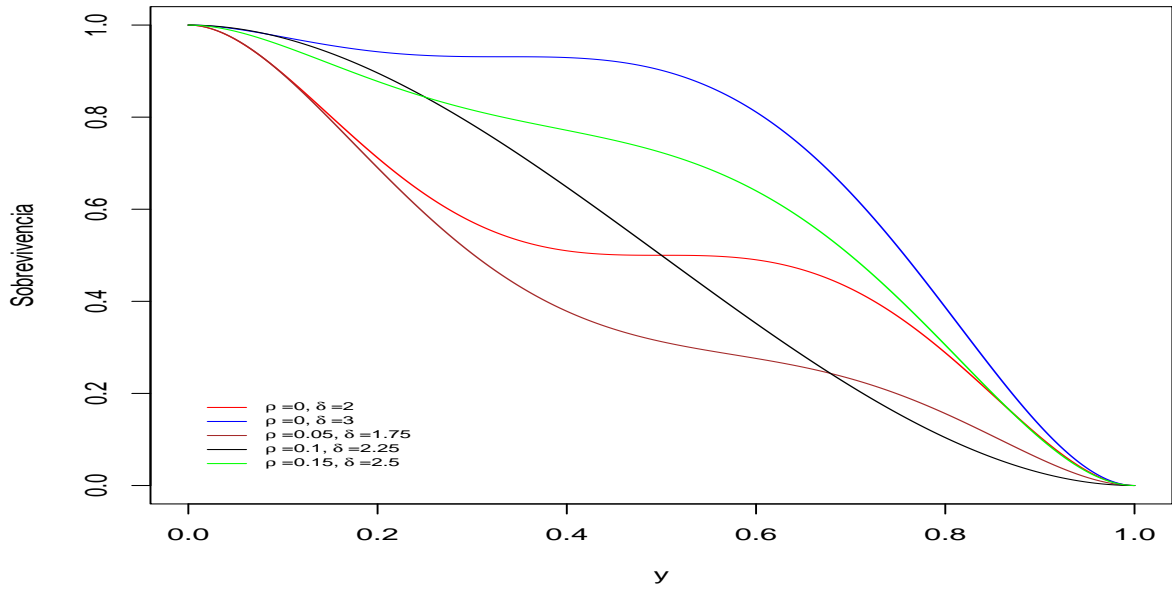


Figura 3.3: Função de Sobrevivência BBeta para diferentes valores dos parâmetros da distribuição com α e $\beta = 2$.

Corolário 3.3.4. Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ então a função de vida média residual de X , definida por $\text{MRL}(x; \boldsymbol{\theta}_\delta) = \int_x^1 S(t; \boldsymbol{\theta}_\delta) dt / S(x; \boldsymbol{\theta}_\delta)$, é escrita como

$$\text{MRL}(x; \boldsymbol{\theta}_\delta) = \frac{\sum_{i=0}^2 c_i \{ [B(\alpha + i + 1, \beta) - xB(\alpha + i, \beta)] - [B_x(\alpha + i + 1, \beta) - xB_x(\alpha + i, \beta)] \}}{\sum_{i=0}^2 c_i [B(\alpha + i, \beta) - B_x(\alpha + i, \beta)]},$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$.

Demonstração. Usando Integração por partes temos

$$\text{MRL}(x, \boldsymbol{\theta}_\delta) = \frac{1}{S(x, \boldsymbol{\theta}_\delta)} \mathbb{E}(X \mathbf{1}_{\{X \geq x\}}) - x. \quad (3.13)$$

Tomando $t = 1$, $a = x$ e $b = 1$ no Teorema 3.3, obtemos

$$\mathbb{E}(X \mathbf{1}_{\{X \geq x\}}) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \left[\frac{B(\alpha + i + 1, \beta)}{B(\alpha, \beta)} - \frac{B_x(\alpha + i + 1, \beta)}{B(\alpha, \beta)} \right].$$

Substituindo a identidade acima em (3.13), a prova segue. \square

Combinando as fórmulas da f.d.a. (3.11) e da definição da f.d.p. (3.1), obtemos a

fórmula (3.12) para a função de risco da função Beta bimodal.

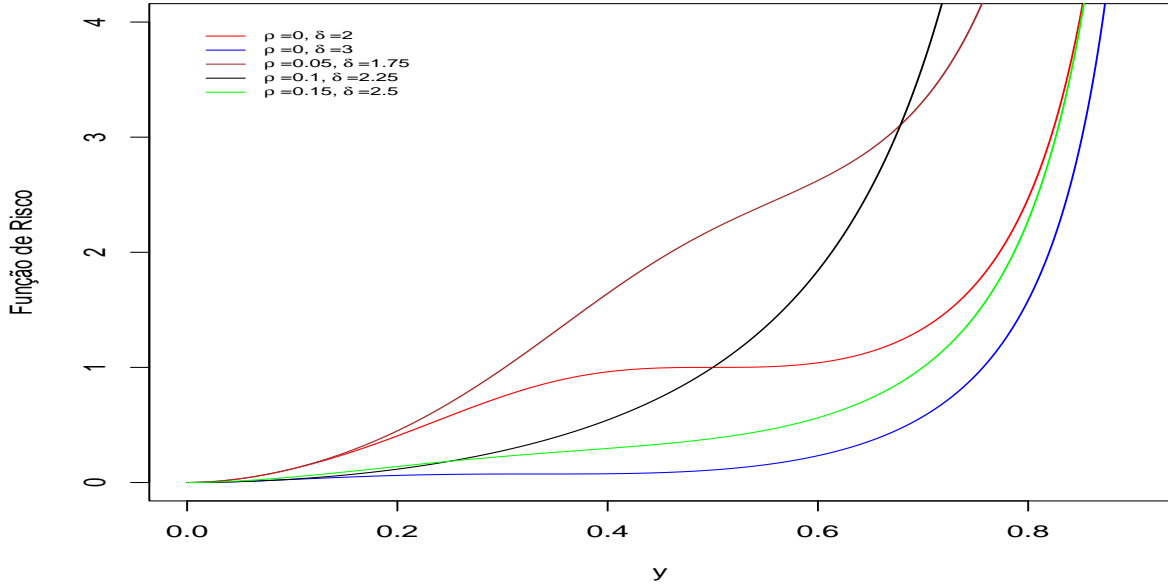


Figura 3.4: Função de risco BBeta para alguns valores dos parâmetros e α e $\beta = 2$.

3.4 Momentos reais

Assim como a distribuição Beta tradicional, a distribuição Beta bimodal apresenta fórmulas fechadas para os momentos, e também, para a função geradora de momentos. Além disso, sobre determinadas restrições nos parâmetros, todos os momentos desta distribuição são finitos. Sendo assim, os resultados desta seção destinam-se a apresentar as fórmulas das medidas relacionadas aos momentos da BBeta.

Corolário 3.4.1 (Momentos reais). *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ e $t > -\alpha$, então*

$$\mathbb{E}(X^t) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \left[(1 + \rho) \frac{B(\alpha + t, \beta)}{B(\alpha, \beta)} - 2\delta \frac{B(\alpha + t + 1, \beta)}{B(\alpha, \beta)} + \delta^2 \frac{B(\alpha + t + 2, \beta)}{B(\alpha, \beta)} \right].$$

Demonstração. Tomando $b = 1$ e $a = 0$ temos o seguinte:

$$\mathbb{E}(X^t) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \frac{B(\alpha + t + i, \beta)}{B(\alpha, \beta)},$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$. □

Corolário 3.4.2 (Momentos). *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ e $k \in [0, +\infty) \cap \mathbb{Z}$, então*

$$\begin{aligned} \mathbb{E}(X^k) &= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \prod_{j=0}^{k+i-1} \frac{\alpha + j}{\alpha + \beta + j} \\ &= \frac{1}{Z(\boldsymbol{\theta}_\delta)} \left(\prod_{j=0}^{k-1} \frac{\alpha + j}{\alpha + \beta + j} \right) \left[1 + \rho - \frac{2\delta(\alpha + k)}{\alpha + \beta + k} + \frac{\delta^2(\alpha + k)(\alpha + k + 1)}{(\alpha + \beta + k)(\alpha + \beta + k + 1)} \right], \end{aligned}$$

em que estamos adotando a convenção de que $\prod_{j=0}^{-1} (\alpha + j) / (\alpha + \beta + j) = 1$.

Demonstração. Tomando $t = k$ no Corolário 3.4.1 e usando a seguintes simples relação de recorrência:

$$B(x + k, y) = B(x, y) \prod_{j=0}^{k-1} \frac{x + j}{x + y + j}. \quad (3.14)$$

temos

$$\mathbb{E}(X^k) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \prod_{j=0}^{k+i-1} \frac{\alpha + j}{\alpha + \beta + j},$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$. Assim, concluímos a prova. \square

Como consequência do Corolário 3.4.2, as expressões fechadas para os momentos, variância, assimetria e curtose da BBeta são facilmente obtidas. As Figuras 3.5 e 3.6 fornecem a assimetria e curtose, respectivamente, para os diferentes valores do parâmetro de bimodalidade δ .

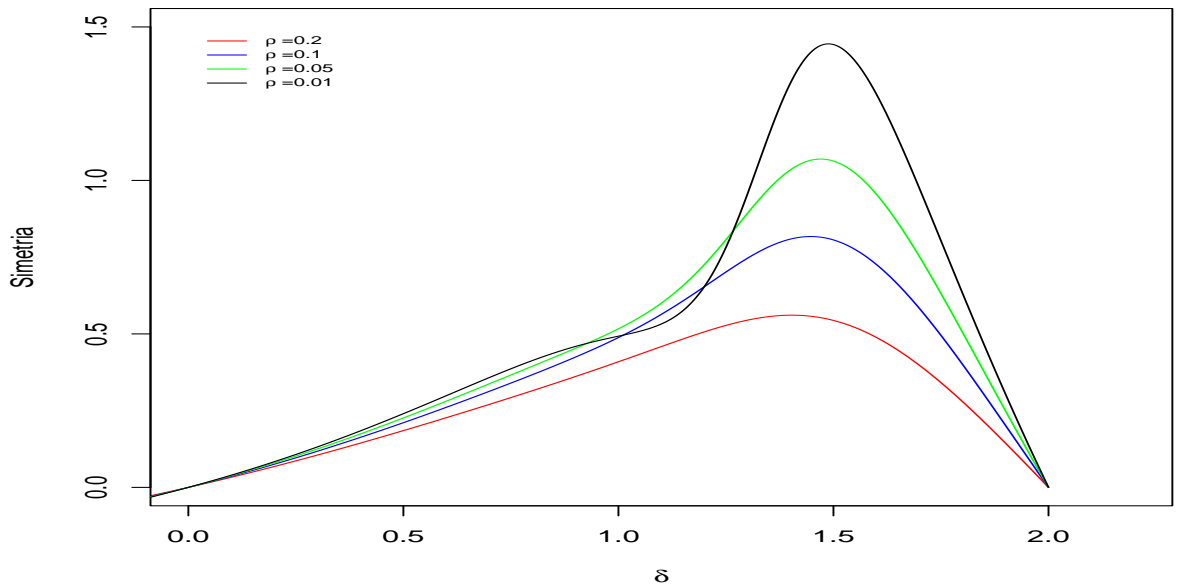


Figura 3.5: Assimetria para alguns valores de ρ , α e β são valores fixos iguais a 2.

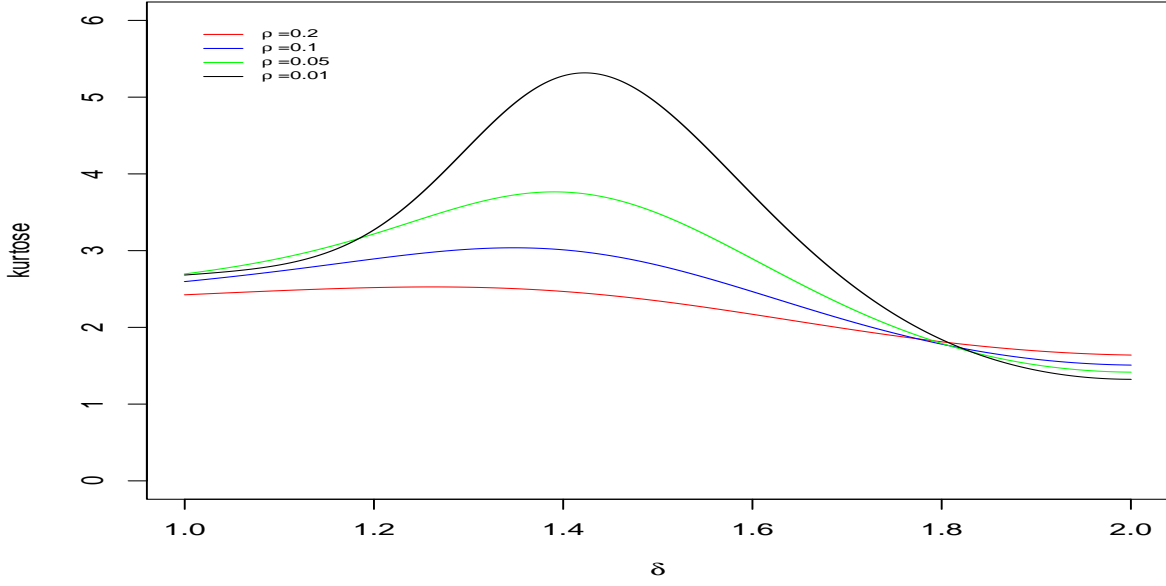


Figura 3.6: Curtose para alguns valores de ρ , α e β são valores fixos iguais a 2.

Considerando α e β fixos de valor 2 é, para alguns valores de ρ , observamos que quando os valores de δ são pouco maiores que 1, obtemos uma grande escala de assimetria e curtose. Podemos concluir que o parâmetro δ contribui para a flexibilidade da função.

Proposição 3.4.3. Tomando $\delta = 0$ nos Corolários 3.4.1 e 3.4.2 obtemos as seguintes fórmulas conhecidas:

$$\mathbb{E}(Y^t) = \frac{B(\alpha + t, \beta)}{B(\alpha, \beta)}, \quad \text{para } t > -\alpha;$$

$$\mathbb{E}(Y^k) = \prod_{j=0}^{k-1} \frac{\alpha + j}{\alpha + \beta + j}, \quad \text{para } k \in [0, +\infty) \cap \mathbb{Z};$$

em que $Y \stackrel{d}{=} X \sim \text{BBeta}(\boldsymbol{\theta}_0) = \text{Beta}(\alpha, \beta)$.

Uma aplicação imediata do Corolário 3.4.2 fornece o seguinte resultado:

Corolário 3.4.4. Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ então

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

$$\mathbb{E}(X^2) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} \left[1 + \rho - \frac{2\delta(\alpha + 2)}{\alpha + \beta + 2} + \frac{\delta^2(\alpha + 2)(\alpha + 3)}{(\alpha + \beta + 2)(\alpha + \beta + 3)} \right].$$

Proposição 3.4.5. *A q -função geradora de momentos (q -FGM) da v.a. $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$, definida por $M_{q,X}(t) = \mathbb{E}[(1 + (1 - q)tX)^{1/(1-q)}]$, pode ser escrita da seguinte forma:*

$$M_{q,X}(t) = \frac{\Gamma(\alpha)\Gamma(\beta) \left[(1 + \rho)H_2\left(\frac{1}{q-1}, \alpha, \alpha + \beta, t(q-1)\right) + A \right]}{B(\alpha, \beta)\Gamma(\alpha + \beta)Z(\boldsymbol{\theta}_\delta)}, \quad q \in [0, 1), t \geq 0,$$

em que $H_2(a, b, c, z)$ denota a função hipergeométrica ${}_2F_1(a, b; c; z)$ e

$$A = \frac{-2(1 + \alpha + \beta)H_2\left(\frac{1}{q-1}, 1 + \alpha, 1 + \alpha + \beta, t(q-1)\right) + (1 + \alpha)\delta H_2\left(\frac{1}{q-1}, 2 + \alpha, 2 + \alpha + \beta, t(q-1)\right)}{(\alpha + \beta)(1 + \alpha + \beta)/(\delta\alpha)}.$$

Usando a regra de L'Hôpital para o cálculo de limites, note que, a medida que $q \rightarrow 1$, a q -FGM da Proposição 3.4.5 se aproxima da Função Geradora de Momentos (FGM).

Corolário 3.4.6 (Função Geradora de Momentos). *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ e $t \geq 0$, então a FGM de X , definida por $M_X(t) = \mathbb{E}[\exp(tX)]$, é dada por*

$$M_X(t) = \Gamma(\alpha + \beta)t^{-1} \left[1 + \rho + \frac{\alpha\delta(\delta(1 + \alpha) - 2(1 + \alpha + \beta))}{(\alpha + \beta)(1 + \alpha + \beta)} \right]^{-1} \\ \times \{ t[(\delta - 1)^2 + \rho] - \beta\delta^2 H_1(\alpha, \alpha + \beta, t) + \beta\delta[\delta(\alpha + \beta) + t(2 - \delta)]H_1(\alpha, 1 + \alpha + \beta, t) \},$$

em que $H_1(a, b, c)$ é a função hipergeométrica confluyente regularizada ${}_1F_1(a; b; z)/\Gamma(b)$.

3.5 Medidas de entropia

Como foi abordado na Seção 1.7, a entropia é uma medida fundamental que informa a quantidade de incerteza em uma amostra que segue uma certa distribuição. Assim, esta seção dedica-se a apresentar expressões matemáticas para a Entropia Tsallis, Entropia Quadrática e para a Entropia de Shannon associadas à distribuição BBeta.

Teorema 3.5.1 (Entropia de Tsallis, Definição 16). *Seja $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$, $\alpha_q = q(\alpha - 1) + 1 > 0$, $\beta_q = q(\beta - 1) + 1 > 0$ e $0 \leq q < 1$. Então*

$$\int_0^1 f^q(x; \boldsymbol{\theta}_\delta) dx \leq \frac{B(\alpha_q, \beta_q)}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^q} \left[1 + q\rho + q\delta^2 \frac{B(\alpha_q + 2, \beta_q)}{B(\alpha_q, \beta_q)} - 2q\delta \frac{B(\alpha_q + 1, \beta_q)}{B(\alpha_q, \beta_q)} \right].$$

Além disso, a igualdade se mantém se e somente se q for suficientemente próximo de 1. Em particular, para $\alpha_q > 0$, $\beta_q > 0$ e $0 \leq q < 1$, a Entropia de Tsallis sempre existe.

Demonstração. Pela definição (3.1) da p.d.f. da distribuição BBeta, temos

$$\int_0^1 f^q(x; \boldsymbol{\theta}_\delta) dx = \frac{1}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^q} \int_0^1 [\rho + (1 - \delta x)^2]^q x^{q(\alpha-1)} (1-x)^{q(\beta-1)} dx. \quad (3.15)$$

Usando a desigualdade (Hardy et al., 1934),

$$a^b \leq 1 + (a-1)b, \quad \text{para } b \in [0, 1], \quad a \geq 1,$$

a expressão do lado direito em (3.15) é no máximo

$$\begin{aligned} & \frac{1}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^q} \int_0^1 \{1 + [\rho - 1 + (1 - \delta x)^2]q\} x^{q(\alpha-1)} (1-x)^{q(\beta-1)} dx \\ &= \frac{B(q(\alpha-1) + 1, q(\beta-1) + 1)}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^q} \mathbb{E}\{1 + [\rho - 1 + (1 - \delta Y)^2]q\} \\ &= \frac{B(q(\alpha-1) + 1, q(\beta-1) + 1)}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^q} [1 + q\rho + q\delta^2\mathbb{E}(Y^2) - 2q\delta\mathbb{E}(Y)], \end{aligned}$$

em que $Y \sim \text{BBeta}(q(\alpha-1) + 1, q(\beta-1) + 1)$. Usando a Proposição 3.4.3 para obtermos os momentos de $\mathbb{E}(Y)$ e $\mathbb{E}(Y^2)$, a prova do teorema segue. \square

Proposição 3.5.2 (Entropia quadrática, Definição 15). *Seja $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ com $\alpha > 1/2$ e $\beta > 1/2$. Então, a Entropia Quadrática é expressada como segue:*

$$H_2(X) = -\log B(2\alpha - 1, 2\beta - 1) + 2\log Z(\boldsymbol{\theta}_\delta) + 2\log B(\alpha, \beta) - \log \sum_{i=0}^4 \tilde{c}_i \prod_{j=0}^{i-1} \frac{\alpha + j}{\alpha + \beta + j},$$

com $\tilde{c}_0 = (1 + \rho)^2$, $\tilde{c}_1 = -4\delta(1 + \rho)$, $\tilde{c}_2 = 2\delta^2(3 + \rho)$, $\tilde{c}_3 = -4\delta^3$ e $\tilde{c}_4 = \delta^4$.

Demonstração. Como $\alpha > 1/2$ e $\beta > 1/2$, usando as definições de densidade f e de esperança, temos

$$\begin{aligned} \int_0^1 f^2(x; \boldsymbol{\theta}_\delta) dx &= \frac{B(2\alpha - 1, 2\beta - 1)}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^2} \int_0^1 [\rho + (1 - \delta x)^2]^2 \frac{x^{2(\alpha-1)} (1-x)^{2(\beta-1)}}{B(2\alpha - 1, 2\beta - 1)} dx \\ &= \frac{B(2\alpha - 1, 2\beta - 1)}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^2} \mathbb{E}\{[\rho + (1 - \delta Y)^2]^2\}, \end{aligned} \quad (3.16)$$

em que $Y \sim \text{Beta}(2\alpha - 1, 2\beta - 1)$. Desenvolvendo o fator quadrático acima,

$$\mathbb{E}\{[\rho + (1 - \delta Y)^2]^2\} = (1 + \rho)^2 - 4\delta(1 + \rho)\mathbb{E}(Y) + 2\delta^2(3 + \rho)\mathbb{E}(Y^2) - 4\delta^3\mathbb{E}(Y^3) + \delta^4\mathbb{E}(Y^4)$$

e substituindo em (3.16), obtemos que

$$\int_0^1 f^2(x; \boldsymbol{\theta}_\delta) dx = \frac{B(2\alpha - 1, 2\beta - 1)}{[Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)]^2} \sum_{i=0}^4 \tilde{c}_i \mathbb{E}(Y^i),$$

com $\tilde{c}_0 = (1 + \rho)^2$, $\tilde{c}_1 = -4\delta(1 + \rho)$, $\tilde{c}_2 = 2\delta^2(3 + \rho)$, $\tilde{c}_3 = -4\delta^3$ e $\tilde{c}_4 = \delta^4$.

Portanto, a partir da Proposição 3.4.3 e da Definição 15 de Entropia Quadrática, a prova segue. \square

O seguinte teorema, o qual nos fornece informação completa sobre o momento logarítmico de uma v.a. com distribuição BBeta, joga um rol de vital importância na obtenção de uma expressão fechada para a Entropia de Shannon.

Lema 3.5.3 (O 1-ésimo momento logarítmico na origem). *Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ é uma v.a. tomando valores em $(0, 1)$, com $\alpha \geq 2$, então*

$$\mathbb{E}[\log(X)] = \frac{1}{Z(\boldsymbol{\theta}_\delta)B(\alpha, \beta)} \sum_{i=0}^2 c_i \frac{\partial B(\alpha + i, \beta)}{\partial \alpha},$$

em que $c_0 = 1 + \rho$, $c_1 = -2\delta$ e $c_2 = \delta^2$, e $\psi(x) = \Gamma'(x)/\Gamma(x)$ é a função Digama.

Demonstração. Usando a definição do valor esperado de uma v.a. com distribuição BBeta, temos

$$\mathbb{E}[\log(X)] = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \mathbb{E}[Y^i \log(Y)], \quad (3.17)$$

com $Y \sim \text{BBeta}(\boldsymbol{\theta}_0) = \text{Beta}(\alpha, \beta)$.

No que segue verificamos a seguinte identidade:

$$\mathbb{E}[Y^i \log(Y)] = \frac{1}{B(\alpha, \beta)} \frac{\partial B(\alpha + i, \beta)}{\partial \alpha}, \quad i = 0, 1, 2, \dots \quad (3.18)$$

Note que a validade de (3.18) junto à identidade (3.17), garante a veracidade do teorema. Portanto, no que resta da prova, mostraremos a identidade (3.18).

De fato, uma vez que $\partial y^{\alpha-1}/\partial \alpha = (\log y)y^{\alpha-1}$, obtemos

$$\begin{aligned} \mathbb{E}[Y^i \log(Y)] &= \frac{1}{B(\alpha, \beta)} \int_0^1 \log(y) y^{\alpha+i-1} (1-y)^{\beta-1} dy \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 \frac{\partial}{\partial \alpha} [y^{\alpha+i-1} (1-y)^{\beta-1}] dy. \end{aligned} \quad (3.19)$$

Observe que as seguintes condições (C1), (C2) e (C3) são verdadeiras:

(C1) A derivada

$$\frac{\partial}{\partial \alpha} [y^{\alpha+i-1}(1-y)^{\beta-1}] = \log(y) y^{\alpha+i-1}(1-y)^{\beta-1}.$$

existe e é contínua em α , para todo y e todo α em um intervalo aberto $S \subset (0, +\infty)$.

(C2) Como $\alpha \geq 2$ e $\alpha \mapsto y^{\alpha+i-1}$ é uma função decrescente, temos

$$\left| \frac{\partial}{\partial \alpha} [y^{\alpha+i-1}(1-y)^{\beta-1}] \right| \leq |\log(y)| y^{i+1}(1-y)^{\beta-1} \quad \text{on } S,$$

onde

$$\begin{aligned} \int_0^1 |\log(y)| y^{i+1}(1-y)^{\beta-1} dy &= - \int_0^1 \log(y) y^{i+1}(1-y)^{\beta-1} dy \\ &\leq - \int_0^1 \left(1 - \frac{1}{y}\right) y^{i+1}(1-y)^{\beta-1} dy \\ &\leq B(i+1, \beta+1) < \infty. \end{aligned}$$

Aqui, na segunda desigualdade acima, usamos a desigualdade $\log(x) \geq 1 - 1/x$, para cada $x > 0$.

(C3) A integral

$$\int_0^1 y^{\alpha+i-1}(1-y)^{\beta-1} dy = B(\alpha+i, \beta)$$

existe em S .

Sob as condições (C1), (C2) e (C3), pela regra da integral de Leibniz, podemos intercambiar a derivada com a integral em (3.19). Logo

$$\mathbb{E}[Y^i \log(Y)] = \frac{1}{B(\alpha, \beta)} \frac{\partial}{\partial \alpha} \int_0^1 y^{\alpha+i-1}(1-y)^{\beta-1} dy = \frac{1}{B(\alpha, \beta)} \frac{\partial B(\alpha+i, \beta)}{\partial \alpha},$$

e a identidade (3.18) é verificada. Portanto, completamos a prova do lema. \square

Note que, usando o Lema 3.5.3, a identidade

$$\frac{\partial \log B(\alpha, \beta)}{\partial \alpha} = \psi(\alpha) - \psi(\alpha + \beta) \quad \text{com } \psi(x) = \Gamma'(x)/\Gamma(x),$$

e a relação de recorrência (3.14), obtemos

$$\mathbb{E}(\log X) = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \sum_{i=0}^2 c_i \left\{ [\psi(\alpha) - \psi(\alpha + \beta)] \prod_{j=0}^{i-1} \frac{\alpha + j}{\alpha + \beta + j} + \frac{\partial}{\partial \alpha} \prod_{j=0}^{i-1} \frac{\alpha + j}{\alpha + \beta + j} \right\}. \quad (3.20)$$

Teorema 3.5.4 (Entropia de Shannon, Definição 14). *Seja $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ uma v.a. tomando valores no intervalo $(0, 1)$, com $\rho = 0$, $\alpha \geq 2$ e $\delta = 1$. Então*

$$\begin{aligned} H_1(X) &= \log \Gamma(\alpha) + \log \Gamma(\beta) - \log \Gamma(\alpha + \beta) + \log \left[1 - 2 \frac{\alpha}{\alpha + \beta} + \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} \right] \\ &\quad - \frac{(\alpha - 1)\beta(\beta + 1)}{(\alpha + \beta)(\alpha + \beta + 1) - 2\alpha(\alpha + \beta + 1) + \alpha(\alpha + 1)} \left[\psi(\alpha) - \psi(\alpha + \beta) - \frac{2(\alpha + \beta) + 1}{(\alpha + \beta)(\alpha + \beta + 1)} \right] \\ &\quad + \frac{(\alpha + \beta)(\alpha + \beta + 1)(\beta + 1)}{(\alpha + \beta)(\alpha + \beta + 1) - 2\alpha(\alpha + \beta + 1) + \alpha(\alpha + 1)} \sum_{i=0}^2 c_i \sum_{k=1}^{\infty} \frac{1}{k} \prod_{j=0}^{k+i-1} \frac{\alpha + j}{\alpha + \beta + j}, \end{aligned}$$

sempre que a série acima converge absolutamente. Aqui, $c_0 = c_2 = 1$ e $c_1 = -2$, e $\psi(x) = \Gamma'(x)/\Gamma(x)$ é a função Digama.

Demonstração. Como $\rho = 0$ e $\delta = 1$, um simples cálculo mostra que

$$\begin{aligned} \int_0^1 f(x; \boldsymbol{\theta}_1) \log f(x; \boldsymbol{\theta}_1) dx &= \mathbb{E}[\log f(X; \boldsymbol{\theta}_1)] \\ &= -\log Z(\boldsymbol{\theta}_1) - \log B(\alpha, \beta) + (\alpha - 1) \mathbb{E}(\log X) + (\beta + 1) \mathbb{E}[\log(1 - X)]. \end{aligned} \quad (3.21)$$

Tomando $c_0 = c_2 = 1$ e $c_1 = -2$ em (3.20), obtemos

$$\mathbb{E}(\log X) = \frac{\beta(\beta + 1)}{Z(\boldsymbol{\theta}_1) (\alpha + \beta)(\alpha + \beta + 1)} \left[\psi(\alpha) - \psi(\alpha + \beta) - \frac{2(\alpha + \beta) + 1}{(\alpha + \beta)(\alpha + \beta + 1)} \right]. \quad (3.22)$$

A seguir, fornecemos uma expressão fechada para a expectativa $\mathbb{E}[\log(1 - X)]$.

De fato, usando a representação em série da função $\log(1 - x)$; também chamada de série Newton-Mercator:

$$\log(1 - x) = - \sum_{k=1}^{\infty} \frac{x^k}{k},$$

a qual converge para $0 < x < 1$; temos que

$$\begin{aligned} \mathbb{E}[\log(1 - X)] &= - \sum_{k=1}^{\infty} \frac{\mathbb{E}(X^k)}{k} \\ &= - \frac{1}{Z(\boldsymbol{\theta}_1)} \sum_{i=0}^2 c_i \sum_{k=1}^{\infty} \frac{1}{k} \prod_{j=0}^{k+i-1} \frac{\alpha + j}{\alpha + \beta + j}, \end{aligned} \quad (3.23)$$

em que na segunda igualdade usamos o Corolário 3.4.2.

Combinando (3.21), (3.22) e (3.23), e usando as definições de constante de normalização $Z(\boldsymbol{\theta}_1)$ e da função Beta, $B(\alpha, \beta)$, a prova segue. \square

3.6 Representação estocástica

Nesta seção apresentamos uma estrutura eficiente sobre como descrever o novo modelo bimodal com outra abordagem inerentemente probabilístico. A abordagem mencionada pode facilitar outras demonstrações e resultados desta nova distribuição.

Dizemos que uma v.a Y tem distribuição Beta limitada no intervalo (a, b) , com parâmetros de forma $\alpha > 0$ e $\beta > 0$, se seu f.d.p. é dada por

$$g(x; \alpha, \beta, a, b) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{B(\alpha, \beta)(b-a)^{\alpha+\beta-1}}, \quad a < x < b.$$

Proposição 3.6.1 (Representação estocástica para $\delta < 0$). *Seja $Y_{k;\alpha,\beta}$ uma v.a que seguindo a distribuição Beta não padrão limitada em $(0, 1/k)$, com $k = 1, 2, 3$ e com parâmetros $\alpha > 0$ e $\beta > 0$. Seja W uma v.a. discreta assumindo valores no conjunto finito $\{1, 2, 3\}$ e com funções de probabilidades dadas por*

$$\begin{aligned} \pi_1 &= \mathbb{P}(W = 1) = \frac{1 + \rho}{Z(\boldsymbol{\theta}_\delta)}, \\ \pi_2 &= \mathbb{P}(W = 2) = -\frac{2\alpha\delta}{Z(\boldsymbol{\theta}_\delta)(\alpha + \beta)}, \quad \delta < 0, \\ \pi_3 &= \mathbb{P}(W = 3) = \frac{\alpha(\alpha + 1)\delta^2}{Z(\boldsymbol{\theta}_\delta)(\alpha + \beta)(\alpha + \beta + 1)}. \end{aligned}$$

Uma simples manipulação algébrica mostra que $\pi_1 + \pi_2 + \pi_3 = 1$.

Assumamos que

$$Y = \sum_{k=1}^3 Y_{k;\alpha+k-1,\beta} \delta_{W,k} \delta_{W,l}, \quad l = 1, 2, 3,$$

e que W é independente de $Y_{k;\alpha,\beta}$, para cada $k = 1, 2, 3$, onde $\delta_{W,k}$ denota a função delta de Kronecker; ou seja, $\delta_{W(\omega),k}$ é 1 se $W(\omega) = k$ para $\omega \in \Omega$, e é 0 caso contrário.

Se $X = WY$ então $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$. Reciprocamente, se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ então $X = WY$.

Demonstração. Pela lei da probabilidade total e pela independência das variáveis envolvidas, obtemos

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(WY \leq x) = \sum_{l=1}^3 \mathbb{P}(WY \leq x | W = l) \mathbb{P}(W = l) \\ &= \sum_{l=1}^3 \mathbb{P}(lY_{l;\alpha+l-1,\beta} \leq x) \mathbb{P}(W = l) = \sum_{l=1}^3 \mathbb{P}(Y_{1;\alpha+l-1,\beta} \leq x) \pi_l, \end{aligned}$$

pois $lY_{l;\alpha+l-1,\beta} = Y_{1;\alpha+l-1,\beta} \sim B(\alpha + l - 1, \beta)$, para $l = 1, 2, 3$. Uma vez que para $Y \sim \text{Beta}(\alpha, \beta)$ sua f.d.a. é dada por $F(x; \alpha, \beta) = I_x(\alpha, \beta)$, $0 \leq x \leq 1$, pela definição das probabilidades π_l 's, a expressão acima é

$$= \sum_{l=1}^3 I_x(\alpha + l - 1, \beta) \pi_l = \frac{1}{Z(\boldsymbol{\theta}_\delta)} \left[(1 + \rho) I_x(\alpha, \beta) - 2\delta \frac{B_x(\alpha + 1, \beta)}{B(\alpha, \beta)} + \delta^2 \frac{B_x(\alpha + 2, \beta)}{B(\alpha, \beta)} \right].$$

Mas por (3.11), o lado direito é igual à f.d.a. $F(x; \boldsymbol{\theta}_\delta)$ do modelo BBeta. Isso completa a prova. \square

3.7 Identificabilidade da distribuição Beta bimodal

Para caracterizar a identificabilidade da distribuição Beta bimodal, uma relação entre a Beta bimodal e a Beta tradicional é mostrada nesta seção. Vamos supor que $f(x; \alpha, \beta)$ é o f.d.p. da distribuição Beta, onde $\alpha > 0$ e $\beta > 0$ são os parâmetros de forma. Uma observação simples mostra que a f.d.p. $f(x; \boldsymbol{\theta}_\delta)$ da Beta bimodal, com vetor de parâmetros $\boldsymbol{\theta}_\delta = (\alpha, \beta, \rho, \delta)$, pode ser escrita como uma mistura finita (generalizada) de três distribuições Beta com diferentes parâmetros de forma. Em outras palavras,

$$f(x; \boldsymbol{\theta}_\delta) = \pi_1 f(x; \alpha, \beta) + \pi_2 f(x; \alpha + 1, \beta) + \pi_3 f(x; \alpha + 2, \beta), \quad 0 \leq x \leq 1, \quad (3.24)$$

em que π_1 , π_2 e π_3 são constantes (que dependem apenas de $\boldsymbol{\theta}_\delta$) dadas na Proposição 3.6.1.

A diferença da Proposição 3.6.1, aqui δ pode ser negativo. Em princípio, misturar pesos não-negativos nem sempre é necessário, pois misturas podem resultar em uma f.d.p mesmo que alguns pesos sejam negativos.

Seja \mathcal{B} a família de distribuições Beta, como segue:

$$\mathcal{B} = \left\{ F : F(x; \alpha, \beta) = \int_0^x f(y; \alpha, \beta) dy, \quad \alpha > 0, \beta > 0, \quad 0 \leq x \leq 1 \right\}. \quad (3.25)$$

Escreva $\mathcal{H}_{\mathcal{B}}$ para denotar à classe de todas as misturas finitas de \mathcal{B} . É bem sabido que a classe $\mathcal{H}_{\mathcal{B}}$ é identificável (este fato é obtido como consequência do principal resultado de (Atienza et al., 2006); veja Proposição .1.2 no Apêndice). O seguinte resultado prova a identificabilidade da distribuição Beta bimodal.

Proposição 3.7.1. *A aplicação $\boldsymbol{\theta}_\delta \mapsto f(\cdot; \boldsymbol{\theta}_\delta)$ é injetiva.*

Demonstração. Suponhamos que $f(x; \boldsymbol{\theta}_\delta) = f(x; \boldsymbol{\theta}'_\delta)$ para todo $0 \leq x \leq 1$. Em outras palavras, por (3.24),

$$\begin{aligned} \pi_1 f(x; \alpha, \beta) + \pi_2 f(x; \alpha + 1, \beta) + \pi_3 f(x; \alpha + 2, \beta) \\ = \pi'_1 f(x; \alpha', \beta') + \pi'_2 f(x; \alpha' + 1, \beta') + \pi'_3 f(x; \alpha' + 2, \beta'). \end{aligned}$$

Como \mathcal{H}_B é identificável, temos que $\pi_i = \pi'_i$, para $i = 1, 2, 3$, e $\alpha = \alpha'$, $\beta = \beta'$. Portanto, das igualdades $\pi_i = \pi'_i$, $i = 1, 2, 3$, segue imediatamente que $\rho = \rho'$ e $\delta = \delta'$. Portanto, $\boldsymbol{\theta}_\delta = \boldsymbol{\theta}'_\delta$, e a prova segue. \square

Capítulo 4

Modelo de regressão, estimativa e análise diagnóstica

Para a aplicação no estudo de relação entre variáveis, o modelo de regressão baseada na distribuição Beta bimodal é introduzido neste capítulo. Por ser um modelo paramétrico, é possível calcular a função de verossimilhança e compará-lo a outros modelos.

Sejam X_1, \dots, X_n , n variáveis aleatórias independentes, em que cada X_i , $i = 1, \dots, n$, segue a f.d.p. BBeta dada em (3.1) com vetor de parâmetros $(\alpha_i, \beta_i, \rho, \delta)$. Assumimos que os parâmetros α_i e β_i satisfazem as seguintes relações funcionais:

$$g_1(\alpha_i) = \eta_{1i} = \mathbf{w}_i^\top \boldsymbol{\gamma} \quad \text{e} \quad g_2(\beta_i) = \eta_{2i} = \mathbf{z}_i^\top \boldsymbol{\zeta}, \quad i = 1, \dots, n, \quad (4.1)$$

em que $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^\top$ e $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_q)^\top$ são vetores de coeficientes de regressão desconhecidos que são considerados funcionalmente independentes, $\boldsymbol{\gamma} \in \mathbb{R}^p$ e $\boldsymbol{\zeta} \in \mathbb{R}^q$, com $p + q < n$, η_{1i} e η_{2i} são os preditores lineares, e $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})^\top$ e $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$ são observações em p e q regressores conhecidos, para $i = 1, \dots, n$. Além disso, assumimos que as matrizes de covariáveis $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^\top$ e $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^\top$ têm dimensão p e q , respectivamente. As funções de ligação $g_1 : \mathbb{R} \rightarrow \mathbb{R}^+$ e $g_2 : \mathbb{R} \rightarrow \mathbb{R}^+$ em (4.1) devem ser estritamente monótonas, positivas e pelo menos duas vezes diferenciável, de forma que $\alpha_i = g_1^{-1}(\mathbf{w}_i^\top \boldsymbol{\gamma})$ e $\beta_i = g_2^{-1}(\mathbf{z}_i^\top \boldsymbol{\zeta})$, com $g_1^{-1}(\cdot)$ e $g_2^{-1}(\cdot)$ sendo as funções inversas de $g_1(\cdot)$ e $g_2(\cdot)$, respectivamente.

A função log-verossimilhança para $\boldsymbol{\theta}_\delta = (\boldsymbol{\gamma}, \boldsymbol{\zeta}, \rho, \delta)$ com base em uma amostra de n observações independentes é dada por

$$\ell(\boldsymbol{\theta}_\delta) = \sum_{i=1}^n \ell(\alpha_i, \beta_i, \rho, \delta), \quad (4.2)$$

em que

$$\begin{aligned} \ell(\alpha_i, \beta_i, \rho, \delta) &= -\log Z(\boldsymbol{\theta}_\delta) - \log B(\alpha_i, \beta_i) \\ &\quad + \log [\rho + (1 - \delta x_i)^2] + (\alpha_i - 1) \log x_i + (\beta_i - 1) \log(1 - x_i) \end{aligned}$$

e $Z(\boldsymbol{\theta}_\delta)$ é a constante de normalização dada em (3.2).

O estimador de máxima verossimilhança (EMV) $\hat{\boldsymbol{\theta}}_\delta = (\hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\zeta}}^\top, \hat{\rho}, \hat{\delta})^\top$ de $\boldsymbol{\theta}_\delta = (\boldsymbol{\gamma}^\top, \boldsymbol{\zeta}^\top, \rho, \delta)^\top$ é obtido pela maximização da função log-verossimilhança (4.2). No entanto, não é possível derivar solução analítica para o EMV $\hat{\boldsymbol{\theta}}$, portanto devemos requerer a solução numérica usando algum algoritmo de otimização, como Newton-Raphson e quase-Newton.

Sob clássicas condições de regularidade, a medida que n cresce, a distribuição assintótica do EMV $\hat{\boldsymbol{\theta}}_\delta = (\hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\zeta}}^\top, \hat{\rho}, \hat{\delta})^\top$ tem comportamento assintótico normal multivariada (de dimensão $p + q + 2$) com vetor valor esperado $\boldsymbol{\theta}_\delta = (\boldsymbol{\gamma}^\top, \boldsymbol{\zeta}^\top, \rho, \delta)^\top$ e matriz de variância-covariância $\mathbf{K}^{-1}(\boldsymbol{\theta}_\delta)$ onde

$$\mathbf{K}(\boldsymbol{\theta}_\delta) = \mathbb{E} \left[- \frac{\partial \ell(\boldsymbol{\theta}_\delta)}{\partial \boldsymbol{\theta}_\delta \partial \boldsymbol{\theta}_\delta^\top} \right]$$

é a matriz de informação esperada de Fisher. Infelizmente, não existe uma expressão de forma fechada para a matriz $\mathbf{K}(\boldsymbol{\theta}_\delta)$. No entanto, um estimador consistente da matriz de informação esperada de Fisher é dado por

$$\mathbf{J}(\hat{\boldsymbol{\theta}}_\delta) = - \frac{\partial \ell(\boldsymbol{\theta}_\delta)}{\partial \boldsymbol{\theta}_\delta \partial \boldsymbol{\theta}_\delta^\top} \Big|_{\boldsymbol{\theta}_\delta = \hat{\boldsymbol{\theta}}_\delta},$$

que é a matriz de informação de Fisher observada. Portanto, para n grande, podemos substituir $\mathbf{K}(\boldsymbol{\theta}_\delta)$ por $\mathbf{J}(\hat{\boldsymbol{\theta}}_\delta)$.

Seja θ_{δ_r} o r -ésimo componente de $\boldsymbol{\theta}_\delta$. O intervalo de confiança assintótico $100(1-\varphi)\%$ para θ_{δ_r} é dado por

$$\hat{\theta}_r \pm z_{\varphi/2}(\hat{\theta}_{\delta_r}), \quad r = 1, \dots, p + q + 2,$$

onde $z_{\varphi/2}$ é o quantil superior $\varphi/2$ da distribuição normal padrão e $(\hat{\theta}_{\delta_r})$ é o erro padrão assintótico de $\hat{\theta}_{\delta_r}$. Observe que $(\hat{\theta}_{\delta_r})$ é a raiz quadrada do r -ésimo elemento diagonal da matriz $\mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}_\delta)$.

Os resíduos são amplamente utilizados para verificar o ajuste do modelo. Para verificar a qualidade do ajuste do modelo BBeta, propomos usar os resíduos de quantis randomizados introduzido na referência (Dunn and Smyth, 1996). Seja $F(x_i; \boldsymbol{\theta}_\delta)$ a f.d.a. do modelo BBeta, conforme definido em (3.11), em que a estrutura do modelo de regressão

é assumida como em (4.1). O quantil residual randomizado é dado por

$$r_i = \Phi^{-1}(F(x_i; \hat{\boldsymbol{\theta}}_\delta)), \quad i = 1, \dots, n,$$

em que $\Phi^{-1}(\cdot)$ denota a função inversa da f.d.a. da distribuição normal padrão. Se o modelo assumido para os dados for bem ajustado, esses resíduos têm distribuição normal padrão (Dunn and Smyth, 1996).

Capítulo 5

Estudo de simulação

Neste capítulo, simulações de Monte Carlo são realizadas (i) para avaliar o comportamento das estimativas de máxima verossimilhança dos coeficientes de regressão considerando diferentes tamanhos das amostras, (ii) investigar a distribuição empírica dos resíduos dos quantis randomizados.

Os experimentos de Monte Carlo foram realizados considerando a seguinte estrutura de regressão:

$$\begin{aligned}\log(\alpha_i) &= \gamma_0 + \gamma_1 z_i, \\ \log(\beta_i) &= \zeta_0 + \zeta_1 z_i, \quad i = 1, \dots, n.\end{aligned}$$

Os estudos de simulação foram realizadas a partir de amostras de tamanho $n = 50, 100, 200$ e 300 , e com os seguintes valores dos parâmetros $\gamma_0 = -1.8, \gamma_1 = 5.9, \zeta_0 = 3.8, \zeta_1 = -2.4, \rho = 0.1$ e $\delta = 2.4$, pois esses valores serão usados na aplicação da regressão no próximo capítulo. Para cada um desses cenários foram simuladas 5000 amostras e os valores da covariável de z_i foram gerados a partir da distribuição uniforme padrão. Todas as simulações foram conduzidas em R usando o Algoritmo de Broyden-Fletcher-Goldfarb-Shanno (BFGS) disponível na função `optim`.

5.1 Estimativas dos parâmetros

Para cada tamanho de amostra, os valores empíricos para o viés e Erro Quadrático Médio (EQM) dos estimadores de $\gamma_0, \gamma_1, \zeta_0, \zeta_1, \rho$ e δ pelo método da máxima verossimilhança são apresentados na Tabela 5.1 e pela abordagem do Box-plot na Figura 5.1.

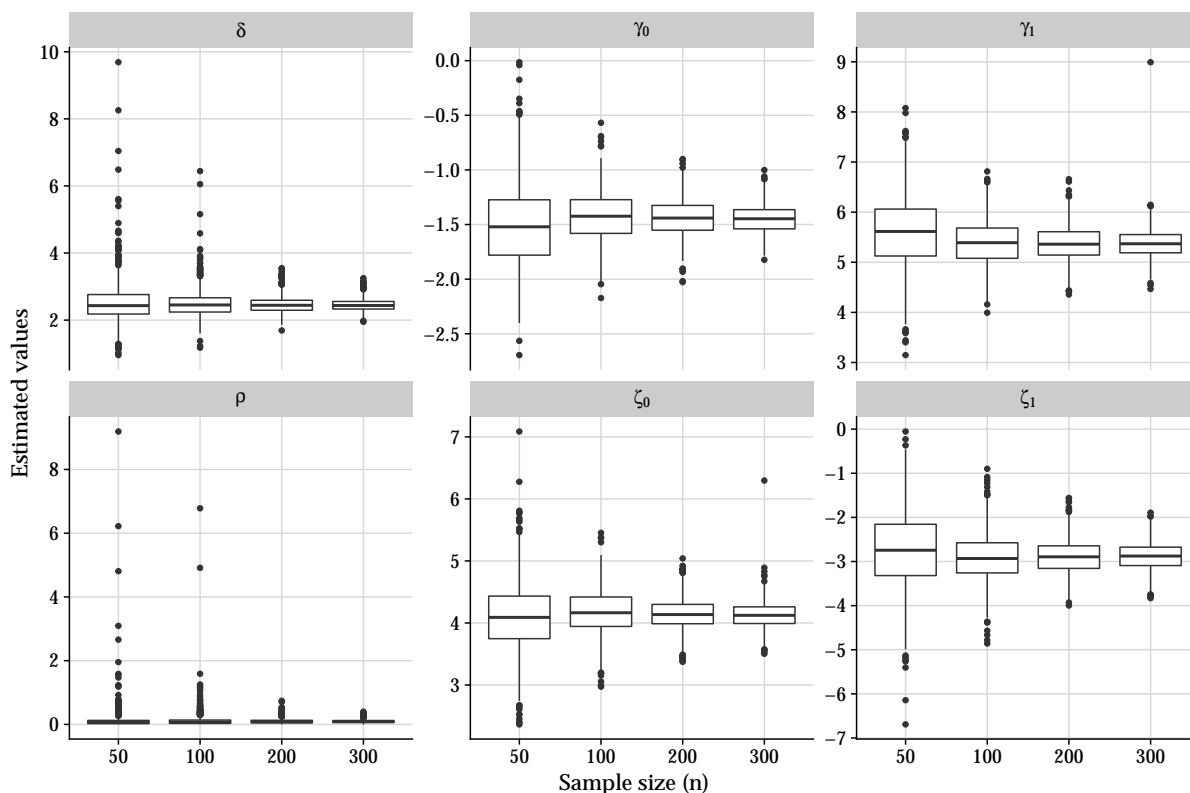
Com base na tabela, descobrimos que as estimativas são convergentes para seus valores. Como esperado, aumentar o tamanho da amostra reduz substancialmente tanto viés e o EQM. As descobertas anteriores são confirmadas pelos gráficos de Box-plot mostrados na Figura 5.1.

Tabela 5.1: Viés estimado e erro quadrático médio.

n	Viés						EQM					
	γ_0	γ_1	ζ_0	ζ_1	δ	ρ	γ_0	γ_1	ζ_0	ζ_1	δ	ρ
50	0.212	0.106	0.132	0.299	0.177	1.306	0.234	0.634	0.417	0.839	0.488	0.235
100	0.213	0.099	0.114	0.254	0.120	0.938	0.192	0.475	0.276	0.558	0.183	0.091
200	0.202	0.093	0.095	0.215	0.081	0.543	0.157	0.390	0.181	0.381	0.068	0.006
300	0.195	0.091	0.088	0.200	0.061	0.414	0.139	0.353	0.152	0.313	0.037	0.003

Fonte: (Vila et al., 2021b).

Figura 5.1: Boxplots dos parâmetros estimados obtidos em experimentos de Monte Carlo para diferentes tamanhos de amostra



Fonte: (Vila et al., 2021b).

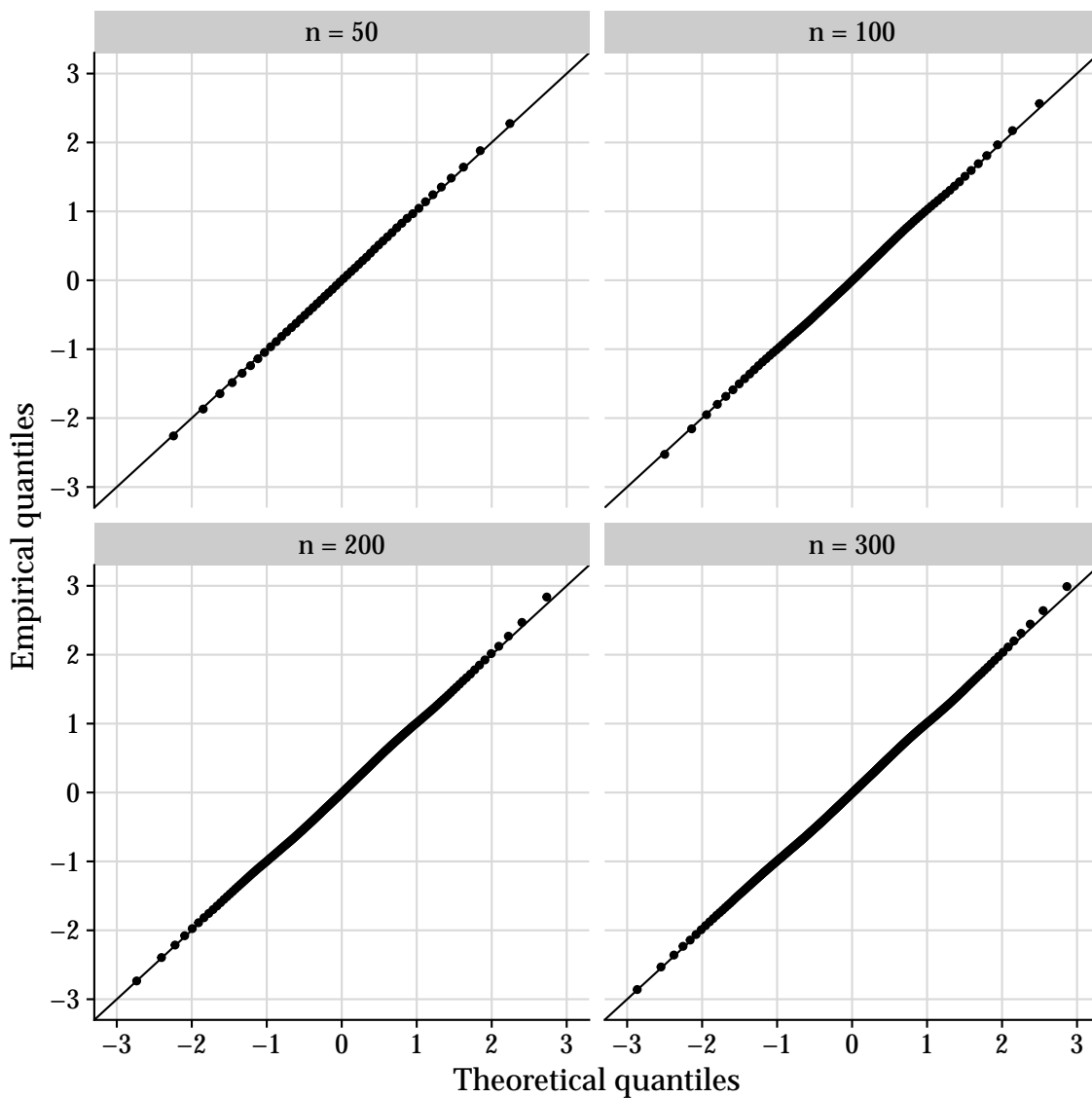
5.2 Resíduos

O estudo seguinte de simulação foi realizado para averiguar quão bem as distribuições dos resíduos de quantis randomizados são aproximadas pela distribuição normal padrão. A avaliação dos resíduos dos quantis randomizados foi baseada nos gráficos de

probabilidade normal das estatísticas de ordem, média e medidas descritivas. Os resultados são apresentados na Tabela 5.2 e na Figura 5.2.

Na Tabela 5.2, apresentamos a média, desvio padrão (DP), assimetria e curtose do resíduo de quantis randomizados. Para todos os cenários os resíduos têm aproximadamente média zero e desvio padrão 1, têm assimetria próxima a zero e a curtose é próxima a três.

Figura 5.2: Gráficos de probabilidade normal das estatísticas de ordem média.



Fonte: (Vila et al., 2021b).

A Figura 5.2 exibe os quantílicos amostrais versus quantílicos teóricos dos resíduos. Nota-se na Figura 5.2 que os resíduos dos quantis randomizados seguem aproximadamente uma distribuição normal padrão. O que indica um ajuste satisfatório.

Tabela 5.2: Medidas descritivas dos resíduos de quantis randomizados.

n	Média	DP	Assimetria	Curtose
50	- 0.001	0.999	0.028	2.854
100	- 0.002	0.999	0.054	2.976
200	- 0.003	0.997	0.077	3.002
300	- 0.003	0.997	0.084	3.025

Fonte: (Vila et al., 2021b).

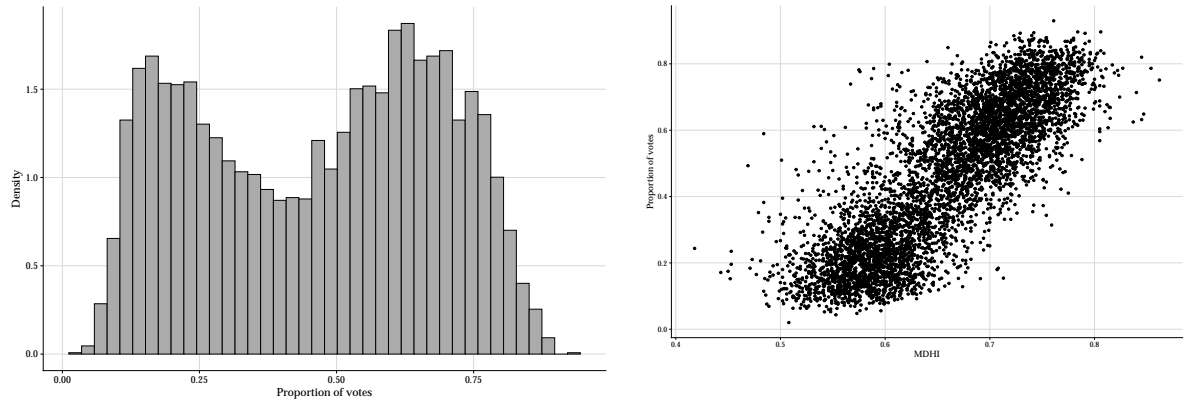
Capítulo 6

Aplicação de dados reais

Neste capítulo será apresentado uma aplicação em dados reais do modelo beta bimodal proposto em esta dissertação. Os dados utilizados nesta aplicação são da proporção de votos que Jair Bolsonaro recebeu no segundo turno das eleições brasileiras de 2018, na qual se encaixou na proposta de este trabalho devido ao seu comportamento bimodal. Para avaliar a aplicabilidade do modelo, comparamos a potencialidade da regressão BBeta com o modelo de regressão Beta tradicional. Para estimar os parâmetros do modelo, adotamos o método EMV, conforme discutido no Capítulo 4. Os erros padrão assintóticos e os intervalos de confiança foram calculados a partir da matriz de informação observada. As avaliações numéricas necessárias para a análise de dados foram implementadas usando o software R.

O objetivo desta análise de dados é descrever a proporção de votos que Jair Bolsonaro recebeu no segundo turno das eleições brasileiras em 2018 para todas as 5.565 cidades dado uma variável explicativa. A variável de resposta X_i é a proporção de votos dado o desenvolvimento humano municipal (mhdi). A Figura 6.1 traça o histograma da variável de resposta usada na aplicação e nos gráficos de dispersão do desenvolvimento humano municipal contra proporção de votos. Na Figura 6.1, podemos ver que a variável de resposta tem bimodalidade. Além disso, há evidências de uma tendência da proporção de votos com o aumento do desenvolvimento humano municipal.

Figura 6.1: Gráficos empíricos de dados



Fonte: (Vila et al., 2021b).

Para explicar essa proporção de votos, consideramos o modelo de regressão Beta bimodal, definido como $Y_i \sim \text{BBeta}(\boldsymbol{\theta}_i)$,

$$\begin{aligned}\log(\alpha_i) &= \gamma_0 + \gamma_1 \text{mhdi}_i, \\ \log(\beta_i) &= \zeta_0 + \zeta_1 \text{mhdi}_i,\end{aligned}$$

em que $i = 1, 2, \dots, 5.565$ cidades e mhdi_i é o desenvolvimento humano municipal das cidades i . Para fins de comparação, o modelo de regressão Beta foi ajustado, assumindo que $Y_i \sim \text{Beta}(\alpha_i, \beta_i)$,

$$\begin{aligned}\text{logit}\left(\frac{\alpha_i}{\alpha_i + \beta_i}\right) &= \gamma_0 + \gamma_1 \text{mhdi}_i, \\ \log(\alpha_i + \beta_i) &= \zeta_0 + \zeta_1 \text{mhdi}_i.\end{aligned}$$

A Tabela 6.1 mostra os parâmetros estimados, erro padrão, limites dos intervalos de confiança com nível de significância de 5% sob os modelos BBeta e Beta. Observe que os coeficientes são estatisticamente significativos no nível de 5%, para os modelos de regressão BBeta e Beta com a estrutura acima.

Tabela 6.1: Estimativas de ML, erros padrão e intervalo de confiança de 95 %.

Modelo	Parâmetro	Estimativa	S.E.	2,5 %	97,5 %
BBeta	γ_0	- 1,8999	0,1963	- 2,2846	- 1,5152
	γ_1	5,9471	0,3044	5,3505	6,5437
	ζ_0	3,8341	0,1915	3,4587	4,2095
	ζ_1	- 2,4232	0,2862	- 2,9842	- 1,8622
	ρ	0,1096	0,0090	0,0920	0,1273
	δ	2,4092	0,0351	2,3405	2,4780
Beta	γ_0	- 7,5343	0,0749	- 7,6810	- 7,3875
	γ_1	11,1820	0,1105	10,9654	11,3987
	ζ_0	1,0029	0,1675	0,6746	1,3312
	ζ_1	2,5214	0,2528	2,0260	3,0169

Fonte: (Vila et al., 2021b).

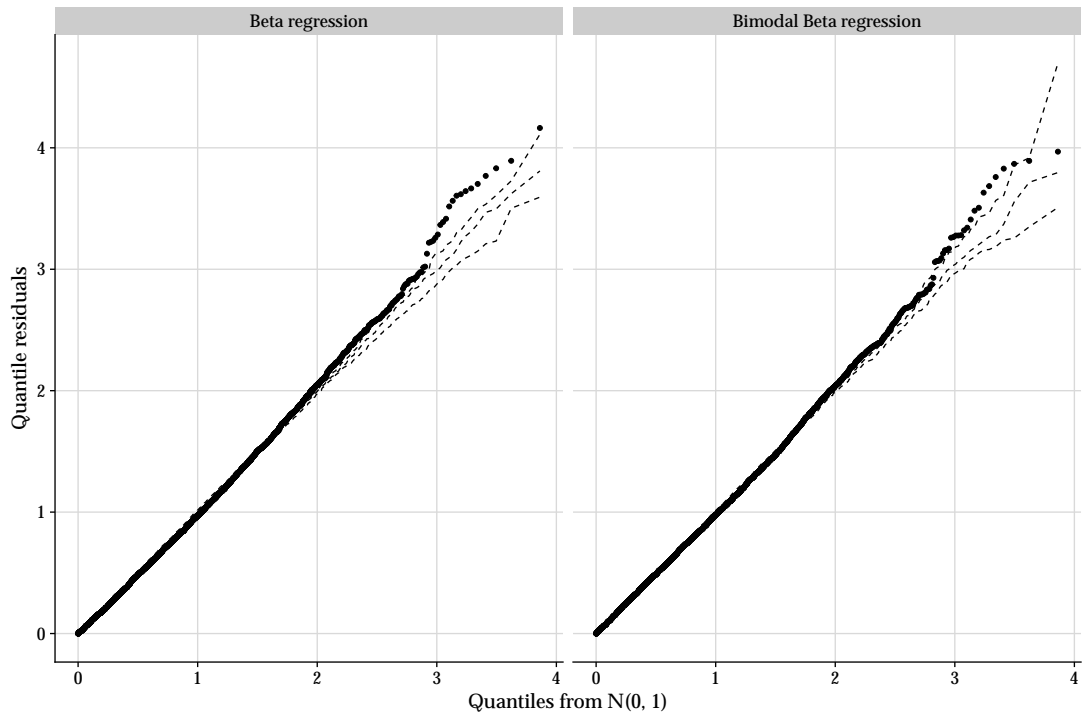
A Tabela 6.2 mostra o critério de informação de Akaike (AIC), o critério de informação Bayesiano (BIC) e a estatística de Kolmogorov-Smirnov (KS) para os modelos ajustados. Em geral, é esperado que o melhor modelo para ajustar os dados apresente os menores valores para essas grandezas (AIC e BIC). Com base nos critérios AIC e BIC, o modelo BBeta parece ser mais razoável em comparação com o modelo Beta no ajuste desses dados. Os resultados dos gráficos de resíduos na Figura 6.2 são condizentes com a afirmação apresentada pela abordagem dos critérios AIC e BIC.

Tabela 6.2: Medidas de adequação.

Modelo	KS	AIC	BIC
Beta	0,0203 (0,2014)	- 8238	- 8212
BBeta	0,0149 (0,5659)	- 8786	- 8746

Fonte: (Vila et al., 2021b).

Figura 6.2: Gráfico Half-normal dos resíduos de quantis randomizados com envelope simulado.



Fonte: (Vila et al., 2021b).

Conclusão

Nesta dissertação, a conhecida distribuição Beta de dois parâmetros é estendida pela introdução de dois parâmetros extras, definindo-se assim a distribuição Beta bimodal (BBeta), com base em uma técnica de transformação quadrática usada para gerar funções bimodais. Ao longo deste trabalho foram apresentadas diversas propriedades da distribuição Beta bimodal. Foi matematicamente descrito em que situações a distribuição assume forma unimodal ou bimodal e em quais intervalos a função cresce e decresce. Foi visto que é possível obter fórmulas fechadas para os momentos reais 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 as entropias de Shannon e Quadrática. Já em relação as medidas de sobrevivência, encontramos expressões para a função de sobrevivência, f.d.a., 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 de máximo da função de verossimilhança. Adquirimos uma estrutura de regressão para os parâmetros α e β . A estimativa dos parâmetros do modelo foi aplicada pela abordagem da máxima verossimilhança, seu bom desempenho foi avaliado por meio de simulações de Monte Carlo, na qual os estimadores se mostraram consistentes, visto que seu viés e variância diminuem na medida em que o número de amostras aumenta. Além disso, foi proposto resíduos para o modelo e conduzimos um estudo de simulação para estabelecer suas propriedades empíricas para avaliar seus desempenhos. O modelo proposto foi ajustado à proporção de votos que Jair Bolsonaro recebeu no segundo turno das eleições brasileiras de 2018. O modelo BBeta mostrou um desempenho um pouco melhor que o modelo de regressão baseado na Beta clássica. Portanto, a distribuição Beta bimodal mostrou ser uma distribuição bimodal flexível que pode ser aplicado em dados reais oriundos de vários campos do conhecimento. Em trabalhos futuros pode-se buscar aplicações para esta nova distribuição.

Bibliografia

- Ahmed, 2008 Ahmed, S. E.; Hussein, A. (2008). *Gamma Mixture: Bimodality, Inflexions, and L-Moments*. Communications in Statistics, **37**, pp. 1147-116.
- Atienza et al., 2006 Atienza, N.; Garcia-Heras, J.; Muñoz-Pichardo, J. M. (2006). *A new condition for identifiability of finite mixture distributions*. Metrika, **63**, p.p. 215–221.
- Casella and Berger, 2010 Casella, G.; Berger, R. L.: Inferência Estatística, 2.ed. São Paulo: Cengage Learning, (2010)
- Dunn and Smyth, 1996 Dunn, P. K.; Smyth, G. K., 1996. Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5, 236–244.
- Elal-Olivero, 2010 Elal-Olivero, D. (2010). *Alpha-skew-normal distribution*. Addison Wesley, Proyecciones Journal of Mathematics, **29**, pp. 224-240.
- Gomez, 2019 Gómez, Y. M.; Gallardo, D. I. (2019). *An Asymmetric Bimodal Distribution with Application to Quantile Regression*. Symmetry, **11**, 899.
- Griffiths, 1947 Griffiths, L. (1947). Introduction to the Theory of Equations. J. Wiley.
- Hardy et al., 1934 Hardy, G. H., Littlewood, J. E., Pólya, G. (1934). Inequalities. Cambridge University Press, Cambridge.
- Hsu, 2008 Hsu, Y.; Pearn, W.; Wu, P. (2008). *Capability adjustment for gamma processes with mean shift consideration in implementing six sigma program*. European Journal of Operational Research, **119**, pp. 517-529.
- Jonhson et al., 1993 Jonhson, N. L.; Kotz, S.; Kemp, A.: Univariate Discrete Distributions. New York: Wiley, (1993).
- Magalhães, 2015 Magalhães, M. N.: Probabilidade e Variáveis Aleatórias, 3. ed. São Paulo: Editora da Universidade de São Paulo, (2015).
- Rao, 1988 Rao, K. S., *A bimodal distribution*. Bulletin of Calcutta Mathematical Society, **80**, pp. 238-240, 1988.
- Rényi, 1961 Rényi, A. (1961). *On measures of information and entropy*. Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960. pp. 547-561.

- Shankar, 2003 Shankar, P. M.; Forsberg, F.; Lown, L. (2003). *Statistical modeling of atherosclerotic plaque in carotid B mode images-a feasibility study*. *Ultrasound in Medicine and Biolog*, **29**, pp. 1305-1309.
- Shannon, 1948 Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, 27, 379–423, 623-656.
- Tsallis, 2009 Tsallis, C. (2009). *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*. Springer, New York.
- Vila et al., 2020a Vila, R.; Leao, J.; Saulo, H.; Shahzad, M. N.; Santos-Neto, M. (2020). *On a bimodal Birnbaum-Saunders distribution with applications to lifetime data*. *Brazilian Journal of Probability and Statistics*, **34**, pp. 495-518.
- Vila et al., 2020b Vila, R.; Ferreira, L.; Saulo, H.; Pratavia, F.; Ortega, E. M. M. (2020). *A bimodal gamma distribution: Properties, regression model and applications*. *Statistics*, **54**, pp. 469–493.
- Vila and Çankaya, 2021a Vila, R.; Çankaya, M. N. (2021). *A Bimodal Weibull Distribution: Properties and Inference*. *Journal of Applied Statistics*, pp. 1–19.
- Vila et al., 2021b Vila, R; Alfaia, L.; Menezes, A. F. B.; Çankaya, M. N.; Bourguignon, M. (2021). *A Model for Bimodal Rates and Proportions*. Preprint, available at [Arxiv.org/2108.07934](https://arxiv.org/2108.07934).
- Vinberg, 2003 Vinberg, E. (2003). *A Course in Algebra*. Graduate studies in mathematics. American Mathematical Society.
- Weisberg, 2005 Weisberg, S. N. (2005). *Applied Linear Regression*. John Wiley & Sons: New York, NY, USA.
- Xue, 2012 Xue, J. (2012). *Loop Tiling for Parallelism*. The Springer International Series in Engineering and Computer Science. Springer US.

Apêndices

.1 Misturas finitas de distribuições Beta é identificável

Em esta seção provamos que $\mathcal{H}_{\mathcal{B}}$, a classe de todas as misturas finitas de \mathcal{B} , onde \mathcal{B} é a família de distribuições Beta introduzido em (3.25), é uma classe identificável. Para este fim, primeiro enunciamos o seguinte corolário.

Corolário .1.1 ((Atienza et al., 2006)). *Seja \mathcal{F} uma família de distribuições. Seja M uma aplicação linear a qual transforma qualquer $F \in \mathcal{F}$ numa função real ϕ_F com domínio $S(F) \subseteq \mathbb{R}^d$. Seja $S_0(F) = \{t \in S(F) : \phi_F(t) \neq 0\}$ e suponha que exista um ponto t_0 satisfazendo*

$$t_0 \in \left[\bigcap_{i=1}^k S_0(F_i) \right]^c \quad (1)$$

para qualquer coleção finita de distribuições $F_1, \dots, F_k \in \mathcal{F}$. Se a relação

$$F_1 \prec F_2 \text{ if and only if } \lim_{t \rightarrow t_0} \frac{\phi_{F_2}(t)}{\phi_{F_1}(t)} = 0 \quad (2)$$

define uma ordem total sobre \mathcal{F} , então a classe \mathcal{H} de todas as misturas finitas de distribuições de \mathcal{F} é identificável.

Proposição .1.2. *A classe $\mathcal{H}_{\mathcal{B}}$ de todas as misturas finitas de \mathcal{B} é identificável.*

Demonstração. Seja $F \in \mathcal{B} \mapsto M(F) = \mathbb{E}_F[\exp(t \log X)]$ uma aplicação a qual transforma a função de distribuição $F \in \mathcal{B}$ na função geradora de momentos de $\log X$. Veja que M é linear e que

$$M[F(\cdot; \alpha, \beta)] = \phi_{F(\alpha, \beta)}(t) = \frac{B(\alpha + t, \beta)}{B(\alpha, \beta)}, \quad t \in (-\alpha, +\infty), \quad F \in \mathcal{B}.$$

Em este caso $S_0(F) = (-\alpha, +\infty)$. O ponto $t_0 = +\infty$ satisfaz a condição (1) do Corolário .1.1. A prova é completada se verificamos que a relação binária em (2) define uma ordem total sobre a família \mathcal{B} .

De fato, usando a aproximação de Stirling's temos a seguinte formula assintótica $B(x, y) \sim \Gamma(y)x^{-y}$ para $x \rightarrow +\infty$ (isto é, o limite do cociente é 1). Logo

$$\phi_{F(\alpha, \beta)}(t) \sim \frac{\Gamma(\beta)}{B(\alpha, \beta)} (\alpha + t)^{-\beta}.$$

Portanto, para $t \rightarrow +\infty$ é fácil verificar que

$$\frac{\phi_{F(\alpha', \beta')}(t)}{\phi_{F(\alpha, \beta)}(t)} \sim \frac{\Gamma(\beta')B(\alpha, \beta)}{\Gamma(\beta)B(\alpha', \beta')} \frac{(\alpha' + t)^{-\beta'}}{(\alpha + t)^{-\beta}}.$$

Daí, $F(\cdot; \alpha, \beta) \prec F(\cdot; \alpha', \beta')$ se e somente se $[\alpha = \alpha'$ and $\beta < \beta']$, a qual define uma ordem total em \mathcal{B} . Logo, por aplicar o Corolário .1.1, a prova segue. \square

.2 Estimativa de máxima verossimilhança

Seja (X_1, \dots, X_n) uma amostra aleatória da v.a $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$ que depende de um vetor de parâmetros $\boldsymbol{\theta}_\delta = (\alpha, \beta, \rho, \delta)$, onde $\boldsymbol{\theta}_\delta$ pertence a um subconjunto aberto de \mathbb{R}^4 , de forma que valores distintos de $\boldsymbol{\theta}_\delta$ geram distribuições distintas para X_1 ; veja a Seção 3.7. Se $\mathbf{x} = (x_1, \dots, x_n)$ denota os valores observados correspondentes de (X_1, \dots, X_n) , a função log-verossimilhança para $\boldsymbol{\theta}_\delta$ é dada por

$$l(\boldsymbol{\theta}_\delta; \mathbf{x}) = -n \log Z(\boldsymbol{\theta}_\delta) - n \log B(\alpha, \beta) + \sum_{i=1}^n \left\{ \log [\rho + (1 - \delta x_i)^2] + (\alpha - 1) \log x_i + (\beta - 1) \log(1 - x_i) \right\},$$

onde $Z(\boldsymbol{\theta}_\delta)$ é constante de normalização na f.d.p. da Beta bimodal (3.1).

As equações de log-verossimilhança máxima para os estimadores $\hat{\alpha}$, $\hat{\beta}$, $\hat{\rho}$ e $\hat{\delta}$ são as seguintes:

$$\begin{aligned} n[\psi(\hat{\alpha}) - \psi(\hat{\alpha} + \hat{\beta})] - \frac{1}{(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1)} \frac{n\hat{\beta}\hat{\delta}[2(\hat{\alpha} + \hat{\beta} + 1)^2 - \hat{\delta}(2\hat{\alpha}^2 + 2\hat{\alpha}\hat{\beta} + 2\hat{\alpha} + \hat{\beta} + 1)]}{(1 + \hat{\rho})(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1) - 2\hat{\alpha}(\hat{\alpha} + \hat{\beta} + 1)\hat{\delta} + \hat{\alpha}(\hat{\alpha} + 1)\hat{\delta}^2} &= \sum_{i=1}^n \log x_i; \\ n[\psi(\hat{\beta}) - \psi(\hat{\alpha} + \hat{\beta})] + \frac{1}{(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1)} \frac{n\hat{\alpha}\hat{\delta}[2(\hat{\alpha} + \hat{\beta} + 1)^2 - \hat{\delta}\hat{\alpha}(\hat{\alpha} + 1)(2\hat{\alpha} + 2\hat{\beta} + 1)]}{(1 + \hat{\rho})(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1) - 2\hat{\alpha}(\hat{\alpha} + \hat{\beta} + 1)\hat{\delta} + \hat{\alpha}(\hat{\alpha} + 1)\hat{\delta}^2} &= \sum_{i=1}^n \log(1 - x_i); \\ \frac{n(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1)}{(1 + \hat{\rho})(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1) - 2\hat{\alpha}(\hat{\alpha} + \hat{\beta} + 1)\hat{\delta} + \hat{\alpha}(\hat{\alpha} + 1)\hat{\delta}^2} &= \sum_{i=1}^n \frac{1}{\hat{\rho} + (1 - \hat{\delta}x_i)^2}; \\ \frac{n\hat{\alpha}[(\hat{\alpha} + \hat{\beta} + 1) - \hat{\delta}(\hat{\alpha} + 1)]}{(1 + \hat{\rho})(\hat{\alpha} + \hat{\beta})(\hat{\alpha} + \hat{\beta} + 1) - 2\hat{\alpha}(\hat{\alpha} + \hat{\beta} + 1)\hat{\delta} + \hat{\alpha}(\hat{\alpha} + 1)\hat{\delta}^2} &= \sum_{i=1}^n \frac{(1 - \hat{\delta}x_i)x_i}{\hat{\rho} + (1 - \hat{\delta}x_i)^2}; \end{aligned}$$

onde $\psi(x) = \Gamma'(x)/\Gamma(x)$ é a função Digamma.

.3 Matriz de informação

Se x denota a uma única observação da distribuição BBeta, então a função de log-verossimilhança pode ser escrita como segue

$$l(\boldsymbol{\theta}_\delta; x) = -\log Z(\boldsymbol{\theta}_\delta) + \log [\rho + (1 - \delta x)^2] + l_Y(\boldsymbol{\theta}_0; x),$$

onde $Z(\boldsymbol{\theta}_\delta)$ é a constante de normalização dada em (3.2), e

$$l_Y(\boldsymbol{\theta}_0; x) = -\log B(\alpha, \beta) + \{(\alpha - 1) \log x + (\beta - 1) \log(1 - x)\}$$

é a função log-verossimilhança para $\boldsymbol{\theta}_0 = (\alpha, \beta)$, dado o resultado x de $Y \sim \text{BBeta}(\boldsymbol{\theta}_0)$.

As derivadas parciais de segunda ordem são:

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \alpha^2} &= -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \alpha^2} + \frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \alpha^2}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \beta^2} &= -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \beta^2} + \frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \beta^2}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \rho^2} &= -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \rho^2} - \frac{1}{[\rho + (1-\delta x)^2]^2}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \delta^2} &= -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \delta^2} + \frac{2x^2[\rho - (1-\delta x)^2]}{[\rho + (1-\delta x)^2]^2}.
\end{aligned} \tag{3}$$

Pelo Teorema de Schwarz, as derivadas parciais mistas são simétricas. Então, as derivadas parciais mistas de segunda ordem são dadas por:

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \alpha \partial \beta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \beta \partial \alpha} = -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \alpha \partial \beta} + \frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \alpha \partial \beta}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \alpha \partial \rho} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \rho \partial \alpha} = -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \alpha \partial \rho}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \alpha \partial \delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \delta \partial \alpha} = -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \alpha \partial \delta}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \beta \partial \rho} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \rho \partial \beta} = -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \beta \partial \rho}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \beta \partial \delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \delta \partial \beta} = -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \beta \partial \delta}, \\
\frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \rho \partial \delta} &= \frac{\partial^2 l(\boldsymbol{\theta}_\delta; x)}{\partial \delta \partial \rho} = -\frac{\partial^2 \log Z(\boldsymbol{\theta}_\delta)}{\partial \rho \partial \delta} + \frac{2x(1-\delta x)}{[\rho + (1-\delta x)^2]^2};
\end{aligned} \tag{4}$$

onde as derivadas parciais $\partial^2 \log Z(\boldsymbol{\theta}_\delta) / \partial x \partial y$, $x, y \in \{\alpha, \beta, \rho, \delta\}$, são facilmente obtidas a partir de (3.2), e

$$\begin{aligned}
\frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \alpha^2} &= \frac{\Gamma(\alpha+\beta)\Gamma''(\alpha+\beta) - \{\Gamma(\alpha+\beta)\}^2}{\Gamma^2(\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma''(\alpha) - \{\Gamma'(\alpha)\}^2}{\Gamma^2(\alpha)}, \\
\frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \alpha \partial \beta} &= \frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \beta \partial \alpha} = \frac{\Gamma(\alpha+\beta)\Gamma''(\alpha+\beta) - \{\Gamma(\alpha+\beta)\}^2}{\Gamma^2(\alpha+\beta)}, \\
\frac{\partial^2 l_Y(\boldsymbol{\theta}_0; x)}{\partial \beta^2} &= \frac{\Gamma(\alpha+\beta)\Gamma''(\alpha+\beta) - \{\Gamma(\alpha+\beta)\}^2}{\Gamma^2(\alpha+\beta)} - \frac{\Gamma(\beta)\Gamma''(\beta) - \{\Gamma'(\beta)\}^2}{\Gamma^2(\beta)}.
\end{aligned}$$

Se $X \sim \text{BBeta}(\boldsymbol{\theta}_\delta)$, os elementos da matriz de informação de Fisher associada a X são definidos por

$$[I_F(\boldsymbol{\theta}_\delta)]_{jk} = \mathbb{E} \left[\frac{\partial l(\boldsymbol{\theta}_\delta; X)}{\partial \theta_j} \frac{\partial l(\boldsymbol{\theta}_\delta; X)}{\partial \theta_k} \right], \quad j, k = 1, \dots, n,$$

para $\theta_j, \theta_k \in \{\alpha, \beta, \rho, \delta\}$. Sob condições de regularidade clássicas, os elementos da matriz de informação de Fisher são dados por $\mathbb{E}[-\partial^2 l(\boldsymbol{\theta}_\delta; X) / \partial \theta_j \partial \theta_k]$, onde as derivadas parciais $\partial^2 l(\boldsymbol{\theta}_\delta; x) / \partial \theta_j \partial \theta_k$ são fornecidas nos Itens (3) e (4).