



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

**Discriminante para mistura de distribuições GEV**

por

Evelyn de Castro Cruvinel

Orientadora: Prof<sup>ª</sup>. Dra. Cira Etheowalda Guevara Otiniano

10 de Abril de 2017

**Evelyn de Castro Cruvinel**

## **Discriminante para mistura de distribuições GEV**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

**Universidade de Brasília**

**Brasília, 10 de Abril de 2017**

*À minha querida avó Geralda,  
exemplo de força e determinação.*

# Agradecimentos

Agradeço aos meus pais, Fátima e Cruvinel, por todo amor e dedicação, por terem se esforçado para me proporcionar uma boa educação. Devo tudo a vocês.

A minha irmã, Bruna, por estar comigo em todos os momentos, bons e ruins, na alegria e no mau humor. Eu sei que o nosso amor de irmã resiste a tudo e você jamais vai soltar da minha mão.

Ao Felipe por estar ao meu lado por todo esse percurso e me ajudar a enfrentar cada obstáculo com muito amor e carinho. Você tornou tudo mais fácil.

A minha orientadora Cira pela oportunidade, pela dedicação, pela paciência e por tudo que aprendi. Tenho a você toda a minha gratidão e carinho.

Ao meu gerente, Romanatto, pela oportunidade, pelo apoio, pelo incentivo e por todos os conselhos. Também quero agradecer as minhas amigas, Clécia e Sueide, por todo apoio e compreensão. Trabalhar ao lado de pessoas tão companheiras e generosas, é um grande privilégio.

Enfim, a todos que de algum modo me ajudaram para que esse objetivo fosse alcançado.

# Sumário

Agradecimentos

Lista de Figuras 2

Resumo 3

Abstract 4

Introdução 5

**1 Conceitos preliminares 8**

1.1 Distribuição de valor extremo generalizada . . . . . 8

1.2 Mistura de distribuições . . . . . 12

1.3 Identificabilidade . . . . . 15

1.4 Análise Discriminante . . . . . 20

1.4.1 Função discriminante para mistura de distribuições . . . . . 20

1.4.2 Erros de classificação . . . . . 22

**2 Estudo do discriminante de mistura de GEV 23**

2.1 Estimação por máxima verossimilhança . . . . . 23

2.1.1 Estimação por máxima verossimilhança para amostra não clas-  
sificada . . . . . 24

2.1.2 Estimação por máxima verossimilhança para amostra classifi-  
cada . . . . . 32

2.2 Estudo do discriminante para mistura de GEV . . . . . 37

2.2.1 Função discriminante mistura de GEV . . . . . 37

2.3	Erros de Classificação . . . . .	47
2.3.1	Forma analítica: Caso 1 ( $\gamma_1 = 0, \gamma_2 = 0$ )(ver Ahmad <i>et al.</i> (2010)). . . . .	47
2.3.2	Forma empírica . . . . .	49
<b>3</b>	<b>Simulações e Aplicação</b>	<b>50</b>
3.1	Simulações . . . . .	50
3.2	Aplicações . . . . .	63
3.2.1	Dados da vazão máxima anual das usinas da bacia do rio Pa- ranapanema . . . . .	63
3.2.2	Dados históricos do índice S&P 500 . . . . .	67
<b>4</b>	<b>Conclusões</b>	<b>70</b>
	<b>Referências Bibliográficas</b>	<b>73</b>
	<b>Apêndice</b>	<b>75</b>

# Lista de Figuras

1.1	Função de densidade de probabilidade das distribuições de valores extremos padrão (com $\alpha = 1$ ). . . . .	11
1.2	Função de densidade da GEV para $\gamma$ maior, menor e igual a 0. . . . .	11
1.3	Curvas de densidade da mistura de funções GEV: <b>(a)</b> , <b>(b)</b> , <b>(c)</b> , <b>(d)</b> , <b>(e)</b> , <b>(f)</b> . . . . .	14
3.1	Gráfico das densidades $h(x; \Theta)$ (curva contínua, $h(x; \hat{\Theta})$ (curva tracejada e em negrito) e $h(x; \tilde{\Theta})$ (curva pontilhada) para $\Theta_{1,1}$ e $\Theta_{1,2}$ . . . . .	57
3.2	Gráfico das densidades $h(x; \Theta)$ (curva contínua), $h(x; \hat{\Theta})$ (curva tracejada e em negrito) e $h(x; \tilde{\Theta})$ (curva pontilhada) para $\Theta_{2,1}$ e $\Theta_{2,2}$ . . . . .	57
3.3	Gráfico das densidades $h(x; \Theta)$ (curva contínua), $h(x; \hat{\Theta})$ (curva tracejada e em negrito) e $h(x; \tilde{\Theta})$ (curva pontilhada) para $\Theta_{3,1}$ e $\Theta_{3,2}$ . . . . .	58
3.4	Gráfico das densidades $h(x; \Theta)$ (curva contínua), $h(x; \hat{\Theta})$ (curva tracejada e em negrito) e $h(x; \tilde{\Theta})$ (curva pontilhada) para $\Theta_{4,1}$ e $\Theta_{4,2}$ . . . . .	58
3.5	Gráfico das densidades $h(x; \Theta)$ (curva contínua) , $h(x; \hat{\Theta})$ (curva tracejada e em negrito) e $h(x; \tilde{\Theta})$ (curva pontilhada) para $\Theta_{5,1}$ e $\Theta_{5,2}$ . . . . .	59
3.6	Gráfico das densidades $h(x; \Theta)$ (curva contínua) , $h(x; \hat{\Theta})$ (curva tracejada e em negrito) e $h(x; \tilde{\Theta})$ (curva pontilhada) para $\Theta_{6,1}$ e $\Theta_{6,2}$ . . . . .	59
3.7	Histograma dos dados vazão máxima anual das usinas da bacia do rio Paranapanema e as curvas de distribuição GEV em cada região modal. . . . .	64
3.8	Histograma dos dados e curva estimada . . . . .	65
3.9	(a)QQ-plot e (b) Função de distribuição acumulada da mistura de GEV estimada. . . . .	65
3.10	Gráfico período de retorno estimado e pela vazão. . . . .	67

3.11	Histograma dos maiores valores mensais do índice S&P 500 e esboços das curvas de distribuição GEV em cada região modal. . . . .	68
3.12	Histograma dos maiores valores mensais do índice S&P 500 e sua curva estimada . . . . .	69
3.13	(a) QQ-plot e (b) Função de distribuição acumulada da mistura de GEV estimada. . . . .	69
1	Gráficos <i>boxplot</i> das estimativas de $\Theta_{1,1}$ para amostra não classificada e amostra classificada . . . . .	77
2	Gráficos <i>boxplot</i> das estimativas de $\Theta_{1,2}$ para amostra não classificada e amostra classificada . . . . .	78
3	Gráficos <i>boxplot</i> das estimativas de $\Theta_{2,1}$ para amostra não classificada e amostra classificada . . . . .	79
4	Gráficos <i>boxplot</i> das estimativas de $\Theta_{2,2}$ para amostra não classificada e amostra classificada . . . . .	80
5	Gráficos <i>boxplot</i> das estimativas de $\Theta_{3,1}$ para amostra não classificada e amostra classificada . . . . .	81
6	Gráficos <i>boxplot</i> das estimativas de $\Theta_{3,2}$ para amostra não classificada e amostra classificada . . . . .	82
7	Gráficos <i>boxplot</i> das estimativas de $\Theta_{4,1}$ para amostra não classificada e amostra classificada . . . . .	83
8	Gráficos <i>boxplot</i> das estimativas de $\Theta_{4,2}$ para amostra não classificada e amostra classificada . . . . .	84
9	Gráficos <i>boxplot</i> das estimativas de $\Theta_{5,1}$ para amostra não classificada e amostra classificada . . . . .	85
10	Gráficos <i>boxplot</i> das estimativas de $\Theta_{5,2}$ para amostra não classificada e amostra classificada . . . . .	86
11	Gráficos <i>boxplot</i> das estimativas de $\Theta_{6,1}$ para amostra não classificada e amostra classificada . . . . .	87
12	Gráficos <i>boxplot</i> das estimativas de $\Theta_{6,2}$ para amostra não classificada e amostra classificada . . . . .	88

# Resumo

Este trabalho apresenta o estudo de um discriminante não linear da mistura de duas distribuições de valor extremo generalizada, conhecidas como GEV, com o parâmetro de escala comum. Como o parâmetro de forma da GEV pode assumir valor positivo, negativo ou nulo foram considerados seis casos possíveis para mistura de duas distribuições GEV. Para uma amostra classificada e não classificada de uma população com distribuição mistura de duas GEV são obtidas as expressões a serem resolvidas para obter os estimadores de máxima verossimilhança dos parâmetros do modelo, bem como sua função de discriminante. A avaliação do modelo proposto é feita por meio de simulações Monte Carlo utilizando amostras de tamanho  $n = 50$  e  $n = 100$  para dois conjuntos de parâmetros para cada caso possível de mistura de duas distribuições GEV. São apresentadas duas aplicações em dados reais que ilustram a eficiência da análise discriminante do modelo de mistura de duas distribuições GEV em situações bimodais.

**Palavras-chave:** Mistura; GEV; Discriminante; Amostra classificada; Amostra não-classificada.

# Abstract

This work presents the study of a nonlinear discriminant of the mixture of two generalized extreme value distributions, known as GEV, with the common scale parameter. As the GEV shape parameter can assume positive, negative or zero value, six possible cases were considered for mixing two GEV distributions. For a classified and unclassified sample of a population with mixed distribution of two GEV, the expressions to be solved are obtained to obtain the maximum likelihood estimators of the parameters of the model, as well as its discriminant function. The evaluation of the proposed model is done using Monte Carlo simulations using samples of size  $n = 50$  and  $n = 100$  for two sets of parameters for each possible case of mixing of two GEV distributions. Two applications are presented in real data that illustrates the efficiency of the discriminant analysis of the mixing model of two GEV distributions in bimodal situations.

**Keywords:**Mixture; GEV; Discriminant; Classified sample; Unclassified sample.

# Introdução

Mistura finita de densidades é uma abordagem matemática de modelos estatísticos com ampla aplicabilidade por ser tratar de uma ferramenta extremamente flexível. Tem sido utilizada principalmente onde há heterogenidade populacional, para modelar dados constituídos por uma densidade assimétrica e para aproximação de densidades complexas, como as de aspectos multimodais. Esses modelos são aplicados em várias áreas da estatística, por exemplo, análise de agrupamento, análise discriminante, análise de sobrevivência, métodos não-paramétricos ou semi-paramétricos e em processamento de imagens, e em outras áreas tais como astronomia, biologia, genética, medicina, psiquiatria, economia e engenharia ( ver McLachlan e Peel (2000)).

A mistura de  $k$  densidades (populações) é a combinação convexa de  $k$  densidades de certa família, sendo cada densidade denominada de componente da mistura e os coeficientes chamados de pesos da mistura. O vetor de parâmetros do modelo de mistura tem no mínimo  $2k - 1$  parâmetros a serem estimados, em que  $k - 1$  são referentes aos pesos e os outros  $k$  são os componentes da mistura. Esses parâmetros podem ser estimados, por exemplo, por máxima verossimilhança via algoritmo EM ou por inferência bayesiana.

O modelo de misturas finitas cujas componentes são a densidade família de distribuição de Valor Extremo Generalizada (GEV) abrangem diversos fenômenos. Escalante-Sandoval (2007) utiliza misturas das componentes GEV para avaliar dados sobre a frequência de fluídos de rios de uma região do Noroeste do México, estimando os parâmetros por máxima verossimilhança. No trabalho de Otiniano e Teixeira (2014) os estimadores dos parâmetros da mistura de duas componentes GEV são obtidos via algoritmo EM, e a aplicação foi para dados de energia.

No entanto, antes de discutir problemas de estimação é importante investigar a propriedade de identificabilidade da mistura, que garante a unicidade dos parâmetros. A identificabilidade de misturas é discutida por vários autores, incluindo Teicher (1963), Chandra (1977), Al-Hussaini e Ahmad (1981), Ahmad (1988), Atienza *et al.* (2006). Especificamente, Otiniano *et al.* (2016) mostra a identificabilidade para misturas arbitrárias de distribuições do valor extremo.

Outro problema interessante relacionado aos modelos de mistura de distribuições é o estudo do discriminante, pois ele permite classificar de qual população um valor amostral pertence. A análise de discriminante, no contexto de mistura de distribuições, quando o número de classes é conhecido, está baseada na suposição de que as observações na  $i$ -ésima classe são geradas por uma distribuição de probabilidade  $F_i$ ,  $i = 1, 2, \dots, k$ . Então, se  $\pi_i$  é a proporção de indivíduos da população relativa à classe  $i$ , o teorema de Bayes diz que a probabilidade a posteriori de que a observação  $y$  pertença à classe  $i$  é

$$P(y \in \pi_i) = \frac{P(y|\pi_i) P(\pi_i)}{\sum_{i=1}^k P(y|\pi_i) P(\pi_i)} = \frac{p_i f_i(y)}{\sum_{i=1}^k p_i f_i(y)},$$

atribuindo  $y$  à classe que possui a maior probabilidade a posteriori, minimizando a taxa de erro esperado. Esse é o chamado classificador de Bayes. Quando o tamanho da amostra é pequeno, vários autores utilizam esse classificador, entre eles pode-se citar Amoh (1984) (para mistura de duas distribuições gaussianas inversas), Mahmoud e Moustafa (1993) (para uma mistura de duas distribuições gama), Ahmad e Abdelrahman (1994) (para uma mistura de duas distribuições Weibull) e, recentemente, Ahmad *et al.* (2010) (para uma mistura de duas distribuições Gumbel).

Como nesses trabalhos citados, nesta dissertação utiliza-se o classificador de Bayes para a análise de discriminante não linear para a mistura de distribuições GEV. Assim, organiza-se este trabalho em três capítulos. No Capítulo 1, descreve-se os três tipos de distribuições de valor extremo e a distribuição GEV, também são apontadas as relações entre a GEV e essas três distribuições. Em seguida, aborda-se os conceitos básicos de mistura finita de distribuições, a propriedade de identificabilidade e a análise discriminante.

No segundo capítulo está o estudo do discriminante da mistura de duas componentes GEV, com a abordagem da estimação por máxima verossimilhança.

No Capítulo 3, são realizadas simulações de dados de distribuições GEV com intuito de estudar o comportamento e desempenho das funções discriminantes apresentadas no Capítulo 2. Além disso, retrata-se dois estudos de casos reais.

Por fim, apresenta-se as conclusões, apontando limitações do trabalho e recomendações para trabalhos futuros.

# Capítulo 1

## Conceitos preliminares

### 1.1 Distribuição de valor extremo generalizada

O aquecimento global é uma das causas de eventos climáticos extremos, tais como grandes inundações ou secas extremas, altas temperaturas, furações e tsunamis. Por esse motivo, torna-se importante compreender o comportamento dos valores extremos de determinadas grandezas ou de determinados processos. Por exemplo, estudar os níveis máximos e mínimos de precipitação pluviométrica ao longo de um ano em uma determinada região geográfica, ou as temperaturas máximas mensais em determinada cidade, velocidades máximas de ventos durante um ano em determinada localidade ou ainda os máximos diários dos níveis de maré em determinado trecho do litoral.

A teoria dos valores extremos tem aplicações em outras áreas além de hidrologia e por isso tem atraído a atenção de muitos estudiosos de diferentes ramos. São encontrados problemas relacionados com valores extremos nas áreas de astronomia, engenharias, geociências, economia e ciências atuariais. Em Kotz e Nadajarah (2000) encontra-se algumas aplicações tais como terremotos, corrida de cavalo, tempestades, filas de supermercados, correntes marítimas, velocidades do vento e recordes em corridas.

Nesse sentido, o estudo estatístico de valores extremos torna-se uma ferramenta muito útil, permitindo modelar tais situações.

A teoria de valores extremos trata do estudo da distribuição limite das estatís-

ticas do máximo ou mínimo normalizadas de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Note que, se  $X_1, X_2, \dots, X_n$  são variáveis aleatórias i.i.d. com função de distribuição comum  $F(x)$ , a função de distribuição de  $X_{(n)} = \max \{X_1, X_2, \dots, X_n\}$  é dada por

$$F_{X_{(n)}}(x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = [F(x)]^n,$$

e

$$F^n(x) \rightarrow H(x) = \begin{cases} 0, & x < x_0 \\ 1, & x \geq x_0 \end{cases}, n \rightarrow \infty,$$

sendo  $x_0$  o menor valor tal que  $F(x) = 1$ .

A função  $H(x)$  representa uma função de distribuição de uma variável aleatória degenerada  $Y$ , tal que  $P(Y = x_0) = 1$ .

Desse modo, a fim de se obter uma distribuição limite não-degenerada de  $X_{(n)}$ , define-se uma normalização linear da variável  $X_{(n)}$ :

$$X_{(n)}^* = \frac{X_{(n)} - b_n}{a_n},$$

em que  $a_n$  e  $b_n$  são seqüências de constantes apropriadas. Se existem seqüências de constantes  $a_n > 0$  e  $b_n$  tais que

$$P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) \rightarrow G(x), \text{ quando } n \rightarrow \infty,$$

em que  $G(x)$  é uma função de distribuição não degenerada, então  $G(x)$  pertence à família de distribuições extremas. Gnedenko (1943) apresentou expressões para a função  $G$ , que é de um dos três tipos:

$$\text{I- Gumbel: } \Lambda(x; \mu, \sigma) = \exp\left\{-\exp\left[-\left(\frac{x - \mu}{\sigma}\right)\right]\right\}, -\infty < x < \infty;$$

$$\text{II-Fr\'echet: } \Phi(x; \alpha, \mu, \sigma) = \begin{cases} 0, & x < \mu \\ \exp \left\{ - \left( \frac{x-\mu}{\sigma} \right)^{-\alpha} \right\}, & x \geq \mu \end{cases};$$

$$\text{III-Weibull: } \Psi(x; \alpha, \mu, \sigma) = \begin{cases} \exp \left\{ - \left[ - \left( \frac{x-\mu}{\sigma} \right)^\alpha \right] \right\}, & x < \mu \\ 1, & x > \mu \end{cases};$$

em que  $\sigma > 0$ ,  $\mu \in \mathbb{R}$  e  $\alpha > 0$ .

Dessa maneira, quando ocorrer a convergência em distribuição da variável aleatória  $X_{(n)}^*$ , a distribuição limite  $G(x)$  irá pertencer a uma das três famílias das distribuições de valor extremo dos tipos I, II ou III, também conhecidas como famílias Gumbel, Fréchet e Weibull, respectivamente.

Neste sentido, para modelar um determinado fenômeno ou processo de eventos extremos (de máximos ou mínimos) há a necessidade de se escolher a aproximação por uma distribuição dentro dessas famílias. Jenkinson (1955) obteve a distribuição GEV que engloba essas três distribuições. A forma padrão da GEV é

$$F(x; \gamma, \mu, \sigma) = \begin{cases} \exp \left\{ - \left[ 1 + \gamma \left( \frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\}, & \text{se } \gamma \neq 0 \text{ e } 1 + \gamma \left( \frac{x-\mu}{\sigma} \right) > 0, \\ \exp \left\{ - \exp \left[ - \left( \frac{x-\mu}{\sigma} \right) \right] \right\}, & \text{se } \gamma = 0. \end{cases}$$

A distribuição GEV possui três parâmetros:  $\mu$  de localização,  $\sigma$  de escala e  $\gamma$  de forma.

O parâmetro  $\gamma$  é denominado índice caudal e tem relação com o parâmetro de forma  $\alpha$  dada por:

- $\gamma = 0$ , tem-se a Distribuição Gumbel, com  $F(x; \gamma, \mu, \sigma) = \Lambda(x; \mu, \sigma)$ ;
- $\gamma = \alpha^{-1}$ , a Distribuição Fréchet, com  $F(x; \gamma, \mu, \sigma) = \Phi(\sigma + \gamma x; \frac{1}{\gamma}, \gamma\mu, \sigma)$ ;
- $\gamma = -\alpha^{-1}$ , a Distribuição Weibull, com  $F(x; \gamma, \mu, \sigma) = \Psi(-\sigma - \gamma x; \frac{-1}{\gamma}, \gamma\mu, \sigma)$ .

A Figura 1.1 apresenta a função de densidade de probabilidade para cada caso das distribuições de valores extremos padrão ( $\mu = 0$  e  $\sigma = 1$ ) com  $\alpha = 1$ . Já na Figura 1.2 tem-se a representação da função de densidade da GEV com as três possibilidades para  $\gamma$ :  $\gamma < 0$ ,  $\gamma > 0$  e  $\gamma = 0$ .

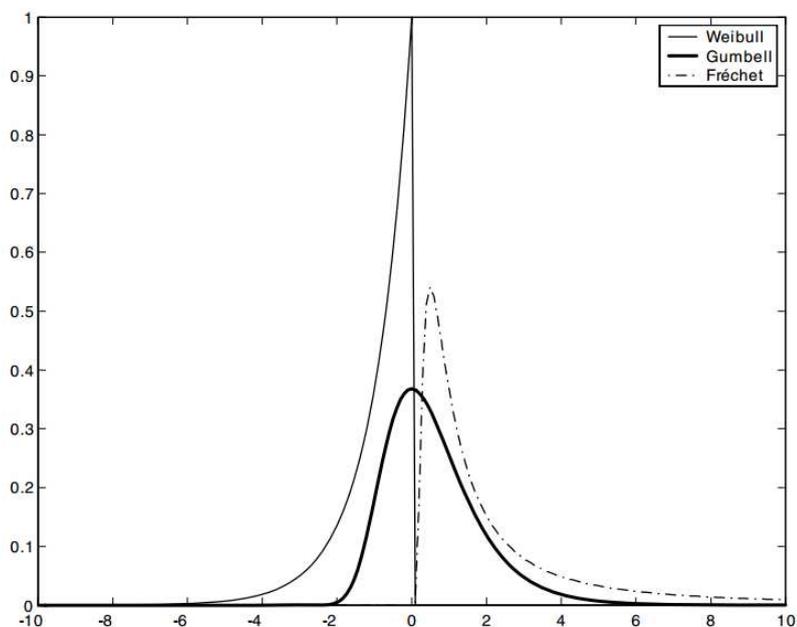


Figura 1.1: Função de densidade de probabilidade das distribuições de valores extremos padrão (com  $\alpha = 1$ ).

Fonte: (Jondeau *et al.*, 2007)

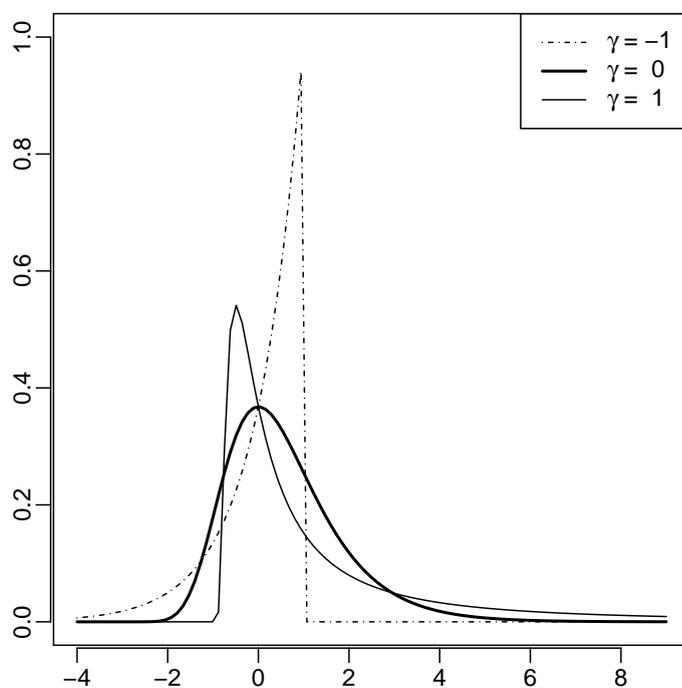


Figura 1.2: Função de densidade da GEV para  $\gamma$  maior, menor e igual a 0.

No Capítulo 2, estuda-se o discriminante de um mistura com duas componentes GEV. Por esse motivo, neste capítulo ainda será discutido sobre os conceitos básicos mistura finita de distribuições, sobre a propriedade de identificabilidade e análise

discriminante.

## 1.2 Mistura de distribuições

Misturas finitas de densidades são modelos de probabilidade que podem ser utilizados em uma ampla variedade de fenômenos. Tem sido utilizado principalmente onde há heterogeneidade populacional, para modelar dados constituídos por uma densidade assimétrica e para aproximação de densidades complexas, por exemplo, as com aspectos multimodais.

Esses modelos são aplicados em várias áreas da estatística: análise de agrupamento, análise discriminante, análise de sobrevivência, métodos não-paramétricos ou semi-paramétricos e processamento de imagens. Também são utilizados em várias outras áreas tais como astronomia, biologia, genética, medicina, psiquiatria, economia e engenharia. Teoria e aplicações de modelos de mistura finita podem ser encontrados em, por exemplo, McLachlan e Basford (1988), McLachlan e Peel (2000) e Titterington *et al.* (1985).

Um modelo de mistura finita é uma combinação convexa das funções de distribuição acumuladas. Mais precisamente, se  $\mathcal{F}$  é uma família de função de distribuição acumuladas e  $F_1, F_2, \dots, F_k \in \mathcal{F}$ , então

$$H(x; \Theta) = \sum_{i=1}^k \pi_i F_i(x; \theta_i) \quad (1.1)$$

é uma mistura finita de  $k$  componentes  $F_1, F_2, \dots, F_k$  com os pesos

$$\pi_1 > 0, \pi_2 > 0, \dots, \pi_k > 0, \text{ tal que } \sum_{i=1}^k \pi_i = 1,$$

sendo que  $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$  é o vetor de parâmetros do modelo.

A mistura (1.1) também pode ser caracterizada por

$$h(x; \Theta) = \sum_{i=1}^k \pi_i f_i(x; \theta_i), \quad (1.2)$$

sendo  $f_1, f_2, \dots, f_k$  as densidades correspondentes às  $F_1, F_2, \dots, F_k$ .

A classe de mistura finita da família  $\mathcal{H}$  é dada por:

$$\mathcal{H} = \left\{ H : H(x; \Theta) = \sum_{i=1}^k \pi_i F_i(x; \theta_i); F_i(x; \theta_i) \in \mathcal{F} \right\}. \quad (1.3)$$

Neste caso, as componentes  $F_i$  podem pertencer a uma mesma família  $\mathcal{F}$ , mas em geral  $\mathcal{F}$  pode ser a união de diferentes famílias de funções de distribuição.

O modelo de misturas finitas cujas componentes são as densidades da distribuição GEV tem ampla aplicabilidade. Escalante-Sandoval (2007) utiliza misturas das componentes GEV para avaliar dados sobre a frequência de fluídos de rios de uma região do Noroeste do México, estimando os parâmetros por máxima verossimilhança. No estudo de Otiniano e Teixeira (2014) os estimadores dos parâmetros da mistura de duas componentes GEV são obtidos via algoritmo EM, e também é apresentado uma aplicação para dados reais.

Na Figura 1.3 tem-se exemplos da mistura de GEVs para os seguintes casos:

- (a)  $\gamma_1 = 0$  e  $\gamma_2 = 0$ ;
- (b)  $\gamma_1 > 0$  e  $\gamma_2 > 0$ ;
- (c)  $\gamma_1 < 0$  e  $\gamma_2 < 0$ ;
- (d)  $\gamma_1 > 0$  e  $\gamma_2 = 0$ ;
- (e)  $\gamma_1 < 0$  e  $\gamma_2 = 0$ ;
- (f)  $\gamma_1 > 0$  e  $\gamma_2 < 0$ ;

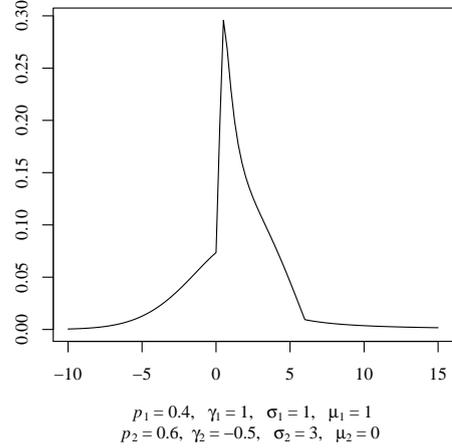
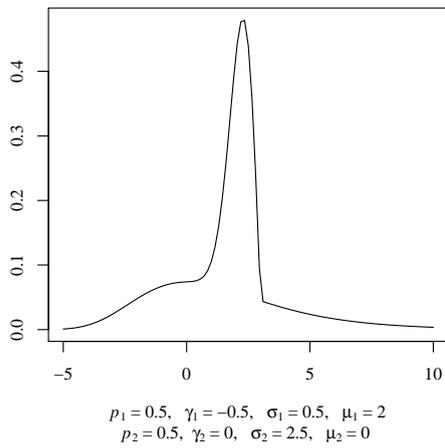
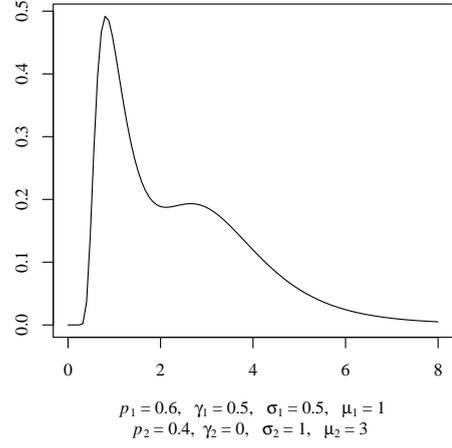
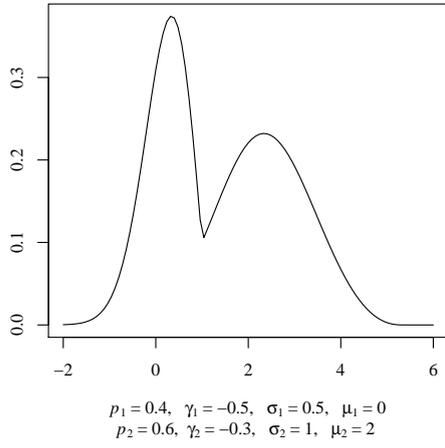
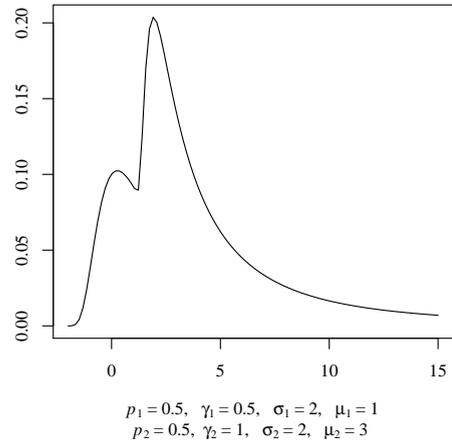
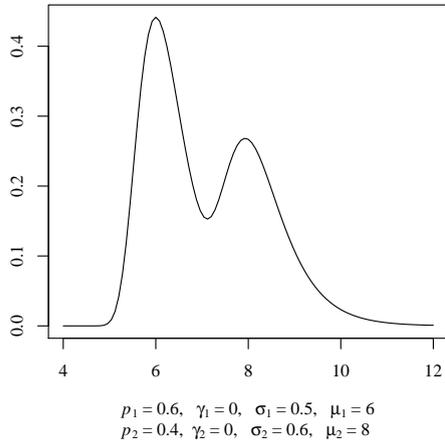


Figura 1.3: Curvas de densidade da mistura de funções GEV: (a), (b), (c), (d), (e), (f)

## 1.3 Identificabilidade

Identificabilidade é uma propriedade relativa ao modelo que garante a unicidade dos parâmetros. Desta maneira, antes de discutir problemas de estimação é importante investigar a propriedade de identificabilidade da mistura. Em termos matemáticos, a classe de funções  $\mathcal{F} = \{F(\cdot; \Theta) : F \text{ é distribuição de probabilidade}\}$  é identificável se para  $F \in \mathcal{F}$ ,  $\Theta \neq \Theta^*$  tem-se que  $F(x; \Theta) \neq F(x; \Theta^*)$  para  $\forall x$ .

A propriedade de identificabilidade para mistura de distribuições é definida a seguir. A classe  $\mathcal{H}$ , definida (1.3), de todas as misturas finitas de  $\mathcal{F}$  é identificável se, e somente se, para  $H, H' \in \mathcal{H}$ ,

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

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

Este tema é discutido por vários autores, incluindo Teicher (1963), Chandra (1977), Al-Hussaini e Ahmad (1981), Ahmad (1988), Atienza *et al.* (2006).

A identificabilidade da classe de mistura de distribuições GEV ainda não é conhecida, porém a identificabilidade da classe de mistura de distribuições extremais foi provada no trabalho de Otiniano *et al.* (2016). Como há uma relação estreita entre distribuições extremais e distribuições GEV, nesta seção é apresentado a identificabilidade da classe de distribuições extremais, conforme Otiniano *et al.* (2016).

Para realizar essa demonstração é necessário apresentar alguns resultados preliminares. Primeiro, considere o resultado dado por Atienza *et al.* (2006), que fornece uma condição necessária e suficiente para identificabilidade de uma classe de misturas finitas.

**Lema 1.3.1.** (Atienza *et al.*, 2006) Seja  $\mathcal{F}$  uma família de funções de distribuição. Seja  $M$  uma aplicação linear que transforma qualquer  $F \in \mathcal{F}$  numa função real  $\varphi_F$  com domínio  $S(F) \in \mathbb{R}^d$ . Seja  $S_0(F) = \{t \in S(F) : \varphi_F \neq 0\}$  e suponha que exista  $t_0$  satisfazendo

$$t_0 \in \left[ \bigcap_{1 \leq i \leq k} S_0(F_i) \right]', \quad (1.4)$$

para qualquer coleção finita de distribuições  $F_1, F_2, \dots, F_k \in \mathcal{F}$ . Se a ordem

$$F_1 \prec F_2 \text{ se e somente se } \lim_{x \rightarrow t_0} \frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = 0$$

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

Para provar a identificabilidade também utiliza-se a transformada de Laplace das distribuições de valores extremos do tipo 1 e tipo 2. Essa expressão depende da função de Kräzel (Kilbas *et al.*, 2010):

$$Z_\rho^\nu = \int_0^\infty s^{\nu-1} \exp\{-s^\rho - xs^{-1}\} ds,$$

definida para  $x > 0$ ,  $\rho \in \mathbb{R}$ ,  $\nu \in \mathbb{C}$ , com  $Re(\nu) < 0$  quando  $\rho \leq 0$ , sendo que  $Re(\nu)$  é a parte real de  $\nu$ .

Utilizando essa função, foi obtida a transformada de Laplace das distribuições extremas. Conforme Otiniano *et al.* (2016), define-se  $\mathcal{F}_{EVD_1}$ , como sendo

$$\mathcal{F}_{EVD_1} = \mathcal{F}_\Lambda \cup \mathcal{F}_\Phi$$

com

$$\mathcal{F}_\Lambda = \{F : F(x) = \Lambda(x; \sigma, \mu)\}$$

e

$$\mathcal{F}_\Phi = \{F : F(x) = \Phi(x; \alpha, \sigma, \mu)\}.$$

Desta maneira, considere  $X$  uma variável aleatória com função de distribuição  $F(\cdot; \alpha, \sigma, \mu) \in \mathcal{F}_{EDV_1}$ . Então, a transformada de Laplace de  $X$  é dado por

$$\varphi_X(t) = \exp\{-t\mu\} \mathcal{K}_{\alpha, \sigma(t)}, \quad (1.5)$$

sendo que

$$\mathcal{K}_{\alpha,\sigma(t)} = \begin{cases} \alpha Z_{\alpha}^{\alpha}(\sigma t), & t > 0 \text{ se } F \in \mathcal{F}_{\Phi} \\ \Gamma(1 + \sigma t), & t > -\frac{1}{\sigma} \text{ se } F \in \mathcal{F}_{\Lambda} \end{cases}. \quad (1.6)$$

(em que  $\Gamma(x) = \int_0^{\infty} x^{t-1} e^{-x} dx$ ).

A função de Kräzel possui algumas propriedades assintóticas que são descritas pelo Lema 1.3.2. Para isto, considere a definição

$$f(x) = O(g(x)), \text{ quando } x \rightarrow +\infty,$$

se, e somente se, existe  $C > 0$  tal que  $|\frac{f(x)}{g(x)}| \leq C$ , para  $x$  suficientemente grande.

**Lema 1.3.2.** (Kilbas *et al.*, 2010)

(a) Seja  $z \in \mathbb{C}$ , ( $z \neq 0$ ),  $\nu \in \mathbb{C}$ ,  $\rho > 0$  e  $\epsilon \in \mathbb{R}$  tal que  $0 < \epsilon < (\rho + 1)\pi/(2\rho)$ .

$$Z_{\rho}^{\nu} = a z^{(2\nu-\rho)/[2(\rho+1)]} \exp[-b z^{\rho/(\rho+1)}] [1 + O(z^{-\rho/(\rho+1)})], \text{ quando } z \rightarrow +\infty, \quad (1.7)$$

sendo que

$$a = [2\pi/(\rho + 1)]^{1/2} \rho^{-\frac{(2\nu+1)}{[2(\rho+1)]}}, b = (1 + 1/\rho)\rho^{1/(\rho+1)}.$$

Aqui, a convergência é uniforme para  $|\arg(x)| < (\rho + 1)\pi/(2\rho) - \epsilon$  para  $\epsilon \in (0, (\rho + 1)\pi/(2\rho))$ .

(b) Seja  $z \in \mathbb{C}$ ,  $\nu \in \mathbb{C}$  ( $Re(\nu) < 0$ ),  $\rho < 0$  e  $|\arg(x)| < (\rho + 1)\pi/(2\rho) - \epsilon$ , então

$$Z_{\rho}^{\nu}(z) = O(1), \text{ para } z \rightarrow 0$$

e

$$Z_{\rho}^{\nu}(z) = O(z^{Re(\nu)}), \text{ para } z \rightarrow \infty.$$

Na proposição a seguir, utiliza-se o Lemma 1.3.1, o Lema 1.3.2 e os outros resultados apresentados nessa seção para provar a identificabilidade de todas as classes de mistura finita com componentes GEV do tipo I e tipo II.

**Proposição 1.3.1.** (Otiniano *et al.*, 2016) A classe  $\mathcal{H}_{EDV_1}$  de todas as misturas finitas de  $\mathcal{F}_{EDV_1}$  é identificável.

**Prova da Proposição 1.3.1.** Como  $\mathcal{F}_{EDV_1} = \mathcal{F}_\Lambda \cup \mathcal{F}_\Phi$ , há três casos possíveis para provar. Para isso, considere  $M$  uma transformação linear que transforma  $F \in \mathcal{F}$  na transformada de Laplace  $\varphi_F$ .

Para todos os casos,  $S_0(F)$  é da forma  $(a(F), +\infty)$ , então o ponto  $t_0 = +\infty$  verifica a condição (1.4) do Lemma 1.3.2.

De fato,

(i) Se  $F_1, F_2 \in \mathcal{F}_\Phi$  então  $S_0(F_i) = (0, +\infty), i = 1, 2$ , utilizando (1.5) e (1.6). Para provar que (1.4) é satisfeita, utiliza-se a representação (1.7) em que  $Z_{\alpha_i}^{\alpha_i}(\sigma_i t)$ . Então, quando  $t \rightarrow +\infty$

$$\varphi_{F_i}(t) = a_i \alpha_i (\sigma_i t)^{c_i/2} \exp \{ -\mu t - b_i (\sigma_i t)^{c_i} \} [1 + O(t^{-c_1})], \quad (1.8)$$

sendo que

$$a_i = \left( \frac{2\pi}{\alpha_i + 1} \right)^{(1/2)} \alpha_i^{-\frac{2\alpha_i+1}{2(\alpha_i+1)}}, \quad b_i = (1 + 1/\alpha_i) \alpha_i^{c_i}, \quad e \quad c_i = \frac{1}{\alpha_i + 1}, \quad i = 1, 2.$$

Agora, quando  $t \rightarrow +\infty$ , partir de (1.8) tem-se

$$\frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = \frac{\alpha_2 a_2 (\sigma_2)^{c_2/2} [1 + O(t^{-c_2})]}{\alpha_1 a_1 (\sigma_1)^{c_1/2} [1 + O(t^{-c_1})]} \times \exp \left\{ -t \left[ (\mu_2 - \mu_1) - b_2 (\sigma_2 t)^{c_2} \times \right. \right. \\ \left. \left. \times \left( t^{-1} - \left( \frac{c_2 - c_1}{2b_2(\sigma_2)^{c_2}} \right) \times \frac{\ln(t)}{t^{c_2+1}} - \frac{b_1(\sigma_1)^{c_1} t^{c_1-1}}{(\sigma_2)^{c_2} t^{c_2}} \right) \right] \right\}$$

e, como  $0 < c_i < 1, i = 1, 2$ , tem-se  $\lim_{x \rightarrow \infty} \frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = 0$  se  $[\mu_1 < \mu_2]$  ou  $[\mu_1 = \mu_2$  e  $\alpha_2 > \alpha_1]$  ou  $[\mu_1 = \mu_2, \alpha_1 = \alpha_2$  e  $\sigma_2 > \sigma_1]$ .

(ii) Se  $F_1, F_2 \in \mathcal{F}_\Lambda$  então  $S_0(F_i) = \left(-\frac{1}{\sigma_i}, +\infty\right), i = 1, 2$ , utilizando (1.5) e (1.6).

Usando a fórmula de Stirling:  $\Gamma(z+1) \sim \sqrt{2\pi} z^{z+1/2} e^{-z}$  em (1.6), tem-se que

$$\varphi_{F_i}(t) \sim \sqrt{2\pi} \exp\{-\mu_i - \sigma_i t\} (\sigma_i t)^{\sigma_i t + 1/2}, \text{ quando } t \rightarrow +\infty \quad (1.9)$$

e

$$\lim_{x \rightarrow \infty} \frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = \lim_{x \rightarrow \infty} \exp \left\{ -t \ln t \left[ (\sigma_1 - \sigma_2) - \frac{\mu_1 - \mu_2}{t \ln t} - \frac{(\sigma_1 - \sigma_2) + \sigma_2 \ln \sigma_2 - \sigma_1 \ln \sigma_1}{\ln t} \right] \right\}.$$

Assim,  $\lim_{x \rightarrow \infty} \frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = 0$  se e somente se  $[\sigma_2 < \sigma_1]$  ou  $[\mu_1 < \mu_2$  e  $\sigma_2 = \sigma_1]$ .

(iii) Se  $F_1 \in \mathcal{F}_\Lambda$  e  $F_2 \in \mathcal{F}_\Phi$  então  $S_0(F_1) = \left(-\frac{1}{\sigma_1}, +\infty\right)$  e  $S_0(F_2) = (0, +\infty)$  ao utilizar (1.5) e (1.6).

Por outro lado, o comportamento assintótico de  $\varphi_{F_1}(t)$  e  $\varphi_{F_2}(t)$  é dado por (1.9) e (1.8), respectivamente. Portanto, quando  $t \rightarrow +\infty$

$$\begin{aligned}
\frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} &= \frac{\alpha_2 a_2 \sigma_2^{c_2/2} \exp\{-\mu_2 t - b_2 \sigma_2^{c_2} t^{c_2}\} t^{c_2/2} [1 + O(t^{-c_2})]}{\sqrt{2\pi} \exp\{-\mu_1 t - \sigma_1 t\} (\sigma_1 t)^{\sigma_1 t + 1/2}} \\
&= \frac{\alpha_2 a_2 \sigma_2^{c_2/2}}{\sqrt{2\pi} (\sigma_1)^{c_2/2}} [1 + O(t^{-c_2})] \times \\
&\times \exp\{-(\mu_2 - \mu_1)t + \sigma_1 t - b_2 \sigma_2^{c_2} t^{c_2}\} (\sigma_1 t)^{c_2/2 - \sigma_1 t - 1/2} \\
&= \frac{\alpha_2 a_2 \sigma_2^{c_2/2}}{\sqrt{2\pi} (\sigma_1)^{c_2/2}} [1 + O(t^{-c_2})] \times \\
&\times \exp\left\{-(\mu_2 - \mu_1)t + \sigma_1 t - b_2 \sigma_2^{c_2} t^{c_2} + \left(\frac{c_2}{2} - \sigma_1 t - 1/2\right) \ln(\sigma_1 t)\right\} \\
&= \frac{\alpha_2 a_2 \sigma_2^{c_2/2}}{\sqrt{2\pi} (\sigma_1)^{c_2/2}} [1 + O(t^{-c_2})] \times \\
&\times \exp\left\{-(\mu_2 - \mu_1)t - \sigma_1 t \ln(\sigma_1 t) \left[1 - \frac{1}{\ln(\sigma_1 t)}\right] - \left(\frac{1 - c_2}{2} \ln(\sigma_1 t)\right) - b_2 \sigma_2^{c_2} t^{c_2}\right\}
\end{aligned}$$

e  $\lim_{x \rightarrow \infty} \frac{\varphi_{F_2}(t)}{\varphi_{F_1}(t)} = 0$  se  $\mu_2 \geq \mu_1$ , porque  $0 < c_2 < 1$ .

## 1.4 Análise Discriminante

A análise discriminante é uma técnica utilizada para classificação de elementos de uma amostra ou população em grupos predefinidos, ou seja, conhecidos a priori. Esse conhecimento permite a elaboração de uma função matemática chamada regra de classificação ou discriminação, que será fundamentada na teoria de probabilidade (Mingoti, 2005).

No livro de Mingoti (2005) são citadas muitas situações nas quais o interesse está na discriminação de grupos, como na agricultura, em que é importante identificar as áreas de maior potencial para plantação de certa cultura; em marketing, na identificação de mercados potenciais e não potenciais para determinado produto; em finanças, diferenciar se uma pessoa que está pleitando um empréstimo bancário será inadimplente ou não, no ensino, discriminar os estudantes com maior potencial para concluir um determinado curso e os com menor potencial.

No entanto, este trabalho trata da análise discriminante no contexto de misturas

de distribuições. O estudo de discriminante para misturas foi introduzido por O'Neill (1978), que obteve a função discriminante para mistura de normais.

### 1.4.1 Função discriminante para mistura de distribuições

A função discriminante de uma mistura de distribuições (1.2) é extremamente útil para classificar de qual população (componente) um valor amostral pertence. Suponha que há duas populações  $\pi_1$  e  $\pi_2$  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  $\pi_i$ ,  $i = 1, 2$  é dada por:

$$P(x \in \pi_i) = \frac{p_i f_i(x)}{p_1 f_1(x) + p_2 f_2(x)}. \quad (1.10)$$

A função logística, bem como uma função de probabilidade, assume valores entre 0 a 1. Desta maneira, a função de probabilidade a posteriori, (1.10), é expressa em termos da função logística, com interesse de simplificar a função discriminante. Desta maneira, tem-se:

$$P(x \in \pi_1) = \frac{1}{1 + \exp(a - bz)} \quad (1.11)$$

e

$$P(x \in \pi_2) = \frac{\exp(a - bz)}{1 + \exp(a - bz)}, \quad (1.12)$$

sendo  $a$ ,  $b$  constantes 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. Assim, classifica-se  $x$  em  $\pi_1$  se  $P(x \in \pi_1) > P(x \in \pi_2)$  ou  $a - bz > 0$ , (1.13), pois as seguintes desigualdades são equivalentes.

$$\begin{aligned}
P(x \in \pi_1) &> P(x \in \pi_2) \\
\frac{1}{1 + \exp(a - bz)} &> \frac{\exp(a - bz)}{1 + \exp(a - bz)} \\
\exp(a - bz) &< 1 \\
a - bz &< 0.
\end{aligned} \tag{1.13}$$

De forma análoga, uma observação é classificada em  $\pi_2$  se  $a - bz > 0$ . Desse modo, a função discriminante de uma mistura de distribuições é definida por:

$$NL(x) = a - bz, \tag{1.14}$$

que permite classificar  $x$  em  $\pi_1$  se  $NL(x) < 0$  e em  $\pi_2$  se  $NL(x) \geq 0$ . Além disso, pode-se observar que  $x$  será classificado em  $\pi_1$  se  $P(x \in \pi_1) > \frac{1}{2}$ , pois

$$\begin{aligned}
P(x \in \pi_1) > \frac{1}{2} &\Leftrightarrow \{1 + \exp\{NL(x)\}\}^{-1} > \frac{1}{2} \\
&\Leftrightarrow 1 + \exp\{NL(x)\} < 2 \\
&\Leftrightarrow \exp\{NL(x)\} < 1 \\
&\Leftrightarrow NL(x) < 0.
\end{aligned}$$

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

## 1.4.2 Erros de classificação

Todo processo de tomada de decisões traz consigo um possível erro de decisão. Portanto, é importante definir o erro de classificar uma observação em uma população quando, na realidade, ela pertence a outra.

Classifica-se uma observação  $x$  em  $\pi_1$ , 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  $\pi_1$ , utilizando a função discriminante  $NL(x)$ , é dada por:

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

Analogamente, a probabilidade classificar de forma errada uma observação  $x$  para  $\pi_2$ , por meio da função discriminante  $NL(x)$  é obtido por:

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

## Capítulo 2

# Estudo do discriminante de mistura de GEV

Neste capítulo apresenta-se o estudo do discriminante da mistura de duas componentes GEV. Inicialmente, aborda-se a estimação por máxima verossimilhança da mistura de duas distribuições GEV, pois para estimar a função discriminante é necessário utilizar as estimativas dos parâmetros da mistura. São obtidas as estimativas para uma amostra classificada e não classificada.

### 2.1 Estimação por máxima verossimilhança

Seja  $X$  uma variável aleatória com função de distribuição  $H$ , mistura de duas componentes GEV, com parâmetro de escala comum, cuja a função de densidade de probabilidade (f.d.p.) é dada por:

$$h(x; \Theta) = \frac{1}{\sigma} [p_1 g_1(x; \gamma_1, \mu_1, \sigma) + p_2 g_2(x; \gamma_2, \mu_2, \sigma)], p_1 + p_2 = 1, \quad (2.1)$$

sendo  $g_i, i = 1, 2$ , a  $i$ -ésima componente da mistura, dado por

$$g_i(x; \gamma_i, \mu_i, \sigma) = \begin{cases} \left[ 1 + \frac{\gamma_i}{\sigma}(x - \mu_i) \right]^{-\frac{1}{\gamma_i} - 1} \exp \left\{ - \left[ 1 + \frac{\gamma_i}{\sigma}(x - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\}, & \gamma_i \neq 0 \\ \exp \left\{ - \left( \frac{x - \mu_i}{\sigma} \right) - \exp \left[ - \left( \frac{x - \mu_i}{\sigma} \right) \right] \right\}, & \gamma_i = 0 \end{cases}, \quad (2.2)$$

para  $1 + \frac{\gamma_i}{\sigma}(x - \mu_i) > 0$ .

O modelo (2.1) possui seis parâmetros:  $\gamma_i \in \mathbb{R}$  de forma,  $\mu_i \in \mathbb{R}$  de localização,  $\sigma > 0$  de escala,  $i=1,2$ ,  $p_1$  e  $p_2$  são as proporções das componentes  $g_1$  e  $g_2$  da mistura. Dessa maneira, o vetor de parâmetros de interesse é dado por  $\Theta = (p_1, \gamma_1, \mu_1, \gamma_2, \mu_2, \sigma)$ .

Geralmente, os parâmetros de uma população são desconhecidos. Os dados disponíveis são utilizados para estimar os parâmetros da função de densidade. Deve-se considerar dois tipos de dados para fazer essa estimação:

- **Amostra não classificada(m)**: Quando os dados são obtidos de uma amostragem da população de uma mistura, mas cada observação é de origem desconhecida.
- **Amostra classificada(c)**: Neste caso, os dados também são amostrados de uma mistura, mas a origem de cada observação é determinada após a amostragem.

Inicia-se apresentando a estimação por máxima verossimilhança para o caso da amostra não classificada.

### 2.1.1 Estimação por máxima verossimilhança para amostra não classificada

Para calcular o estimador de máxima verossimilhança para amostra não classificada, considere os valores  $x_1, x_2, \dots, x_n$  de uma amostra aleatória de tamanho  $n$  de  $X$ . A função de verossimilhança da amostra pode ser escrita como:

$$L(\underline{X}, \Theta) = \prod_{j=1}^n \frac{1}{\sigma} Q(x_j), \quad (2.3)$$

em que  $Q(x_j) = p_1 g_1(x_j; \theta_1) + p_2 g_2(x_j; \theta_2)$ , com  $g_i(x_j; \theta_i)$  dado por (2.2), para  $j = 1, 2, \dots, n$ , e  $i = 1, 2$ .

O logaritmo da função de (2.3) é dado por:

$$l(\underline{X}, \Theta) = -n \ln(\sigma) + \sum_{j=1}^n \ln [p_1 g_1(x_j; \theta_1) + p_2 g_2(x_j; \theta_2)]. \quad (2.4)$$

Como o parâmetro de forma da GEV  $\gamma$  pode assumir valor maior, menor ou igual a zero, tem-se seis modelos correspondentes a (2.1), quando combina-se cada componente com os diferentes sinais de  $\gamma_i$ ,  $i = 1, 2$ .

O primeiro caso corresponde à mistura de duas componentes Gumbel. Este caso também é tratado por Ahmad *et al.* (2010), sendo o parâmetro de escala ( $\sigma$ ) comum nas duas componentes da mistura.

**Caso 1:** ( $\gamma_1 = 0, \gamma_2 = 0$ )

Tomando a derivada parcial de (2.4) com relação ao parâmetro  $p_1$  e igualando a zero, obtém-se

$$\sum_{j=1}^n \frac{g_1(x_j; \theta_1) - g_2(x_j; \theta_2)}{Q(x_j)} = 0. \quad (2.5)$$

Agora, derivando (2.4) com respeito a  $\mu_i$ ,  $i = 1, 2$  e simplificando essa expressão ao igualar a zero,

$$\sum_{j=1}^n \frac{p_i g_i(x_j; \theta_i)}{Q(x_j)} \left\{ \frac{1}{\sigma} - \frac{1}{\sigma} \exp \left[ - \left( \frac{x_j - \mu_i}{\sigma} \right) \right] \right\} = 0. \quad (2.6)$$

Em seguida, ao derivar (2.4) com respeito  $\sigma$ , com algumas simplificações ao igualar a zero, obtém-se:

$$\frac{n}{\sigma} - \sum_{j=1}^n \left\{ \sum_{i=1}^2 \frac{g_i(x_j; \theta_i)}{Q(x_j)} \left\{ \left( \frac{x_j - \mu_i}{\sigma^2} \right) - \left( \frac{x_j - \mu_i}{\sigma^2} \right) \exp \left[ - \left( \frac{x_j - \mu_i}{\sigma} \right) \right] \right\} \right\} = 0.$$

Para simplificar ainda mais essa última equação, escreve-se  $\frac{1}{\sigma^2}$  em evidência. Assim,

$$n\sigma - \sum_{j=1}^n \left\{ \sum_{i=1}^2 \frac{g_i(x_j; \theta_i)}{Q(x_j)} \left\{ (x_j - \mu_i) \times \left\{ 1 - \exp \left[ - \left( \frac{x_j - \mu_i}{\sigma} \right) \right] \right\} \right\} \right\} = 0. \quad (2.7)$$

As estimativas de  $\mu_i$  e  $\sigma$ ,  $i = 1, 2$ , devem ser obtidas ao resolver as equações (2.6) e (2.7), respectivamente, via algum método numérico.

Na verdade, essas estimativas foram obtidas por Escalante-Sandoval (2007) utilizando-se um método de otimização não linear multivariado e com restrições.

Outra alternativa para resolver essas equações é simplificá-las ao definir a função  $w_{ij}$  e, em seguida, resolver as novas equações iterativamente.

Dada uma observação  $x_j$ , a função  $w_{ij}$  é a probabilidade dessa observação vir da  $i$ -ésima componente. Assumindo que as componentes da mistura existem em proporções fixas, tem-se que,

$$w_{1j} = \frac{p_1 g_1(x_j; \theta_1)}{Q(x_j)} \quad e \quad w_{2j} = 1 - w_{1j}. \quad (2.8)$$

Substituindo (2.8) em (2.5) e efetuando algumas operações simples obtém-se:

$$p_1 = \sum_{j=1}^n \frac{w_{1j}}{n}. \quad (2.9)$$

Ao substituir (2.8) em (2.6) e (2.7), segue que as novas equações a serem resolvidas para se obter os estimadores de  $\mu_i$  e  $\sigma$ ,  $i = 1, 2$ , são

$$\sum_{j=1}^n w_{ij}(x_j) \left\{ \frac{1}{\sigma} - \frac{1}{\sigma} \exp \left[ - \left( \frac{x_j - \mu_i}{\sigma} \right) \right] \right\} = 0 \quad (2.10)$$

e

$$n\sigma - \sum_{j=1}^n \left\{ \sum_{i=1}^2 w_{ij}(x_j) \left\{ (x_j - \mu_i) \times \left\{ 1 - \exp \left[ - \left( \frac{x_j - \mu_i}{\sigma} \right) \right] \right\} \right\} \right\} = 0. \quad (2.11)$$

As estimativa dos parâmetros  $\hat{p}_1$ ,  $\hat{\mu}_i$  e  $\hat{\sigma}$ ,  $i=1,2$  são obtidas ao resolver as equações (2.9) até (2.11), iterativamente.

O segundo caso corresponde à mistura de duas componentes GEV com  $\gamma_i > 0$ ,  $i = 1, 2$ . A estimação de um modelo particular desse caso é tratado por Sultan *et al.* (2007), através do algoritmo EM. Eles mostram a estimação do modelo (2.1) quando  $\mu_1 = \mu_2 = 0$  (mistura de duas componentes Inversa Weibull).

**Caso 2:** ( $\gamma_1 > 0, \gamma_2 > 0, \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

A derivada parcial de (2.4) em relação a  $p_1$  é dada pela equação (2.5).

Já a derivada parcial de (2.4) com respeito aos parâmetros  $\mu_i, i = 1, 2$  com algumas simplificações quando igualada a zero, é igual a

$$\sum_{j=1}^n \frac{p_i}{Q(x_j)} \left\{ \left( \frac{1}{\sigma} + \frac{\gamma_i}{\sigma} \right) \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i} - 2} \times \exp \left\{ - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\} - \left( \frac{1}{\sigma} \right) \times \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{2}{\gamma_i} - 2} \exp \left\{ - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\} \right\} = 0.$$

Para simplificar ainda mais essa última equação, escreve-se em evidência  $g_i(x_j; \theta_i)$  na última equação, obtendo-se:

$$\sum_{j=1}^n \frac{p_i g_i(x_j; \theta_i)}{Q(x_j)} \left\{ \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \left\{ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\} \right\} = 0. \quad (2.12)$$

Agora derivando (2.4) com respeito ao parâmetro  $\sigma$ , e simplificando essa expressão quando igualada a zero,

$$\begin{aligned} & - \frac{n}{\sigma} + \sum_{j=1}^n \left\{ \left( -\frac{1}{\gamma_i} - 1 \right) \left( -\frac{\gamma_i}{\sigma^2} (x_j - \mu_i) \right) \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i} - 2} \times \right. \\ & \times \exp \left\{ - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\} - \left( -\frac{1}{\gamma_i} \right) \left( -\frac{\gamma_i}{\sigma^2} (x_j - \mu_i) \right) \times \\ & \left. \times \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{2}{\gamma_i} - 2} \exp \left\{ - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i}} \right\} \right\} = 0. \end{aligned}$$

Para se obter uma expressão ainda mais simples, escreve-se  $\frac{1}{\sigma^2}$  em evidência na

equação acima, e alcança-se:

$$\begin{aligned}
& - n\sigma + \sum_{j=1}^n \left\{ \sum_{i=1}^2 \left\{ \frac{g_i(x_j; \theta_i)}{Q(x_j)} (x_j - \mu_i) \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \right. \right. \\
& \times \left. \left. \left\{ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right\} \right\} \right\} = 0. \tag{2.13}
\end{aligned}$$

A fim de facilitar o cálculo da derivada em relação aos parâmetros  $\gamma_i$ , foi preciso reescrever a função (2.2),

$$\begin{aligned}
g_i(x) &= \exp \left\{ \left( -\frac{1}{\gamma_i} - 1 \right) \ln \left( 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right) - \right. \\
& \left. - \exp \left[ -\frac{1}{\gamma_i} \ln \left( 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right) \right] \right\}.
\end{aligned}$$

Desta maneira, o resultado da derivada parcial de (2.4) em relação o parâmetro  $\gamma_i$  após algumas simplificações quando igualada a zero é igual a

$$\begin{aligned}
& \sum_{j=1}^n \frac{p_i g_i(x_j; \theta_i)}{Q(x_j)} \left\{ \frac{1}{\gamma_i^2} \ln \left( 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right) + \left( -\frac{1}{\gamma_i} \right) \left( \frac{x_j - \mu_i}{\sigma} \right) \times \right. \\
& \times \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} - \left( \frac{x_j - \mu_i}{\sigma} \right) \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} - \\
& - \left\{ \frac{1}{\gamma_i^2} \ln \left( 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right) + \left( -\frac{1}{\gamma_i} \right) \left( \frac{x_j - \mu_i}{\sigma} \right) \times \right. \\
& \times \left. \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \right\} \times \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-\frac{1}{\gamma_i}} \left. \right\} = 0.
\end{aligned}$$

Equivalentemente,

$$\begin{aligned}
& \sum_{j=1}^n \frac{p_i g_i(x_j; \theta_i)}{Q(x_j)} \left\{ (\gamma_i)^{-2} \ln \left( 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right) \times \left[ 1 - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right] - \right. \\
& \left. - \left( \frac{x_j - \mu_i}{\gamma_i \sigma} \right) \times \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \left[ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right] \right\} = 0.
\end{aligned} \tag{2.14}$$

Da mesma maneira que no caso anterior, para obter as estimativas de  $\mu_i$ ,  $\sigma$ ,  $\gamma_i$ ,  $i = 1, 2$  deve-se resolver as equações (2.12), (2.13), (2.14), respectivamente, via algum método numérico. Novamente, como no Caso 1, utiliza-se a função  $w_{ij}$ ,

$$w_{1j}(x_j) = \begin{cases} \frac{p_1 f_1(x_j; \theta_1)}{p_1 f_1(x_j; \theta_1)} = 1, & \mu_1 - \frac{\sigma}{\gamma_1} \leq x_j < \mu_2 - \frac{\sigma}{\gamma_2} \\ \frac{p_1 f_1(x_j; \theta_1)}{Q(x_j)}, & x_j \geq \mu_2 - \frac{\sigma}{\gamma_2} \end{cases}, \text{ e } w_{2j} = 1 - w_{1j}. \tag{2.15}$$

Desta maneira, substituindo (2.15) em (2.5) e efetuando-se algumas operações simples, obtém-se:

$$p_1 = \sum_{j=1}^n \frac{w_{1j}}{n}. \tag{2.16}$$

Ainda, pode-se reescrever as equações (2.12), (2.13) e (2.14) como

$$\sum_{j=1}^n w_{ij}(x_j) \left\{ \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \left\{ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right\} \right\} = 0; \tag{2.17}$$

$$\begin{aligned}
& - n\sigma + \sum_{j=1}^n \left\{ \sum_{i=1}^2 \left\{ w_{ij}(x_j) (x_j - \mu_i) \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \right. \right. \\
& \left. \left. \times \left\{ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right\} \right\} \right\} = 0.
\end{aligned} \tag{2.18}$$

$$\sum_{j=1}^n w_{ij}(x_j) \left\{ (\gamma_i)^{-2} \ln \left( 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right) \times \left[ 1 - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right] - \right. \\ \left. - \left( \frac{x_j - \mu_i}{\gamma_i \sigma} \right) \times \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \left[ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right] \right\} = 0. \quad (2.19)$$

As estimativas dos parâmetros  $p_1$ ,  $\mu_i$ ,  $\sigma$  e  $\gamma_i$   $i=1,2$  devem ser obtidas ao se resolver as equações (2.16) até (2.19), iterativamente, considerando um "chute" inicial.

O caso seguinte corresponde ao último modelo de mistura de componentes da mesma família ( $\gamma_1, \gamma_2 < 0$ , Família Weibull). Um estudo particular desse modelo corresponde a Ahmad e Abd-Elrahman (1994), pois eles consideram componentes Weibull não negativas com  $\mu_1 = \mu_2 = 0$ , que é a versão simétrica deste Caso 3 no suporte positivo.

**Caso 3:** ( $\gamma_1 < 0, \gamma_2 < 0; \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

A derivada parcial de (2.4) em relação os parâmetros  $p_1$ ,  $\mu_i$ ,  $\sigma$  e  $\gamma_i, i = 1, 2$  é dada respectivamente pelas equações (2.5), (2.12), (2.13) e (2.14), sendo que  $w_{ij}$  é dado por:

$$w_{1j}(x_j) = \begin{cases} 0, & \mu_1 - \frac{\sigma}{\gamma_1} < x_j \leq \mu_2 - \frac{\sigma}{\gamma_2} \\ \frac{p_1 f_1(x_j; \theta_1)}{Q(x_j)}, & x_j \leq \mu_1 - \frac{\sigma}{\gamma_1} \end{cases}, \text{ e } w_{2j} = 1 - w_{1j}.$$

Os casos 4, 5 e 6 correspondem à mistura de componentes de famílias distintas.

**Caso 4:** ( $\gamma_1 > 0, \gamma_2 = 0$ )

Neste caso, o resultado da derivada parcial de (2.4) para os parâmetros  $p_1$ ,  $\mu_1$  e  $\gamma_1$  são respectivamente (2.5), (2.12) e (2.14), enquanto, o parâmetro  $\mu_2$  é representado pela equação (2.6).

Já a derivada parcial de (2.4) com respeito a  $\sigma$ , é dado por:

$$\begin{aligned}
& - n\sigma + \sum_{j=1}^n \frac{p_1 g_1(x_j; \theta_1)}{Q(x_j)} (x_j - \mu_i) \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{-1} \times \left\{ 1 + \gamma_i - \left[ 1 + \frac{\gamma_i}{\sigma} (x_j - \mu_i) \right]^{\frac{-1}{\gamma_i}} \right\} \\
& + \sum_{j=1}^n \frac{p_2 g_2(x_j; \theta_2)}{Q(x_j)} \left\{ (x_j - \mu_i) \times \left\{ 1 - \exp \left[ - \left( \frac{x_j - \mu_i}{\sigma} \right) \right] \right\} \right\}
\end{aligned} \tag{2.20}$$

Sendo que  $w_{ij}$  é representado por:

$$w_{1j}(x_j) = \begin{cases} \frac{p_1 f_1(x_j; \theta_1)}{Q(x_j)}, & x_j \geq \mu_1 - \frac{\sigma}{\gamma_1} \\ 0, & x_j < \mu_1 - \frac{\sigma}{\gamma_1} \end{cases}, \text{ e } w_{2j} = 1 - w_{1j}.$$

**Caso 5:** ( $\gamma_1 < 0, \gamma_2 = 0$ )

Obtém-se a derivada parcial de (2.4) com respeito os parâmetros  $p_1, \mu_1$  e  $\gamma_1$  por meio equações (2.5), (2.12), (2.13) e (2.14) e os parâmetros  $\mu_2$  e  $\sigma$  através das equações (2.6) e (2.20), em que  $w_{ij}$  é dado por:

$$w_{1j}(x_j) = \begin{cases} \frac{p_1 f_1(x_j; \theta_1)}{Q(x_j)}, & x_j \leq \mu_1 - \frac{\sigma}{\gamma_1} \\ 0, & x_j > \mu_1 - \frac{\sigma}{\gamma_1} \end{cases}, \text{ e } w_{2j} = 1 - w_{1j}.$$

**Caso 6:** ( $\gamma_1 > 0, \gamma_2 < 0; \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

A derivada parcial de (2.4) em relação os parâmetros  $p_1, \mu_i, \sigma$  e  $\gamma_i, i = 1, 2$  são dados respectivamente pelas equações (2.5), (2.12), (2.13) e (2.14). Analogamente ao dois casos anteriores, no qual  $w_{ij}$  é dado por:

$$w_{1j}(x_j) = \begin{cases} 0, & x_j < \mu_1 - \frac{\sigma_1}{\gamma_1}, \\ \frac{p_1 f_1(x_j; \theta_1)}{Q(x_j)}, & \mu_1 - \frac{\sigma_1}{\gamma_1} \leq x_j \leq \mu_2 - \frac{\sigma_2}{\gamma_2}, \\ \frac{p_1 f_1(x_j; \theta_1)}{p_1 f_1(x_j; \theta_1)} = 1, & x_j > \mu_2 - \frac{\sigma_2}{\gamma_2}, \end{cases} \text{ e } w_{2j} = 1 - w_{1j}.$$

Agora será apresentada a estimação por máxima verossimilhança para o caso da amostra classificada.

### 2.1.2 Estimação por máxima verossimilhança para amostra classificada

Considere uma amostra aleatória de tamanho  $n$  de  $X$  com função de densidade de probabilidade  $h(x; \Theta)$  dada em (2.1). Suponha que há  $n_i$  observações iniciais avaliadas para cada população  $\pi_i$ , com  $n = n_1 + n_2$ . Dessa forma, a amostra aleatória de  $X$  de tamanho  $n$  está dividida da seguinte maneira:  $x_{11}, x_{12}, \dots, x_{1n_1} \in \pi_1$ ,  $x_{21}, x_{22}, \dots, x_{2n_2} \in \pi_2$ .

Assim, a estimativa do parâmetro da proporção de mistura  $p_1$  pode ser obtida de forma explícita:

$$\tilde{p}_1 = \frac{n_1}{n} \quad \text{e} \quad \tilde{p}_2 = 1 - \tilde{p}_1. \quad (2.21)$$

A estimação dos outros parâmetros,  $\Theta = (\gamma_1, \mu_1, \mu_2, \gamma_2, \sigma)$ , é obtida por máxima verossimilhança. Desta maneira, considere a função de verossimilhança dada por:

$$L(\underline{X}, \Theta) = L_1(\underline{X}_1, \Theta^1) \times L_2(\underline{X}_2, \Theta^2),$$

sendo que  $\Theta = (\Theta^1, \Theta^2)$ ,

$$L_1(\underline{X}_1, \Theta^1) = \prod_{j=1}^{n_1} \frac{1}{\sigma} g_1(x_{1j}; \Theta^1)$$

e

$$L_2(\underline{X}_2, \Theta^2) = \prod_{j=1}^{n_2} \frac{1}{\sigma} g_2(x_{2j}; \Theta^2),$$

em que  $\Theta^1 = (\mu_1, \sigma, \gamma_1)$  e  $\Theta^2 = (\mu_2, \sigma, \gamma_2)$ . As densidades  $g_1(x_{1j}; \Theta^1)$  e  $g_2(x_{2j}; \Theta^2)$  são dadas por (2.2).

Dessa maneira, o logaritmo de verossimilhança é dado por

$$l(\underline{X}, \Theta) = l_1(\underline{X}_1, \Theta^1) + l_2(\underline{X}_2, \Theta^2)$$

em que

$$l_1(\underline{X}_1, \Theta^1) = -n_1 \ln(\sigma) + \sum_{j=1}^{n_1} \ln g_1(x_{1j}; \Theta^1)$$

e

$$l_2(\underline{X}_2, \Theta^2) = -n_2 \ln(\sigma) + \sum_{j=1}^{n_2} \ln g_2(x_{2j}; \Theta^2)$$

obtendo

$$l(\underline{X}, \Theta) = -n \ln(\sigma) + \sum_{j=1}^{n_1} \ln g_1(x_{1j}; \Theta^1) + \sum_{j=1}^{n_2} \ln g_2(x_{2j}; \Theta^2) \quad (2.22)$$

A função será calculada para seis casos distintos de acordo com o valor que o parâmetro  $\gamma_i$ ,  $i = 1, 2$  assume em cada componente da mistura.

**Caso 1:** ( $\gamma_1 = 0, \gamma_2 = 0$ )

A derivada parcial de (2.22) com respeito ao parâmetro  $\mu_1$  simplificando a expressão e igualando a zero é dada por:

$$\sum_{j=1}^{n_1} \left\{ \frac{1}{\sigma} - \frac{1}{\sigma} \exp \left[ - \left( \frac{x_{1j} - \mu_1}{\sigma} \right) \right] \right\} = 0; \quad (2.23)$$

De modo análogo, a derivada parcial do logaritmo da função de verossimilhança (2.22) em relação a  $\mu_2$  é dado por:

$$\sum_{j=1}^{n_2} \left\{ \frac{1}{\sigma} - \frac{1}{\sigma} \exp \left[ - \left( \frac{x_{2j} - \mu_2}{\sigma} \right) \right] \right\} = 0; \quad (2.24)$$

Já a derivada parcial de (2.22) com respeito ao parâmetro  $\sigma$  simplificando a expressão e igualando a zero é dada por:

$$\begin{aligned}
& - \frac{n}{\sigma} + \sum_{j=1}^{n_1} \left\{ \left( \frac{x_{1j} - \mu_1}{\sigma^2} \right) \times \left\{ 1 - \exp \left[ - \left( \frac{x_{1j} - \mu_1}{\sigma} \right) \right] \right\} \right\} + \\
& + \sum_{j=1}^{n_2} \left\{ \left( \frac{x_{2j} - \mu_2}{\sigma^2} \right) \times \left\{ 1 - \exp \left[ - \left( \frac{x_{2j} - \mu_2}{\sigma} \right) \right] \right\} \right\}
\end{aligned} \tag{2.25}$$

**Caso 2:**  $(\gamma_1 > 0, \gamma_2 > 0, \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2})$

Tem-se que a derivada parcial de (2.22) em relação  $\mu_1$  com algumas simplificações e igualada a zero é dada por:

$$\begin{aligned}
& \sum_{j=1}^{n_1} \left( -\frac{1}{\gamma_1} - 1 \right) \left( -\frac{\gamma_1}{\sigma} \right) \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-1} - \\
& - \sum_{j=1}^{n_1} \left( -\frac{1}{\gamma_1} \right) \left( -\frac{\gamma_1}{\sigma} \right) \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-\frac{1}{\gamma_1}-1} = 0.
\end{aligned}$$

Equivalentemente,

$$\sum_{j=1}^{n_1} \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-1} \times \left\{ 1 + \gamma_1 - \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} = 0. \tag{2.26}$$

Ao derivar (2.22) em relação a  $\gamma_1$ , com algumas simplificações ao igualar a zero, obtém-se:

$$\begin{aligned}
& \sum_{j=1}^{n_1} \left\{ \frac{1}{\gamma_1^2} \ln \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right] - \frac{1}{\gamma_1} \left( \frac{x_{1j} - \mu_1}{\sigma} \right) \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{-1} - \right. \\
& - \left. \left( \frac{x_{1j} - \mu_1}{\sigma} \right) \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{-1} - \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{-\frac{1}{\gamma_1}} \times \right. \\
& \left. \times \left\{ \frac{1}{\gamma_1^2} \ln \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right] - \frac{1}{\gamma_1} \left( \frac{x_{1j} - \mu_1}{\sigma} \right) \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{-1} \right\} \right\} = 0.
\end{aligned}$$

Equivalentemente,

$$\begin{aligned}
& \sum_{j=1}^{n_1} \left\{ (\gamma_1)^{-2} \ln \left( 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right) \times \left[ 1 - \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{\frac{-1}{\gamma_1}} \right] - \right. \\
& - \left. \left( \frac{x_{1j} - \mu_1}{\gamma_1 \sigma} \right) \times \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{-1} \times \left[ 1 + \gamma_1 - \left[ 1 + \frac{\gamma_1}{\sigma} (x_{1j} - \mu_1) \right]^{\frac{-1}{\gamma_1}} \right] \right\} = 0.
\end{aligned} \tag{2.27}$$

De maneira análoga, a derivada parcial do logaritmo da função de verossimilhança dos parâmetros  $\mu_2$  e  $\gamma_2$  são dados, respectivamente, por:

$$\sum_{j=1}^{n_2} \left[ 1 + \frac{\gamma_2}{\sigma_2} (x_{2j} - \mu_2) \right]^{-1} \times \left\{ 1 + \gamma_2 - \left[ 1 + \frac{\gamma_2}{\sigma_2} (x_{2j} - \mu_2) \right]^{-\frac{1}{\gamma_2}} \right\} = 0 \tag{2.28}$$

e

$$\begin{aligned}
& \sum_{j=1}^{n_2} \left\{ (\gamma_2)^{-2} \ln \left( 1 + \frac{\gamma_2}{\sigma} (x_{2j} - \mu_2) \right) \times \left[ 1 - \left[ 1 + \frac{\gamma_2}{\sigma} (x_{2j} - \mu_2) \right]^{\frac{-1}{\gamma_2}} \right] - \right. \\
& - \left. \left( \frac{x_{2j} - \mu_2}{\gamma_2 \sigma} \right) \times \left[ 1 + \frac{\gamma_2}{\sigma} (x_{2j} - \mu_2) \right]^{-1} \times \left[ 1 + \gamma_2 - \left[ 1 + \frac{\gamma_2}{\sigma} (x_{2j} - \mu_2) \right]^{\frac{-1}{\gamma_2}} \right] \right\} = 0.
\end{aligned} \tag{2.29}$$

Agora derivando (2.22) com respeito a  $\sigma$ , simplificando e igualando a expressão a zero:

$$\begin{aligned}
& - \frac{n}{\sigma} + \sum_{j=1}^{n_1} \left( -\frac{1}{\gamma_1} - 1 \right) \left[ -\frac{\gamma_1(x_{1j} - \mu_1)}{\sigma^2} \right] \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-1} - \\
& - \sum_{j=1}^{n_1} \left( -\frac{1}{\gamma_1} \right) \left[ -\frac{\gamma_1(x_{1j} - \mu_1)}{\sigma^2} \right] \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-\frac{1}{\gamma_1}-1} + \\
& + \sum_{j=1}^{n_2} \left( -\frac{1}{\gamma_2} - 1 \right) \left[ -\frac{\gamma_2(x_{2j} - \mu_2)}{\sigma^2} \right] \left[ 1 + \frac{\gamma_2}{\sigma}(x_{2j} - \mu_2) \right]^{-1} - \\
& - \sum_{j=1}^{n_2} \left( -\frac{1}{\gamma_2} \right) \left[ -\frac{\gamma_2(x_{2j} - \mu_2)}{\sigma^2} \right] \left[ 1 + \frac{\gamma_2}{\sigma}(x_{2j} - \mu_2) \right]^{-\frac{1}{\gamma_2}-1} = 0.
\end{aligned}$$

Efetuada algumas manipulações algébricas obtém-se:

$$\begin{aligned}
& - n\sigma + \sum_{j=1}^{n_1} \left\{ (x_{1j} - \mu_1) \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-1} \times \left\{ 1 + \gamma_1 - \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} \right\} + \\
& + \sum_{j=1}^{n_2} \left\{ (x_{2j} - \mu_2) \left[ 1 + \frac{\gamma_2}{\sigma}(x_{2j} - \mu_2) \right]^{-1} \times \left\{ 1 + \gamma_2 - \left[ 1 + \frac{\gamma_2}{\sigma}(x_{2j} - \mu_2) \right]^{-\frac{1}{\gamma_2}} \right\} \right\} \quad (2.30)
\end{aligned}$$

**Caso 3:** ( $\gamma_1 < 0, \gamma_2 < 0; \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

Este caso é análogo ao Caso 2 e o resultado da derivada parcial do logaritmo da função de verossimilhança dada por (2.22) para os parâmetros  $p_1, \mu_i, \gamma_i$ , e  $\sigma$ ,  $i = 1, 2$ , são, respectivamente, (2.21), (2.26), (2.28), (2.27), (2.29) e (2.30).

**Caso 4:** ( $\gamma_1 > 0, \gamma_2 = 0$ )

Neste caso, o resultado da derivada parcial do logaritmo da função de verossimilhança dada por (2.22) para os parâmetros  $p_1, \mu_i$  e  $\gamma_i$ ,  $i = 1, 2$ , são respectivamente (2.21), (2.26), (2.24) e (2.27).

Ainda, derivando (2.22) com respeito a  $\sigma$ , simplificando e igualando a expressão a zero:

$$\begin{aligned}
& - n\sigma + \sum_{j=1}^{n_1} \left\{ (x_{1j} - \mu_1) \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-1} \times \left\{ 1 + \gamma_1 - \left[ 1 + \frac{\gamma_1}{\sigma}(x_{1j} - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} + \right. \\
& \left. + \sum_{j=1}^{n_2} \left\{ (x_{2j} - \mu_2) \times \left\{ 1 - \exp \left[ -\left( \frac{x_{2j} - \mu_2}{\sigma} \right) \right] \right\} \right\} \right\}
\end{aligned} \tag{2.31}$$

**Caso 5:** ( $\gamma_1 < 0, \gamma_2 = 0$ )

Para este caso, análogo ao Caso 4, o resultado da derivada parcial do logaritmo da função de verossimilhança dada por (2.22) para os parâmetros  $p_1, \mu_i, \gamma_i$  e  $\sigma, i = 1, 2$ , são respectivamente (2.21), (2.26),(2.24), (2.27) e (2.31).

**Caso 6:** ( $\gamma_1 > 0, \gamma_2 < 0; \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

Aqui também tem-se uma caso análogo ao Caso 2 e o resultado da derivada parcial do logaritmo da função de verossimilhança dada por (2.1.2) para os parâmetros  $p_1, \mu_i, \gamma_i$  e  $\sigma, i = 1, 2$ , são respectivamente (2.21), (2.26),(2.28), (2.27),(2.29) e (2.30).

## 2.2 Estudo do discriminante para mistura de GEV

Nesta seção, o estudo do discriminante não linear para pequenas amostras da mistura de duas componentes GEV está de acordo com os trabalhos Ahmad e Abd-Elrahman (1994), Ahmad *et al.* (2010) e Sultan e Al-Moisheer (2013) conforme foi apresentado na Seção 1.4. São tratados os casos de função discriminante ótima (o), quando todos os parâmetros do modelo de mistura são conhecidos e os casos estimados para amostra classificada (c) e não classificada (m).

### 2.2.1 Função discriminante mistura de GEV

Observa-se que a probabilidade dada na equação (1.10) coincide com a expressão de  $w_{ij}$ , definida na Seção 2.1.1. Ambas as expressões representam a probabilidade da observação  $x_j$  pertencer a  $i$ -ésima componente da mistura. Logo, também pode-se relacionar  $w_{ij}$  com a função logística, conforme equações (1.11) e (1.12). Assim,

nesta seção determina-se as funções discriminantes (1.14) a partir de  $w_{ij}$ , para os seis possíveis casos descritos na Seção 2.1.

**Caso 1:** ( $\gamma_1 = 0, \gamma_2 = 0$ )

Em geral, a probabilidade  $w_{1j}$  é:

$$w_{1j} = \frac{p_1 g_1(x_j; \theta_1)}{Q(x_j)} = \left\{ \frac{p_1 g_1(x_j; \theta_1) + p_2 g_2(x_j; \theta_2)}{p_1 g_1(x_j; \theta_1)} \right\}^{-1}. \quad (2.32)$$

Substituindo (2.2) em (2.32), tem-se:

$$w_{1j} = \left\{ 1 + \frac{p_2 \exp \left\{ - \left( \frac{x_j - \mu_2}{\sigma} \right) - \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] \right\}}{p_1 \exp \left\{ - \left( \frac{x_j - \mu_1}{\sigma} \right) - \exp \left[ - \left( \frac{x_j - \mu_1}{\sigma} \right) \right] \right\}} \right\}^{-1}.$$

Equivalentemente,

$$w_{1j} = \left\{ 1 + \frac{p_2}{p_1} \exp \left\{ \frac{\mu_2}{\sigma} - \frac{\mu_1}{\sigma} - \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] + \exp \left[ - \left( \frac{x_j - \mu_1}{\sigma} \right) \right] \right\} \right\}^{-1}. \quad (2.33)$$

Para obter uma equação do tipo (1.11), após algumas manipulações algébricas em (2.33) tem-se:

$$w_{1j} = \left\{ 1 + \exp \left\{ \ln \left( \frac{p_2}{p_1} \right) + \frac{\mu_2}{\sigma} - \frac{\mu_1}{\sigma} - \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] + \exp \left[ - \left( \frac{x_j - \mu_1}{\sigma} \right) \right] \right\} \right\}^{-1}. \quad (2.34)$$

Assim, pode-se reescrever (2.34) como:

$$w_{1j} = \{1 + \exp \{a_1 - b_1 z_1\}\}^{-1},$$

e

$$NL(x) = a_1 - b_1 z_1, \quad (2.35)$$

sendo

$$a_1 = \ln \left( \frac{p_2}{p_1} \right) + \frac{\mu_2}{\sigma} - \frac{\mu_1}{\sigma};$$

$$b_1 = 1;$$

$$z_1 = \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] - \exp \left[ - \left( \frac{x_j - \mu_1}{\sigma} \right) \right] \Bigg\}.$$

Esses resultados coincidem com encontrados por Ahmad *et al.* (2010). Então,  $x$  será classificado em  $\pi_1$  se  $NL(x) < 0$  e em  $\pi_2$  se  $NL(x) \geq 0$ .

Se todos os parâmetros do modelo de mistura são conhecidos trata-se da função de discriminante ótima.

Para estimar (2.35) pode-se utilizar as estimativas da amostra não classificada  $\hat{p}_1$ ,  $\hat{\mu}_i$  e  $\hat{\sigma}$ ,  $i = 1, 2$ , obtidas resolvendo numericamente o sistema de equações formado por (2.5), (2.6) e (2.7).

Também há opção de aplicar as estimativas da amostra classificada  $\tilde{p}_1$ ,  $\tilde{\mu}_1$ ,  $\tilde{\mu}_2$  e  $\tilde{\sigma}$  que são calculadas via algum método numérico a partir das equações (2.21), (2.23), (2.24) e (2.25).

**Caso 2:** ( $\gamma_1 > 0, \gamma_2 > 0, \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

A probabilidade  $w_{1j}$ , para  $x_j \geq \mu_2 - \frac{\sigma}{\gamma_2}$ , e dada por :

$$w_{1j} = \frac{p_1 g_1(x_j; \theta_1)}{Q(x_j)} = \left\{ \frac{p_1 g_1(x_j; \theta_1) + p_2 g_2(x_j; \theta_2)}{p_1 g_1(x_j; \theta_1)} \right\}^{-1}. \quad (2.36)$$

Ao substituir (2.2) em (2.36), obtém-se:

$$w_{1j} = \left\{ \frac{p_2 \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}-1} \exp \left\{ - \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}} \right\}}{p_1 \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}-1} \exp \left\{ - \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\}} \right\}^{-1}. \quad (2.37)$$

Com algumas manipulações algébricas em (2.37) apresenta-se:

$$\begin{aligned} w_{1j} = & \left\{ 1 + \exp \left\{ \ln \left( \frac{p_2}{p_1} \right) - \left( \frac{1}{\gamma_2} + 1 \right) \ln \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right] + \right. \right. \\ & + \left. \left( \frac{1}{\gamma_1} + 1 \right) \ln \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right] - \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}} + \right. \\ & \left. \left. + \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} \right\}^{-1}. \end{aligned}$$

Equivalentemente,

$$\begin{aligned} w_{1j} = & \left\{ 1 + \exp \left\{ \ln \left( \frac{p_2}{p_1} \right) - \left( \frac{1}{\gamma_2} + 1 \right) \ln \left[ \frac{\sigma + \gamma_2(x_j - \mu_2)}{\sigma} \right] + \right. \right. \\ & + \left. \left( \frac{1}{\gamma_1} + 1 \right) \ln \left[ \frac{\sigma + \gamma_1(x_j - \mu_1)}{\sigma} \right] - \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}} + \right. \\ & \left. \left. + \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} \right\}^{-1}. \end{aligned}$$

Novamente, com interesse de encontrar uma equação do tipo (1.11) são efetuados cálculos para obter:

$$w_{1j} = \left\{ 1 + \exp \left\{ \ln \left( \frac{p_2}{p_1} \right) + \left( \frac{1}{\gamma_2} - \frac{1}{\gamma_1} \right) \ln(\sigma) - \left( \frac{1}{\gamma_2} + 1 \right) \ln [\sigma + \gamma_2(x_j - \mu_2)] + \right. \right. \\ \left. \left. + \left( \frac{1}{\gamma_1} + 1 \right) \ln [\sigma + \gamma_1(x_j - \mu_1)] - \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}} + \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} \right\}^{-1}.$$

Logo, como  $w_{ij}$  é dada por

$$w_{1j}(x_j) = \begin{cases} 1, & \mu_1 - \frac{\sigma}{\gamma_1} \leq x_j < \mu_2 - \frac{\sigma}{\gamma_2}, \\ \{1 + \exp \{a_2 - b_2 z_2\}\}^{-1}, & x_j \geq \mu_2 - \frac{\sigma}{\gamma_2}, \end{cases},$$

e

$$NL(x) = a_2 - b_2 z_2, \quad (2.38)$$

sendo

$$a_2 = \ln \left( \frac{p_2}{p_1} \right) + \left( \frac{1}{\gamma_2} - \frac{1}{\gamma_1} \right) \ln(\sigma);$$

$$b_2 = 1;$$

$$z_2 = \left( \frac{1}{\gamma_2} + 1 \right) \ln [\sigma + \gamma_2(x_j - \mu_2)] - \left( \frac{1}{\gamma_1} + 1 \right) \ln [\sigma + \gamma_1(x_j - \mu_1)] + \\ + \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}} - \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \Big\}.$$

Destaca-se que a função discriminante não precisa ser avaliada quando  $x$  pertence ao intervalo  $\left[ \mu_1 - \frac{\sigma}{\gamma_1}; \mu_2 - \frac{\sigma}{\gamma_2} \right)$ , pois, com probabilidade um,  $x$  pertence a primeira população. Sendo assim, para  $x \geq \mu_2 - \frac{\sigma}{\gamma_2}$ , tem-se que  $x$  pertence a  $\pi_1$  se  $NL(x) < 0$  e a  $\pi_2$  se  $NL(x) \geq 0$ .

Para obter a estimativa de (2.38) utiliza-se das estimativas da amostra não classificada  $\hat{p}_1$ ,  $\hat{\mu}_i$ ,  $\hat{\gamma}_i$  e  $\hat{\sigma}$ ,  $i = 1, 2$ , obtidas resolvendo numericamente o sistema de equações formado por (2.5), (2.12), (2.14) e (2.13).

Além disso, há como alternativa as estimativas da amostra classificada  $\tilde{p}_1$ ,  $\tilde{\mu}_1, \tilde{\gamma}_1$ ,  $\tilde{\mu}_2$ ,  $\tilde{\gamma}_2$  e  $\tilde{\sigma}$  que são calculadas via algum método numérico a partir das equações (2.21), (2.26), (2.27), (2.28), (2.29) e (2.30).

Quando todos os parâmetros do modelo de mistura são conhecidos trata-se da função de discriminante ótima.

**Caso 3:**  $\left(\gamma_1 < 0, \gamma_2 < 0; \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}\right)$

Neste caso  $w_{1j}$  é dada por

$$w_{1j}(x_j) = \begin{cases} 0, & \mu_1 - \frac{\sigma}{\gamma_1} < x_j \leq \mu_2 - \frac{\sigma}{\gamma_2}, \\ \{1 + \exp\{a_3 - b_3 z_3\}\}^{-1}, & x_j \leq \mu_1 - \frac{\sigma}{\gamma_1}, \end{cases},$$

e

$$NL(x) = a_3 - b_3 z_3, \quad (2.39)$$

sendo

$$a_3 = \ln\left(\frac{p_2}{p_1}\right) + \left(\frac{1}{\gamma_2} - \frac{1}{\gamma_1}\right) \ln(\sigma);$$

$$b_3 = 1;$$

$$z_3 = \left(\frac{1}{\gamma_2} + 1\right) \ln[\sigma + \gamma_2(x_j - \mu_2)] - \left(\frac{1}{\gamma_1} + 1\right) \ln[\sigma + \gamma_1(x_j - \mu_1)] + \left[1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2)\right]^{-\frac{1}{\gamma_2}} - \left[1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1)\right]^{-\frac{1}{\gamma_1}} \Bigg\}.$$

É importante salientar que a função discriminante não precisa ser avaliada para  $\mu_1 - \frac{\sigma}{\gamma_1} < x \leq \mu_2 - \frac{\sigma}{\gamma_2}$ , já que a probabilidade assume zero, portanto  $x$  pertence a

segunda população. À vista disso,  $x \leq \mu_1 - \frac{\sigma}{\gamma_1}$  pertence a  $\pi_1$  se  $NL(x) < 0$  e  $\pi_2$  se  $NL(x) \geq 0$ .

Do mesmo modo, para estimar (2.39) emprega-se as estimativas da amostra não classificada  $\hat{p}_1, \hat{\mu}_i, \hat{\gamma}_i$  e  $\hat{\sigma}$ ,  $i = 1, 2$ , obtidas resolvendo numericamente o sistema de equações formado por (2.5), (2.12), (2.14) e (2.13).

Ainda há possibilidade de usar as estimativas da amostra classificada  $\tilde{p}_1, \tilde{\mu}_1, \tilde{\gamma}_1, \tilde{\mu}_2, \tilde{\gamma}_2$  e  $\tilde{\sigma}$  que são calculadas via algum método numérico a partir das equações (2.21), (2.26), (2.27), (2.28), (2.29) e (2.30). E se os parâmetros do modelo de mistura são conhecidos trata-se da função de discriminante ótima.

**Caso 4:** ( $\gamma_1 > 0, \gamma_2 = 0$ )

A probabilidade  $w_{1j}$ , para  $x_j \geq \mu_1 - \frac{\sigma}{\gamma_1}$ , pode ser escrita como:

$$w_{1j} = \frac{p_1 g_1(x_j; \theta_1)}{Q(x_j)} = \left\{ \frac{p_1 g_1(x_j; \theta_1) + p_2 g_2(x_j; \theta_2)}{p_1 g_1(x_j; \theta_1)} \right\}^{-1}. \quad (2.40)$$

Ao fazer a substituição de (2.2) em (2.40), tem-se:

$$w_{1j} = \left\{ 1 + \frac{p_2 \exp \left\{ - \left( \frac{x_j - \mu_2}{\sigma} \right) - \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] \right\}}{p_1 \left[ 1 + \frac{\gamma_1}{\sigma} (x_j - \mu_1) \right]^{-\frac{1}{\gamma_1} - 1} \exp \left\{ - \left[ 1 + \frac{\gamma_1}{\sigma} (x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\}} \right\}^{-1}. \quad (2.41)$$

Efetuada algumas manipulações algébricas em (2.41), obtém-se:

$$w_{1j} = \left\{ 1 + \exp \left\{ \ln \left( \frac{p_2}{p_1} \right) + \left( \frac{1}{\gamma_1} + 1 \right) \ln \left[ \frac{\sigma + \gamma_1 (x_j - \mu_1)}{\sigma} \right] - \left( \frac{x_j - \mu_2}{\sigma} \right) - \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] + \left[ 1 + \frac{\gamma_1}{\sigma} (x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \right\} \right\}^{-1}. \quad (2.42)$$

Com objetivo de obter uma equação do tipo (1.11) são realizadas algumas simplificações matemáticas e (2.42) é reescrita por

$$w_{1j} = \left\{ 1 + \exp \left\{ \ln \left( \frac{p_2}{p_1} \right) - \left( \frac{1}{\gamma_1} + 1 \right) \ln(\sigma) + \frac{\mu_2}{\sigma} - \right. \right. \\ \left. \left. - \left\{ \frac{x_j}{\sigma} + \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] - \left[ 1 + \frac{\gamma_1}{\sigma} (x_j + \mu_1) \right]^{-\frac{1}{\gamma_1}} - \right. \right. \right. \\ \left. \left. \left. - \left( \frac{1}{\gamma_1} + 1 \right) \ln \left[ \sigma + \gamma_1 (x_j - \mu_1) \right] \right\} \right\}^{-1} .$$

Desta maneira, como  $w_{ij}$  é dada por

$$w_{1j}(x_j) = \begin{cases} \{1 + \exp \{a_4 - b_4 z_4\}\}^{-1}, & x_j \geq \mu_1 - \frac{\sigma}{\gamma_1}, \\ 0, & x_j < \mu_1 - \frac{\sigma}{\gamma_1}, \end{cases} .$$

e

$$NL(x) = a_4 - b_4 z_4, \quad (2.43)$$

com

$$a_4 = \ln \left( \frac{p_2}{p_1} \right) - \left( \frac{1}{\gamma_1} + 1 \right) \ln(\sigma) + \frac{\mu_2}{\sigma};$$

$$b_4 = 1;$$

$$z_4 = \frac{x_j}{\sigma} + \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] - \left( \frac{1}{\gamma_1} + 1 \right) \ln \left[ \sigma + \gamma_1 (x_j - \mu_1) \right] - \left[ 1 + \frac{\gamma_1}{\sigma} (x_j + \mu_1) \right]^{-\frac{1}{\gamma_1}} .$$

Para  $x < \mu_1 - \frac{\sigma}{\gamma_1}$ , a probabilidade assume o valor zero, ou seja,  $x$  pertence a segunda população, portanto a função discriminante não precisa ser avaliada nesse intervalo. Por isto,  $x \geq \mu_1 - \frac{\sigma}{\gamma_1}$  pertence a  $\pi_1$  se  $NL(x) < 0$  e a  $\pi_2$  se  $NL(x) \geq 0$ .

A estimação de (2.43) pode ser obtida por meio das estimativas da amostra não classificada  $\hat{p}_1, \hat{\mu}_1, \hat{\gamma}_1, \hat{\mu}_2$  e  $\hat{\sigma}$  obtidas resolvendo numericamente o sistema de equações formado por (2.5), (2.12), (2.14), (2.6) e (2.7).

Outra escolha seria optar pela estimativas da amostra classificada  $\tilde{p}_1, \tilde{\mu}_1, \tilde{\gamma}_1, \tilde{\mu}_2$  e  $\tilde{\sigma}$  que são calculadas via algum método numérico a partir das equações (2.21), (2.26), (2.27), (2.24) e (2.30).

Para o caso de todos os parâmetros do modelo de mistura serem conhecidos trata-se da função de discriminante ótima.

**Caso 5:** ( $\gamma_1 < 0, \gamma_2 = 0$ )

Ao considerar

$$w_{1j}(x_j) = \begin{cases} \{1 + \exp \{a_5 - b_5 z_5\}\}^{-1}, & x_j \leq \mu_1 - \frac{\sigma}{\gamma_1}, \\ 0, & x_j > \mu_1 - \frac{\sigma}{\gamma_1}, \end{cases}.$$

obtem-se

$$NL(x) = a_5 - b_5 z_5, \quad (2.44)$$

com

$$a_5 = \ln \left( \frac{p_2}{p_1} \right) - \left( \frac{1}{\gamma_1} + 1 \right) \ln(\sigma) + \frac{\mu_2}{\sigma},$$

$$b_5 = 1,$$

e

$$z_5 = \frac{x_j}{\sigma} + \exp \left[ - \left( \frac{x_j - \mu_2}{\sigma} \right) \right] - \left( \frac{1}{\gamma_1} + 1 \right) \ln \left[ \sigma + \gamma_1 (x_j - \mu_1) \right] - \left[ 1 + \frac{\gamma_1}{\sigma} (x_j + \mu_1) \right]^{-\frac{1}{\gamma_1}}.$$

Se  $x > \mu_1 - \frac{\sigma}{\gamma_1}$ , a probabilidade assume zero, logo  $x$  pertence a segunda população e a função discriminante não precisa ser avaliada nesse intervalo. Desta maneira, tem-se que  $x \leq \mu_1 - \frac{\sigma}{\gamma_1}$  pertence a  $\pi_1$  se  $NL(x) < 0$  e a  $\pi_2$  se  $NL(x) \geq 0$ .

Semelhantemente, para estimar (2.44) pode-se utilizar as estimativas da amostra

não classificada  $\hat{p}_1, \hat{\mu}_1, \hat{\gamma}_1, \hat{\mu}_2$  e  $\hat{\sigma}$  obtidas resolvendo numericamente o sistema de equações formado por (2.5), (2.12), (2.14), (2.6) e (2.13).

Ainda, há a opção de aplicar as estimativas da amostra classificada  $\tilde{p}_1, \tilde{\mu}_1, \tilde{\gamma}_1, \tilde{\mu}_2$  e  $\tilde{\sigma}$  que são calculadas via algum método numérico a partir das equações (2.21), (2.26), (2.27), (2.24) e (2.30).

Quando todos os parâmetros do modelo de mistura serem conhecidos tem-se da função de discriminante ótima.

**Caso 6:** ( $\gamma_1 > 0, \gamma_2 < 0; \mu_1 - \frac{\sigma}{\gamma_1} < \mu_2 - \frac{\sigma}{\gamma_2}$ )

Finalmente, para a mistura de componentes com maior suporte em lados opostos, tem -se

$$w_{1j}(x_j) = \begin{cases} 0, & x_j < \mu_1 - \frac{\sigma}{\gamma_1} \\ \{1 + \exp \{a_6 - b_6 z_6\}\}^{-1}, & \mu_1 - \frac{\sigma}{\gamma_1} \leq x_j \leq \mu_2 - \frac{\sigma}{\gamma_2}, \\ 1, & x_j > \mu_2 - \frac{\sigma}{\gamma_2}, \end{cases} .$$

e

$$NL(x) = a_6 - b_6 z_6, \quad (2.45)$$

sendo

$$a_6 = \ln \left( \frac{p_2}{p_1} \right) + \left( \frac{1}{\gamma_2} - \frac{1}{\gamma_1} \right) \ln(\sigma);$$

$$b_6 = 1;$$

e

$$z_6 = \left( \frac{1}{\gamma_2} + 1 \right) \ln [\sigma + \gamma_2(x_j - \mu_2)] - \left( \frac{1}{\gamma_1} + 1 \right) \ln [\sigma + \gamma_1(x_j - \mu_1)] + \left[ 1 + \frac{\gamma_2}{\sigma}(x_j - \mu_2) \right]^{-\frac{1}{\gamma_2}} - \left[ 1 + \frac{\gamma_1}{\sigma}(x_j - \mu_1) \right]^{-\frac{1}{\gamma_1}} \Bigg\}.$$

Note que a probabilidade assume zero no intervalo  $x < \mu_1 - \frac{\sigma}{\gamma_1}$ , isso implica que  $x$  pertence a segunda população. Já no intervalo,  $x > \mu_2 - \frac{\sigma}{\gamma_2}$ , a probabilidade assume um, ou seja,  $x$  pertence a primeira população. Portanto, a função discriminante é avaliada apenas no intervalo,  $\mu_1 - \frac{\sigma}{\gamma_1} \leq x \leq \mu_2 - \frac{\sigma}{\gamma_2}$ , logo  $x$  pertence a  $\pi_1$  se  $NL(x) < 0$  e a  $\pi_2$  se  $NL(x) \geq 0$ .

Para obter a estimativa (2.45) pode-se aplicar as estimativas da amostra não classificada  $\hat{p}_1, \hat{\mu}_1, \hat{\gamma}_1, \hat{\mu}_2, \hat{\gamma}_2$  e  $\hat{\sigma}_2$  obtidas resolvendo numericamente o sistema de equações formado por (2.5), (2.12), (2.14) e (2.13) .

Por outro lado, pode-se empregar também as estimativas da amostra classificada  $\tilde{p}_1, \tilde{\mu}_1, \tilde{\gamma}_1, \tilde{\mu}_2, \tilde{\gamma}_2$  e  $\tilde{\sigma}$  que são calculadas via algum método numérico a partir das equações (2.21), (2.26), (2.27), (2.28), (2.29) e (2.30).

Se são conhecidos todos os parâmetros do modelo de mistura tem-se da função de discriminante ótima.

## 2.3 Erros de Classificação

Os erros de classificação são obtidos analiticamente, conforme descrito na seção 1.4.2, apenas para o Caso 1. Para os demais casos não é possível calcular os erros de classificação de forma analítica. Desta maneira, apresenta-se uma maneira empírica para calcular os erros de classificação para os demais casos e que também será para o Caso 1.

Destaca-se que a função discriminante não linear é definida para os casos: ótimo (o), quando utiliza-se os parâmetros, estimação por amostra não classificada (m) e estimação por amostra classificada(c).

### 2.3.1 Forma analítica: Caso 1 ( $\gamma_1 = 0, \gamma_2 = 0$ )(ver Ahmad *et al.* (2010)).

Tem-se que a função discriminante não linear é dada por:

$$\begin{aligned}
 NL_k(x) &= \ln\left(\frac{p_2}{p_1}\right) + \frac{\mu_2 - \mu_1}{\sigma} - \exp\left[-\left(\frac{x - \mu_2}{\sigma}\right)\right] + \exp\left[-\left(\frac{x - \mu_1}{\sigma}\right)\right] \Big\} \\
 &= \ln\left(\frac{p_2}{p_1}\right) + \frac{\mu_2 - \mu_1}{\sigma} - \exp\left(-\frac{x}{\sigma}\right) \times \exp\left(\frac{\mu_2}{\sigma}\right) + \exp\left(-\frac{x}{\sigma}\right) \times \exp\left(\frac{\mu_1}{\sigma}\right) \\
 &= \ln\left(\frac{p_2}{p_1}\right) + \frac{\mu_2 - \mu_1}{\sigma} - \exp\left(-\frac{x}{\sigma}\right) \times \left\{\exp\left(\frac{\mu_2}{\sigma}\right) - \exp\left(\frac{\mu_1}{\sigma}\right)\right\}.
 \end{aligned}$$

Desta maneira, pode-se escrever a função discriminante não linear como:

$$NL_k(x) = A_k - \exp\left(-\frac{x}{\sigma}\right) B_k,$$

sendo que

$$A_k = \ln\left(\frac{p_2}{p_1}\right) + \frac{\mu_2 - \mu_1}{\sigma},$$

e

$$B_k = \exp\left(\frac{\mu_2}{\sigma}\right) - \exp\left(\frac{\mu_1}{\sigma}\right)$$

A partir dessas considerações pode-se calcular os erros de classificação. Inicia-se erro  $e_{1k}$ ,  $k = o, m, c$ :

$$e_{1k} = P\left(A_k - B_k \exp\left(-\frac{x}{\sigma}\right) > 0 \mid x \in \pi_1\right).$$

Utilizando algumas operações algébricas, tem-se

$$\begin{aligned}
e_{1j} &= P\left(x > \sigma \ln\left(\frac{B_k}{A_k}\right) \mid x \in \pi_1\right) \\
e_{1j} &= \begin{cases} \bar{F}\left(\sigma \ln\left(\frac{B_k}{A_k}\right)\right), & \mu_2 \geq \mu_1 \text{ e } p_1 \leq 0.5 \\ F\left(\sigma \ln\left(\frac{B_k}{A_k}\right)\right), & \mu_2 < \mu_1 \text{ e } p_1 > 0.5 \end{cases}, \quad (2.46)
\end{aligned}$$

$F(\cdot)$  é a função de distribuição acumulada da GEV. Analogamente,  $e_{2k}$ , em que  $k = o, m, c$ , é dado por:

$$e_{2j} = \begin{cases} F\left(\sigma \ln\left(\frac{B}{A}\right)\right), & \mu_2 \geq \mu_1 \text{ e } p_1 \leq 0.5 \\ \bar{F}\left(\sigma \ln\left(\frac{B}{A}\right)\right), & \mu_2 < \mu_1 \text{ e } p_1 > 0.5 \end{cases}. \quad (2.47)$$

Ainda o erro de classificação global ponderado pela proporção de mistura,  $e_k$  é dado por

$$e_k = p_1 e_{1k} + p_2 e_{2k}. \quad (2.48)$$

De modo que, para estimar os valores  $e_{1k}$ ,  $e_{2k}$  e  $e_k$ , para  $k = m, c$  pode-se simplesmente substituir os parâmetros das equações (2.46), (2.47) e (2.48) por suas respectivas estimativas.

### 2.3.2 Forma empírica

Para os demais casos não é possível obter os erros de classificação de maneira analítica, esses são, portanto, calculados empiricamente.

Seja  $X$  uma variável aleatória com função de densidade de probabilidade  $h(x; \Theta)$  dada em (2.1). Com uma amostra aleatória de tamanho  $n$  de  $X$ ,  $X_1, X_2, \dots, X_n$ , pode-se calcular empiricamente  $e_{1k}$  e  $e_{2k}$ ,  $k = o, m, c$ :

$$\widehat{e}_{1k} = \frac{1}{n} \sum_{j=1}^n I_{(NL_k(x_j) \leq 0 \mid x_j \in \pi_1) > 0}(x_j), \quad (2.49)$$

e

$$\widehat{e}_{2k} = \frac{1}{n} \sum_{j=1}^n I_{(NL_k(x_j) < 0 | x_j \in \pi_2)}(x_j). \quad (2.50)$$

Sendo que

$$I_A(x) = \begin{cases} 1, & \text{se } x \in A \\ 0, & \text{se } x \notin A \end{cases}. \quad (2.51)$$

O erro de classificação global ponderado pela proporção de mistura,  $\widehat{e}_k$  é dado por

$$\widehat{e}_k = p_1 \widehat{e}_{1k} + p_2 \widehat{e}_{2k}. \quad (2.52)$$

# Capítulo 3

## Simulações e Aplicação

### 3.1 Simulações

Com o interesse de mostrar a utilidade das técnicas propostas neste trabalho, são realizados experimentos de simulação. Os principais objetivos deste capítulo são:

- Testar o comportamento dos estimadores de máxima verossimilhança para os casos de amostra não classificada (m) e amostra classificada (c).
- Investigar o desempenho da função discriminante não linear  $NL_i(x)$ ,  $i=m, c$  e o, comparando os casos de amostra não classificada (m), amostra classificada (c) e ótimo (o) por meio do erro de classificação.

Para obter os estimadores de máxima verossimilhança do modelo de mistura (2.1) são utilizadas amostras classificadas e não classificadas. Como nas Seções 2.1.1 e 2.1.2 a estimação dos parâmetros contempla os seis casos possíveis para mistura de GEV, são escolhidos doze conjuntos de parâmetros de tal forma a considerar dois conjuntos para cada caso. Na Tabela 3.1, são apresentados esses parâmetros. Para  $j = 1, 2$ ,  $\Theta_{1,j}$ ,  $\Theta_{2,j}$ ,  $\Theta_{3,j}$ ,  $\Theta_{4,j}$ ,  $\Theta_{5,j}$  e  $\Theta_{6,j}$  corresponde aos casos (1), (2), (3), (4), (5) e (6), respectivamente.

Tabela 3.1: Parâmetros para simulações

$\Theta$	$p_1$	$p_2$	$\gamma_1$	$\gamma_2$	$\sigma$	$\mu_1$	$\mu_2$
$\Theta_{1,1}$	0.6	0.4	0	0	1	6	8
$\Theta_{1,2}$	0.6	0.4	0	0	1	1	4
$\Theta_{2,1}$	0.5	0.5	0.5	1	1	1	3
$\Theta_{2,2}$	0.5	0.5	1	1	1	1	2
$\Theta_{3,1}$	0.4	0.6	-0.5	-0.5	1	-1	1
$\Theta_{3,2}$	0.4	0.6	-0.5	-0.3	1	0	2
$\Theta_{4,1}$	0.6	0.4	0.5	0	1	1	3
$\Theta_{4,2}$	0.6	0.4	1	0	1	1	1
$\Theta_{5,1}$	0.5	0.5	-0.5	0	1	2	0
$\Theta_{5,2}$	0.5	0.5	-0.5	0	1	1	3
$\Theta_{6,1}$	0.4	0.6	1	-0.5	1	1	0
$\Theta_{6,2}$	0.4	0.6	0.5	-0.5	1	-1	1

Para a estimação utiliza-se a função `newtonRaphson(.)` o software R versão 3.3.1. Para o caso de amostra não classificada, segue-se os seguintes procedimento:

1. Gera-se uma amostra dos tamanhos  $n = 100$  e  $n = 50$  para cada escolha do vetor  $\Theta$ .
2. A amostra aleatória da variável  $X$  cuja densidade é a mistura (2.2) é gerada da seguinte maneira:
  - (a) Gerar duas variáveis uniformes  $u_1$  e  $u_2$  no intervalo  $(0, 1)$ ;
  - (b) Se  $u_1 < p_1$ , então usa-se  $u_2$  para gerar um valor de  $x$  da variável aleatória  $X$ , sendo que  $x = F_1^{-1}(u_2)$  e  $F_1$  é a distribuição acumulada de  $f_1$  dada por (2.2);
  - (c) Se  $u_1 \geq p_1$ , então usa-se  $u_2$  para gerar um valor de  $x$  da variável aleatória  $X$ , sendo que  $x = F_2^{-1}(u_2)$  e  $F_2$  é a distribuição acumulada de  $f_2$  dada por (2.2).
3. Calcula-se interativamente os estimadores de  $\Theta$  utilizando as expressões dadas por (2.9)-(2.11) e (2.17)-(2.19).
4. Obtém-se estimativas de  $\Theta$  utilizando 100 amostras dos tamanhos  $n = 100$  e  $n = 50$ , então calcula-se a média e o Erro do Quadrado Médio (EQM) de  $\Theta$  por Monte Carlo.

Os resultados das estimações e do EQM para o caso de amostra não classificado (m) são apresentados nas Tabelas 3.2 e 3.3, respectivamente.

Para o caso de amostra classificada (c) o procedimento é parecido com o caso anterior, a diferença que os valores de  $n_1$  e  $n_2$ , para  $n_1 + n_2 = 100$ , já estão fixados e no passo (3) utiliza-se as Equações (2.26)-(2.30) e (2.27)-(2.29). Os resultados das estimações e do EQM são apresentados nas Tabelas 3.4 e 3.5, respectivamente.

Com base nos resultados apresentados na Tabela 3.2 nota-se que as médias das estimativas para amostra não classificado estão próximas aos seus verdadeiros valores. Ainda é possível verificar uma pequena tendência de sobrestimar ou subestimar os parâmetros  $\gamma_i$ ,  $i = 1, 2$ . Na maior parte dos casos, a sobrestimação ocorre quando o valor de  $\gamma_i$  ( $i = 1, 2$ ) é positivo, como nos casos 2, 4 e 6 (apenas primeira componente). Já a subestimação é verificada quando o valor para  $\gamma_i$  ( $i = 1, 2$ ) é negativo, como nos casos 3, 5 e 6 (segunda componente). Apenas em três conjuntos de parâmetros  $\Theta_{3,1}$ ,  $\Theta_{5,1}$  e  $\Theta_{6,2}$ , que representa os casos 3, 5 e 6 respectivamente, para amostra de tamanho 50, demandam maior atenção para esse problema.

A Tabela 3.3 certifica os bons resultados alcançados pelo algoritmo, pois, verifica-se que o EQM é pequeno para todos os casos. As estimativas de  $p_i$ ,  $\sigma$  e  $\mu_i$  são mais precisas do que as de  $\gamma_i$  por apresentarem valores de EQM menores, para  $i = 1, 2$ . Os gráficos *boxplot* de cada parâmetro, de acordo com a Tabela (3.1), para amostra de tamanho 100 são apresentados no Apêndice. Para construir esses gráficos utiliza-se as 100 estimativas de cada parâmetro do Método de Monte Carlo, possibilitando compreender melhor porque as estimativas de  $\gamma_i$ ,  $i = 1, 2$ , apresenta resultados piores que dos outros parâmetros.

Também, nota-se que os valores do EQM para todos os parâmetros aumentam para amostra de tamanho menor, entretanto essas diferenças são pequenas, o que mostra que há pouca alteração no resultado para amostra menor.

Tabela 3.2: Média de  $\hat{\Theta}$ 

$\hat{\Theta}$	$n$	$\hat{p}_1$	$\hat{p}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
$\hat{\Theta}_{1,1}$	50	0.6032	0.3968	-	-	1.0020	6.0052	8.0057
$\hat{\Theta}_{1,1}$	100	0.5976	0.4024	-	-	1.0057	5.9978	8.0126
$\hat{\Theta}_{1,2}$	50	0.6044	0.3956	-	-	1.0003	1.0051	4.0103
$\hat{\Theta}_{1,2}$	100	0.5992	0.4008	-	-	1.0065	0.9980	4.0132
$\hat{\Theta}_{2,1}$	50	0.5081	0.4919	0.5707	1.0626	0.9904	1.0160	3.0268
$\hat{\Theta}_{2,1}$	100	0.5007	0.4993	0.5318	1.0065	0.9936	1.0099	3.0064
$\hat{\Theta}_{2,2}$	50	0.5045	0.4955	1.1246	1.0180	0.9974	1.0259	2.0972
$\hat{\Theta}_{2,2}$	100	0.5024	0.4976	1.0659	1.0098	0.9969	1.0191	2.0051
$\hat{\Theta}_{3,1}$	50	0.3952	0.6048	-0.5064	-0.8226	0.9985	-1.0164	1.0018
$\hat{\Theta}_{3,1}$	100	0.3945	0.6055	-0.5047	-0.5377	0.9919	-0.9943	0.9822
$\hat{\Theta}_{3,2}$	50	0.3957	0.6043	-0.5812	-0.4882	1.0044	-0.0117	2.0209
$\hat{\Theta}_{3,2}$	100	0.3934	0.6066	-0.5071	-0.3557	0.9959	0.0014	2.0061
$\hat{\Theta}_{4,1}$	50	0.6036	0.3964	0.5430	-	0.8045	1.0234	3.0114
$\hat{\Theta}_{4,1}$	100	0.5986	0.4014	0.5211	-	0.8251	1.0086	3.0176
$\hat{\Theta}_{4,2}$	50	0.6021	0.3979	1.0487	-	0.9900	1.0164	1.0306
$\hat{\Theta}_{4,2}$	100	0.6018	0.3982	1.0027	-	1.0002	1.0040	1.0323
$\hat{\Theta}_{5,1}$	50	0.5081	0.4919	-0.7388	-	0.9921	1.9818	0.0585
$\hat{\Theta}_{5,1}$	100	0.5019	0.4981	-0.5918	-	0.9956	1.9918	0.0269
$\hat{\Theta}_{5,2}$	50	0.5048	0.4952	-0.5043	-	1.0070	0.9892	3.0025
$\hat{\Theta}_{5,2}$	100	0.5006	0.4994	-0.5087	-	1.0024	0.9934	3.0055
$\hat{\Theta}_{6,1}$	50	0.4015	0.5985	0.9708	-0.5801	1.0009	1.0070	-0.0011
$\hat{\Theta}_{6,1}$	100	0.3977	0.6023	0.9898	-0.5102	1.0005	0.9992	-0.0084
$\hat{\Theta}_{6,2}$	50	0.4002	0.5998	0.7153	-0.6239	0.9920	-0.9607	0.9900
$\hat{\Theta}_{6,2}$	100	0.3968	0.6032	0.5153	-0.5408	0.9984	-0.9852	1.0038

A Tabela 3.4 apresenta o resultado da estimação para amostra classificada, também, são obtidas médias para as estimativas próximas aos seus verdadeiros valores. Assim como na amostra não classificada, foi verificado que os valores de  $\gamma_1$  e  $\gamma_2$  algumas vezes são sobrestimados e outras subestimados. Esse problema destaca-se para os conjuntos de parâmetros  $\Theta_{3,1}$ ,  $\Theta_{3,2}$ ,  $\Theta_{5,1}$ , e  $\Theta_{5,2}$ , que representa os casos 3 e 5, para amostra de tamanho 50. Desta maneira, nota-se que a maior dificuldade está em estimar o parâmetro  $\gamma_1$  com valor negativo.

A Tabela 3.5 representa os valores do EQM da estimação para amostra classificada, que aponta valores pequenos para todos os casos. Também, como ocorreu

Tabela 3.3: Erro quadrático médio de  $\hat{\Theta}$ 

$\hat{\Theta}$	$n$	$\hat{p}_1$	$\hat{p}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
$\hat{\Theta}_{1,1}$	50	0.0536	0.0536	-	-	0.0774	0.1749	0.1516
$\hat{\Theta}_{1,1}$	100	0.0303	0.0303	-	-	0.0618	0.1351	0.0836
$\hat{\Theta}_{1,2}$	50	0.0643	0.0643	-	-	0.0860	0.1749	0.1610
$\hat{\Theta}_{1,2}$	100	0.0394	0.0394	-	-	0.0634	0.1371	0.1064
$\hat{\Theta}_{2,1}$	50	0.0555	0.0555	0.1881	0.3407	0.0720	0.1362	0.0769
$\hat{\Theta}_{2,1}$	100	0.0344	0.0344	0.0945	0.0890	0.0355	0.0761	0.0322
$\hat{\Theta}_{2,2}$	50	0.0365	0.0365	0.2670	0.1461	0.0482	0.0800	0.8874
$\hat{\Theta}_{2,2}$	100	0.0297	0.0297	0.1511	0.0818	0.0341	0.0661	0.0288
$\hat{\Theta}_{3,1}$	50	0.0533	0.0533	0.1612	1.9412	0.0358	0.0866	0.1178
$\hat{\Theta}_{3,1}$	100	0.0363	0.0363	0.0274	0.1258	0.0278	0.0658	0.0823
$\hat{\Theta}_{3,2}$	50	0.0546	0.0546	0.4513	0.6494	0.0446	0.0784	0.1567
$\hat{\Theta}_{3,2}$	100	0.0350	0.0350	0.0427	0.1228	0.0307	0.0696	0.1088
$\hat{\Theta}_{4,1}$	50	0.0523	0.0523	0.2553	-	0.1124	0.1255	0.1634
$\hat{\Theta}_{4,1}$	100	0.0328	0.0328	0.1237	-	0.0966	0.0993	0.0871
$\hat{\Theta}_{4,2}$	50	0.0293	0.0293	0.3691	-	0.0730	0.0570	0.2139
$\hat{\Theta}_{4,2}$	100	0.0198	0.0198	0.0787	-	0.0381	0.0326	0.1481
$\hat{\Theta}_{5,1}$	50	0.0491	0.0491	1.0284	-	0.0608	0.1206	0.1929
$\hat{\Theta}_{5,1}$	100	0.0366	0.0366	0.3872	-	0.0420	0.0699	0.1269
$\hat{\Theta}_{5,2}$	50	0.0623	0.0623	0.0644	-	0.0416	0.0897	0.1535
$\hat{\Theta}_{5,2}$	100	0.0403	0.0403	0.0334	-	0.0274	0.0670	0.0901
$\hat{\Theta}_{6,1}$	50	0.0462	0.0462	0.1993	0.4563	0.0318	0.0452	0.1100
$\hat{\Theta}_{6,1}$	100	0.0324	0.0324	0.0821	0.0411	0.0205	0.0232	0.0520
$\hat{\Theta}_{6,2}$	50	0.0440	0.0440	1.7505	0.5246	0.0696	0.1632	0.1304
$\hat{\Theta}_{6,2}$	100	0.0360	0.0360	0.1684	0.0975	0.0410	0.1198	0.0680

para a amostra não classificada, as estimativas de  $\sigma$  e  $\mu_i$  são mais precisas do que as de  $\gamma_i$  por apresentarem valores de EQM menores, para  $i = 1, 2$ . Também, são apresentado no Apêndice os gráficos *boxplot* de cada parâmetro, de acordo com a Tabela (3.1), para amostra de tamanho 100.

Assim, como para o caso classificado, os valores do EQM para todos os parâmetros aumentam para amostra de tamanho menor, mas essas diferenças são pequenas, indicando que a estimação para amostra menor aproxima-se da estimação com amostra maior.

Tabela 3.4: Média de  $\tilde{\Theta}$ 

$\tilde{\Theta}$	$n$	$\tilde{p}_1$	$\tilde{p}_2$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$	$\tilde{\sigma}$	$\tilde{\mu}_1$	$\tilde{\mu}_2$
$\hat{\Theta}_{1,1}$	50	0.6000	0.4000	-	-	0.9823	6.0027	8.0522
$\hat{\Theta}_{1,1}$	100	0.6000	0.4000	-	-	0.9942	6.0105	8.0106
$\hat{\Theta}_{1,2}$	50	0.6000	0.4000	-	-	0.9823	1.0027	4.0522
$\hat{\Theta}_{1,2}$	100	0.6000	0.4000	-	-	0.9942	1.0105	4.0106
$\hat{\Theta}_{2,1}$	50	0.5000	0.5000	0.5397	1.1504	0.9712	1.0525	3.1157
$\hat{\Theta}_{2,1}$	100	0.5000	0.5000	0.5163	1.0579	0.9852	1.0266	3.0170
$\hat{\Theta}_{2,2}$	50	0.5000	0.5000	1.2602	1.0951	0.9880	1.0319	2.0371
$\hat{\Theta}_{2,2}$	100	0.5000	0.5000	1.0851	1.0601	0.9843	1.0251	2.0170
$\hat{\Theta}_{3,1}$	50	0.4000	0.6000	-0.9433	-0.5517	1.0408	-0.9977	0.9836
$\hat{\Theta}_{3,1}$	100	0.4000	0.6000	-0.6287	-0.5526	1.0310	-1.0143	0.9952
$\hat{\Theta}_{3,2}$	50	0.4000	0.6000	-1.0122	-0.3108	1.0496	-0.0017	1.9713
$\hat{\Theta}_{3,2}$	100	0.4000	0.6000	-0.6294	-0.3303	1.0462	-0.0138	2.0138
$\hat{\Theta}_{4,1}$	50	0.6000	0.4000	0.5243	-	0.9799	1.0269	3.0495
$\hat{\Theta}_{4,1}$	100	0.6000	0.4000	0.5251	-	0.9897	1.0193	3.0106
$\hat{\Theta}_{4,2}$	50	0.6000	0.4000	1.0863	-	0.9778	1.0418	1.0332
$\hat{\Theta}_{4,2}$	100	0.6000	0.4000	1.0649	-	0.9839	1.0192	1.0106
$\hat{\Theta}_{5,1}$	50	0.5000	0.5000	-0.9612	-	1.0028	1.9794	0.0326
$\hat{\Theta}_{5,1}$	100	0.5000	0.5000	-0.6756	-	0.9834	1.9705	0.0100
$\hat{\Theta}_{5,2}$	50	0.5000	0.5000	-0.8432	-	0.9934	0.9989	3.0246
$\hat{\Theta}_{5,2}$	100	0.5000	0.5000	-0.6571	-	0.9931	0.9695	3.0247
$\hat{\Theta}_{6,1}$	50	0.4000	0.6000	1.3329	-0.5843	1.0477	1.0445	-0.0197
$\hat{\Theta}_{6,1}$	100	0.4000	0.6000	1.0891	-0.5987	1.0293	1.0299	-0.0089
$\hat{\Theta}_{6,2}$	50	0.4000	0.6000	0.5963	-0.5855	1.0696	-0.9697	0.9784
$\hat{\Theta}_{6,2}$	100	0.4000	0.6000	0.5219	-0.5973	1.0456	-0.9752	0.9909

Tabela 3.5: Erro quadrático médio de  $\tilde{\Theta}$ 

$\tilde{\Theta}$	$n$	$\tilde{p}_1$	$\tilde{p}_2$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$	$\tilde{\sigma}_1$	$\tilde{\sigma}_2$	$\tilde{\mu}_1$	$\tilde{\mu}_2$
$\hat{\Theta}_{1,1}$	50	-	-	-	-	0.1108	0.2007	0.2196	
$\hat{\Theta}_{1,1}$	100	-	-	-	-	0.0710	0.1378	0.1714	
$\hat{\Theta}_{1,2}$	50	-	-	-	-	0.1108	0.2007	0.2196	
$\hat{\Theta}_{1,2}$	100	-	-	-	-	0.0710	0.1378	0.1714	
$\hat{\Theta}_{2,1}$	50	-	-	0.3222	0.2965	0.0833	0.1543	0.4397	
$\hat{\Theta}_{2,1}$	100	-	-	0.1548	0.1526	0.0552	0.1037	0.0577	
$\hat{\Theta}_{2,2}$	50	-	-	0.8148	0.3202	0.0581	0.0779	0.2076	
$\hat{\Theta}_{2,2}$	100	-	-	0.2106	0.1568	0.0497	0.0597	0.0577	
$\hat{\Theta}_{3,1}$	50	-	-	<b>1.7274</b>	0.1128	0.0560	0.1580	0.1080	
$\hat{\Theta}_{3,1}$	100	-	-	0.5181	0.1058	0.0414	0.1200	0.0755	
$\hat{\Theta}_{3,2}$	50	-	-	<b>1.9314</b>	0.1731	0.0679	0.1588	0.1801	
$\hat{\Theta}_{3,2}$	100	-	-	0.5180	0.0961	0.0516	0.1194	0.1076	
$\hat{\Theta}_{4,1}$	50	-	-	0.2352	-	0.1019	0.1365	0.2142	
$\hat{\Theta}_{4,1}$	100	-	-	0.1271	-	0.0683	0.0887	0.1714	
$\hat{\Theta}_{4,2}$	50	-	-	0.2280	-	0.0805	0.1314	0.2106	
$\hat{\Theta}_{4,2}$	100	-	-	0.1573	-	0.0560	0.0519	0.1714	
$\hat{\Theta}_{5,1}$	50	-	-	<b>2.8333</b>	-	0.0725	0.1763	0.2008	
$\hat{\Theta}_{5,1}$	100	-	-	0.8565	-	0.0568	0.1122	0.1462	
$\hat{\Theta}_{5,2}$	50	-	-	<b>2.5456</b>	-	0.0682	0.1425	0.2162	
$\hat{\Theta}_{5,2}$	100	-	-	0.8897	-	0.0525	0.1058	0.1386	
$\hat{\Theta}_{6,1}$	50	-	-	<b>1.2275</b>	0.6233	0.0634	0.1448	0.1308	
$\hat{\Theta}_{6,1}$	100	-	-	0.2468	0.4851	0.0481	0.0830	0.0816	
$\hat{\Theta}_{6,2}$	50	-	-	0.4956	0.6231	0.0741	0.1583	0.1315	
$\hat{\Theta}_{6,2}$	100	-	-	0.1609	0.4845	0.0554	0.1182	0.0815	

As Figuras de 3.1 a 3.6 comparam a curva de densidade  $h(x; \Theta)$  com as curvas de densidade  $h(x; \hat{\Theta})$  e  $h(x; \tilde{\Theta})$ . Em cada gráfico a curva contínua representa a densidade  $h(x; \Theta)$ , os valores de  $\Theta$  estão na Tabela 3.1. Já a curva tracejada em negrito representa a densidade da estimação por amostra não classificada  $h(x; \hat{\Theta})$ , denominada estimação 1. E a curva pontilhada representa a densidade da estimação por amostra classificada  $h(x; \tilde{\Theta})$ , denominada de estimação 2. Destaca-se que são utilizadas as estimativas para amostra de tamanho 100 apresentadas nas Tabelas 3.2 e 3.4.

Por meio da Figura 3.1, que trás a representação gráfica para o Caso 1 de mistura de GEV, nota-se que as curvas estimadas estão bem próximas da curva de parâmetros

para o primeiro conjunto de parâmetros  $\Theta_{1,1}$ , indicando que ambas as estimações obtêm bons ajustes. Já para o segundo conjunto de parâmetros  $\Theta_{1,2}$  observa-se que as curvas estimadas estão uma sobre a outra, de forma que não é possível distinguir uma da outra, e apresentam certa distancia da curva dos parâmetros.

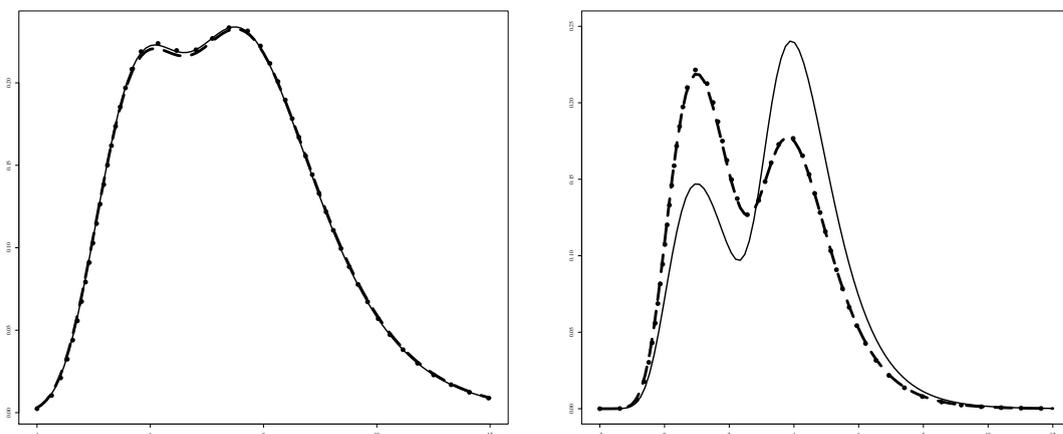


Figura 3.1: Gráfico das densidades  $h(x; \Theta)$  (curva contínua),  $h(x; \hat{\Theta})$  (curva tracejada e em negrito) e  $h(x; \tilde{\Theta})$  (curva pontilhada) para  $\Theta_{1,1}$  e  $\Theta_{1,2}$ .

A Figura 3.2 representa o Caso 2 de mistura GEV. As curvas estimadas e a curva de parâmetros para ambos os conjunto de parâmetros  $\Theta_{2,1}$  e  $\Theta_{2,2}$ , estão bem próximas entre si, indicando que ambas as estimações obtêm bons ajustes.

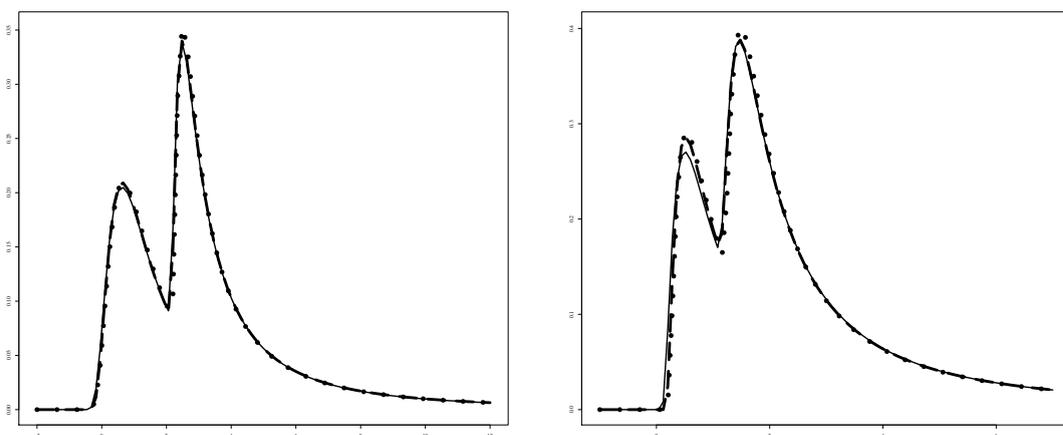


Figura 3.2: Gráfico das densidades  $h(x; \Theta)$  (curva contínua),  $h(x; \hat{\Theta})$  (curva tracejada e em negrito) e  $h(x; \tilde{\Theta})$  (curva pontilhada) para  $\Theta_{2,1}$  e  $\Theta_{2,2}$ .

Tem-se na Figura 3.3 a representação gráfica do Caso 3 de mistura de GEV e verifica-se que a estimaco 1 tem um ajuste melhor que a estimaco 2, para ambos

os conjuntos de parâmetros  $\Theta_{3,1}$  e  $\Theta_{3,2}$ . Uma possível explicação para esse comportamento é sobestimação do parâmetro  $\gamma_1$  para a amostra classificada, como pode ser verificado na Tabela(3.4).

Já a Figura 3.4 retrata o Caso 4 de mistura de GEV, para o primeiro conjunto de parâmetros,  $\Theta_{4,1}$ , nota-se que a curva estimada pela amostra classificada acompanha melhor a curva de densidade  $h(x; \Theta)$ . Para o segundo conjunto de parâmetros,  $\Theta_{4,2}$ , verifica-se que as curvas estimadas estão bem próximas e ambas obtêm bons ajustes a curva de densidade  $h(x; \Theta)$ .

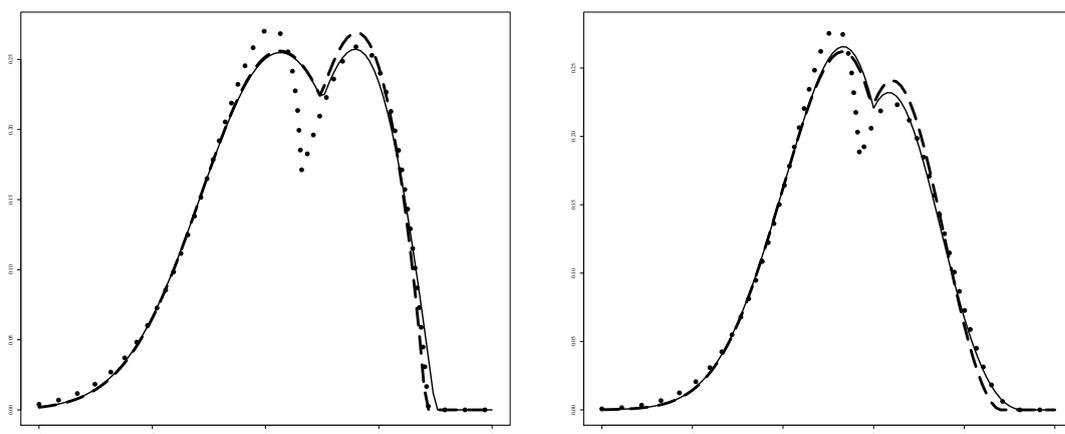


Figura 3.3: Gráfico das densidades  $h(x; \Theta)$ (curva contínua),  $h(x; \hat{\Theta})$  (curva tracejada e em negrito) e  $h(x; \tilde{\Theta})$  (curva pontilhada) para  $\Theta_{3,1}$  e  $\Theta_{3,2}$ .

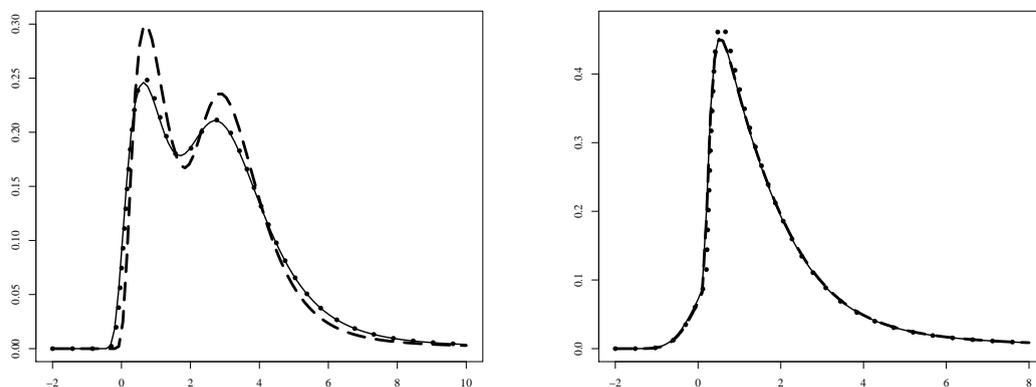


Figura 3.4: Gráfico das densidades  $h(x; \Theta)$ (curva contínua),  $h(x; \hat{\Theta})$  (curva tracejada e em negrito) e  $h(x; \tilde{\Theta})$  (curva pontilhada) para  $\Theta_{4,1}$  e  $\Theta_{4,2}$ .

A Figura 3.5 descreve o Caso 5 de mistura de GEV, constata-se que a estimação

1 obtém um ajuste melhor que a estimação 2 para ambos os conjuntos de parâmetros  $\Theta_{5,1}$  e  $\Theta_{5,2}$ . Pode-se explicar esse comportamento pela sobestimação do parâmetro  $\gamma_1$  para a amostra classificada, como pode ser verificado na Tabela(3.4).

Para finalizar, a Figura 3.6 representa o Caso 6 de mistura de GEV, as curvas estimadas estão bem próximas da curva de parâmetros  $h(x; \Theta)$  para ambos os conjuntos de parâmetros  $\Theta_{6,1}$  e  $\Theta_{6,2}$ .

Conclui-se que ambas as estimações obteve bons ajustes, no entanto, a estimação com amostra classificada (estimação 2) apresenta maiores problemas quando o parâmetro  $\gamma_1$  é negativo.

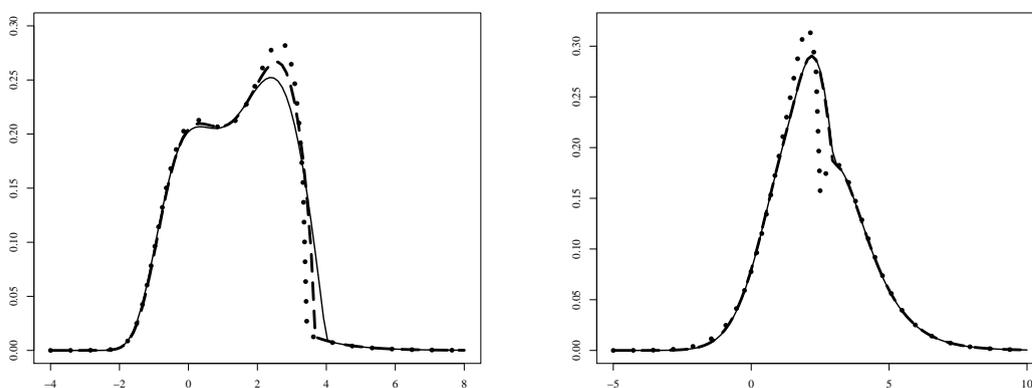


Figura 3.5: Gráfico das densidades  $h(x; \Theta)$ (curva contínua) ,  $h(x; \hat{\Theta})$  (curva tracejada e em negrito) e  $h(x; \tilde{\Theta})$  (curva pontilhada) para  $\Theta_{5,1}$  e  $\Theta_{5,2}$ .

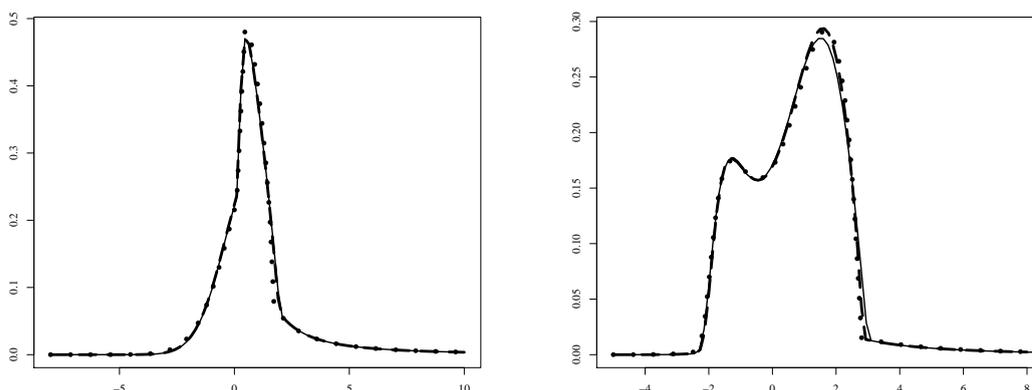


Figura 3.6: Gráfico das densidades  $h(x; \Theta)$ (curva contínua) ,  $h(x; \hat{\Theta})$  (curva tracejada e em negrito) e  $h(x; \tilde{\Theta})$  (curva pontilhada) para  $\Theta_{6,1}$  e  $\Theta_{6,2}$ .

Agora com interesse de investigar o desempenho da função discriminante não linear  $NL_i(x)$ ,  $i=m, c$  e  $o$ , de acordo com o tamanho da amostra e o tipo estimação são calculados os erros de classificação. São usados dois tamanhos de amostra  $n = 50$  e  $n = 100$ . Além disso, a estimação foi dividida em dois casos: para amostra não classificada (m) e amostra classificada (c), também usa-se os verdadeiros parâmetros, sendo esse caso chamado de ótimo(o).

A Tabela 3.6 apresenta as estimativas dos erros de classificação  $\widehat{e}_{1k}$  e  $\widehat{e}_{2k}$ ,  $k = o, m$  e  $c$ , para os dois conjuntos de parâmetros de cada caso da mistura de GEV. Esses erros são calculados empiricamente conforme é apresentado na Seção 2.3.2, nas Equações (2.49) e (2.50).

Tabela 3.6: Erros de Classificação de acordo com as Equações (2.49) e (2.50).

$\Theta$	n	$\widehat{e}_{1o}$	$\widehat{e}_{1m}$	$\widehat{e}_{1c}$	$\widehat{e}_{2o}$	$\widehat{e}_{2m}$	$\widehat{e}_{2c}$
$\widehat{\Theta}_{1,1}$	50	0.2170	0.2187	0.2060	0.1520	0.1588	0.1600
$\widehat{\Theta}_{1,1}$	100	0.2240	0.2256	0.2167	0.1519	0.1486	0.1568
$\widehat{\Theta}_{1,2}$	50	0.1254	0.1262	0.1213	0.0661	0.0645	0.0600
$\widehat{\Theta}_{1,2}$	100	0.1267	0.1264	0.1218	0.0580	0.0597	0.0658
$\widehat{\Theta}_{2,1}$	50	0.3146	0.3194	0.3144	0.0059	0.0123	0.0212
$\widehat{\Theta}_{2,1}$	100	0.3187	0.3211	0.3158	0.0052	0.0093	0.0152
$\widehat{\Theta}_{2,2}$	50	0.5480	0.4935	0.5048	0.0150	0.0818	0.0576
$\widehat{\Theta}_{2,2}$	100	0.5506	0.5119	0.5032	0.0176	0.0685	0.0566
$\widehat{\Theta}_{3,1}$	50	0.1204	0.1146	0.1040	0.1509	0.1484	0.1343
$\widehat{\Theta}_{3,1}$	100	0.1222	0.1270	0.1000	0.1512	0.1480	0.1497
$\widehat{\Theta}_{3,2}$	50	0.1359	0.1193	0.1125	0.1305	0.1318	0.1083
$\widehat{\Theta}_{3,2}$	100	0.1342	0.1371	0.1140	0.1281	0.1289	0.1332
$\widehat{\Theta}_{4,1}$	50	0.2524	0.2439	0.2260	0.1395	0.1742	0.1305
$\widehat{\Theta}_{4,1}$	100	0.2609	0.2505	0.2502	0.1321	0.1550	0.1368
$\widehat{\Theta}_{4,2}$	50	0.1884	0.1813	0.1707	0.5921	0.6138	0.5830
$\widehat{\Theta}_{4,2}$	100	0.1826	0.1819	0.1677	0.5876	0.6076	0.6110
$\widehat{\Theta}_{5,1}$	50	0.1570	0.1543	0.1424	0.2255	0.2248	0.2132
$\widehat{\Theta}_{5,1}$	100	0.1592	0.1593	0.1504	0.2298	0.2266	0.2138
$\widehat{\Theta}_{5,2}$	50	0.0000	0.0008	0.0012	0.3480	0.3482	0.2888
$\widehat{\Theta}_{5,2}$	100	0.0002	0.0006	0.0004	0.3484	0.3485	0.2692
$\widehat{\Theta}_{6,1}$	50	0.5768	0.5116	0.4385	0.0168	0.0470	0.0293
$\widehat{\Theta}_{6,1}$	100	0.5704	0.5499	0.5327	0.0149	0.0300	0.0165
$\widehat{\Theta}_{6,2}$	50	0.2748	0.2636	0.2600	0.0821	0.0833	0.0850
$\widehat{\Theta}_{6,2}$	100	0.2907	0.2949	0.2657	0.0872	0.0894	0.0840

Nota-se que no geral os erros de classificação são menores para a amostra de tamanho 50, para os erros  $e_{1k}$ ,  $i=m$ ,  $c$  e  $o$ . Enquanto, na maior parte dos casos, os erros  $e_{2k}$ ,  $i=m$ ,  $c$  e  $o$ , são menores na amostra de tamanho 100. Sendo que a diferença é muito pequena, demonstrando que o tamanho da amostra não tem grande influência no valor do erro de classificação.

Na maior parte dos exemplos, tanto o erro de classificação  $e_{1k}$ , quanto o  $e_{2k}$ ,  $i=m$ ,  $c$  e  $o$ , obtém-se resultados próximos para os casos ótimo, amostra classificada e não classificada. Apontando que o tipo de estimação, também, não tem grande influência no valor do erro de classificação.

A Tabela 3.7 apresenta as estimativas dos erro de classificação total para os três casos em estudo,  $\widehat{e}_k$ ,  $i=m$ ,  $c$  e  $o$ , conforme Equação (2.52). Também são exibidos os valores do viés absoluto relativo para amostra não classificada,

$$B(\widehat{e}_m) = \frac{|\widehat{e}_m - \widehat{e}_o|}{\widehat{e}_o}, \quad (3.1)$$

e para amostra classificada,

$$B(\widehat{e}_c) = \frac{|\widehat{e}_c - \widehat{e}_o|}{\widehat{e}_o}. \quad (3.2)$$

Ainda, tem-se o valor do viés absoluto relativo de  $\widehat{e}_m$  em relação a  $\widehat{e}_c$ ,

$$B = \frac{|\widehat{e}_m - \widehat{e}_c|}{\widehat{e}_c}. \quad (3.3)$$

Por meio, da Tabela 3.7, nota-se que os erros de classificação total apresentam valores próximos para os três casos: ótimo, amostra não classifica e amostra classificada. De modo geral, as maiores diferenças são verificadas quando se utiliza o tamanho de amostra  $n = 50$ , conforme mostra os resultados do cálculo dos viés absoluto relativo. Os maiores valores do viés padronizado em relação ao erro de classificação ótimo,  $\widehat{e}_o$ , são para o caso de amostra classificada e que apresentam um dos parâmetros  $\gamma_i$ , com valor negativo, para  $i = 1, 2$ , especificamente os casos 3 e 5. Quando a comparação é feita entre os métodos para amostra não classificada e classificada se verifica, na maior parte dos casos, que o viés é maior para amostra

Tabela 3.7: Erros de Classificação total e cálculos do viés.

$\Theta$	n	$\widehat{e}_o$	$\widehat{e}_m$	$\widehat{e}_c$	$B(\widehat{e}_m)$	$B(\widehat{e}_c)$	$B$
$\widehat{\Theta}_{1,1}$	50	0.1910	0.1950	0.1876	0.0684	0.0150	0.0393
$\widehat{\Theta}_{1,1}$	100	0.1952	0.1946	0.1927	0.0145	0.0338	0.0101
$\widehat{\Theta}_{1,2}$	50	0.1017	0.1018	0.0968	0.0011	0.0474	0.0514
$\widehat{\Theta}_{1,2}$	100	0.0992	0.0996	0.0994	0.0160	0.0659	0.0023
$\widehat{\Theta}_{2,1}$	50	0.1602	0.1683	0.1678	0.1664	0.0330	0.0032
$\widehat{\Theta}_{2,1}$	100	0.1620	0.1654	0.1655	0.0975	0.0464	0.0007
$\widehat{\Theta}_{2,2}$	50	0.2815	0.2895	0.2812	0.1270	0.0726	0.0296
$\widehat{\Theta}_{2,2}$	100	0.2841	0.2913	0.2799	0.2009	0.0344	0.0407
$\widehat{\Theta}_{3,1}$	50	0.1387	0.1350	0.1222	0.0762	0.4612	0.1050
$\widehat{\Theta}_{3,1}$	100	0.1396	0.1397	0.1298	0.0044	0.3349	0.0766
$\widehat{\Theta}_{3,2}$	50	0.1327	0.1269	0.1100	0.1183	0.5123	0.1533
$\widehat{\Theta}_{3,2}$	100	0.1306	0.1321	0.1255	0.0399	0.2718	0.0527
$\widehat{\Theta}_{4,1}$	50	0.2072	0.2163	0.1878	0.1633	0.1436	0.1515
$\widehat{\Theta}_{4,1}$	100	0.2094	0.2121	0.2048	0.0706	0.0386	0.0358
$\widehat{\Theta}_{4,2}$	50	0.3499	0.3534	0.3356	0.0539	0.1713	0.0530
$\widehat{\Theta}_{4,2}$	100	0.3446	0.3514	0.3450	0.1517	0.1098	0.0184
$\widehat{\Theta}_{5,1}$	50	0.1912	0.1890	0.1778	0.0354	0.1587	0.0628
$\widehat{\Theta}_{5,1}$	100	0.1945	0.1929	0.1821	0.0356	0.1809	0.0591
$\widehat{\Theta}_{5,2}$	50	0.1740	0.1728	0.1450	0.0191	0.3113	0.1919
$\widehat{\Theta}_{5,2}$	100	0.1743	0.1743	0.1348	0.0008	0.5420	0.2933
$\widehat{\Theta}_{6,1}$	50	0.2408	0.2335	0.1930	0.1105	0.5172	0.2101
$\widehat{\Theta}_{6,1}$	100	0.2371	0.2367	0.2230	0.0109	0.3203	0.0616
$\widehat{\Theta}_{6,2}$	50	0.1592	0.1555	0.1550	0.0685	0.1722	0.0031
$\widehat{\Theta}_{6,2}$	100	0.1686	0.1710	0.1567	0.0671	0.0685	0.0910

de tamanho 50.

## 3.2 Aplicações

### 3.2.1 Dados da vazão máxima anual das usinas da bacia do rio Paranapanema

No contexto atual de mudanças climáticas, as inundações são responsáveis por muitas tragédias, causando perdas econômicas e até mesmo perdas de vidas humanas. Pode-se citar o que ocorreu no Rio de Janeiro em 2011, quando em menos de 24 horas choveu o que costuma chover em um mês. Não só as cidades ficaram debaixo d'água como o solo das encostas cedeu, causando deslizamentos que passaram por cima do que estava no caminho.

As cheias são a principal causa das inundações, por esse motivo, seu comportamento vem sendo estudado durante as últimas décadas. Nesse cenário, é importante aperfeiçoar os métodos estatísticos para o controle desse fenômeno natural, de modo a preservar vidas, impactos socioeconômicos e ambientais.

O conjunto de dados usado para aplicação é a vazão máxima anual das usinas da bacia do rio Paranapanema para os anos de 1931 a 2002. Essas séries são disponibilizadas pelo ONS (Operador Nacional do Sistema). A bacia do rio Paranapanema, foi escolhida para aplicar a metodologia estudada por apresentar mais de um período de cheia anual, máximos anuais que podem ocorrer tanto no inverno como no verão, as quais seriam advindas de diferentes populações.

Destaca-se que o Teste Ljung-Box foi aplicado para verificar a hipótese nula de independência serial, sendo que a estatística de teste não rejeita a hipótese nula.

Para determinar a função mistura de GEV que melhor adequa-se aos dados, inicialmente busca-se entender o comportamento dos dados através de um histograma. A vazão máxima anual das usinas distribui-se em um histograma, Figura 3.7(a), com um possível padrão de uma mistura com um  $\gamma$  positivo e outro negativo. Na Figura 3.7(b) são colocadas dois esboços das distribuição GEV em cada região modal no histograma dos dados. A curva com linha contínua possui  $\gamma$  positivo e a curva com linha tracejada possui  $\gamma$  negativo. Com o suporte desses dois esboços da Figura 3.7(b) é possível inferir que a mistura a ser estimada é a curva do Caso 6, como é

apresentada na Seção 2.2.1 desta dissertação. Ou seja, tem-se uma curva com  $\gamma_1 > 0$  e  $\gamma_2 < 0$ .

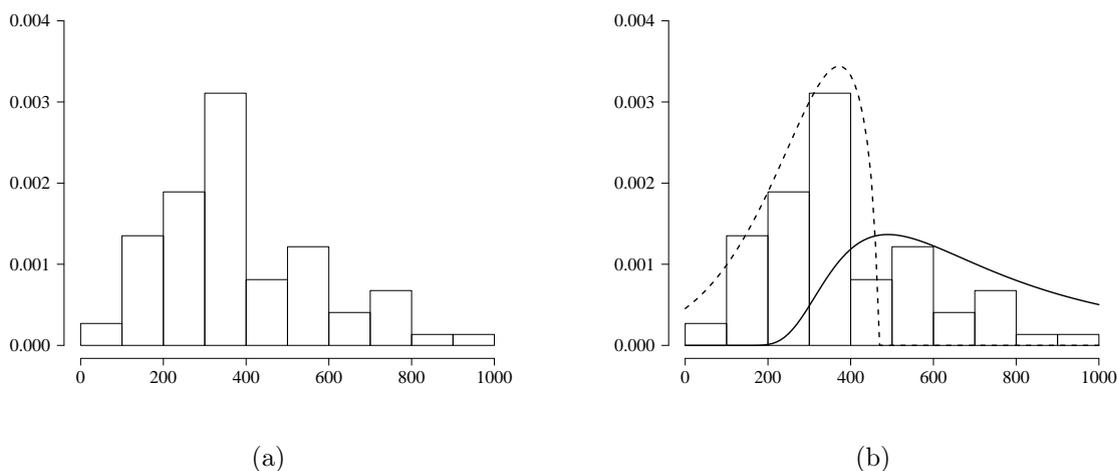


Figura 3.7: Histograma dos dados vazão máxima anual das usinas da bacia do rio Paranapanema e as curvas de distribuição GEV em cada região modal.

Em posse dessa informação, o passo seguinte é estimar por máxima verossimilhança os parâmetros da função mistura de GEV, a partir das expressões (2.5), (2.12), (2.13) e (2.14) dadas no Caso 6, do Capítulo 2, que resulta em uma curva com as estimativas dos parâmetros apresentado na Tabela 3.8. A Figura 3.8 contém a representação da curva com os parâmetros da Tabela 3.8. Observa-se que se teve um bom ajuste, com o traço da curva próximo do comportamento dos dados explicitado no histograma. E com suas duas regiões modais bem caracterizadas no ajuste.

Tabela 3.8: Estimativa e desvio padrão dos parâmetros da função mistura de GEV para os dados da vazão máxima anual das usinas.

$\hat{p}_1$	$\hat{p}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
0,36	0,64	0,12	-0,69	103,83	508,65	248,17
0,0614	0,0614	0,2165	0,0148	10,1937	28,3135	13,2789

Ainda foi realizado o teste de Kolmogorov-Smirnov para verificar a hipótese nula de que os dados seguem a distribuição em questão, obtendo um p-valor de 0,9661, ou seja, não rejeita-se a hipótese nula. Portanto, os dados tem distribuição da mistura de GEV estudada.

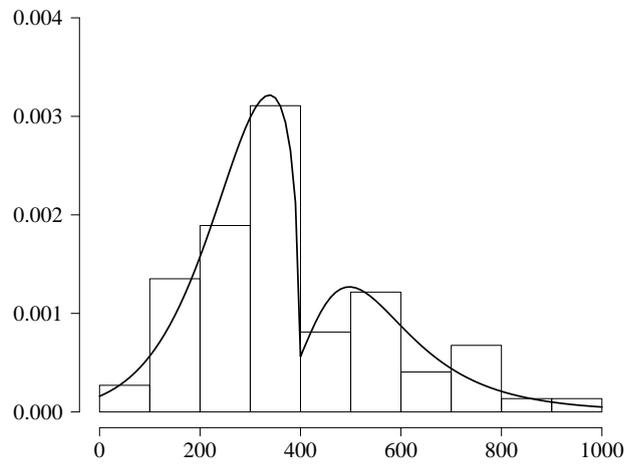


Figura 3.8: Histograma dos dados e curva estimada

O QQ-plot na Figura 3.9(b) também indica uma boa estimação, com seus quantis quase todos em um perfeito alinhamento, exceto nos quantis mais altos, que possivelmente sofrem influência por causa da concentração de dados do intervalo [700;800]. No entanto, a distribuição acumulada da função estimada, Figura 3.9(a), aproxima-se bastante da distribuição acumulada empírica.

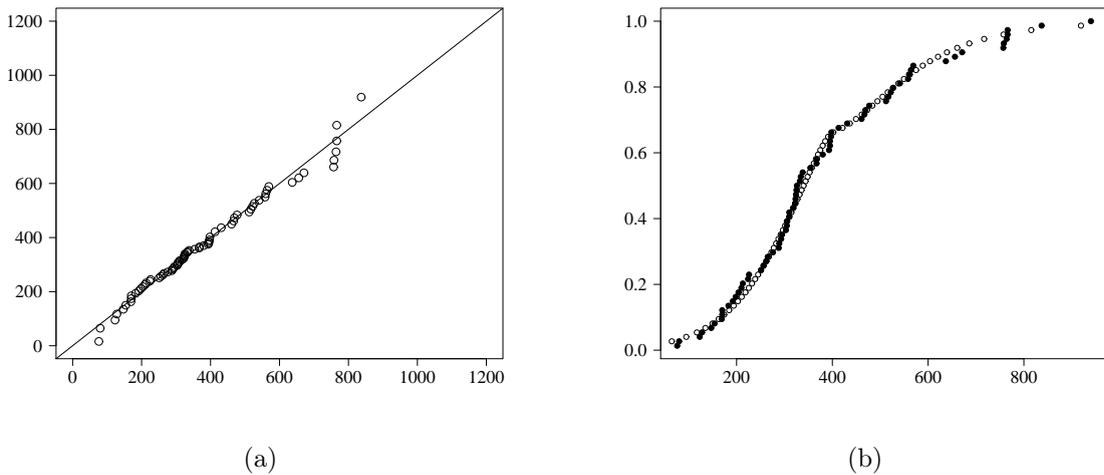


Figura 3.9: (a) QQ-plot e (b) Função de distribuição acumulada da mistura de GEV estimada.

A função discriminante não linear pode ser encontrada pela expressão (2.45), utilizando as estimativas obtidas.

Calcula-se, então, para cada observação da amostra, o valor da função discriminante correspondente, se  $NL_m(x_j) < 0$  então  $x_j$  é classificada na primeira população e se  $NL_m(x_j) > 0$  é classificada na segunda população. Sendo assim, há 74 observações, das quais 26 são classificadas na primeira população e 48 na segunda população. Além disso, como a amostra havia sido dividida previamente em duas populações, máximos anuais que ocorreram no inverno e no verão, é possível calcular os erros de classificação. Com isso, a probabilidade de cometer o erro de classificar uma observação na segunda população e ela pertencer a primeira é de  $e_{1m} = 0,5833$ . Enquanto a probabilidade de classificar na primeira população e ela pertencer a segunda é de  $e_{2m} = 0,0263$ . Desta maneira, obtém-se um erro de classificação global de  $e_m = 0,2245$ .

Uma informação de muita relevância para hidrologia é o período de retorno, trata-se do intervalo de tempo estimado de ocorrência de um determinado evento. O período de retorno é definido como o inverso da probabilidade de um evento ser igualado ou ultrapassado e pode ser expresso por:

$$T(x) = \frac{1}{1 - F(x; \Theta)}. \quad (3.4)$$

Enquanto, a função tempo de retorno empírica é dada por:

$$T(\hat{x}) = \left[ \left( \frac{1}{N} \right) \times \sum_{i=1}^N I_{(X_i \geq x)} \right]^{-1}. \quad (3.5)$$

Para cada observação da amostra estima-se o período de retorno, utilizando a Equação (3.4) e os parâmetros estimados (Tabela 3.8), sendo  $F(x; \Theta)$  a distribuição acumulada do modelo de mistura de duas componentes GEV. Além disso, calcula-se a função tempo de retorno empírica para cada valor do conjunto de dados, por meio da Equação (3.5). Na Figura 3.10 os pontos cinza representam o período de retorno estimado e os pontos pretos correspondem a função de retorno empírica de cada valor da amostra pela vazão.

Nota-se pela Figura 3.10 que quanto maior a vazão, mais longo é o tempo de retorno estimado e o valor da função de retorno empírica.

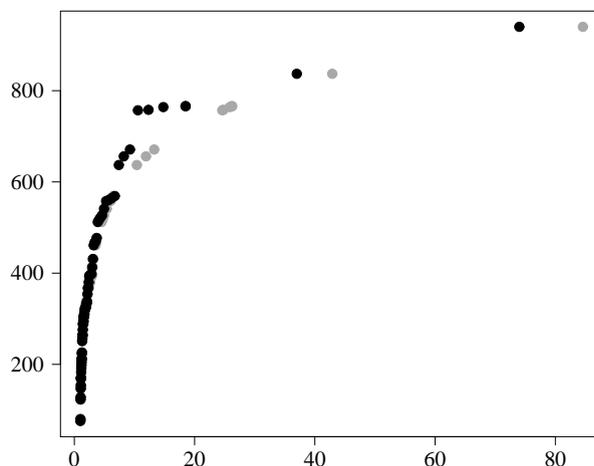


Figura 3.10: Gráfico período de retorno estimado e pela vazão.

### 3.2.2 Dados históricos do índice S&P 500

O Índice *Standard & Poor* 500 (S&P 500) é um índice de 500 ações norte-americanas de maior representatividade no mercado de capitais, pois capta cerca de 80% de cobertura de capitalização de mercado disponível. É considerado o melhor indicador individual do mercado acionário americano de grande capitalização. Fora coletado para análise, o maior índice registrado de cada mês, no período de 01/01/2010 a 31/12/2016, somando portanto 84 observações. No histograma da Figura 3.11(a) vê-se o comportamento bimodal nos dados. Nota-se na Figura 3.11(b), por meio do esboço das curvas de distribuição GEV, que novamente os dados adequam-se ao caso da mistura de GEV com um gama positivo e outro negativos. No entanto, ao contrário do que aconteceu no exemplo anterior, as caudas longas das distribuições se encontram.

As estimativas da função mistura de GEV, explicitadas na Tabela 3.9, concebem uma curva com características muito próximas do histograma dos dados, como pode ser visto na Figura 3.12. Com isso, a função estimada apresenta um ajuste muito bom, dado que seu comportamento é muito semelhante ao histograma, principalmente representando as duas regiões modais com exatidão.

Também, foi realizado o teste de Kolmogorov-Smirnov para verificar a hipótese

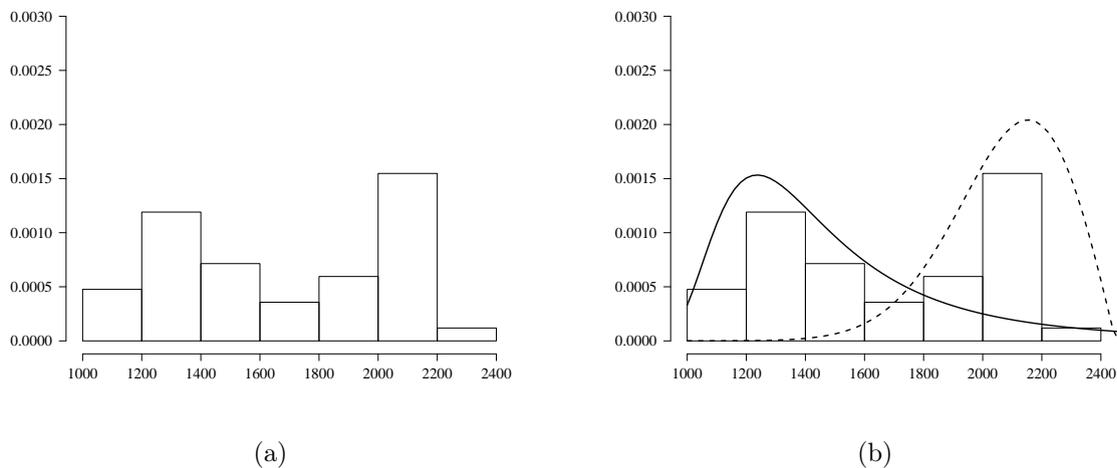


Figura 3.11: Histograma dos maiores valores mensais do índice S&P 500 e esboços das curvas de distribuição GEV em cada região modal.

Tabela 3.9: Estimativa e desvio padrão dos parâmetros da função mistura de GEV para os maiores valores mensais do índice S&P 500.

$\hat{p}_1$	$\hat{p}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\sigma}$	$\hat{\mu}_1$	$\hat{\mu}_2$
0,55	0,45	0,11	-0,52	136,06	1289	2011,51
0,0710	0,0710	0,1102	0,8273	13,5744	22,9349	27,7996

nula de que os dados seguem a distribuição em questão, obtendo um p-valor de 0.8466, ou seja, não rejeita-se a hipótese nula. Portanto, os dados tem distribuição da mistura de GEV estudada.

Tanto o QQ-plot quanto a distribuição acumulada da função estimada, Figura 3.13, apresentam bons resultados, com um bom alinhamento dos pontos em ambas situações. Diferentemente do caso analisado na Seção 3.2.1, o QQ-plot da Figura 3.13(a) não possui pontos discrepantes.

Nesta aplicação, como as populações não estavam previamente separadas, não é possível calcular o erro de classificação. Mas, pode-se comparar a proporção que foi classificada para cada população por meio da função discriminante com os valores estimados para os parâmetros  $p_1$  e  $p_2$ . Assim, há 84 observações, das quais 45 são classificadas na primeira população e 38 na segunda população, obtendo proporções de 0,55 e 0,45, respectivamente. Sendo que,  $\hat{p}_1 = 0,59$  e  $\hat{p}_2 = 0,41$ , logo são valores próximos das proporções obtidas pela função discriminante.

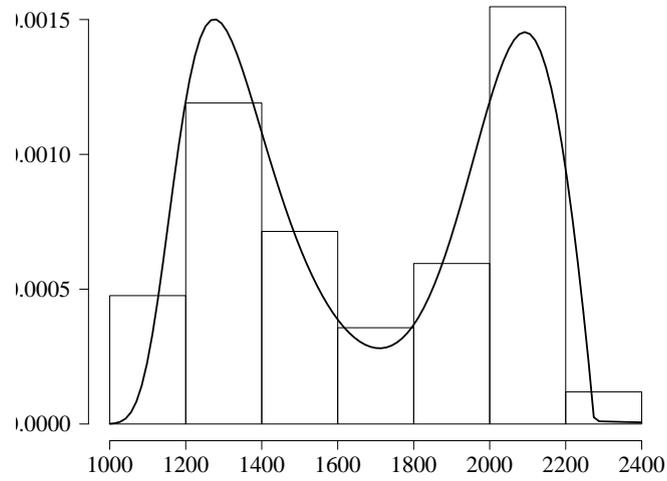


Figura 3.12: Histograma dos maiores valores mensais do índice S&P 500 e sua curva estimada

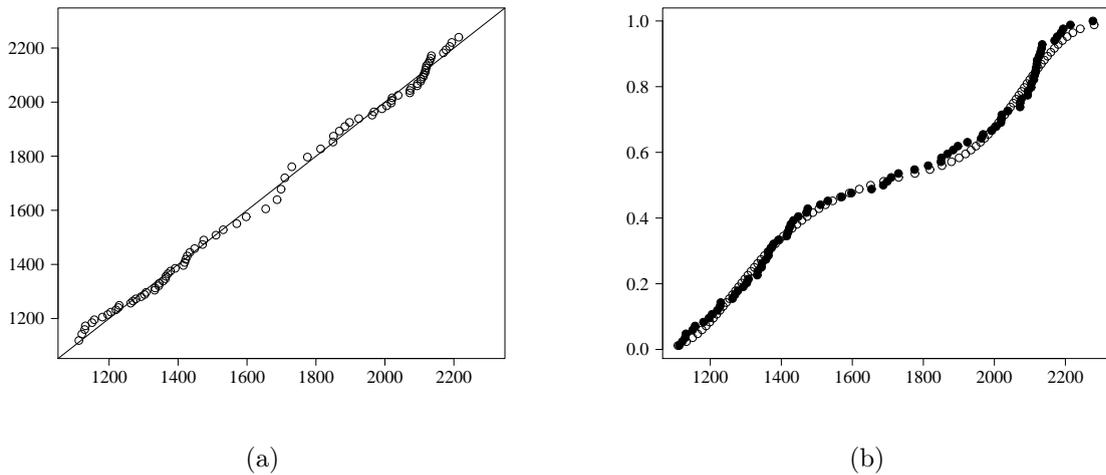


Figura 3.13: (a) QQ-plot e (b) Função de distribuição acumulada da mistura de GEV estimada.

Ainda, destaca-se que ao aplicar o Teste Ljung-Box, a estatística de teste rejeitou a hipótese nula, indicando que os dados podem ser autocorrelacionados.

# Capítulo 4

## Conclusões

Neste trabalho foram apresentados os métodos para estimação dos parâmetros de uma mistura de duas distribuições GEV e suas respectivas funções de discriminante. Foram analisados os comportamentos das estimativas do modelo e as estimativas da função discriminante, para amostras classificadas e não classificadas, para diferentes tamanhos de amostra. Além disso, como o parâmetro de forma da GEV pode assumir valor positivo, negativo ou nulo, foram analisados os seis modelos possíveis de mistura de duas distribuições GEV, quando se combina cada componente com os diferentes sinais de  $\gamma_i$ .

Inicialmente, o interesse foi o de comparar o comportamento dos estimadores de máxima verossimilhança para os casos de amostra não classificada e amostra classificada. Os experimentos de simulação mostraram que os dois tipos de estimações apresentaram bons resultados. Em ambos os casos de estimação, as estimativas de  $p_i$ ,  $\sigma_i$  e  $\mu_i$  são mais precisas do que as de  $\gamma_i$  por apresentarem valores de EQM menores, para  $i = 1, 2$ . Também, nota-se que os valores do EQM para todos os parâmetros aumentam para amostra de tamanho menor, entretanto essas diferenças são pequenas, o que mostra que há pouca alteração no resultado para amostra menor, para os dois tipos de estimação.

Outro objetivo era investigar o desempenho da função discriminante não linear, comparando os casos de amostra não classificada, amostra classificada e ótimo por meio do erro de classificação. Com base nos resultados dos experimentos de simulação verificou-se que, no geral, mesmo havendo diferença nas estimativas do modelo

quando as amostras são de tamanho 50 e 100, os erros de classificação para esses diferentes tamanhos de amostras são bastante próximos. Demonstrando que o uso de uma amostra menor não tem grande influência no valor do erro de classificação. Ainda, na maior parte dos exemplos, o erro de classificação  $e_{1k}$  obteve resultados próximos para os casos ótimo, amostra classificada e não classificada, e o mesmo foi visto para o erro de classificação  $e_{2k}$ . Apontando que o tipo de estimação, também, não tem grande influência no valor do erro de classificação.

Em relação aos erros de classificação global, de modo geral, as maiores diferenças foram verificadas quando se utilizou o tamanho de amostra  $n=50$ , independente do tipo de estimação. Os maiores valores do viés padronizado em relação ao erro de classificação ótimo são para o caso de amostra classificada e que apresentam um dos parâmetros  $\gamma_i$ , com valor negativo, para  $i = 1, 2$ , especificamente os casos 3 e 5. Quando a comparação é feita entre os métodos para amostra não classificada e classificada, na maioria dos exemplos, o viés é maior para amostra de tamanho 50.

Destaca-se que houve dificuldade para a obtenção de expressões fechadas para os erros de classificação, dado que teriam que ser feitas muitas restrições. Desta maneira, optou-se por calcular os erros de classificação de forma empírica.

As aplicações em dados reais ilustraram a eficiência do modelo de mistura de duas distribuições GEV para situações bimodais. A primeira aplicação utilizou a vazão máxima anual das usinas da bacia do rio Paranapanema para os anos de 1931 a 2002. Essa aplicação apresentou um bom ajuste, ainda que seu QQ-plot exibiu alguns pontos discrepantes, possivelmente por influência da concentração de dados do intervalo [700;800]. Além disso, o erro de classificação global foi de  $e_m = 0,2245$ . Esse alto valor pode estar relacionado com a proximidade das duas populações.

Já a segunda aplicação utilizou o maior valor registrado de cada mês para o Índice S&P 500, no período de 01/01/2010 a 31/12/2016. Esse caso, obteve um ajuste ainda melhor que o primeiro, pois, o QQ-plot apresentou um bom alinhamento dos pontos sem nenhum ponto discrepante. Não foi possível calcular o erro de classificação para essa aplicação. No entanto, a proporção que foi classificada para cada população por meio da função discriminante ficou próxima dos valores estimados para os parâmetros  $p_1$  e  $p_2$ . E também, há indicação que pode haver autocorrelação nos dados.

Para proposta de trabalhos futuros, poderia ser levantada a questão da prova da identificabilidade da classe de mistura de distribuições GEV que ainda não é conhecida e por esse motivo foi apresentado a identificabilidade da classe de mistura de distribuições extremas. Ainda, utilizar a mistura de distribuições GEV em série temporais para tratar dados autocorrelacionados com bimodalidade.

# Referências Bibliográficas

- Ahmad, K. E. (1988). Identifiability of finite mixtures using a new transform. *Ann. Inst. Stat. Math.*, 40.
- Ahmad, K. E. e Abd-Elrahman, A. M. (1994). Updating a nonlinear discriminant function estimated from a mixture of two weibull distributions. *Mathematical and Computer Modelling*, 19(11):41–51.
- Ahmad, K. E., Jaheen, Z. F., e Modhesh, A. A. (2010). Estimation of a discriminant function based on small sample size from a mixture of two gumbel distributions. *Communications in Statistics - Simulation and Computation*, 39(4):713–725.
- Al-Hussaini, E. K. e Ahmad, K. E. (1981). On the identifiability of finite mixture of distributions. *IEEE Trans. Inform. Theor.*, 27:664—668.
- Amoh, R. K. (1984). Estimation of a discriminant function from a mixture of two inverse gaussian distributions when the sample size is small. *Journal of Statistical Computation and Simulation*, 20(4):275–286.
- Atienza, N., Garcia-Heras, J., e Muñoz-Pichardo, J. M. (2006). A new condition for identifiability of finite mixture distributions. *Metrika*, 63:215–221.
- Chandra, S. (1977). On the mixtures of probability distributions. *Scand. J. Stat.*, 4:105–112.
- Escalante-Sandoval, C. (2007). A mixed distribution with ev1 and gev components for analyzing heterogeneous samples. *Ingeniería Investigación y Tecnología*, 8(3):123–133.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44:423–453.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorology Society*, 87:145–158.

- Jondeau, E., Poon, S.-H., e Rockinger, M. (2007). *Financial Modeling Under Non-Gaussian Distributions*. Springer-Verlag London.
- Kilbas, A. A., Rodríguez-Germá, L., Saxena, R. K., e Trujillo, J. J. (2010). The krätzel function and evaluation of integrals. *Computers and Mathematics with Applications*, 59:1790–1800.
- Kotz, S. e Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Application*. Imperial College Press.
- Mahmoud, M. e Moustafa, H. (1993). Estimation of a discriminant function from a mixture of two gamma distributions when the sample size is small. *Mathematical and Computer Modelling*, 18(5):87–95.
- McLachlan, G. J. e Basford, K. E. (1988). *Mixture models*. Marcel Dekker.
- McLachlan, G. J. e Peel, D. (2000). *Finite mixture models*. wiley.
- Mingoti, S. A. (2005). *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG.
- Nadarajah, S. e Pogány, T. K. (2012). On the characteristic functions for extreme value distributions. *Extremes*, 16(1):27–39.
- O’Neill, T. J. (1978). Normal discrimination with unclassified observations. *J. Am. Stat. Assoc.*, 73:821–826.
- Otiniano, C. E. G., Gonçalves, C. R., e Dorea, C. C. Y. (2016). Mixture of extreme value distributions: identifiability and estimation. *Communications in Statistics - Theory and Methods*.
- Otiniano, C. G. e Teixeira, E. (2014). Estimação dos parâmetros da mistura de duas componentes gev via algoritmo em. *Tendências em Matemática Aplicada e Computacional*, 15(1):59–71.
- Sultan, K. S. e Al-Moisheer, A. (2013). Estimation of a discriminant function from a mixture of two inverse weibull distributions. *Journal of Statistics Computation and Simulation*, 83(3):405–416.
- Sultan, K. S., Ismail, M. A., e Al-Moisheer, A. S. (2007). Mixture of two inverse weibull distributions: Properties and estimation. *Computational Statistics & Data Analysis*, 51:5377–5387.

Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Stat*, 34:1265—1269.

Titterington, D. M., F., S. A. M., e Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. wiley.

# Apêndice

São apresentados os gráfico *boxplot* das estimativas para os conjunto de parâmetros dos seis casos para amostra classifica e não classificada de tamanho 100. Para construir esses gráficos utiliza-se as 100 estimativas de cada parâmetro do Método de Monte Carlo.

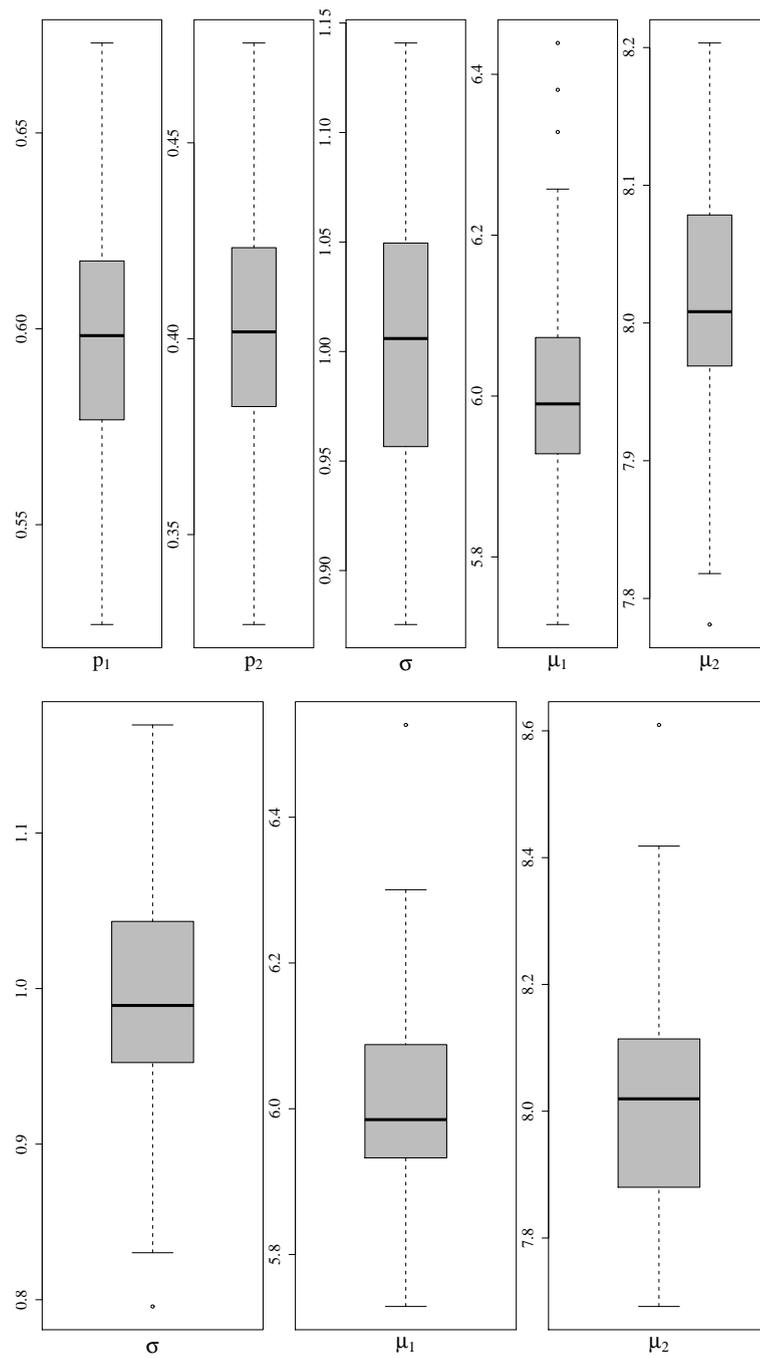


Figura 1: Gráficos *boxplot* das estimativas de  $\Theta_{1,1}$  para amostra não classificada e amostra classificada

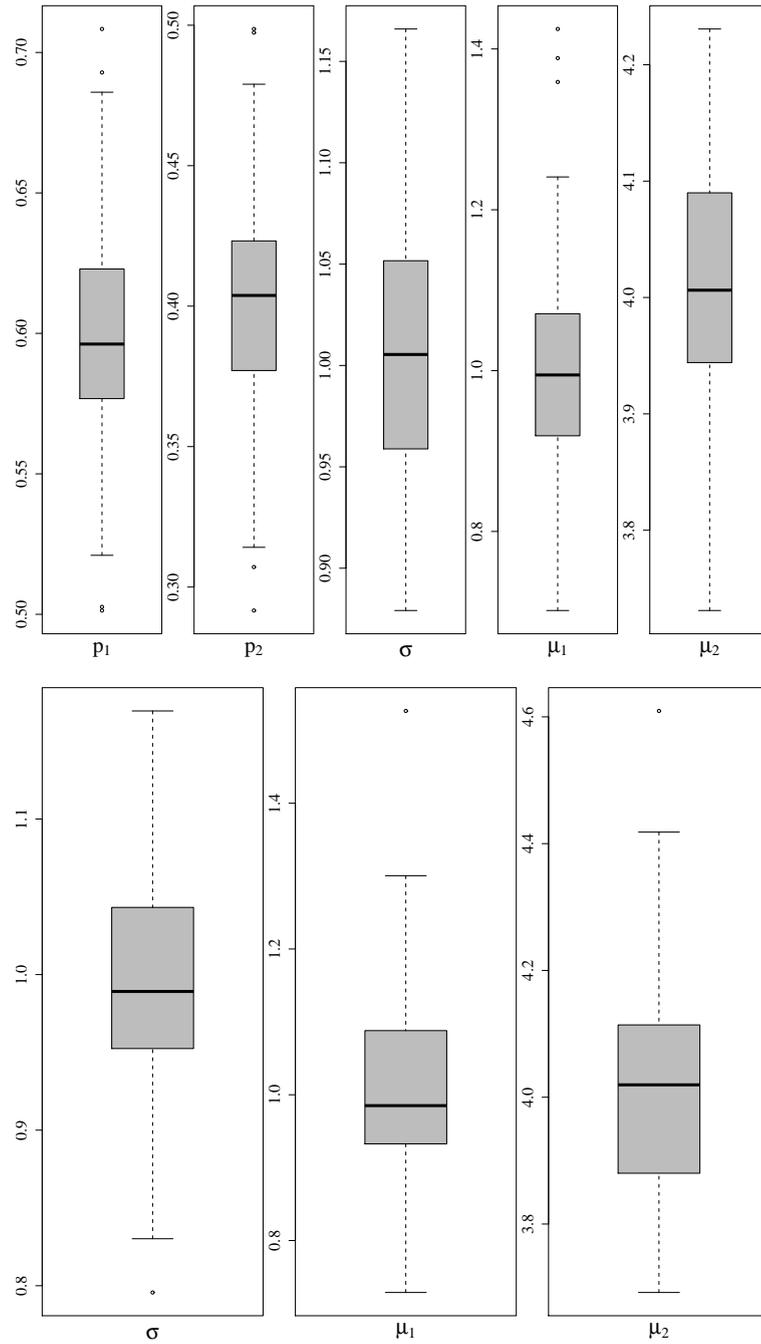


Figura 2: Gráficos *boxplot* das estimativas de  $\Theta_{1,2}$  para amostra não classificada e amostra classificada

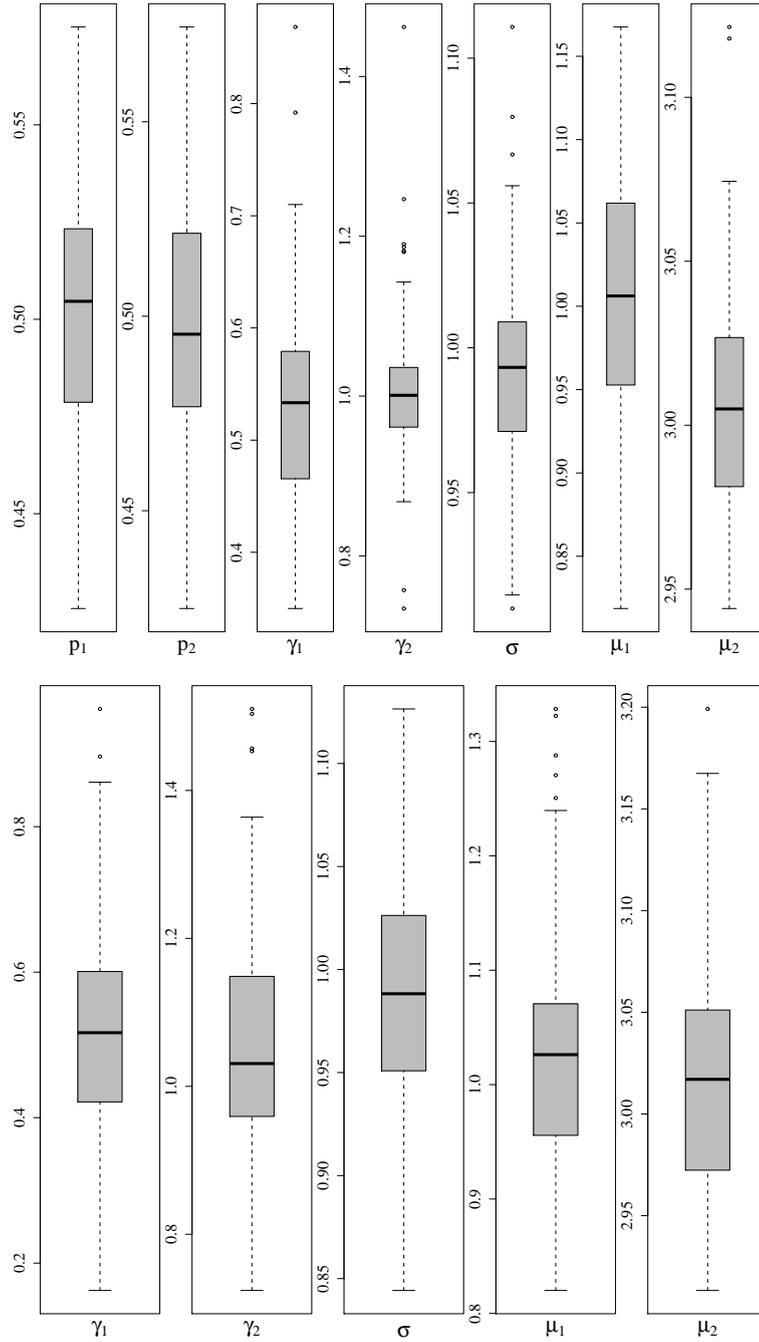


Figura 3: Gráficos *boxplot* das estimativas de  $\Theta_{2,1}$  para amostra não classificada e amostra classificada

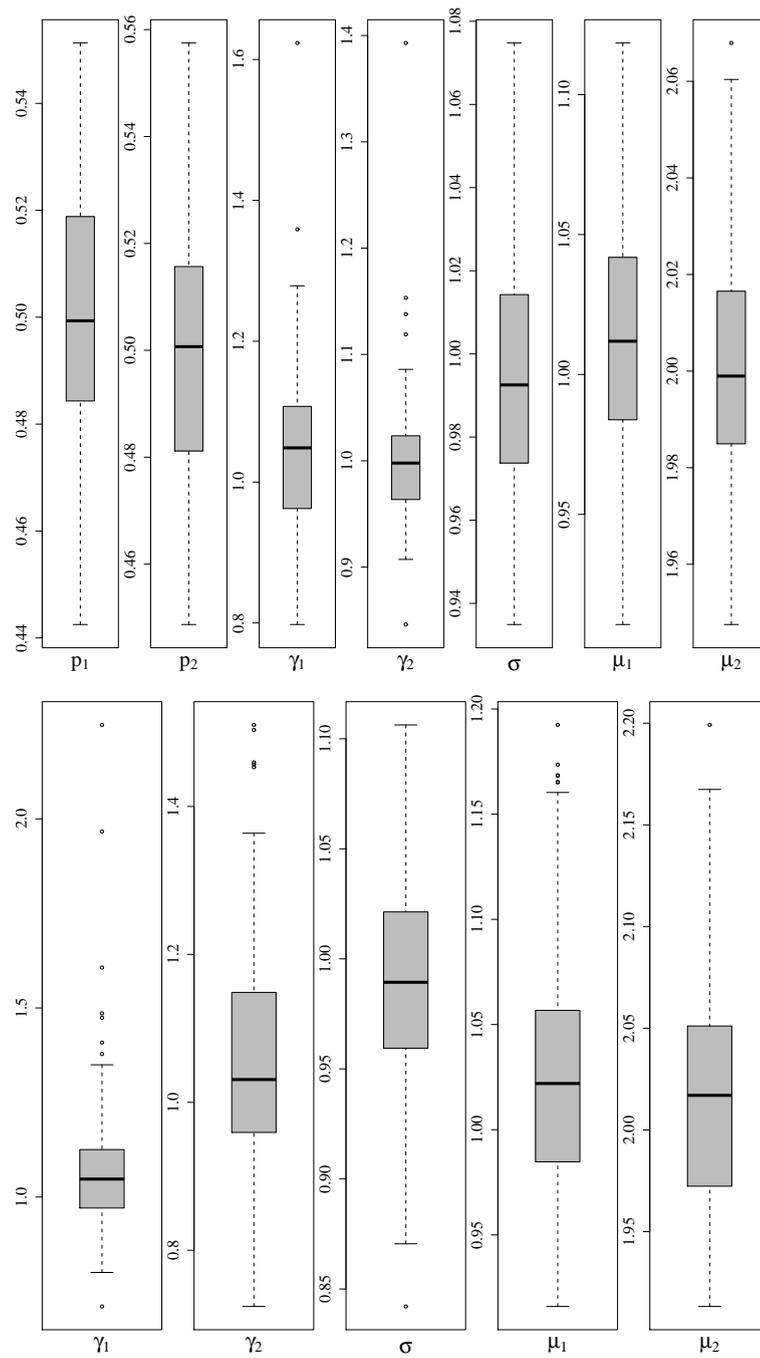


Figura 4: Gráficos *boxplot* das estimativas de  $\Theta_{2,2}$  para amostra não classificada e amostra classificada

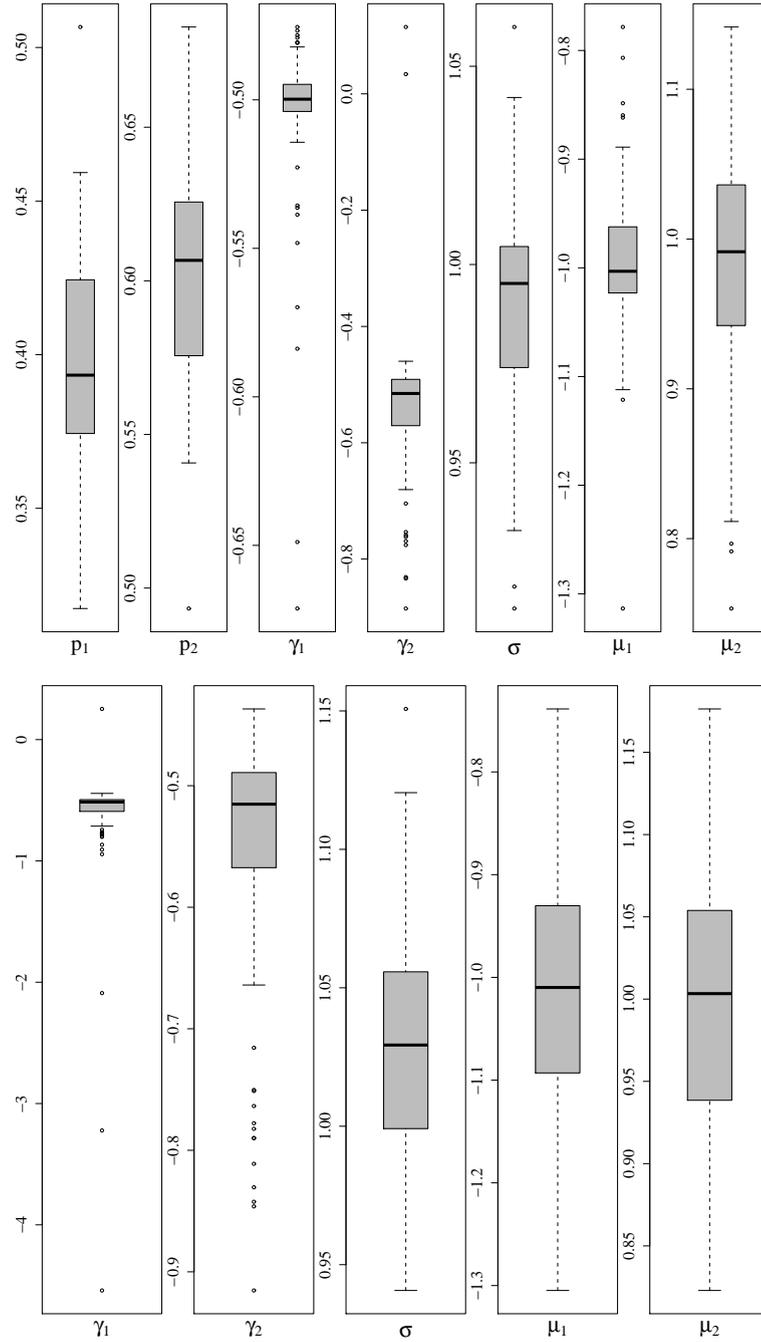


Figura 5: Gráficos *boxplot* das estimativas de  $\Theta_{3,1}$  para amostra não classificada e amostra classificada

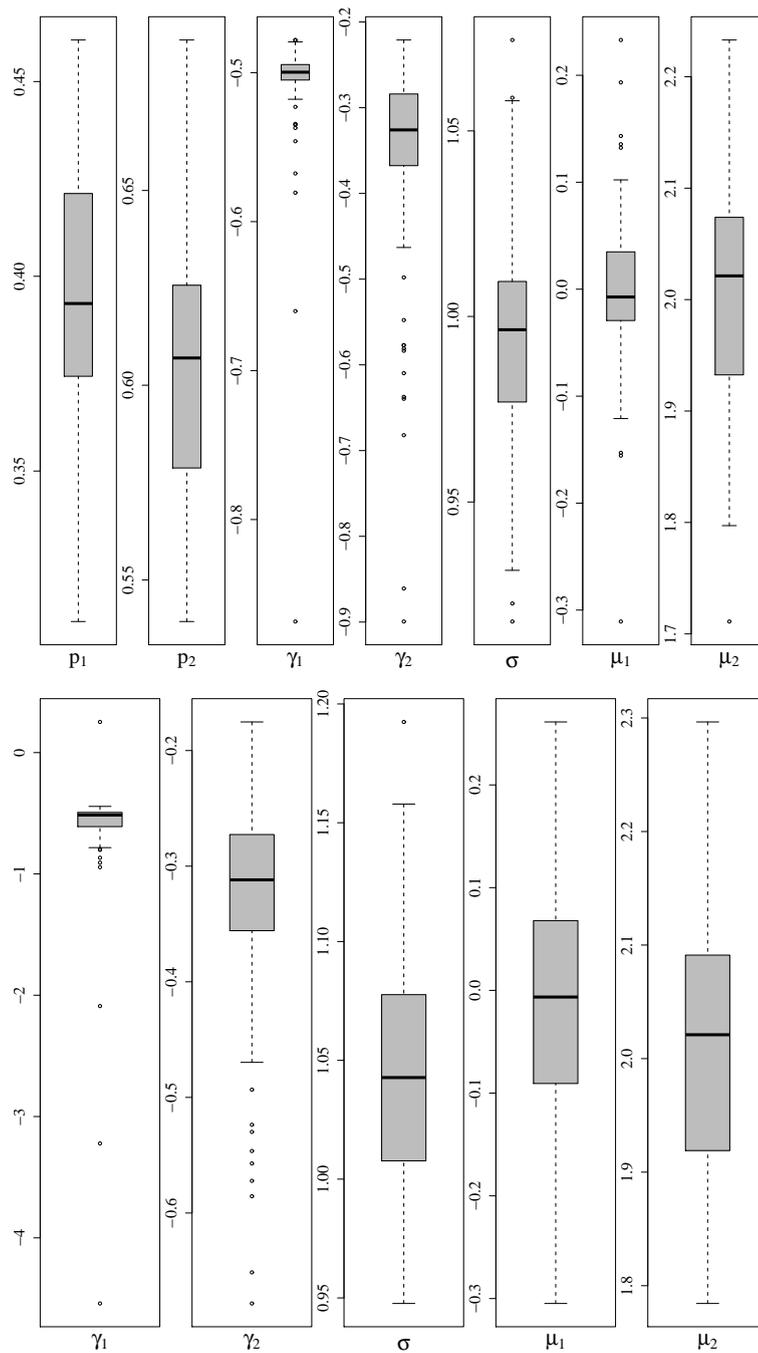


Figura 6: Gráficos *boxplot* das estimativas de  $\Theta_{3,2}$  para amostra não classificada e amostra classificada

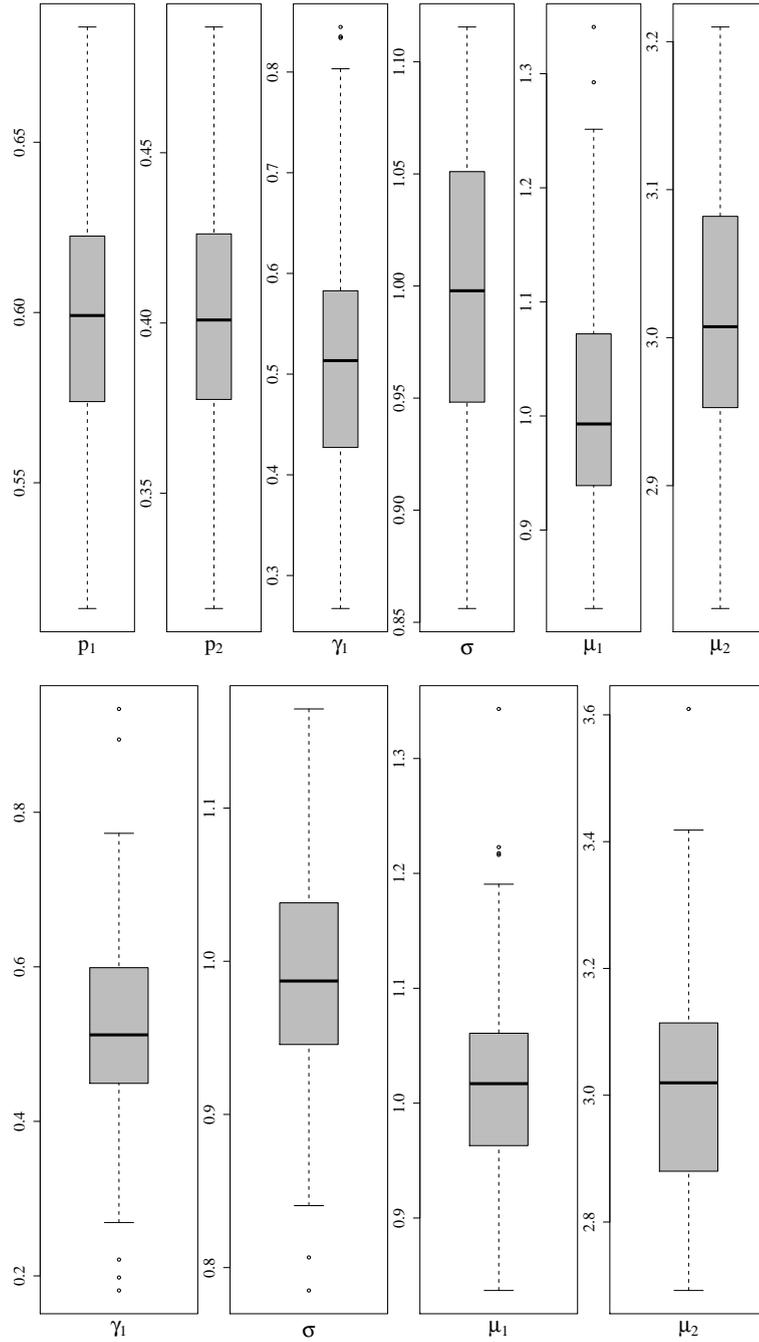


Figura 7: Gráficos *boxplot* das estimativas de  $\Theta_{4,1}$  para amostra não classificada e amostra classificada

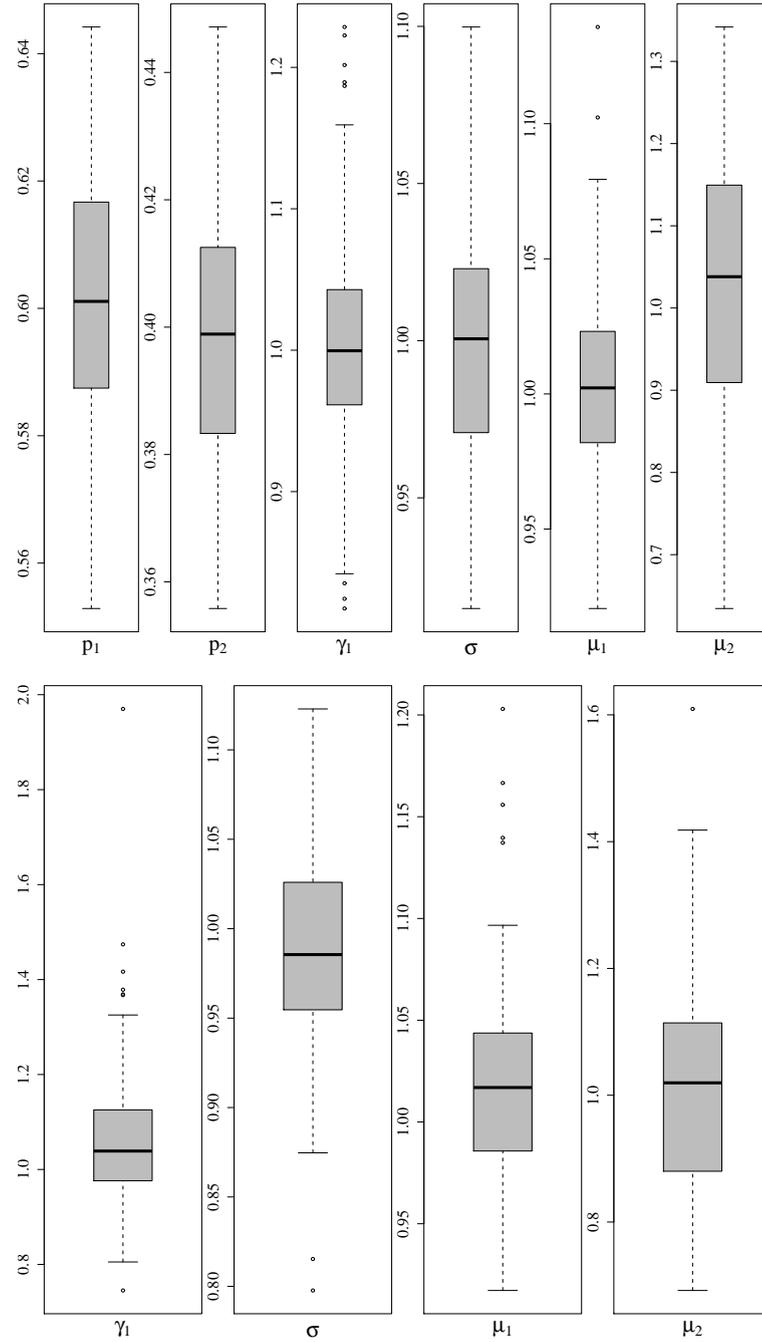


Figura 8: Gráficos *boxplot* das estimativas de  $\Theta_{4,2}$  para amostra não classificada e amostra classificada

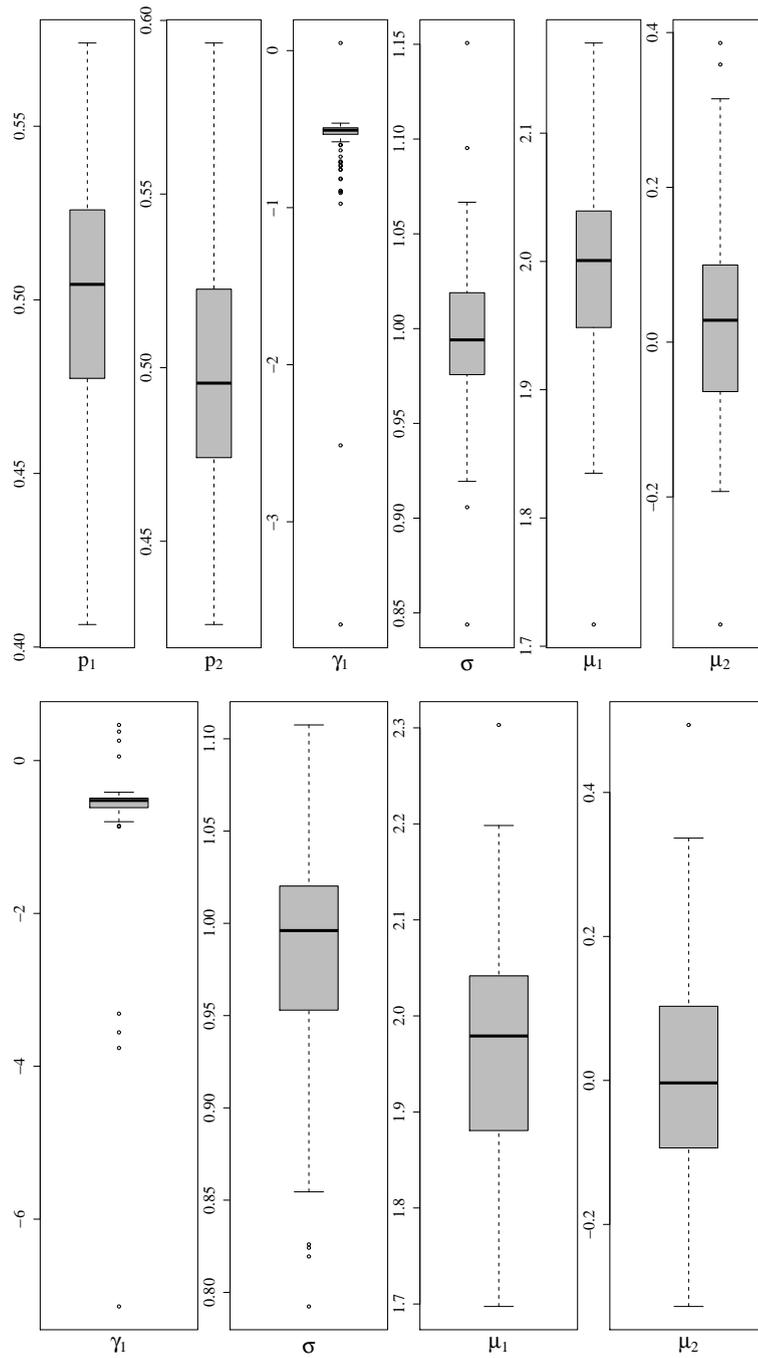


Figura 9: Gráficos *boxplot* das estimativas de  $\Theta_{5,1}$  para amostra não classificada e amostra classificada

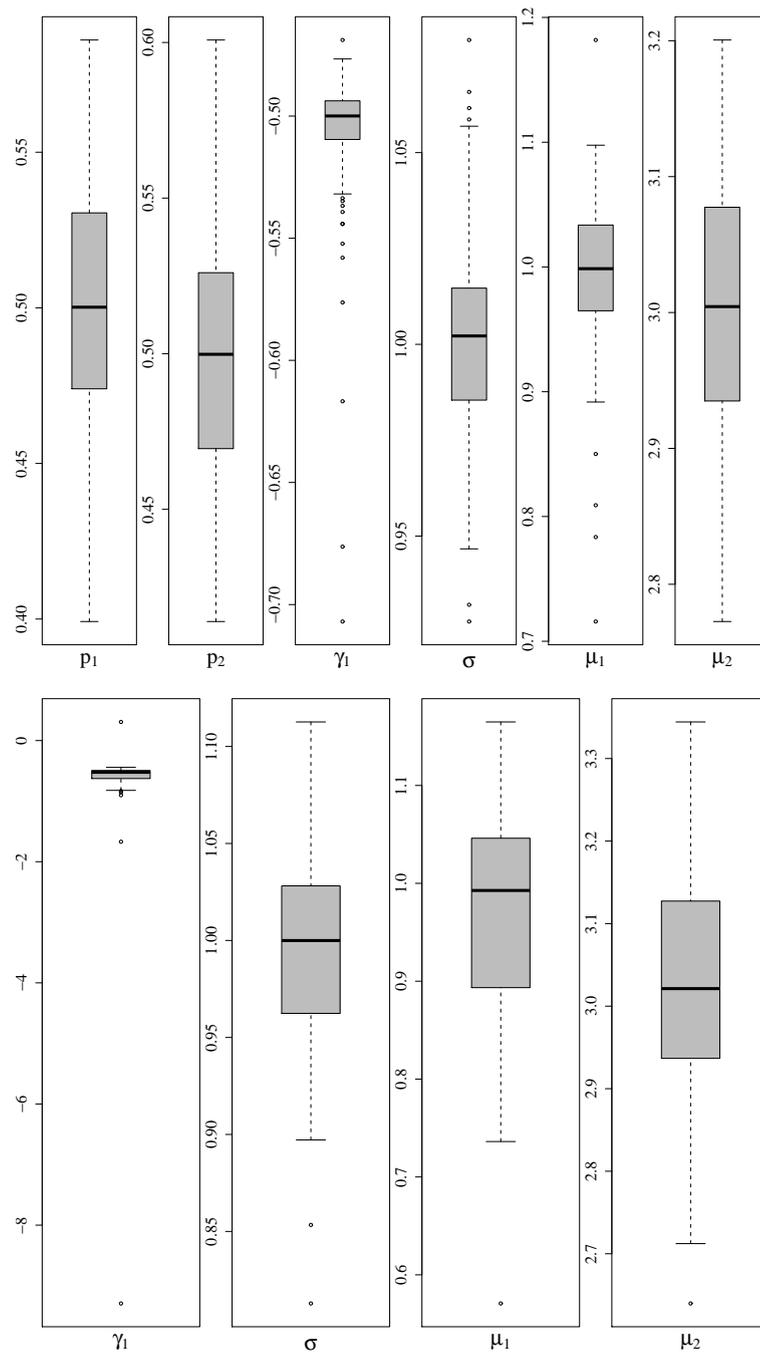


Figura 10: Gráficos *boxplot* das estimativas de  $\Theta_{5,2}$  para amostra não classificada e amostra classificada

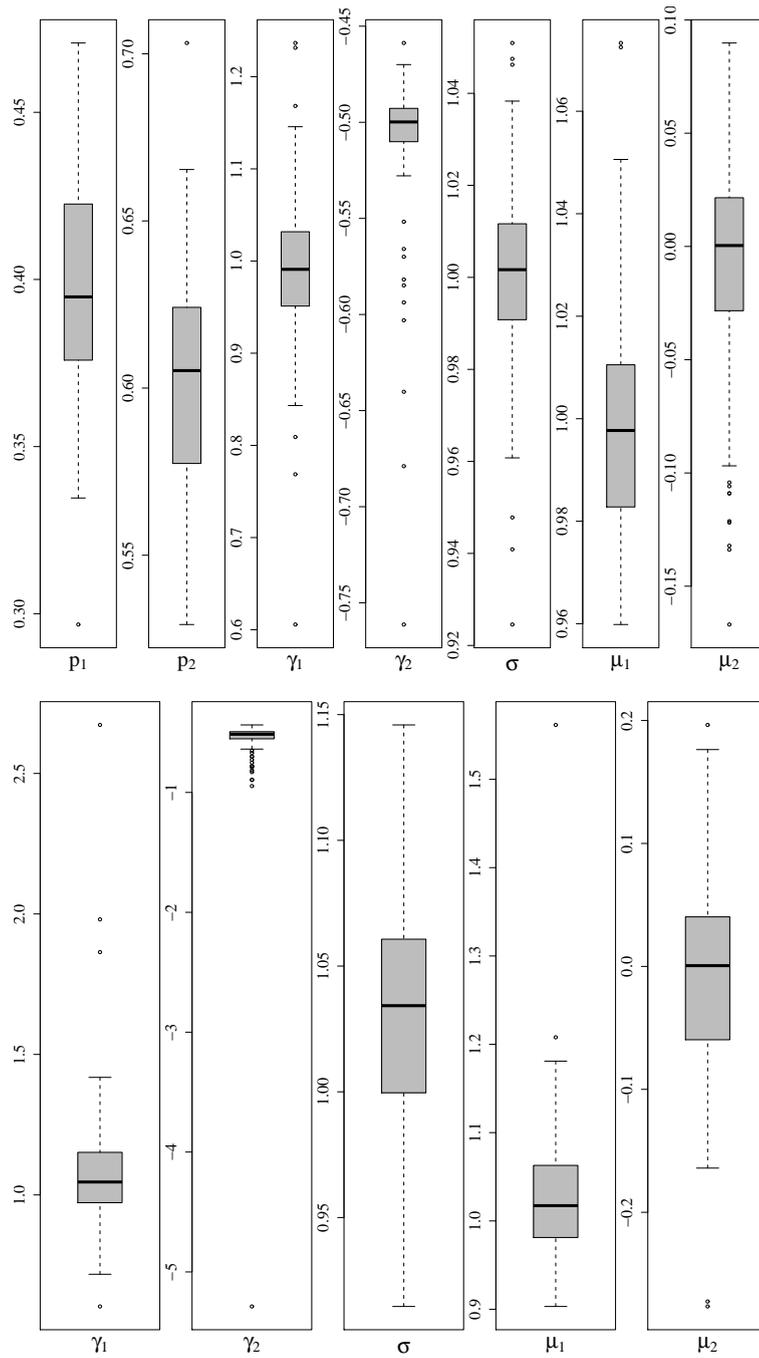


Figura 11: Gráficos *boxplot* das estimativas de  $\Theta_{6,1}$  para amostra não classificada e amostra classificada

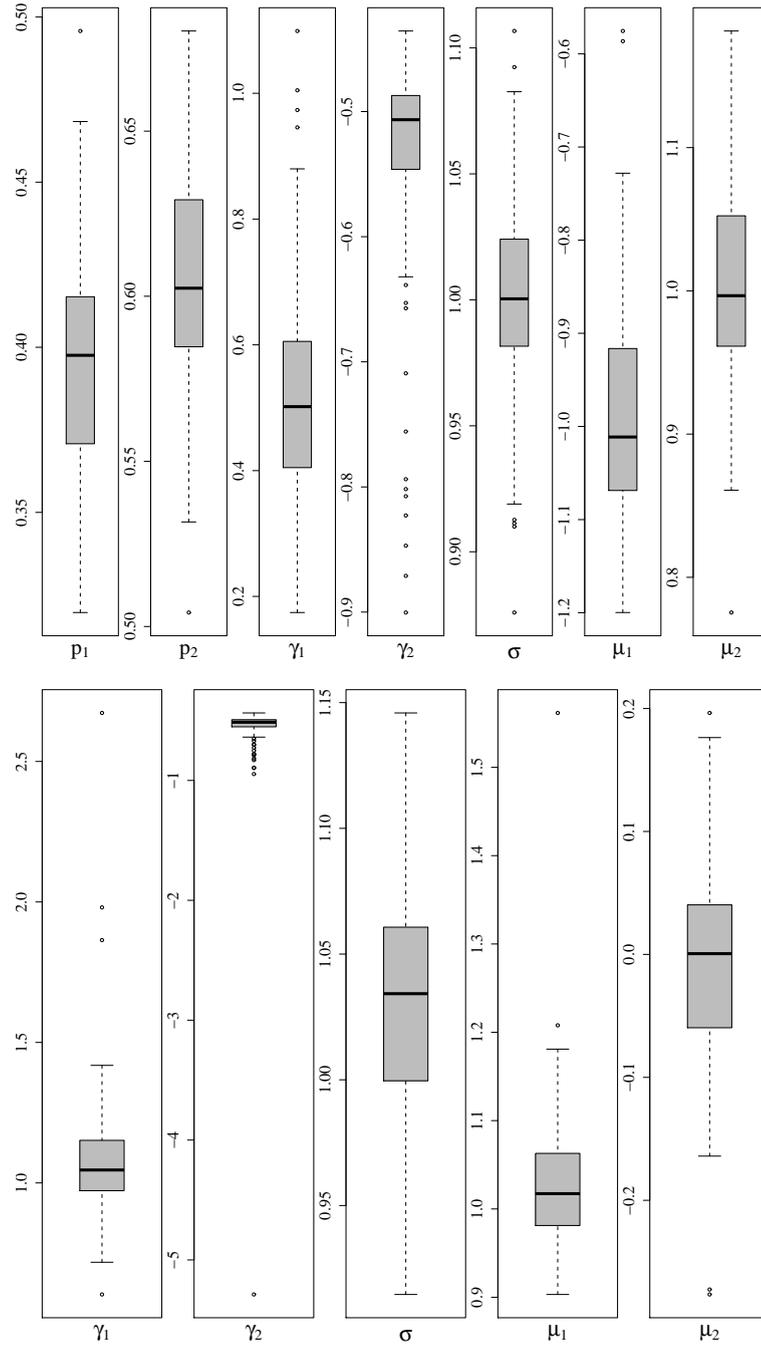


Figura 12: Gráficos *boxplot* das estimativas de  $\Theta_{6,2}$  para amostra não classificada e amostra classificada