

Universidade de Brasília
Instituto de Exatas
Dissertação de Mestrado em Estatística

**Discriminante não-linear para
mistura de distribuições Beta**

Por Euler Rodrigues de Alencar

Orientadora: Prof^ª. Dr^ª. Cira Etheowalda Guevara Otiniano

Brasília, Maio de 2018

Dissertação

Discriminante não-linear para mistura de distribuições Beta

Dissertação submetida à banca examinadora designada pelo Colegiado do Curso de Mestrado em Estatística da Universidade de Brasília (UnB).

Brasília, Maio de 2018

Resumo

Em geral, nos modelos de mistura finita, há dois tipos de problemas que devem ser resolvidos. O primeiro consiste em estimar o número de componentes da mistura ou estimar os parâmetros da mistura. O segundo problema consiste em estimar a função discriminante da mistura para dados não classificados. Neste trabalho, estuda-se a estimação dos parâmetros da mistura de duas distribuições Beta por máxima verossimilhança. Para o estudo da função discriminante considera-se dados não classificados (mistura) e dados classificados (classificados). Quando trata-se de dados de mistura a estimação dos parâmetros utiliza o algoritmo EM. Para avaliação do desempenho da função discriminante utiliza-se dados simulados e calcula-se os erros de discriminação. Por último, aplica-se a metodologia para um conjunto de dados reais.

Abstract

Usually, in finite mixture models, there are two types of problems that must be solved. The first one is to estimate the number of components or to estimate the parameters of the mixture. The second one is to estimate a discriminant function for the unclassified data. This essay study parameter estimation of mixture of two Beta distributions by maximum likelihood. For the search of the discriminative function, consider the unclassified and classified data. When the data set is unclassified EM algorithm is used. To evaluate discrimination performance, simulated data is used and calculate the discrimination errors. Finally, the methodology is applied to real data.

Agradecimentos

Aos meus pais, pelo amor, incentivo e apoio incondicional. Agradeço à minha mãe, Laélia Alencar, heroína que me deu todo apoio, incentivo nas horas difíceis, de desânimo e cansaço. Ao meu pai, Milton Silva, pela inspiração neste caminho.

Obrigado aos meus irmãos Yuri Alencar e Natália Alencar, pelo apoio e pelos meus momentos de maiores ansiedades, pois sempre estiverem ali para auxiliar-me. À minha sobrinha Layla pela alegria do dia a dia.

Meus agradecimentos aos amigos da vida, companheiros de trabalho e irmãos que fiz durante minha formação. Na certeza que irão continuar em minha vida como inspiração. Agradeço ao amigo Carlos Levicoy que me auxiliou nos momentos de dúvida.

Por fim, agradeço a Universidade de Brasília pelo suporte através do Departamento de Estatística, em especial à Professora Cira Guevara, minha orientadora, por toda paciência e sabedoria na condução deste trabalho. Um agradecimento especial à Professora Cibele, que sempre serviu de inspiração por seu grande profissionalismo. Além da contribuição valiosa dos professores do departamento que fizeram parte dessa jornada.

Sumário

Resumo	i
Abstract	iii
Agradecimentos	iv
Lista de Figuras	vii
Lista de Tabelas	ix
1 Introdução	1
1.1 Motivação, Objetivo e Metodologia	1
2 Revisão de literatura	5
2.1 Modelo Beta	5
2.2 Mistura de densidades	7
2.3 Estimador de Máxima Verossimilhança	7
2.4 Algoritmo EM - caso geral	9
2.5 Algoritmo EM para amostras misturadas	10
2.6 Identificabilidade na forma geral	11
2.7 Discriminante não-Linear	11
2.7.1 Função discriminante para mistura de distribuições	12
2.7.2 Erros de classificação	14
3 Discriminante não-linear para mistura de Beta	17
3.1 Mistura de Betas	17
3.2 Identificabilidade	18
3.3 Máxima verossimilhança da mistura de Betas não classificadas	19
3.4 EMV para mistura de Betas em amostras classificadas	22
3.5 Função discriminante ótima para mistura de Betas	24
3.6 Função discriminante - caso particular	25
3.7 Erro de classificação	26
4 Simulação e Aplicação	29
4.1 Geração dos dados e estimação	29

5	Aplicações	44
5.1	Tempo entre erupções de geysers	44
5.2	Vazão máxima anual das usinas do rio Paranapanema	46
5.3	Proeficiência no ENEM 2017	48
5.4	Conclusão	50
A	Resultados da Estimação	52
B	Código R para mistura de normais	55
C	Código R para geração dos dados	57

Lista de Figuras

2.1	Gráficos da densidade de $X \sim \text{Beta}(\alpha, \beta)$	6
3.1	Mistura de distribuições	17
4.1	Densidades $h(x, \theta)$, conforme (3.1), com Θ_1 a Θ_9 definidos na Tabela 4.1	30
4.2	Comparação da densidade teórica, classificada e misturada para Θ_1	33
4.3	Comparação da densidade teórica, classificada e misturada para Θ_3	34
4.4	Comparação da densidade teórica, classificada e misturada para Θ_4	34
4.5	Comparação da densidade teórica, classificada e misturada para Θ_5	35
4.6	Comparação da densidade teórica, classificada e misturada para Θ_2	37
4.7	Comparação da densidade teórica, classificada e misturada para Θ_6	40
4.8	Comparação da densidade teórica, classificada e misturada para Θ_7	40
4.9	Comparação da densidade teórica, classificada e misturada para Θ_8	41
4.10	Comparação da densidade teórica, classificada e misturada para Θ_9	41
5.1	Histograma dos dados de erupção de geysers (min/100)	44
5.2	Histograma dos dados de erupção de geysers (min/100) e a distribuição Beta estimada.	45
5.3	Histograma dos dados de vazão máxima anual das usinas da bacia do rio Paranapanema.	46
5.4	Histograma dos dados de vazão máxima anual das usinas da bacia do rio Paranapanema e a estimativa da mistura de duas Betas	47
5.5	Histograma da proporção de acertos no ENEM 2017, a densidade empírica e densidade estimada.	49
5.6	QQ-plot da distribuição estimada.	49
A.1	Comparação dos erros quadráticos do parâmetro p	52
A.2	Comparação dos erros quadráticos do parâmetro α	53
A.3	Comparação dos erros quadráticos do parâmetro β_1	53
A.4	Comparação dos erros quadráticos do parâmetro β_2	54

Lista de Tabelas

2.1	Tabela das frequências dos erros de classificação	15
2.2	Processo do método Holdout	15
2.3	Processo do método Lachenbruch	16
4.1	Parâmetros da simulação.	29
4.2	Estimativa dos parâmetros de $h(x, \Theta_1)$	32
4.3	Estimativa dos parâmetros de $h(x, \Theta_3)$	32
4.4	Estimativa dos parâmetros de $h(x, \Theta_4)$	33
4.5	Estimativa dos parâmetros de $h(x, \Theta_5)$	33
4.6	Erros de classificação com Θ_1	36
4.7	Erros de classificação com Θ_3	36
4.8	Erros de classificação com Θ_4	36
4.9	Erros de classificação com Θ_5	36
4.10	Estimativa dos parâmetros de $h(x, \Theta_2)$	37
4.11	Erros de classificação com Θ_2	38
4.12	Estimativa dos parâmetros de $h(x, \Theta_6)$	38
4.13	Estimativa dos parâmetros de $h(x, \Theta_7)$	39
4.14	Estimativa dos parâmetros de $h(x, \Theta_8)$	39
4.15	Estimativa dos parâmetros de $h(x, \Theta_9)$	39
4.16	Erros de classificação com Θ_6	42
4.17	Erros de classificação com Θ_7	42
4.18	Erros de classificação com Θ_8	42
4.19	Erros de classificação com Θ_9	43
5.1	Estimação Geysler.	45

Capítulo 1

Introdução

1.1 Motivação, Objetivo e Metodologia

Uma mistura finita de densidades é uma abordagem matemática de modelos estatísticos com ampla aplicabilidade por se tratar de uma ferramenta muito flexível. Ela tem sido utilizada onde há heterogeneidade populacional ou para aproximação de densidades complexas com aspectos multimodais, veja, por exemplo, McLachlan e Peel (2000).

Define-se a mistura de k densidades como a combinação convexa de k densidades de uma mesma família ou até de família distintas, sendo cada densidade denominada de componente da mistura e os coeficientes chamados de pesos da mistura. O vetor dos parâmetros de um modelo de mistura tem no mínimo $2k - 1$ parâmetros a serem estimados, sendo $k - 1$ referentes aos pesos e k às componentes. Esses parâmetros podem ser estimados por máxima verossimilhança, via algoritmo EM ou por métodos bayesianos.

Em geral, esses modelos são aplicados em diversas áreas como análise de agrupamento, análise discriminante, análise de sobrevivência, métodos não paramétricos ou semiparamétricos e processamento de imagens. Há também aplicações em outras áreas tais como astronomia, biologia, genética, medicina, psiquiatria, economia e engenharia, conforme McLachlan e Peel (2000), McLachlan e Basford (1988) e Titterton *et.al.* (1985).

Na década de 60, a estimação dos parâmetros dos modelos de mistura finita utilizando estimação por máxima verossimilhança (EMV) foi estudada por diversos autores, entre eles Day (1969) e Wolfe (1965, 1967, 1970). Entretanto, após a publicação do trabalho de Dempster *et. al.* (1977), a estimação dos parâmetros desses modelos por máxima verossimilhança via algoritmo EM tem tornado a tarefa de estimação mais fácil.

Apesar dos avanços em relação a estimação dos parâmetros, o estudo de modelos de mistura finita apresenta ainda dois problemas. O primeiro é a estimação dos parâmetros do modelo e o outro é a análise de discriminantes para dados não classificados. Categoriza-se os discriminantes entre lineares e não lineares, tal que este trata de discriminantes com função não linear, como uma função logarítmica e aquele utiliza função linear.

Este trabalho foca no uso de uma função não linear para discriminação de elementos. Para pequenas amostras esta metodologia tem sido analisada por vários autores, entre eles Amoh (1984), que estudou a mistura de duas distribuições gaussianas inversas, Ahmad e Adb-Elrahman (1994) que estudaram discriminantes não lineares para mistura de duas Weibulls, e recentemente Ahmad *et.al.* (2010), que estudaram duas distribuições Gumbel.

Neste contexto, é importante definir dois conceitos que serão amplamente usados para

testar a performance da função discriminante. Esses conceitos referem-se à amostra utilizada no processo de estimação. Eles são:

- **Amostra não classificada (m):** Quando os dados são obtidos de uma amostragem de uma população de uma mistura, mas a observação de origem é desconhecida;
- **Amostra classificada (c):** Os dados também são amostrados de uma mistura, mas a origem de cada observação é conhecida.

Para cada uma dessas amostras haverá um método de estimação diferente para os parâmetros, por isso a importância de comparar os valores da função discriminante considerando os estimadores dos dois tipos.

Outra propriedade importante para tratar ambos os problemas é a unicidade da classe de misturas finitas. A propriedade da identificabilidade de uma classe de distribuições garante a unicidade dos parâmetros das distribuições. Assim, o conhecimento dessa propriedade para uma classe de modelos de mistura é importante tanto na estimação dos parâmetros quanto na estimação da função discriminante associada à mistura. Com o intuito de facilitar a prova de identificabilidade de classes de modelos de mistura, várias caracterizações têm sido propostas. Entre elas temos a de Teicher (1963), Chandra (1977), Al-Hussaini e Ahmad (1981), Ahmad (1988) e Atienza *et.al.* (2006).

Outro ponto do trabalho é a escolha da distribuição Beta para modelagem do problema. Quando uma população de interesse apresenta suporte no intervalo $(0, 1)$, a densidade Beta surge naturalmente como uma possibilidade. Esse modelo é conhecido por sua ampla aplicabilidade em diversas áreas. Na estatística, por exemplo, é utilizado na inferência bayesiana como distribuição a priori para parâmetros que variem entre 0 e 1. Em outras áreas, é utilizado para modelar porcentagens, dadas suas características de variação no intervalo 0 e 1, e também é utilizado para estudos sobre correlações.

Ji *et al.* (2005) utilizaram, na área de bioinformática, um modelo de mistura de Betas para resolver problemas relacionados à correlação de expressões de níveis de certos genes. Já Heisterkamp e Pennings (2004) utilizam a mistura finita de distribuições Beta para análise de MicroArrays relacionado à genes. O estudo utilizou a distribuição Beta para separar grupos ao invés de usar ANOVA ou teste-t, por exemplo. Schröder e Rahmann (2017) utilizaram a mistura de Betas para estudar os níveis de metilação do DNA.¹ O trabalho é interessante por propor algoritmo que combina variáveis latentes com métodos dos momentos, em vez de utilizar algoritmo EM, especialmente porque trata do problema da singularidade da função de log-verossimilhança.

Este trabalho está organizado em 5 (cinco) capítulos. No Capítulo 2, é feita uma revisão de literatura, apresentando a metodologia, o modelo Beta e sua função geratriz de momentos. Há uma introdução sobre mistura de densidades, o estimador de máxima verossimilhança para o caso geral e o algoritmo EM na sua forma geral. No Capítulo 3, é apresentado o modelo de mistura de duas distribuições Beta, calculam-se seus momentos e prova-se a identificabilidade da classe de misturas finitas. Em seguida obtêm-se as expressões a serem resolvidas para estimar, por máxima verossimilhança, os parâmetros da mistura, no caso de amostra não classificada usa-se o algoritmo EM. Quanto à função discriminante, calcula-se

¹A metilação do DNA é um tipo de modificação química do DNA que pode ser herdada e subsequentemente removida sem mudar a sequência original do DNA

a expressão e seus respectivos erros de classificação. O Capítulo 4 trata basicamente da avaliação do modelo utilizando-se de uma simulação e aplicação em dados reais. Em primeiro lugar foi avaliada a performance do discriminante não linear, foi testada sua eficácia através do cálculo dos erros de classificação com dados simulados. O Capítulo 5 apresenta 3 aplicações da metodologia para estimação e discriminação de dados de tempo de erupção de geysers, para dados de vazão máxima anual das usinas da bacia do rio Paranapanema e para dados das notas do ENEM 2017. Por fim, apresenta-se a conclusão do trabalho, apontando limitações da metodologia e recomendações para trabalhos futuros.

Capítulo 2

Revisão de literatura

Neste capítulo será apresentada a distribuição Beta, sua função geratriz e seus momentos. É definido o conceito geral de mistura de densidades e o estimador de máxima verossimilhança para o caso de misturas. Dada a necessidade de simplificação do processo de estimação, apresenta-se o Algoritmo-EM geral para mistura de densidades. Outro conceito relevante é a identificabilidade, que procura garantir a unicidade da estimação. Por fim, faz-se uma apresentação sobre a função discriminante não-linear utilizada para classificação de elementos da amostra.

2.1 Modelo Beta

Os resultados desta seção foram baseados nos trabalhos de Johnson, Kotz e Balakrishnan (1995) e Arjun (2004). Eles trazem resultados sobre a distribuição Beta e suas aplicações.

Uma variável aleatória (v.a) X com distribuição Beta é frequentemente usada para modelar proporções ou variáveis que pertençam ao intervalo $(0,1)$. Por ser uma distribuição flexível, é possível modelar funções de densidade de probabilidade restrita a um intervalo finito (m,n) , tal que $Y = (n - m)X + m$. Assim, o modelo Beta tem inúmeras aplicações para representar quantidades físicas cujos valores estejam restritos a determinado intervalo.

Uma v.a X tem distribuição Beta se sua função de densidade de probabilidade (f.d.p) para valores $0 < x < 1$ é dada por

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad x \in (0, 1) \quad \text{e} \quad \alpha, \beta > 0, \quad (2.1)$$

sendo α e β parâmetros de forma e $\Gamma(\cdot)$ é a função gama definida por

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx, \quad t > 0.$$

A partir de $\Gamma(\cdot)$, define-se a função Beta, $B(\cdot, \cdot)$, como sendo

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dx = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

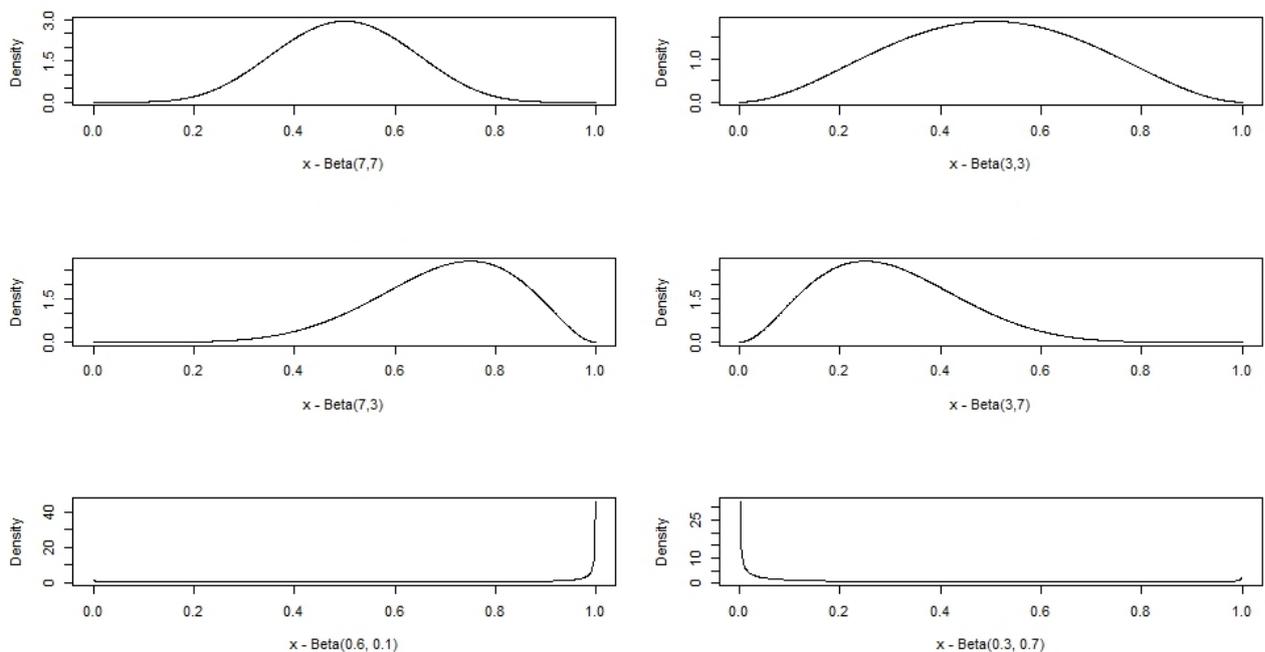
Assim, (2.1) pode ser reescrita como

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad x \in (0, 1) \quad \text{e} \quad \alpha, \beta > 0.$$

logo, neste trabalho, será utilizado a notação $X \sim \text{Beta}(\alpha, \beta)$ para indicar que uma v.a X tem densidade definida em (2.1), tal que os parâmetros α e β definem a forma da distribuição. Se $\alpha = \beta$, a distribuição é simétrica, se $\alpha > \beta$, a assimetria é negativa e, no caso de $\alpha < \beta$, sua assimetria é positiva.

A figura abaixo mostra o comportamento da densidade de uma v.a Beta para diferentes tipos de parâmetros.

Figura 2.1: Gráficos da densidade de $X \sim \text{Beta}(\alpha, \beta)$



Função Geradora de Momentos (f.g.m) e Momentos da Beta

Se X é uma variável aleatória (v.a) com distribuição Beta, então sua função geradora de momentos é dada por

$$M_X(t) := \mathbb{E}[e^{tX}] = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{t^k}{k!}, \quad t \in \mathbb{R},$$

Assim, é possível obter o k -ésimo momento da densidade Beta

$$\mathbb{E}(X^k) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+k-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + k)}{\Gamma(\alpha + \beta + k)\Gamma(\alpha)}. \quad (2.2)$$

De (2.2), com $k = 1$, obtém-se

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

e para $k = 2$, o segundo momento é dado por

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

Portanto, a variância é obtida por:

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

Importante frisar que o k -ésimo momento é especialmente importante para obter a condição de identificabilidade da classe de mistura finita de densidades Beta, a ser tratada no capítulo 3.

2.2 Mistura de densidades

Um modelo de mistura finita de k densidades $h(\cdot; \cdot)$ de uma família \mathcal{F} é uma combinação linear convexa dessas densidades de probabilidade. Formalmente, isso quer dizer que $h(\cdot; \cdot)$ é uma mistura de duas densidades f_1 e $f_2 \in \mathcal{F}$, se

$$h(x; \Theta) = p_1 f_1(x, \theta_1) + p_2 f_2(x, \theta_2), \quad \text{em que } p_1 + p_2 = 1, \quad (2.3)$$

sendo $\Theta = (\theta_1, \theta_2)$. Dependendo da escolha dos parâmetros, a mistura pode modelar desde simples gráficos unimodais até gráficos que possuem várias modas. Por possuir esta flexibilidade, estes modelos são bastante estudados no ramo da estatística.

Uma condição importante para tratar destes tipos de modelos é que a mistura de densidade deve ser uma combinação convexa. Em geral, combinações lineares de densidades de probabilidade não são necessariamente uma densidade de probabilidade, isto é, pode acontecer de ser um valor negativo ou a integração pode não ser igual a 1. Contudo, uma combinação linear convexa de funções de densidade de probabilidade preserva as duas propriedades.

2.3 Estimador de Máxima Verossimilhança

O método de estimação de máxima verossimilhança (EMV) é um dos procedimentos mais utilizados para se obter estimadores. Considere uma população ou uma variável aleatória X com função de probabilidade $P(X = x)$ (se X é uma variável aleatória discreta) ou função densidade de probabilidade $f(x, \theta)$ (se X é uma variável aleatória contínua), sendo

θ o parâmetro desconhecido. Retira-se uma amostra aleatória simples de X , de tamanho n , X_1, \dots, X_n , e sejam x_1, \dots, x_n os valores observados.

A função de verossimilhança L é definida por

$$L(\theta; x_1, \dots, x_n) = f(x_1; \theta) \times \dots \times f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta). \quad (2.4)$$

O estimador de máxima verossimilhança de θ é o valor que maximiza $L(\theta; x_1, \dots, x_n)$. Ao substituir (2.3) na equação (2.4), obtém-se

$$L(\Theta; X) = \prod_{j=1}^n [h(x, \Theta)] = \prod_{i=1}^n \left[p_1 f_1(x_i; \theta_1) + p_2 f_2(x_i; \theta_2) \right]. \quad (2.5)$$

Já a função log-verossimilhança é dada por

$$\log [L(\Theta; X)] = \log \left(\prod_{i=1}^n [h(x, \Theta)] \right) = \sum_{j=1}^n \log \left(p_1 f_1(x_i; \theta_1) + p_2 f_2(x_i; \theta_2) \right). \quad (2.6)$$

A depender da densidade $f(\cdot)$ é possível agrupar algumas quantidades e obter formas da log-verossimilhança mais simples.

A resolução deste tipo de equação se dá a partir da derivação da função $\log [L(\Theta; X)]$ em relação a cada um dos parâmetros que compõe o vetor Θ . De forma geral, diz-se que a solução é a resolução do sistema de equações gerada por

$$\frac{\partial L(\Theta; X)}{\partial \Theta} = 0. \quad (2.7)$$

A aplicação da derivada em relação a cada um dos parâmetros leva a um sistema de equações, conforme abaixo:

$$\begin{cases} \frac{\partial L(\Theta; X)}{\partial \theta_1} = 0; \\ \frac{\partial L(\Theta; X)}{\partial \theta_2} = 0; \\ \dots \\ \frac{\partial L(\Theta; X)}{\partial \theta_n} = 0. \end{cases}$$

A resolução do sistema de equações acima leva às estimativas $\hat{\Theta}$. Como a expressão (2.6) envolve o logaritmo de somas, as expressões correspondente a (2.7) em geral não são simples, o que dificulta o processo de estimação. Por esse motivo, o algoritmo EM surge como uma ferramenta que viabiliza a estimação de Θ de forma iterativa.

2.4 Algoritmo EM - caso geral

O algoritmo EM (Dempster et al. 1977) é uma ferramenta computacional utilizada para o cálculo do estimador de máxima verossimilhança de forma iterativa. Este método é a principal ferramenta utilizada para resolver problemas envolvendo dados incompletos.

Há duas principais aplicações envolvendo o algoritmo EM:

- i. Quando os dados são incompletos;
- ii. Quando a maximização da verossimilhança é complexa demais.

O primeiro caso diz respeito a problemas ou limitações no processo de observação dos dados (dados faltantes). O segundo é mais comum e a utilização do método depende da simplificação da função de probabilidade através de algum método computacional. Segundo Casella e Berger (2010), o algoritmo EM tem como princípio substituir uma difícil maximização da verossimilhança por uma sequência de maximizações mais fáceis.

O processo do algoritmo EM é o seguinte: Considere y um conjunto de dados observados e z um conjunto de dados faltantes. O dado completo é representado por $x = (y, z)$ tal que a densidade de x é $\mathbb{P}(X|\theta)$, com $\theta \in \Theta \subset \mathbb{R}^n$. Denota-se por $\mathcal{L}_c(\theta|x)$ a função log-verossimilhança dos dados completos, e por $\mathcal{L}(\theta|y)$ a função log-verossimilhança dos dados observados. Segundo Zhu e Lee (2001), a função de log-verossimilhança dos dados completos geralmente tem forma mais simples que a log-verossimilhança dos dados observados.

Assim, cada iteração do algoritmo EM envolve dois passos:

1. Passo E (Esperança):

Calcule $Q(\theta|\theta^{(k)}) = \mathbb{E} [\mathcal{L}_c(\theta|x)|y, \theta^{(k)}]$, em que a esperança é tomada com respeito aos dados observados y .

2. Passo M (Maximização):

Encontre $\theta^{(k+1)}$ que maximiza $Q(\theta|\theta^{(k)})$.

Simplificando, o passo **E** calcula o valor esperado do logaritmo da verossimilhança e o passo **M** encontra o valor de θ que maximiza a função Q . Espera-se com esse procedimento obter o valor ótimo para θ . Nota-se, no entanto, que os máximos encontrados no algoritmo EM garantem apenas máximos locais, não há garantia sobre máximos globais.

Os passos desse algoritmo devem ser repetidos até atingir a convergência, adotando-se algum critério de parada como, por exemplo, $|\theta^{(k+1)} - \theta^{(k)}| \leq \epsilon$, no qual ϵ é um valor suficientemente pequeno maior que zero.

Assim, seja uma variável aleatória X com sua f.d.p $f(x|\theta)$ e seja $X = (X_1, \dots, X_n)$ uma amostra aleatória de X , tal que a variável aleatória X é chamada de dados incompletos. A função de verossimilhança e sua log-verossimilhança são dadas por

- a. $L(\Theta|x) = \prod_{i=1}^n f(x_i|\theta)$;
- b. $\mathcal{L}(\theta|x) = \log [L(\Theta|x)] = \sum_{i=1}^n \log [f(x_i|\theta)]$.

Se $Y = (Y_1, \dots, Y_m)$ são dados faltantes, então $Z = (X, Y)$ será o vetor de dados completos. Logo

$$f(z|\theta) = f(x, y|\theta) = f(y|x, \theta)f(x|\theta).$$

A função de verossimilhança é:

$$\mathcal{L}(\theta|z) = \log [L(\Theta|x, y)] = \log \left[\prod_{n+m=1}^{n+m} f(x, y|\theta) \right]. \quad (2.8)$$

O algoritmo EM consiste no cálculo em dois passos:

PASSO E: Aqui X e $\theta^{(k)}$ são conhecidos, então a função a ser obtida Q depende de Y .

$$Q(\theta, \theta^{(k)}) = E [\log(f(X, Y|\theta)|X, \theta^{(k)})] = \sum_{y_i \in \{1,2\}} \log [f(X, y_i|\theta)] P(y_i|X, \theta^{(k)}) \quad (2.9)$$

PASSO M: Consiste em maximizar Q com respeito ao parâmetro θ , isto é, resolver a equação 2.9 usando a derivada $\frac{\partial Q^{(k)}}{\partial \theta} = 0$.

2.5 Algoritmo EM para amostras misturadas

Seja X uma variável aleatória com densidade $h(x|\Theta)$ e (X_1, \dots, X_n) uma amostra aleatória. Quando $h(x|\Theta) = p_1 f_1(x|\theta_1) + p_2 f_2(x|\theta_2)$ e $\sum_{l=1}^2 p_l = 1$, com $\Theta = (p_1, p_2, \theta_1, \theta_2)$, então a função de verossimilhança dos dados completos, aplicados à equação (2.8), fica

$$\mathcal{L}(\theta|z) = \log \left[\prod_{i=1}^n h(x_i, y_i|\theta) \right] = \sum_{i=1}^n \log [h(y_i|x_i, \theta)h(x_i|\theta)],$$

De (2.9) obtém-se

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= \sum_{y_i} \log [h(x_i, y_i|\theta)] P(y_i|x_i, \theta^{(k)}) \\ &= \sum_{l=1}^2 \sum_{i=1}^n \log [p_l f_l(x_i|\theta_l)] P(Y = l|x_i, \theta^{(k)}), \end{aligned}$$

sendo

$$P(Y = l|x_i, \theta^{(k)}) = \frac{p_l f_l(x_i|\theta_l)}{h(x_i|\theta)}, \quad l = 1, 2.$$

Logo, a função Q é dada em forma explícita por:

$$\begin{aligned}
Q(\theta, \theta^{(k)}) &= \sum_{l=1}^2 \sum_{i=1}^n \ln(p_l) P(Y = l | x_i, \theta^{(k)}) + \sum_{l=1}^2 \sum_{i=1}^n \ln(f_l(x_i | \theta)) P(Y = l | x_i, \theta^{(k)}) \\
&= \sum_{i=1}^n \ln(p_1) P(Y = 1 | x_i, \theta_1^{(k)}) + \sum_{i=1}^n \ln(p_2) P(Y = 2 | x_i, \theta_2^{(k)}) \\
&\quad + \sum_{i=1}^n \ln(f_1(x_i | \theta_1)) P(Y = 1 | x_i, \theta_1^{(k)}) + \sum_{i=1}^n \ln(f_2(x_i | \theta_2)) P(Y = 2 | x_i, \theta_2^{(k)}).
\end{aligned} \tag{2.10}$$

2.6 Identificabilidade na forma geral

Identificabilidade é uma propriedade importante de se obter quando se fala de mistura de distribuições. As condições para obtenção de tal propriedade garantem a unicidade dos parâmetros. Dessa maneira, é preciso investigar tal propriedade.

Matematicamente, a classe de funções $\mathcal{F} = \{F(\cdot; \Theta) : F \text{ é distribuição de probabilidade}\}$ é identificável se para $F \in \mathcal{F}$, $\Theta \neq \Theta^*$, implica que $F(x; \Theta) \neq F(x; \Theta^*)$ para qualquer x .

No caso de mistura, ao considerar

$$\mathcal{H} = \left\{ H : H(x, \theta) = \sum_{i=1}^k F_i(x; \theta_i), \quad \theta \in \Theta, \quad F_i \in \mathcal{F}, \quad i = 1 \dots k \right\},$$

a classe de todas as misturas de k componentes $F_i \in \mathcal{F}$, diz-se que \mathcal{H} é identificável se, e somente se, para $H, H' \in \mathcal{H}$, tal que

$$H = \sum_{i=1}^k \pi_i F_i, \quad H' = \sum_{i=1}^{k'} \pi_i F'_i,$$

a igualdade $H = H'$ implica $k = k'$ e $(\pi_1, F_1), (\pi_2, F_2), \dots, (\pi_k, F_k)$ são permutações de $(\pi'_1, F'_1), (\pi'_2, F'_2), \dots, (\pi'_k, F'_k)$.

A propriedade da identificabilidade de diversas classes de misturas são discutidas por vários autores, incluindo Teicher (1963), Yakowitz e Spragins (1968), Chandra (1977), AL-Hussaini e Ahmad (1981), Ahmad (1988), Atienza (2006), Sultan *et. al.* (2007), Ahmad *et. al.* (2010), entre outros.

2.7 Discriminante não-Linear

Neste seção, apresenta-se uma breve descrição sobre discriminante não-linear, o instrumental matemático envolvido e a função utilizada para classificação dos elementos de uma amostra. A sua aplicação depende da pré-classificação dos elementos, isto é, do conhecimento a priori de suas características gerais.

Busca-se através dos dados já conhecidos, a construção de uma regra objetiva de classificação de tal forma que ao se obter um novo elemento (não se conhece de que grupo pertence a priori), seja possível classificá-lo em algum dos grupos.

A análise discriminante é procedimento exploratório de dados utilizado para classificação de elementos de uma amostra ou população em grupos predefinidos. Permite-se a elaboração de uma função matemática chamada regra de classificação ou discriminação baseada na teoria de probabilidade (Mingoti, 2005).

A terminologia é creditada ao matemático e estatístico R. A. Fisher, que a utilizou em problemas relativos a separação de elementos. O principal objetivo da criação desta ferramenta é a separação de elementos, porém pode ser utilizada para classificação e alocação também.

Mingoti (2005) cita muitas situações nas quais o interesse está na discriminação de grupos, como na ecologia, ao classificar espécies muito parecidas ou na medicina, ao identificar doenças com sintomas muito parecidos. Atualmente, há grande aplicação no setor bancário, no qual se busca segmentar clientes e identificar aqueles de maior risco. Além disso, há diversas áreas com interesse em classificar elementos.

Martínez (2001) estudou método de reconhecimento de objetos utilizando-se de análise de discriminante e demonstrou a utilidade para aplicações em reconhecimento facial através de análise de componentes principais e análise de discriminante linear. Pohar *et al.* discutem a utilização de dois métodos: discriminante linear e função logística. Eles concluem que o primeiro é mais eficiente quando há variáveis independentes normalmente distribuídas, porém tem sua eficácia diminuída quando se trata de poucas categorias (2 ou 3). Por outro lado, o segundo é um pouco pior, mas sempre muito próximo do primeiro, sem contudo necessitar da condição de normalidade.

Dessa forma, este trabalho trata da análise de discriminante em misturas de distribuições a partir da hipótese que os elementos advêm de uma distribuição Beta. Este assunto foi introduzido por O'Neill (1978), que obteve a função discriminante para mistura de distribuições normais, mais recentemente os trabalhos de Dean *et al.* (2006), Chapelle *et al.* (2006) and Liang *et al.* (2007), que tratam da aplicação de discriminantes em Machine Learning.

2.7.1 Função discriminante para mistura de distribuições

A função discriminante de uma mistura de distribuições é extremamente útil para classificar de qual população (componente) um valor amostral pertence. Suponha que há duas populações π_1 e π_2 com as correspondentes densidades $f_i(x)$, $i = 1, 2$. Ahmad e Abd-Elrahman (1994), Ahmad *et al.* (2010) e Sultan e Al-Moisheer (2013) utilizaram a função discriminante para uma mistura de distribuições com duas componentes, descrita a seguir:

A probabilidade do indivíduo x de origem desconhecida pertencer a π_i , $i = 1, 2$ é dada por:

$$P(x \in \pi_i) = \frac{p_i f_i(x)}{p_1 f_1(x) + p_2 f_2(x)}.$$

A função logística, devido a sua construção, assume valores entre 0 a 1. Desta maneira,

a função de probabilidade é expressa em termos da função logística, com interesse de simplificar a função discriminante. Desta maneira, tem-se:

$$P(x \in \pi_1) = \frac{1}{1 + \exp(a - bz)},$$

$$P(x \in \pi_2) = \frac{\exp(a - bz)}{1 + \exp(a - bz)},$$

sendo a, b funções do parâmetro das componentes e z uma função de x . Se z for uma função linear de x tem-se uma função discriminante linear, caso z seja função não linear de x , a função discriminante é não linear. Para definir de qual população provém um indivíduo x de origem desconhecida, utiliza-se uma regra de classificação. Classifica-se x em π_1 se $P(x \in \pi_1) > P(x \in \pi_2)$ ou $a - bz < 0$, pois as desigualdades são equivalentes:

$$P(x \in \pi_1) > P(x \in \pi_2) \Leftrightarrow$$

$$\frac{1}{1 + \exp(a - bz)} > \frac{\exp(a - bz)}{1 + \exp(a - bz)} \Leftrightarrow$$

$$\exp(a - bz) < 1 \rightarrow a - bz < 0.$$

De forma análoga, uma observação é classificada em π_2 se $a - bz > 0$. Desse modo, a função discriminante de uma mistura de distribuições é definida por:

$$NL(x) = a - bz, \quad (2.11)$$

que permite classificar x em π_1 se $NL(x) < 0$ e em π_2 se $NL(x) \geq 0$. Além disso, pode-se observar que x será classificado em π_1 se $P(x \in \pi_1) > \frac{1}{2}$, pois

$$P(x \in \pi_1) > \frac{1}{2} \iff (1 + \exp(NL(x)))^{-1} > \frac{1}{2}$$

$$\iff 1 + \exp(NL(x)) < 2$$

$$\iff \exp(NL(x)) < 1 \iff NL(x) < 0.$$

Analogamente, x será classificado em π_2 se $P(x \in \pi_2) \leq \frac{1}{2}$.

2.7.2 Erros de classificação

Todo processo de classificação traz consigo um possível erro de classificação. Portanto, é importante definir o erro de se classificar uma observação em uma população quando, na realidade, ela pertence a outra. Classifica-se uma observação x em π_1 , primeira população (componente), se a função discriminante é menor que zero, $NL(x) < 0$. Desta maneira, a probabilidade de classificar erroneamente uma observação x para π_1 , utilizando a função discriminante $NL(x)$, é dada por

$$e_1 = P(NL(x) > 0 | x \in \pi_1).$$

Analogamente, a probabilidade de classificar de forma errada uma observação x para π_2 , por meio de função discriminante $NL(x)$, é obtida por

$$e_2 = P(NL(x) < 0 | x \in \pi_2).$$

Essa é a medida que verifica o erro tipo I e II no modelo. Quanto menor forem essas probabilidades, melhor será a função de discriminação. Existem vários procedimentos para estimar as probabilidades de classificação incorretas. Mingoti (2015) cita 3 métodos usuais em softwares estatísticos:

1. Método da Resubstituição;
2. Validação Cruzada - Método de Holdout;
3. Validação Cruzada - Método de Lachenbruch.

Método da Resubstituição

Este método utiliza-se da frequência de dados mal classificados. A tabela 2.1 possui o resultado de determinada classificação. A quantidade n_{11} são os elementos do grupo de origem 1 e que foram corretamente classificados a este grupo. Já a quantidade n_{12} refere-se aos elementos do grupo de origem 1 que foram incorretamente classificados no grupo 2. O mesmo raciocínio vale para as quantidades n_{21} e n_{22} .

Com base nesses dados, obtêm-se as estimativas das probabilidades de ocorrência dos erros tipo 1 e 2,

$$\hat{p}(2|1) = \frac{n_{12}}{n_1} \quad e \quad \hat{p}(1|2) = \frac{n_{21}}{n_2}. \quad (2.12)$$

Define-se a probabilidade global de erro como

$$\hat{p}(erro) = \frac{n_{21} + n_{12}}{n}. \quad (2.13)$$

Este método em geral é o mais comum na avaliação da qualidade do estimador ou da regra de classificação, além de ser o método escolhido para avaliar os modelos simulados neste trabalho, em virtude de sua simplicidade na aplicação quando o dado é classificado. Os métodos alternativos a seguir são procedimentos de validação cruzada, que tem como objetivo melhorar as estimativas (2.12) e (2.13).

Tabela 2.1: Tabela das frequências dos erros de classificação

População de Origem	População Classificada			
	Grupo	1	2	Total
	1	n_{11}	n_{12}	n_1
2	n_{21}	n_{22}	n_2	

Validação cruzada - método Holdout

Este método é a divisão do conjunto total de dados em dois subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros) e outro para teste (validação). É possível separar em partes iguais ou não. Uma proporção muito comum é considerar 70% dos dados para treinamento e os 30% restantes para teste. Ou seja, a primeira servirá para construção da regra de discriminação e a segunda para validação, que será a estimação das probabilidades de classificação incorretas.

É possível utilizar as medidas de erro do tipo I em (2.12) e erro tipo II em (2.13) na amostra de validação, para verificar se a regra construída foi bem calibrada.

As estimativas das probabilidades de classificações incorretas obtidas por este procedimento são não viciadas. A desvantagem é a redução do número da amostra original, que poderá diminuir acentuadamente a confiabilidade da regra de classificação. Johnson e Wichern (2002) recomendam que para melhor eficiência do método seja deixado de 25% a 50% dos elementos da amostra original para validação. Este método não é aconselhável quando trata-se de pequenas amostras, pois é possível que haja quantidade de elementos necessário para fazer a divisão dos elementos.

Tabela 2.2: Processo do método Holdout

Processo	70% (n_1+n_2)	30% (n_1+n_2)
Constrói-se	$NL(x)$	-
Valida-se	-	$NL(x)$

Validação Cruzada - Método de Lachenbruch

Este método também é conhecido como validação cruzada ou pseudo-jackknife, conforme Lachenbruch (1968). Este método consiste nos seguintes passos:

1. Retira-se um vetor de observações da amostra conjunta e utilizam-se os $(n_1 + n_2 - 1)$ elementos amostrais restantes para construir a função de discriminação;
2. Utiliza-se a regra de discriminação construída no passo 1 para classificar o elemento que ficou à parte na construção da regra de discriminação, verificando se a regra de discriminação conseguiu acertar a sua real procedência ou não;
3. Retorna-se o elemento amostral que foi retirado no passo 1 à amostra original e retira-se o segundo elemento amostral. Repete-se os passos 1 e 2.

Os passos 1, 2 e 3 devem ser repetidos para todos os $(n_1 + n_2)$ elementos, conforme a Tabela 2.3.

Tabela 2.3: Processo do método Lachenbruch

Etapas	Construção da Regra	Elemento Retirado para Avaliação	Avaliação da Regra
Etapa I	Constrói-se a $NL(x)$ com n_1+n_2-1 elementos	Último elemento	Bem Classificado ou Mal Classificado
Etapa II	Constrói-se a $NL(x)$ com n_1+n_2-1 elementos	Penúltimo elemento	Bem Classificado ou Mal Classificado
...
Etapa n_1+n_2	Constrói-se a $NL(x)$ com n_1+n_2-1 elementos	1º elemento	Bem Classificado ou Mal Classificado

Verifica-se a proporção de elementos bem classificados utilizando-se da soma das avaliações em cada etapa, isto é, a quantidade de elementos bem classificados nas $n_1 + n_2$ etapas. É possível calcular a taxa de erro da mesma forma que (2.13), tal que n_{12} refere-se a quantidade de itens classificados em no grupo 1, mas pertencem ao grupo 2 e n_{21} a quantidade de itens classificados no grupo 2, mas pertencem ao grupo 1. Assim, na etapa I o elemento retirado para avaliação, caso tenha sido classificado erroneamente, será contabilizado ou em n_{12} ou em n_{21} .

Há também o método chamado de "verdadeiro jackknife" proposto por Efron e Gong (1983). No entanto, Davison e Hall (1992) e Gananesanandam e Krzanowski (1990) mostram que a validação cruzada tem desempenho semelhante ao "verdadeiro jackknife", porém com menor esforço computacional.

Probabilidade global de acerto

A probabilidade global de acerto nada mais é do que a proporção de observações bem classificadas. Dada a tabela abaixo com o quantitativo dos dados bem e mal classificados.

Utilizando da Tabela 2.1 onde n_{11} e n_{22} são a quantidade de elementos corretamente classificados na população 1 e população 2, respectivamente, calcula-se a probabilidade de acerto global dada pela fórmula

$$\hat{p}(\text{acerto}) = \frac{n_{11} + n_{22}}{n_1 + n_2}.$$

Recomenda-se a estimação tanto das ocorrências dos erros tipo 1 e 2 de classificação incorretas, quanto a probabilidade de acerto global para avaliação da qualidade do discriminante. Assim, é possível ter alto acerto global, mas com alto erro em uma das funções de discriminação.

Capítulo 3

Discriminante não-linear para mistura de Beta

3.1 Mistura de Betas

Para uma população heterogênea, um modelo de mistura de distribuições é utilizado para representar sub-populações dentro de uma população geral. Graficamente, espera-se que tais distribuições tenham o formato a seguir.

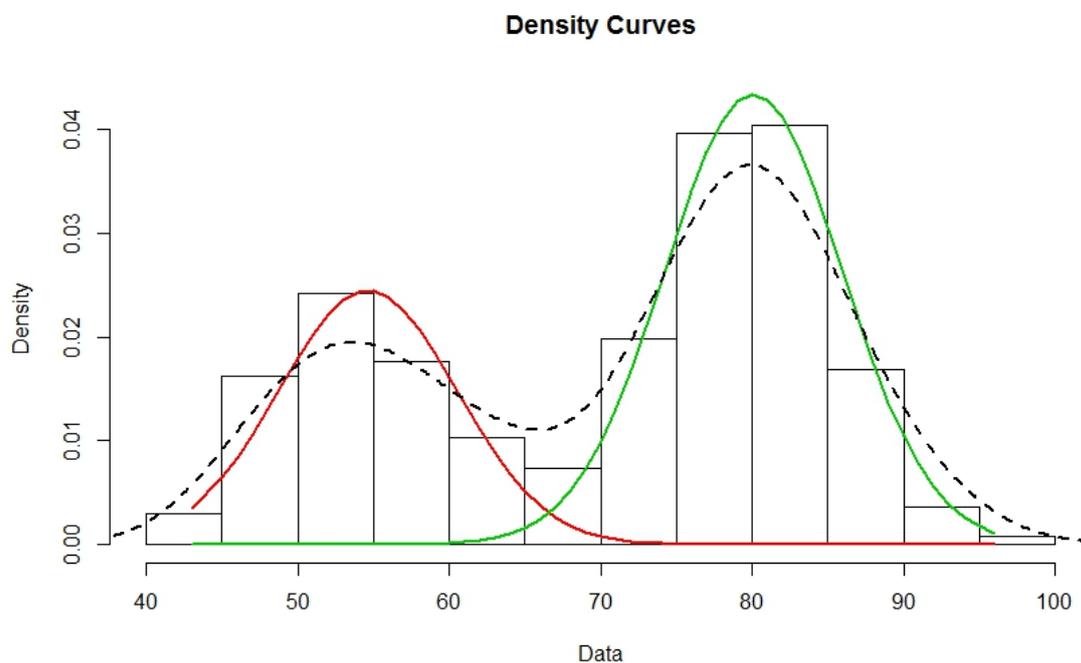


Figura 3.1: Mistura de distribuições

As linhas em verde e vermelho mostram duas distribuições, tal que a linha tracejada é o modelo de mistura, baseado na combinação convexa das componentes.

Para definir o modelo de mistura de distribuições, considere uma população X cuja f.d.p. é construída conforme (2.1) e (2.3),

$$h(x; \Theta) = p_1 \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} x^{\alpha_1-1} (1-x)^{\beta_1-1} + p_2 \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} x^{\alpha_2-1} (1-x)^{\beta_2-1}, \quad (3.1)$$

em que $\Theta = (p_1, p_2, \theta_1, \theta_2)$ e $\theta_i = (\alpha_i, \beta_i)$, $i = 1, 2$.

Momentos da mistura de Betas

Seja X um v.a com f.d.p $h(x, \Theta)$ de mistura de betas, o k -ésimo momento pode ser calculado por

$$\mathbb{E}(X^k) = \int_0^1 x^k h(x, \theta) dx.$$

Conforme (3.1) obtém-se o momento da variável X

$$\mathbb{E}(X^k) = \int_0^1 x^k \left(p_1 \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} x^{\alpha_1-1} (1-x)^{\beta_1-1} + p_2 \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} x^{\alpha_2-1} (1-x)^{\beta_2-1} \right) dx. \quad (3.2)$$

Separando as integrais de (3.2) e usando a Equação (2.2), obtém-se

$$\mathbb{E}(X^k) = p_1 \frac{\Gamma(\alpha_1 + \beta_1)\Gamma(\alpha_1 + k)}{\Gamma(\alpha_1 + \beta_1 + k)\Gamma(\alpha_1)} + p_2 \frac{\Gamma(\alpha_2 + \beta_2)\Gamma(\alpha_2 + k)}{\Gamma(\alpha_2 + \beta_2 + k)\Gamma(\alpha_2)}. \quad (3.3)$$

Logo, ao substituir $k = 1$ em (3.3), tem-se

$$\mathbb{E}(X) = p_1 \frac{\alpha_1}{\alpha_1 + \alpha_2} + p_2 \frac{\alpha_2}{\alpha_1 + \alpha_2}.$$

Analogamente se obtém o segundo momento e então a variância dada por

$$\text{Var}(X) = p_1^2 \frac{\alpha_1 \beta_1}{(\alpha_1 + \beta_1 + 1)(\alpha_1 + \beta_1)^2} + p_2^2 \frac{\alpha_2 \beta_2}{(\alpha_2 + \beta_2 + 1)(\alpha_2 + \beta_2)^2}.$$

3.2 Identificabilidade

Nesta seção utiliza-se uma caracterização da identificabilidade de uma classe de mistura finita de distribuições, dada por Atienza et al. (2006) no Lema a seguir.

3.3. MÁXIMA VEROSSIMILHANÇA DA MISTURA DE BETAS NÃO CLASSIFICADAS 19

Lema 3.2.1. (Atienza et al (2006)) Seja \mathcal{F} uma família de distribuições. Seja M uma transformação linear que transforma $F \in \mathcal{F}$ em uma função real ϕ_F com o domínio $S(F) \in \mathbb{R}^d$. Seja $S_0(F) = \{t \in S(F): \varphi_F(t) \neq 0\}$ e suponha que exista um ponto t_0 que verifique

$$t_0 \in \left[\bigcap_{1 \leq i \leq k} S_0(F_i) \right]', \quad (3.4)$$

para uma coleção finita de distribuições $F_1, \dots, F_k \in \mathcal{F}$, em que S' denota um conjunto de pontos de acumulação de S . Se a ordem

$$F_1 \prec F_2 \text{ se e somente se } \lim_{t \rightarrow t_0} \frac{\phi_{F_2}(t)}{\phi_{F_1}(t)} = 0, \quad (3.5)$$

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

Proposição 3.2.2. A família $\mathcal{H}_B =$ família de misturas com $\beta_i \neq \beta_j, i \neq j$ é identificável e seja M a aplicação linear que transforma $F \in \mathcal{F}$ no respectivo t -ésimo momento, em que $\phi_F(t) = \int_0^1 x^t dF(t)$. Então, de (3.4) tem-se que:

$$M[F(., \alpha, \beta)] = \phi_F(t) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + t)}{\Gamma(\alpha + \beta + t)}, \quad t \in (-\infty, +\infty),$$

isto é, $S_0(F)$ do Lema 3.2.1 é $(-\infty, +\infty)$ e o t_0 que satisfaz a condição (3.4) é $+\infty, t_0 = +\infty$.

Demonstração. Para verificar a condição (3.5) do Lema 3.2.1, reescrevemos $\phi_F(t)$ em termos da função Beta e usamos aproximação $B(x, y) \sim \Gamma(y)x^y$ para $x \Rightarrow +\infty$. Assim,

$$\phi_F(t) = \frac{1}{B(\alpha, \beta)} B(\alpha + t, \beta) \sim \frac{1}{B(\alpha, \beta)} \Gamma(\beta)(\alpha + t)^{-\beta}, \quad t \rightarrow +\infty. \quad (3.6)$$

Com a última expressão de (3.6) calcula-se o limite

$$\lim_{t \rightarrow +\infty} \frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = \lim_{t \rightarrow +\infty} \frac{B(\alpha_1, \beta_1)\Gamma(\alpha_2)(\alpha_2 + t)^{-\beta_2}}{B(\alpha_2, \beta_2)\Gamma(\alpha_1)(\alpha_1 + t)^{-\beta_1}} = 0,$$

se, e somente se, $\beta_1 < \beta_2$, ou seja, tem-se uma ordem total. Portanto pelo Lema 3.2.1 segue que \mathcal{H}_B é identificável. \square

3.3 Máxima verossimilhança da mistura de Betas não classificadas

A função máxima verossimilhança da mistura de densidades Beta, é obtida ao substituir (3.1) em (2.5) e é dada por:

$$L(\Theta; X) = \prod_{j=1}^n [h(x, \Theta)] = \prod_{j=1}^n \left[p_1 \frac{\Gamma(\alpha + \beta_1)}{\Gamma(\alpha)\Gamma(\beta_1)} x^{\alpha-1} (1-x)^{\beta_1-1} + p_2 \frac{\Gamma(\alpha + \beta_2)}{\Gamma(\alpha)\Gamma(\beta_2)} x^{\alpha-1} (1-x)^{\beta_2-1} \right].$$

Por conveniência, utilizaremos $\alpha_1 = \alpha_2 = \alpha$, tal que a verossimilhança torna-se

$$L(\Theta; X) = \frac{1}{\Gamma(\alpha)^n} \prod_{j=1}^n x_j^{\alpha-1} Q_j, \quad (3.7)$$

em que $Q_j = p_1 g_1(x_j; \alpha, \beta_1) + p_2 g_2(x_j; \alpha, \beta_2)$, e as funções $g(x_j; \alpha, \beta_i)$ são representadas por

$$g_i(x_j; \alpha, \beta_i) = \frac{\Gamma(\alpha + \beta_i)}{\Gamma(\beta_i)} (1 - x_j)^{\beta_i-1}, \quad i = 1, 2.$$

Aplicando a função logarítmica à Equação (3.7), obtém-se

$$\log L(\Theta; X) = -n \log(\Gamma(\alpha)) + (\alpha - 1) \sum_{j=1}^n \log(x_j) + \sum_{j=1}^n \log(Q_j).$$

A partir desta equação obtêm-se as derivadas parciais, conforme parâmetros do vetor $\Theta = (\alpha, \beta_1, \beta_2, p_1)$. As derivadas parciais obtidas são dadas por

$$\begin{aligned} \frac{\partial}{\partial p_1} \log(L(\Theta; X)) &= \sum_{j=1}^n \frac{g_1(x_j, \beta_1) - g_2(x_j, \beta_2)}{Q_j} = 0; \\ \frac{\partial}{\partial \beta_1} \log(L(\Theta; X)); &= \sum_{j=1}^n p_1 \frac{\partial g_1(x_j, \beta_1)}{\partial \beta_1} \frac{1}{Q_j} = 0; \\ \frac{\partial}{\partial \beta_2} \log(L(\Theta; X)); &= \sum_{j=1}^n p_2 \frac{\partial g_2(x_j, \beta_2)}{\partial \beta_2} \frac{1}{Q_j} = 0; \\ \frac{\partial}{\partial \alpha} \log L(\Theta; X); &= -n \frac{\partial}{\partial \alpha} \log(\Gamma(\alpha)) + \sum_{j=1}^n \log(x_j) + \frac{\partial}{\partial \alpha} \sum_{j=1}^n \log(Q_j) = 0. \end{aligned}$$

Por não ser possível resolver tais equações analiticamente, procura-se um método de estimação para facilitar os cálculos. Para isso, utilizar-se-á o algoritmo EM, pois permite encontrar estimativas a partir de dois passos. Os passos E e M, para modelos de mistura de distribuições, foram definidos na Seção 2.5, tal que repetem-se iterativamente até atingir um determinado critério de parada.

A função $Q(\cdot)$ definida em (2.10) é construída a partir de mistura de Betas, conforme (3.1).

3.3. MÁXIMA VEROSSIMILHANÇA DA MISTURA DE BETAS NÃO CLASSIFICADAS²¹

$$\begin{aligned}
 Q(\theta, \theta^{(k)}) &= \sum_{i=1}^n \log(p_1) P(Y = 1 | x_i, \alpha^{(k)}, \beta_1^{(k)}) \\
 &+ \sum_{i=1}^n \log(p_2) P(Y = 2 | x_i, \alpha^{(k)}, \beta_2^{(k)}) \\
 &+ \sum_{i=1}^n \log(f_1(x_i | \alpha^{(k)}, \beta_1)) P(Y = 1 | x_i, \alpha^{(k)}, \beta_1^{(k)}) \\
 &+ \sum_{i=1}^n \log(f_2(x_i | \alpha^{(k)}, \beta_2)) P(Y = 2 | x_i, \alpha^{(k)}, \beta_2^{(k)}).
 \end{aligned}$$

Observe que as quantidades $P(Y = 1 | x_i, \alpha^{(k)}, \beta_1^{(k)})$ e $P(Y = 2 | x_i, \alpha^{(k)}, \beta_2^{(k)})$ são valores que dependem somente da amostra, assim, foram trocados, respectivamente, por w_1 e w_2 . No passo M, se obtém as derivadas de $Q(\theta, \theta^{(k)})$ em relação a cada um dos parâmetros de interesse e iguala-se a zero. Então, ao derivar $Q(\theta, \theta^{(k)})$ com respeito a p_1 e resolvendo a equação, tem-se

$$\hat{p}_1^{k+1} = \frac{\sum_{i=1}^n P(Y = 1 | x_i, \alpha^{(k)}, \beta_1^{(k)})}{N}. \quad (3.8)$$

As equações obtidas para os parâmetros β_1 , β_2 , e α são, respectivamente:

$$\sum_{i=1}^n \frac{\partial \log [f_1(x_i | \alpha^{(k)}, \beta_1^{(k)})] P(Y = 1 | x_i, \alpha^{(k)}, \beta_1^{(k)})}{\partial \beta_1^{(k)}} = 0; \quad (3.9)$$

$$\sum_{i=1}^n \frac{\partial \log [f_2(x_i | \alpha^{(k)}, \beta_2^{(k)})] P(Y = 2 | x_i, \alpha^{(k)}, \beta_2^{(k)})}{\partial \beta_2^{(k)}} = 0; \quad (3.10)$$

$$\sum_{i=1}^n w_1 \frac{\partial \log [f_1(x_i | \alpha^{(k)}, \beta_1^{(k)})]}{\partial \alpha} + \sum_{i=1}^n w_2 \frac{\partial \log [f_2(x_i | \alpha^{(k)}, \beta_2^{(k)})]}{\partial \alpha} = 0. \quad (3.11)$$

Desenvolvendo a derivada parcial de (3.9), obtém-se

$$\begin{aligned}
 &\sum_{i=1}^n w_1 \left[\frac{\partial \Gamma(\alpha^{(k)} + \beta_1^{(k+1)})}{\partial \beta_1^{(k+1)}} - \psi(\beta_1^{(k+1)}) + \log(1 - x_i) \right] \\
 &= \left[\sum_{i=1}^n w_1 \psi(\alpha^{(k)} + \beta_1^{(k+1)}) - \sum_{i=1}^n w_1 \psi(\beta_1^{(k+1)}) + \sum_{i=1}^n w_1 \log(1 - x_i) \right] = 0. \quad (3.12)
 \end{aligned}$$

Analogamente, desenvolvendo as derivadas parciais de (3.10) e (3.11), obtém-se

$$\begin{aligned}
& \sum_{i=1}^n w_2 \left[\frac{\partial \Gamma(\alpha^{(k)} + \beta_2^{(k+1)})}{\partial \beta_2^{(k+1)}} - \psi(\beta_2^{(k+1)}) + \log(1 - x_i) \right] \\
&= \left[\sum_{i=1}^n w_2 \psi(\alpha^{(k)} + \beta_2^{(k+1)}) - \sum_{i=1}^n w_2 \psi(\beta_2^{(k+1)}) + \sum_{i=1}^n w_2 \log(1 - x_i) \right] = 0 \quad (3.13)
\end{aligned}$$

e

$$\begin{aligned}
&= \left[\sum_{i=1}^n w_1 \psi(\alpha^{(k+1)} + \beta_1^{(k)}) - \sum_{i=1}^n w_1 \psi(\alpha^{(k+1)}) + \sum_{i=1}^n w_1 \log(x_i) \right] \\
&+ \left[\sum_{i=1}^n w_2 \psi(\alpha^{(k+1)} + \beta_2^{(k)}) - \sum_{i=1}^n w_2 \psi(\alpha^{(k+1)}) + \sum_{i=1}^n w_2 \log(x_i) \right] = 0. \quad (3.14)
\end{aligned}$$

Assim, no passo $(k + 1)$, as atualizações de $p_1^{(k+1)}$, $\beta_1^{(k+1)}$, $\beta_2^{(k+1)}$ e $\alpha^{(k+1)}$ são obtidos de (3.8) e ao resolver as equações (3.12), (3.13) e (3.14), respectivamente. Para isso, é necessário escolher valores iniciais para os parâmetros p^0 , β_1^0 , β_2^0 e α^0 .

A matriz de covariância dos parâmetros $\theta = (p_1, \beta_1, \beta_2, \alpha)$ pode ser obtido através do valor esperado das segundas derivadas do log-verossimilhança. Note que essa matriz é a inversa da matriz de informação de Fisher.

$$\Sigma = I^{-1}(\theta) = \left[-E \left(\frac{\partial \log(L(\theta))}{\partial \theta_i \partial \theta_j} \right) \right]^{-1}, \quad i, j = 1, 2, 3, 4. \quad (3.15)$$

3.4 EMV para mistura de Betas em amostras classificadas

Dada uma amostra de tamanho n do modelo de mistura em (3.1) na qual sabemos que n_1 observações são provenientes da primeira componente π_1 e n_2 da segunda π_2 , em que $n = n_1 + n_2$, então conhecemos o grupo no qual cada elemento da amostra proveio.

Nesta situação, o procedimento de estimação paramétrica tratará a função de verossimilhança da mistura como o produto de duas verossimilhanças, em que cada uma conterá somente os elementos associados a cada densidade. Se não houvesse o parâmetro em comum nas duas componentes, então o processo de estimação poderia ser separado ao resolver o problema de maximização para cada densidade Beta separadamente. Porém, pelo fato de haver essa restrição, as estimativas de máxima verossimilhança terão que ser obtidas utilizando a função de verossimilhança referente a mistura. A função de verossimilhança para esse tipo de amostragem é dada por

$$\begin{aligned}
L(\theta) &= L_1(\theta_1)L_2(\theta_2) \\
&= \prod_{j=1}^{n_1} f_1(x_j; \theta_1) \prod_{j=1}^{n_2} f_2(x_j; \theta_2) \\
&= \prod_{j=1}^{n_1} B_1(\alpha, \beta_1)x_j^{\alpha-1}(1-x_j)^{\beta_1-1} \prod_{j=1}^{n_2} B_2(\alpha, \beta_2)x_j^{\alpha-1}(1-x_j)^{\beta_2-1}.
\end{aligned}$$

Aplicando logaritmo na verossimilhança, tem-se

$$\begin{aligned}
\mathcal{L}(\theta) &= -n_1 \log(B_1(\alpha, \beta_1)) + (\alpha - 1) \sum_{j=1}^{n_1} \log(x_j) + (\beta_1 - 1) \sum_{j=1}^{n_1} \log(1 - x_j) \\
&\quad - n_2 \log(B_2(\alpha, \beta_2)) + (\alpha - 1) \sum_{j=1}^{n_2} \log(x_j) + (\beta_2 - 1) \sum_{j=1}^{n_2} \log(1 - x_j) \\
&= -n_1 \log(\Gamma(\alpha)) - n_1 \log(\Gamma(\beta_1)) + n_1 \log(\Gamma(\alpha + \beta_1)) \\
&\quad + (\alpha - 1) \sum_{j=1}^{n_1} \log(x_j) + (\beta_1 - 1) \sum_{j=1}^{n_1} \log(1 - x_j) \\
&\quad - n_2 \log(\Gamma(\alpha)) - n_2 \log(\Gamma(\beta_2)) + n_2 \log(\Gamma(\alpha + \beta_2)) \\
&\quad + (\alpha - 1) \sum_{j=1}^{n_2} \log(x_j) + (\beta_2 - 1) \sum_{j=1}^{n_2} \log(1 - x_j). \tag{3.16}
\end{aligned}$$

O estimador de p_1 é obtido ao igualar a zero a derivada da equação (3.16) em relação a p_1 e resolvê-la.

$$\frac{\partial l(\theta)}{\partial p_1} = \frac{n_1}{p_1} - \frac{n_2}{1-p_1} = 0 \quad \Rightarrow \quad \hat{p}_1 = \frac{n_1}{n}. \tag{3.17}$$

Em relação a β_1 e β_2 aplicam-se a derivada em (3.16) em relação a estes parâmetro, respectivamente, e obtém-se

$$\frac{\partial l(\theta)}{\partial \beta_1} = n_1 \psi(\alpha + \beta_1) - n_1 \psi(\beta_1) + \sum_{j=1}^{n_1} \log(1 - x_j) = 0, \tag{3.18}$$

$$\frac{\partial l(\theta)}{\partial \beta_2} = n_2 \psi(\alpha + \beta_2) - n_2 \psi(\beta_2) + \sum_{j=1}^{n_2} \log(1 - x_j) = 0, \tag{3.19}$$

e em relação a α

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \alpha} = & n_1 \psi(\alpha + \beta_1) - n_1 \psi(\alpha) + \sum_{j=1}^{n_1} \log(x_j) \\ & + n_2 \psi(\alpha + \beta_2) - n_2 \psi(\alpha) + \sum_{j=1}^{n_2} \log(x_j) = 0. \end{aligned} \quad (3.20)$$

Portanto a estimativa de p é igual ao número de observações da primeira componente n_1 sobre o tamanho da amostra n . Contudo para achar as estimativas de α , β_1 e β_2 é necessário utilizar um método numérico como Newton-Rhapson.

3.5 Função discriminante ótima para mistura de Betas

A função discriminante para mistura de duas Betas, definida em (2.11), será utilizada para classificar observações das subpopulações presentes na mistura.

A construção da função discriminante $NL(x) = a + bz$ será o objetivo desta seção, dado que através dela será possível classificar elementos em seus respectivos componentes. A aplicação desta metodologia para construção da regra, conforme Equação (2.10) e resultados obtidos em (3.7), resulta em

$$\begin{aligned} P(Y = 1 | x_j, \theta^{(k)}) = P(x_j \in \pi_1) &= \frac{p_1 g_1(x_j; \beta_1)}{p_1 g_1(x_j; \beta_1) + p_2 g_2(x_j; \beta_2)} \\ &= \frac{1}{1 + \frac{\frac{p_2 \Gamma(\alpha + \beta_2) (1 - x_j)^{\beta_2 - 1}}{\Gamma(\beta_2)}}{\frac{p_1 \Gamma(\alpha + \beta_1) (1 - x_j)^{\beta_1 - 1}}{\Gamma(\beta_1)}}}. \end{aligned}$$

De forma equivalente

$$P(x_j \in \pi_1) = \frac{1}{1 + \left(\exp \left[\log \left(\frac{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)}{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)} \right) + (\beta_2 - \beta_1) \log(1 - x_j) \right] \right)}. \quad (3.21)$$

Daí, $NL(x) = a + bz$, sendo:

1. $a = \log \left(\frac{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)}{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)} \right)$;
2. $b = \beta_2 - \beta_1$;
3. $z = \log(1 - x_j)$.

Isto é, a função discriminante ótima, que é usada quando conhecemos todos os parâmetros da densidade estudada, será dada por:

$$NL(x) = \log \left(\frac{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)}{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)} \right) + (\beta_2 - \beta_1) \log(1 - x_j). \quad (3.22)$$

Como z é uma função logarítmica, então trata-se de um discriminante não-linear. Assim, caso $NL(x) \leq 0$, então classificamos o elemento x_j à distribuição π_1 . Caso $NL(x) > 0$, então o elemento x_j será classificado à distribuição π_2 .

3.6 Função discriminante - caso particular

Quando os parâmetros da distribuição são desconhecidos, a função de discriminante utilizada dependerá dos parâmetros estimados da densidade. A estimativa dos parâmetros é feita através dos métodos da máxima verossimilhança, porém dependendo de como os dados são amostrados a função de verossimilhança acaba tendo formas diferentes e conseqüentemente as estimativas dos parâmetros ficam diferentes.

Nos módulos anteriores foram apresentados dois tipos de amostragem diferentes para obtenção das estimativas dos parâmetros que podem ser obtidos por amostras classificadas e amostras misturadas (não classificadas).

A obtenção da função discriminante não linear para amostras classificadas é obtida da mesma expressão do discriminante não linear ótimo, através das estimativas das equações (3.17), (3.18), (3.19) e (3.20). A função de discriminante não linear para amostras classificadas é dada por

$$NL(x_j) = \tilde{a} + \tilde{b}\tilde{z}, \quad (3.23)$$

sendo

1. $\tilde{a} = \log \left(\frac{\tilde{p}_2 \Gamma(\tilde{\alpha} + \tilde{\beta}_2) \Gamma(\tilde{\beta}_1)}{\tilde{p}_1 \Gamma(\tilde{\alpha} + \tilde{\beta}_1) \Gamma(\tilde{\beta}_2)} \right)$;
2. $\tilde{b} = \tilde{\beta}_2 - \tilde{\beta}_1$;
3. $\tilde{z} = \log(1 - x_j)$.

A regra de classificação diz que, para determinado elemento x_j , caso $NL(x_j) > 0$, então o elemento será classificado na segunda população π_2 . Caso contrário, $NL(x_j) < 0$, o elemento será classificado na primeira população.

Quando se fala em uma amostra não classificada significa que há uma amostra de n elementos e não se sabe de qual grupo ele é originado. O que se sabe é que, para cada observação, o valor da densidade da mistura assume determinado valor conhecido. Diferentemente, na amostra classificada, sabe-se de qual subpopulação a observação foi gerada.

Percebe-se, então, que o discriminante não linear para amostras misturadas (não classificadas) é dado pela mesma expressão que a amostra classificada, porém com os parâmetros

sendo estimados pelas equações (3.8), (3.9), (3.10) e (3.11). Estas são provenientes das estimativas utilizando o algoritmo EM. Assim, o discriminante para o modelo com amostras misturadas é dado por

$$NL(x_j) = \hat{a} + \hat{b}\hat{z}, \quad (3.24)$$

sendo

1. $\hat{a} = \log \left(\frac{\hat{p}_2 \Gamma(\hat{\alpha} + \hat{\beta}_2) \Gamma(\hat{\beta}_1)}{\hat{p}_1 \Gamma(\hat{\alpha} + \hat{\beta}_1) \Gamma(\hat{\beta}_2)} \right)$;
2. $\hat{b} = \hat{\beta}_2 - \hat{\beta}_1$;
3. $\hat{z} = \log(1 - x_j)$.

Os elementos x_j serão classificados em π_2 , caso $NL(x_j) > 0$, e em π_1 caso contrário.

3.7 Erro de classificação

Anteriormente, foram encontradas as funções do discriminante não linear ótimo, para amostras misturadas e para amostras classificadas. A classificação dos elementos em determinado grupo será determinada pela probabilidade de pertencer àquele grupo. Portanto, por se tratar de probabilidade, estamos sujeitos a erros de estimação. Para o mesmo conjunto de dados é possível construir várias funções para discriminar as observações com objetivo de minimizar o erro.

Uma dada observação é classificada em π_1 , se $NL(x_j) \leq 0$, para $k = \{\text{o (ótimo), c (classificada), m (mistura)}\}$, e é classificada em π_2 , caso contrário. Dessa forma, o erro de classificar um elemento na componente 2, dado que ele pertence à primeira distribuição, é obtido da seguinte maneira

$$\begin{aligned} e_{1k} &= P(NL(x_j) > 0 | x_j \in \pi_1) \\ &= P \left(\log \left(\frac{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)}{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)} \right) + (\beta_2 - \beta_1) \log(1 - x_i) > 0 | x_j \in \pi_1 \right) \\ &= P \left((\beta_2 - \beta_1) \log(1 - x_i) > -\log \left(\frac{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)}{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)} \right) \right) \\ &= P \left(\log(1 - x_i) > \frac{-1}{\beta_2 - \beta_1} \left[\log \left(\frac{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)}{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)} \right) \right] \right) \\ &= P \left(\log(1 - x_i) > \left[\log \left(\frac{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)}{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)} \right) \right]^{\frac{1}{\beta_2 - \beta_1}} \right) \\ &= P \left(1 - x_i > \left[\frac{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)}{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)} \right]^{\frac{1}{\beta_2 - \beta_1}} \right) \\ &= P \left(x_i < 1 - \left[\frac{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)}{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)} \right]^{\frac{1}{\beta_2 - \beta_1}} \right). \end{aligned}$$

Sendo assim, podemos dizer que a acumulada $F(\cdot)$ de Beta pode ser escrita

$$e_{1k} = F_{\beta}(w_k),$$

sendo $w_k = 1 - \left[\frac{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)}{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)} \right]$. A função de distribuição Beta (Incompleta) F_{β} é dada por

$$F_{\beta}(w_k, \alpha, \beta_i) = \frac{\int_0^{w_k} t^{\alpha} (1-t)^{\beta_i-1} dt}{B(\alpha, \beta_i)}.$$

Isto é,

$$\left[\frac{p_1 B_1(\alpha + \beta_1, \beta_1)}{p_2 B_2(\alpha + \beta_2, \beta_2)} \right]^{\frac{1}{\beta_2 - \beta_1}} > 1,$$

tal que $B_i(\alpha + \beta_i, \beta_i)$, para $i = \{1, 2\}$, são funções Beta. Por questões de simplificação será utilizado B_i para denotar a função $B(\alpha, \beta_i)$. Assim, para atender a condição de $w_k > 1$ é necessário satisfazer duas condições abaixo:

1. $\frac{p_1 B_1}{p_2 B_2} > 1$ e $\beta_1 < \beta_2$;
2. $\frac{p_1 B_1}{p_2 B_2} < 1$ e $\beta_1 > \beta_2$.

As condições acima são equivalentes à:

1. $B_1 - \frac{p_2 B_2}{p_1} > 0$ e $\beta_1 - \beta_2 < 0$;
2. $B_2 - \frac{p_1 B_1}{p_2} > 0$ e $\beta_1 - \beta_2 > 0$.

Com essas condições satisfeitas, o erro de classificação na primeira componente (e_{1k}), onde $k = \{\text{o (ótimo), c (classificada), m (mista)}\}$, é dado por

$$e_{1k} = P \left(x_j < 1 - \left[\frac{p_1 \Gamma(\alpha + \beta_1) \Gamma(\beta_2)}{p_2 \Gamma(\alpha + \beta_2) \Gamma(\beta_1)} \right]^{\frac{1}{\beta_2 - \beta_1}} \right). \quad (3.25)$$

$$e_{1k} = F_{\beta}(x_j; \beta_{1k}, \alpha_k).$$

Observe que o erro e_{1k} depende dos parâmetros $(p_1, \beta_{1k}, \alpha_k)$, pois a probabilidade calculada é dado que x_j pertença a π_1 . Para que erro de classificação na segunda componente (e_{2k}) esteja bem definido, é preciso das mesmas condições que definimos para e_{1k} . Dessa forma, e_{2k} é dado pela expressão abaixo

$$e_{2k} = 1 - F(x_j; \beta_{2k}, \alpha_k). \quad (3.26)$$

Como na prática dificilmente saberemos de qual subpopulação cada observação provém, calcular somente e_{1k} e e_{2k} não ajudariam no problema de classificação. Dessa maneira, seria útil calcular a probabilidade de classificar uma observação como pertencendo a uma subpopulação e ela pertencer à outra. Para calcular esse tipo de erro, define-se ϵ_{1k} como

a probabilidade de classificarmos uma observação em π_2 e ela pertencer a π_1 , e ϵ_{2k} como a probabilidade de classificar em π_1 , quando na verdade pertence a π_2 . A seguir, apresentamos como obter ϵ_{1k} em função de e_{1k} .

$$\begin{aligned}\epsilon_{1k} &= P(NL(x_j) > 0 \cap x_j \in \pi_1) \\ \epsilon_{1k} &= P(x_j \in \pi_1)P(NL(x_j) > 0 | x_j \in \pi_1) \\ \epsilon_{1k} &= p_k e_{1k}.\end{aligned}$$

Analogamente, se obtém ϵ_{2k} , que é dado pela expressão

$$\epsilon_{2k} = (1 - p_k)e_{2k}.$$

Portanto, a probabilidade total de cometer um erro de classificação pode ser definida como classificar uma observação de π_1 em π_2 ou o contrário. Logo, obtém-se

$$\epsilon_k = \epsilon_{1k} + \epsilon_{2k} = p_k e_{1k} + (1 - p_k)e_{2k} \quad (3.27)$$

Capítulo 4

Simulação e Aplicação

4.1 Geração dos dados e estimação

No intuito de avaliar a eficácia do discriminante $NL(x_j)$ dado pela equação (2.11), para o modelo $h(x, \theta)$, definido em (3.1), foram realizados experimentos com dados simulados. A simulação tem como objetivo averiguar a qualidade das estimativas dos parâmetros da mistura para amostras classificadas e misturadas (não classificadas), além de verificar as taxas de erros de classificação obtidas através das diferentes funções de discriminante: ótima, classificada e mistura.

O software utilizado foi o R na versão 3.4.3 (<https://www.r-project.org/>). Há a necessidade de utilizar alguns pacotes específicos para resolver as estimações exigidas pelo método como *rootSolve*, *VGAM* e *xtable*. As programações utilizadas na simulação e aplicação dos dados podem ser conferidas em anexo.

A simulação foi realizada para 9 (nove) diferentes combinações de parâmetros, variando o tamanho da amostra entre 50, 100, 500 e 1000, com a finalidade de obter os mais diferentes tipos de distribuição.

Os valores de α , β_1 e β_2 foram escolhidos de forma a obter gráficos bimodais bem definidos, nos quais pudéssemos analisar os resultados de forma mais objetiva possível, obtendo gráficos das mais variadas formas, conforme gráfico abaixo. Para a estimação de Θ do modelo $h(x, \Theta)$ foram considerados os seguintes parâmetros.

Tabela 4.1: Parâmetros da simulação.

Modelo	p	α	β_1	β_2
1	0.1	20	5	40
2	0.2	50	15	40
3	0.4	10	10	30
4	0.4	10	5	40
5	0.6	25	10	40
6	0.6	5	10	30
7	0.7	2	2	20
8	0.8	10	10	30
9	0.9	10	5	40

Os modelos teóricos dos parâmetros escolhidos são representados graficamente pelas densidades abaixo. Define-se modelo 1 aquele referente a Θ_1 , modelo 2 aquele referente a Θ_2 e assim sucessivamente.

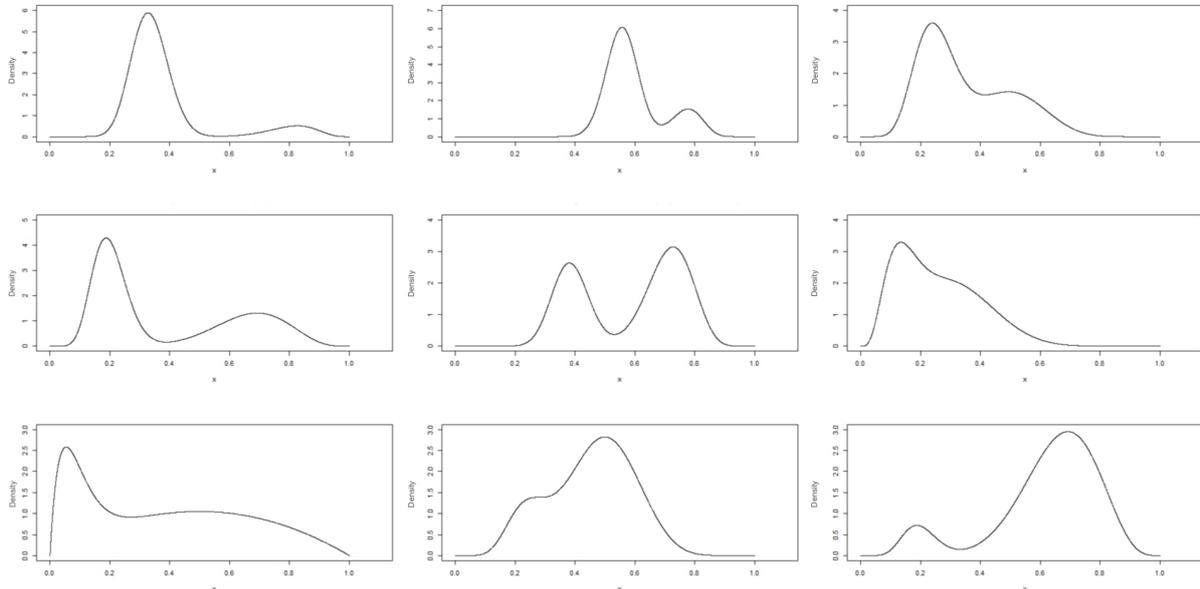


Figura 4.1: Densidades $h(x, \theta)$, conforme (3.1), com Θ_1 a Θ_9 definidos na Tabela 4.1

O processo de simulação de dados envolve a geração de valores de uma uniforme e a obtenção da função inversa da distribuição Beta, conforme algoritmo a seguir:

1. Gerar duas observações U_1 e U_2 que seguem a distribuição $U(0, 1)$;
2. Se $U_1 < p$, então é gerada uma observação da primeira componente utilizando-se da função inversa, isto é, $F_1^{-1}(U_2, \alpha, \beta_1)$;
3. Se $U_1 > p$, então é gerada uma observação da segunda componente $F_2^{-1}(U_2, \alpha, \beta_2)$.

Repete-se n vezes o processo para gerar n dados. As observações que foram geradas pela primeira componente são juntadas em um vetor chamado de $vetor_1$ e as observações geradas pela segunda num $vetor_2$. A junção das duas informações gera as observações de dados misturados na qual, em tese, não há conhecimento sobre a origem do dado. Este vetor de misturas será usado para a aplicação do algoritmo EM, pois estamos trabalhando com uma amostra não classificada, enquanto os vetores $vetor_1$ e $vetor_2$ serão usados para as estimativas via EMV para amostras classificadas.

Com as n observações geradas acima, (x_1, \dots, x_n) , a estimação por máxima verossimilhança de Θ_n via algoritmo EM consiste em:

1. Escolher os valores iniciais para os parâmetros $p_1^0, \alpha^0, \beta_1^0$ e β_2^0 ;
2. **Passo E:** Calcular as proporções em cada iteração $(k+1)$: $P(Y = 1|x_i, \alpha^{(k)}, \beta_1^{(k)})$ e $P(Y = 2|x_i, \alpha^{(k)}, \beta_2^{(k)})$;

3. **Passo M:** No passo $(k + 1)$ as atualizações de $p_1^{(k+1)}$, $\beta_1^{(k+1)}$, $\beta_2^{(k+1)}$ e $\alpha^{(k+1)}$ são obtidas de (3.8) e ao resolver as equações (3.12), (3.13) e (3.14), respectivamente.
4. O algoritmo EM chegará ao fim quando $|\hat{\theta}^k - \hat{\theta}^{k-1}| < 0.01$, sendo k o número de iterações realizadas.

Observe que não há informação sobre a quantidade de dados em cada componente, por isso a necessidade de obter as proporções no passo E da mistura (3.1).

Por outro lado, para amostras classificadas, a quantidade de observações geradas de cada subpopulação é conhecida e as observações correspondentes a subpopulação 1 e 2 são dadas pelos vetores $vetor_1$ e $vetor_2$, respectivamente. Assim, as estimativas de Θ_n são obtidas ao solucionar as Equações (3.17), (3.18), (3.19) e (3.20).

Os resultados da estimação de Θ_1 a Θ_9 , bem como o erro quadrático médio (EQM) e o viés são utilizados para avaliar a estimação de Θ , conforme Tabelas (4.2) a (4.5), (4.10) e (4.12) a (4.15). O EQM e o viés são calculados por

$$MSE_k = \sum_{k=1}^n \frac{(\hat{\theta}_k - \theta)^2}{n}, \quad (4.1)$$

$$Vies_k = \sum_{k=1}^n \frac{(\hat{\theta}_k - \theta)}{n}, \quad (4.2)$$

em que $k = \{m \text{ (mistura), } c \text{ (classificada)}\}$. Além disso, foi construído o gráfico com a comparação entre a densidade correspondente aos parâmetros escolhidos e as densidades ajustadas para amostras classificadas e não classificadas.

Os resultados mostram que os métodos de estimação utilizando amostras não classificadas (Mistura) e classificadas (Classificada) produziram estimativas próximas, com base no viés e EQM.

Observou-se que para as amostras com $n = 50$, as estimativas distanciaram-se do verdadeiro valor, conforme Tabelas (4.2) a (4.5), Tabela (4.10) e Tabelas (4.12) a (4.15). Assim, sugere-se que as estimações tenham aproximadamente 100 elementos para que se obtenha bons resultados.

Por fim, o ponto central do estudo diz respeito a avaliação da função de discriminante. Cada observação x_j será avaliada pela função discriminante da amostra classificada e da amostra misturada, conforme função dada em (3.23) e (3.24), respectivamente.

Deve-se, então, avaliar a qualidade da classificação dos elementos através da probabilidade de erro da classificação, erro tipo I e erro tipo II, conforme equações (3.25) e (3.26), respectivamente. Define-se este erro de **erro teórico**. Como é possível rastrear a origem do dado, calcula-se, também, as proporções de elementos mal classificados pela função discriminante. Define-se este erro de **erro real**, definido em (2.13), tal que

$$E_{ik} = \frac{\text{n}^\circ \text{ observações mal classificadas}}{n}, \quad (4.3)$$

em que $k = \{m \text{ (mistura), } c \text{ (classificada)}\}$ e $i = \{1, 2\}$ para cada componente.

As quantidades acima são avaliadas componente por componente, assim é necessário uma medida que avalie conjuntamente o erro tipo I e o erro tipo II. Esta medida é definida como **erro total**, dada pela Equação (3.27).

As Tabelas (4.6) a (4.9), a Tabela (4.11) e as Tabelas (4.16) a (4.19) mostram a magnitude de cada um destes erros, de forma a avaliar a qualidade da classificação da função discriminante.

De forma geral, observa-se que o discriminante conseguiu classificar efetivamente os dados, com exceção do modelo 2, no caso da mistura. A proporção de erros das mistura oscilou entre 5% e 15%, o que mostra a qualidade da discriminação.

Análise das estimativas de $h(x, \Theta)$ para $\Theta_1, \Theta_3, \Theta_4$ e Θ_5

Estes modelos demonstraram bom ajuste das estimativas, exceto quando se utiliza amostra $n = 50$, pois ficaram longe do alvo (parâmetro), conforme mostrado pelo viés e EQM. Isso indica que a quantidade de dados deve ser aumentada para melhoria da estimação. Tanto a estimativa pelo algoritmo EM, quanto a estimativa por EMV obtiveram resultados semelhantes, porém, os resultados com amostras classificadas tendem a ser ligeiramente melhores, pois possuem mais informação.

Por outro lado, as estimativas para $n = \{100, 500, 1000\}$ se mostraram acuradas e precisas. As Tabelas (4.2) a (4.5) e as Figuras 4.2 a 4.5 apresentam os resultados das estimativas e do erro quadrático médio que corroboram com a análise. Os gráficos da evolução do EQM, anexos (A.1) a (A.4), demonstram a melhora dos resultados quando n cresce.

Tabela 4.2: Estimativa dos parâmetros de $h(x, \Theta_1)$.

Estimativas	mistura				classificada			
	\mathbf{p}	α	β_1	β_2	\mathbf{p}	α	β_1	β_2
	0.1	20	5	40	0.1	20	5	40
Θ_1 50	0.047	10.349	2.633	20.646	0.047	10.447	2.656	20.857
MSE 50	0.011	408.805	27.087	1635.293	0.011	407.841	26.576	1632.175
viés 50	-0.106	-19.303	-4.733	-38.707	-0.106	-19.105	-4.688	-38.286
Θ_1 100	0.097	20.630	5.186	41.117	0.097	20.878	5.224	41.607
MSE 100	0.001	7.866	1.088	31.936	0.001	8.443	1.073	34.039
viés 100	-0.003	0.630	0.185	1.117	-0.003	0.878	0.224	1.607
Θ_1 500	0.100	20.392	5.196	40.750	0.100	20.638	5.243	41.247
MSE 500	0.001	8.372	1.142	35.200	0.001	8.828	1.150	36.962
viés 500	0.000	0.392	0.196	0.750	0.000	0.638	0.243	1.247
Θ_1 1000	0.100	20.566	5.199	41.118	0.100	20.823	5.262	41.647
MSE 1000	0.001	9.522	1.178	38.926	0.001	10.057	1.238	41.168
viés 1000	-0.000	0.566	0.200	1.118	-0.000	0.823	0.262	1.647

Tabela 4.3: Estimativa dos parâmetros de $h(x, \Theta_3)$.

Estimativas	mistura				classificada			
	\mathbf{p}	α	β_1	β_2	\mathbf{p}	α	β_1	β_2
	0.4	10	10	30	0.4	10	10	30
Θ_3 50	0.206	5.453	5.467	16.485	0.208	5.322	5.369	16.075
MSE 50	0.163	105.220	106.043	959.385	0.162	103.229	103.873	933.085
Vies 50	-0.389	-9.094	-9.066	-27.029	-0.384	-9.356	-9.262	-27.850
Θ_3 100	0.396	10.564	10.503	32.130	0.392	10.238	10.199	30.872
MSE 100	0.006	4.203	4.291	46.108	0.003	2.420	2.816	24.050
Vies 100	-0.004	0.564	0.503	2.130	-0.008	0.238	0.199	0.872
Θ_3 500	0.403	10.729	10.698	32.456	0.401	10.328	10.329	30.974
MSE 500	0.004	4.102	4.264	46.270	0.002	2.213	2.467	21.820
Vies 500	0.003	0.729	0.698	2.456	0.001	0.328	0.329	0.974
Θ_3 1000	0.402	10.713	10.626	32.328	0.402	10.310	10.291	30.969
MSE 1000	0.004	4.129	4.158	43.660	0.002	2.160	2.518	20.772
Vies 1000	0.002	0.713	0.626	2.328	0.002	0.310	0.291	0.969

Tabela 4.4: Estimativa dos parâmetros de $h(x, \Theta_4)$.

Estimativas	mistura				classificada			
	p	α	β_1	β_2	p	α	β_1	β_2
	0.4	10	5	40	0.4	10	5	40
Θ_4 50	0.199	5.330	2.699	21.419	0.199	5.219	2.650	20.986
MSE 50	0.163	102.764	25.900	1645.838	0.163	102.234	25.689	1639.422
Vies 50	-0.403	-9.340	-4.602	-37.162	-0.402	-9.561	-4.700	-38.027
Θ_4 100	0.389	10.637	5.271	42.387	0.390	10.385	5.171	41.400
MSE 100	0.002	2.377	0.717	41.533	0.002	1.662	0.551	28.262
Vies 100	-0.011	0.637	0.271	2.387	-0.010	0.385	0.171	1.400
Θ_4 500	0.398	10.459	5.228	41.829	0.399	10.283	5.154	41.128
MSE 500	0.003	2.543	0.706	45.006	0.003	2.248	0.634	39.487
Vies 500	-0.002	0.459	0.228	1.829	-0.001	0.283	0.154	1.128
Θ_4 1000	0.402	10.580	5.302	42.395	0.402	10.440	5.235	41.802
MSE 1000	0.002	2.746	0.799	47.896	0.002	2.445	0.721	42.221
Vies 1000	0.002	0.580	0.302	2.394	0.002	0.440	0.235	1.802

Tabela 4.5: Estimativa dos parâmetros de $h(x, \Theta_5)$.

Estimativas	mistura				classificada			
	p	α	β_1	β_2	p	α	β_1	β_2
	0.6	25	10	40	0.6	25	10	10
Θ_5 50	0.295	13.491	5.351	21.612	0.295	13.646	5.414	21.857
MSE 50	0.363	646.274	103.299	1655.171	0.362	649.009	103.830	1663.360
Vies 50	-0.611	-23.018	-9.297	-36.777	-0.611	-22.708	-9.171	-36.285
Θ_5 100	0.598	26.107	10.431	41.783	0.598	26.419	10.554	42.299
MSE 100	0.002	17.952	2.619	45.965	0.002	18.028	2.742	46.937
Vies 100	-0.002	1.107	0.431	1.783	-0.002	1.419	0.554	2.299
Θ_5 500	0.600	25.592	10.206	40.898	0.600	25.914	10.336	41.453
MSE 500	0.002	16.807	2.548	45.558	0.002	15.994	2.482	44.195
Vies 500	-0.000	0.592	0.205	0.898	-0.000	0.914	0.336	1.453
Θ_5 1000	0.602	25.807	10.288	41.397	0.602	26.101	10.411	41.904
MSE 1000	0.002	16.214	2.542	44.673	0.002	15.645	2.503	42.967
Vies 1000	0.002	0.807	0.288	1.397	0.002	1.101	0.411	1.904

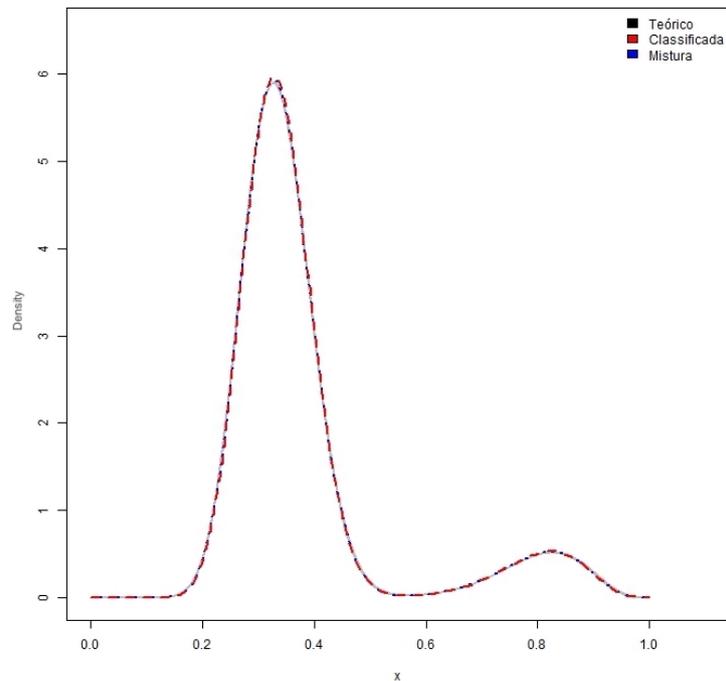


Figura 4.2: Comparação da densidade teórica, classificada e misturada para Θ_1

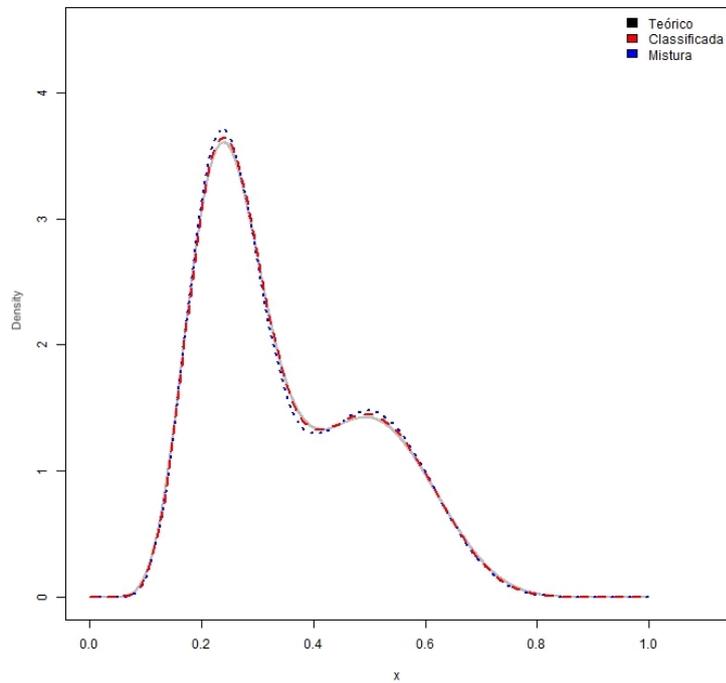


Figura 4.3: Comparação da densidade teórica, classificada e misturada para Θ_3

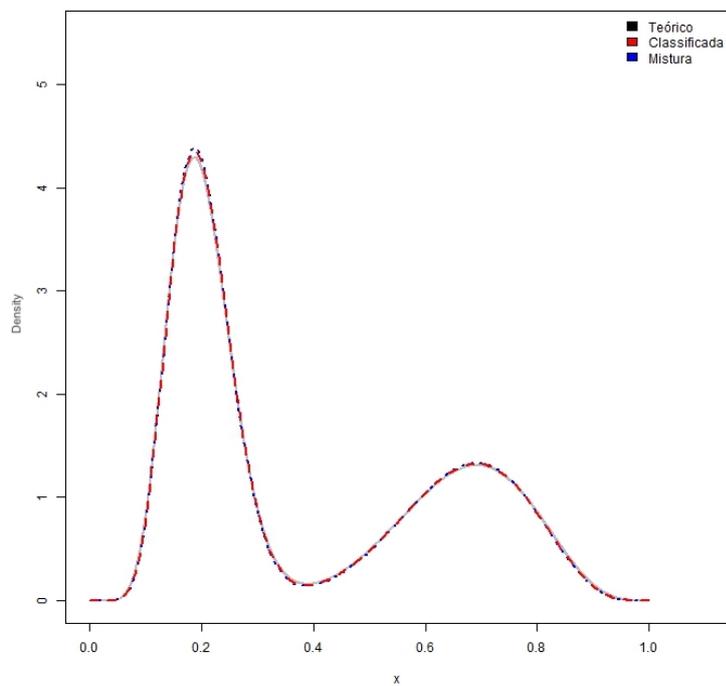


Figura 4.4: Comparação da densidade teórica, classificada e misturada para Θ_4

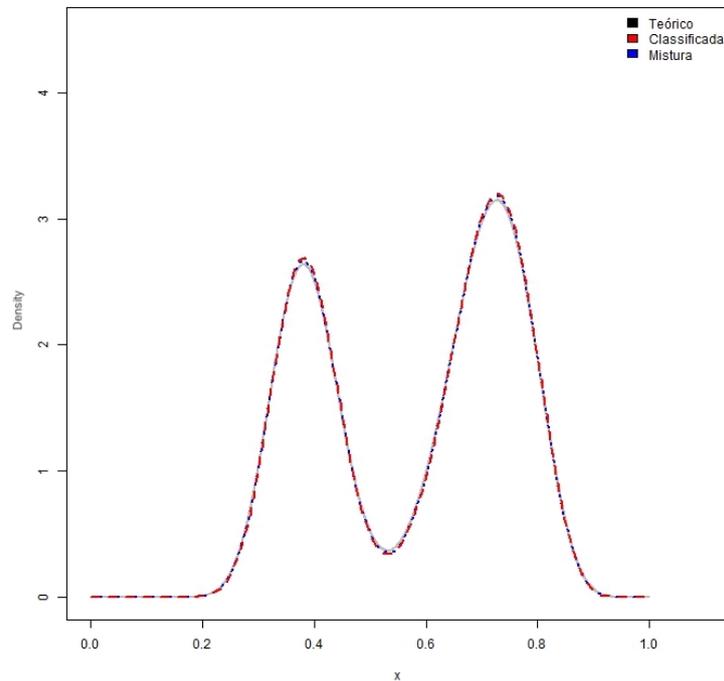


Figura 4.5: Comparação da densidade teórica, classificada e misturada para Θ_5

Os gráficos acima foram gerados a partir da amostra com $n=500$. Além da avaliação da estimativa, é necessário analisar o resultado da função discriminante, que será utilizada conforme Equações (3.23) e (3.24). As Tabelas (4.6) a (4.9) trazem os resultados do erro da função discriminante associados à classificação de dados. Nela foram estimados os erros em cada componente (e_{ik}), o erro total (ϵ_k) e o erro real (E_k).

Quando $n = 50$, o erro real é muito grande. A proporção de dados em cada componente apresenta diferenças significativas em relação à distribuição real, isso impacta diretamente na classificação dos dados, já que a quantidade de elementos classificados em cada componente é afetado diretamente pela proporção p .

Como há a informação da origem dos dados é possível apurar a qualidade da classificação do discriminante obtido em (3.23) e (3.24), sendo possível aferir a verdadeira dimensão do erro. Assim, o erro real (ϵ_k) para $n = 50$ é de aproximadamente 50%, comportamento que pôde ser percebido em todos os modelos com $n = 50$, conforme Tabelas (4.6) a (4.9).

Para as outras amostras percebem-se que os erros ϵ_c e ϵ_m variavam entre 1% e 6%, o que indica a boa adequação da função discriminante. A aproximação ao erro ótimo é um bom indicativo da qualidade das estimativas, além disso o erro real confirma que o discriminante é eficaz.

Tabela 4.6: Erros de classificação com Θ_1 .

Tipo Erro / Amostras	n=50	n=100	n=500	n=1000
Erro Ótimo 1ª Comp. (e_1)	0.0043	0.0043	0.0043	0.0043
Erro Ótimo 2ª Comp. (e_2)	0.0002	0.0002	0.0002	0.0002
Erro Mistura 1ª Comp. (\hat{e}_1)	0.0027	0.0053	0.0063	0.0057
Erro Mistura 2ª Comp. (\hat{e}_2)	0.0001	0.0003	0.0003	0.0003
Erro Classificada 1ª Comp. (\hat{e}_1)	0.0026	0.0050	0.0059	0.0053
Erro Classificada 2ª Comp. (\hat{e}_2)	0.0001	0.0002	0.0003	0.0003
Erro Total Ótimo (ϵ_o)	0.0006	0.0006	0.0006	0.0006
Erro Total Mistura (ϵ_m)	0.0004	0.0007	0.0009	0.0008
Erro Total Classificada (ϵ_c)	0.0004	0.0007	0.0008	0.0008
Erro Real Mistura (E_m)	0.5004	0.0004	0.0008	0.0008
Erro Real Classificada (E_c)	0.5004	0.0011	0.0015	0.0013

Tabela 4.7: Erros de classificação com Θ_3 .

Tipo Erro / Amostras	n=50	n=100	n=500	n=1000
Erro Ótimo 1ª Comp. (e_1)	0.12140	0.12140	0.12140	0.12140
Erro Ótimo 2ª Comp. (e_2)	0.04760	0.04760	0.04760	0.04760
Erro Mistura 1ª Comp. (\hat{e}_1)	0.05600	0.11680	0.11480	0.11440
Erro Mistura 2ª Comp. (\hat{e}_2)	0.02360	0.04680	0.04660	0.04590
Erro Classificada 1ª Comp. (\hat{e}_1)	0.05620	0.12230	0.12060	0.11960
Erro Classificada 2ª Comp. (\hat{e}_2)	0.02430	0.04570	0.04790	0.04750
Erro Total Ótimo (ϵ_o)	0.07710	0.07710	0.07710	0.07710
Erro Total Mistura (ϵ_m)	0.03620	0.07230	0.07230	0.07180
Erro Total Classificada (ϵ_c)	0.03730	0.07470	0.07620	0.07540
Erro Real Mistura (E_m)	0.54220	0.08420	0.08450	0.08250
Erro Real Classificada (E_c)	0.53890	0.07740	0.07810	0.07800

Tabela 4.8: Erros de classificação com Θ_4 .

Tipo Erro / Amostras	n=50	n=100	n=500	n=1000
Erro Ótimo 1ª Comp. (e_1)	0.01000	0.01000	0.01000	0.01000
Erro Ótimo 2ª Comp. (e_2)	0.00340	0.00340	0.00340	0.00340
Erro Mistura 1ª Comp. (\hat{e}_1)	0.00490	0.00950	0.01030	0.00980
Erro Mistura 2ª Comp. (\hat{e}_2)	0.00170	0.00320	0.00350	0.00340
Erro Classificada 1ª Comp. (\hat{e}_1)	0.00520	0.01010	0.01080	0.01020
Erro Classificada 2ª Comp. (\hat{e}_2)	0.00190	0.00340	0.00370	0.00360
Erro Total Ótimo (ϵ_o)	0.00610	0.00610	0.00610	0.00610
Erro Total Mistura (ϵ_m)	0.00290	0.00560	0.00620	0.00590
Erro Total Classificada (ϵ_c)	0.00320	0.00600	0.00650	0.00620
Erro Real Mistura (E_m)	0.50330	0.00680	0.00760	0.00650
Erro Real Classificada (E_c)	0.50290	0.00700	0.00700	0.00580

Tabela 4.9: Erros de classificação com Θ_5 .

Tipo Erro / Amostras	n=50	n=100	n=500	n=1000
Erro Ótimo 1ª Comp. (e_1)	0.01020	0.01020	0.01020	0.01020
Erro Ótimo 2ª Comp. (e_2)	0.01110	0.01110	0.01110	0.01110
Erro Mistura 1ª Comp. (\hat{e}_1)	0.00460	0.01030	0.01110	0.01050
Erro Mistura 2ª Comp. (\hat{e}_2)	0.00470	0.01140	0.01230	0.01160
Erro Classificada 1ª Comp. (\hat{e}_1)	0.00450	0.00990	0.01050	0.01000
Erro Classificada 2ª Comp. (\hat{e}_2)	0.00460	0.01080	0.01160	0.01100
Erro Total Ótimo (ϵ_o)	0.01060	0.01060	0.01060	0.01060
Erro Total Mistura (ϵ_m)	0.00460	0.01060	0.01140	0.01080
Erro Total Classificada (ϵ_c)	0.00450	0.01020	0.01080	0.01030
Erro Real Mistura (E_m)	0.50550	0.01080	0.01140	0.01080
Erro Real Classificada (E_c)	0.50510	0.01100	0.01040	0.01040

Análise das estimativas de $h(x, \Theta)$ para Θ_2

O modelo $\Theta_2 = (0.2, 50, 15, 40)$ apresentou diferentes estimativas para cada tipo de amostra. Enquanto a estimação para dados classificados mostrou-se eficaz e eficiente, a estimação para mistura forneceu parâmetros que se aproximavam de uma distribuição unimodal ($n=500$), conforme figura abaixo.

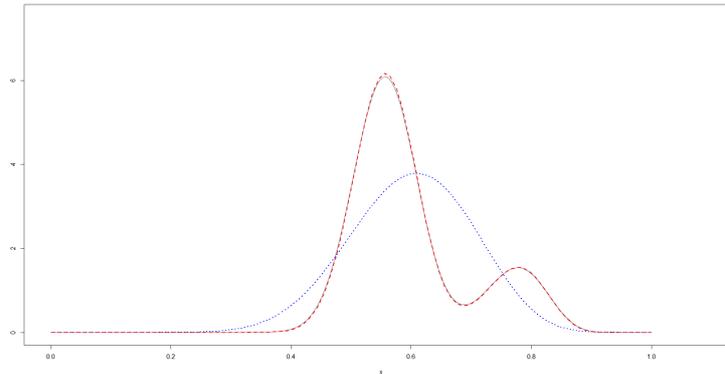


Figura 4.6: Comparação da densidade teórica, classificada e misturada para Θ_2 .

É possível notar que embora o algoritmo EM tenha perdido a precisão na estimativa, a densidade unimodal estimada busca a diminuição do erros ao envolver as duas componentes da mistura.

Tabela 4.10: Estimativa dos parâmetros de $h(x, \Theta_2)$.

Estimativas	mistura				classificada			
	p 0.2	α 50	β_1 15	β_2 40	p 0.2	α 50	β_1 15	β_2 40
estimativa 50	0.498	6.511	4.331	5.119	0.102	26.419	7.843	21.100
MSE 50	0.673	3870.878	267.500	2487.959	0.042	2554.967	229.571	1629.253
Vies 50	0.596	-86.977	-21.337	-69.761	-0.196	-47.162	-14.315	-37.800
estimativa 100	0.996	13.369	8.962	10.559	0.200	51.789	15.694	41.467
MSE 100	0.633	1345.966	39.255	869.528	0.002	69.382	6.732	44.562
Vies 100	0.795	-36.631	-6.038	-29.441	0.000	1.789	0.694	1.467
estimativa 500	0.996	13.183	8.805	10.357	0.199	51.414	15.389	41.078
MSE 500	0.634	1358.972	40.569	880.760	0.001	55.349	5.466	35.727
Vies 500	0.796	-36.817	-6.195	-29.643	-0.001	1.414	0.389	1.078
estimativa 1000	0.996	13.177	8.804	10.362	0.201	51.898	15.604	41.536
MSE 1000	0.633	1359.542	40.762	880.673	0.002	63.608	6.483	41.105
Vies 1000	0.796	-36.823	-6.196	-29.638	0.001	1.897	0.604	1.536

Em outros estudos é possível observar características semelhantes quando há duas distribuições muito próximas. Há a possibilidade do algoritmo interpretar a existência de uma única distribuição. O estudo de Lima, *et.al*, (2016) utiliza modelo de mistura para valores máximos de chuva em Seoul. Apesar de utilizar 4 modelos diferentes, em situações semelhantes, os modelos estimaram densidades unimodais.

Tabela 4.11: Erros de classificação com Θ_2 .

	n = 50	n = 100	n = 500	n = 1000
Erro Ótimo 1ª Comp. (ϵ_1)	0.04990	0.04990	0.04990	0.04990
Erro Ótimo 2ª Comp. (ϵ_2)	0.00750	0.00750	0.00750	0.00750
Erro Mistura 1ª Comp. ($\hat{\epsilon}_1$)	0.00000	0.00000	0.00000	0.00000
Erro Mistura 2ª Comp. ($\hat{\epsilon}_2$)	0.50000	1.00000	1.00000	1.00000
Erro Classificada 1ª Comp. ($\hat{\epsilon}_1$)	0.02330	0.05400	0.05260	0.05150
Erro Classificada 2ª Comp. ($\hat{\epsilon}_2$)	0.00350	0.00780	0.00760	0.00760
Erro Total Ótimo (ϵ_o)	0.01600	0.01600	0.01600	0.01600
Erro Total Mistura (ϵ_m)	0.00220	0.00450	0.00390	0.00420
Erro Total Classificada (ϵ_c)	0.00740	0.01660	0.01620	0.01610
Erro Real Mistura (E_m)	0.89810	0.79960	0.80130	0.79860
Erro Real Classificada (E_c)	0.51810	0.04230	0.04160	0.04170

Com relação a análise de discriminante, o erro real foi muito grande, em torno de 80%. Observe que o discriminante erra completamente o Componente 1 e completamente o Componente 2. Como cerca de 80% dos dados devem estar na primeira componente o erro real tende a ser de 80%.

Análise das estimativas de $h(x, \Theta)$ para $\Theta_6, \Theta_7, \Theta_8$ e Θ_9

Estes modelos, assim como anteriores, demonstram bom ajuste das estimativas, exceto quando se utilizam amostra $n = 50$, pois ficaram longe do alvo (parâmetro), mostrado pelo viés e pelo EQM, conforme Gráficos nos anexos (A.1) a (A.4). As estimativas dos parâmetros das amostras classificadas e das amostras misturadas obtiveram resultados semelhantes, conforme Tabelas (4.12) a (4.15)

As estimativas para $n = \{100, 500, 1000\}$ se mostraram acuradas e precisas. A diferença destes modelos para os anteriores é o alto erro quadrático médio obtido nas estimativas de β_2 (em negrito), conforme Tabela (4.12) a (4.15). As Figuras 4.7 a 4.10 mostram que há um desvio maior da densidade dos dados misturados em relação a densidade teórica, comparado aos primeiros modelos.

Tabela 4.12: Estimativa dos parâmetros de $h(x, \Theta_6)$.

Estimativas	mistura				classificada			
	\mathbf{p}	α	β_1	β_2	\mathbf{p}	α	β_1	β_2
	0.6	5	10	30	0.6	5	10	30
$\hat{\Theta}_6$ 50	0.299	3.059	5.979	19.545	0.294	2.761	5.459	16.716
MSE 50	0.367	27.675	108.206	1058.191	0.363	26.004	104.247	944.029
Vies 50	-0.603	-3.881	-8.042	-20.911	-0.611	-4.478	-9.083	-26.568
$\hat{\Theta}_6$ 100	0.594	5.739	11.273	35.470	0.598	5.248	10.550	31.514
MSE 100	0.010	1.966	5.918	97.727	0.002	0.471	2.346	22.399
Vies 100	-0.006	0.739	1.273	5.470	-0.002	0.248	0.550	1.514
$\hat{\Theta}_6$ 500	0.589	5.609	10.992	34.554	0.601	5.202	10.447	31.220
MSE 500	0.012	1.831	6.302	106.500	0.002	0.589	2.779	26.848
Vies 500	-0.011	0.609	0.992	4.554	0.001	0.203	0.447	1.220
$\hat{\Theta}_6$ 1000	0.584	5.718	11.135	35.428	0.599	5.193	10.433	31.317
MSE 1000	0.012	2.114	6.765	112.795	0.002	0.561	2.714	24.802
Vies 1000	-0.016	0.719	1.135	5.428	-0.001	0.193	0.433	1.317

Tabela 4.13: Estimativa dos parâmetros de $h(x, \Theta_7)$.

Estimativas	mistura				classificada			
	\mathbf{p} 0.7	α 2	β_1 2	β_2 20	\mathbf{p} 0.7	α 2	β_1 2	β_2 20
$\hat{\Theta}_7$ 50	0.340	1.090	1.033	10.964	0.349	1.015	1.001	10.303
MSE 50	0.494	4.209	4.157	442.554	0.492	4.083	4.101	412.305
Vies 50	-0.719	-1.819	-1.934	-18.073	-0.701	-1.969	-1.998	-19.395
$\hat{\Theta}_7$ 100	0.688	2.234	2.194	23.130	0.697	2.086	2.110	21.399
MSE 100	0.003	0.194	0.151	40.387	0.002	0.076	0.104	12.753
Vies 100	-0.012	0.234	0.194	3.130	-0.003	0.086	0.110	1.399
$\hat{\Theta}_7$ 500	0.689	2.198	2.122	22.665	0.699	2.064	2.052	20.924
MSE 500	0.004	0.197	0.133	41.452	0.002	0.083	0.096	15.469
Vies 500	-0.011	0.198	0.122	2.665	-0.001	0.064	0.052	0.924
$\hat{\Theta}_7$ 1000	0.691	2.235	2.150	23.414	0.701	2.068	2.056	21.061
MSE 1000	0.005	0.245	0.155	52.845	0.002	0.084	0.091	17.561
Vies 1000	-0.009	0.235	0.150	3.414	0.001	0.068	0.056	1.061

Tabela 4.14: Estimativa dos parâmetros de $h(x, \Theta_8)$.

Estimativas	mistura				classificada			
	\mathbf{p} 0.8	α 10	β_1 10	β_2 30	\mathbf{p} 0.8	α 10	β_1 10	β_2 30
$\hat{\Theta}_8$ 50	0.391	5.415	5.331	16.321	0.399	5.063	5.087	15.367
MSE 50	0.645	103.921	102.773	938.459	0.641	101.685	101.909	920.691
Vies 50	-0.819	-9.171	-9.337	-27.358	-0.802	-9.874	-9.825	-29.267
$\hat{\Theta}_8$ 100	0.787	10.656	10.490	32.628	0.797	10.012	9.999	30.576
MSE 100	0.005	3.626	2.599	45.129	0.002	1.534	1.359	25.743
Vies 100	-0.013	0.656	0.490	2.628	-0.003	0.012	-0.001	0.576
$\hat{\Theta}_8$ 500	0.789	10.925	10.749	33.205	0.802	10.244	10.244	31.009
MSE 500	0.004	5.291	4.160	58.174	0.002	2.122	2.199	24.716
Vies 500	-0.011	0.925	0.749	3.205	0.002	0.244	0.244	1.009
$\hat{\Theta}_8$ 1000	0.787	11.099	10.940	33.698	0.800	10.413	10.426	31.570
MSE 1000	0.004	5.334	4.207	64.037	0.002	2.342	2.404	29.890
Vies 1000	-0.012	1.099	0.940	3.698	0.000	0.413	0.426	1.570

Tabela 4.15: Estimativa dos parâmetros de $h(x, \Theta_9)$.

Estimativas	mistura				classificada			
	\mathbf{p} 0.9	α 10	β_1 5	β_2 40	\mathbf{p} 0.9	α 10	β_1 5	β_2 40
$\hat{\Theta}_9$ 50	0.452	5.414	2.715	21.933	0.452	5.292	2.659	21.321
MSE 50	0.811	103.715	25.904	1675.560	0.811	103.073	25.757	1662.819
Vies 50	-0.896	-9.171	-4.569	-36.134	-0.896	-9.417	-4.683	-37.358
$\hat{\Theta}_9$ 100	0.901	10.570	5.247	43.484	0.903	10.379	5.164	43.079
MSE 100	0.001	2.897	0.668	103.975	0.001	2.583	0.589	87.851
Vies 100	0.001	0.570	0.247	3.484	0.003	0.379	0.164	3.079
$\hat{\Theta}_9$ 500	0.900	10.454	5.212	42.128	0.901	10.231	5.115	41.343
MSE 500	0.001	2.333	0.553	64.466	0.001	2.022	0.496	55.082
Vies 500	-0.000	0.455	0.211	2.128	0.001	0.231	0.115	1.343
$\hat{\Theta}_9$ 1000	0.898	10.619	5.307	42.747	0.899	10.428	5.223	42.071
MSE 1000	0.001	3.028	0.688	72.363	0.001	2.557	0.606	64.440
Vies 1000	-0.002	0.619	0.307	2.747	-0.001	0.428	0.223	2.071

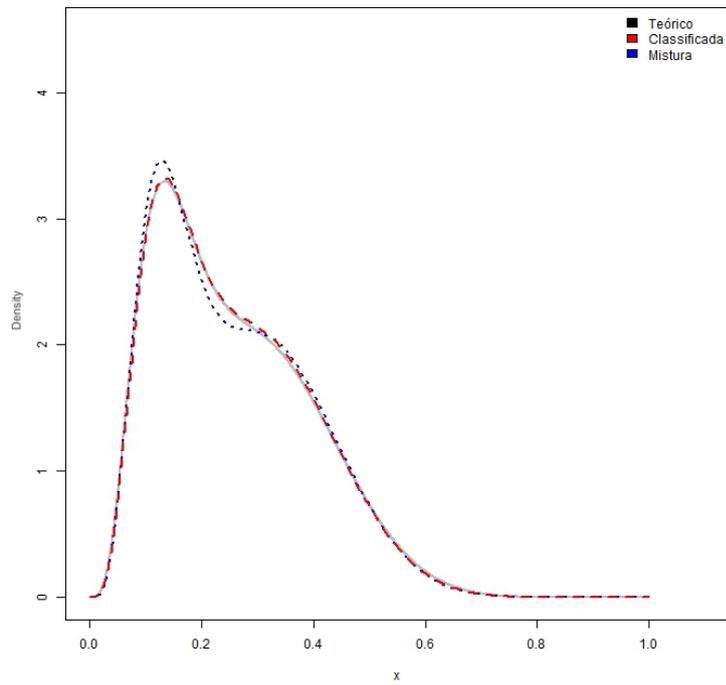


Figura 4.7: Comparação da densidade teórica, classificada e misturada para Θ_6

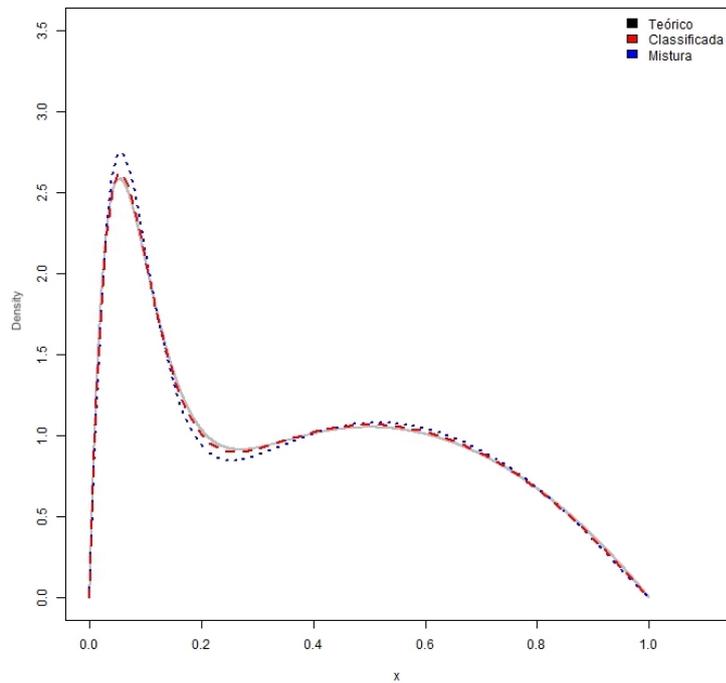
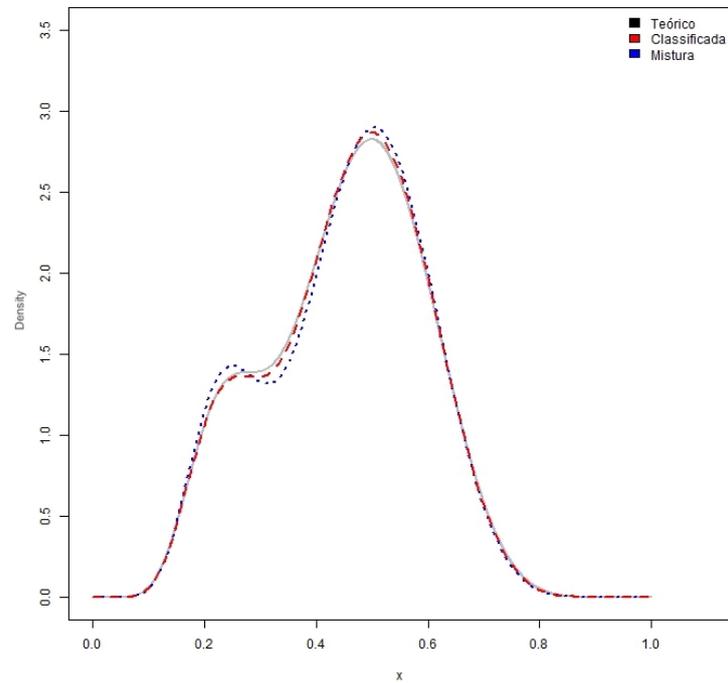
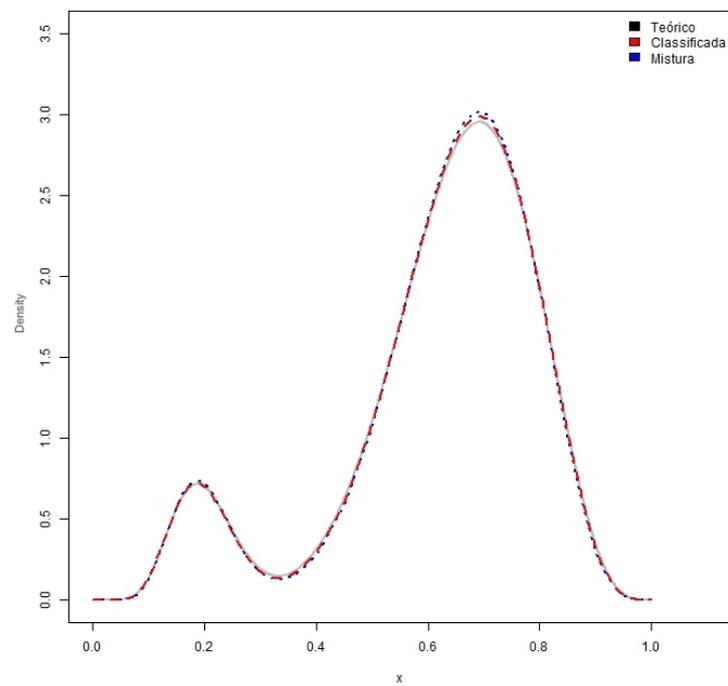


Figura 4.8: Comparação da densidade teórica, classificada e misturada para Θ_7

Figura 4.9: Comparação da densidade teórica, classificada e misturada para Θ_8 Figura 4.10: Comparação da densidade teórica, classificada e misturada para Θ_9

Da mesma forma que os gráficos anteriores, estes foram gerados a partir das estimativas com $n=500$ para efeito de comparação. Assim, Com relação a avaliação do discriminante, os erros ϵ_c e ϵ_m , variavam entre 7% e 14%, ou seja, percentuais maiores do que aqueles encontrados nos modelos 1, 3, 4 e 5. Este resultado corrobora com a análise de que estes modelos obtiveram resultados de estimação piores em relação aos 4 (quatro) primeiros.

O erro total da amostra misturada e classificada ficou próximo do erro total ótimo, indicando que a regra de classificação pode ser usada adequadamente.

Tabela 4.16: Erros de classificação com Θ_6 .

	n = 50	n = 100	n = 500	n = 1000
Erro Ótimo 1ª Comp. (e_1)	0.13560	0.13560	0.13560	0.13560
Erro Ótimo 2ª Comp. (e_2)	0.15300	0.15300	0.15300	0.15300
Erro Mistura 1ª Comp. (\hat{e}_1)	0.05290	0.11880	0.12710	0.12450
Erro Mistura 2ª Comp. (\hat{e}_2)	0.05890	0.14300	0.14480	0.13660
Erro Classificada 1ª Comp. (\hat{e}_1)	0.06350	0.13370	0.13340	0.13350
Erro Classificada 2ª Comp. (\hat{e}_2)	0.06750	0.15170	0.15620	0.15340
Erro Total Ótimo (ϵ_o)	0.14250	0.14250	0.14250	0.14250
Erro Total Mistura (ϵ_m)	0.05300	0.11930	0.12240	0.11800
Erro Total Classificada (ϵ_c)	0.06390	0.13880	0.14030	0.13920
Erro Real Mistura (E_m)	0.57010	0.16250	0.16540	0.16120
Erro Real Classificada (E_c)	0.56950	0.14600	0.14710	0.14640

Tabela 4.17: Erros de classificação com Θ_7 .

	n = 50	n = 100	n = 500	n = 1000
Erro Ótimo 1ª Comp. (e_1)	0.07870	0.07870	0.07870	0.07870
Erro Ótimo 2ª Comp. (e_2)	0.10150	0.10150	0.10150	0.10150
Erro Mistura 1ª Comp. (\hat{e}_1)	0.03710	0.06960	0.07150	0.06870
Erro Mistura 2ª Comp. (\hat{e}_2)	0.04840	0.09090	0.09390	0.09240
Erro Classificada 1ª Comp. (\hat{e}_1)	0.03860	0.07520	0.07650	0.07580
Erro Classificada 2ª Comp. (\hat{e}_2)	0.05050	0.09930	0.10190	0.10310
Erro Total Ótimo (ϵ_o)	0.08560	0.08560	0.08560	0.08560
Erro Total Mistura (ϵ_m)	0.03930	0.07440	0.07580	0.07310
Erro Total Classificada (ϵ_c)	0.04160	0.08140	0.08290	0.08250
Erro Real Mistura (E_m)	0.54960	0.08990	0.09250	0.09330
Erro Real Classificada (E_c)	0.54590	0.08780	0.09060	0.09190

Tabela 4.18: Erros de classificação com Θ_8 .

	n = 50	n = 100	n = 500	n = 1000
Erro Ótimo 1ª Comp. (e_1)	0.04140	0.04140	0.04140	0.04140
Erro Ótimo 2ª Comp. (e_2)	0.18260	0.18260	0.18260	0.18260
Erro Mistura 1ª Comp. (\hat{e}_1)	0.02110	0.04030	0.04030	0.03970
Erro Mistura 2ª Comp. (\hat{e}_2)	0.08220	0.17070	0.16750	0.16480
Erro Classificada 1ª Comp. (\hat{e}_1)	0.02050	0.04170	0.03980	0.03950
Erro Classificada 2ª Comp. (\hat{e}_2)	0.09270	0.18650	0.18720	0.18330
Erro Total Ótimo (ϵ_o)	0.06970	0.06970	0.06970	0.06970
Erro Total Mistura (ϵ_m)	0.03310	0.06380	0.06400	0.06340
Erro Total Classificada (ϵ_c)	0.03440	0.06990	0.06750	0.06670
Erro Real Mistura (E_m)	0.54010	0.08220	0.08010	0.08000
Erro Real Classificada (E_c)	0.54600	0.09260	0.09440	0.08900

Tabela 4.19: Erros de classificação com Θ_g .

	n = 50	n = 100	n = 500	n = 1000
Erro Ótimo 1ª Comp. (e_1)	0.00320	0.00320	0.00320	0.00320
Erro Ótimo 2ª Comp. (e_2)	0.02170	0.02170	0.02170	0.02170
Erro Mistura 1ª Comp. (\hat{e}_1)	0.00140	0.00320	0.00330	0.00320
Erro Mistura 2ª Comp. (\hat{e}_2)	0.01150	0.02290	0.02540	0.02390
Erro Classificada 1ª Comp. (\tilde{e}_1)	0.00150	0.00310	0.00340	0.00330
Erro Classificada 2ª Comp. (\tilde{e}_2)	0.01260	0.02290	0.02660	0.02480
Erro Total Ótimo (e_o)	0.00510	0.00510	0.00510	0.00510
Erro Total Mistura (ϵ_m)	0.00220	0.00510	0.00520	0.00510
Erro Total Classificada (ϵ_c)	0.00240	0.00490	0.00540	0.00530
Erro Real Mistura (E_m)	0.50330	0.00600	0.00670	0.00600
Erro Real Classificada (E_c)	0.50390	0.00780	0.00900	0.00820

Capítulo 5

Aplicações

5.1 Tempo entre erupções de geysers

No intuito de analisar a qualidade do ajuste do modelo de mistura finita com distribuição Beta utilizou-se os dados de 272 intervalos (em minutos) que o geysers Old Faithful, que fica em Yellowstone National Park, Wyoming (USA), demora para entrar em erupção. Esse conjunto de dados pode ser obtido através do software R, através do pacote MASS.

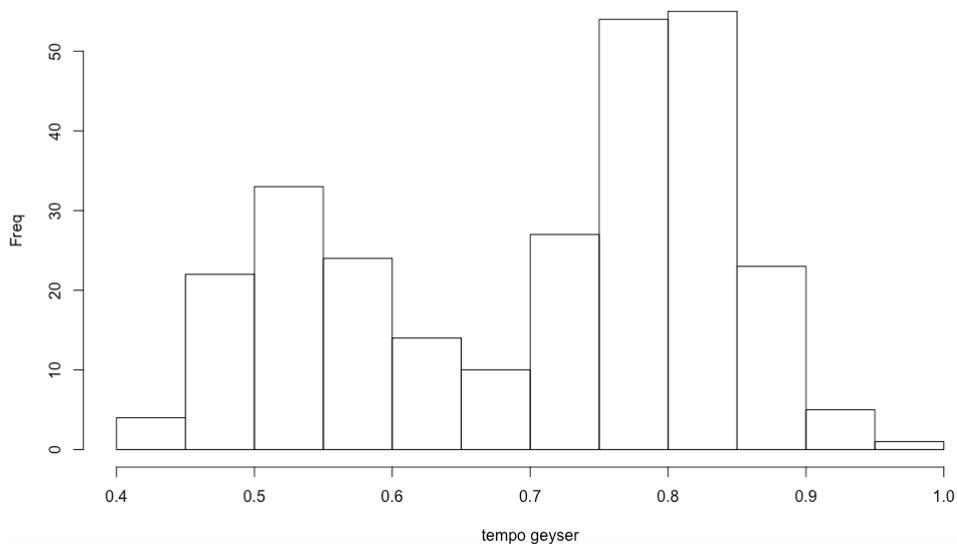


Figura 5.1: Histograma dos dados de erupção de geysers (min/100)

Os dados apresentam valor máximo de 96 minutos e mínimo de 43 minutos. Assim, não seria possível fazer a estimativa utilizando modelo Beta, já que o suporte da distribuição é no intervalo (0,1). Há, então, duas possibilidades: a primeira seria aplicar transformação dos dados através da seguinte fórmula

$$z_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}, \quad (5.1)$$

em que o vetor $x = (x_1, \dots, x_n)$ são os dados originais e z_i é o i -ésimo valor da transformação, de forma que $z_i \in (0,1)$. Como a estimação se utiliza de métodos numéricos envolvendo a função digamma, os valores máximos e mínimos de z_i acarretam problemas de singularidade. Assim, os valores 0 e 1 da distribuição z_i são substituídos pela segunda estatística de ordem e o máximo pela $(n - 1)$ -ésima estatística de ordem, isto é, o vetor transformado é dado por $z = (z_{(2)}, z_{(2)}, z_3, \dots, z_{(n-2)}, z_{(n-1)}, z_{(n-1)})$. Outra possibilidade é dividir os valores por 100 (cem) e tratar os dados como se estivessem em milésimos de segundo. Essa segunda opção, além de mais simples, apresentou resultados melhores para estimação.

A princípio não se conhece o fenômeno que divide os dados em dois grupos diferentes. O interesse neste caso é obter uma regra de classificação para cada um desses grupos. Através da função discriminante vamos estimar a quantidade de observações geradas do grupo 1 e do grupo 2 e calcular a probabilidade de erro, respectivamente.

A seguir, apresenta-se o histograma dos dados transformados e a curva ajustada através da mistura de duas Betas. Além disso, segue a tabela com os valores encontrados em suas estimativas.

Tabela 5.1: Estimação Geysler.

\hat{p}	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.35	33.14	27.77	8.29

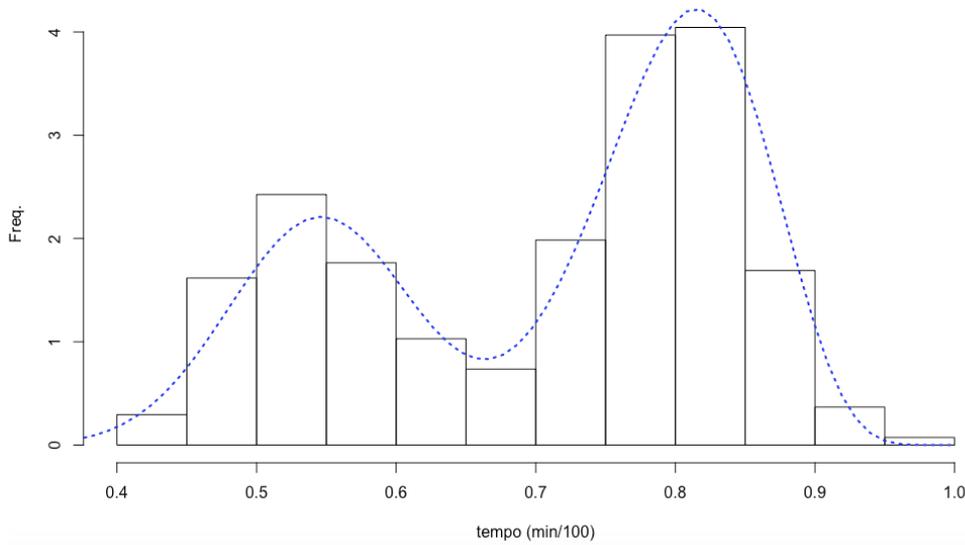


Figura 5.2: Histograma dos dados de erupção de geysler (min/100) e a distribuição Beta estimada.

A função de discriminante não linear é encontrada através da expressão (3.22), utilizando-se das estimativas obtidas conforme (3.8), (3.12), (3.13) e (3.14).

$$NL(x_j) = -21.05 - 19.47(1 - x_j). \quad (5.2)$$

Para cada observação x_j foi calculado o valor da função discriminante correspondente. Se $NL(x_j) \leq 0$, então classifica-se a observação em π_1 , caso contrário em π_2 . Assim, das 271 observações 99 foram classificadas na primeira componente e 173 na segunda componente.

Outra estatística importante é a probabilidade de erro de classificação e_1 e e_2 , definidos em (3.25) e (3.26). Assim, a probabilidade e_1 de se classificar no grupo 1, quando na verdade pertence ao grupo 2 é de 2,06%, enquanto a probabilidade e_2 de se classificar no grupo 2, quando na verdade pertence ao grupo 1, é de 3,10%. A probabilidade de se cometer qualquer erro $\hat{\epsilon}_m$, definido em (3.27), é de 2,74%.

No estudo de Levicoy (2016) estima-se uma mistura de Kumaraswamy destes mesmos dados e obtém-se probabilidade de erro $\hat{\epsilon}_m$ igual a 3,36%.

5.2 Vazão máxima anual das usinas do rio Paranapanema

O contexto atual da ciência traz preocupações recorrentes com as mudanças climáticas e as inundações são responsáveis por muitos prejuízos financeiros, além das perdas de vida que não podem ser avaliadas.

As cheias são a principal causa das inundações, motivo pelo qual é alvo de estudo nas últimas décadas. Os métodos estatísticos são importantes ferramentas para o estudo desse fenômeno, além da possibilidade de previsão, que poderia resultar no impacto socioeconômico das regiões estudadas.

Assim, para aplicação da técnica, foi escolhido o conjunto de dados da vazão máxima anual das usinas da bacia do rio Paranapanema entre os anos de 1931 e 2002. Os dados podem ser obtidos pelo Operador Nacional do Sistema (ONS). Esses dados foram escolhidos, pois apresentam mais de um período de cheia anual, isto é, tem-se máximos anuais que podem ocorrer tanto no inverno quanto no verão.

Inicialmente buscou-se verificar a distribuição dos dados através do histograma abaixo.

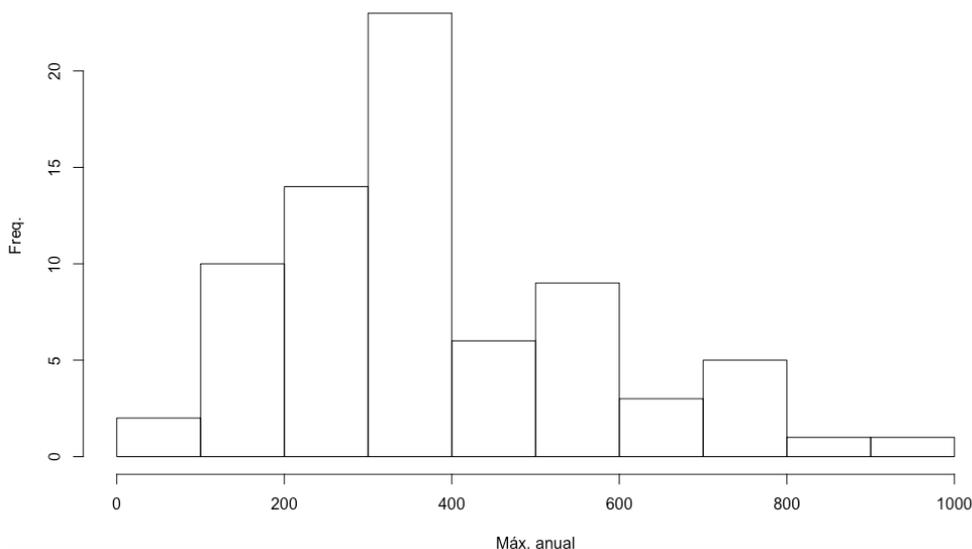


Figura 5.3: Histograma dos dados de vazão máxima anual das usinas da bacia do rio Paranapanema.

Observa-se que os valores do histograma variam entre 0 e 1000. Por isso, utilizou-se os valores da vazão divididos por 1000, o que fornece dados entre 0 e 1.

Dessa forma, foi Utilizado o algoritmo EM para estimar os parâmetros da mistura de duas Betas, conforme equações (3.8), (3.12), (3.13) e (3.14). O histograma com a curva estimada é apresentada abaixo, com o valor das estimativas abaixo.

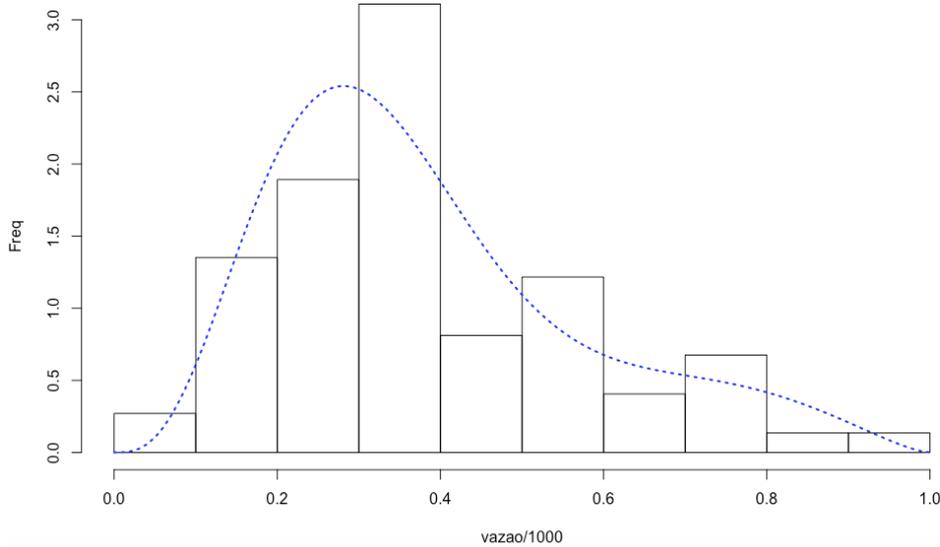


Figura 5.4: Histograma dos dados de vazão máxima anual das usinas da bacia do rio Paranapanema e a estimativa da mistura de duas Betas

\hat{p}	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.76	4.19	9.44	2.54

A função de discriminante não linear é encontrada através da Equação 3.22, utilizando-se das estimativas obtidas conforme (3.8), (3.12), (3.13) e (3.14).

$$NL(x_j) = -5.41 - 6.90(1 - x_j). \quad (5.3)$$

Para cada observação x_j foi calculado o valor da função discriminante correspondente. Se $NL(x_j) \leq 0$, então classifica-se a observação em π_1 , caso contrário em π_2 . Assim, das 74 observações 60 foram classificadas na primeira componente e 14 na segunda componente.

A probabilidade de erro de classificação e_1 e e_2 , definidos em (3.25) e (3.26). Assim, a probabilidade e_1 de se classificar no grupo 1, quando na verdade pertence ao grupo 2, é de 32.07%, enquanto a probabilidade e_2 de se classificar no grupo 2, quando na verdade pertence ao grupo 1 é de 3.58%. A probabilidade de se cometer qualquer erro $\hat{\epsilon}_m$, definido em (3.27) é de 25.17%.

Cruvinel (2017) fez o mesmo estudo para misturas de distribuição GEV. É possível comparar as estimativas obtidas com a distribuição Beta. Nela, nota-se que das 74 observações, 48 estão na primeira componente e 26 na segunda. Além disso, estima-se um erro global de 22,45% para o modelo de discriminante, isto é, houve um melhor ajustamento com a GEV.

5.3 Proeficiência no ENEM 2017

O Exame Nacional do Ensino Médio (ENEM) ganhou importância no contexto nacional ao conseguir unificar o vestibular entre as principais universidades do Brasil. O exame trouxe como grande novidade a utilização da Teoria de Resposta ao Item (TRI) na avaliação do aluno.

O exame tem como objetivo medir o conhecimento (proficiência) do participante em quatro grandes áreas: Matemática e suas Tecnologias; Linguagens, Códigos e suas Tecnologias; Ciências da Natureza e suas Tecnologias; e Ciências Humanas e suas Tecnologias.

Na avaliação do conhecimento, a unidade de medida se expressa por meio do conjunto de itens pertencentes a uma escala de proficiência; assim, os parâmetros dos itens são estabelecidos previamente. A proficiência é verificada a partir da análise do perfil das respostas do participante a esse conjunto de itens.

A nota do aluno estará associada a uma escala (régua) que pode assumir qualquer valor no conjunto dos números reais, como, por exemplo, 621, 476, ou outro valor qualquer. Outro aspecto importante no é que as notas mínima (nenhum acerto) e máxima (45 acertos) dependem do grau de dificuldade da prova aplicada. Em outras palavras, o mínimo e o máximo em cada prova dependem somente das questões que compõem a prova e não de quem as responde. Porém mesmo com mínimos e máximos diferentes, seus resultados são comparáveis.

Assim, analisou-se dados do ENEM 2017 na prova de Ciências Humanas e suas Tecnologias. Foram selecionados os 10.000 primeiros dados (há aleatoriedade no banco) e retirados as linhas que não possuíam informação da nota do candidato (NA) em qualquer uma das quatro provas. Por fim foram selecionados os dados dos alunos do DF, totalizando 172 unidades amostrais.

Como as notas não possuem nota mínima e máxima fixas, foi utilizada a nota relativa do aluno como proporção de acertos. Assim, utilizou-se a transformação a seguir

$$z_i = \frac{\text{nota.aluno}_i - \min(\text{prova}_j)}{\max(\text{prova}_j) - \min(\text{prova}_j)}$$

Como o ENEM utiliza o conceito de proficiência não é possível utilizar somente a proporção de acertos brutos, pois seria possível obter notas diferentes mesmo com proporções de acertos iguais.

Como não há classificação prévia dos dados, utilizá-se o algoritmo EM para estimação dos dados. Assim, obteve-se o histograma com curva suavizada dos dados (vermelho) e a densidade estimada (azul).

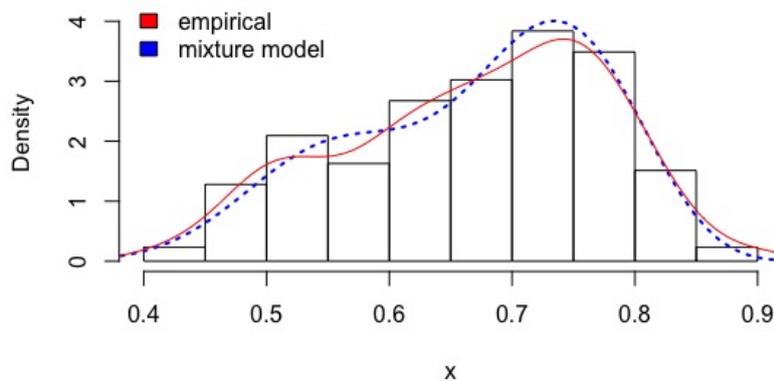


Figura 5.5: Histograma da proporção de acertos no ENEM 2017, a densidade empírica e densidade estimada.

Os parâmetros associados a densidade estimada são apresentados na Tabela a seguir:

\hat{p}	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.31	29.64	24.40	11.25

A função discriminante estimada é dada por $NL(x_j) = -12,58 - 13,15z$. Através desta função obteve-se que dos 172 dados utilizados para estimação, 51 foram classificadas na primeira componente e a 121 na segunda componente. Os erros associados a primeira componente corresponde a 6,55% e a segunda corresponde a 15,99%. O erro total (ϵ_m) associado ao modelo é de 13,11%.

Analisa-se também o QQ-plot para verificar a existência de pontos discrepantes, pois é possível que tais elementos influenciem na construção da função de discriminante. Observa-se na Figura abaixo o bom ajustamento dos dados.

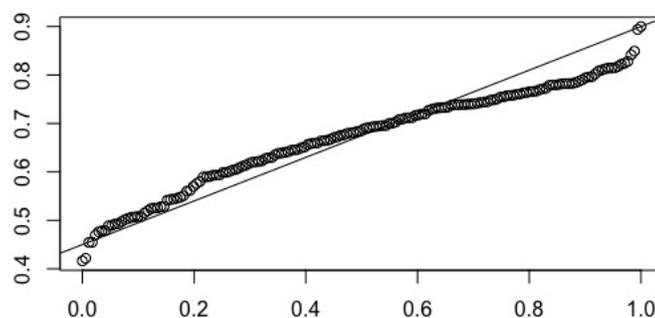


Figura 5.6: QQ-plot da distribuição estimada.

O resultado da estimação sugere a existência de dois grupos de candidatos com habilidade em conhecimentos humanos diferentes ou que a construção da avaliação tenha apresentado algum problema, pois entre o percentual 55% a 60% observa-se um quebra do padrão.

Este resultado é especialmente importante em virtude da prova do ENEM ser construída utilizando a Teoria de Resposta ao Item (TRI). Logo, o resultado acima indica grupos de alunos diferentes ou a necessidade de ajustar os itens da prova, com a finalidade de ter resultados com aspecto unimodal.

5.4 Conclusão

O trabalho apresentou o algoritmo EM e a EMV para estimação dos parâmetros de uma mistura de duas Betas e suas respectivas funções de discriminante. Analisou-se o comportamento da estimação e dos discriminantes em 9 (nove) modelos simulados entre dados classificados e misturados.

O resultado esperado era que as estimativas, a partir de dados classificados, fossem mais precisas do que as da misturada, pois há mais informação. De fato, o que pôde ser observado é que, em geral, as amostras classificados obtiveram melhor desempenho, pois as estimativas foram mais próximas dos valores ótimos, conforme viés e EQM. Isto é, os parâmetros, em geral, se aproximam mais do verdadeiro valor quando a amostra classificada é utilizada.

Tal fato é evidenciado pelos gráficos das estimativas das distribuições e pelos gráficos do EQM para cada parâmetro. A quantidade da amostra também influenciou a estimação. Sugere-se utilizar pelo menos dados com aproximadamente 100 elementos de forma a obter melhores resultados.

Na análise do desempenho da função discriminante foram observadas as probabilidades de se cometer erros de classificação para diferentes combinações de parâmetros. É possível perceber que os erros apresentaram variações mínimas a partir da amostra de tamanho $n=100$, conforme tabela (A.1) a (A.4) em anexo. Isto é, independe se foi usada amostra classificada ou misturada. O fato dos erros teóricos ficarem bem próximos do erros das componentes (erros teóricos) indica bom ajustamento do modelo. Isso implica que a estimação de modelos misturados, mais comuns de ocorrer, podem ser utilizados com eficácia semelhante dos dados classificados.

No que diz respeito a aplicação da metodologia em dados reais, observou-se eficácia na utilização de misturas de duas Betas para situações bimodais. A primeira aplicação, que trata do tempo médio de erupção de geysers, apresentou bom ajuste dos dados. A quantidade de elementos classificados em cada componente ficou bem próxima a de outro estudo semelhante. A segunda aplicação, que trata da vazão de água no rio Paranapanema, também apresentou bom ajuste de dados. Verificou-se maior diferença na quantidade de elementos classificados em cada componente, contudo erros totais se aproximaram comparado a estudo semelhante. A terceira aplicação demonstra a utilidade da metodologia na avaliação dos resultados da prova do ENEM. Observa-se da mesma forma que nas outras aplicações eficácia na estimação e erro estimado baixo.

Assim, a proposta inicial de modelar o comportamento dos dados por meio de uma mistura de distribuições Beta e utilizá-lo em casos reais mostrou-se bastante útil. A utilização do algoritmo EM simplificou a estimação dos parâmetros, já que as situações reais estão mais próximas deste caso.

Por fim, como sugestão para trabalhos futuro a possibilidade de estudar o erro padrão das estimativas via bootstrap.

Apêndice A

Resultados da Estimação

Erro Quadrático Médio

Nas Figuras A.1 a A.4, mostra-se o comportamento do EQM de p , α , β_1 e β_2 , respectivamente. Nota-se que quando o tamanho da amostra aumenta o EQM diminui para todos os casos.

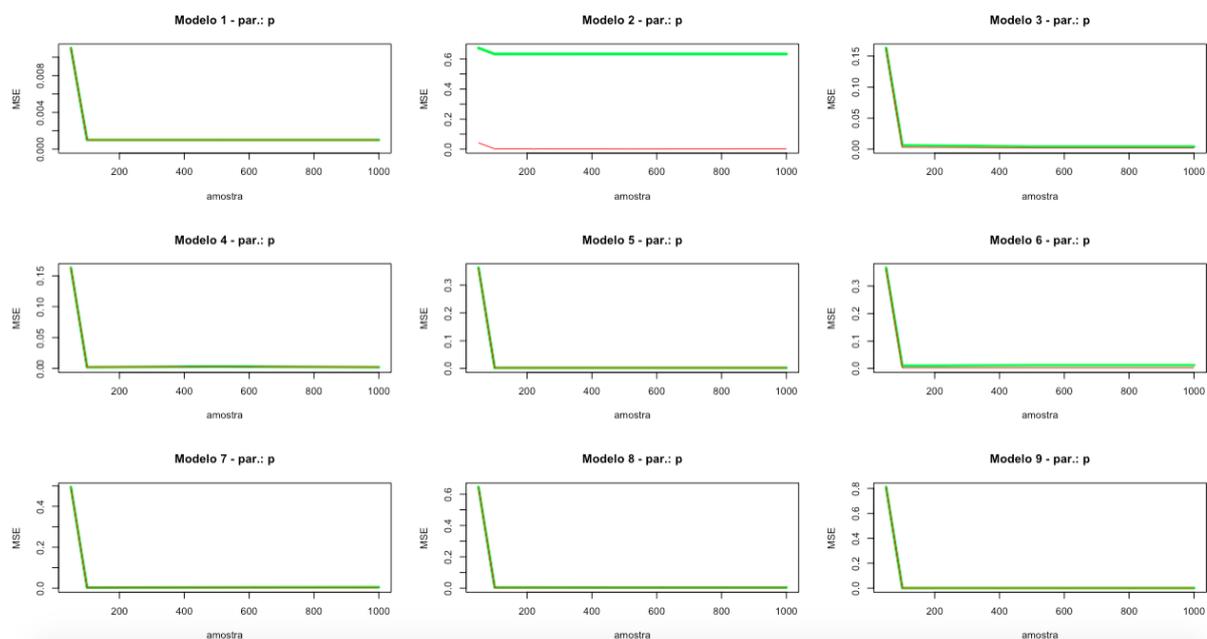


Figura A.1: Comparação dos erros quadráticos do parâmetro p

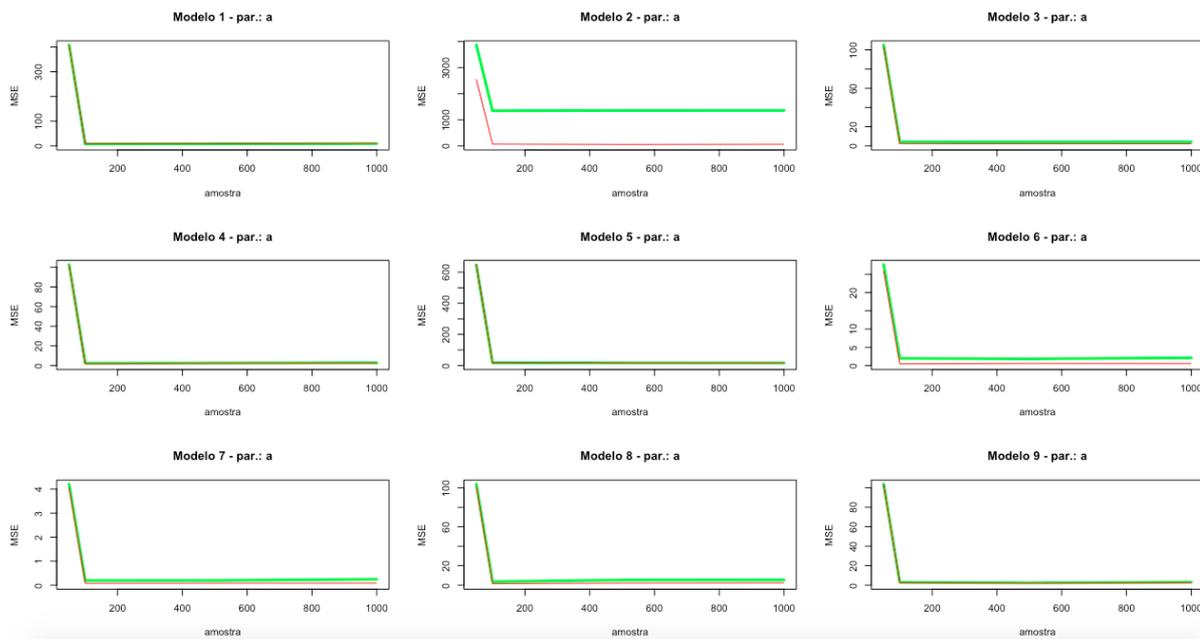


Figura A.2: Comparação dos erros quadráticos do parâmetro α

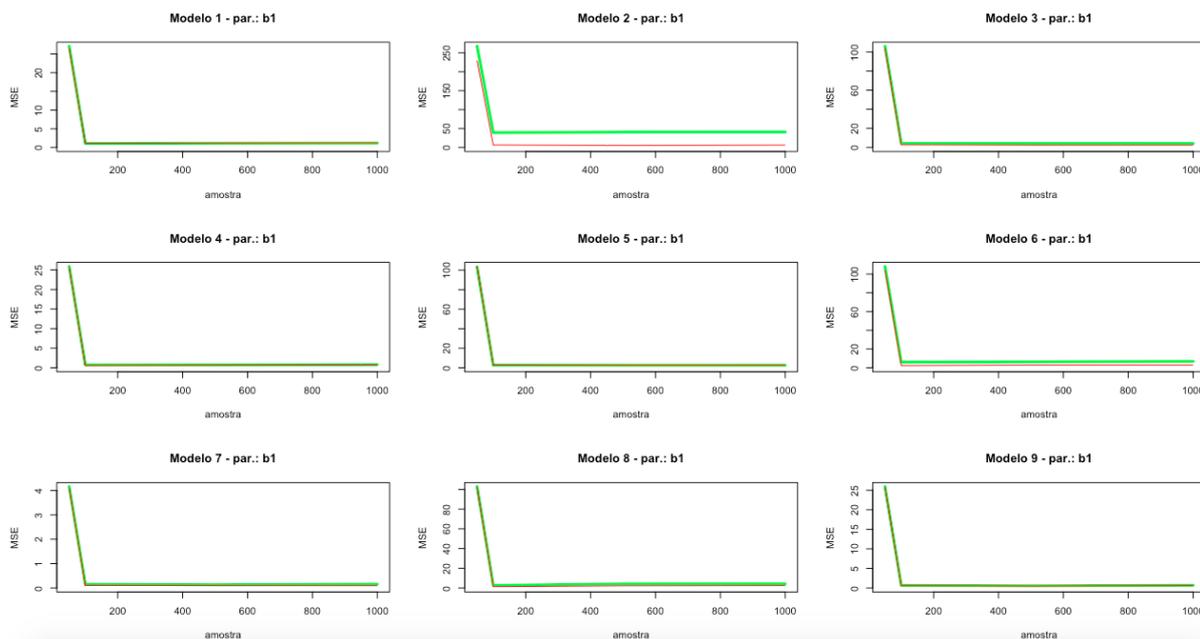
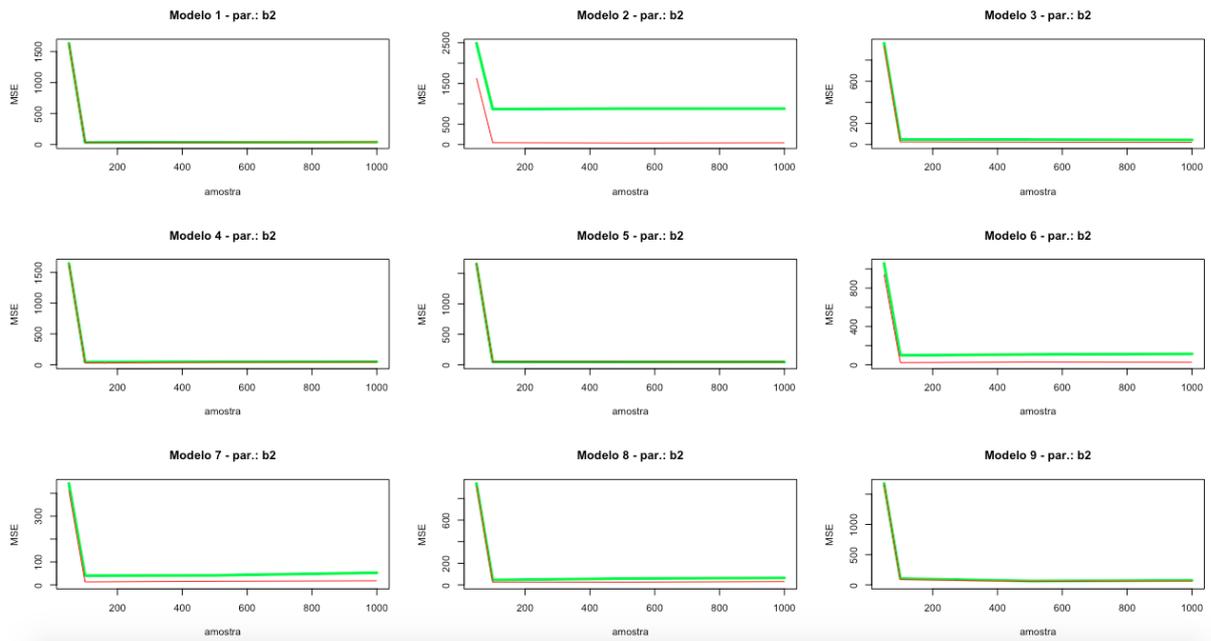


Figura A.3: Comparação dos erros quadráticos do parâmetro β_1

Figura A.4: Comparação dos erros quadráticos do parâmetro β_2

Apêndice B

Código R para mistura de normais

```
data(faithful)
attach(faithful)
##EM Algorithm
W = waiting
s = c(0.5, 40, 90, 16, 16)
em = function(W,s) {
  Ep = s[1]*dnorm(W, s[2], sqrt(s[4]))/(s[1]*dnorm(W, s[2], sqrt(s[4]))
    (1-s[1])*dnorm(W, s[3], sqrt(s[5])))
  s[1] = mean(Ep)
  s[2] = sum(Ep*W) / sum(Ep)
  s[3] = sum((1-Ep)*W) / sum(1-Ep)
  s[4] = sum(Ep*(W-s[2])^2) / sum(Ep)
  s[5] = sum((1-Ep)*(W-s[3])^2) / sum(1-Ep)
  s
}

iter = function(W, s) {
  s1 = em(W,s)
  for (i in 1:5) {
    if (abs(s[i]-s1[i]) > 0.0001) {
      s=s1
      iter(W,s)
    }
    else s1
  }
  s1
}

iter(W,s)
[1] 0.3507784 54.2179838 79.9088649 29.8611799 35.9824271

p = iter(W, s)
p1<-p[1]
```

```
p2<-p[2]
p3<-p[3]
p4<-p[4]
p5<-p[5]
Boot<-function(B) {
  r<-0
  k<-0
  bootmean1 <-rep(0, B)+ bootvar1<-rep(0, B)
  bootmean2<-rep(0, B)
  bootvar2<-rep(0, B)
  for(i in 1:B){
    p<-runif(1, 0, 1)
    if(p<p1){
      boot1<-rnorm(p1*272, p2, sqrt(p4))
      bootmean1[i]<-mean(boot1)
      bootvar1[i]<-var(boot1)
      r<-r+1
    }
    else{
      boot2<-rnorm((1-p1)*272, p3, sqrt(p5))
      bootmean2[i]<-mean(boot2)
      bootvar2[i]<-var(boot2)
      k<-k+1
    }
  }
  meanbootm1<-sum(bootmean1)/r
  meanbootvar1<-sum(bootvar1)/r
  meanbootm2<-sum(bootmean2)/k
  meanbootvar2<-sum(bootvar2)/k
  list(meanbootm1= meanbootm1, meanbootvar1= meanbootvar1,
  meanbootm2= meanbootm2, meanbootvar2= meanbootvar2 )
}
>
Boot(1000)
$meanbootm1
[1] 54.26174
$meanbootvar1
[1] 29.74411
$meanbootm2
[1] 79.88433
$meanbootvar2
[1] 36.02818
```

Fonte: procurar no evernote

Apêndice C

Código R para geração dos dados

```
#install.packages('rootSolve')
require(rootSolve)
#install.packages('VGAM')
require(VGAM)
#install.packages('xtable')
require(xtable)
par(mfrow=c(1,1))

#TODOS OS PARAMETROS A SEREM ESTIMADOS
# Todos os Parâmetros a serem estimados
for (jil in 1:1){
  p1 <- c(0.1, 20, 5, 40) #escolhido
  p2 <- c(0.2, 50, 15, 40) #escolhido
  p3 <- c(0.4, 10, 10, 30) #escolhido
  p4 <- c(0.4, 10, 5, 40) #escolhido
  p5 <- c(0.6, 25, 10, 40) #escolhido
  p6 <- c(0.6, 5, 10, 30) #escolhido
  p7 <- c(0.7, 2, 2, 20) #escolhido
  p8 <- c(0.8, 10, 10, 30) #escolhido
  p9 <- c(0.9, 10, 5, 40) #escolhido
  theta <- t(matrix(c(p1,p2,p3,p4,p5,p6,p7,p8,p9), 4, 9))
  #theta <- t(round(matrix(vetor, 4, 9), 2))
}

model <- 1 #Basta mudar o modelo aqui para rodar a programação

##### ESCOLHER OS VERDADEIROS VALORES DE CADA PARAMETRO #####
for (j in 1:1){
  par.p = theta[model,1]
  par.a = theta[model,2]
```

```

par.b1 = theta[model,3]
par.b2 = theta[model,4]
parametros <- c(par.p,par.a,par.b1,par.b2,NA)
}

#TITULO DO GR?FICO
for (j in 1:1){
titulo <- "NA"
param <- paste("(", parametros[1], ",", parametros[2], ",",
parametros[3], ",", parametros[4], ")")
nome <- c("p, alfa, beta1, beta2")
titulo <- paste("Mistura de Betas : ", nome, ") = ", param, sep="")
}

#OBTER O VALOR MÁXIMO DA FUNCAO
for (j in 1:1){
densidade <- par.p*dbeta(seq(0,1,0.0001),par.a,par.b1) +
(1-par.p)*dbeta(seq(0,1,0.0001),par.a,par.b2)
max_dens <- which.max(densidade)
max <- ifelse(is.infinite(round(densidade[max_dens]+0.5,0)),
6,round(densidade[max_dens]+0.5,0))
}

# PLOT DOS VALORES EM UM GRAFICO
par(mfrow=c(1,1))
plot(seq(0,1,0.0001), (par.p*dbeta(seq(0,1,0.0001),par.a,par.b1) +
(1-par.p)*dbeta(seq(0,1,0.0001),par.a,par.b2)), type="l"
, xlim=c(0,1.1), ylim=c(0,max), ylab="Densidade", xlab="x", main=titulo)

#Criando vetores vazios que ir?o armazenar as estimativas
for(j in 1:1){
mixed.p <- rep(0,100)
mixed.a <- rep(0,100)
mixed.b1 <- rep(0,100)
mixed.b2 <- rep(0,100)
mixed.e1 <- rep(0,100)
mixed.e2 <- rep(0,100)
mixed.acertos <- rep(0,100)
mixed.erros1 <- rep(0,100)
mixed.erros2 <- rep(0,100)
class.p <- rep(0,100)
}

```

```

class.a <- rep(0,100)
class.b1 <- rep(0,100)
class.b2 <- rep(0,100)
class.e1 <- rep(0,100)
class.e2 <- rep(0,100)
class.acertos <- rep(0,100)
class.erros1 <- rep(0,100)
class.erros2 <- rep(0,100)
}

#Criar uma tabela (antes de começar o FOR)
tabelaestimativas <- matrix(c(rep(c(50,100,500,1000),each=4),
rep(NA,128)),16,9, dimnames = list(c(rep(c("parametro",
"estimativa", "MSE", "Vies"),4)),
tabelaerros <- matrix(c(c(50,100,500,1000),rep(NA,68)),4,18,
dimnames = list(c("amostra n=50", "amostra n=100",
"amostra n=500", "amostra n=1000"), c("n", "erro.1.o", "erro.2.o",
"errol.m", "erro2.m", "errol.c", "erro2.c", "epsilon.o", "epsilon.m",
"MSE", "Vies", "epsilon.c", "MSE", "Vies", "erro.real.m", "erro.m",
"erro.real.c", "erro.c")))

#####
#####
##### COMEÇAR A RODAR O CODIGO A PARTIR DAQUI#####
#####
#####

#TABELA DE LISTS CRIADAS
for(j in 1:1){
tabelaest <- list()
tabelaerro <- list()
tabelaestlatex <- list()
tabelaerrolatex <- list()
}

# FOR CRIADO PARA FAZER O LOOPING DE TODAS AS TABELAS
for (quatro in 1:4){
amostra = c(50,100,500,1000)
n <- amostra[quatro]
linha = c(1,5,9,13) # para auxiliar na construção da
tabela ___tabelaestimativas___

# FOR PARA FAZER AS ESTIMATIVAS
for(y in 1:n){ #Numero de repeticoes para cada estima??o

```

```

#for(y in 1:amostra){
N = 100      #Tamanho da amostra gerada
U =runif(N) # gera??o de 100 uniformes(0,1)
mist = rep(NA,N)
mist.1 = rep(NA,N)
mist.2 = rep(NA,N)

#mist é o vetor de todas observações geradas e mist.1 e mist.2
#as observações geradas pela primeira e segunda componente.

for(i in 1:N){
  if(U[i]<par.p){
    mist[i] = rbeta(1,par.a,par.b1)
    mist.1[i] = mist[i]
  }else{
    mist[i] = rbeta(1,par.a,par.b2)
    mist.2[i] = mist[i]
  }
}

mist.11 <- as.numeric(na.omit(mist.1))
mist.22 <- as.numeric(na.omit(mist.2))

for (a in 1:1){
m=0
ak=0
b1k=0
b2k=0
pk<-0
juju<-0
}

#CHUTES INICIAIS PARA CADA UM DOS PARAMETROS

a<- ainic <- 15
b1<-b1inic <- 40
b2<-b2inic <- 10
p<-pinic <- .3

##### Algoritmo EM - Inicio

while(m==0){
  #beta = W_1j = p1g1/(p1g1+p2g2)

  beta <- ((p*dbeta(mist,a,b1))/(p*dbeta(mist,a,b1)+
(1-p)*dbeta(mist,a,b2)))

```

```

fb1 <- function(b1){
  deriv_b1 <- sum(beta*(digamma(a+b1)))-
  sum(beta*digamma(b1))+sum(beta*log(1-mist))
}

b1 <- multiroot(fb1,2)$root

fb2 <- function(b2){
  deriv_b2 <- sum((1-beta)*(digamma(a+b2)))-
  sum((1-beta)*digamma(b2))+sum((1-beta)*log(1-mist))
}

fa <- function(a){ #Poss?vel erro ln(x) seria ln(1-x)
  b1 <- rep(b1,100)
  b2 <- rep(b2,100)
  deriv_a <- sum(beta*(digamma(a+b1)))-
  sum(beta*digamma(a))+sum(beta*log(mist)) +
  sum((1-beta)*(digamma(a+b2)))-
  sum((1-beta)*digamma(a))+sum((1-beta)*log(mist))
}

# Como não há forma fechada, então podemos utilizar a função multroot
para obter os parâmetros
b1k <- multiroot(fb1,2,positive=TRUE)$root
b2k <- multiroot(fb2,2,positive=TRUE)$root
ak <- multiroot(fa,2,positive=TRUE)$root

pk <-sum(beta)/length(mist) #Conforme pag17 (2.16)

##### CRITERIO DE PARADA

if(abs(ak-a)<0.01 && abs(b1k-b1)<0.01 && abs(b2k-b2)<0.01 &&
abs(pk-p)<0.01){      m<-1      }

juju<-juju+1
b1 <- b1k
b2 <- b2k
a <- ak
p <- pk

}

##### Algoritmo EM - FIM

```

```

mixed.p[y] = 1-p
mixed.a[y] = a
mixed.b1[y] = b2
mixed.b2[y] = b1

##### Erros para amostra n?o classificada
numwj <- (p*gamma(a+b1)*gamma(b2))
denwj <- (((1-p)*gamma(a+b2)*gamma(b1)))
wj <- 1-(numwj/denwj)^(1/(b2-b1))

e1<-pbeta(wj,a,b2)
e2<-1-pbeta(wj,a,b1)

mixed.e1[y] = e1
mixed.e2[y] = e2
#####

#####erro otimo (verdadeiros valores dos parametros)
numwj.otimo <- (par.p*gamma(par.a+par.b1)*gamma(par.b2))
denwj.otimo <- (((1-par.p)*gamma(par.a+par.b2)*gamma(par.b1)))
wjotimo <- 1-(numwj.otimo/denwj.otimo)^(1/(par.b2-par.b1))

elotimo<-pbeta(wjotimo,par.a,par.b1)
e2otimo<-1-pbeta(wjotimo,par.a,par.b2)
otimo.erro <- par.p*elotimo+(1-par.p)*e2otimo

#####

for (fechar in 1:1){
#ca<-0
#cb1<-0
#cb2<-0

#cb1<-b1inic
#cb2<-b2inic
#ca<- ainic
#cp<-pinic
}

##### Estimacao para amostras classificadas (Inicio)

for (rep in 1:1) {

  cp <- length(mist.11)/length(mist)

  fpar <- function(x){

```

```

f_b1 =length(mist.11)*((digamma(x[3]+x[1]))-
digamma(x[1]))+sum(log(1-mist.11))
f_b2 = length(mist.22)*((digamma(x[3]+x[2]))-
digamma(x[2]))+sum(log(1-mist.22))
f_a = length(mist.11)*(digamma(x[3]+x[1]))-
digamma(x[3]))+sum(log(mist.11))+
length(mist.22)*(digamma(x[3]+x[2]))-
digamma(x[3]))+sum(log(mist.22))
c(f_b1=f_b1,f_b2=f_b2,b_a=f_a)
}

emv <- multiroot(f = fpar, start = c(1,1,1))

cb1 <- emv$root[1]
cb2 <- emv$root[2]
ca <- emv$root[3]

}

##### Estimacao para amostras classificadas (FIM)

class.p[y]=cp
class.a[y]=ca
class.b1[y]=cb1
class.b2[y]=cb2

#Erro para amostras classificadas
cnumwj <- (cp*gamma(ca+cb1)*gamma(cb2))
cdenwj <- (((1-cp)*gamma(ca+cb2)*gamma(cb1)))
cwj <- 1-(cnumwj/cdenwj)^(1/(cb2-cb1))

ce1<-pbeta(cwj,ca,cb1)
ce2<-1-pbeta(cwj,ca,cb2)

class.e1[y]=ce1
class.e2[y]=ce2

#####

#atuacao do discriminante

verdadeiros<- as.numeric(!is.na(mist.1))

#discriminante amostras misturadas
#a1 = log(((1-p)*gamma(a+b2)*gamma(b1))/(p*gamma(a+b1)*gamma(b2)))

```

```

al = log((p)*gamma(a+b1)*gamma(b2))/((1-p)*gamma(a+b2)*gamma(b1))
#b1 = b2 - b1
b1 = b1 - b2
z1 = log(1-mist)
nld.m<- al+b1*z1

```

```

classificador.m<-as.numeric()
for(i in 1:length(mist)){

```

```

  if(nld.m[i]<=0){
    classificador.m[i]<-1
  }
  else{
    classificador.m[i]<-0
  }
}

```

```

mixed.erros1[y] <- sum(as.numeric(verdadeiros==1 &
classificador.m==0))/length(mist)
mixed.erros2[y] <- sum(as.numeric(verdadeiros==0 &
classificador.m==1))/length(mist)

```

```

mixed.acertos1 <- sum(as.numeric(verdadeiros==1 &
classificador.m==1))/length(mist)
mixed.acertos2 <- sum(as.numeric(verdadeiros==0 &
classificador.m==0))/length(mist)

```

```

mixed.acertos[y] <- mixed.acertos1+mixed.acertos2

```

```

#discriminante amostras classificadas
al.c = log(((1-cp)*gamma(ca+cb2)*gamma(cb1))/
(p*gamma(ca+cb1)*gamma(cb2)))
bl.c = cb2 - cb1
z1.c = log(1-mist)
nld.c<- al.c + bl.c*z1.c

```

```

classificador.c<-as.numeric()
for(i in 1:length(mist)){

```

```

  if(nld.c[i]<=0){
    classificador.c[i]<-1
  }
  else{
    classificador.c[i]<-0
  }
}

```

```

class.erros1[y] <- sum(as.numeric(verdadeiros==1 &
classificador.c==0))/length(mist)
class.erros2[y] <- sum(as.numeric(verdadeiros==0 &
classificador.c==1))/length(mist)

class.acertos1 <- sum(as.numeric(verdadeiros==1 &
classificador.c==1))/length(mist)
class.acertos2 <- sum(as.numeric(verdadeiros==0 &
classificador.c==0))/length(mist)

#classificador.c<-as.numeric(verdadeiros==classificador.c)
#acertos2 <- sum(classificador.c)/length(mist)

class.acertos[y] <- class.acertos1+class.acertos2
}

#####FINAL DO LOOP#####
#construção das tabelas
for (cont in 1:1){
jj1<-mean(mixed.p)
jj2<-sum((mixed.p-par.p)^2)/n
vies1<-sum((mixed.p-par.p))/n

jj3<-mean(mixed.a)
jj4<-sum((mixed.a-par.a)^2)/n
vies2<-sum((mixed.a-par.a))/n

jj5<-mean(mixed.b1)
jj6<-sum((mixed.b1-par.b1)^2)/n
vies3<-sum((mixed.b1-par.b1))/n

jj7<-mean(mixed.b2)
jj8<-sum((mixed.b2-par.b2)^2)/n
vies4<-sum((mixed.b2-par.b2))/n

mixed.erro<-mixed.p*mixed.e1 + (1-mixed.p)*mixed.e2

jj9<-mean(mixed.erro) #teorico
jj10<-0
jj11<-mean(mixed.acertos) #pratico
jj12<-sum((mixed.erro-otimo.erro)^2)/n
## difere mais que os outros pq não faz parte do EM
vies5<-sum((mixed.erro-otimo.erro))/n

```

```

jj13<-mean(class.p)
jj14<-sum((class.p-par.p)^2)/n
vies6<-sum((class.p-par.p))/n

jj15<-mean(class.a)
jj16<-sum((class.a-par.a)^2)/n
vies7<-sum((class.a-par.a))/n

jj17<-mean(class.b1)
jj18<-sum((class.b1-par.b1)^2)/n
vies8<-sum((class.b1-par.b1))/n

jj19<-mean(class.b2)
jj20<-sum((class.b2-par.b2)^2)/n
vies9<-sum((class.b2-par.b2))/n

class.erro<-class.p*class.e1 + (1-class.p)*class.e2

jj21<-mean(class.erro) #teorico
jj22<-0
jj23<-mean(class.acertos) #pratico
jj24<-sum((class.erro-otimo.erro)^2)/n
## difere mais que os outros pq não faz parte do EM
vies10<-sum((class.erro-otimo.erro))/n

matrizinha <- matrix(c(par.p, jj1, jj2, vies1, par.a, jj3, jj4, vies2,
par.b1, jj5, jj6, vies3, par.b2, jj7, jj8, vies4, par.p, jj13, jj14, vies6,
par.a, jj15, jj16, vies7, par.b1, jj17, jj18, vies8, par.b2, jj19, jj20,
vies9), nrow=4, ncol=8, dimnames = list(c("parametro", "estimativa",
"MSE", "Vies"), c("p.mist", "a.mist",
b1.mist", "b2.mist", "p.class", "a.class", "b1.class", "b2.class")))

matriz.eros <- matrix(c(otimo.erro, jj9, jj12, vies5, jj21, jj24,
vies10), nrow=1, ncol=7, dimnames = list(c(""), c("epsilon.o",
"epsilon.m", "MSE", "Vies", "epsilon.c", "MSE", "Vies")))

otimo.erro; jj9; jj12; vies5; jj21; jj24; vies10

estimativas<- round(matrizinha, digits=4)
erros <- round(matriz.eros, digits=4)
e1.e2<-round(matrix(c(elotimo, e2otimo, mean(mixed.e1), mean(mixed.e2),
mean(class.e1), mean(class.e2)), nrow=1, dimnames = list(c(""),
c("erro.1.o", "erro.2.o", "erro1.m", "erro2.m", "erro1.c", "erro2.c"))),
digits=4)

```

```
pratico<-round(matrix(c(1-jj11, jj9, 1-jj23, jj21), nrow=1, dimnames =
list(c(""), c("erro.real.m", "erro.m", "erro.real.c", "erro.c"))),
digits=4)
```

```
m.pzin<-estimativas[1,1]
c.pzin<-estimativas[1,5]
e1.e2.pratico<-round(matrix(c(mean(mixed.erros1),
m.pzin*mean(mixed.e1), mean(mixed.erros2),
(1-m.pzin)*mean(mixed.e2), mean(class.erros1),
c.pzin*mean(class.e1) ,mean(class.erros2),
(1-c.pzin)*mean(class.e2)), nrow=1, dimnames =
list(c(""), c("erro.real.m1", "epslon1",
"erro.real.m2", "epslon2", "erro.real.c1", "epslon1",
"erro.real.c2", "epslon2"))), digits=4)
}
```

```
#TABELA COM AS ESTIMATIVAS PARA TODAS AS AMOSTRAS
line = linha[quatro]
tabelaestimativas[(line):(line+3),2:9] <- estimativas
tabelaerros[(quatro),2:7] <- e1.e2
tabelaerros[(quatro),8:14] <- erros
tabelaerros[(quatro),15:18] <- pratico
```

```
#TABELA COM AS ESTIMATIVAS GUARDADAS EM UMA LISTA
tabelaest[[quatro]] <- estimativas
tabelaerro[[quatro]] <- erros
tabelaestlatex[[quatro]] <- xtable(estimativas, digits=3)
tabelaerrolatex[[quatro]] <- xtable(erros, digits=3)
```

```
}
```

```
#rep(c("a", "b", "c"), each=2)
#c(rep(c("a", "b", "c"), 2))
```

```
#####
#FIM #####
#####
```

```
#####
#####
#####
# TABELAS PARA SEREM COLOCADOS NO LATEX
#####
#####
#####

tabelaestimativas
tabelaerros2 <- t(tabelaerros)
tabelaerros2 <- tabelaerros2[-c(1,10,11,13,14),]
tabelaerros2

nomelinhatabela <- row.names(tabelaestimativas)
reflinhatabela <- tabelaestimativas[,1]
row.names(tabelaestimativas) <- paste(nomelinhatabela,
reflinhatabela, sep=" ")
tabelaestimativas2 <- tabelaestimativas[,-1]

tabela_latex_est <- xtable(tabelaestimativas2, digits=3)
tabela_latex_err <- xtable(tabelaerros2, digits=5)

#tabelas para gerar o latex
tabela_latex_est;tabela_latex_err

#tabelas para serem colocadas nos resultados
tabelares <- matrix(rep(NA,12),3,4, dimnames =
list(c("parâmetro", "mistura", "classificada"),
c("p", "alfa", "beta1", "beta2")))
tabelares[1,] <- round(tabelaestimativas[9,2:5],1)
tabelares[2,] <- round(tabelaestimativas[10,2:5],5)
tabelares[3,] <- round(tabelaestimativas[10,6:9],5)

#tabela de resultados
tabela_latex_res <- xtable(tabelares, digits=5)

#tabela de erros
tabelares.e <- matrix(round(tabelaerros2[10:13,3],4),2,2,
dimnames = list(c("real", "teórico"),c("mistura", "classificada")))
tabela_latex_res.e <- xtable(tabelares.e, digits=4)

tabela_latex_res;tabela_latex_res.e
```

Referências Bibliográficas

- [1] Davison A. C. and Hall P. On the bias and variability of bootstrap and cross-validation estimates of error rate in discrimination problems. *Biometrika*, 1992.
- [2] Martínez A. M. and Kak A. C. Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.
- [3] Dempster A. P., N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, B(39), 1977.
- [4] Efron B. and Goin G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, 1983.
- [5] Schroder C. and Rahmann S. A hybrid parameter estimation algorithm for beta mixtures and applications to methylation state classification. *Algorithms for Molecular Biology*, 2017.
- [6] Robertson C. A. and Fryer J. G. Some descriptive properties of normal mixtures. *Journal Scandinavian Actuarial*, 1969.
- [7] Levicoy C. E. L. Estimação da função discriminante da mistura de duas distribuições kumaraswamy. *UnB*, 2016.
- [8] Lima C. H. R., Kwon H. H., and Kim J. Y. A bayesian beta distribution model for estimating rainfall idf curves in a changing climate. *Journal of Hydrology*, 2016.
- [9] Titterington D.M. and Makov U. E. *Statistical analysis of finite mixture distributions*. Wiley, 1985.
- [10] Cruvinel E. C. Discriminante para mistura de distribuições gev. *Universidade de Brasília*, 2017.
- [11] Casella G. and Berger R. L. *Inferência Estatística - Tradução da 2ª edição norte-americana*, volume 2. Centage Learning, 2010.
- [12] McLachlan G. J. and Peel D. *Finite mixture models*, volume 1. Wiley, 2000.
- [13] McLachlan G. J. and Basford K. E. *Mixture Models*. Marcel Dekker, 1988.
- [14] Teicher H. Identifiability of finite mixtures. the annals of mathematical statistics. *Metrika*, (1265–1269), 1963.

- [15] Arjun K. *Handbook of Beta Distribution and Its Applications.*, volume 1. CRC Press, 2004.
- [16] Ahmad K. E. and Abd-Elrahman A. M. Updating a nonlinear discriminant function estimated from a mixture of two weibull distributions. *Mathematical and Computer Modelling*, 1994.
- [17] Ahmad K. E., Jaheen Z. F., and A. A. Modhesh. Estimation of discriminant function based on small sample size from a mixture of two gumbel distributions. *Communications in Statistics*, 2010.
- [18] F. Liang, Mukherjee S., and M. West. *The Use of Unlabeled Data in Predictive Modeling*. Statistical Science, 2007.
- [19] Pohar M., Blas M., and Turk S. Comparison of logistic regression and linear discriminant analysis: A simulation study. *Metodološki zvezki*, 2004.
- [20] Atienza N., Garcia-Heras J., and Munoz-Pichardo J. M. A new condition for identifiability of finite mixture distributions. *Metrika*, (63(2)), 2006.
- [21] Dean N., Murphy T. B., and G. Downey. Updating classification rules using unlabelled data with applications in food authenticity studies. *Journal of the Royal Statistical Society*, 2006.
- [22] Johnson N. L., Kotz S., and Balakrishnan N. *Continuous Univariate Distributions*, volume 2. Wiley, 1995.
- [23] Chapelle O., Scholkopf B., and A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.
- [24] Ganeshanandam S. and Krzanowski W. J. Error-rate estimation in two-group discriminant analysis using the linear discriminant function. *Journal of Statistical Computation and Simulation*, 1990.
- [25] Mingoti S. A. *Análise de dados através de métodos de estatística multivariada*, volume 1. Editora UFMG, 2005.
- [26] Heisterkamp S.H and J.L.A. Pennings. The use of a finite mixture of beta distributions in the analysis of microarray data. *Kwantitatieve Methoden*, 2004.
- [27] Carvalho T.M. Modelagem da lgd via mistura de distribuições kumaraswam. *UnB*, 2015.
- [28] Ji Y., Wu C., Liu P., J. Wang, and Coomers K. R. Applications of beta-mixture models in bioinformatics. *Bioinformatics*, 21(9), 2005.