

UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA  
DEPARTAMENTO DE ESTATÍSTICA

**Estimação paramétrica e semi-paramétrica em misturas  
uniforme-beta generalizada.  
Uma aplicação em dados de microarranjos.**

**Gabriel Coelho Gonçalves de Abreu**

Orientador: Prof. Dr. Aluísio de Souza Pinheiro

Dissertação apresentada junto ao Departamento de Estatística do Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas, para obtenção do Título de Mestre em Estatística.

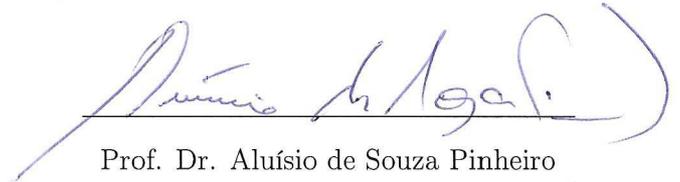
Campinas - SP

2007

**Estimação paramétrica e semi-paramétrica em misturas  
uniforme-beta generalizada.  
Uma aplicação em dados de microarranjos.**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Gabriel Coelho Gonçalves de Abreu e aprovada pela comissão julgadora.

Campinas, 19 de Janeiro de 2007.



Prof. Dr. Aluísio de Souza Pinheiro  
Orientador

Banca examinadora:

1. Prof. Dr. Aluísio de Souza Pinheiro (Orientador) - IMECC/UNICAMP
2. Profa. Dra. Gisela Tunes da Silva - IME/USP
3. Prof. Dr. Jesus Enrique Garcia - IMECC/UNICAMP

Dissertação apresentada junto ao Departamento de Estatística do Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Estatística.

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO IMECC DA UNICAMP**  
Bibliotecária: Maria Júlia Milani Rodrigues – CRB8a / 2116

Abreu, Gabriel Coelho Gonçalves de

Ab86e      Estimação paramétrica e semi-paramétrica em misturas uniforme-beta generalizada. Uma aplicação em dados de microarranjos / Gabriel Coelho Gonçalves de Abreu -- Campinas, [S.P.:s.n.], 2007.

Orientador: Aluísio de Souza Pinheiro

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Estimativa de parâmetro. 2. Estatística não paramétrica. 3. Bioestatística. I. Pinheiro, Aluísio de Souza . II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Parametric and semi-parametric estimation in uniform-generalized beta mixtures. An application in microarray data.

Palavras-chave em inglês (Keywords): 1. Parameter estimation. 2. Nonparametric statistics. 3. Biostatistics.

Área de concentração: Probabilidade e Estatística Aplicada

Titulação: Mestre em Estatística

Banca examinadora: Prof. Dr. Aluísio de Souza Pinheiro (IMECC-UNICAMP)

Profa. Dra. Gisela Tunes da Silva (IME-USP)

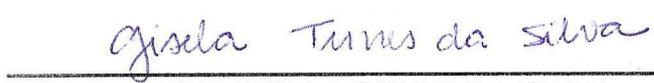
Prof. Dr. Jesus Enrique Garcia (IMECC-UNICAMP)

Data da defesa: 17/01/2007

Programa de Pós-Graduação: Mestrado em Estatística

Dissertação de Mestrado defendida em 17 de janeiro de 2007 e aprovada pela Banca Examinadora composta pelos Profs. Drs.

  
Prof (a). Dr (a). ALUÍSIO DE SOUZA PINHEIRO

  
Prof (a). Dr (a). GISELA TUNES DA SILVA

  
Prof (a). Dr (a). JESUS ENRIQUE GARCIA

Aos meus pais José Mauricio e Maria Luiza,  
e minha esposa Flavia.

# Agradecimentos

Agradeço aos meus pais, José Mauricio e Maria Luiza, por sempre acreditarem em mim, me apoiando em todas as escolhas. Ao meu irmão Gustavo. A minha esposa, Flavia, por seu constante bom humor, apoio e compreensão.

Ao meu orientador, Prof. Dr. Aluísio de Souza Pinheiro, por sua enorme atenção, interesse e grande paciência. Tenho certeza que sua experiência e conhecimento foram essenciais para este trabalho.

Aos professores do Imecc, em especial aos professores Dr. Mauro Sérgio de Freitas Marques, Dra. Nancy Lopes Garcia, Dr. Benjamin Bordin, Dr. Antônio Carlos Moretti, Dra. Hildete Prisco Pinheiro, Dr. Ronaldo Dias, Dra. Eliana Heiser de Freitas Marques e Dr. Filidor Edilson Vilca Labra, por seu interesse e atenção.

Aos amigos no Imecc, em especial a Tatiana Bordin pela essencial ajuda durante todo o curso, a Marina Canina pelo apoio, e também a Camila Estevam, Daniela Fonseca, Rafael Pimentel Maia, Tatiana Benaglia, Benilton, Helder Pataro, Rafael Moraes. Aos eternos amigos Renato Nardi Pedro, Fernando José Jammal, Fernanda Maria Vecchi.

Aos meus tios Marcelo e Nívia, Maria Lúcia e Sérgio, Maria Helena, meus primos Guilherme, Felipe e Rogério.

Ao Laboratório de Genoma Funcional do Cbmeg (Centro de Biologia Molecular e Engenharia Genética da Unicamp), especialmente ao Prof. Dr. Marcelo Menossi e ao Dr. Rodrigo Duarte Drummond, pela disponibilização de dados e orientações técnicas.

À banca examinadora, Profa. Dra. Gisela e Prof. Dr. Jesus.

Ao Imecc, Capes e Cenapad São Paulo (Centro Nacional e Processamento de Alto Desempenho em São Paulo) pelo apoio.

Acima de tudo, ao meu avô Arlindo, que tenho certeza está olhando por nós.

*“Mestre não é quem sempre ensina, mas quem de repente aprende.”*

João Guimarães Rosa

## *Resumo*

A análise de dados de expressão gênica tem sido de grande importância nas mais variadas áreas do desenvolvimento humano, como agricultura, melhoramento animal e medicina. Apesar dos avanços na área de estatística genética, a análise desse tipo de dados pode ser complexa e de difícil execução. Os investimentos já feitos nos últimos anos em pesquisa laboratorial podem levar a resultados concretos (melhoramento genético, vacinas genéticas, patentes) em pouco tempo, sob a correta interpretação dos resultados. Como a análise é feita em milhares de genes, existem problemas de comparações múltiplas, excedendo substancialmente o valor nominal de cada teste. Atualmente, em biologia, o problema de testes múltiplos se tornou uma norma, e não uma exceção. Assim, soluções sugeridas englobam o controle da taxa de erro, como o FDR (*False discovery rate*). O estudo da distribuição empírica dos p-valores, obtidos através dos testes estatísticos, pode ser realizado sob um modelo de mistura finita de distribuições beta. Sugere-se a utilização da distribuição beta generalizada com três parâmetros, mais flexível que a beta padrão. Faz-se um estudo da estimação paramétrica e semi-paramétrica no modelo proposto. São feitos estudos de simulação e aplicação a dados reais.

## *Abstract*

The analysis of gene expression data has been of great importance in many fields of human knowledge, as agriculture, animal breeding, and medicine. Despite the continuous progress of statistical genetics, the analysis of such data can be complex and of difficult evaluation. The investments done in the last years in laboratorial research can lead to important results in short time under the correct interpretation of data. As the analysis are done under a huge amount of data, multiple comparison problems are present, resulting in a reduction of the nominal confidence of each test. Nowadays, in biological and related fields, multiple testing problems has become a reality. Thus, among possible solutions is the control of error rates, such the FDR (False discovery rate). The estimation of the p-values distribution, obtained through statistical tests, can be evaluated by a finite mixture model of beta distributions. The use of generalized beta distribution with three parameters, a more flexible distribution, is suggested. The parametric and semi-parametric estimations are studied over the proposed model. Simulations and an application to real data are considered.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Distribuição Beta Generalizada</b>	<b>5</b>
2.1	Estimação . . . . .	7
2.2	Distribuição Beta Generalizada . . . . .	8
2.3	Momentos . . . . .	9
2.3.1	Função Hipergeométrica Gaussiana . . . . .	9
2.3.2	Momentos de $G3B$ . . . . .	11
2.3.3	Moda de $G3B$ . . . . .	12
2.3.4	Função de Distribuição Acumulada - $G3B$ . . . . .	13
2.3.5	Estimação dos Parâmetros . . . . .	13
<b>3</b>	<b>Arranjos de DNA</b>	<b>17</b>
3.1	Estrutura do DNA . . . . .	17
3.1.1	Fundamentos de Expressão Gênica . . . . .	18
3.1.2	Estrutura e Organização dos Genes . . . . .	20
3.2	Genoma Funcional . . . . .	22
3.2.1	Microarranjos . . . . .	23
3.2.2	Macroarranjos . . . . .	25
<b>4</b>	<b>Dados de DNA e Misturas de Densidades</b>	<b>27</b>

4.1	Considerações estatísticas sobre o estudo de expressão gênica em arranjos de DNA . . . . .	27
4.2	Taxa de Falsos Positivos . . . . .	29
4.3	Misturas de Densidades . . . . .	34
4.3.1	Número de Componentes . . . . .	35
4.3.2	Interpretação de Modelos de Misturas . . . . .	37
4.3.3	Identificabilidade . . . . .	39
<b>5</b>	<b>Estimação Paramétrica em Misturas de Densidades</b>	<b>41</b>
5.1	Estimação de máxima verossimilhança . . . . .	43
5.2	Métodos de Busca . . . . .	48
5.2.1	Busca direta . . . . .	48
5.2.2	Busca de Fibonacci . . . . .	48
5.3	Método <i>EM</i> . . . . .	50
5.4	Métodos numéricos . . . . .	54
5.4.1	Método de Newton-Raphson . . . . .	54
5.4.2	Método de Davidon-Fletcher-Powell . . . . .	56
<b>6</b>	<b>Uso do SiZer para Estimação Semi-paramétrica em Misturas de Densidades</b>	<b>59</b>
6.1	Estimação de densidades através de função núcleo . . . . .	59
6.1.1	Propriedades Assintóticas . . . . .	60
6.2	SiZer . . . . .	62
6.2.1	Desenvolvimento do SiZer . . . . .	65
6.2.2	Implementação numérica . . . . .	66
6.3	Convergência . . . . .	68
<b>7</b>	<b>Simulação e Aplicação</b>	<b>71</b>
7.1	Simulação de modelos de misturas . . . . .	71

<i>SUMÁRIO</i>	xix
7.2 Aplicação a dados reais . . . . .	74
<b>8 Considerações Finais</b>	<b>81</b>
<b>A Simulação</b>	<b>83</b>
Índice Onomástico . . . . .	125
Índice Remissivo . . . . .	128

# Lista de Figuras

2.0.1 Fdp beta padrão (2.0.1) (Johnson et al., 1995a). . . . .	8
2.2.1 Fdp beta padrão (2.0.1) (A) e G3B (2.2.1) (B). . . . .	10
3.1.1 Parte de uma dupla hélice de DNA e a respectiva seqüência linear codificada (Thompson et al., 1993). . . . .	17
3.1.2 Esquema simplificado da codificação protéica (Thompson et al., 1993). . .	19
3.1.3 Fluxo de informações DNA → RNA para a proteína de um gene hipotético. As etapas incluem a transcrição, processamento e emenda do RNA, trans- porte do RNA do núcleo ao citoplasma e tradução (Thompson et al., 1993).	19
3.1.4 Estrutura de um gene humano típico (Thompson et al., 1993). . . . .	20

- 3.2.1 Para preparar o arranjo, fragmentos de DNA, cada um correspondendo a um gene, são fixados sobre a lâmina por um robô. Neste exemplo, o mRNA é coletado de dois tipos diferentes de células para uma comparação direta de seus níveis de expressão. As amostras são convertidas em cDNA e marcadas, uma com fluorocromo vermelho e outra com fluorocromo verde. As amostras marcadas são misturadas e então aplicadas sobre a lâmina. Após incubação, o arranjo é lavado e a fluorescência registrada. Na porção mostrada do microarranjo, que representa a expressão de 110 genes de levedura, “spots” vermelhos indicam que o gene é mais expresso na amostra 1 que o correspondente gene na amostra 2; “spots” verdes indicam que ocorre o inverso. “Spots” escuros indicam que ocorre pouca ou nenhuma expressão em ambas amostras. As diferentes intensidades de cores representam, então, o nível de expressão em cada amostra (Alberts et al., 2002). . . . . 24
- 4.2.1 Região A corresponde à ocorrência de positivos verdadeiros; região B corresponde à ocorrência de falsos negativos; região C corresponde à ocorrência de falsos positivos; e a região D corresponde à ocorrência de negativos verdadeiros (Pounds & Morrins, 2003). . . . . 32
- 5.2.1 Etapa da busca de Fibonacci (Everitt, 1987). . . . . 50
- 6.1.1 Estimativa e núcleos individuais com  $h = 0,4$  (A),  $h = 0,2$  (B) e  $h = 0,8$  (C) (Silverman, 1993). . . . . 60
- 6.2.1 O painel (A) mostra a densidade real da mistura. Combinação das estimativas de densidade por função núcleo (B) e o mapa gerado pelo SiZer (C), usando nível de significância de  $\alpha = 0,05$ , também são apresentados. A janela com  $h = 0,4449$  é destacada. As linhas pontilhadas no mapa SiZer mostram tamanhos de janela efetivos para cada  $h$ , com intervalos representando  $\pm 2h$  (ou seja,  $\pm 2h$  desvios padrões do núcleo gaussiano). . . 63

7.1.1 Misturas usadas na simulação, com  $p = 0,95$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  
 (1,1;3,0;2,0)(A); (1,1;5,0;2,0)(B); (1,1;5,0;5,0)(C); (1,1;2,0;5,0)(D); (2,0;1,9;5,0)(E);  
 (2,0;2,0;6,1)(F); (2,1;2,0;5,0)(G); (1,1;4,0;1,3)(H); (1,1;1,2;4,3)(I); (1,1;4,2;4,3)(J);  
 (0,4;5,0;2,5)(K); (0,6;1,2;4,3)(L). . . . . 72

7.2.1 Histogramas dos p-valores analisados: tempo 1 (A), tempo 2 (B), tempo 3  
 (C) e tempo 4 (D). . . . . 75

7.2.2 Densidades estimadas segundo o método de busca: tempo 1 (A), tempo 2  
 (B), tempo 3 (C) e tempo 4 (D). . . . . 77

7.2.3 Densidades estimadas segundo o método de Fibonacci: tempo 1 (A), tempo  
 2 (B), tempo 3 (C) e tempo 4 (D). . . . . 78

7.2.4 Densidades estimadas segundo o SiZer: tempo 1 (A), tempo 2 (B), tempo  
 3 (C) e tempo 4 (D). . . . . 79

A.0.1 Misturas usadas na simulação, com  $p = 0,25$  e parâmetros  $(\alpha; \beta; \lambda)$   
 sendo: (1,1;3,0;2,0)(A); (1,1;5,0;2,0)(B); (1,1;5,0;5,0)(C); (1,1;2,0;5,0)(D);  
 (2,0;1,9;5,0)(E); (2,0;2,0;6,1)(F); (2,1;2,0;5,0)(G); (1,1;4,0;1,3)(H); (1,1;1,2;4,3)(I);  
 (1,1;4,2;4,3)(J); (0,4;5,0;2,5)(K); (0,6;1,2;4,3)(L). . . . . 83

A.0.2 Misturas usadas na simulação, com  $p = 0,50$  e parâmetros  $(\alpha; \beta; \lambda)$   
 sendo: (1,1;3,0;2,0)(A); (1,1;5,0;2,0)(B); (1,1;5,0;5,0)(C); (1,1;2,0;5,0)(D);  
 (2,0;1,9;5,0)(E); (2,0;2,0;6,1)(F); (2,1;2,0;5,0)(G); (1,1;4,0;1,3)(H); (1,1;1,2;4,3)(I);  
 (1,1;4,2;4,3)(J); (0,4;5,0;2,5)(K); (0,6;1,2;4,3)(L). . . . . 84

A.0.3 Misturas usadas na simulação, com  $p = 0,70$  e parâmetros  $(\alpha; \beta; \lambda)$   
 sendo: (1,1;3,0;2,0)(A); (1,1;5,0;2,0)(B); (1,1;5,0;5,0)(C); (1,1;2,0;5,0)(D);  
 (2,0;1,9;5,0)(E); (2,0;2,0;6,1)(F); (2,1;2,0;5,0)(G); (1,1;4,0;1,3)(H); (1,1;1,2;4,3)(I);  
 (1,1;4,2;4,3)(J); (0,4;5,0;2,5)(K); (0,6;1,2;4,3)(L). . . . . 85

A.0.4 Misturas usadas na simulação, com  $p = 0,75$  e parâmetros  $(\alpha; \beta; \lambda)$   
 sendo: (1,1;3,0;2,0)(A); (1,1;5,0;2,0)(B); (1,1;5,0;5,0)(C); (1,1;2,0;5,0)(D);  
 (2,0;1,9;5,0)(E); (2,0;2,0;6,1)(F); (2,1;2,0;5,0)(G); (1,1;4,0;1,3)(H); (1,1;1,2;4,3)(I);  
 (1,1;4,2;4,3)(J); (0,4;5,0;2,5)(K); (0,6;1,2;4,3)(L). . . . . 86

- A.0.5 Misturas usadas na simulação, com  $p = 0,80$  e parâmetros  $(\alpha; \beta; \lambda)$   
sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  
 $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  
 $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L). . . . . 87
- A.0.6 Misturas usadas na simulação, com  $p = 0,85$  e parâmetros  $(\alpha; \beta; \lambda)$   
sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  
 $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  
 $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L). . . . . 88
- A.0.7 Misturas usadas na simulação, com  $p = 0,90$  e parâmetros  $(\alpha; \beta; \lambda)$   
sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  
 $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  
 $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L). . . . . 89
- A.0.8 Misturas usadas na simulação, com  $p = 0,97$  e parâmetros  $(\alpha; \beta; \lambda)$   
sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  
 $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  
 $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L). . . . . 90

# Lista de Tabelas

2.3.1 Exemplos da inversa da Informação de Fisher ( $-I^{-1}$ ). . . . .	16
3.1.1 Codificação dos códons . . . . .	21
4.2.1 Número de erros cometidos quando testadas $m$ hipóteses nulas. . . . .	29
5.2.1 Taxa de redução no intervalo de busca para diferentes $n$ 's (Everitt, 1987). . . . .	49
7.1.1 Combinações de parâmetros usadas na simulação. . . . .	74
7.2.1 Estimativas dos parâmetros para os dados reais. . . . .	76
7.2.2 FDR para dados fosfato. A coluna <i>Método</i> identifica o método de estimacão, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci) e SiZer (estimacão através do SiZer). . . . .	76
A.0.1 $p=0,25$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimacão, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimacão através do SiZer) e EM (Algoritmo EM). . . . .	91
A.0.2 $p=0,25$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimacão, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimacão através do SiZer) e EM (Algoritmo EM). . . . .	92

A.0.3	$p=0,50$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	93
A.0.4	$p=0,50$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	94
A.0.5	$p=0,70$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	95
A.0.6	$p=0,70$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	96
A.0.7	$p=0,75$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	97
A.0.8	$p=0,75$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	98
A.0.9	$p=0,80$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	99
A.0.10	$p=0,80$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	100

A.0.11	$p=0,85$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	101
A.0.12	$p=0,85$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	102
A.0.13	$p=0,90$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	103
A.0.14	$p=0,90$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	104
A.0.15	$p=0,95$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	105
A.0.16	$p=0,95$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	106
A.0.17	$p=0,97$ . Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	107
A.0.18	$p=0,97$ . (Quartil1) Mediana (Quartil3). A coluna <i>Método</i> identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).	108

A.0.19	Número de NaN e Inf em EM. . . . .	109
A.0.20	Distância Hellinger média. . . . .	110
A.0.21	Distância Hellinger (desvios). . . . .	111
A.0.22	FDR - Média (Desvio padrão). A coluna <i>Método</i> identifica o método de estimaco, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimaco atravs do SiZer) e EM (Mtodo EM). . .	112
A.0.23	FDR - (Q1) (Mediana) (Q3). A coluna <i>Mtodo</i> identifica o mtodo de estimaco, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimaco atravs do SiZer) e EM (Mtodo EM). . .	114

# Capítulo 1

## Introdução

A análise de expressão gênica através de arranjos de DNA representa um grande avanço na tecnologia genética, pela qual é possível analisar milhares de genes em um único conjunto de experimentos a um custo razoável. Entretanto, devido à enorme quantidade de dados e sutil natureza da expressão gênica no uso de uma técnica que ainda não se apresenta totalmente controlada, a análise e incerteza das decisões recaem sobre a teoria e modelos estatísticos. O problema fica cada vez mais complexo com o aumento no número de genes estudados. Nesse caso, milhares de genes são testados contra um conjunto de hipóteses nulas, esperando-se que um subconjunto dos mesmos seja significativo. Isso resulta em problemas de comparações múltiplas, excedendo substancialmente o valor nominal de cada teste. Soluções sugeridas englobam o controle da taxas de erro, como o FWER (*Familywise error rate*) e a FDR (*False discovery rate*).

Sendo de interesse, no caso de estudos de genoma funcional por macro e microarranjos, selecionar aqueles genes potencialmente mais interessantes para análises posteriores, o custo causado pelas incorretas seleções pode tornar-se proibitivo. Assim, a literatura recente mostra uma tendência na análise da FDR para o controle do número de falsos positivos. O objetivo é identificar o máximo de genes significantes, implicando em uma baixa proporção relativa de falsos positivos. Como os testes são realizados em cada gene, os valores de corte tradicionais para os p-valores (0,01 e 0,05) devem ser corrigidos para

evitar uma abundância de falsos positivos. Dessa forma, se nenhum gene é diferentemente expresso e a FDR é controlado em algum nível  $q$ , então a probabilidade de detecção errônea de algum gene diferentemente expresso é menor do que ou igual à  $q$ . Exemplos de aplicação prática para a FDR podem ser encontrados em Norris & Kahn (2006) e Unger et al. (2005).

O corrente interesse nesse tipo de medida torna disponível na literatura inúmeros artigos descrevendo métodos de estimação, análises e interpretações possíveis. Entre eles é possível citar Wu, Guan & Zhao (2006), Jung & Jang (2006) e Li et al. (2005).

Os genes com p-valores pequenos são selecionados como genes responsáveis pela diferenciação da expressão do RNA entre os tratamentos estudados. Uma questão de interesse é qual deve ser o valor de corte para considerar um p-valor como pequeno. A literatura recente mostra que a FDR é um critério poderoso e razoável para selecionar tais genes (Yang & Yang, 2006).

O interesse principal do trabalho é considerar situações em que os testes estatísticos sejam bem definidos, a cada um dos milhares de genes testados, avaliando uma hipótese nula contra uma hipótese alternativa. O objetivo é propor uma estimativa da significância para cada gene, de maneira que a mesma tenha interpretação na avaliação de testes simultâneos.

A estimação da distribuição dos p-valores, obtidos através dos testes estatísticos, pode ser realizada sob um modelo de mistura finita de distribuições beta (Parker & Rothenberg, 1988). Uma classe muito rica de densidades é a classe beta generalizada de três parâmetros ( $G3B$ ) definida por Libby & Novick (1982). Pela inclusão de um terceiro parâmetro, de deformação, a distribuição torna-se mais flexível, ampliando o espaço em que a mesma pode assumir média, variância, assimetria e curtose (Chen et al., 1984; Pham-Gia & Duong, 1989).

No modelo beta-uniforme usado na análise de dados de microarranjo, o parâmetro  $p$  pode ser interpretado como a proporção de genes que são diferencialmente expressos. A vantagem do modelo é que qualquer teste que produza um p-valor válido pode ser usado.

O teste-t usual é o mais comum, mas métodos não-paramétricos e o bootstrap também já foram propostos (Neuhauser & Lam, 2004; Tusher et al., 2001). O número de observações (p-valores) usadas e a proximidade dos parâmetros dos limites do espaço paramétrico têm grande importância no desempenho dos estimadores, assim como a proximidade da componente beta da componente uniforme. O número de testes usados num experimento de microarranjos é importante na precisão das estimativas dos parâmetros a partir do momento que esse tipo de experimento pode envolver entre 1.000 e 60.000 testes (Xiang et al., in press).

Foram realizados estudos de simulação em estimação dos parâmetros da mistura uniforme-beta generalizada. Selecionou-se conjuntos de parâmetros que aproximassem a distribuição dos dados, assim como outras configurações da mistura.

Um dos problemas encontrados na estimação de misturas refere-se à detecção de modas da mistura. Nessa área, uma aplicação muito interessante foi introduzida por Chaudhuri & Marron (1999), denominada **SiZer** (*SIgnificant ZERo crossings of derivatives*). A idéia é destacar características significativas, como modas, mostrando locais onde a curva cresce ou decresce significativamente. O estudo é feito sobre estimativas da densidade através de função núcleo.

O modelo foi aplicado a dados reais gerados pelo Laboratório de Genoma Funcional do Cbmeg (Centro de Biologia Molecular e Engenharia Genética da Unicamp), sob orientação do Prof. Dr. Marcelo Menossi. Os dados são referentes a um experimento com cana-de-açúcar cultivada sob condições de estresse biológico, observadas em quatro tempos diferentes.

## Capítulo 2

# Distribuição Beta Generalizada

A distribuição beta é de grande importância na estatística, com muita versatilidade, o que leva a inúmeras aplicações tanto em Estatística Clássica quanto Bayesiana (Pham-Gia & Turkkan, 1998).

Segundo Johnson (1995a), a família de distribuições beta é composta por todas as distribuições com função densidade de probabilidade (fdp) da forma:

$$f(y) = \frac{1}{B(p, q)} \frac{(y - a)^{p-1} (b - y)^{q-1}}{(b - a)^{p+q-1}}, \quad a \leq y \leq b,$$

com  $p > 0, q > 0$ , denominada por  $Beta(p, q)$ , sendo (Pham-Gia & Turkkan, 1998; Abramowitz & Stegun, 1972)

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)},$$

com

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha-1} dt.$$

Se  $q = 1$ , a distribuição é chamada uma distribuição “power-function”.

Com a transformação

$$X = \frac{Y - a}{b - a},$$

obtém-se a fdp

$$f(x) = \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1}, \quad 0 \leq x \leq 1. \quad (2.0.1)$$

Essa é a *forma padrão* (ou *central*) da distribuição beta (Tipo I) com parâmetros  $p$  e  $q$ , e representa a distribuição da razão  $X_1/(X_1 + X_2)$  em que  $X_i$  ( $i = 1, 2$ ) são variáveis aleatórias independentes com distribuição gama, sendo  $X_1 \sim \text{Gama}(p, \eta)$  e  $X_2 \sim \text{Gama}(q, \eta)$ . A “power-function” padrão é dada por

$$f(x) = px^{p-1}, \quad 0 \leq x \leq 1.$$

Harter (1978) introduziu a família beta simétrica ( $p = q$ ) padronizada, de fdp

$$f_X(x) = \left[ \frac{\Gamma(2p)}{\Gamma^2(p)(2\sqrt{2p+1})^{2p-1}} \right] (2p+1-x^2)^{(p-1)}, \\ -\sqrt{2p+1} \leq x \leq \sqrt{2p+1}.$$

É claro que  $\mathbb{E}[X] = 0$  e  $\text{Var}[X] = 1$ . A fdp de uma distribuição beta simétrica de parâmetro  $p$ , média  $\mu$  e desvio padrão  $\sigma$  é dada por

$$f_X(x) = \frac{\Gamma(2p)}{\sigma[\Gamma(p)]^2(2\sqrt{2p+1})^{2p-1}} \left[ 2p+1 - \left( \frac{x-\mu}{\sigma} \right)^2 \right]^{(p-1)},$$

$$\mu - \sigma\sqrt{2p+1} \leq x \leq \mu + \sigma\sqrt{2p+1}.$$

A função distribuição acumulada (fda) da forma padrão, até  $x$ , é dada pela razão da função beta incompleta denotada por  $I_x(p, q)$ , da forma (Johnson et al., 1995a)

$$I_x(p, q) = \frac{1}{B(p, q)} \int_0^x t^{p-1} (1-t)^{q-1} dt. \quad (2.0.2)$$

Ainda sobre a distribuição padrão, tem-se que

$$\mathbb{E}[X^r] = \frac{B(p+r, q)}{B(p, q)} = \frac{\Gamma(p+r)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+r)} = \frac{p^{[r]}}{(p+q)^{[r]}},$$

ocorrendo a última igualdade caso  $r$  seja inteiro, sendo  $y^{[r]} = y(y+1)\dots(y+r-1)$ .

Quando  $p > 1$  e  $q > 1$ , a fdp é unimodal em  $x = \frac{p-1}{p+q-2}$ , e  $\text{Var}[X] < 1/12$ . Caso

$p < 1$  e  $q < 1$ , então a fdp apresenta a forma de “U”, com antimoda em  $x = \frac{p-1}{p+q-2}$  e  $Var[X] > 1/12$ . Se  $(p-1)(q-1) \leq 0$  então a fdp não possui moda ou antimoda, apresentando a forma de “J” ou “J” reverso (Johnson et al., 1995a; Pham-Gia & Turkkan, 1998).

Fazendo  $\lambda = \frac{1}{p+q}$  e  $\theta = \frac{p}{p+q}$ , decorre a seguinte relação de recorrência entre os momentos centrais:

$$\mathbb{E}[(X - \mathbb{E}(X))^{s+1}] = \mu_{s+1} = \frac{-s\lambda}{1+s\lambda}\mu_s + \theta \sum_{j=1}^s \binom{s}{j} \frac{\lambda^j (1-\theta)^j j!}{(1+s\lambda)\dots(1+[s-j]\lambda)} \mu_{s-j},$$

sendo que  $\mu_0 = 1$  e  $\mu_1 = 0$  (Johnson et al., 1995a).

Para todo  $p, q > 0$  existem pontos de inflexão em  $\frac{p-1}{p+q-2} \pm \frac{1}{p+q-2} \sqrt{\frac{(p-1)(q-1)}{p+q-3}}$  desde que estes valores sejam reais e entre 0 e 1. Note que os pontos são equidistantes das modas (quando as mesmas existem) (Johnson et al., 1995a).

Quando  $p$  e  $q$  são trocados de posição, ou seja, define-se  $B(q, p)$ , a distribuição é refletida em torno de  $1/2$ , o que pode ser visto na Figura 2.0.1 (Johnson et al., 1995a).

A estimação dos parâmetros não pode ser obtida por máxima verossimilhança diretamente, havendo necessidade do uso de métodos numéricos (Johnson et al., 1995a).

## 2.1 Estimação

Johnson (1995a) afirma que as dificuldades envolvidas na estimação por máxima verossimilhança estão relacionadas ao emprego de técnicas de busca eficientes para sua maximização. O método de Newton-Raphson é extremamente sensível aos valores iniciais de  $\tilde{p}$  e  $\tilde{q}$ , não havendo garantias de convergência. O autor cita trabalhos em que o método dos momentos obteve melhores resultados que o de máxima verossimilhança, provavelmente devido à erros computacionais atribuídos ao método de determinação da verossimilhança.

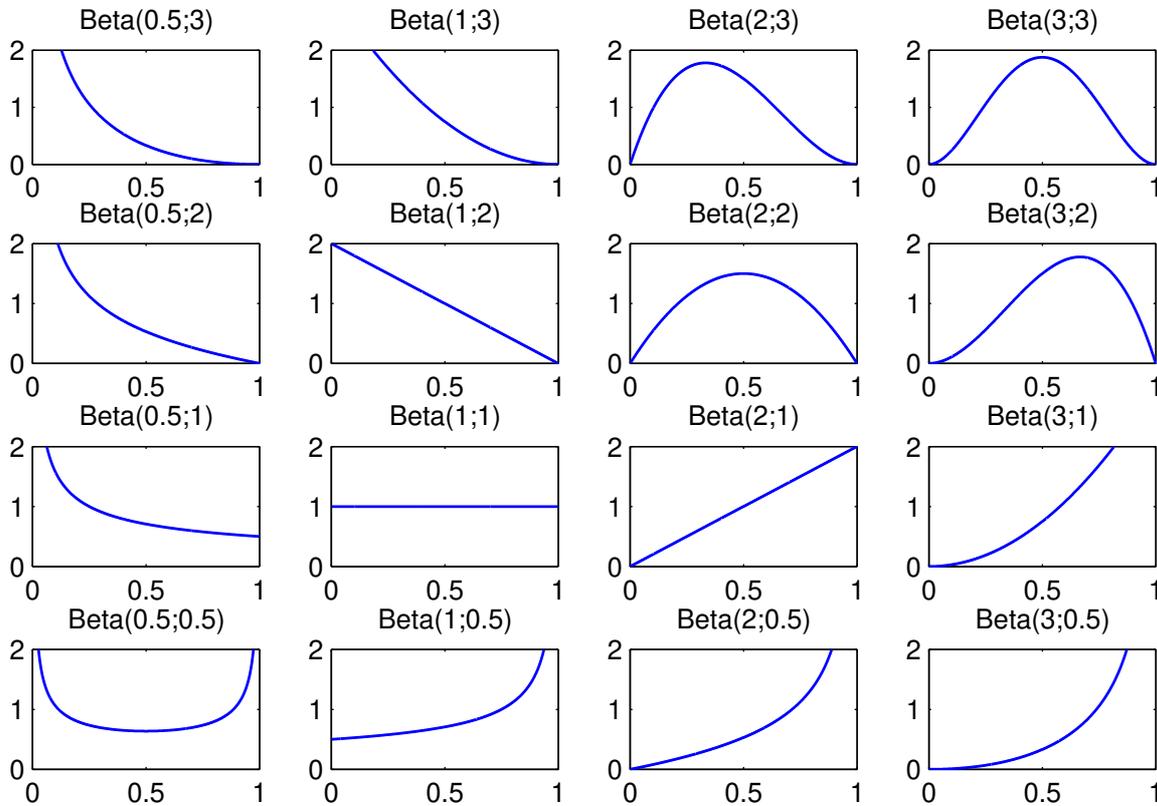


Figura 2.0.1: Fdp beta padrão (2.0.1) (Johnson et al., 1995a).

## 2.2 Distribuição Beta Generalizada

Uma classe muito rica de densidades é a classe beta unidimensional de três parâmetros definida por Libby & Novick (1982). Pela inclusão de um único parâmetro de deformação, a distribuição torna-se mais flexível, ampliando o espaço em que a mesma pode assumir média, variância, assimetria e curtose (Chen et al., 1984).

Pham-Gia & Duong (1989) investigou a distribuição beta generalizada com três parâmetros,  $G3B(\alpha, \beta, \lambda)$  com fdp

$$f_Y(y|\alpha, \beta, \lambda) = \frac{\lambda^\alpha y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta) [1 - (1-\lambda)y]^{\alpha+\beta}}, \quad 0 \leq y \leq 1; \quad \alpha, \beta, \lambda > 0, \quad (2.2.1)$$

que é a distribuição da razão  $X_1/(X_1 + X_2)$  em que  $X_i$  ( $i = 1, 2$ ) são independentes com distribuição gama, sendo  $X_1 \sim \text{Gama}(\alpha, \eta_1)$  e  $X_2 \sim \text{Gama}(\beta, \eta_2)$ , em que  $\lambda = \eta_1/\eta_2$ . Se  $Y \sim G3B(\alpha, \beta, \lambda)$  então  $(1 - Y) \sim G3B(\beta, \alpha, \lambda^{-1})$ .

A presença do parâmetro  $\lambda$  permite a  $G3B$  assumir uma variedade de formas maior que a distribuição beta padrão, o que pode ser comparado na Figura 2.2.1. Em geral, quando  $0 < \lambda < 1$ , a fdp de  $G3B(\alpha, \beta, \lambda)$  está abaixo da fdp da distribuição padrão, mas cruza a última, tornando-se maior, em  $y_0 = [1 - \lambda^{\alpha/(\alpha+\beta)}]^{-1} - (1 - \lambda)^{-1}$ . Para  $\lambda > 1$ , o contrário é verdadeiro, cruzando no mesmo ponto.

## 2.3 Momentos

### 2.3.1 Função Hipergeométrica Gaussiana

A função hipergeométrica gaussiana, ou apenas função hipergeométrica, possui a seguinte forma:

$$\begin{aligned} {}_2F_1(a, b; c; x) &= 1 + \frac{ab}{c}x + \frac{a(a+1)b(b+1)}{c(c+1)2!}x^2 + \dots = \\ &= \sum_{j=0}^{\infty} \frac{(a)_j(b)_j x^j}{(c)_j j!}, \quad c \neq 0, -1, -2, \dots, \end{aligned} \quad (2.3.1)$$

em que  $(a)_j$  é o símbolo de Pochhammer,

$$(n)_k = n(n+1)\dots(n+k-1) = \frac{\Gamma(n+k)}{\Gamma(n)}, \quad k \geq 1,$$

sendo que  $(n)_0 = 1$ . É fácil ver que  ${}_2F_1(a, b; c; x) = {}_2F_1(b, a; c; x)$  (Johnson et al., 1995a).

Os casos de interesse são quando  $a$ ,  $b$ ,  $c$  e  $x$  são reais. Se  $a$  é um inteiro não-positivo, então  $(a)_j$  é zero para  $j > -a$ , e a série é finita. Quando a série é infinita, é absolutamente convergente para  $|x| < 1$  e divergente para  $|x| > 1$ . Para  $|x| = 1$ , é

1. absolutamente convergente se  $c - a - b > 0$ ,
2. condicionalmente convergente se  $-1 < c - a - b \leq 0$ ,  $x = -1$ ,
3. divergente se  $c - a - b \leq -1$ .

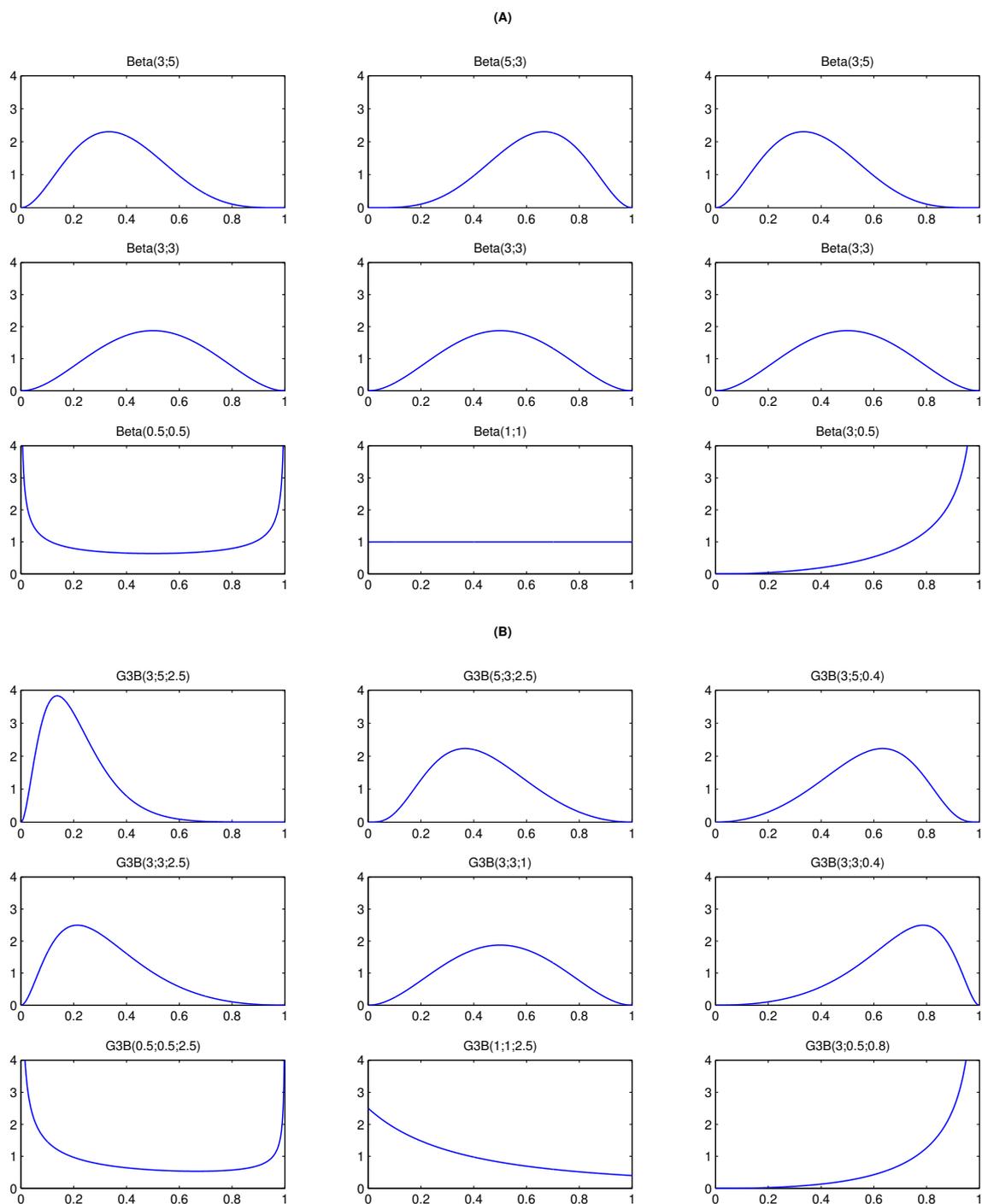


Figura 2.2.1: Fdp beta padrão (2.0.1) (A) e G3B (2.2.1) (B).

Quando  $a = 1$  e  $b = c$  (ou  $b = 1$  e  $a = c$ ), a série se torna  $1 + x + x^2 + \dots$

Se  $x = 1$ , o *Teorema da Soma de Gauss* implica que

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} = \frac{B(c, c-a-b)}{B(c-a, c-b)},$$

em que  $c - a - b > 0$ ,  $c \neq 0, -1, -2, \dots$

A função hipergeométrica satisfaz a equação diferencial de segunda ordem (Pham-Gia & Duong, 1989)

$$x(1-x)\frac{d^2y}{dx^2} + [c - (a+b+1)x]\frac{dy}{dx} - aby = 0.$$

As derivadas são

$$\frac{d}{dx} {}_2F_1(a, b; c; x) = \frac{ab}{c} {}_2F_1(a+1, b+1; c+1; x) \text{ e}$$

$$\frac{d^n}{dx^n} {}_2F_1(a, b; c; x) = \frac{(a)_n(b)_n}{(c)_n} {}_2F_1(a+n, b+n; c+n; x).$$

A *Integral de Euler* para a função hipergeométrica (Exton, 1976; Chen et al., 1984; Johnson et al., 1995b) é dada por

$${}_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{a-1}(1-u)^{c-a-1}(1-xu)^{-b} du,$$

quando  $c > a > 0$ .

Um grande número de funções especiais podem ser representadas pela função hipergeométrica, sendo a função beta incompleta uma delas (Johnson et al., 1995a):

$$B_p(a, b) = \int_0^p x^{a-1}(1-x)^{b-1} dx = a^{-1}p^a {}_2F_1(a, 1-b; a+1; p).$$

### 2.3.2 Momentos de $G3B$

Suponha que  $Y \sim G3B(\alpha, \beta, \lambda)$ . O  $m$ -ésimo momento de  $Y$  é dado por (Chen et al., 1984)

$$\begin{aligned} \mathbb{E}[Y^m] &= \int_0^1 \frac{y^m}{B(\alpha, \beta)} \frac{\lambda^\alpha y^{\alpha-1} (1-y)^{\beta-1}}{(1-(1-\lambda)y)^{\alpha+\beta}} dy \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \frac{\lambda^\alpha y^{\alpha+m-1} (1-y)^{\beta-1}}{(1-(1-\lambda)y)^{\alpha+\beta}} dy. \end{aligned}$$

A integral pode ser representada pela série hipergeométrica (2.3.1), com  $a = \alpha + m$ ,  $b = \alpha + \beta$ ,  $c = \alpha + \beta + m$  e  $x = 1 - \lambda$ . Dessa forma,

$$\begin{aligned}\mathbb{E}[Y^m] &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + m)}{\Gamma(\alpha)\Gamma(\alpha + \beta + m)} \lambda^\alpha \frac{\Gamma(\alpha + \beta + m)}{\Gamma(\alpha + m)\Gamma(\beta)} \int_0^1 \frac{y^{\alpha+m-1}(1-y)^{\beta-1}}{(1-(1-\lambda)y)^{\alpha+\beta}} dy \\ &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + m)}{\Gamma(\alpha)\Gamma(\alpha + \beta + m)} \lambda^\alpha {}_2F_1(\alpha + m, \alpha + \beta; \alpha + \beta + m; 1 - \lambda).\end{aligned}\tag{2.3.2}$$

Segundo Exton (1976), a Integral de Euler é invariável às transformações  $y = 1 - x$  e  $y = \frac{1-x}{1-zx}$ . Assim,  $\mathbb{E}[Y^m]$  também pode ser escrita como (Chen et al., 1984)

$$\mathbb{E}[Y^m] = \lambda^{-m} \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + m)}{\Gamma(\alpha)\Gamma(\alpha + \beta + m)} {}_2F_1\left(\alpha + m, m; \alpha + \beta + m; 1 - \frac{1}{\lambda}\right),$$

quando  $\lambda \geq 1$ , e

$$\mathbb{E}[Y^m] = \lambda^{-m} \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + m)}{\Gamma(\alpha)\Gamma(\alpha + \beta + m)} {}_2F_1(\beta, m; \alpha + \beta + m; 1 - 1/\lambda),$$

para  $\lambda < 1$ . Note que, para todo  $a$ ,  $b$  e  $c$ , a série hipergeométrica tem  ${}_2F_1(a, b; c; 0) = 1$ . Portanto, se  $\lambda = 1$ , ambas séries são reduzidas ao caso especial dos momentos de uma distribuição beta usual,

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

Para qualquer valor de  $\lambda$  com  $\alpha$ ,  $\beta$  grandes,  ${}_2F_1(\beta, m; \alpha + \beta + m; 1 - \lambda)$  aproxima-se de

$$\left(\frac{\alpha + \beta}{\alpha + \lambda\alpha_2}\right)^m,$$

fornecendo uma aproximação de  $\mathbb{E}[Y^m]$ ,

$$\mathbb{E}[Y^m] \doteq \lambda^{-m} \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + m)}{\Gamma(\alpha)\Gamma(\alpha + \beta + m)} \left(\frac{\alpha + \beta}{\alpha + \lambda\alpha_2}\right)^m.$$

### 2.3.3 Moda de $G3B$

Suponha que  $Y \sim G3B(\alpha, \beta, \lambda)$ . A moda de  $Y$  pode ser encontrada maximizando o logaritmo da função de densidade de probabilidade, ou seja, fazendo (Chen et al., 1984)

$$\frac{\alpha - 1}{y} - \frac{\beta - 1}{1 - y} + \frac{(\alpha + \beta)(1 - \lambda)}{1 - (1 - \lambda)y} = 0.$$

Para  $\alpha, \beta > 1$ , a moda é dada pela raiz em  $(0,1)$  da equação quadrática

$$2(\lambda - 1)y^2 + [(3 - \alpha) - \lambda(\beta - 1)]y + (\alpha - 1) = 0,$$

obtida fazendo  $\partial f/\partial y = 0$  (Pham-Gia & Duong, 1989).

Uma propriedade interessante dessa distribuição é que se  $\lambda = \alpha/\beta$ , então a moda =  $1/2$ . Portanto, diferentemente da distribuição beta usual, a distribuição beta generalizada pode ser não simétrica com moda  $0,5$ . Em geral, o fato de  $G3B$  apresentar três parâmetros possibilita especificar a “skewness” independentemente da média e variância.

### 2.3.4 Função de Distribuição Acumulada - $G3B$

Probabilidades de  $G3B$  podem ser calculadas a partir da distribuição beta padrão. Suponha que  $Y \sim G3B(\alpha, \beta, \lambda)$ , então

$$X = \frac{\lambda Y}{1 - (1 - \lambda)Y} \quad (2.3.3)$$

tem distribuição beta com parâmetros  $\alpha$  e  $\beta$  (Chen et al., 1984). Conseqüentemente,

$$\mathbb{P}[Y \leq y] = \mathbb{P}\left[X \leq \frac{\lambda y}{1 - (1 - \lambda)y}\right] = I_{\lambda Y/(1-(1-\lambda)Y)}(\alpha, \beta),$$

em que  $I_p(\alpha, \beta)$  é a função beta incompleta definida em (2.0.2).

### 2.3.5 Estimação dos Parâmetros

#### Método dos momentos

A distribuição beta padrão, como em (2.0.1), origina expressões fechadas em termos da esperança e da variância para os parâmetros, sendo assim, o método dos momentos fornece fórmulas exatas para a estimação dos mesmos. Entretanto, os momentos de uma variável aleatória  $X \sim G3B(\alpha, \beta, \lambda)$  dependem da função hipergeométrica gaussiana (2.3.2). Como essa função representa uma série infinita, os momentos de  $X$  não podem ser representados como expressões finitas. Para resolver o problema, Pham-Gia & Duong (1989)

mostram que a transformação  $Y = \frac{X}{1-X}$  dá origem à variável aleatória  $Y \sim G3F(\alpha, \beta, \lambda)$ , ou seja, com distribuição  $F$  generalizada, cuja densidade é dada por:

$$f_Y(y|\alpha, \beta, \lambda) = \frac{\lambda^\alpha y^{\alpha-1}}{B(\alpha, \beta)(1 + \lambda y)^{\alpha+\beta}}, \quad y > 1; \quad \alpha, \beta, \lambda > 0.$$

Os momentos da variável aleatória  $Y$  são obtidos por:

$$\mathbb{E}(Y^m) = \frac{\Gamma(\alpha + m)\Gamma(\beta - m)}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{\lambda^m} = \frac{(\alpha)_m}{(\beta - m)_m} \frac{1}{\lambda^m}, \quad \beta \geq m.$$

Seja  $X_1, \dots, X_n$  uma amostra aleatória de  $X \sim G3B(\alpha, \beta, \lambda)$ . Usando a transformação descrita acima, obtém-se  $Y_1, \dots, Y_n$ , sendo os três primeiros momentos amostrais dados por

$$m_1 = \frac{1}{n} \sum_{i=1}^n Y_i, \quad m_2 = \frac{1}{n} \sum_{i=1}^n Y_i^2 \quad e \quad m_3 = \frac{1}{n} \sum_{i=1}^n Y_i^3.$$

Agora, igualando os momentos amostrais aos teóricos, tem-se que

$$m_1 = \frac{1}{\lambda} \frac{\alpha}{\beta - 1}, \quad m_2 = \frac{1}{\lambda^2} \frac{\alpha(\alpha + 1)}{(\beta - 1)(\beta - 2)} \quad e \quad m_3 = \frac{1}{\lambda^3} \frac{\alpha(\alpha + 1)(\alpha + 2)}{(\beta - 1)(\beta - 2)(\beta - 3)}.$$

Resolvendo o sistema obtém-se que

$$\hat{\alpha} = \frac{2m_1(m_2^2 - m_1m_3)}{2m_1^2m_3 - m_1m_2^2 - m_2m_3},$$

$$\hat{\beta} = \frac{m_1^2m_2 + 3m_1m_3 - 4m_2^2}{m_1^2m_2 - 2m_2^2 + m_1m_3},$$

$$\hat{\lambda} = \frac{-m_1^2m_2 + 2m_2^2 - m_1m_3}{2m_1^2m_3 - m_1m_2^2 - m_2m_3}.$$

Pham-Gia & Duong (1989) afirma que  $\hat{\alpha}$ ,  $\hat{\beta}$  e  $\hat{\lambda}$  são estimadores consistentes mas viciados para os verdadeiros parâmetros. São assintoticamente normais com suas médias variando dos parâmetros por uma quantidade da ordem de  $1/n$  e com variância  $c/n$ .

Os autores ainda afirmam que resultados de simulação mostraram que, para valores pequenos de  $\alpha$ ,  $\beta$ ,  $\lambda$  e  $n$ , os parâmetros  $\alpha$  e  $\lambda$  são superestimados enquanto  $\beta$ , em geral, é subestimado. Entretanto, para  $\alpha = \beta \geq 10$ , com  $\lambda$  próximo de 1 e uma amostra grande ( $n \geq 200$ ), as estimativas são próximas dos valores reais.

### Máxima verossimilhança

Seja  $X_1, \dots, X_n$  uma amostra aleatória de  $X \sim G3B(\alpha, \beta, \lambda)$ , a log-verossimilhança é dada por

$$l(\alpha, \beta, \lambda | x_1, \dots, x_n) = \sum_{i=1}^n \log \left\{ \frac{\lambda^\alpha x_i^{\alpha-1} (1-x_i)^{\beta-1}}{B(\alpha, \beta) [1 - (1-\lambda)x_i]^{\alpha+\beta}} \right\}$$

e o método de máxima verossimilhança resulta no sistema

$$\Psi(\alpha + \beta) - \Psi(\alpha) = -\frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{x_i \lambda}{1 - (1-\lambda)x_i} \right\},$$

$$\Psi(\alpha + \beta) - \Psi(\beta) = -\frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{1 - x_i}{1 - (1-\lambda)x_i} \right\}$$

e

$$\frac{n\alpha}{\lambda} = (\alpha + \beta) \sum_{i=1}^n \left\{ \frac{x_i}{1 - (1-\lambda)x_i} \right\},$$

em que  $\Psi(w) = d/dw[\ln \Gamma(w)]$  representa a função digamma.

Observa-se, então, que não existem estatísticas suficientes para os parâmetros, e o sistema deve ser resolvido através de um método numérico iterativo.

A matriz de informação de Fisher é dada por

$$I = n \begin{bmatrix} \Psi'(\alpha + \beta) - \Psi'(\alpha) & \Psi'(\alpha + \beta) & \frac{\beta}{\lambda(\alpha + \beta)} \\ & \Psi'(\alpha + \beta) - \Psi'(\beta) & \frac{-\alpha}{\lambda(\alpha + \beta)} \\ & & \frac{-\alpha\beta}{\lambda^2(\alpha + \beta + 1)} \end{bmatrix}, \text{ e}$$

a matriz assintótica de variâncias e covariâncias dos estimadores de  $(\alpha, \beta, \lambda)$  é dada por  $-I^{-1}/n$ .

Para ilustrar o comportamento de  $-I^{-1}$ , considere os seguintes valores para os parâmetros:

$$\begin{cases} \alpha \in [0, 3; 2, 0], \\ \beta \in [0, 7; 1, 4] \text{ e} \\ \lambda \in [0, 1; 0, 9; 1, 1; 2, 0; 5, 0]. \end{cases}$$

Tabela 2.3.1: Exemplos da inversa da Informação de Fisher ( $-I^{-1}$ ).

$\lambda$	$(\alpha; \beta)$											
	(0,3;0,7)			(0,3;1,4)			(2,0;0,7)			(2,0;1,4)		
0,1	0,09	0,13	0,03	0,09	0,30	0,01	5,06	0,95	0,16	3,30	1,65	0,06
		0,84	-0,24		4,31	-0,49		0,42	-0,08		1,46	-0,14
			0,10			0,06			0,03			0,02
0,9	0,09	0,13	0,23	0,09	0,30	0,11	5,06	0,95	1,45	3,30	1,65	0,55
		0,84	-2,16		4,31	-4,40		0,42	-0,74		1,46	-1,22
			7,71			5,21			2,14			1,27
1,1	0,09	0,13	0,28	0,09	0,30	0,14	5,06	0,95	1,77	3,30	1,65	0,67
		0,84	-2,64		4,31	-5,38		0,42	-0,90		1,46	-1,49
			11,52			7,78			3,20			1,90
2,0	0,09	0,13	0,51	0,09	0,30	0,25	5,06	0,95	3,22	3,30	1,65	1,22
		0,84	-4,81		4,31	-9,77		0,42	-1,64		1,46	-2,70
			38,10			25,71			10,57			6,29
5,0	0,09	0,13	1,28	0,09	0,30	0,62	5,06	0,95	8,05	3,30	1,65	3,05
		0,84	-12,01		4,31	-24,44		0,42	-4,10		1,46	-6,76
			238,10			160,71			66,07			39,29

Os valores de  $-I^{-1}$  são mostrados na Tabela 2.3.1. Observa-se uma correlação negativa entre os estimadores de máxima verossimilhança de  $\beta$  e  $\lambda$ .

# Capítulo 3

## Arranjos de DNA

### 3.1 Estrutura do DNA

O DNA (ácido desoxirribonucleico) é uma molécula polimérica, em que cada unidade, o nucleotídeo, é composto por três sub-unidades: uma ose de cinco carbonos, a desoxirribose; uma base contendo nitrogênio (adenina (A), guanina (G), citosina (C) ou timina (T)); e um grupo fosfato. Os nucleotídeos se polimerizam em longas cadeias de polinucleotídeos, em forma de dupla hélice (em que A de uma das hélices se combina com T da outra, e G com C). Essa dupla hélice, contínua e única, representa um cromossomo. A estrutura do DNA especifica a seqüência de aminoácidos das cadeias polipeptídicas e a resultante síntese das proteínas (Thompson et al., 1993; Griffiths et al., 1999).

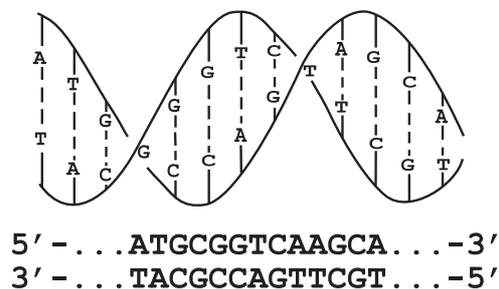


Figura 3.1.1: Parte de uma dupla hélice de DNA e a respectiva seqüência linear codificada (Thompson et al., 1993).

### 3.1.1 Fundamentos de Expressão Gênica

Thompson et al. (1993) define gene como sendo uma “seqüência de DNA cromossômico necessária para a elaboração de um produto funcional”.

As várias células especializadas em um organismo multicelular se apresentam tão diferentes umas das outras quanto possível, o que parece paradoxal, sendo que todas são derivadas de uma mesma célula. Em alguns poucos casos, a especialização celular envolve a perda de material genético, mas a maioria das células preserva todo o material original. Assim, a diferenciação depende da *expressão gênica*, e não da perda ou ganho de genes, sendo possível dizer que os genes podem ser *ligados* ou *desligados* (Alberts et al., 2002).

Técnicas de análise de DNA mostram que, em regra geral, as mudanças na expressão gênica em organismos multicelulares não são acompanhadas por alterações da seqüência de DNA dos genes correspondentes (Alberts et al., 2002; Brown, 2002).

As informações genéticas contidas no DNA dos cromossomos se localizam no núcleo celular (em células eucariontes, como as humanas), mas a síntese de proteínas, resultado da expressão gênica, ocorre no citoplasma celular. O elo molecular existente entre o código de DNA dos genes e o código de aminoácidos das proteínas é o **RNA** (ácido ribonucléico). As estruturas químicas destes dois ácidos são semelhantes, com exceção da substituição da desoxirribose pela ribose, e da timina pela uracila (U). Outra diferença, é que o RNA normalmente existe como fita única, enquanto o DNA é uma fita dupla (Thompson et al., 1993).

As relações de informação entre DNA, RNA e proteína são circulares: o DNA dirige a síntese e seqüência do RNA; este dirige a síntese e seqüência dos polipeptídeos; e proteínas específicas estão envolvidas na síntese e metabolismo de DNA e RNA (Thompson et al., 1993).

O código genético armazenado no DNA determina, através da seqüência de bases, a seqüência de aminoácidos na proteína codificada (Thompson et al., 1993; Griffiths et al., 1999; Alberts et al., 2002).

O processo pode ser descrito de maneira simplificada como: o RNA é sintetizado

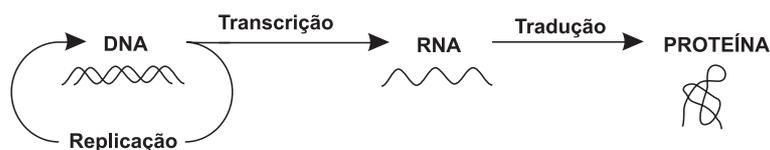


Figura 3.1.2: Esquema simplificado da codificação proteica (Thompson et al., 1993).

a partir do modelo de DNA pelo processo de transcrição; o RNA mensageiro (**mRNA**) gerado é transportado para o citoplasma (células eucariontes), onde sua seqüência é lida nos ribossomos (o qual possui, como uma de suas sub-estruturas, o RNA ribossômico - **rRNA**), pelo processo de tradução, determinando a seqüência de aminoácidos na proteína que está sendo sintetizada; o elo molecular entre a seqüência de bases codificadas do mRNA e a seqüência de aminoácidos da proteína é dado por um terceiro tipo de RNA, o transportador - **tRNA** (Thompson et al., 1993).

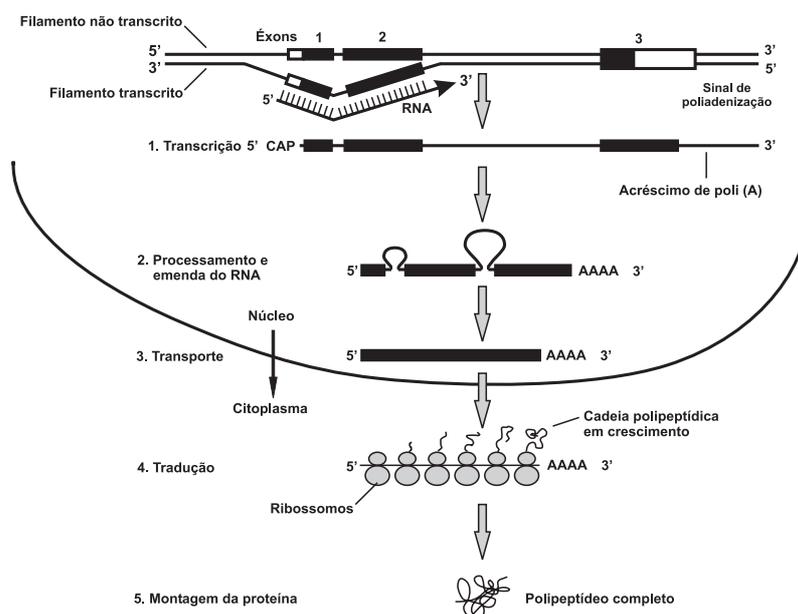


Figura 3.1.3: Fluxo de informações DNA  $\rightarrow$  RNA para a proteína de um gene hipotético. As etapas incluem a transcrição, processamento e emenda do RNA, transporte do RNA do núcleo ao citoplasma e tradução (Thompson et al., 1993).

### 3.1.2 Estrutura e Organização dos Genes

Os genes, em geral, não se apresentam como uma sequência codificadora contínua, mas sim interrompida por uma ou mais regiões não codificadoras. Essas regiões, chamadas de *íntrons*, não estão presentes no mRNA que será traduzido, não sendo, então, representadas no produto protéico final. As regiões do DNA que são codificadas são chamadas de *éxons* (Thompson et al., 1993).

O gene possui, além de éxons e íntrons, sequências adjacentes de nucleotídeos essenciais a sua expressão correta, como a região **promotora** (na extremidade 5') e a **finalizadora** (na extremidade 3'), indicando o início e o término da síntese, respectivamente. Assim, torna-se possível a produção de uma molécula de mRNA funcional (Thompson et al., 1993; Alberts et al., 2002).

O fluxo de informações do gene para o polipeptídeo envolve várias etapas. A transcri-

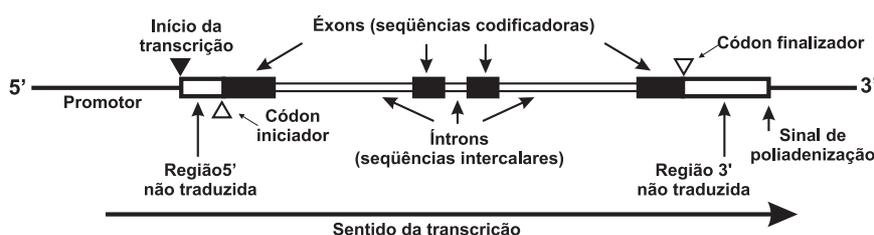


Figura 3.1.4: Estrutura de um gene humano típico (Thompson et al., 1993).

ção do gene é iniciada num sítio do DNA cromossômico (que corresponde à extremidade 5' do RNA final), estendendo-se por centenas (ou até milhares) de pares de bases, incluindo íntrons e éxons. A síntese do RNA primário transcrito prossegue na direção 5'-para-3', enquanto o filamento do gene que está sendo transcrito é lido na direção 3'-para-5' (Thompson et al., 1993; Brown, 2002).

O transcrito primário é processado acrescentando-se à extremidade 5' uma cobertura (*CAP*) e clivando-se a extremidade 3' onde é acrescentada uma cauda poli-A (sequência de adenosinas). Tal cauda tem a função de assinalar a extremidade 3' e, aparentemente, aumentar a estabilidade do RNA. Após isso, são eliminadas as sequências correspondentes aos íntrons (Thompson et al., 1993).

Tabela 3.1.1: Codificação dos códons

1ª letra do códon	2ª letra do códon				3ª letra do códon
	U	C	A	G	
U	fenilalanina	serina	tirosina	cisteína	U
	fenilalanina	serina	tirosina	cisteína	C
	leucina	serina	STOP	STOP	A
	leucina	serina	STOP	triptofano	G
C	leucina	prolina	histidina	arginina	U
	leucina	prolina	histidina	arginina	C
	leucina	prolina	glutamina	arginina	A
	leucina	prolina	glutamina	arginina	G
A	isoleucina	treonina	asparagina	serina	U
	isoleucina	treonina	asparagina	serina	C
	isoleucina	treonina	lisina	arginina	A
	metionina	treonina	lisina	arginina	G
G	valina	alanina	ác. aspártico	glicina	U
	valina	alanina	ác. aspártico	glicina	C
	valina	alanina	ác. glutâmico	glicina	A
	valina	alanina	ác. glutâmico	glicina	G

No citoplasma, o mRNA é traduzido em proteína pela ação de várias moléculas de tRNA, cada uma específica para um aminoácido. Essas moléculas têm a tarefa de transferir os aminoácidos corretos do citoplasma para suas posições ao longo do modelo de mRNA, a fim de serem acrescentados à cadeia polipeptídica em crescimento (Thompson et al., 1993; Brown, 2002).

Cada conjunto de três bases constitui um **códon**, específico para determinado aminoácido. Variações enormes são teoricamente possíveis no arranjo das bases ao longo de uma cadeia de polinucleotídeos. Em qualquer posição, há quatro possibilidades (A, T, C ou G); logo, existem  $4^n$  combinações possíveis numa seqüência de  $n$  bases. Para cada códon, há  $4^3$  combinações possíveis. Como existem apenas 20 aminoácidos mais o sinal de parada (STOP) e 64 códons possíveis, a maioria dos aminoácidos é especificada por mais de um códon, podendo-se dizer que o código é **degenerado** (Thompson et al., 1993), como pode ser observado na Tabela 3.1.1 (Alberts et al., 2002).

Após a tradução, a proteína ainda pode sofrer modificações, como o acréscimo de moléculas em sítios específicos ou clivagens de determinadas regiões (Thompson et al.,

1993).

Assim, os tipos celulares em um organismo multicelular se tornam diferentes uns dos outros devido à síntese e ao acúmulo de diferentes tipos de RNA e proteínas (Alberts et al., 2002; Brown, 2002).

Existem vários mecanismos através dos quais as células podem controlar a expressão gênica, destacando-se (Alberts et al., 2002):

1. quando e com qual frequência um gene é transcrito (controle de transcrição);
2. como o RNA transcrito é processado (controle no processamento do RNA);
3. seleção de quais mRNAs são transportados do núcleo para o citoplasma (controle no transporte de RNA);
4. seleção de quais mRNAs presentes no citoplasma são traduzidos pelos ribossomos (controle de tradução);
5. desestabilização seletiva de algumas moléculas de mRNA presentes no citoplasma (controle de degradação do mRNA); e
6. ativação, desativação ou compartimentalização seletiva de proteínas específicas após sua produção (controle da atividade protéica).

## 3.2 Genoma Funcional

As diferentes respostas de um organismo aos estímulos externos são devidas à alterações no conjunto de proteínas das células. Esse processo permite a adaptação dos organismos às situações mais adversas possíveis. Assim, a quantidade de RNA transcrito em uma célula indica o quanto o gene correspondente é expresso (Felix et al., 2001).

Com base nessas informações, foram desenvolvidas tecnologias (arranjos de DNA) para a quantificação do material transcrito. Dessa forma, a grande quantidade de informações

geradas pelo projeto *Genoma* pode ser analisada e cada gene pode ter sua função identificada, de acordo com as situações sob as quais são expressos (Felix et al., 2001).

Um fator importante na análise desse tipo de dados é o custo. Tanto o experimento de micro quanto de macroarranjo são caros, envolvendo equipamentos e materiais de valores elevados. Quando deseja-se selecionar aqueles genes, entre os milhares estudados, que possuam um comportamento de expressão específico para posterior estudo mais detalhado dos mesmos, o custo se eleva ainda mais. Assim sendo, o pesquisador espera selecionar poucos genes. Dessa maneira, a técnica de análise deve ser capaz de determinar apenas os genes que apresentem comportamento muito bem determinado. Podemos dizer então que, nesse caso, a subseleção é mais interessante do que a superseleção dos genes.

### 3.2.1 Microarranjos

No passado, a análise genética se restringia ao estudo de poucos genes em cada experimento. Com os projetos genomas identificando grandes quantidades de genes, surgiu a necessidade de novas técnicas capazes de analisar estas centenas ou milhares ao mesmo tempo. Hoje, é possível analisar mais de 300.000 genes em um arranjo com aproximadamente  $1,64 \text{ cm}^2$ . Existem inúmeras aplicações para os microarranjos, no entanto, as duas mais usadas são o estudo de expressão gênica (RNA) e o estudo de variação do DNA (Strachan & Read, 1999; Freeman et al., 2000).

Em experimentos de microarranjos, o mRNA é extraído de tecidos específicos sob determinadas condições biológicas, que são funcionalmente associadas ao mecanismo estudado (Reiner et al., 2003).

A análise da expressão gênica sob diversas condições biológicas, estágios do ciclo celular, tecidos e indivíduos pode ajudar a identificar genes diferentemente expressos. Os microarranjos de DNA têm sido usados para monitorar os níveis de expressão de milhares de genes simultaneamente, identificando aqueles diferentemente expressos (Reiner et al., 2003).

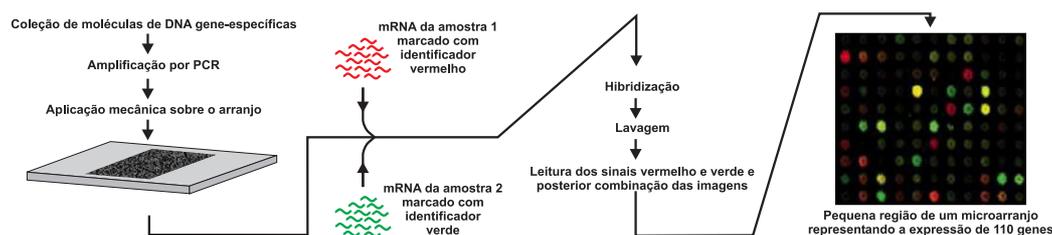


Figura 3.2.1: Para preparar o arranjo, fragmentos de DNA, cada um correspondendo a um gene, são fixados sobre a lâmina por um robô. Neste exemplo, o mRNA é coletado de dois tipos diferentes de células para uma comparação direta de seus níveis de expressão. As amostras são convertidas em cDNA e marcadas, uma com fluorocromo vermelho e outra com fluorocromo verde. As amostras marcadas são misturadas e então aplicadas sobre a lâmina. Após incubação, o arranjo é lavado e a fluorescência registrada. Na porção mostrada do microarranjo, que representa a expressão de 110 genes de levedura, “spots” vermelhos indicam que o gene é mais expresso na amostra 1 que o correspondente gene na amostra 2; “spots” verdes indicam que ocorre o inverso. “Spots” escuros indicam que ocorre pouca ou nenhuma expressão em ambas amostras. As diferentes intensidades de cores representam, então, o nível de expressão em cada amostra (Alberts et al., 2002).

A arquitetura de um microarranjo é relativamente simples: consiste em uma grande quantidade de porções parciais ou completas de genes (*sondas*), com seqüências diferentes, sendo imobilizadas ordenadamente e em posições específicas (*spots*) sobre uma superfície sólida (comumente vidro ou náilon). No caso de estudos de expressão gênica, tais sondas consistem em **cDNA**, que é o produto da transcrição reversa do mRNA. O arranjo é então incubado com o DNA alvo marcado com elementos fluorescentes, para que ocorra a hibridização. Após isso, a superfície do arranjo é analisada com uso de laser ou microscópio, identificando as posições onde o marcador foi detectado (Brown, 2002; Strachan & Read, 1999).

Um experimento sobre expressão gênica pode ser assim descrito:

- Suponha um experimento para determinar a diferença da expressão gênica de determinada célula sob diversas situações, por exemplo, a comparação de células fúngicas crescendo em ambientes com e sem oxigênio;

1. o mRNA é extraído das células separadamente;

2. sintetiza-se o cDNA, usando-se nucleotídeos marcados com elementos fluorescentes, uma cor para células que cresceram em ambiente com oxigênio (por exemplo, verde) e outra para as que cresceram em ambiente sem oxigênio (por exemplo, vermelho);
3. os cDNAs marcados são misturados;
4. a mistura é aplicada sobre o microarranjo, em que cada *spot* possui uma seqüência de DNA diferente. Após a incubação, moléculas do cDNA se encontram ligadas (hibridizadas) a moléculas de DNA presentes nos *spots*;
5. o microarranjo deve ser lavado para eliminar o cDNA não hibridizado;
6. o microarranjo é posicionado em uma câmara escura, onde a aplicação de lasers verde e vermelho detecta os *spots* em que ocorreu a hibridização. Cada laser é aplicado separadamente e as imagens armazenadas; e
7. ambas as imagens são sobrepostas, gerando níveis diferentes de verde, vermelho e amarelo, representando respectivamente a expressão de genes em ambiente com oxigênio, sem oxigênio e em ambos.

### 3.2.2 Macroarranjos

Os macroarranjos são arranjos de náilon com densidade aproximada de 20 genes/mm<sup>2</sup>. Tais arranjos são uma alternativa aos arranjos de vidro (microarranjos), tanto pelo custo quanto pela facilidade de implementação (Felix et al., 2001).

No caso de macroarranjos, não se utilizam marcadores fluorescentes, mas sim radioativos. O processo é muito semelhante, com algumas diferenças (Felix et al., 2001):

- normalmente, os genes aparecem em duplicata no arranjo; e
- a leitura do resultado final é obtida através da exposição de uma placa composta por material sensível à radiação. A imagem resultante é digitalizada e armazenada.

# Capítulo 4

## Dados de DNA e Misturas de Densidades

### 4.1 Considerações estatísticas sobre o estudo de expressão gênica em arranjos de DNA

A análise de expressão gênica através de arranjos de DNA representa um grande avanço na tecnologia genética, pela qual é possível analisar milhares de genes em um único conjunto de experimentos a um custo razoável. Entretanto, devido a enorme quantidade de dados e sutil natureza da expressão gênica, a análise e incerteza das decisões recaem sobre a teoria e modelos estatísticos. O problema fica cada vez mais complexo com o aumento no número de genes estudados. Nesse caso, milhares de genes são testados contra um conjunto de hipóteses nulas, esperando-se que um subconjunto dos mesmos seja significativo. Isso resulta em problemas de comparações múltiplas. Soluções sugeridas englobam o controle da taxas de erro, como a FWER (*Familywise error rate*) e a FDR (*False discovery rate*).

Sendo de interesse, no caso de estudos de genoma funcional por macro e microarranjos, selecionar aqueles genes com características mais interessantes para análises posteriores,

o custo causado pela superseleção pode tornar-se proibitivo. Assim, a literatura recente mostra uma tendência na análise da FDR para o controle do número de falsos positivos. O objetivo é identificar o máximo de genes significantes, implicando em uma baixa proporção relativa de falsos positivos. Como os testes são realizados em cada gene, os valores de corte tradicionais para os p-valores (0,01 e 0,05) devem ser corrigidos para evitar uma abundância de falsos positivos. Dessa forma, se nenhum gene é diferentemente expresso e a FDR é controlada em algum nível  $q$ , então a probabilidade de detecção errônea de algum gene diferentemente expresso é menor ou igual à  $q$  (Norris & Kahn, 2006; Unger et al., 2005).

Quando muitos testes estatísticos são realizados simultaneamente, a probabilidade total de erro tipo I (rejeitar a hipótese nula quando a mesma é verdadeira) pode exceder substancialmente o valor nominal usado em cada teste. Por exemplo, se  $\alpha = 0,05$ , a probabilidade de ao menos um erro do tipo I entre 10 testes independentes é de aproximadamente 0,40, enquanto a mesma probabilidade entre 100 testes independentes é maior que 0,99 (Parker & Rothenberg, 1998).

A necessidade de solucionar tal problema gerou uma série de procedimentos de comparações múltiplas. Assim, inúmeras técnicas visam o ajuste dos resultados individuais para controlar o erro total. Em geral, tais técnicas aplicam a cada teste um critério mais rigoroso (um p-valor menor) do que o necessário individualmente. No entanto, em análises que buscam resultados para experimentos futuros, tal método conservativo pode não ser apropriado. Alguns resultados podem ser interessantes, mesmo que não estatisticamente significantes ao nível nominal. Portanto, são necessárias técnicas para identificar tais resultados interessantes, estatisticamente significantes ou não (Parker & Rothenberg, 1998).

## 4.2 Taxa de Falsos Positivos

Segundo Benjamini & Hochberg (1995), a abordagem comum para o problema de multiplicidade tem por objetivo controlar a “familywise error rate” (FWER), ou seja, controlar a probabilidade de cometer qualquer erro em múltiplos testes. No entanto, o autor aponta falhas nesse tipo de procedimento, apresentando uma nova abordagem que envolve o controle da proporção esperada de hipóteses erroneamente rejeitadas, a “false discovery rate” (FDR). Essa medida é equivalente à FWER quando todas as hipóteses são verdadeiras, mas menor caso contrário.

O controle a partir da FDR é baseado na idéia de que se pode tolerar um maior número de falsos positivos se o número de testes é grande (Yang & Yang, 2006).

Considere o problema que testa simultaneamente  $m$  hipóteses, das quais  $m_0$  são verdadeiras. Seja  $\mathbf{R}$  o número de hipóteses rejeitadas. A Tabela 4.2.1 resume a situação na forma tradicional. A variável aleatória  $\mathbf{R}$  é observável, enquanto  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{S}$  e  $\mathbf{T}$  não são observáveis. Se cada hipótese nula é testada separadamente ao nível  $\alpha$ , então  $\mathbf{R}=\mathbf{R}(\alpha)$  é crescente em  $\alpha$  (Benjamini & Hochberg, 1995).

Em termos dessas variáveis aleatórias, a taxa de erro por comparação (PCER) é dada

Tabela 4.2.1: Número de erros cometidos quando testadas  $m$  hipóteses nulas.

	Não Rejeitadas	Rejeitadas	Total
Hipótese nula verdadeira	$\mathbf{U}$	$\mathbf{V}$	$m_0$
Hipótese nula falsa	$\mathbf{T}$	$\mathbf{S}$	$m - m_0$
	$m - \mathbf{R}$	$\mathbf{R}$	$m$

por  $\mathbb{E}(\mathbf{V}/m)$  e a FWER é dada por  $\mathbb{P}(\mathbf{V} \geq 1)$ . O teste individual de cada hipótese ao nível  $\alpha$  garante que  $\mathbb{E}(\mathbf{V}/m) \leq \alpha$ . Testando individualmente cada hipótese ao nível  $\alpha/m$  garante que  $\mathbb{P}(\mathbf{V} \geq 1) \leq \alpha$  (Benjamini & Hochberg, 1995).

A proporção de erros cometidos por rejeitar erroneamente hipóteses nulas pode ser representada pela variável aleatória  $\mathbf{Q}=\mathbf{V}/(\mathbf{V}+\mathbf{S})$ , ou seja, a proporção das hipóteses rejeitadas que são erroneamente rejeitadas. Naturalmente,  $\mathbf{Q}=0$  quando  $\mathbf{V}+\mathbf{S}=0$ . A vari-

ável aleatória  $\mathbf{Q}$  também é não observável. Assim, o FRD  $Q_e$  é definido como a esperança de  $\mathbf{Q}$ ,

$$Q_e = \mathbb{E}(\mathbf{Q}) = \mathbb{E}[\mathbf{V}/(\mathbf{V} + \mathbf{S})] = \mathbb{E}(\mathbf{V}/\mathbf{R}).$$

A FDR tem duas propriedades facilmente demonstráveis mas ainda assim muito importantes:

1. quando todas as hipóteses nulas são verdadeiras, a FDR é equivalente à FWER. Assim, controlar a FDR implica em controlar a FWER.
2. quando  $m_0 < m$ , a FDR é menor ou igual à FWER. Assim, qualquer procedimento que controla a FWER também controla a FDR. O potencial de aumento no poder é maior quanto mais das hipóteses não são verdadeiras.

O autor ainda apresenta um procedimento para controlar a FDR. Considere o teste de  $H_1, H_2, \dots, H_m$  baseado nos p-valores correspondentes  $P_1, P_2, \dots, P_m$ . Sejam  $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$  os p-valores ordenados, denotando por  $H_{(i)}$  a hipótese nula correspondente à  $P_{(i)}$ . Defina o seguinte procedimento de testes múltiplos tipo Bonferroni:

$$\begin{aligned} &\text{seja } k \text{ o maior } i \text{ tal que } P_{(i)} \leq \frac{i}{m} \alpha; \\ &\text{então rejeite } H_{(i)}, i = 1, 2, \dots, k. \end{aligned}$$

O autor mostra que para estatísticas de teste independentes e para qualquer configuração das hipóteses nulas, o procedimento controla a FDR ao nível  $\alpha$ .

Segundo Reiner et al. (2003), considerando a análise de dados de microarranjos através do teste de cada gene, o problema de testes múltiplos torna-se uma preocupação. A probabilidade de um falso positivo (erro tipo I) pode crescer drasticamente quando o número de genes testados aumenta (Reiner et al., 2003; Xiang et al., in press).

A FWER representa a probabilidade de cometer ao menos um erro na família de hipóteses. Quando algumas das hipóteses testadas são de fato falsas, a FDR é menos estrita que a FWER, sendo potencialmente mais poderosa. Quando a rejeição de uma hipótese resulta em uma ação, como aprovar uma nova droga, o controle da FWER é mais

adequado. Caso contrário, o controle da FDR é suficiente, como é o caso da análise de dados de microarranjo, que está interessando em extrair genes que serão possíveis candidatos à análises posteriores. Nesse caso, um número de rejeições errôneas, desde que pequeno, não distorce as conclusões nesse estágio da análise. Portanto, controlar a probabilidade de ao menos um falso positivo (FWER) parece ser super-conservativo, resultando em uma redução da eficiência do experimento devido a uma perda desnecessária do poder. O controle da FDR permite controlar a proporção de esforços investidos em vão, em média, na próxima etapa da investigação (Reiner et al., 2003).

Em termos de microarranjos, se nenhum gene é diferentemente expresso e a FDR é controlada em algum nível  $q$ , então a probabilidade de detecção errônea de algum gene diferentemente expresso é menor ou igual à  $q$  (Reiner et al., 2003).

Pounds & Morrins (2003) mostra que a ocorrência de falsos positivos e falsos negativos em análises de microarranjos pode ser facilmente estimada se a distribuição dos p-valores é aproximada e então expressa como uma mistura de densidades nula e alternativa. Essencialmente, qualquer distribuição de p-valores pode ser expressa como tal mistura pela remoção da densidade de uma uniforme.

Duas propriedades formam a base para análise de grande conjunto de p-valores: (1) p-valores provenientes das hipóteses nulas são distribuídos uniformemente entre 0 e 1; (2) a distribuição de um conjunto de p-valores pode ser expressa como uma mistura de uma uniforme em  $(0,1)$  com outra componente.

O autor introduz um modelo que descreve a distribuição dos p-valores originários de uma análise de microarranjo. O modelo é usado para obter uma distribuição estimada que é facilmente expressa como uma mistura de densidades nula e alternativa. Fornecendo um limite de significância, a distribuição estimada é particionada em regiões correspondendo a ocorrência de falsos positivos, falsos negativos, positivos verdadeiros e negativos verdadeiros, como pode ser visto na Figura 4.2.1. Na figura,  $\tau$  representa o ponto de corte (*threshold*) para a significância. O teste é dito significativo quando o p-valor é menor que  $\tau$ . Tendo selecionado o valor de  $\tau$ , a densidade estimada  $\hat{f}(x)$  pode ser particionada

como mostrado. A região acima da linha  $y = \pi$  representa a componente alternativa da distribuição, a região abaixo, representa a porção proveniente da componente nula.

O modelo apresentado pelo autor é de uma mistura beta-uniforme (BUM) da forma

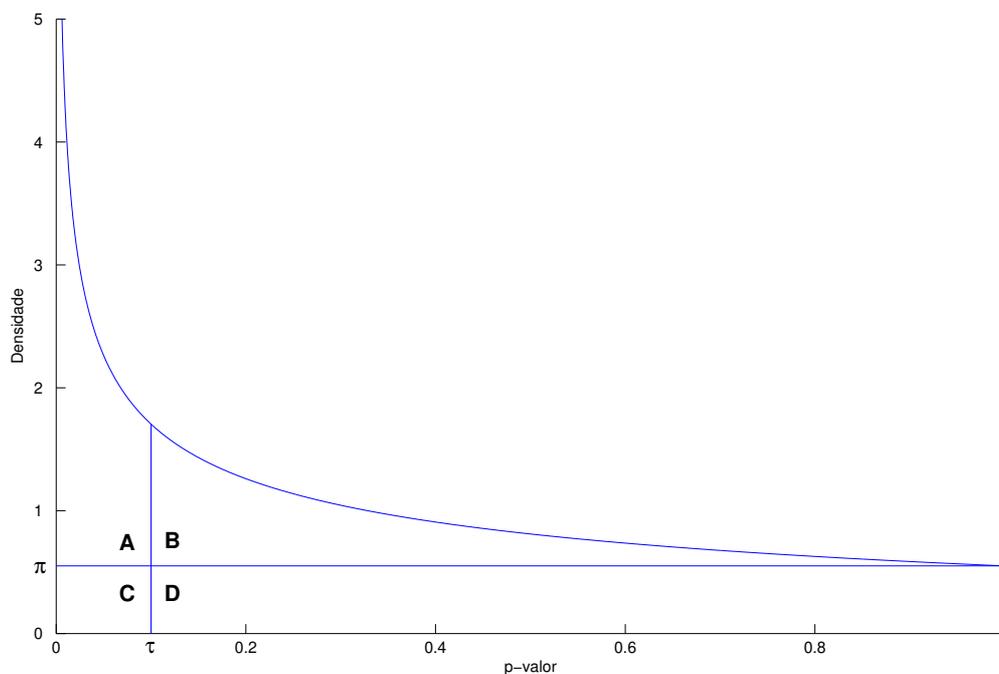


Figura 4.2.1: Região A corresponde à ocorrência de positivos verdadeiros; região B corresponde à ocorrência de falsos negativos; região C corresponde à ocorrência de falsos positivos; e a região D corresponde à ocorrência de negativos verdadeiros (Pounds & Morrins, 2003).

$$f(x|p, \lambda) = \lambda + (1 - \lambda)px^{p-1},$$

em que  $0 < x \leq 1$ ,  $0 < \lambda < 1$  e  $0 < p < 1$ . A função densidade de probabilidade (fdp)  $f(x|p, \lambda)$  é uma curva com assíntota em  $x = 0$  e que decresce monotonamente ao mínimo de  $\lambda + (1 - \lambda)p$  em  $x = 1$ . Tal curva prevê a distribuição dos p-valores provenientes de uma análise de microarranjo. Sob a hipótese nula, os p-valores terão uma densidade uniforme correspondente à linha horizontal. Sob a hipótese alternativa, os p-valores terão uma distribuição com alta densidade para valores pequenos, decrescendo conforme os valores

crecem.

Os parâmetros  $p$  e  $\lambda$  podem ser estimados por máxima verossimilhança, possibilitando a estimação do *threshold*  $\tau$  por

$$\hat{\tau} = \hat{\lambda} + (1 - \hat{\lambda})\hat{p},$$

e que representa a máxima proporção de p-valores originados da hipótese nula.

Segundo Dalmasso et al. (2005), o critério mais usado atualmente na análise de dados de microarranjos é a FDR. O autor propõe uma nova família de estimadores para  $\tau$ , possibilitando o cálculo da FDR. É mostrado que o estimador proposto possui bom comportamento sob amostras finitas além de melhor erro quadrático médio na maior parte dos casos quando comparado a outros procedimentos, inclusive o BUM,

$$\hat{\pi}_{0(n)} = \frac{(1/m) \sum_{i=1}^m [-\log(1 - p_i)]^n}{n!},$$

em que  $n \in \mathbb{N}$  e  $m$  é o tamanho da amostra. O autor usa que

$$\hat{\pi}_{0(n)} \approx N \left[ \frac{E(\phi(P)^n)}{n!}, \frac{1}{(n!)^2} \frac{\sigma_{(n)}^2}{m} \right],$$

em que  $\sigma^2(n)$  é a variância da variável aleatória  $\phi(P)^n$ , sendo  $\phi(x) = -\ln(1 - x)$ . O autor define  $\pi_0$  como a probabilidade da hipótese nula ser verdadeira.

Xiang (in press) usa três estratégias computacionais para avaliar a performance da estimação de intervalos em modelos beta-uniformes. O número de p-valores e a proximidade de um parâmetro dos limites de seu espaço exercem papéis importantes na precisão das estimações, assim como a proximidade da componente beta com a componente uniforme.

Parker & Rothenberg (1988) afirma que qualquer distribuição no intervalo  $[0,1]$  pode ser aproximada por uma mistura finita de distribuições beta. Portanto, sempre é possível usar uma mistura de distribuições beta como modelo genérico para um conjunto de p-valores, independentemente do(s) teste(s) estatístico(s) usado(s) para obter tais p-valores. Por exemplo, essa abordagem genérica pode ser usada quando a significância estatística é determinada usando uma variedade de diferentes testes no mesmo experimento, como

testes qui-quadrados em tabelas de contingência ou testes t para diferenças entre grupos de variáveis contínuas. Dessa maneira, o modelo poderia ser assim representado

$$f(x_i) = p_0\beta(1, 1) + \sum_{j=1}^k p_j\beta(r_j, s_j)(x_i),$$

em que  $p_j$  representa a proporção relativa do conjunto  $\{x_i\}$  na  $j$ -ésima distribuição,  $j = 1, \dots, k$ , com  $0 < p_j \leq 1$ ,  $\sum_{j=1}^k p_j = 1$  e  $\beta(r, s)$  representando a densidade de uma variável aleatória beta.

Todo  $x_i$  que representa uma hipótese nula verdadeira, não importando o valor de  $x_i$ , vem do primeiro termo da densidade,  $p_0\beta(1, 1)(\bullet)$ . Os  $x_i$  restantes, que representam falsos positivos, vêm do segundo termo. Portanto, se todos os resultados vêm de hipóteses nulas verdadeiras, então  $k = 0$  e  $p_0 = 1$ .

### 4.3 Misturas de Densidades

Titterington et al. (1985) define uma *mistura finita* da seguinte forma: suponha uma variável ou vetor aleatório absolutamente contínuo,  $\mathbf{X}$ , que assume valores no espaço  $\mathbb{X}$ , e que sua distribuição pode ser representada pela função densidade de probabilidade da forma

$$f(\mathbf{x}) = \sum_{i=1}^k p_i f_i(\mathbf{x}),$$

em que

$$p_j > 0, j = 1, \dots, k; p_1 + \dots + p_k = 1$$

e

$$f_j(\cdot) \geq 0, \int_{\mathbb{X}} f_j(\mathbf{x}) d\mathbf{x} = 1, j = 1, \dots, k.$$

Dessa maneira,  $\mathbf{X}$  tem a distribuição de uma mistura finita e  $f(\cdot)$  é a função de densidade da mistura (Titterington et al., 1985).

Os parâmetros  $p_1, \dots, p_k$  são chamados de *pesos da mistura* e  $f_1(\cdot), \dots, f_k(\cdot)$  os *componentes* da mistura. Em muitas aplicações, as componentes  $f_i(\mathbf{x})$  são especificadas como pertencentes a alguma família paramétrica. Nesse caso, tais densidades têm uma representação paramétrica definida, e a densidade da mistura possui uma representação mais explícita (Titterington et al., 1985; McLachlan & Peel, 2000):

$$f(\mathbf{x}|\boldsymbol{\psi}) = p_1 f_1(\mathbf{x}|\boldsymbol{\theta}_1) + \dots + p_k f_k(\mathbf{x}|\boldsymbol{\theta}_k), \quad (4.3.1)$$

em que  $\boldsymbol{\theta}_j$  representa os parâmetros em  $f_j(\cdot)$ . Podemos representar todos os parâmetros que ocorrem nas componentes da mistura por  $\boldsymbol{\theta}$  e a coleção completa de parâmetros que ocorrem em  $f(\cdot)$  por  $\boldsymbol{\psi}$ .

Não existe a restrição de que todas as componentes de  $f(\cdot)$  devam pertencer à mesma família de densidades, mas em muitas aplicações este será o caso. Dessa forma, a função de densidade da mistura finita terá a forma:

$$f(\mathbf{x}|\boldsymbol{\psi}) = \sum_{j=1}^k p_j f(\mathbf{x}|\boldsymbol{\theta}_j),$$

em que  $f(\cdot|\theta)$  denota um membro genérico da família paramétrica.

No caso dessa mistura, cada  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k$  é um elemento do mesmo espaço paramétrico, digamos,  $\Theta$ . Assim, pode-se pensar em  $\boldsymbol{\pi} = (p_1, \dots, p_k)$  como definindo uma distribuição de probabilidade sobre  $\Theta$ , com  $p_j = P(\theta = \boldsymbol{\theta}_j), j = 1, \dots, k$ . Se  $G_{\boldsymbol{\pi}}(\cdot)$  denota a medida de probabilidade sobre  $\Theta$  definida por  $\boldsymbol{\pi}$ , então  $f(\mathbf{x}|\boldsymbol{\psi})$  pode ser formalmente definida por

$$p(\mathbf{x}|\boldsymbol{\psi}) = \int_{\Theta} f(\mathbf{x}|\boldsymbol{\theta}) dG_{\boldsymbol{\pi}}(\boldsymbol{\theta}).$$

Modelos de misturas finitas têm sido muito usados em biologia e medicina com objetivo de modelagem e agrupamento (Miloslavsky & van der Laan, 2003).

### 4.3.1 Número de Componentes

McLachlan & Basford (1988) afirma que o teste para o número de componentes  $k$  de uma mistura é um problema importante mas muito difícil, não tendo sido ainda completamente

resolvido. Uma maneira óbvia de abordar o problema é usar a estatística da razão de verossimilhança  $\lambda$ , testando o menor valor de  $k$  compatível com os dados. Infelizmente, com modelos de misturas, as condições de regularidade não estão presentes para que  $-2\log \lambda$  tenha sua distribuição nula assintótica qui-quadrada com graus de liberdade iguais à diferença do número de parâmetros entre as duas hipóteses. Uma maneira rápida de ilustrar isso é: considere uma mistura de duas fdp normais, ou seja,  $k = 2$ . A hipótese nula de que haveria apenas uma componente seria dada por

$$H_0 : k = 1,$$

podendo ser obtida testando se  $p_1 = 1$ , que está no limite do espaço paramétrico, violando à condição de regularidade. A abordagem poderia ser feita testando se  $\mu_1 = \mu_2$  e  $\sigma_1^2 = \sigma_2^2$ , tornando o valor de  $p_1$  irrelevante. Mas, mesmo nesse caso, as condições de regularidade não são obtidas (Quinn & McLachlan, 1987). Ghosh & Sen (1985) apresenta problemas com as condições de regularidade na teoria assintótica clássica para o teste da razão de verossimilhança.

Miloslavsky & van der Laan (2003) afirma que a estimação do número de componentes  $k$  de uma mistura é um problema não resolvido. Métodos disponíveis para tal estimação incluem o bootstrap da razão de verossimilhança e a otimização de uma variedade de funções de validação. O autor investiga a minimização da distância entre o modelo estimado para a mistura (de normais univariadas) e a verdadeira densidade, usando as distâncias de Kullback-Leibler e  $\mathbb{L}_2$ , calculadas através de validação cruzada. Utiliza AIC, BIC/MDL e ICOMP para comparação dos resultados.

McLachlan (1987) e Feng (1996) usam o bootstrap da razão de verossimilhança.

Parker & Rothenberg(1988) afirma que, devido a violações causadas à teoria assintótica do teste da razão de verossimilhança, não existe um método formal para determinar o valor de  $k$ . Dessa maneira, o autor determina o número de componentes na mistura através do “grau de ajuste” (goodness-of-fit) de cada modelo, selecionando aquele  $k$  que mais “adequadamente” ajuste os dados. A maneira de comparar os modelos é através da estatística Cramer-von Mises (CVM).

Polymenis & Titterington (1998) apresenta uma técnica, para determinação do  $k$ , baseada nos autovalores da matriz de informação de Fisher.

Vlassis et al. (1999) propõe uma técnica que emprega o uso da verossimilhança, testes envolvendo média, variância e curtose para decidir o número de *kernels* a ser usado quando da estimação da densidade de uma mistura.

James et al. (2001) apresenta uma metodologia semiparamétrica, afirmando uma convergência quase certa para o número real de parâmetros.

Allison et al. (2002) apresenta um método comparativo sobre o ganho na verossimilhança, utilizando bootstrap, para obter a complexidade da mistura. O autor utiliza o modelo de mistura para análise de dados de microarranjos.

Steinholff et al. (2003) utiliza misturas de normais para estudar a intensidade da expressão gênica em microarranjos. O autor utiliza o *Critério de Informação Bayesiano* (BIC) para estimar a complexidade do modelo.

### 4.3.2 Interpretação de Modelos de Misturas

Uma maneira clara de gerar uma amostra aleatória  $\mathbf{X}_1, \dots, \mathbf{X}_n$  da densidade (4.3.1) pode ser assim descrita. Seja  $Z_j$  uma variável aleatória categorizada, assumindo valores  $1, 2, \dots, k$ , com probabilidades  $p_1, p_2, \dots, p_k$ , respectivamente. Suponha que a densidade condicional de  $\mathbf{X}_j$  dado  $Z_j = i$  é  $f_i(\mathbf{x}_j)$ ,  $i = 1, 2, \dots, k$ . Então, a densidade incondicional de  $\mathbf{X}_j$ , ou seja, sua densidade marginal, é dada por  $f(\mathbf{x}_j)$ . Nesse contexto, a variável  $Z_j$  pode ser interpretada como um “rótulo” do vetor  $\mathbf{X}_j$ . Posteriormente, será conveniente trabalhar com um vetor  $k$ -dimensional  $\mathbf{Z}_j$ , ao invés da variável categorizada  $Z_j$ , na qual o  $i$ -ésimo elemento,  $Z_{ij} = (Z_j)_i$  é definido como um ou zero quando  $\mathbf{X}_j$  pertence ou não à componente  $i$ ,  $i = 1, 2, \dots, k$ . Assim,  $\mathbf{Z}_j$  tem distribuição multinomial da forma:

$$\mathbb{P}(\mathbf{Z}_j = \mathbf{z}_j) = p_1^{z_{1j}} p_2^{z_{2j}} \dots p_k^{z_{kj}},$$

sendo representada como  $\mathbf{Z}_j \sim \text{Mult}_k(1, \boldsymbol{\pi})$ , com  $\boldsymbol{\pi} = (p_1, p_2, \dots, p_k)'$  (McLachlan & Peel, 2000).

Uma situação clara em que a mistura pode ser interpretada como acima é quando  $\mathbf{X}_j$  são amostradas de uma população  $G$  constituída por  $k$  grupos,  $G_1, \dots, G_k$ , com proporções  $p_1, \dots, p_k$ . Em aplicações biológicas, pode-se considerar como fontes de heterogeneidade a idade, sexo, espécie e origem geográfica. Por exemplo, a população  $G$  pode ser constituída por dois grupos,  $G_1$  e  $G_2$ , correspondendo a indivíduos com e sem determinada doença. O problema pode ser estimar a prevalência da doença, ou seja, a proporção  $p_1$  da mistura, com base em um vetor amostral selecionado da população. Um exemplo assim pode ser encontrado em Do & McLachlan (1984) (McLachlan & Peel, 2000).

No entanto, existem situações em que o número de componentes não é conhecido *a priori*. Em alguns casos, as componentes são introduzidas no modelo para permitir maior flexibilidade na modelagem de populações heterogêneas que não sejam bem estimadas através de uma componente apenas. Num caso extremo, usando  $k = n$  e proporções iguais para cada componente  $p_i = 1/n$ , pode-se obter uma estimação por função núcleo, da forma

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_j}{h}\right),$$

fazendo  $f_i(x) = h^{-1}K((x - x_i)h^{-1})$  em (4.3.1). A função  $K$  representa o a função núcleo e  $H$  o parâmetro de alisamento, como definido na Seção 6.1 (McLachlan & Peel, 2000).

Assim, para valores de  $k$  entre 1 e  $n$ , modelos de mistura podem ser vistos um “equilíbrio semiparamétrico” entre (1) o modelo paramétrico completo, representado por uma família paramétrica ( $k = 1$ ), e (2) um modelo não paramétrico representado no caso de  $k = n$  pelo método de estimação de densidade por função núcleo. Por exemplo, embora uma única distribuição normal é claramente inadequada para modelar uma distribuição contínua assimétrica, a mistura de duas distribuições normais pode fornecer um ajuste adequado (McLachlan & Peel, 2000).

### 4.3.3 Identificabilidade

Em geral, uma família paramétrica de funções densidade de probabilidade é dita identificável se valores distintos dos parâmetros determinam distintos membros da família. No caso de famílias de misturas, é necessária uma interpretação especial. Podemos dizer que a função densidade da mistura  $f(x|\boldsymbol{\psi})$ , definida em (4.3.1), é “economicamente representada” se, para cada par de inteiros  $i$  e  $j$ , entre 1 e  $k$ , tem-se que  $f_i(x|\boldsymbol{\psi}_i) = f_j(x|\boldsymbol{\psi}_j)$  para quase todo  $x \in \mathbb{R}$ , se e somente se  $i = j$  ou um de  $f_i$  e  $f_j$  for nulo. Assim é suficiente dizer que a família de densidades da mistura dada por (4.3.1) é identificável para  $\boldsymbol{\psi} \in \Theta$  se para todo par  $\boldsymbol{\psi}' = (p'_1, \dots, p'_k, \theta'_1, \dots, \theta'_k)$  e  $\boldsymbol{\psi}'' = (p''_1, \dots, p''_k, \theta''_1, \dots, \theta''_k)$  em  $\Theta$  determinam densidades  $p(x|\boldsymbol{\psi}')$  e  $p(x|\boldsymbol{\psi}'')$  representadas economicamente. Tem-se que  $f(x|\boldsymbol{\psi}') = f(x|\boldsymbol{\psi}'')$  para quase todo  $x \in \mathbb{R}$  apenas se existe uma permutação  $\delta$  de  $(1, \dots, k)$  tal que  $p'_i = p''_{\delta(i)}$  e, se  $p'_i \neq 0$ ,  $\theta'_i = \theta''_{\delta(i)}$ ,  $i = 1, \dots, k$  (Redner & Walker, 1984; McLachlan & Basford, 1988).

Dessa forma, se todas as componentes  $f_i(x|\theta_i)$  pertencem a mesma família paramétrica, então  $f(x|\boldsymbol{\psi})$  é invariante sobre  $k!$  permutações dos rótulos das componentes em  $\boldsymbol{\psi}$ . Nesse caso, se  $\hat{\boldsymbol{\psi}}$  corresponde a um máximo local,  $L(\mathbf{x}|\boldsymbol{\psi})$  terá ao menos  $k!$  máximos locais de mesmo valor. Sendo assim, se todas as componentes pertencerem a mesma família paramétrica, basta acrescentar a restrição  $p_1 \geq p_2 \geq \dots \geq p_k$  (McLachlan & Basford, 1988).

## Capítulo 5

# Estimação Paramétrica em Misturas de Densidades

Uma grande variedade de métodos foram desenvolvidos ao longo dos anos. Podemos citar métodos gráficos, momentos, distância mínima, máxima verossimilhança e bayesianos. Provavelmente, a existência de tantos métodos se deve ao fato de que fórmulas fechadas para os parâmetros normalmente não são possíveis. Por exemplo, não é possível encontrar fórmulas fechadas para os estimadores de máxima verossimilhança (EMV) para as proporções da mistura, médias e variâncias/covariâncias das componentes de uma mistura de normais. Esses estimadores têm de ser encontrados iterativamente, como discutido a frente (McLachlan & Peel, 2000).

Uma maneira de estimar o vetor de parâmetros  $\psi$  é através da minimização de  $\delta(F_n, F_\psi)$ , a distância entre a distribuição da mistura  $F_\psi$  e a distribuição empírica  $F_n$ . Este método é conhecido como estimação por distância mínima (McLachlan & Peel, 2000).

O objetivo é estimar uma densidade desconhecida a partir de uma seqüência de amostras. Uma proposta é aproximar a densidade com uma mistura finita de *kernels* gaussianos cujos parâmetros e pesos são estimados iterativamente, a partir da amostra, usando máxima verossimilhança (Vlassis et al., 1999).

Em 1978, Murray & Titterington apresentou abordagens paramétricas e não-paramé-

tricas para o problema de estimação. Para o caso paramétrico, o autor afirma que uma solução explícita para as equações de verossimilhança geralmente não é possível e muitos casos numericamente estáveis recaem sobre o algoritmo EM. O autor também apresenta a possibilidade do uso de uma abordagem bayesiana. Já na abordagem não paramétrica, discute o uso de métodos de *kernel*. Apresenta também um algoritmo EM-tipo para funções *kernel*.

Aitkin & Rubin (1985) impõe uma distribuição *a priori* nos pesos da mistura,  $\boldsymbol{\pi}$ , estimando os parâmetros  $\boldsymbol{\theta}$  por máxima verossimilhança a partir de  $L(\boldsymbol{\theta}|\mathbf{X})$ , obtido pela integração de  $L(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{X})$  sobre a distribuição a priori de  $\boldsymbol{\pi}$ .

Titterington et al. (1985) apresenta uma variedade de técnicas de estimação de densidade de misturas: (1) método dos momentos; (2) máxima verossimilhança, introduzindo o algoritmo EM, além de Newton-Raphson e o Método de Escore; (3) métodos bayesianos; (4) distância mínima baseada em funções de distribuição.

McLachlan & Basford (1988) apresenta a estimação dos pesos da mistura através da função de verossimilhança, de análise discriminante e método dos momentos. O autor também utiliza, sobre esses procedimentos, o método de distância mínima através da escolha de uma distância apropriada.

Helmbold et al. (1995) investiga o problema de estimar o vetor de proporções que maximiza a verossimilhança de uma dada amostra para uma mistura de densidades conhecidas. O autor faz uma comparação entre alguns algoritmos usados para tanto.

Bartolucci & Scaccia (2005) faz uso da estimação de misturas para tratar erros de regressão não-normais. Utiliza uma abordagem semiparamétrica para modelar tais erros. O algoritmo EM é empregado para estimação dos parâmetros e o bootstrap para construção de intervalos de confiança.

Karunamunia et al. (2006) investiga as propriedades de convergência de uma versão prática do estimador adaptativo de Cutler & Cordero-Braña, sob condições de regularidade, comparando-as com as do estimador não-adaptativo. O estimador de Cutler & Cordero-Braña é um estimador adaptativo por *kernel* para a distância de Hellinger em

modelos mistos.

Karvanen (2006) utiliza “L-moments” para a estimação de quantis de misturas.

Pokhyl’ko (2005) apresenta estimadores para densidade usando *wavelet* sobre dados originados de uma mistura de várias componentes. O autor constrói estimadores lineares e adaptativos, provando sua convergência em norma quadrática e em norma uniforme.

## 5.1 Estimação de máxima verossimilhança

Casella & Berger (2002) afirma que o método de máxima verossimilhança é o mais popular para estimação dos parâmetros de uma distribuição. Sendo  $X_1, \dots, X_n$  uma amostra aleatória independente e identicamente distribuída de uma população com função densidade de probabilidade  $f(x|\theta_1, \dots, \theta_k)$ , a função de verossimilhança é definida por

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

Para cada conjunto amostrado  $\mathbf{x}$ , seja  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  o valor dos parâmetros em que  $L(\boldsymbol{\theta}|\mathbf{x})$  assume o máximo como função de  $\boldsymbol{\theta}$ , com  $\mathbf{x}$  fixo. Portanto, os autores definem o *estimador de máxima verossimilhança* (EMV) de um parâmetro  $\boldsymbol{\theta}$  baseado numa amostra  $\mathbf{X}$  como  $\hat{\boldsymbol{\theta}}(\mathbf{X})$ .

Intuitivamente, o EMV é uma escolha razoável dos parâmetros. O EMV representa o ponto no espaço paramétrico em que a amostra é mais provável. Em geral, o estimador de máxima verossimilhança é uma boa estimativa, possuindo boas propriedades de otimalidade (Casella & Berger, 2002).

Quando a função de verossimilhança se mostra difícil de trabalhar, opta-se pela maximização do logaritmo da mesma, ou seja,  $l(\boldsymbol{\theta}|\mathbf{x}) = \log L(\boldsymbol{\theta}|\mathbf{x})$ . Sendo  $L(\boldsymbol{\theta}|\mathbf{x}) > 0$  e o logaritmo uma função monótona crescente, a maximização de  $L(\boldsymbol{\theta}|\mathbf{x})$  equivale à maximização de  $l(\boldsymbol{\theta}|\mathbf{x})$ .

Para mostrar que o estimador de máxima verossimilhança é consistente, se pode recor-

rer ao Teorema 1 (Lehmann & Casella, 2001). Seja  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_s)'$  e considere as seguintes suposições.

- (A) As distribuições  $\mathbb{P}_\theta$  das observações são distintas (caso contrário,  $\theta$  não pode ser estimado consistentemente).
- (B) As distribuições  $\mathbb{P}_\theta$  têm suporte comum.
- (C) As observações são  $\mathbf{X} = (X_1, \dots, X_n)$ , em que  $X_i$  são variáveis aleatórias independentes e identicamente distribuídas (iid) com função densidade de probabilidade  $f(x_i|\boldsymbol{\theta})$ .
- (D) O espaço paramétrico  $\Omega$  contém um aberto  $\omega$  de qual o verdadeiro valor do parâmetro  $\boldsymbol{\theta}_0$  é ponto interior, tal que para quase todo  $x$ , a densidade  $f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i|\boldsymbol{\theta})$  admite todas as terceiras derivadas  $(\partial^3/\partial\theta_j\partial\theta_k\partial\theta_l)f(\mathbf{x}|\boldsymbol{\theta})$ ,  $\forall \boldsymbol{\theta} \in \omega$ .
- (E) As primeiras e segundas derivadas do logaritmo de  $f(\mathbf{x}|\boldsymbol{\theta})$  satisfazem as equações

$$\mathbb{E}_\theta \left[ \frac{\partial}{\partial\theta_j} \log f(\mathbf{X}|\boldsymbol{\theta}) \right] = 0; \quad j = 1, \dots, s$$

$$\begin{aligned} I_{jk}(\boldsymbol{\theta}) &= \mathbb{E}_\theta \left[ \frac{\partial}{\partial\theta_j} \log f(\mathbf{X}|\boldsymbol{\theta}) \cdot \frac{\partial}{\partial\theta_k} \log f(\mathbf{X}|\boldsymbol{\theta}) \right] \\ &= \mathbb{E}_\theta \left[ -\frac{\partial^2}{\partial\theta_j\partial\theta_k} \log f(\mathbf{X}|\boldsymbol{\theta}) \right]. \end{aligned}$$

Claramente, essas duas equações implicam em

$$I_{jk}(\boldsymbol{\theta}) = Cov \left[ \frac{\partial}{\partial\theta_j} \log f(\mathbf{X}|\boldsymbol{\theta}), \frac{\partial}{\partial\theta_k} \log f(\mathbf{X}|\boldsymbol{\theta}) \right]$$

- (F) Sendo  $I(\boldsymbol{\theta})_{s \times s}$  uma matriz de variâncias e covariâncias, deve ser definida positiva. Devemos assumir que  $I_{jk}(\boldsymbol{\theta})$  é finito e que a matriz  $I(\boldsymbol{\theta})$  é positiva definida para todo  $\boldsymbol{\theta}$  em  $\omega$ , então, as  $s$  estatísticas

$$\frac{\partial}{\partial\theta_1} \log f(\mathbf{X}|\boldsymbol{\theta}), \dots, \frac{\partial}{\partial\theta_s} \log f(\mathbf{X}|\boldsymbol{\theta})$$

são independentes com probabilidade 1.

(G) Supõe-se que existam funções  $M_{jkl}$  tais que

$$\left| \frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \log f(\mathbf{X}|\boldsymbol{\theta}) \right| \leq M_{jkl}(\mathbf{x}), \quad \forall \boldsymbol{\theta} \in \omega$$

em que

$$m_{jkl} = \mathbb{E}_{\boldsymbol{\theta}} [M_{jkl}(\mathbf{X})] < \infty, \quad \forall j, k, l.$$

**Teorema 1** (Lehmann & Casella, 2001). *Sejam  $X_1, \dots, X_n$  variáveis aleatórias iid, cada com função densidade de probabilidade  $f(x|\boldsymbol{\theta})$ , satisfazendo (A)-(G) acima. Então, com probabilidade tendendo à 1 quando  $n \rightarrow \infty$ , existem soluções  $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(X_1, \dots, X_n)$  das equações de máxima verossimilhança tais que*

(a)  $\hat{\theta}_{j_n}$  é consistente para estimar  $\theta_j$ ,

(b)  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  é assintoticamente normal com vetor de médias nulo e matriz de variâncias-covariâncias  $[I(\boldsymbol{\theta})]^{-1}$ , e

(c)  $\hat{\theta}_{j_n}$  é assintoticamente eficiente no sentido de

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N \left\{ 0, [I(\boldsymbol{\theta})]_{jj}^{-1} \right\}$$

No caso de  $X \sim G3B(\alpha, \beta, \lambda)$ , com  $\boldsymbol{\theta} = (\alpha, \beta, \lambda)'$ , a função densidade de probabilidade é dada por

$$f(x|\boldsymbol{\theta}) = \frac{\lambda^\alpha x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta) (1 - (1-\lambda)x)^{\alpha+\beta}}, \quad (5.1.1)$$

em que

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)},$$

sendo  $\Gamma(y)$  a função gama aplicada no ponto  $y$ .

As equações abaixo mostram as primeiras derivadas de  $f(x|\boldsymbol{\theta})$  nos parâmetros.

$$\frac{\partial}{\partial \alpha} f(x|\boldsymbol{\theta}) = f(x|\boldsymbol{\theta}) \left[ \Psi(\alpha + \beta) - \Psi(\alpha) + \log \left( \frac{\lambda x}{1 - (1-\lambda)x} \right) \right],$$

$$\frac{\partial}{\partial \beta} f(x|\boldsymbol{\theta}) = f(x|\boldsymbol{\theta}) \left[ \Psi(\alpha + \beta) - \Psi(\beta) + \log \left( \frac{(1-x)}{1 - (1-\lambda)x} \right) \right] e$$

$$\frac{\partial}{\partial \lambda} f(x|\boldsymbol{\theta}) = f(x|\boldsymbol{\theta}) \left[ \frac{\alpha}{\lambda} - \frac{(\alpha + \beta)x}{1 - (1 - \lambda)x} \right].$$

Os cálculos foram realizados para segundas e terceiras derivadas, mas as equações são aqui suprimidas.

As equações de verossimilhança são dadas por

$$\Psi(\alpha + \beta) - \Psi(\alpha) = \frac{1}{n} \sum_{i=1}^n \log \left[ \frac{\lambda x_i}{1 - (1 - \lambda)x_i} \right],$$

$$\Psi(\alpha + \beta) - \Psi(\beta) = \frac{1}{n} \sum_{i=1}^n \log \left[ \frac{1 - x_i}{1 - (1 - \lambda)x_i} \right] \text{ e}$$

$$\frac{n\alpha}{\lambda} = (\alpha + \beta) \sum_{i=1}^n \frac{\lambda x_i}{1 - (1 - \lambda)x_i}.$$

É fácil ver que a matriz de informação de Fisher, como apresentada por Pham-Gia & Duong (1989), é dada por

$$I(\boldsymbol{\theta}) = n \begin{bmatrix} \Psi'(\alpha) - \Psi'(\alpha + \beta) & -\Psi'(\alpha + \beta) & \frac{-\beta}{\lambda(\alpha + \beta)} \\ & \Psi'(\beta) - \Psi'(\alpha + \beta) & \frac{\alpha}{\lambda(\alpha + \beta)} \\ & & \frac{\alpha\beta}{\lambda^2(\alpha + \beta + 1)} \end{bmatrix}.$$

As terceiras derivadas para função de log-verossimilhança são mostradas abaixo.

$$\frac{\partial^3}{\partial \alpha^3} \log f(\mathbf{x}|\boldsymbol{\theta}) = n\Psi''(\alpha + \beta) - n\Psi''(\alpha)$$

$$\frac{\partial^3}{\partial \alpha^2 \partial \beta} \log f(\mathbf{x}|\boldsymbol{\theta}) = n\Psi''(\alpha + \beta)$$

$$\frac{\partial^3}{\partial \alpha^2 \partial \lambda} \log f(\mathbf{x}|\boldsymbol{\theta}) = 0$$

$$\frac{\partial^3}{\partial \beta^3} \log f(\mathbf{x}|\boldsymbol{\theta}) = n\Psi''(\alpha + \beta) - n\Psi''(\beta)$$

$$\frac{\partial^3}{\partial \beta^2 \partial \alpha} \log f(\mathbf{x}|\boldsymbol{\theta}) = n\Psi''(\alpha + \beta)$$

$$\frac{\partial^3}{\partial \beta^2 \partial \lambda} \log f(\mathbf{x}|\boldsymbol{\theta}) = 0$$

$$\frac{\partial^3}{\partial \lambda^3} \log f(\mathbf{x}|\boldsymbol{\theta}) = \frac{2n\alpha}{\lambda^2} - (\alpha + \beta) \sum_{i=1}^n \frac{2x_i^3}{(1 - (1 - \lambda)x_i)^3}$$

$$\mathbb{E}_{\boldsymbol{\theta}} \left[ \frac{\partial^3}{\partial \lambda^3} \log f(\mathbf{x}|\boldsymbol{\theta}) \right] = \frac{2n\alpha}{\lambda^2} - \frac{2n\alpha(\alpha + 1)(\alpha + 2)}{\lambda^3(\alpha + \beta + 1)(\alpha + \beta + 2)}$$

$$\frac{\partial^3}{\partial \lambda^2 \partial \alpha} \log f(\mathbf{x}|\boldsymbol{\theta}) = -\frac{1}{\lambda^2} + \sum_{i=1}^n \frac{x_i^2}{(1 - (1 - \lambda)x_i)^2}$$

$$\mathbb{E}_{\boldsymbol{\theta}} \left[ \frac{\partial^3}{\partial \lambda^2 \partial \alpha} \log f(\mathbf{x}|\boldsymbol{\theta}) \right] = -\frac{1}{\lambda^2} - \frac{2n\alpha(\alpha + 1)}{\lambda^2(\alpha + \beta + 1)}$$

$$\frac{\partial^3}{\partial \lambda^2 \partial \beta} \log f(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^n \frac{x_i^2}{(1 - (1 - \lambda)x_i)^2}$$

$$\mathbb{E}_{\boldsymbol{\theta}} \left[ \frac{\partial^3}{\partial \lambda^2 \partial \beta} \log f(\mathbf{x}|\boldsymbol{\theta}) \right] = \frac{2n\alpha(\alpha + 1)}{\lambda^2(\alpha + \beta + 1)}$$

Pode-se mostrar que existem funções  $M_{jkl}$  tais que  $|(\partial^3/\partial\theta_j\partial\theta_k\partial\theta_l) \log f(\mathbf{x}|\boldsymbol{\theta})| \leq M_{jkl}(\mathbf{x})$ ,  $\forall \boldsymbol{\theta} \in \omega$ , tal que  $m_{jkl} = \mathbb{E}_{\boldsymbol{\theta}} [M_{jkl}(\mathbf{X})] < \infty$ ,  $\forall j, k, l$ .

Sendo assim, as suposições necessárias para validade do Teorema 1 são atendidas, resultando em boas propriedades do estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ . No caso de misturas de uma uniforme em  $(0, 1)$  e outra  $G3B$ , pode-se mostrar essas boas características para os estimadores de máxima verossimilhança.

Quanto à família de misturas, sua identificabilidade já foi discutida na Seção 4.3.3.

## 5.2 Métodos de Busca

### 5.2.1 Busca direta

O método de busca direta, na procura dos estimadores de máxima verossimilhança, testa inúmeras combinações dos parâmetros, selecionando aqueles que maximizam a verossimilhança. Não é um bom método, pois deveria testar todas as combinações possíveis dos parâmetros, o que é inviável, sendo o espaço paramétrico contínuo. Considere uma densidade com três parâmetros,  $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]$ , sendo que  $\theta_1, \theta_2, \theta_3 > 0$ . Se variássemos os valores de 0,5 em 0,5, limitando a busca a  $[0, 10] \times [0, 10] \times [0, 10]$ , teríamos que calcular a verossimilhança em  $21^3$  (9.261) pontos. Aumentando mais um parâmetro, teríamos que testar 194.481 pontos. No entanto, um incremento de 0,5 é muito grande. Usando incremento de 0,05, com três parâmetros, seriam 8.120.601 avaliações.

### 5.2.2 Busca de Fibonacci

A busca pelo mínimo de uma função  $f(x)$  pode ser refinada pelo uso da *regra de ouro*, que emprega os números de Fibonacci para a secção do intervalo de busca. Supondo que o mínimo esteja no intervalo  $(x, y)$  conhecido, é possível escolher dois pontos,  $z$  e  $w$ , tal que  $x < z < w < y$ . Assumindo que a função  $f(x)$  é unimodal em  $(x, y)$ , é fácil ver que, se  $f(z) \geq f(w)$ , então o mínimo se encontra em  $(z, y)$ . Caso  $f(z) \leq f(w)$ , então o mínimo se encontra em  $(x, w)$ . Isso é ilustrado na Figura 5.2.1. Avaliando a função em outros pontos, é possível reduzir, iterativamente, o intervalo de busca (Everitt, 1987).

O problema agora é selecionar os pontos em que a função é avaliada. Se especificarmos o número de vezes que a função será avaliada,  $n$ , pode-se mostrar que o método mais eficiente, em termos de redução do intervalo de busca (Tabela 5.2.1), é a busca de Fibonacci (Walsh, 1975). Faz-se necessária a definição dos números de Fibonacci (Everitt, 1987):

$$F_0 = F_1 = 1$$

Tabela 5.2.1: Taxa de redução no intervalo de busca para diferentes  $n$ 's (Everitt, 1987).

Valor de $n$	$F_n$	Redução intervalo
0	1	1,0
1	1	1,0
2	2	0,50
3	3	0,3333
4	5	0,20
5	8	0,1250
10	89	0,01124
11	144	0,006944
12	233	0,004292
13	377	0,002653
14	610	0,001639
15	987	0,001013
20	10946	0,00009136

$$F_i = F_{i-1} + F_{i-2}, \quad i \geq 2,$$

resultando na seqüência 1, 1, 2, 3, 5, 8, 13, 21, .... Sendo  $(x^{(1)}, y^{(1)})$  o intervalo inicial, então o primeiro passo da busca é fazer

$$I_1 = \frac{F_{n-1}}{F_n} I_0$$

em que  $I_0 = y^{(1)} - x^{(1)}$  é o comprimento do intervalo inicial. O próximo intervalo em que  $f(x)$  será avaliada é dado por

$$z^{(1)} = x^{(1)} + I_1 \text{ e}$$

$$w^{(1)} = y^{(1)} - I_1.$$

Usando a unimodalidade, parte do intervalo é descartado, como descrito acima, gerando o método iterativo:

$$I_i = \frac{F_{n-i}}{F_n} I_0$$

$$z^{(i)} = x^{(i)} + I_i,$$

$$w^{(i)} = y^{(i)} - I_i, \quad 0 < i \leq n, \quad i \in \mathbb{N}.$$

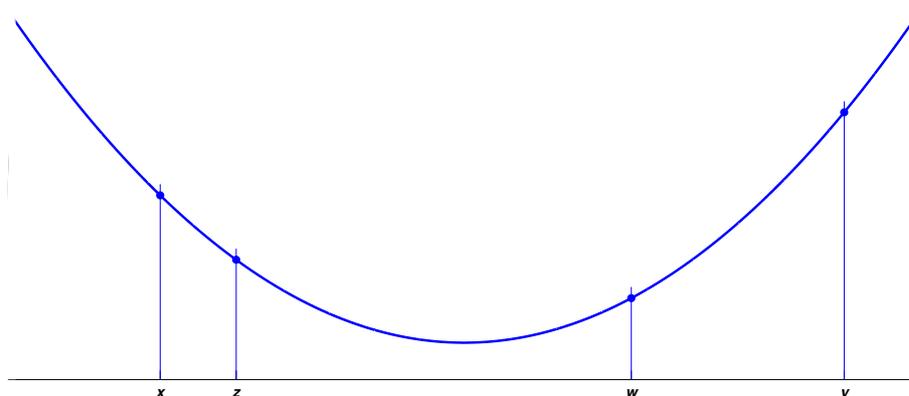


Figura 5.2.1: Etapa da busca de Fibonacci (Everitt, 1987).

### 5.3 Método *EM*

O algoritmo EM (Esperança-Maximização) é muito aplicado na obtenção iterativa de estimativas de máxima verossimilhança (MV), muito útil em problemas de dados incompletos, nos quais métodos como o de Newton-Raphson se tornam muito complicados. A cada iteração do algoritmo dois passos ocorrem: um passo de *esperança* e outro passo de *maximização*. O algoritmo foi primeiro descrito por Dempster, Laird & Rubin (1977). O procedimento é muito similar ao de estimação com dados incompletos, no qual os parâmetros são estimados após os dados serem inicialmente completados. Estes são então atualizados com seus valores preditos usando as estimativas iniciais dos parâmetros. Os parâmetros são re-estimados, repetindo-se o procedimento até que haja convergência (McLachlan & Krishnan, 1997).

Embora a primeira vista um problema possa não parecer com um de dados incompletos, pode existir um grande ganho computacional virtualmente formulando-o como um, facilitando a estimação de máxima verossimilhança. Isso ocorre pois o algoritmo EM explora a redução na complexidade do estimador de máxima verossimilhança dado os dados completos. Para muitos problemas estatísticos, a verossimilhança para dados completos apresenta-se mais simples. O passo E consiste e “produzir” dados que completem o pro-

blema, usando as observações existentes para o problema incompleto e os valores atuais dos parâmetros. Assim, o passo M, mais simples, pode ser aplicado a esse conjunto “completo”. Mais precisamente, é a log-verossimilhança dos dados completos que é “gerada” no passo E. Como é parcialmente baseada em dados não observados, é substituída pela esperança condicionada nos dados observados, usando as estimativas iniciais dos parâmetros. Começando com valores iniciais adequados, os passos E e M são repetidos até que haja convergência (McLachlan & Krishnan, 1997).

Segundo McLachlan & Krishnan (1997), suponha  $f(\mathbf{y}|\boldsymbol{\theta})$  uma função densidade de probabilidade do vetor aleatório  $\mathbf{Y}$  de uma mistura de  $k$  componentes, ou seja,

$$f(\mathbf{y}|\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{i=1}^k p_i f_i(\mathbf{y}|\boldsymbol{\theta}_i),$$

em que  $\boldsymbol{\pi} = (p_1, \dots, p_{k-1})'$  é o vetor contendo as proporções da mistura, logo,  $p_k = 1 - \sum_{i=1}^{k-1} p_i$ , e  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_k)'$ , o vetor de parâmetros das componentes. Nesse caso, a função de log-verossimilhança é dada por

$$\log L(\boldsymbol{\Psi}) = \sum_{j=1}^n \log f(\mathbf{y}_j|\boldsymbol{\Psi}),$$

em que  $\boldsymbol{\Psi} = (\boldsymbol{\pi}', \boldsymbol{\theta}')'$ . Pela diferenciação com respeito à  $\boldsymbol{\pi}$  e  $\boldsymbol{\theta}$ , pode-se equacionar as funções de verossimilhança

$$\sum_{j=1}^n \left\{ \frac{f_i(\mathbf{y}_j|\boldsymbol{\theta}_i)}{f(\mathbf{y}_j|\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})} - \frac{f_k(\mathbf{y}_j|\boldsymbol{\theta}_k)}{f(\mathbf{y}_j|\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})} \right\}, \quad i = 1, \dots, k-1,$$

as quais, visivelmente, não resultarão em solução explícita para

$$\hat{\boldsymbol{\pi}} = (\hat{p}_1, \dots, \hat{p}_{k-1})' \quad \text{e} \quad \hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_k)'.$$

Como apresentado na Seção 4.3.2, a introdução do vetor indicador  $\mathbf{Z}_j$ , especifica a qual componente da mistura a variável  $j$  está determinada. A construção do problema de mistura dessa forma, ou seja, em termos de  $\mathbf{Y}_j$  e  $\mathbf{Z}_j$  é mais útil no processo de estimação por máxima verossimilhança, através do algoritmo EM. Nesse método, os dados são vistos

como provenientes de um problema de dados incompletos, sendo que os vetores indicadores  $\mathbf{z}_1, \dots, \mathbf{z}_n$  não são observáveis (McLachlan & Peel, 2002).

Para desenhar tal problema como de dados incompletos, são introduzidas os dados faltantes, não observáveis,

$$\mathbf{z} = (\mathbf{z}'_1, \dots, \mathbf{z}'_n)',$$

em que  $\mathbf{z}_j$  é um vetor  $k$ -dimensional de 0's e 1's, sendo  $z_{ij} = (\mathbf{z}_j)_i$  igual a 1 se  $\mathbf{x}_j$  vem da componente  $i$ , ou 0 caso contrário ( $i = 1, \dots, k; j = 1, \dots, n$ ). Os vetores aleatórios  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ , sob independência, assumem a distribuição multinomial  $\mathbf{Z}_j \sim \text{Mult}_k(1, \boldsymbol{\pi})$  (McLachlan & Peel, 2002).

A  $i$ -ésima proporção da mistura,  $p_i$ , pode ser interpretada como a probabilidade *a priori* de que a observação pertença à  $i$ -ésima componente da mistura ( $i = 1, \dots, k$ ). A probabilidade *a posteriori* de que pertença à componente  $i$ , sendo que  $\mathbf{y}_j$  foi observado é dada por

$$\begin{aligned} \tau_i(\mathbf{y}_j) &= \mathbb{P}(\text{pertença à } i\text{-ésima componente} | \mathbf{y}_j) \\ &= \mathbb{P}(Z_{ij} = 1 | \mathbf{y}_j) \\ &= \frac{p_i f_i(\mathbf{y}_j)}{f(\mathbf{y}_j)}, \quad i = 1, \dots, k; j = 1, \dots, n. \end{aligned}$$

Caso  $z_{ij}$  fosse observável, o estimador de máxima verossimilhança de  $p_i$  seria dado por

$$\sum_{j=1}^n z_{ij} / n, \quad i = 1, \dots, k,$$

assim como os estimadores de  $\boldsymbol{\theta}_i$  poderiam ser calculados diretamente, a partir das observações pertencentes à cada componente  $i$ .

O algoritmo EM lida com a adição de dados não observáveis ao problema através da esperança condicional da log-verossimilhança completa dado os dados observados. O vetor de dados completos pode ser definido como  $\mathbf{x} = (\mathbf{y}', \mathbf{z}')$ , e a log-verossimilhança completa como

$$\log L(\boldsymbol{\Psi} | \mathbf{y}, \mathbf{z}) = \sum_{i=1}^k \sum_{j=1}^n z_{ij} [\log p_i + \log f_i(\mathbf{y}_j; \boldsymbol{\theta}_i)].$$

Na estimação de máxima verossimilhança, o objetivo é maximizar  $L(\Psi|\mathbf{y})$ , mas o problema seria muito mais simples completando os dados e trabalhando com a distribuição conjunta  $f(\mathbf{y}, \mathbf{z}|\Psi) = L(\Psi|\mathbf{y}, \mathbf{z})$  e a densidade condicional de  $\mathbf{Z}$  dado  $\mathbf{y}$ , ou seja,

$$q(\mathbf{z}|\boldsymbol{\pi}, \mathbf{y}) = \frac{f(\mathbf{y}, \mathbf{z}|\Psi)}{g(\mathbf{y}|\boldsymbol{\theta})},$$

resultando em

$$\log L(\Psi|\mathbf{y}) = \log L(\Psi|\mathbf{y}, \mathbf{z}) - \log q(\mathbf{z}|\boldsymbol{\pi}, \mathbf{y}). \quad (5.3.1)$$

Como  $\mathbf{z}$  não está disponível, o lado direito da equação (5.3.1) é substituída pela sua esperança, usando a distribuição condicional de  $\mathbf{Z}$  dado  $\mathbf{y}$ . Com chute inicial  $\Psi_0 = (\boldsymbol{\pi}'_0, \boldsymbol{\theta}'_0)'$  definimos

$$Q(\Psi|\Psi_0, \mathbf{y}) = \int \log L(\Psi|\mathbf{y})q(\mathbf{z}|\boldsymbol{\pi}_0, \mathbf{y})d\mathbf{z}, \quad (5.3.2)$$

$$H(\Psi|\Psi_0, \mathbf{y}) = \int \log q(\mathbf{z}|\boldsymbol{\pi}, \mathbf{y})q(\mathbf{z}|\boldsymbol{\pi}_0, \mathbf{y})d\mathbf{z}.$$

Como o lado esquerdo da equação (5.3.1) não depende de  $\mathbf{Z}$ , o valor esperado de  $L(\Psi|\mathbf{y})$  é dado por

$$L(\Psi|\mathbf{y}) = Q(\Psi|\Psi_0, \mathbf{y}) - H(\Psi|\Psi_0, \mathbf{y}). \quad (5.3.3)$$

Seja  $\widehat{\Psi}_{(1)}$  o valor de  $\Psi$  que maximiza  $Q(\Psi|\Psi_0, \mathbf{y})$ . O processo é repetido com  $\Psi_0$  em (5.3.2) substituído pelo valor atualizado  $\widehat{\Psi}_{(1)}$ . Dessa forma, uma seqüência de estimadores  $\widehat{\Psi}_{(w)}$ ,  $w = 1, 2, \dots$ , é obtida iterativamente, de forma que  $\widehat{\Psi}_{(w)}$  é definido como o valor de  $\Psi$  que maximiza  $Q(\Psi|\widehat{\Psi}_{(w-1)}, \mathbf{y})$ , ou seja, (Lehmann & Casella, 2001)

$$Q\left(\widehat{\Psi}_{(w)}|\widehat{\Psi}_{(w-1)}, \mathbf{y}\right) = \max_{\Psi} Q\left(\Psi|\widehat{\Psi}_{(w-1)}, \mathbf{y}\right).$$

McLachlan & Krishnan (1997) mostram que a função de verossimilhança dos dados incompletos,  $L(\boldsymbol{\theta})$ , é não decrescente após uma iteração do algoritmo EM, ou seja,

$$L(\Psi_{(k+1)}) > L(\Psi_k) \quad (5.3.4)$$

para  $k = 1, 2, \dots$ . Assim, para uma seqüência limitada de valores da verossimilhança  $\{L(\Psi_{(k)})\}$ ,  $L(\Psi_{(k)})$  converge monotonicamente para algum  $L^*$ . Uma conseqüência de 5.3.4 é a auto-consistência do algoritmo EM.

Em situações em que  $Q(\Psi|\Psi_0, \mathbf{y})$  não existe como uma fórmula fechada, o método de Newton-Raphson (Seção 5.4.1) pode ser considerado para obter um resultado numérico iterativo para  $\Psi^{(k+1)}$ . Como mostrado acima, basta que a função  $Q$  seja maximizada para que a função de verossimilhança seja não-decrescente. Sendo assim, é suficiente que

$$Q(\Psi_{(k+1)}|\Psi_{(k)}, \mathbf{y}) \geq Q(\Psi_{(k)}|\Psi_{(k)}, \mathbf{y})$$

para garantir a monotonicidade da seqüência  $\{L(\Psi_{(k)})\}$ . O algoritmo, quando empregado o procedimento de Newton-Raphson, também é conhecido com algoritmo EM-gradiente (McLachlan & Krishnan, 1997).

## 5.4 Métodos numéricos

### 5.4.1 Método de Newton-Raphson

Everitt (1987) faz uma descrição sobre o método. Uma função quadrática definida como

$$f(\boldsymbol{\theta}) = a + \boldsymbol{\theta}'\mathbf{b} + \frac{1}{2}\boldsymbol{\theta}'\mathbf{B}\boldsymbol{\theta},$$

em que  $a$  define um escalar,  $\mathbf{b} = \mathbf{b}_{d \times 1}$  define um vetor coluna de constantes e  $\mathbf{B} = \mathbf{B}_{d \times d}$  é uma matriz simétrica positiva definida. Tal função tem mínimo no ponto  $\boldsymbol{\theta}^*$  dado por

$$\boldsymbol{\theta}^* = -\mathbf{B}^{-1}\mathbf{b}.$$

Usando a expansão de Taylor, sob certas condições de continuidade, a função  $f(\boldsymbol{\theta})$  pode ser aproximada, na região de um ponto  $\boldsymbol{\theta}_0$ , pela função  $h(\boldsymbol{\theta})$  dada por

$$h(\boldsymbol{\theta}) = f(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{G}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

em que  $\mathbf{g}$  é o vetor gradiente dado por

$$\mathbf{g}' = \left[ \frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, \dots, \frac{\partial f}{\partial \theta_m} \right]$$

e  $\mathbf{G}$  é a matriz ( $m \times m$ ) de segundas derivadas de  $f$  dada por

$$\mathbf{G} = \begin{bmatrix} \frac{\partial^2 f}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_1} & \cdots & \frac{\partial^2 f}{\partial \theta_m \partial \theta_1} \\ \frac{\partial^2 f}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial^2 f}{\partial \theta_m \partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \theta_1 \partial \theta_m} & \frac{\partial^2 f}{\partial \theta_2 \partial \theta_m} & \cdots & \frac{\partial^2 f}{\partial \theta_m \partial \theta_m} \end{bmatrix},$$

ou seja, a matriz Hessiana.

Uma aproximação razoável para o mínimo de  $f(\boldsymbol{\theta})$  seria o mínimo de  $h(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta}^*$ , que pode ser mostrada como sendo

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 - \mathbf{G}^{-1}(\boldsymbol{\theta}_0)\mathbf{g}(\boldsymbol{\theta}_0).$$

O método de Newton-Raphson usa  $\boldsymbol{\theta}^*$  como a próxima aproximação para o mínimo, gerando o esquema de iteração

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \lambda_i \mathbf{G}^{-1}(\boldsymbol{\theta}_i)\mathbf{g}(\boldsymbol{\theta}_i),$$

em que  $\lambda_i$  representa o tamanho do passo e é determinado através de uma busca linear de  $\boldsymbol{\theta}_i$  em direção a  $-\mathbf{G}^{-1}\mathbf{g}$ .

A convergência do método é bastante rápida quando  $\boldsymbol{\theta}$  está próximo do mínimo, sendo que, em geral,  $h(\boldsymbol{\theta})$  dará uma boa aproximação para  $f(\boldsymbol{\theta})$  nesta região. As desvantagens do método são que é necessária a avaliação e inversão da matriz Hessiana a cada iteração, o que demanda grande esforço computacional, e, se  $\boldsymbol{\theta}_i$  não está próximo do mínimo,  $\mathbf{G}$  pode tornar-se negativa definida, o que pode impedir a convergência do método. Além disso, exige-se a resolução de um sistema de  $d$  equações lineares, o que, em geral, apresenta um custo da ordem de  $O(d^3)$  equações aritméticas. Ainda existe, dependendo do valor inicial escolhido, o risco de o método convergir para pontos de sela ou mínimos locais (Everitt, 1987; McLachlan & Krishnan, 1997).

No caso da estimação de máxima verossimilhança, a função  $f(\boldsymbol{\theta})$  considerada é a função de log-verossimilhança  $l(\boldsymbol{\theta}|\mathbf{x})$ . Caso esta função seja côncava e unimodal, a seqüência  $\{\boldsymbol{\theta}^{(k)}\}$  converge para o estimador de máxima verossimilhança, se a função é quadrática em

$\theta$ . Quando a função de log-verossimilhança não é côncava, não há garantia de convergência do método de Newton-Raphson partindo de um valor arbitrário. Sob suposições razoáveis de  $l(\theta|\mathbf{x})$  e um bom valor inicial, a seqüência  $\{\theta^{(k)}\}$  produzida pelo método apresenta convergência quadrática local para uma solução  $\theta^*$  da equação

$$\frac{\partial l(\theta|\mathbf{x})}{\partial \theta} = \mathbf{0}.$$

Ou seja, dada uma norma  $\|\bullet\|$  em  $\Omega$ , existe uma constante  $h$  tal que se  $\theta^{(0)}$  está suficientemente próximo de  $\theta^*$ , então (McLachlan & Krishnan, 1997)

$$\|\theta^{(k+1)} - \theta^*\| \leq h \|\theta^{(k)} - \theta^*\|, \quad k = 0, 1, 2, \dots$$

#### 5.4.2 Método de Davidon-Fletcher-Powell

Segundo Everitt (1987), este método é um dos mais poderosos da classe de procedimentos *quasi-Newton* de minimização. Essencialmente, o método começa como “steepest descent” mudando para o método de Newton durante o curso das interações, pela contínua atualização de uma aproximação da inversa da matriz de segundas derivadas no mínimo, de tal forma a assegurar que as matrizes  $\{\mathbf{H}_i\}$  sejam positivas definidas. As iterações são dadas por

$$\theta_{i+1} = \theta_i - \lambda_i \mathbf{H}_i \mathbf{g}_i,$$

em que  $\mathbf{g}_i$  é o vetor gradiente avaliado em  $\theta_i$  e  $\mathbf{H}_i$  é a  $i$ -ésima aproximação da inversa da matriz de segundas derivadas. A escolha inicial de  $\mathbf{H}$  é arbitrária, desde que positiva definida, sendo a matriz identidade usualmente escolhida com a conseqüência de que o primeiro passo é dado em direção de “steepest descent”.

A aproximação da inversa da matriz de segundas derivadas é atualizada usando a relação

$$\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{A}_i + \mathbf{B}_i,$$

em que

$$\mathbf{A}_i = \frac{\mathbf{z}\mathbf{z}'}{\mathbf{z}'\mathbf{u}}$$
$$\mathbf{B}_i = \frac{\mathbf{H}_i\mathbf{u}\mathbf{u}'\mathbf{H}_i}{\mathbf{u}'\mathbf{H}_i\mathbf{u}}$$

e

$$\mathbf{z} = -\lambda_i\mathbf{H}_i\mathbf{g}_i$$

$$\mathbf{u} = \mathbf{g}_{i+1} - \mathbf{g}_i.$$

Essencialmente, a matriz  $\mathbf{A}_i$  garante que a seqüência de de aproximações matriciais convirja para  $\mathbf{G}^{-1}$  e  $\mathbf{B}_i$  garante que cada estimação  $\mathbf{H}_i$  seja positiva definida. Fletcher & Powell (1963) mostra que, para uma função quadrática,  $\mathbf{H}$  será igual a  $\mathbf{G}^{-1}$  após  $m$  iterações.

# Capítulo 6

## Uso do SiZer para Estimação

## Semi-paramétrica em Misturas de Densidades

### 6.1 Estimação de densidades através de função núcleo

A estimação não-paramétrica de densidades com uso de função núcleo é dada por

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

em que a função núcleo (*kernel*)  $K$  satisfaz

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

Usualmente,  $K$  é uma função densidade de probabilidade simétrica (como a normal). O parâmetro  $h$  representa o *tamanho da janela*, também chamado de *parâmetro de alisamento* (Silverman, 1993).

Silverman (1993) afirma que o estimador por função núcleo é uma soma de “picos”, centrados nas observações. A função núcleo  $K$  determina a forma do pico, e o tamanho da janela  $h$ , seu comprimento. O exemplo mostrado na Figura 6.1.1 apresenta a estimativa

$\hat{f}_h(x)$  e os núcleos individuais  $n^{-1}h^{-1}K[(x - X_i)/h]$ . Na mesma figura são mostrados os efeitos da variação do tamanho da janela  $h$ .

Sendo  $K$  uma função não negativa em toda parte, além de integrar 1, ou seja, uma

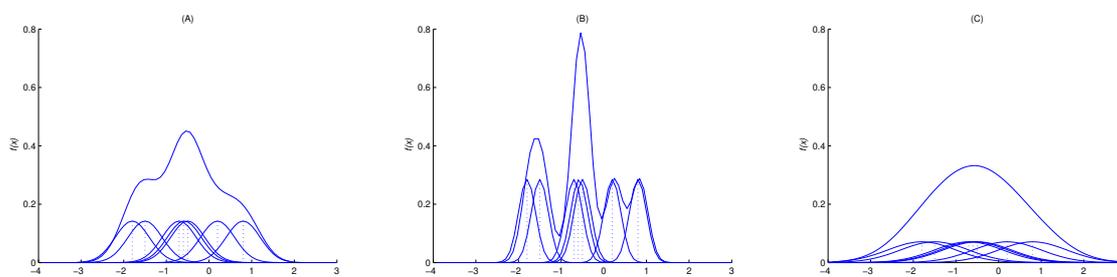


Figura 6.1.1: Estimativa e núcleos individuais com  $h = 0,4$  (A),  $h = 0,2$  (B) e  $h = 0,8$  (C) (Silverman, 1993).

densidade de probabilidade, a estimativa herdará as características de continuidade e diferenciabilidade da função núcleo, assim,  $\hat{f}_n(x)$  também será uma função densidade de probabilidade. Um problema encontrado nesse tipo de estimativa de densidade é quando aplicado à distribuições com longas caudas. Como o tamanho da janela é fixo para toda a amostra, existe a possibilidade de ruídos espúrios aparecerem nas caudas da estimativa. Reduzindo-se o tamanho da janela para corrigir tal problema, corre-se o risco de sobre-alisar características importantes (Silverman, 1993).

### 6.1.1 Propriedades Assintóticas

O suporte usual sobre o qual os teoremas assintóticos relacionados à estimação de densidade por função núcleo são provados assumem que o núcleo  $K$  e a densidade desconhecida  $f$  são fixas e satisfazem condições de regularidade. As estimativas da densidade são construídas a partir das primeiras  $n$  observações de uma seqüência independente e identicamente distribuída  $X_1, X_2, \dots$  de  $f$ . Assume-se que o tamanho da janela  $h$  dependa, de alguma maneira, do tamanho amostral  $n$ . Resultados assintóticos são então obtidos sobre o comportamento da estimativa quando  $n$  tende à infinito. As condições para consistência são fracas, implicando que a taxa de convergência da densidade estimada para o real seja

extremamente lenta (Silverman, 1993).

Parzen (1962) mostra que, quando  $K$  é uma função Borel limitada satisfazendo

$$\int |K(t)| dt < \infty \quad \text{e} \quad \int K(t) dt = 1$$

e

$$|tK(t)| \rightarrow 0 \quad \text{quanto} \quad |t| \rightarrow \infty,$$

e  $h = h(n) = h_n$  satisfaz

$$h_n \rightarrow 0 \quad \text{e} \quad nh_n \rightarrow \infty, \quad \text{quando} \quad n \rightarrow \infty, \quad (6.1.1)$$

então, se  $f$  é contínua em  $x$ ,

$$\hat{f}(x) \rightarrow f(x) \quad \text{em probabilidade quando} \quad n \rightarrow \infty.$$

As condições dadas por (6.1.1) são típicas de consistência. Elas implicam que, enquanto o tamanho da janela fica menor quando  $n$  cresce, o mesmo não deve convergir para zero mais rápido que  $n^{-1}$ . Isso significa dizer que o número esperado de pontos amostrados que caem no intervalo  $x \pm h_n$  tende a infinito, embora lentamente, quando  $n \rightarrow \infty$  (Silverman, 1993).

Silverman (1993) afirma que, para generalizar os resultados de consistência feitos para um único ponto, é necessário determinar em que sentido a curva estimada  $\hat{f}$  aproxima a função  $f$  verdadeira. Parzen (1962) e Nadaraya (1965) consideram a consistência uniforme, ou seja, que  $\sup_x |\hat{f}(x) - f(x)| \xrightarrow{n \rightarrow \infty} 0$ .

**Teorema 2** (Silverman, 1978). *Suponha que  $K(x)$  satisfaça as seguintes condições:*

(a)  $K$  é uniformemente contínua em  $(-\infty, \infty)$ ;

(b)  $K$  é de variação limitada em  $(-\infty, \infty)$ ;

(c)  $\int |K(x)| dx < \infty$  e  $\int K(x) dx = 1$ ;

(d)  $\int |x \log |x||^{\frac{1}{2}} |dK(x)| < \infty$ .

Suponha que  $f$  é uniformemente contínua e que  $h \rightarrow 0$  e  $(nh)^{-1} \log n \xrightarrow{n \rightarrow \infty} 0$ . Então,

$$\sup_x |\hat{f}(x) - f(x)| \xrightarrow{n \rightarrow \infty} 0 \text{ q.c.}$$

## 6.2 SiZer

A suavização de curvas é muito útil em estatística para a detecção de certas características. Nessa área, uma aplicação muito interessante foi introduzida por Chaudhuri & Marron (1999), denominada **SiZer** (**SI**gnificant **ZER**o crossings of derivatives). A idéia é destacar características significativas, como modas, mostrando locais (com respeito tanto a locação quanto escala) onde a curva cresce ou decresce significativamente. É fácil ver que as modas são localizadas onde a derivada da função cruza o zero, ou seja, entre regiões que a função mostre-se significativamente crescente e decrescente. O estudo é feito sobre estimativas da densidade através de função núcleo (Seção 6.1).

O resultado do procedimento é apresentado na forma de um mapa colorido. O esquema de cores é azul (vermelho) em locais onde a curva é significativamente crescente (decrescente), e a cor púrpura é usada onde não é possível concluir se a curva é crescente ou decrescente. O termo “local” é usado no sentido de escala-espaco de ambos “locação de  $x$ ” e “locação do tamanho da janela”. Cinza é usado para indicar regiões onde os dados são muito esparsos para tirar conclusões sobre a significância, pois não existem pontos suficientes em cada janela. A Figura 6.2.1 mostra um exemplo do mapa resultante da análise. Os dados, pseudo-aleatórios, foram obtidos através da mistura de densidades dada por:

$$f(x) = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3,$$

em que  $X_1 \sim N(-1, 1)$ ,  $X_2 \sim N(5, 1)$  e  $X_3 \sim N(10, 1)$ .

O exemplo mostra que em resoluções mais baixas (tamanhos maiores da janela) os dados são sobre-alisados, eliminando completamente características interessantes dos mesmos. Para valores de  $h$  menores, como 0,5623 (note que  $\log_{10}(0,5623) \approx -0,2500$ ), a estimativa se apresenta significativamente crescente à esquerda e significativamente decres-

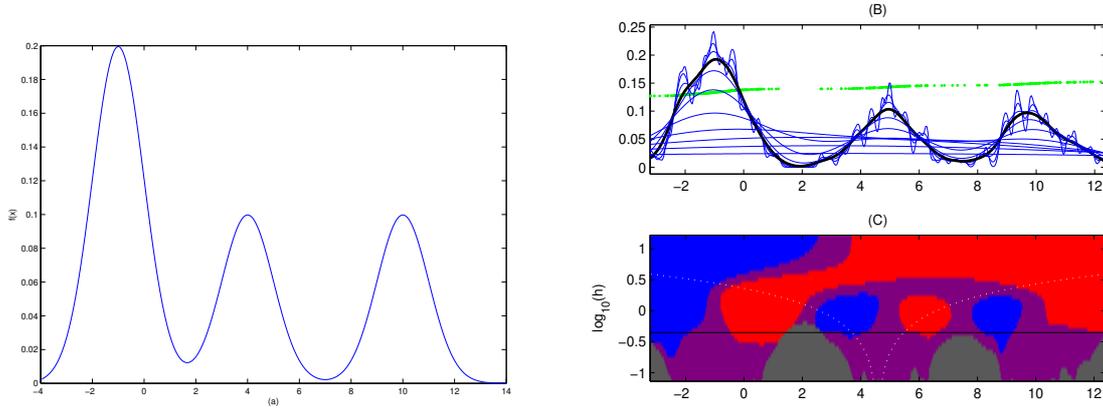


Figura 6.2.1: O painel (A) mostra a densidade real da mistura. Combinação das estimativas de densidade por função núcleo (B) e o mapa gerado pelo SiZer (C), usando nível de significância de  $\alpha = 0,05$ , também são apresentados. A janela com  $h = 0,4449$  é destacada. As linhas pontilhadas no mapa SiZer mostram tamanhos de janela efetivos para cada  $h$ , com intervalos representando  $\pm 2h$  (ou seja,  $\pm 2h$  desvios padrões do núcleo gaussiano).

cente à direita. Pode-se ainda sugerir a existência de três modas, nas regiões de  $x = -1$ ,  $x = 5$  e  $x = 10$ , locais em que ocorrem mudanças de azul ( $\uparrow$ ) para vermelho ( $\downarrow$ ).

Esses mapas gerados podem ser muito úteis durante a análise exploratória. A visualização do problema é feita de um ponto de vista *escala-espaco*. Isto significa observar uma “superfície escala-espaco”, ou seja, a família de todas as estimativas por função núcleo indexadas pelo tamanho da janela  $h$  usada. A idéia essencial é que modelos com  $h$  grande oferecem uma visão macroscópica (distante), em que apenas detalhes em grande escala podem ser observados. Enquanto isso, modelos com  $h$  pequeno propiciam uma visão microscópica (*zoom*), permitindo identificar detalhes finos.

Abordagens clássicas para a significância de características apresentadas nas estimativas, baseadas em bandas de confiança, são muito conservativas para uso em inferência ou inválidas devido ao vício. Os autores assumem uma nova abordagem, o ponto de vista escala-espaco, para tratar o problema do vício. Em particular, ao invés de procurar intervalos de confiança para  $f'(x)$ , são procurados intervalos de confiança para a versão escala-espaco,  $f'_h(x) \triangleq \mathbb{E} [\hat{f}'_h(x)]$ . Desse ponto de vista, a significância de qualquer carac-

terística depende na escala de resolução, ou seja, do valor de  $h$ . A Figura 6.2.1(c) mostra que a estrutura trimodal está presente em alguns níveis de resolução mas desaparece em níveis mais grosseiros, apresentando apenas uma moda em janelas maiores. Uma das vantagens do SiZer é que ele, além de contar o número de modas significantes, também mostra a posição de tais modas. No entanto, segundo os autores, o mapa tende a ser mais conservativo que testes específicos para o número de modas.

Para uma função  $f$  dada (i.e., um sinal), existe um ruído associado (ao menos algum ruído está presente em qualquer sinal real) e que pode ser representado pela convolução  $f * K_h$ , para diferentes valores de  $h$ , em que  $K_h(\bullet) = \frac{1}{h}K\left(\frac{\bullet}{h}\right)$ . O foco da análise é esta família de convoluções, com a idéia de que isso é tudo que está disponível de um conjunto finito de dados na presença de ruído. Tal abordagem difere muito da clássica, em que o foco é a função  $f$ .

Exemplos de características da estimação de curvas incluem pico e vales, que são mostrados no SiZer como pontos em que a derivada cruza o zero. Pode-se dizer que estes pontos são significativos quando a estimação da derivada é significativamente diferente de zero de ambos os lados, com sinais opostos, como mostrado por áreas azul ( $\uparrow$ ) e vermelha ( $\downarrow$ ).

Segundo estudos de estimativas de derivadas em um intervalos de valores para  $h$ , os autores afirmam que o núcleo gaussiano  $K(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$  apresenta importantes vantagens sobre outros núcleos. Em particular, para o alisamento de convoluções, o número de cruzamentos pelo zero das derivadas alisadas é sempre uma função decrescente em  $h$ , o que não é verdade para qualquer outra função núcleo usada nesse tipo de estimação. Em outras palavras, apenas o núcleo gaussiano apresenta monotonicidade de características com respeito a intensidade do alisamento.

### 6.2.1 Desenvolvimento do SiZer

A abordagem de Chaudhuri & Marron (1999) para a determinação da significância de características, como picos e vales, numa família de densidades estimadas

$$\left\{ \hat{f}'_h(x) : h \in [h_{min}, h_{max}] \right\}$$

é baseada em intervalos de confiança para a derivada no espaço escala,  $f'_h(x)$ . O comportamento é mostrado através do mapa SiZer, sendo que regiões azuis indicam locais onde  $\hat{f}'_h(x)$  é significativamente positiva, vermelho onde  $\hat{f}'_h(x)$  é significativamente negativa e púrpura onde  $\hat{f}'_h(x)$  não é significativamente diferente de zero.

Devido à necessidade do repetido cálculo de estimativas de densidades, há necessidade de métodos computacionais rápidos. A proposta usada pelos autores foi apresentada por Fan & Marron (1994), em qual os dados são divididos em intervalos (*bins*), o que é feito apenas uma vez. A idéia é que o cálculo de  $\hat{f}'_h(x)$  pode ser rapidamente obtido por convolução discreta quando os dados são aproximados por contagens em intervalos (*bins*) numa rede igualmente espaçada, resultando em economia de tempo em fatores de 100 (para grandes amostras).

Os intervalos de confiança para  $f'_h(x)$  são da forma

$$\hat{f}'_h(x) \pm q \times \widehat{DP}(\hat{f}'_h(x)), \quad (6.2.1)$$

em que  $q$  é um quantil apropriado e  $DP$  é o desvio padrão, discutido na Seção 6.2.2. Uma locação  $(x, h)$ , no espaço escala, é dita significativamente crescente, decrescente ou não significante quando o zero está abaixo, acima ou dentro do intervalo, respectivamente.

Uma sugestão dos autores para o cálculo de  $q$  é por aproximação simultânea sobre quantis gaussianos, baseada no “número de blocos independentes”. Quando  $x$  e  $x'$  estão suficientemente afastados, tal que as janelas dos núcleos centradas em  $x$  e  $x'$  são essencialmente disjuntas, as estimativas  $\hat{f}'_h(x)$  e  $\hat{f}'_h(x')$  são essencialmente independentes. No entanto, se  $x$  e  $x'$  estão próximos, as estimativas são altamente correlacionadas. O problema do limite de confiança simultâneo é então aproximado por  $m$  intervalos de confiança

independentes, em que  $m$  reflete o número de blocos independentes. O valor de  $m$  pode ser estimado através de um “tamanho amostral efetivo estimado” (ESS), definido para cada  $(x, h)$  como

$$\text{ESS}(x, h) = \frac{\sum_{i=1}^n K_h(x - X_i)}{K_h(0)}.$$

Então,  $m$  é escolhido como sendo o número de blocos independentes de tamanho médio obtido através do tamanho amostral  $n$ ,

$$m(h) = \frac{n}{\text{média}_x \text{ESS}(x, h)}.$$

Agora, assumindo independência dos  $m(h)$  blocos de dados, o quantil simultâneo aproximado é dado por

$$q = q(h) = \Phi^{-1} \left( \frac{1 + (1 - \alpha)^{1/m}}{2} \right).$$

A quantidade ESS também é útil para destacar regiões onde a aproximação normal implícita em (6.2.1) pode ser inadequada (semelhante à  $np$  na aproximação gaussiana da binomial). Assim, regiões onde  $\text{ESS}(x, h) < n_0$  (sendo  $n_0 = 5$ ) são mostradas em cinza, eliminando características espúrias e também para indicar regiões em que o alisamento é baseado essencialmente em dados esparsos.

Outra sugestão para o cálculo de  $q$  é através de bootstrap, sendo, também, simultâneo sobre  $x$ . Para cada reamostra bootstrap, sobre a distribuição empírica, é computada  $\hat{f}'_h(x)^*$  e a versão padronizada

$$Z^*(x, h) = \frac{\hat{f}'_h(x)^* - \hat{f}'_h(x)}{\text{DP}(\hat{f}'_h(x))}.$$

Para cada  $h$ , o quantil bootstrap  $q = q(h)$ , simultâneo sobre  $x$  (onde os dados são razoavelmente densos) é o quantil empírico  $\max_{x \in D_h} |Z^*(x, h)|$  calculado sobre as replicações bootstrap, em que  $D_h = \{x : \text{ESS}(x, h) \geq n_0\}$ .

## 6.2.2 Implementação numérica

A seleção de  $[h_{min}, h_{max}]$  é feita para maximizar o intervalo. Assim,  $h_{min}$  é escolhido como  $h_{min} = 2h$ , em que  $h$  é a menor janela para qual não há distorção substancial na

construção dos intervalos de dados (*bins*), e  $h_{max}$  é tomado como a amplitude dos dados.

A idéia por trás do cálculo do desvio padrão  $\widehat{DP}$  neste contexto é que o estimado da derivada  $\hat{f}'_h(x)$  é uma média das derivadas das funções núcleo. Assim, utiliza-se o desvio padrão amostral correspondente

$$\widehat{Var}(\hat{f}'_h(x)) = \widehat{Var}\left(\frac{1}{n}\sum_{i=1}^n K'_h(x - X_i)\right) = n^{-1}S^2(K'_h(x - X_1), \dots, K'_h(x - X_n))$$

em que  $S^2$  é a variância amostra dos  $n$  valores.

Fan & Marron (1994) afirma que a classificação dos dados em intervalos (*bins*) reduz drasticamente o número de cálculos da função núcleo, pois vários dos valores calculados para estas funções são muito próximos. Isso exige que a rede  $x_1, \dots, x_g$  seja igualmente espaçada. Os dados são então aproximados pelos “dados igualmente espaçados”. Uma maneira simples de fazer isso é classificar os dados em intervalos, substituindo cada  $X_i$  pelo ponto mais próximo na rede  $x_{j(i)}$ , denotado por  $X_i \mapsto x_{j(i)}$ , e fazer a estimação usando este conjunto de dados modificado. O conjunto modificado é convenientemente usado através dos conjuntos de índices

$$I_j \equiv \{i : X_i \mapsto x_j\}, \quad j = 1, \dots, g.$$

Os dados assim classificados podem ser resumidos como

$$\{(x_j, c_j), j = 1, \dots, g\}$$

em que  $c_j = \#\{X_i : i \in I_j\} = \#(I_j)$  representa a contagem de pontos observados classificados como  $x_j$  (Fan & Marron, 1994).

Quando usado o processo de classificação introduzido por Fan & Marron (1994) na construção do SiZer, a diferença é que a função núcleo deve ser substituída por sua derivada. Assim,

$$\hat{f}'_h(x_j) \approx n^{-1}\overline{S}'_0(x_j),$$

em que

$$\overline{S}'_0(x_j) = \sum_{j'=1}^g \kappa'_{j-j'} c_{j'}$$

e

$$\kappa'_{j-j'} = K'_h(x_j - x_{j'}).$$

O desvio padrão pode ser aproximado de maneira similar por

$$\widehat{DP}(x_j) = \frac{1}{\sqrt{2}} \sqrt{\frac{1}{n} \sum_{j'=1}^g (\kappa'_{j-j'})^2 c_{j'} - (\hat{f}'_h(x_j))^2}.$$

### 6.3 Convergência

Karatzas & Schreve (2000) define o *movimento Browniano*, padrão uni-dimensional, como sendo um processo contínuo adaptado  $B = \{B_t, \mathbb{F}_t; 0 \leq t < \infty\}$ , definido num espaço de probabilidade  $(\Omega, \mathbb{F}, P)$ , em que  $B_0 = 0$  quase certamente (q.c.) e para  $0 \leq s < t$ , o incremento  $B_t - B_s$  é independente de  $\mathbb{F}_s$  tendo distribuição normal com média zero e variância  $t - s$ . Pode-se também falar em um movimento Browniano  $B = \{B_t, \mathbb{F}_t; 0 \leq t \leq T\}$  em  $[0, T]$ , para algum  $T > 0$ .

Um processo estocástico  $X = \{X_t, \mathbb{F}_t; 0 \leq t < \infty\}$  em  $\mathbb{R}$  é chamado um *processo Gaussiano* se, para todo  $k \geq 1$  e números reais  $0 \leq t_1 < t_2 < \dots < t_k < \infty$ , o vetor  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  tem distribuição normal multivariada. Caso a distribuição de  $(X_{t+t_1}, X_{t+t_2}, \dots, X_{t+t_k})$  não dependa de  $t$ , diz-se que o processo é *estacionário* (Ross, 1996; Karatzas & Shreve, 2000; Grimmett & Stirzaker, 2004).

Se  $\{B_t, \mathbb{F}_t; 0 \leq t < \infty\}$  é um movimento Browniano, então  $\{Z_t, \mathbb{F}_t; 0 \leq t \leq 1\}$  é um processo de *Ponte Browniana* quando  $Z_t = B_t - tB_1$ . Sendo assim, é imediato ver que  $\{Z_t, \mathbb{F}_t; t \geq 0\}$  é um processo Gaussiano, implicando que  $\mathbb{E}(Z_t) = 0$  e  $Cov(Z_t, Z_s) = s(1-t)$  para  $s \leq t$  (Ross, 1996).

Considerando a estimação da densidade através da função núcleo, como introduzida por Rosenblatt (1956),

$$f_n(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) d\mathbb{F}_n(y) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right),$$

Silverman (1978) propõe, sob a existência de uma ponte Browniana  $W$  num espaço de probabilidade adequado, a decomposição da estimativa por

$$f_n(x) = f(x) + b(x) + n^{-1/2}\rho(x) + \epsilon(x)$$

em que

$$b(x) = \mathbb{E}[f_n(x)] - f(x) = \int \frac{1}{h} K\left(\frac{x-y}{h}\right) d\mathbb{F}(y) - f(x)$$

$$\rho(x) = - \int \frac{1}{h} W[\mathbb{F}(y)] d_y K\left(\frac{x-y}{h}\right)$$

e

$$\epsilon(x) = -\frac{1}{nh} \log n \int Z_n(y) d_y K\left(\frac{x-y}{h}\right),$$

em que, para alguma constante  $C_0$ ,

$$\limsup_n \sup_y |Z_n(y)| \leq C_0 \quad q.c.$$

# Capítulo 7

## Simulação e Aplicação

Segundo a teoria estudada nos capítulos anteriores, foram realizados estudos de simulação para avaliação dos procedimentos de estimação. Uma aplicação a dados reais também é apresentada.

### 7.1 Simulação de modelos de misturas

Foram gerados valores pseudo-aleatórios de modelos de mistura finitos. Os modelos foram constituídos por uma componente uniforme em  $(0, 1)$ , representando os p-valores originários da hipótese nula, e uma componente  $G3B(\alpha; \beta; \lambda)$ , como em (2.3.3). A função de densidade é dada por

$$f(y) = p + (1 - p) \frac{\lambda^\alpha y^{\alpha-1} (1 - y)^{\beta-1}}{B(\alpha, \beta) (1 - (1 - \lambda)y)^{\alpha+\beta}}$$

Os parâmetros foram selecionados como mostrados na Tabela 7.1.1, em que  $p$  representa a proporção da componente uniforme. A Figura 7.1.1 apresenta as densidades utilizadas na simulação, quando  $p = 0,95$ . As demais figuras (A.0.1-A.0.8) podem ser encontradas no Apêndice A. Os valores foram gerados através do método proposto por Pham-Gia & Duong (1989), ou seja, se  $T \sim Beta(\alpha, \beta)$  e  $\lambda > 0$ , então  $X = T/[\lambda + (1 - \lambda)T]$  tem distribuição  $G3B(\alpha, \beta\lambda)$ .

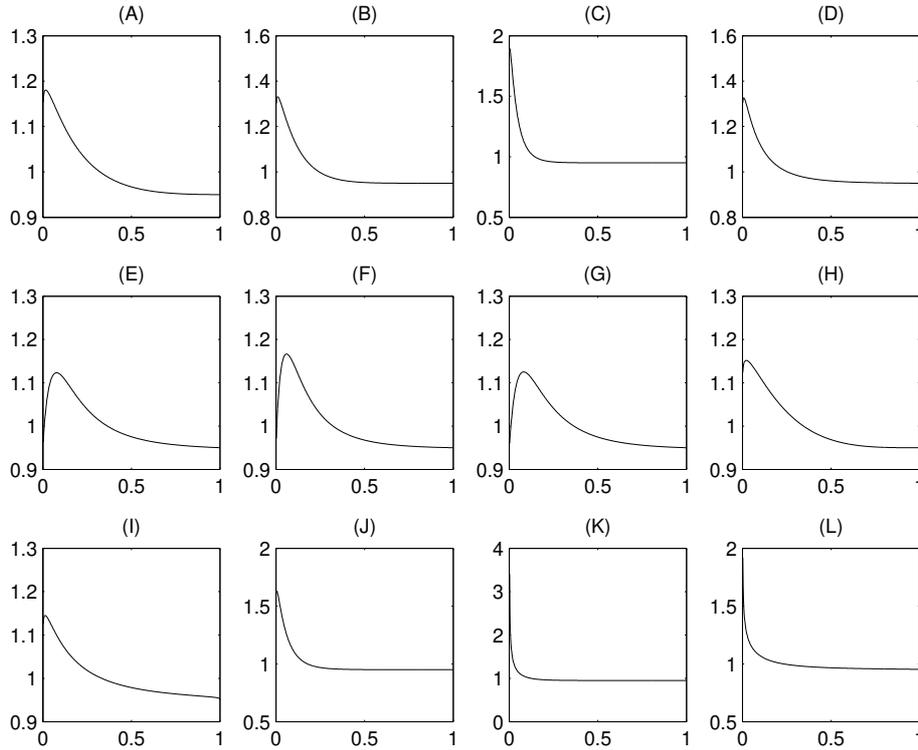


Figura 7.1.1: Misturas usadas na simulação, com  $p = 0,95$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

Segundo o Teorema 1, pode-se mostrar, através de extensivos cálculos, que os estimadores de máxima verossimilhança (EMV) para  $p$ ,  $\alpha$ ,  $\beta$  e  $\lambda$  apresentam boas propriedades. Como não existe forma fechada para os EMV, foram avaliados quatro métodos de estimação segundo a verossimilhança: busca direta (EMVb), busca de Fibonacci (EMVf), através do SiZer e pelo algoritmo EM.

Através dos SiZer, selecionando o tamanho apropriado da janela  $h$ , é possível detectar a presença de uma possível moda na distribuição. Utilizando esse valor, pela modelagem através de uma distribuição beta padrão, valores iniciais dos parâmetros  $\alpha$  e  $\beta$  podem ser encontrados. Pela minimização da distância  $L_2$  entre a curva estimada através de função núcleo e a curva estimada através da mistura uniforme-beta generalizada, é possível atualizar os valores dos parâmetros e encontrar uma estimativa para  $\lambda$ .

Para cada uma das 108 combinações dos parâmetros apresentadas na Tabela 7.1.1 foram gerados 1.000 conjuntos de valores pseudo-aleatórios, com 3.000 observações cada. Os resultados para estimação são mostrados nas Tabelas A.0.1-A.0.18.

No algoritmo EM, passo M, pela ausência de equações fechadas para os EMV, houve necessidade do uso de um método numérico. Foram usados o Newton-Raphson e Davison-Fletcher-Powell. Como as derivadas da função de log-verossimilhança não são bem comportadas, o algoritmo mostra sérios problemas na inversa da matriz Hessiana. O método de Davison-Fletcher-Powell utiliza uma aproximação para essa matriz, mas que converge muito rápido para aquela, também resultando em problemas numéricos, como valores não-numéricos (NaN) e infinitos (Inf). Esses problemas são apresentados na Tabela A.0.19. Os resultados apresentados referem-se ao método de Newton.

Para avaliar os resultados obtidos foi utilizada a distância de Hellinger, como apresentada em Birgé (1986):

$$h(f; g) = \frac{1}{2} \int \left( \sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx,$$

em que  $f$  representa a densidade real e  $g$  a densidade estimada. Tais distâncias são apresentadas nas Tabelas A.0.20 e A.0.21. As menores distâncias são destacadas em negrito.

Para cada densidade estimada, foi calculado o valor da FDR. O cálculo foi feito pela razão das áreas (A) e (A+C) mostradas na Figura 4.2.1, ao nível de 0,05. Os resultados são mostrados nas Tabelas A.0.22 e A.0.23.

O algoritmo EM para estimação dos parâmetros apresentou sérios problemas numéricos em suas derivadas, principalmente nas extremidades da densidade. Associado a isso, foram encontrados problemas de precisão numérica. Por exemplo, quando da realização de somatórios de logaritmos, a simples aplicação da função a um vetor de valores entre 0 e 1 e sua soma, pode levar a resultados não-numéricos ou infinitos. Uma forma de evitar tal problema, quando possível, é a ordenação dos valores e a soma a partir dos extremos, ou seja, somando valores “grandes” com valores “pequenos”. Isso pode compensar a falta de precisão e “balanceamento” da equação. Mesmo com esse tipo de cuidado, não foi possível obter uma convergência consistente do método, levando a resultados desastrosos, como

Tabela 7.1.1: Combinações de parâmetros usadas na simulação.

$p$	$\alpha$	$\beta$	$\lambda$	$p$	$\alpha$	$\beta$	$\lambda$	$p$	$\alpha$	$\beta$	$\lambda$
0,25	1,10	3,00	2,00	0,50	1,10	3,00	2,00	0,70	1,10	3,00	2,00
	1,10	5,00	2,00		1,10	5,00	2,00		1,10	5,00	2,00
	1,10	5,00	5,00		1,10	5,00	5,00		1,10	5,00	5,00
	1,10	2,00	5,00		1,10	2,00	5,00		1,10	2,00	5,00
	2,00	1,90	5,00		2,00	1,90	5,00		2,00	1,90	5,00
	2,00	2,00	6,10		2,00	2,00	6,10		2,00	2,00	6,10
	2,10	2,00	5,00		2,10	2,00	5,00		2,10	2,00	5,00
	1,10	4,00	1,30		1,10	4,00	1,30		1,10	4,00	1,30
	1,10	1,20	4,30		1,10	1,20	4,30		1,10	1,20	4,30
	1,10	4,20	4,30		1,10	4,20	4,30		1,10	4,20	4,30
	0,40	5,00	2,50		0,40	5,00	2,50		0,40	5,00	2,50
	0,60	1,20	4,30		0,60	1,20	4,30		0,60	1,20	4,30
0,75	1,10	3,00	2,00	0,80	1,10	3,00	2,00	0,85	1,10	3,00	2,00
	1,10	5,00	2,00		1,10	5,00	2,00		1,10	5,00	2,00
	1,10	5,00	5,00		1,10	5,00	5,00		1,10	5,00	5,00
	1,10	2,00	5,00		1,10	2,00	5,00		1,10	2,00	5,00
	2,00	1,90	5,00		2,00	1,90	5,00		2,00	1,90	5,00
	2,00	2,00	6,10		2,00	2,00	6,10		2,00	2,00	6,10
	2,10	2,00	5,00		2,10	2,00	5,00		2,10	2,00	5,00
	1,10	4,00	1,30		1,10	4,00	1,30		1,10	4,00	1,30
	1,10	1,20	4,30		1,10	1,20	4,30		1,10	1,20	4,30
	1,10	4,20	4,30		1,10	4,20	4,30		1,10	4,20	4,30
	0,40	5,00	2,50		0,40	5,00	2,50		0,40	5,00	2,50
	0,60	1,20	4,30		0,60	1,20	4,30		0,60	1,20	4,30
0,90	1,10	3,00	2,00	0,95	1,10	3,00	2,00	0,97	1,10	3,00	2,00
	1,10	5,00	2,00		1,10	5,00	2,00		1,10	5,00	2,00
	1,10	5,00	5,00		1,10	5,00	5,00		1,10	5,00	5,00
	1,10	2,00	5,00		1,10	2,00	5,00		1,10	2,00	5,00
	2,00	1,90	5,00		2,00	1,90	5,00		2,00	1,90	5,00
	2,00	2,00	6,10		2,00	2,00	6,10		2,00	2,00	6,10
	2,10	2,00	5,00		2,10	2,00	5,00		2,10	2,00	5,00
	1,10	4,00	1,30		1,10	4,00	1,30		1,10	4,00	1,30
	1,10	1,20	4,30		1,10	1,20	4,30		1,10	1,20	4,30
	1,10	4,20	4,30		1,10	4,20	4,30		1,10	4,20	4,30
	0,40	5,00	2,50		0,40	5,00	2,50		0,40	5,00	2,50
	0,60	1,20	4,30		0,60	1,20	4,30		0,60	1,20	4,30

observados nas Tabelas A.0.1-A.0.18.

A estimação dos parâmetros através do SiZer também não resultou em bons resultados. O método é capaz de determinar a região mais provável para presença de uma moda, mas sua precisão é muito baixa.

Dessa forma, as melhores opções são os métodos de busca, sendo que a busca por Fibonacci parece apresentar melhores resultados.

## 7.2 Aplicação a dados reais

Os procedimentos estudados foram aplicados a dados reais gerados pelo **Laboratório Genoma Funcional, do Centro de Biologia Molecular e Engenharia Genética da Unicamp**, e cedidos pelo Prof. Dr. Marcelo Menossi. São dados relativos a experimentos com cana-de-açúcar cultivada sob situações de estresse biológico. Foram usados quatro microarranjos (representando os tempos 1, 2, 3 e 4 de cultivo), com um total de

1.597 genes em cada. Os genes foram testados segundo metodologia julgada adequada pelos pesquisadores e os p-valores gerados nos foram cedidos (Drummond et al., 2005). A Figura 7.2.1 mostra os histogramas dos p-valores.

A Tabela 7.2.1 mostra as estimativas dos parâmetros segundo cada um dos métodos estudados. Os resultados encontrados sob o algoritmo EM foram incluídos, apesar do mesmo já ter sido mostrado, na seção anterior, ruim. O método é desconsiderado nas análises posteriores. As Figuras 7.2.2, 7.2.3 e 7.2.4 mostram as densidades estimadas.

A Tabela 7.2.2 apresenta os valores da FDR calculados a partir das densidades estimadas. Como observado na simulação, a estimação através do SiZer parece superestimar o valor da FDR.

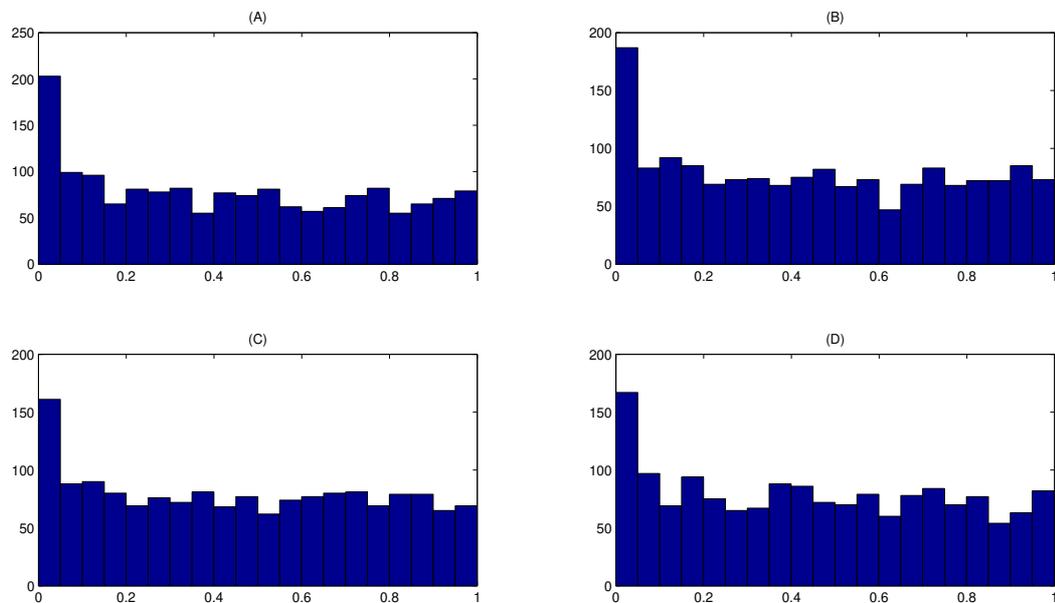


Figura 7.2.1: Histogramas dos p-valores analisados: tempo 1 (A), tempo 2 (B), tempo 3 (C) e tempo 4 (D).

Tabela 7.2.1: Estimativas dos parâmetros para os dados reais.

Tempo	Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
1	EMVb	0,8000	0,2692	0,5923	2,1000
	EMVf	0,8111	0,2335	0,6145	1,8641
	SiZer	0,8976	1,0160	13,7898	1,9100
	EM	1,0000	12374,6209	0,0000	$1,2704 \cdot 10^{45}$
2	EMVb	0,7500	0,2692	1,0231	0,2692
	EMVf	0,7887	0,2537	0,8557	0,4911
	SiZer	0,8871	1,0260	7,9126	2,9100
	EM	1,0000	1,0453	$2,1506 \cdot 10^{11}$	0,0000
3	EMVb	0,9500	0,1615	6,4077	0,5923
	EMVf	0,9454	0,1897	6,6439	0,9880
	SiZer	0,9215	1,0260	5,1346	3,3100
	EM	1,0000	0,5803	$6,2813 \cdot 10^{11}$	0,0000
4	EMVb	0,9000	0,2692	3,6077	0,5923
	EMVf	0,9036	0,2956	3,7490	0,9127
	SiZer	0,9369	1,0160	7,3906	4,2100
	EM	1,0000	4,4703	$1,2888 \cdot 10^{33}$	0,0000

Tabela 7.2.2: FDR para dados fosfato. A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci) e SiZer (estimação através do SiZer).

Método	tempo 1	tempo 2	tempo 3	tempo 4
EMVb	0,0730	0,0939	0,3200	0,2065
EMVf	0,0495	0,0926	0,2634	0,2049
SiZer	0,2946	0,3006	0,4709	0,4025

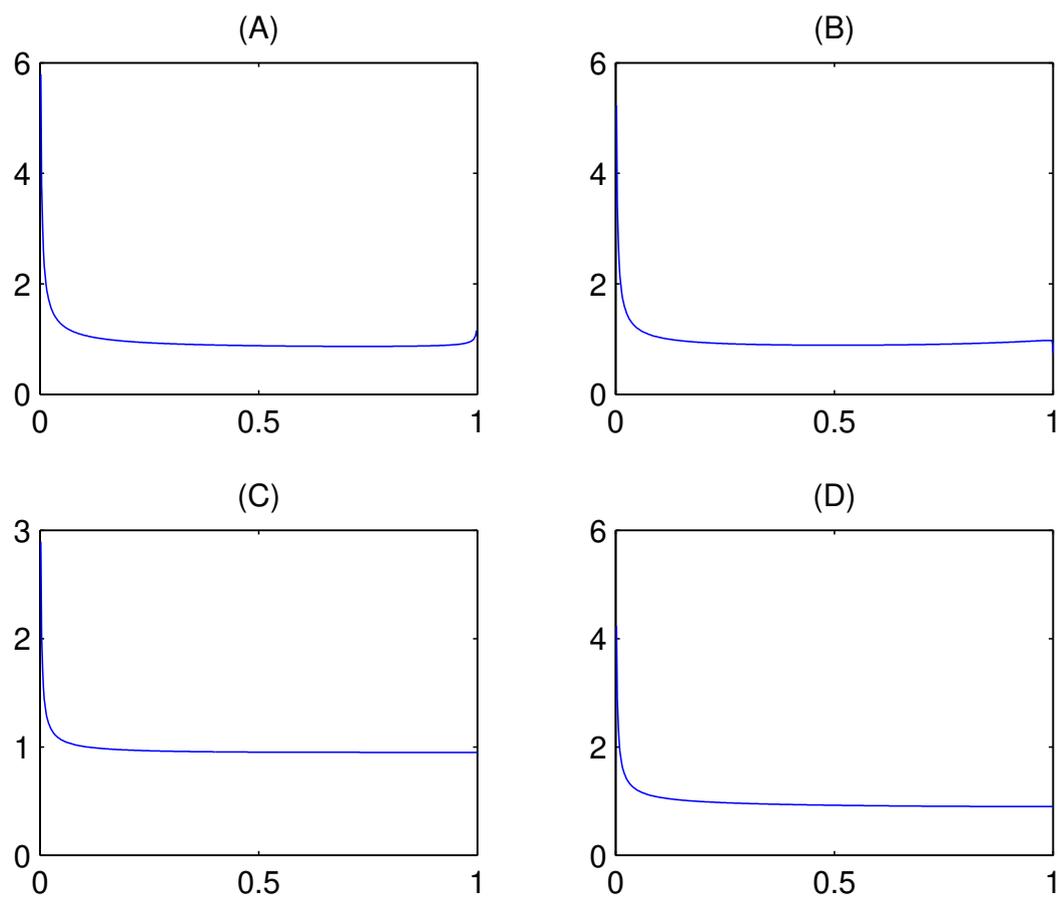


Figura 7.2.2: Densidades estimadas segundo o método de busca: tempo 1 (A), tempo 2 (B), tempo 3 (C) e tempo 4 (D).

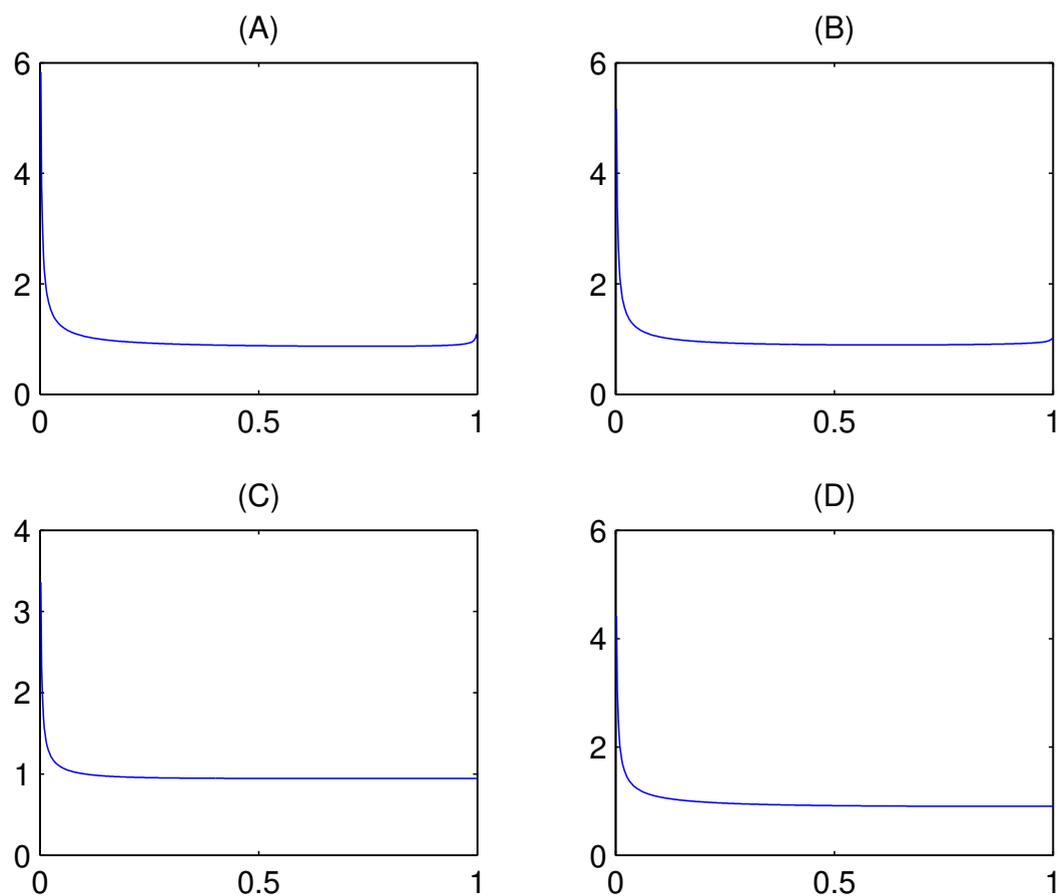


Figura 7.2.3: Densidades estimadas segundo o método de Fibonacci: tempo 1 (A), tempo 2 (B), tempo 3 (C) e tempo 4 (D).

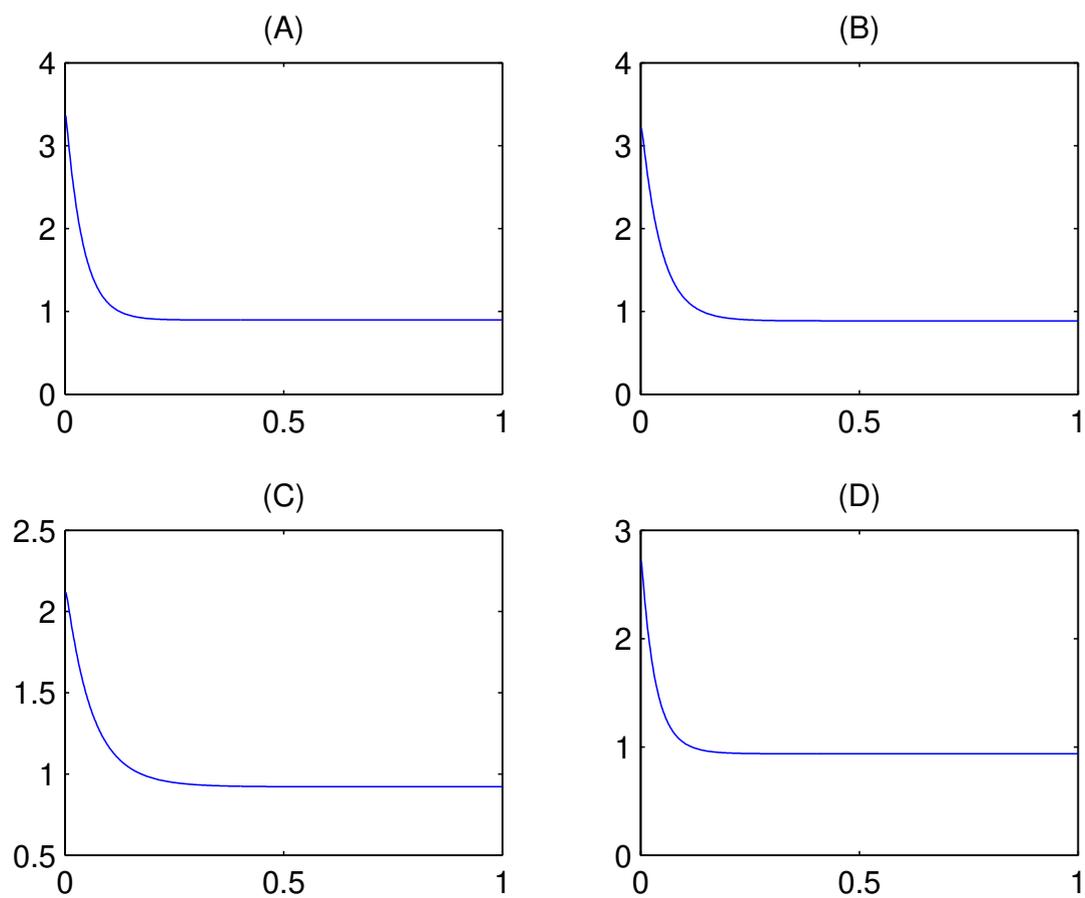


Figura 7.2.4: Densidades estimadas segundo o SiZer: tempo 1 (A), tempo 2 (B), tempo 3 (C) e tempo 4 (D).

# Capítulo 8

## Considerações Finais

O uso de misturas de densidades na análise de dados de expressão gênica tem se mostrado muito útil. Apesar de serem de grande utilidade na estimação de taxas de erro, também podem ser aplicadas na estimação do próprio sinal, como em Steinhilff et al. (2003).

O modelo beta-uniforme se mostra muito adequado nesse tipo de estimação, uma vez que a distribuição dos p-valores pode ser modelada por uma mistura de distribuições beta, de qual a uniforme é um caso particular. Devido a maior flexibilidade da distribuição beta generalizada, sugeriu-se que o modelo uniforme-beta generalizada apresentasse maior adaptabilidade aos dados.

Os estimadores de máxima verossimilhança, ainda que apresentando boas características assintóticas, são de difícil estimação. A complexidade da função de densidade da distribuição beta generalizada, além da inclusão de mais um parâmetro no processo de estimação, resultou em dificuldades maiores nos métodos de estimação numérica. As derivadas da função densidade da mistura não se mostram bem comportadas, principalmente nas caudas. Esse efeito fica claro quando analisado o algoritmo EM, que apresenta estimativas muito ruins. Os problemas de precisão numérica também se mostraram importantes durante o processo de estimação.

O SiZer, apesar de muito interessante na análise exploratória de dados, detectando características da curva de densidade, como modas e vales, não oferece um meio automático

para validação das mesmas ao longo da escala. Isso permite a visualização do comportamento geral da curva, mas não a determinação precisa do parâmetro de suavização.

Pelos resultados da simulação apresentados no Capítulo 7, as estimativas dos parâmetros parecem aproximar melhor os verdadeiros valores quando a proporção  $p$  da mistura cresce. Avaliando as distâncias de Hellinger calculadas, observa-se que a estimação de máxima verossimilhança pelo método de Fibonacci gerou menores distâncias.

Avaliando os valores estimados da FDR, fica claro que a estimação através do SiZer superestima a taxa de falsos positivos. Quando a proporção da mistura é pequena, os métodos de estimação por máxima verossimilhança são afetados, gerando péssimas estimativas. Isso melhora em valores intermediários de  $p$ , mas subestimando a FDR. Em valores mais próximos de 1, essa característica é alterada, passando a superestimar a FDR.

Os resultados obtidos não se mostraram satisfatórios, sendo sugerido, para estudos posteriores, uma melhor avaliação e caracterização do comportamento das funções e suas derivadas.

# Apêndice A

## Simulação

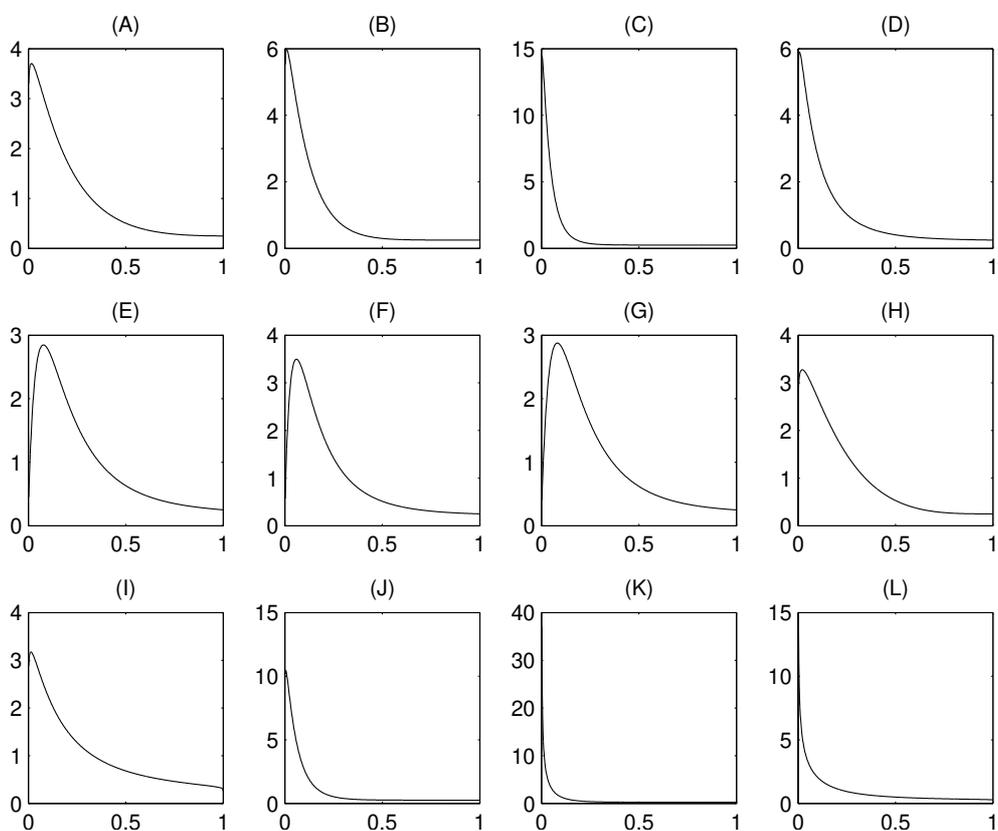


Figura A.0.1: Misturas usadas na simulação, com  $p = 0,25$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

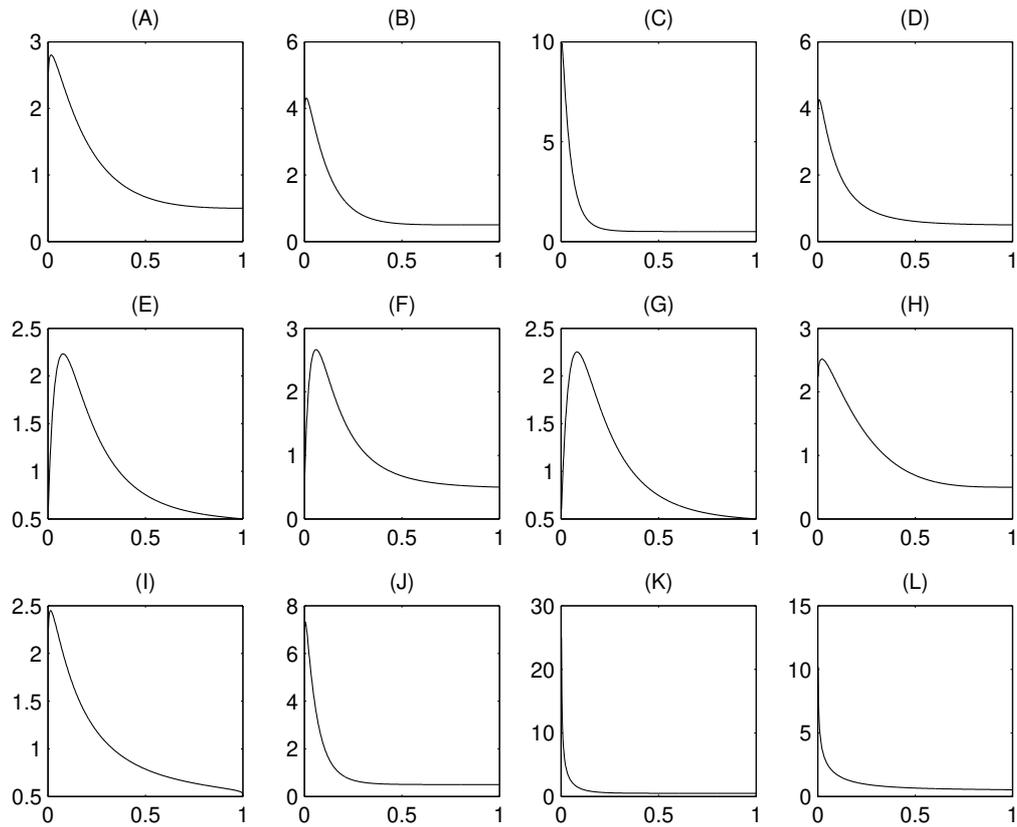


Figura A.0.2: Misturas usadas na simulação, com  $p = 0,50$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

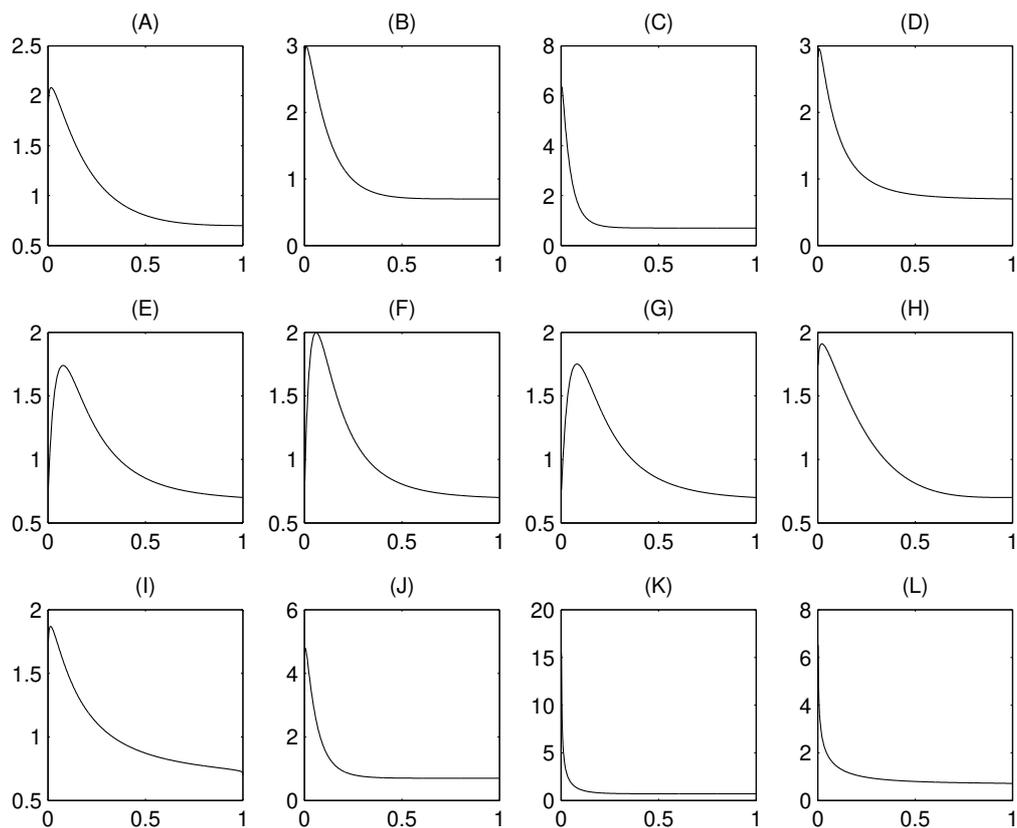


Figura A.0.3: Misturas usadas na simulação, com  $p = 0,70$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

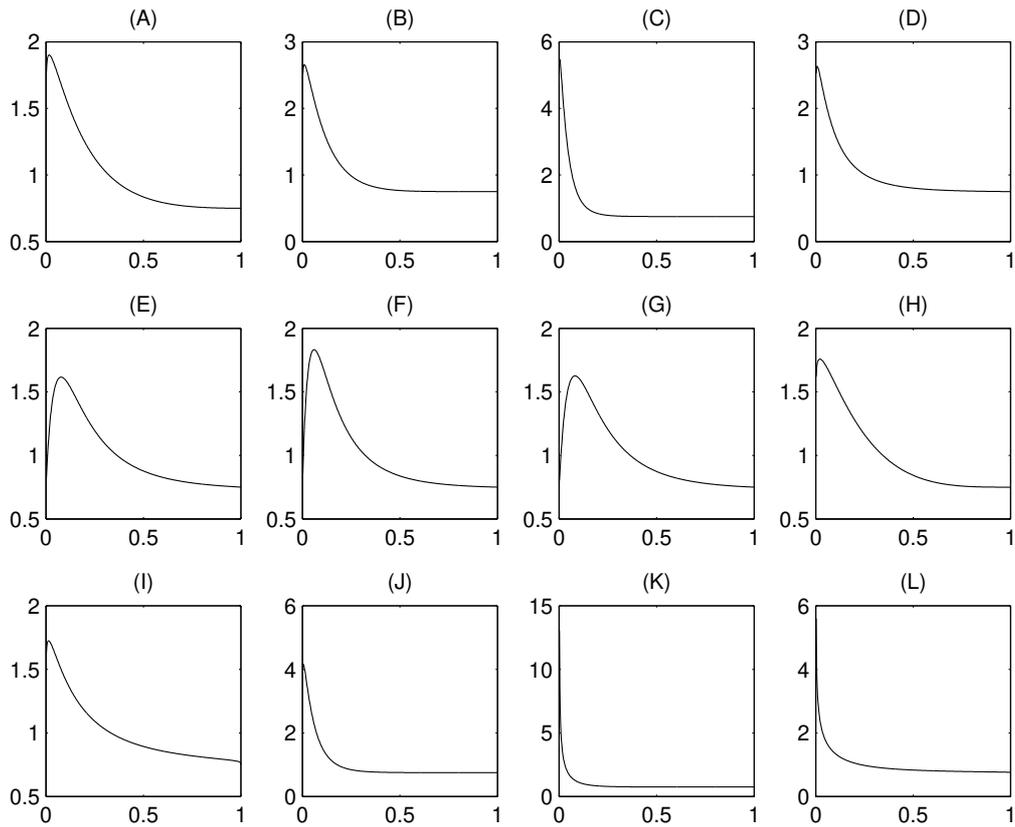


Figura A.0.4: Misturas usadas na simulação, com  $p = 0,75$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

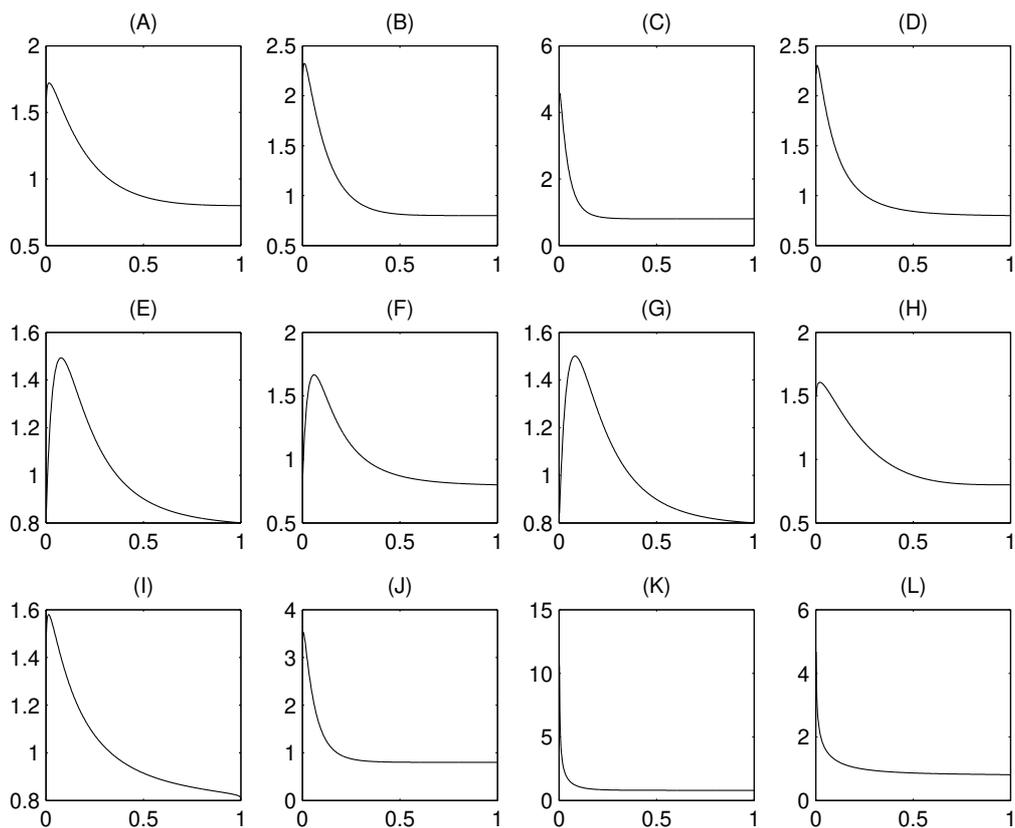


Figura A.0.5: Misturas usadas na simulação, com  $p = 0,80$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

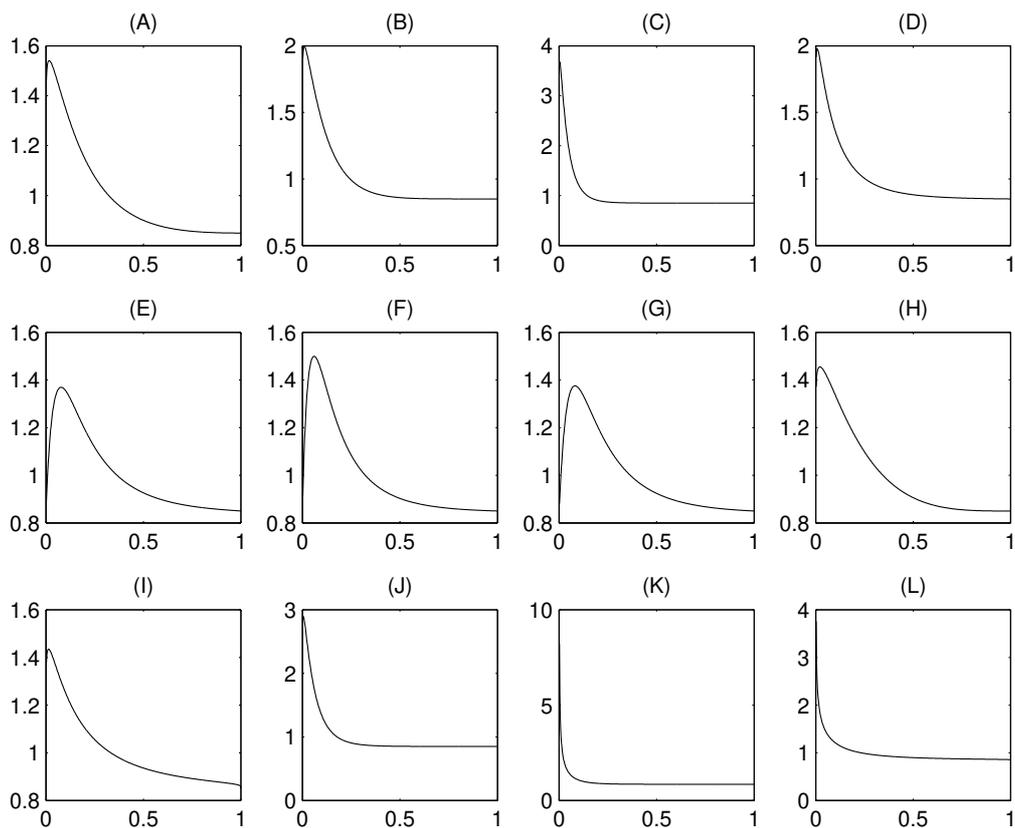


Figura A.0.6: Misturas usadas na simulação, com  $p = 0,85$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

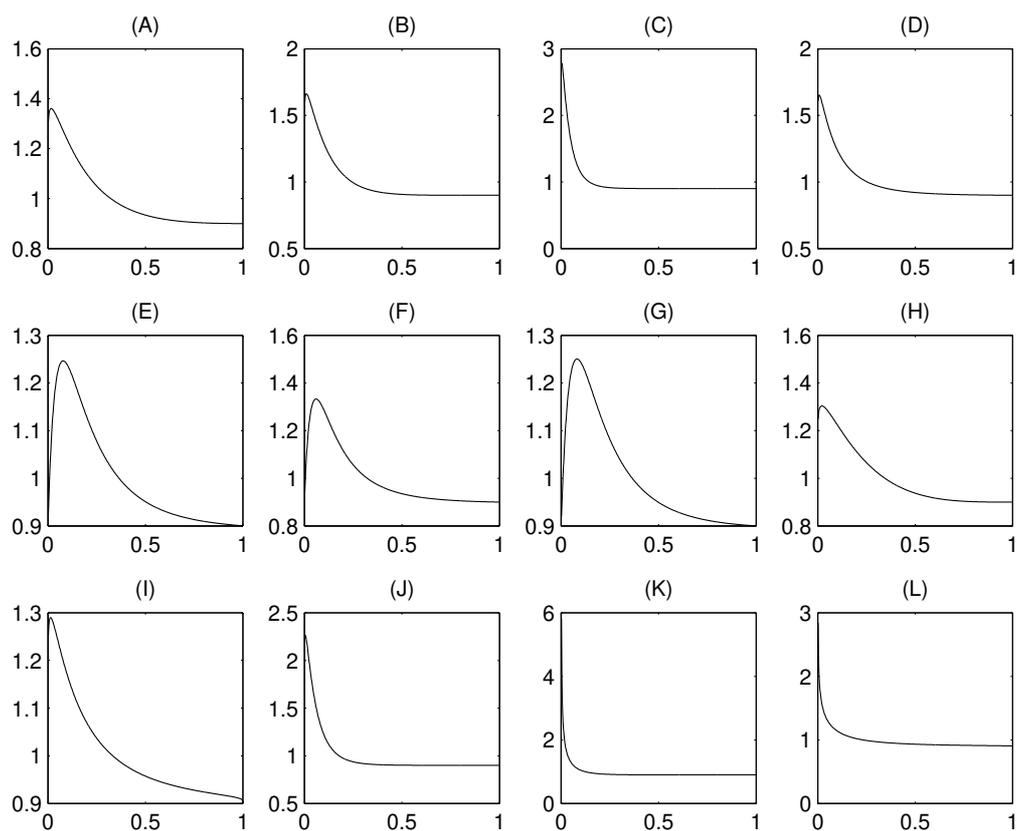


Figura A.0.7: Misturas usadas na simulação, com  $p = 0, 90$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

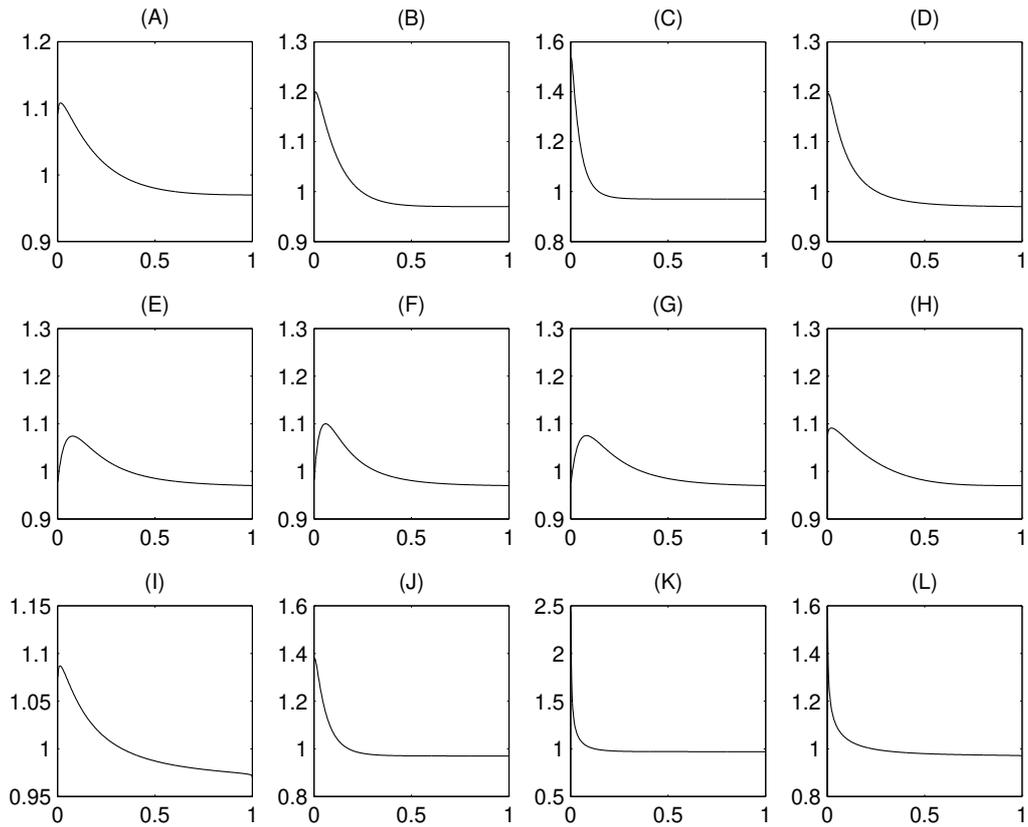


Figura A.0.8: Misturas usadas na simulação, com  $p = 0,97$  e parâmetros  $(\alpha; \beta; \lambda)$  sendo:  $(1,1;3,0;2,0)$ (A);  $(1,1;5,0;2,0)$ (B);  $(1,1;5,0;5,0)$ (C);  $(1,1;2,0;5,0)$ (D);  $(2,0;1,9;5,0)$ (E);  $(2,0;2,0;6,1)$ (F);  $(2,1;2,0;5,0)$ (G);  $(1,1;4,0;1,3)$ (H);  $(1,1;1,2;4,3)$ (I);  $(1,1;4,2;4,3)$ (J);  $(0,4;5,0;2,5)$ (K);  $(0,6;1,2;4,3)$ (L).

Tabela A.0.1:  $p=0,25$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,25	1,10	3,00	2,00
EMVb	0,0969 (0,1326)	0,4708 (0,1057)	1,2806 (1,1094)	2,1263 (1,2113)
EMVf	0,0262 (0,0426)	0,4275 (0,0484)	1,0450 (0,6897)	1,6700 (0,9092)
SiZer	0,6625 (0,0136)	1,3361 (0,0769)	6,6897 (0,4334)	1,6816 (0,1925)
EM	0,0000 (0,0000)	1,2216 (0,0532)	0,9894 (0,0349)	6,4773 (0,6258)
	0,25	1,10	5,00	2,00
EMVb	0,0083 (0,0216)	0,3300 (0,0000)	0,3300 (0,0000)	2,5582 (0,2916)
EMVf	0,0010 (0,0037)	0,2462 (0,0347)	0,4685 (0,0937)	1,6154 (0,2769)
SiZer	0,4411 (0,0115)	1,3347 (0,0952)	9,7466 (0,6316)	1,4843 (0,1993)
EM	0,0022 (0,0467)	1,2791 (0,2094)	4,8306 (85,922)	14,652 (2,6664)
	0,25	1,10	5,00	5,00
EMVb	0,0007 (0,0059)	0,3300 (0,0000)	0,3300 (0,0000)	0,5392 (0,0475)
EMVf	0,0001 (0,0006)	0,1211 (0,0127)	0,2775 (0,0426)	0,2527 (0,1731)
SiZer	0,2760 (0,0080)	1,2765 (0,0990)	20,978 (1,1205)	1,2910 (0,1507)
EM	0,0000 (0,0000)	0,0012 (0,0000)	0,3025 (0,0070)	0,0000 (0,0000)
	0,25	1,10	2,00	5,00
EMVb	0,0650 (0,0820)	0,1815 (0,1643)	0,2083 (0,2005)	1,9252 (2,0306)
EMVf	0,0013 (0,0071)	0,2699 (0,0229)	0,5813 (0,1569)	1,7767 (1,3082)
SiZer	0,5239 (0,0126)	1,2654 (0,0806)	9,0686 (0,4904)	1,6020 (0,1403)
EM	0,0000 (0,0000)	1,1693 (0,0566)	0,8254 (0,0264)	11,319 (1,2028)
	0,25	2,00	1,90	5,00
EMVb	0,0858 (0,0953)	0,8008 (0,0764)	1,7420 (0,8515)	1,6003 (0,6497)
EMVf	0,0096 (0,0239)	0,8246 (0,0622)	1,4620 (0,5525)	1,5642 (0,4371)
SiZer	0,7970 (0,0146)	1,6704 (0,0882)	6,9898 (0,4807)	1,5904 (0,2164)
EM	0,0000 (0,0000)	2,0389 (0,1341)	1,0212 (0,0355)	8,6162 (1,0159)
	0,25	2,00	2,00	6,10
EMVb	0,0082 (0,0332)	0,5500 (0,0000)	1,0362 (0,2068)	1,5607 (0,3458)
EMVf	0,0001 (0,0022)	0,5507 (0,0219)	1,1106 (0,1115)	1,1759 (0,1657)
SiZer	0,6831 (0,0139)	1,5977 (0,0827)	7,8701 (0,4070)	1,5272 (0,0616)
EM	0,0000 (0,0000)	2,1765 (0,1591)	0,9481 (0,0302)	12,874 (1,5910)
	0,25	2,10	2,00	5,00
EMVb	0,0795 (0,0922)	0,7880 (0,0604)	1,7094 (0,8201)	1,5646 (0,5792)
EMVf	0,0079 (0,0215)	0,8133 (0,0615)	1,4454 (0,5561)	1,5455 (0,3898)
SiZer	0,7954 (0,0145)	1,7192 (0,0813)	7,2119 (0,4346)	1,5488 (0,1597)
EM	0,0000 (0,0000)	2,1889 (0,1525)	1,0269 (0,0340)	9,3437 (1,0951)
	0,25	1,10	4,00	1,30
EMVb	0,0007 (0,0059)	0,5500 (0,0000)	0,7392 (0,0777)	2,0104 (0,7507)
EMVf	0,0000 (0,0000)	0,4721 (0,0227)	0,9354 (0,1461)	1,3557 (0,6669)
SiZer	0,7046 (0,0137)	1,3735 (0,0796)	6,3033 (0,3762)	1,6628 (0,1575)
EM	0,0000 (0,0000)	1,2576 (0,0553)	1,0445 (0,0388)	5,7485 (0,5597)
	0,25	1,10	1,20	4,30
EMVb	0,1596 (0,1506)	0,9654 (0,1563)	1,3204 (0,5837)	3,1469 (1,2041)
EMVf	0,1899 (0,1234)	0,9550 (0,1621)	1,3153 (0,5161)	3,2320 (1,2134)
SiZer	0,7707 (0,0147)	1,2541 (0,0653)	5,5287 (0,3917)	1,9890 (0,2139)
EM	1,0000 (0,0000)	1,2960 (0,1202)	1,0615 (0,1028)	6,1890 (1,8457)
	0,25	1,10	4,20	4,30
EMVb	0,0000 (0,0000)	0,3300 (0,0000)	0,3300 (0,0000)	0,7696 (0,0098)
EMVf	0,0007 (0,0023)	0,1312 (0,0106)	0,3542 (0,0297)	0,7609 (0,0914)
SiZer	0,3179 (0,0095)	1,3169 (0,1156)	15,660 (0,9575)	1,4020 (0,2112)
EM	0,0033 (0,0572)	1,3125 (0,2107)	0,8429 (2,9309)	33,436 (5,9270)
	0,25	0,40	5,00	2,50
EMVb	0,0000 (0,0000)	0,3300 (0,0000)	0,3300 (0,0000)	0,3300 (0,0000)
EMVf	0,0036 (0,0051)	0,1222 (0,0075)	0,1552 (0,0237)	0,1445 (0,0703)
SiZer	0,2785 (0,0082)	1,0396 (0,0087)	31,341 (3,0867)	1,0888 (0,0868)
EM	0,0001 (0,0000)	0,0012 (0,0000)	0,2390 (0,0013)	0,0000 (0,0000)
	0,25	0,60	1,20	4,30
EMVb	0,0770 (0,0879)	0,3300 (0,0000)	0,3300 (0,0000)	0,9574 (0,1162)
EMVf	0,0061 (0,0214)	0,1946 (0,0155)	0,3505 (0,0415)	0,9110 (0,2310)
SiZer	0,5651 (0,0129)	1,0478 (0,0054)	13,232 (0,4321)	1,4388 (0,0908)
EM	0,0000 (0,0000)	0,5796 (0,0175)	0,8383 (0,0338)	3,6675 (0,3969)

Tabela A.0.2:  $p=0,25$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,25			1,10			3,00			2,00		
EMVb	(0,0000)	0,0000	(0,2500)	(0,3300)	0,5500	(0,5500)	(0,5500)	0,5500	(1,8700)	(0,7700)	2,7500	(2,9700)
EMVf	(0,0000)	0,0000	(0,0500)	(0,3836)	0,4265	(0,4579)	(0,6411)	0,6693	(1,1302)	(0,6802)	2,2252	(2,3042)
SiZer	(0,6528)	0,6622	(0,6717)	(1,2860)	1,3360	(1,3960)	(6,4836)	6,7251	(6,9218)	(1,6100)	1,7100	(1,7100)
EM	(0,0000)	0,0000	(0,0000)	(1,1865)	1,2181	(1,2552)	(0,9659)	0,9878	(1,0108)	(6,0673)	6,4563	(6,8634)
	0,25			1,10			5,00			2,00		
EMVb	(0,0000)	0,0000	(0,0000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(2,3100)	2,5300	(2,7500)
EMVf	(0,0000)	0,0000	(0,0000)	(0,2025)	0,2721	(0,2760)	(0,3951)	0,3974	(0,5707)	(1,3790)	1,5916	(1,7563)
SiZer	(0,4335)	0,4410	(0,4490)	(1,2560)	1,3260	(1,4160)	(9,4505)	9,7457	(10,055)	(1,4100)	1,4100	(1,6100)
EM	(0,0000)	0,0000	(0,0000)	(1,2614)	1,3098	(1,3503)	(0,8103)	0,8263	(0,8438)	(13,916)	14,889	(15,914)
	0,25			1,10			5,00			5,00		
EMVb	(0,0000)	0,0000	(0,0000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,5500)	0,5500	(0,5500)
EMVf	(0,0000)	0,0000	(0,0000)	(0,1116)	0,1157	(0,1274)	(0,2480)	0,2934	(0,3107)	(0,1490)	0,1601	(0,1987)
SiZer	(0,2705)	0,2755	(0,2815)	(1,2060)	1,2560	(1,3260)	(20,375)	21,124	(21,717)	(1,2100)	1,2100	(1,3100)
EM	(0,0000)	0,0000	(0,0000)	(0,0012)	0,0012	(0,0012)	(0,2982)	0,3029	(0,3070)	(0,0000)	0,0000	(0,0000)
	0,25			1,10			2,00			5,00		
EMVb	(0,0000)	0,0000	(0,1500)	(0,0000)	0,3300	(0,3300)	(0,0000)	0,3300	(0,3300)	(0,0000)	0,9900	(4,0700)
EMVf	(0,0000)	0,0000	(0,0000)	(0,2436)	0,2760	(0,2760)	(0,4203)	0,6226	(0,7375)	(0,6463)	0,8268	(3,0701)
SiZer	(0,5160)	0,5238	(0,5326)	(1,1960)	1,2560	(1,3260)	(8,6441)	9,0623	(9,3772)	(1,5100)	1,6100	(1,7100)
EM	(0,0000)	0,0000	(0,0000)	(1,1294)	1,1676	(1,2033)	(0,8072)	0,8255	(0,8422)	(10,499)	11,226	(12,071)
	0,25			2,00			1,90			5,00		
EMVb	(0,0000)	0,0500	(0,2000)	(0,7700)	0,7700	(0,7700)	(1,2100)	1,4300	(2,0900)	(1,2100)	1,4300	(1,8700)
EMVf	(0,0000)	0,0000	(0,0000)	(0,7784)	0,8151	(0,8751)	(1,1552)	1,2928	(1,4884)	(1,2881)	1,5038	(1,8292)
SiZer	(0,7867)	0,7968	(0,8065)	(1,6260)	1,6760	(1,7260)	(6,7836)	7,0096	(7,2226)	(1,5100)	1,6100	(1,6100)
EM	(0,0000)	0,0000	(0,0000)	(1,9374)	2,0321	(2,1214)	(0,9973)	1,0182	(1,0452)	(7,9043)	8,5724	(9,2522)
	0,25			2,00			2,00			6,10		
EMVb	(0,0000)	0,0000	(0,0000)	(0,5500)	0,5500	(0,5500)	(0,9900)	0,9900	(0,9900)	(1,4300)	1,4300	(1,6500)
EMVf	(0,0000)	0,0000	(0,0000)	(0,5344)	0,5451	(0,5682)	(1,0550)	1,1007	(1,1711)	(1,0679)	1,1613	(1,2264)
SiZer	(0,6738)	0,6831	(0,6926)	(1,5360)	1,6060	(1,6560)	(7,6289)	7,8652	(8,1181)	(1,5100)	1,5100	(1,6100)
EM	(0,0000)	0,0000	(0,0000)	(2,0641)	2,1653	(2,2721)	(0,9286)	0,9475	(0,9688)	(11,764)	12,737	(13,863)
	0,25			2,10			2,00			5,00		
EMVb	(0,0000)	0,0000	(0,1500)	(0,7700)	0,7700	(0,7700)	(1,2100)	1,4300	(2,0900)	(1,2100)	1,6500	(1,8700)
EMVf	(0,0000)	0,0000	(0,0000)	(0,7677)	0,8030	(0,8408)	(1,1631)	1,2719	(1,4628)	(1,3401)	1,5143	(1,7666)
SiZer	(0,7851)	0,7956	(0,8053)	(1,6660)	1,7260	(1,7710)	(7,0101)	7,2190	(7,4289)	(1,5100)	1,5100	(1,6100)
EM	(0,0000)	0,0000	(0,0000)	(2,0750)	2,1718	(2,2852)	(1,0021)	1,0269	(1,0505)	(8,5461)	9,2040	(10,084)
	0,25			1,10			4,00			1,30		
EMVb	(0,0000)	0,0000	(0,0000)	(0,5500)	0,5500	(0,5500)	(0,7700)	0,7700	(0,7700)	(1,6500)	1,6500	(1,8700)
EMVf	(0,0000)	0,0000	(0,0000)	(0,4593)	0,4593	(0,4885)	(0,9146)	0,9756	(1,0193)	(1,0195)	1,0985	(1,2546)
SiZer	(0,6951)	0,7042	(0,7141)	(1,3160)	1,3860	(1,4360)	(6,1138)	6,3016	(6,5292)	(1,6100)	1,6100	(1,7100)
EM	(0,0000)	0,0000	(0,0000)	(1,2168)	1,2581	(1,2914)	(1,0157)	1,0429	(1,0701)	(5,3618)	5,7233	(6,1365)
	0,25			1,10			1,20			4,30		
EMVb	(0,0000)	0,1500	(0,3000)	(0,7700)	0,9900	(0,9900)	(0,9900)	1,2100	(1,4300)	(2,0900)	3,1900	(3,6300)
EMVf	(0,1236)	0,1761	(0,3056)	(0,8171)	0,9523	(1,0743)	(0,9845)	1,1317	(1,4785)	(2,1833)	3,2225	(3,9662)
SiZer	(0,7608)	0,7698	(0,7806)	(1,2060)	1,2460	(1,2960)	(5,3475)	5,5449	(5,7263)	(1,9100)	2,0100	(2,0100)
EM	(1,0000)	1,0000	(1,0000)	(1,2233)	1,2889	(1,3697)	(0,9906)	1,0597	(1,1261)	(5,0977)	5,8958	(7,1633)
	0,25			1,10			4,20			4,30		
EMVb	(0,0000)	0,0000	(0,0000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,7700)	0,7700	(0,7700)
EMVf	(0,0000)	0,0000	(0,0000)	(0,1289)	0,1289	(0,1289)	(0,3221)	0,3607	(0,3732)	(0,6905)	0,7279	(0,8036)
SiZer	(0,3111)	0,3178	(0,3247)	(1,2160)	1,2960	(1,3960)	(15,070)	15,706	(16,412)	(1,2100)	1,3100	(1,5100)
EM	(0,0000)	0,0000	(0,0000)	(1,2928)	1,3397	(1,3873)	(0,6780)	0,6865	(0,6969)	(31,896)	34,101	(36,082)
	0,25			0,40			5,00			2,50		
EMVb	(0,0000)	0,0000	(0,0000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)
EMVf	(0,0000)	0,0003	(0,0059)	(0,1141)	0,1266	(0,1289)	(0,1297)	0,1570	(0,1744)	(0,0751)	0,1308	(0,2037)
SiZer	(0,2722)	0,2785	(0,2839)	(1,0360)	1,0360	(1,0360)	(29,779)	29,805	(29,891)	(1,1100)	1,1100	(1,1100)
EM	(0,0001)	0,0001	(0,0001)	(0,0012)	0,0012	(0,0012)	(0,2380)	0,2390	(0,2398)	(0,0000)	0,0000	(0,0000)
	0,25			0,60			1,20			4,30		
EMVb	(0,0000)	0,0500	(0,1500)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,9900)	0,9900	(0,9900)
EMVf	(0,0000)	0,0000	(0,0000)	(0,1845)	0,1959	(0,2031)	(0,3254)	0,3361	(0,3537)	(0,7500)	0,8819	(1,0735)
SiZer	(0,5567)	0,5644	(0,5738)	(1,0460)	1,0460	(1,0460)	(13,221)	13,229	(13,240)	(1,4100)	1,4100	(1,5100)
EM	(0,0000)	0,0000	(0,0000)	(0,5670)	0,5792	(0,5927)	(0,8159)	0,8356	(0,8610)	(3,4067)	3,6470	(3,9578)

Tabela A.0.3:  $p=0,50$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,50	1,10	3,00	2,00
EMVb	0,4825 (0,0523)	1,1429 (0,1131)	2,9438 (1,2061)	2,6358 (1,2780)
EMVf	0,4858 (0,0591)	1,1540 (0,0830)	2,9580 (1,2952)	2,7198 (1,2573)
SiZer	0,7753 (0,0142)	1,2722 (0,0739)	5,1195 (0,2902)	1,9354 (0,1882)
EM	0,9905 (0,0969)	58,847 (357,03)	58,91 (370,17)	233,25 (953,20)
	0,50	1,10	5,00	2,00
EMVb	0,0238 (0,0558)	0,5502 (0,0070)	0,6125 (0,1425)	2,9203 (0,8820)
EMVf	0,0015 (0,0185)	0,5113 (0,0342)	0,7006 (0,1686)	2,1955 (0,7328)
SiZer	0,6279 (0,0128)	1,2409 (0,0621)	6,8484 (0,4658)	1,8438 (0,2109)
EM	0,0000 (0,0000)	1,0917 (0,0572)	0,6840 (0,0205)	8,5614 (0,9863)
	0,50	1,10	5,00	5,00
EMVb	0,0222 (0,0422)	0,3300 (0,0000)	0,3300 (0,0000)	1,2610 (0,1245)
EMVf	0,0005 (0,0023)	0,2052 (0,0098)	0,4853 (0,0420)	0,4956 (0,1247)
SiZer	0,5174 (0,0104)	1,2102 (0,0709)	13,482 (0,8812)	1,8513 (0,2515)
EM	0,0000 (0,0000)	1,0537 (0,0654)	0,5039 (0,0125)	25,195 (3,4074)
	0,50	1,10	2,00	5,00
EMVb	0,0993 (0,1768)	0,6952 (0,1183)	1,1035 (0,8196)	2,4754 (0,8812)
EMVf	0,0886 (0,1099)	0,6440 (0,0853)	1,0531 (0,5919)	2,1660 (0,9547)
SiZer	0,6815 (0,0124)	1,2030 (0,0557)	6,4778 (0,3345)	2,0042 (0,1614)
EM	0,0000 (0,0000)	0,9752 (0,0456)	0,7254 (0,0259)	5,8481 (0,7083)
	0,50	2,00	1,90	5,00
EMVb	0,5045 (0,0371)	1,9439 (0,1687)	2,1701 (0,5002)	4,3872 (0,9411)
EMVf	0,5049 (0,0319)	1,9521 (0,1715)	2,1591 (0,4846)	4,4442 (1,0128)
SiZer	0,8667 (0,0136)	1,5226 (0,0771)	5,3613 (0,3464)	1,7472 (0,2077)
EM	0,7769 (0,3642)	54,113 (359,23)	146,48 (645,35)	169,98 (788,45)
	0,50	2,00	2,00	6,10
EMVb	0,5081 (0,0293)	1,8784 (0,1435)	2,4438 (0,5703)	4,7326 (0,8375)
EMVf	0,5085 (0,0260)	1,9118 (0,1587)	2,4218 (0,6115)	4,9575 (1,0714)
SiZer	0,7908 (0,0139)	1,4457 (0,0924)	5,8007 (0,3448)	1,7672 (0,2092)
EM	0,9557 (0,2061)	98,131 (492,88)	178,45 (454,07)	162,64 (864,22)
	0,50	2,10	2,00	5,00
EMVb	0,5031 (0,0395)	2,0324 (0,1876)	2,2986 (0,6382)	4,4136 (1,0187)
EMVf	0,5036 (0,0323)	2,0406 (0,1872)	2,2878 (0,6208)	4,4685 (1,0933)
SiZer	0,8653 (0,0149)	1,5521 (0,0609)	5,4964 (0,3032)	1,7246 (0,1497)
EM	0,6859 (0,3278)	42,105 (401,74)	50,624 (363,92)	121,29 (731,62)
	0,50	1,10	4,00	1,30
EMVb	0,4790 (0,0696)	1,1376 (0,1039)	3,5640 (1,2145)	1,8447 (0,9980)
EMVf	0,4846 (0,0611)	1,1482 (0,0902)	3,5966 (1,2938)	1,9010 (1,0378)
SiZer	0,8051 (0,0146)	1,3189 (0,0751)	4,9311 (0,3720)	1,8980 (0,2655)
EM	0,9926 (0,0859)	83,897 (516,35)	69,105 (537,36)	451,33 (1495,5)
	0,50	1,10	1,20	4,30
EMVb	0,5202 (0,0761)	1,0864 (0,1093)	1,6667 (0,8184)	3,6177 (1,3108)
EMVf	0,5525 (0,0656)	1,1265 (0,0906)	1,7652 (0,7984)	3,8072 (1,2508)
SiZer	0,8492 (0,0141)	1,2320 (0,0570)	4,4403 (0,2556)	2,1952 (0,2017)
EM	0,6413 (0,4784)	55,078 (470,15)	37,326 (235,38)	118,96 (748,65)
	0,50	1,10	4,20	4,30
EMVb	0,1877 (0,0496)	0,3300 (0,0000)	0,3300 (0,0000)	3,2274 (0,5735)
EMVf	0,0140 (0,0262)	0,2846 (0,0147)	0,4211 (0,0191)	2,2359 (0,5714)
SiZer	0,5463 (0,0116)	1,2153 (0,0696)	10,2769 (0,7679)	1,8754 (0,2703)
EM	0,0000 (0,0000)	1,0367 (0,0556)	0,5676 (0,0155)	15,6439 (1,9706)
	0,50	0,40	5,00	2,50
EMVb	0,0000 (0,0000)	0,3300 (0,0000)	0,3300 (0,0000)	0,3300 (0,0000)
EMVf	0,0002 (0,0007)	0,2177 (0,0147)	0,3491 (0,0325)	0,2339 (0,0738)
SiZer	0,5204 (0,0117)	1,0270 (0,0051)	21,959 (1,0317)	1,5212 (0,0876)
EM	0,0000 (0,0000)	0,3198 (0,0093)	0,8061 (0,0539)	1,2124 (0,2322)
	0,50	0,60	1,20	4,30
EMVb	0,2988 (0,1071)	0,3344 (0,0309)	0,5993 (0,2484)	1,8005 (0,9578)
EMVf	0,1095 (0,0926)	0,3528 (0,0305)	0,6226 (0,1704)	1,4395 (0,7286)
SiZer	0,7082 (0,0146)	1,0439 (0,0131)	10,2164 (0,5894)	1,8476 (0,2152)
EM	0,0000 (0,0000)	0,5540 (0,0195)	0,8873 (0,0409)	1,5316 (0,1996)

Tabela A.0.4:  $p=0,50$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,50			1,10			3,00			2,00		
EMVb	(0,4500)	0,5000	(0,5000)	(0,9900)	1,2100	(1,2100)	(1,8700)	2,5300	(3,9600)	(1,4300)	2,5300	(3,6300)
EMVf	(0,4644)	0,5038	(0,5172)	(1,0928)	1,1549	(1,2218)	(1,9839)	2,5841	(3,8465)	(1,6266)	2,5069	(3,6896)
SiZer	(0,7642)	0,7754	(0,7848)	(1,2360)	1,2760	(1,3060)	(4,9687)	5,1415	(5,2501)	(1,8100)	1,9100	(2,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,4500)	2,1609	(5,8141)	(0,1564)	1,6783	(9,6266)	(0,6062)	4,1298	(15,346)
	0,50			1,10			5,00			2,00		
EMVb	(0,0000)	0,0000	(0,0000)	(0,5500)	0,5500	(0,5500)	(0,5500)	0,5500	(0,7700)	(2,3100)	2,9700	(3,4100)
EMVf	(0,0000)	0,0000	(0,0000)	(0,4861)	0,5134	(0,5405)	(0,5990)	0,6184	(0,7623)	(1,7894)	2,2896	(2,4761)
SiZer	(0,6188)	0,6277	(0,6371)	(1,1960)	1,2360	(1,2860)	(6,6359)	6,8540	(7,0720)	(1,7100)	1,8100	(1,9100)
EM	(0,0000)	0,0000	(0,0000)	(1,0519)	1,0898	(1,1295)	(0,6698)	0,6824	(0,6984)	(7,8084)	8,5551	(9,2151)
	0,50			1,10			5,00			5,00		
EMVb	(0,0000)	0,0000	(0,0500)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(1,2100)	1,2100	(1,4300)
EMVf	(0,0000)	0,0000	(0,0000)	(0,2025)	0,2025	(0,2025)	(0,4754)	0,4946	(0,5139)	(0,4316)	0,4660	(0,5076)
SiZer	(0,5104)	0,5171	(0,5237)	(1,1560)	1,1960	(1,2560)	(12,939)	13,601	(14,072)	(1,7100)	1,8100	(1,9100)
EM	(0,0000)	0,0000	(0,0000)	(1,0080)	1,0499	(1,0895)	(0,4952)	0,5033	(0,5127)	(22,835)	24,899	(26,936)
	0,50			1,10			2,00			5,00		
EMVb	(0,0000)	0,0000	(0,1000)	(0,5500)	0,7700	(0,7700)	(0,7700)	0,7700	(0,9900)	(1,8700)	2,5300	(2,7500)
EMVf	(0,0000)	0,0764	(0,1236)	(0,5882)	0,6293	(0,6850)	(0,7656)	0,9068	(0,9597)	(1,3978)	1,9370	(3,0125)
SiZer	(0,6733)	0,6813	(0,6896)	(1,1560)	1,1960	(1,2460)	(6,2545)	6,4688	(6,6815)	(1,9100)	2,0100	(2,1100)
EM	(0,0000)	0,0000	(0,0000)	(0,9443)	0,9745	(1,0052)	(0,7069)	0,7239	(0,7422)	(5,3578)	5,8214	(6,2579)
	0,50			2,00			1,90			5,00		
EMVb	(0,5000)	0,5000	(0,5500)	(1,8700)	1,8700	(2,0900)	(1,8700)	2,0900	(2,3100)	(3,6300)	4,7300	(5,1700)
EMVf	(0,4888)	0,5066	(0,5250)	(1,8291)	1,9516	(2,0696)	(1,8483)	2,0462	(2,3217)	(3,7418)	4,6374	(5,2259)
SiZer	(0,8568)	0,8661	(0,8774)	(1,4860)	1,5260	(1,5660)	(5,2236)	5,3632	(5,5205)	(1,7100)	1,7100	(1,8100)
EM	(0,5336)	1,0000	(1,0000)	(0,5955)	1,6381	(5,0517)	(0,7631)	0,8709	(6,2563)	(0,1608)	5,3293	(8,5809)
	0,50			2,00			2,00			6,10		
EMVb	(0,5000)	0,5000	(0,5500)	(1,8700)	1,8700	(1,8700)	(2,0900)	2,3100	(2,5300)	(4,2900)	5,1700	(5,3900)
EMVf	(0,4920)	0,5096	(0,5246)	(1,7713)	1,9458	(2,0236)	(2,0744)	2,2547	(2,5854)	(4,3583)	5,0535	(5,8767)
SiZer	(0,7820)	0,7908	(0,7996)	(1,4060)	1,4460	(1,4860)	(5,6111)	5,8008	(5,9803)	(1,7100)	1,7100	(1,8100)
EM	(1,0000)	1,0000	(1,0000)	(1,6938)	15,685	(26,581)	(0,7858)	11,954	(87,118)	(0,0181)	0,3015	(7,3685)
	0,50			2,10			2,00			5,00		
EMVb	(0,5000)	0,5000	(0,5500)	(1,8700)	2,0900	(2,0900)	(1,8700)	2,0900	(2,5300)	(3,8500)	4,7300	(5,1700)
EMVf	(0,4855)	0,5047	(0,5252)	(1,9224)	2,0451	(2,1593)	(1,8833)	2,1033	(2,4757)	(3,8030)	4,7835	(5,2654)
SiZer	(0,8548)	0,8655	(0,8758)	(1,5160)	1,5560	(1,5960)	(5,3534)	5,4881	(5,6334)	(1,7100)	1,7100	(1,7100)
EM	(0,5339)	0,5371	(1,0000)	(1,5949)	1,7074	(2,5871)	(0,8217)	0,8516	(2,1662)	(1,4760)	6,1526	(6,9852)
	0,50			1,10			4,00			1,30		
EMVb	(0,4500)	0,5000	(0,5000)	(0,9900)	1,2100	(1,2100)	(2,5300)	3,6300	(4,5100)	(0,9900)	1,6500	(2,3100)
EMVf	(0,4730)	0,4972	(0,5112)	(1,0619)	1,1450	(1,2246)	(2,5854)	3,3941	(4,9266)	(1,0282)	1,6506	(2,4274)
SiZer	(0,7962)	0,8052	(0,8147)	(1,2760)	1,3210	(1,3660)	(4,8195)	4,9564	(5,0928)	(1,8100)	1,9100	(1,9100)
EM	(1,0000)	1,0000	(1,0000)	(0,1251)	2,2258	(7,3418)	(0,1793)	1,7502	(7,7895)	(0,1288)	2,7271	(65,713)
	0,50			1,10			1,20			4,30		
EMVb	(0,5000)	0,5500	(0,6000)	(0,9900)	0,9900	(1,2100)	(1,2100)	1,4300	(1,8700)	(2,5300)	3,6300	(4,7300)
EMVf	(0,5164)	0,5643	(0,5973)	(1,0593)	1,1120	(1,2069)	(1,2036)	1,4462	(2,0669)	(2,9287)	3,9403	(4,7487)
SiZer	(0,8389)	0,8493	(0,8583)	(1,1960)	1,2260	(1,2660)	(4,3081)	4,4482	(4,5883)	(2,1100)	2,2100	(2,3100)
EM	(0,0000)	1,0000	(1,0000)	(0,4033)	0,9816	(2,7621)	(0,8332)	0,8949	(4,5158)	(0,3316)	2,4738	(2,9472)
	0,50			1,10			4,20			4,30		
EMVb	(0,1500)	0,2000	(0,2000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(2,7500)	3,1900	(3,6300)
EMVf	(0,0000)	0,0000	(0,0074)	(0,2760)	0,2760	(0,2988)	(0,4089)	0,4232	(0,4333)	(1,7868)	2,1901	(2,6301)
SiZer	(0,5384)	0,5464	(0,5545)	(1,1560)	1,2060	(1,2660)	(9,8748)	10,319	(10,729)	(1,7100)	1,8100	(2,0100)
EM	(0,0000)	0,0000	(0,0000)	(0,9964)	1,0329	(1,0723)	(0,5568)	0,5669	(0,5784)	(14,192)	15,539	(16,967)
	0,50			0,40			5,00			2,50		
EMVb	(0,0000)	0,0000	(0,0000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)
EMVf	(0,0000)	0,0000	(0,0000)	(0,2035)	0,2173	(0,2240)	(0,3215)	0,3520	(0,3767)	(0,1744)	0,2137	(0,2778)
SiZer	(0,5126)	0,5197	(0,5276)	(1,0260)	1,0260	(1,0260)	(21,778)	21,784	(21,794)	(1,5100)	1,5100	(1,6100)
EM	(0,0000)	0,0000	(0,0000)	(0,3135)	0,3197	(0,3258)	(0,7699)	0,8031	(0,8443)	(1,0587)	1,1924	(1,3549)
	0,50			0,60			1,20			4,30		
EMVb	(0,2500)	0,3000	(0,4000)	(0,3300)	0,3300	(0,3300)	(0,5500)	0,5500	(0,5500)	(1,2100)	1,4300	(1,8700)
EMVf	(0,0500)	0,1000	(0,2000)	(0,3373)	0,3562	(0,3668)	(0,5653)	0,5856	(0,6447)	(1,0028)	1,2723	(1,6019)
SiZer	(0,6973)	0,7082	(0,7191)	(1,0360)	1,0410	(1,0460)	(10,154)	10,180	(10,566)	(1,7100)	1,8100	(1,9100)
EM	(0,0000)	0,0000	(0,0000)	(0,5389)	0,5531	(0,5688)	(0,8611)	0,8834	(0,9135)	(1,3655)	1,5195	(1,6574)

Tabela A.0.5:  $p=0,70$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,70	1,10	3,00	2,00
EMVb	0,6617 (0,0942)	1,1290 (0,1182)	2,9273 (1,4804)	2,7238 (1,4747)
EMVf	0,6793 (0,0841)	1,1787 (0,1286)	2,9951 (1,3842)	2,8524 (1,4204)
SiZer	0,8677 (0,0141)	1,2619 (0,0981)	4,0366 (0,2676)	2,0924 (0,2743)
EM	0,8258 (0,3775)	224,32 (447,79)	370,75 (822,54)	213,26 (886,45)
	0,70	1,10	5,00	2,00
EMVb	0,6934 (0,0180)	1,1447 (0,1076)	3,9983 (1,1763)	3,0202 (1,2623)
EMVf	0,6953 (0,0161)	1,1502 (0,0948)	4,0642 (1,2224)	3,0267 (1,2547)
SiZer	0,7775 (0,0134)	1,1902 (0,0629)	4,8641 (0,3235)	2,2506 (0,2069)
EM	0,9988 (0,0345)	7,0351 (43,221)	68,388 (569,53)	129,74 (808,29)
	0,70	1,10	5,00	5,00
EMVb	0,6616 (0,0508)	0,9242 (0,1647)	3,4135 (1,3625)	4,9445 (0,6159)
EMVf	0,6439 (0,0648)	0,9138 (0,1591)	3,5093 (1,3485)	4,7596 (0,5772)
SiZer	0,3735 (0,3555)	0,5981 (0,5701)	4,3661 (4,1945)	1,3857 (1,3433)
EM	0,7762 (0,4172)	7,1430 (86,472)	5,4188 (76,478)	93,863 (657,85)
	0,70	1,10	2,00	5,00
EMVb	0,7018 (0,0318)	1,0886 (0,1095)	2,6805 (1,1033)	4,1083 (1,3092)
EMVf	0,7069 (0,0269)	1,1035 (0,0949)	2,7148 (1,1124)	4,2502 (1,3617)
SiZer	0,8097 (0,0132)	1,1612 (0,0905)	4,6560 (0,3248)	2,4110 (0,2443)
EM	0,7559 (0,4258)	174,44 (607,81)	125,52 (392,60)	319,45 (1086,9)
	0,70	2,00	1,90	5,00
EMVb	0,6836 (0,0868)	1,8977 (0,2766)	2,3188 (1,0111)	4,2328 (1,2410)
EMVf	0,6883 (0,0756)	1,9237 (0,2932)	2,3239 (0,9938)	4,3315 (1,3659)
SiZer	0,9231 (0,0137)	1,4563 (0,1105)	4,2757 (0,2635)	1,8110 (0,2596)
EM	0,8941 (0,3064)	142,96 (424,98)	68,032 (146,25)	164,32 (748,38)
	0,70	2,00	2,00	6,10
EMVb	0,7009 (0,0561)	1,8722 (0,2195)	2,6594 (0,8872)	4,4937 (1,0968)
EMVf	0,7026 (0,0512)	1,8867 (0,2365)	2,6703 (0,8640)	4,5430 (1,1925)
SiZer	0,8762 (0,0134)	1,3583 (0,0885)	4,3876 (0,3896)	1,9636 (0,3367)
EM	0,6450 (0,4767)	157,42 (602,93)	138,07 (502,01)	72,615 (457,46)
	0,70	2,10	2,00	5,00
EMVb	0,6910 (0,0746)	2,0055 (0,2899)	2,5089 (1,0233)	4,1690 (1,3094)
EMVf	0,6943 (0,0652)	2,0163 (0,3072)	2,5142 (0,9881)	4,2200 (1,4378)
SiZer	0,9223 (0,0143)	1,4766 (0,1379)	4,3406 (0,2129)	1,7834 (0,2160)
EM	0,9773 (0,1474)	180,77 (911,69)	90,9344 (411,13)	176,34 (740,61)
	0,70	1,10	4,00	1,30
EMVb	0,6545 (0,0963)	1,1497 (0,1204)	3,2045 (1,5196)	2,2968 (1,4673)
EMVf	0,6696 (0,0880)	1,1863 (0,1209)	3,2741 (1,4942)	2,3630 (1,3864)
SiZer	0,3969 (0,4411)	0,5934 (0,6637)	1,7701 (1,9863)	0,8971 (1,0368)
EM	0,9499 (0,2141)	267,74 (693,74)	223,38 (576,81)	68,612 (668,63)
	0,70	1,10	1,20	4,30
EMVb	0,1757 (0,3066)	0,2677 (0,4676)	0,4847 (1,0639)	0,8362 (1,6074)
EMVf	0,3300 (0,2370)	0,7733 (0,2234)	1,2801 (0,7799)	1,5129 (1,4051)
SiZer	0,9114 (0,0136)	1,2580 (0,1510)	3,6796 (0,2484)	2,1938 (0,4094)
EM	1,0000 (0,0000)	0,1958 (0,0157)	1,0861 (0,1748)	0,1183 (0,0399)
	0,70	1,10	4,20	4,30
EMVb	0,6992 (0,0083)	1,1040 (0,1109)	4,4730 (0,7617)	4,1914 (1,0643)
EMVf	0,7001 (0,0101)	1,1130 (0,0780)	4,4535 (0,7679)	4,2884 (0,9628)
SiZer	0,7272 (0,0127)	1,1461 (0,0412)	6,6882 (0,5081)	2,4896 (0,2622)
EM	0,2450 (0,4306)	1,8925 (11,854)	0,7410 (0,4765)	80,842 (546,24)
	0,70	0,40	5,00	2,50
EMVb	0,2454 (0,1163)	0,3300 (0,0000)	0,3300 (0,0000)	0,5101 (0,0926)
EMVf	0,0764 (0,0954)	0,2678 (0,0545)	0,3262 (0,0389)	0,6461 (0,0928)
SiZer	0,7114 (0,0122)	1,0262 (0,0027)	21,303 (1,7870)	1,5454 (0,2082)
EM	0,1429 (0,3502)	0,3120 (0,1269)	0,9304 (0,3851)	0,7554 (4,0972)
	0,70	0,60	1,20	4,30
EMVb	0,6910 (0,0956)	0,5667 (0,0584)	2,4587 (1,5781)	2,4482 (1,3351)
EMVf	0,6960 (0,0972)	0,5971 (0,0346)	2,5592 (1,7619)	2,7732 (1,3957)
SiZer	0,8261 (0,0138)	1,0473 (0,0205)	6,7191 (1,1519)	2,7060 (0,4549)
EM	0,7391 (0,4490)	1,1131 (1,1365)	59,687 (275,91)	8,1624 (16,342)

Tabela A.0.6:  $p=0,70$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,70			1,10			3,00			2,00		
EMVb	(0,6500)	0,7000	(0,7000)	(0,9900)	1,2100	(1,2100)	(1,6500)	2,5300	(4,5100)	(1,4300)	2,5300	(4,0700)
EMVf	(0,6823)	0,7002	(0,7273)	(1,0818)	1,1529	(1,2969)	(1,8713)	2,7659	(4,2742)	(1,5511)	2,6074	(4,0542)
SiZer	(0,8582)	0,8678	(0,8767)	(1,2260)	1,2660	(1,2960)	(3,9343)	4,0458	(4,1609)	(2,0100)	2,0100	(2,1100)
EM	(1,0000)	1,0000	(1,0000)	(0,9583)	28,679	(319,13)	(0,8868)	69,378	(449,98)	(2,2316)	9,7834	(40,075)
	0,70			1,10			5,00			2,00		
EMVb	(0,7000)	0,7000	(0,7000)	(0,9900)	1,2100	(1,2100)	(2,9700)	4,2900	(5,1700)	(1,8700)	2,7500	(4,0700)
EMVf	(0,6873)	0,6962	(0,7040)	(1,0717)	1,1507	(1,1890)	(2,9467)	4,1734	(5,2328)	(1,9649)	2,6301	(3,9405)
SiZer	(0,7686)	0,7777	(0,7862)	(1,1560)	1,1860	(1,2210)	(4,7161)	4,8636	(5,0369)	(2,1100)	2,2100	(2,3100)
EM	(1,0000)	1,0000	(1,0000)	(0,2574)	0,8998	(1,8134)	(0,3014)	0,6702	(1,7798)	(0,0509)	1,4663	(8,6921)
	0,70			1,10			5,00			5,00		
EMVb	(0,6500)	0,6500	(0,7000)	(0,7700)	0,9900	(0,9900)	(2,0900)	3,1900	(4,9500)	(4,9500)	5,1700	(5,3900)
EMVf	(0,6146)	0,6473	(0,6973)	(0,7891)	0,8908	(1,0550)	(2,3667)	3,2180	(4,8252)	(4,4242)	4,7799	(5,1453)
SiZer	(0,0000)	0,6923	(0,7118)	(0,0000)	1,0860	(1,1360)	(0,0000)	6,9996	(8,5454)	(0,0000)	2,3100	(2,5600)
EM	(1,0000)	1,0000	(1,0000)	(0,0580)	0,5833	(0,9650)	(0,7517)	1,2242	(1,8595)	(0,0037)	1,5629	(4,2349)
	0,70			1,10			2,00			5,00		
EMVb	(0,7000)	0,7000	(0,7000)	(0,9900)	0,9900	(1,2100)	(2,0900)	2,3100	(3,1900)	(2,9700)	4,6200	(5,3900)
EMVf	(0,6962)	0,7060	(0,7195)	(1,0350)	1,0988	(1,1541)	(1,9762)	2,2942	(3,0149)	(3,1610)	4,8116	(5,3507)
SiZer	(0,8006)	0,8099	(0,8179)	(1,1260)	1,1560	(1,1860)	(4,4703)	4,6278	(4,8460)	(2,3100)	2,4100	(2,5100)
EM	(0,5235)	1,0000	(1,0000)	(0,8469)	16,946	(104,22)	(0,8287)	40,246	(96,053)	(2,2113)	7,3042	(50,438)
	0,70			2,00			1,90			5,00		
EMVb	(0,6500)	0,7000	(0,7000)	(1,6500)	1,8700	(2,0900)	(1,6500)	2,0900	(2,5300)	(3,4100)	4,7300	(5,1700)
EMVf	(0,6725)	0,7062	(0,7260)	(1,7322)	1,9076	(2,0771)	(1,6633)	2,0767	(2,6209)	(3,4329)	4,6707	(5,4096)
SiZer	(0,9144)	0,9233	(0,9321)	(1,4060)	1,4460	(1,4960)	(4,1701)	4,2788	(4,3916)	(1,7100)	1,8100	(1,8100)
EM	(1,0000)	1,0000	(1,0000)	(1,1785)	8,8991	(42,584)	(0,8937)	8,6674	(31,402)	(2,6720)	7,1029	(14,687)
	0,70			2,00			2,00			6,10		
EMVb	(0,7000)	0,7000	(0,7000)	(1,6500)	1,8700	(2,0900)	(2,0900)	2,5300	(2,7500)	(4,0700)	4,9500	(5,3900)
EMVf	(0,6879)	0,7137	(0,7265)	(1,7282)	1,9061	(2,0376)	(2,1217)	2,5069	(2,8880)	(3,8684)	4,8620	(5,4056)
SiZer	(0,8676)	0,8758	(0,8850)	(1,3260)	1,3560	(1,3860)	(4,2845)	4,4038	(4,5503)	(1,9100)	1,9100	(2,0100)
EM	(0,0000)	1,0000	(1,0000)	(1,1935)	1,3140	(20,845)	(0,7824)	0,8234	(20,266)	(1,3206)	3,3814	(4,0802)
	0,70			2,10			2,00			5,00		
EMVb	(0,7000)	0,7000	(0,7000)	(1,8700)	2,0900	(2,0900)	(1,8700)	2,0900	(2,7500)	(3,1900)	4,7300	(5,1700)
EMVf	(0,6857)	0,7109	(0,7245)	(1,7698)	2,0058	(2,2093)	(1,8586)	2,2609	(2,9535)	(3,0673)	4,5904	(5,4061)
SiZer	(0,9129)	0,9230	(0,9322)	(1,4260)	1,4760	(1,5160)	(4,2419)	4,3414	(4,4533)	(1,7100)	1,8100	(1,8100)
EM	(1,0000)	1,0000	(1,0000)	(4,8269)	17,790	(37,183)	(1,9904)	14,053	(25,787)	(0,8866)	2,8351	(7,1643)
	0,70			1,10			4,00			1,30		
EMVb	(0,6500)	0,7000	(0,7000)	(0,9900)	1,2100	(1,2100)	(1,8700)	3,1900	(4,7300)	(1,2100)	1,6500	(3,1900)
EMVf	(0,6673)	0,6920	(0,7149)	(1,1016)	1,1649	(1,2777)	(2,1062)	3,2064	(4,5305)	(1,2154)	1,8503	(3,2496)
SiZer	(0,0000)	0,0000	(0,8828)	(0,0000)	0,0000	(1,3060)	(0,0000)	0,0000	(3,9766)	(0,0000)	0,0000	(1,9100)
EM	(1,0000)	1,0000	(1,0000)	(8,2744)	35,433	(213,82)	(9,0540)	29,508	(203,66)	(1,3621)	2,7460	(6,0119)
	0,70			1,10			1,20			4,30		
EMVb	(0,0000)	0,0000	(0,2500)	(0,0000)	0,0000	(0,3850)	(0,0000)	0,0000	(0,3850)	(0,0000)	0,0000	(0,3850)
EMVf	(0,2000)	0,2000	(0,3625)	(0,6216)	0,6216	(0,7798)	(0,9998)	0,9998	(0,9998)	(0,7661)	0,7661	(0,9998)
SiZer	(0,9026)	0,9109	(0,9204)	(1,1960)	1,2410	(1,2860)	(3,6090)	3,6879	(3,7819)	(1,9100)	2,1100	(2,3100)
EM	(1,0000)	1,0000	(1,0000)	(0,1854)	0,1951	(0,2067)	(0,9642)	1,0680	(1,1965)	(0,0903)	0,1126	(0,1401)
	0,70			1,10			4,20			4,30		
EMVb	(0,7000)	0,7000	(0,7000)	(0,9900)	1,2100	(1,2100)	(3,8500)	4,5100	(5,1700)	(2,9700)	4,2900	(5,3900)
EMVf	(0,6943)	0,7001	(0,7062)	(1,0515)	1,1219	(1,1682)	(3,8454)	4,3543	(5,1544)	(3,3178)	4,4261	(5,0428)
SiZer	(0,7182)	0,7266	(0,7354)	(1,1160)	1,1360	(1,1760)	(6,3522)	6,6442	(7,0250)	(2,3100)	2,4100	(2,6100)
EM	(0,0000)	0,0000	(0,0000)	(0,6677)	0,6862	(0,7120)	(0,7328)	0,7649	(0,7896)	(1,9110)	2,1269	(2,4471)
	0,70			0,40			5,00			2,50		
EMVb	(0,1500)	0,2000	(0,3000)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,5500)	0,5500	(0,5500)
EMVf	(0,0006)	0,0315	(0,1021)	(0,2469)	0,2871	(0,3082)	(0,3215)	0,3281	(0,3426)	(0,5983)	0,6395	(0,6942)
SiZer	(0,7030)	0,7123	(0,7200)	(1,0260)	1,0260	(1,0260)	(21,776)	21,780	(21,788)	(1,4100)	1,5100	(1,6100)
EM	(0,0000)	0,0000	(0,0000)	(0,3543)	0,3620	(0,3679)	(1,0081)	1,0661	(1,1175)	(0,3566)	0,4046	(0,4402)
	0,70			0,60			1,20			4,30		
EMVb	(0,7000)	0,7000	(0,7500)	(0,5500)	0,5500	(0,5500)	(1,2100)	1,7600	(4,0700)	(1,2100)	2,3100	(3,1900)
EMVf	(0,6981)	0,7144	(0,7460)	(0,5751)	0,5920	(0,6077)	(1,3271)	1,6146	(4,4151)	(1,3342)	2,9092	(3,7606)
SiZer	(0,8162)	0,8250	(0,8371)	(1,0360)	1,0360	(1,0560)	(5,7389)	6,7227	(7,9135)	(2,3100)	2,7100	(3,0100)
EM	(0,2500)	1,0000	(1,0000)	(0,5835)	0,8015	(1,2443)	(0,3910)	0,9425	(2,3901)	(0,2410)	1,1436	(8,2522)

Tabela A.0.7:  $p=0,75$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,75	1,10	3,00	2,00
EMVb	0,7021 (0,0956)	1,1277 (0,1288)	2,7861 (1,5999)	2,8961 (1,5585)
EMVf	0,7197 (0,0839)	1,1715 (0,1446)	2,8539 (1,5081)	3,0161 (1,5302)
SiZer	0,8919 (0,0133)	1,2699 (0,0882)	3,8028 (0,2613)	2,0736 (0,3271)
EM	0,9974 (0,0506)	59,119 (508,87)	53,496 (554,92)	244,59 (1057,1)
	0,75	1,10	5,00	2,00
EMVb	0,7436 (0,0212)	1,1374 (0,1184)	3,9534 (1,2449)	3,0787 (1,3637)
EMVf	0,7447 (0,0189)	1,1455 (0,1015)	3,9828 (1,2877)	3,1192 (1,3687)
SiZer	0,8154 (0,0134)	1,1837 (0,0703)	4,4118 (0,3187)	2,3556 (0,2740)
EM	1,0000 (0,0000)	29,339 (57,089)	24,288 (22,938)	17,719 (39,551)
	0,75	1,10	5,00	5,00
EMVb	0,7494 (0,0054)	1,0943 (0,1100)	5,1599 (0,4261)	4,7652 (0,6768)
EMVf	0,7504 (0,0083)	1,1058 (0,0753)	5,1815 (0,4621)	4,8843 (0,5746)
SiZer	0,7590 (0,0128)	1,1187 (0,0319)	7,2017 (0,5963)	2,8802 (0,3400)
EM	0,7010 (0,4570)	21,479 (102,31)	7,6818 (52,116)	32,881 (202,26)
	0,75	1,10	2,00	5,00
EMVb	0,7457 (0,0406)	1,0789 (0,1150)	2,6950 (1,1980)	4,0304 (1,3248)
EMVf	0,7525 (0,0333)	1,0972 (0,1039)	2,7100 (1,1511)	4,1906 (1,3660)
SiZer	0,8425 (0,0129)	1,1612 (0,0670)	4,2317 (0,2717)	2,5154 (0,2860)
EM	0,8714 (0,3171)	91,257 (418,84)	72,180 (228,46)	24,755 (92,222)
	0,75	2,00	1,90	5,00
EMVb	0,7243 (0,0986)	1,8964 (0,3211)	2,4262 (1,1682)	4,1219 (1,3497)
EMVf	0,7319 (0,0849)	1,9250 (0,3378)	2,4516 (1,1425)	4,1825 (1,4335)
SiZer	0,9369 (0,0132)	1,4579 (0,1181)	4,0574 (0,1957)	1,7522 (0,2360)
EM	0,9495 (0,2174)	82,720 (168,88)	304,16 (945,53)	54,260 (326,64)
	0,75	2,00	2,00	6,10
EMVb	0,7467 (0,0603)	1,8731 (0,2570)	2,7711 (1,0797)	4,4255 (1,2067)
EMVf	0,7493 (0,0538)	1,8988 (0,2730)	2,7829 (1,0677)	4,5122 (1,2755)
SiZer	0,8971 (0,0138)	1,3460 (0,0973)	4,1047 (0,2786)	1,9404 (0,2679)
EM	0,9203 (0,2692)	149,08 (321,38)	138,31 (276,24)	136,12 (861,61)
	0,75	2,10	2,00	5,00
EMVb	0,7302 (0,0896)	1,9848 (0,3427)	2,6193 (1,2507)	4,0757 (1,4401)
EMVf	0,7364 (0,0754)	2,0088 (0,3697)	2,6193 (1,1779)	4,1390 (1,5513)
SiZer	0,9364 (0,0139)	1,4743 (0,1300)	4,0892 (0,2903)	1,7528 (0,3075)
EM	0,8189 (0,3844)	91,565 (196,07)	268,93 (964,96)	108,09 (618,09)
	0,75	1,10	4,00	1,30
EMVb	0,6983 (0,1009)	1,1528 (0,1329)	3,1038 (1,6306)	2,4240 (1,5308)
EMVf	0,7141 (0,0910)	1,1879 (0,1401)	3,1998 (1,6109)	2,4799 (1,4637)
SiZer	0,9058 (0,0133)	1,3382 (0,1805)	3,8082 (0,2187)	1,9318 (0,3466)
EM	0,9128 (0,2828)	80,418 (437,69)	126,23 (567,88)	102,60 (677,14)
	0,75	1,10	1,20	4,30
EMVb	0,7483 (0,0800)	1,1110 (0,1480)	1,9206 (1,2356)	3,5160 (1,4158)
EMVf	0,7740 (0,0650)	1,1581 (0,1623)	2,0628 (1,1980)	3,7126 (1,4415)
SiZer	0,9270 (0,0134)	1,2898 (0,1516)	3,5515 (0,1720)	2,0462 (0,5256)
EM	1,0000 (0,0000)	22,935 (278,53)	245,79 (978,62)	14,857 (57,959)
	0,75	1,10	4,20	4,30
EMVb	0,7488 (0,0083)	1,1057 (0,1108)	4,4818 (0,7790)	4,2051 (1,0923)
EMVf	0,7499 (0,0098)	1,1113 (0,0824)	4,4810 (0,7852)	4,2573 (0,9879)
SiZer	0,7724 (0,0124)	1,1350 (0,0518)	5,8000 (0,5069)	2,7454 (0,3250)
EM	0,6632 (0,4732)	43,534 (326,55)	3,2386 (5,1501)	56,652 (358,54)
	0,75	0,40	5,00	2,50
EMVb	0,3845 (0,0615)	0,3300 (0,0000)	0,3300 (0,0000)	1,0406 (0,2508)
EMVf	0,2049 (0,0666)	0,2389 (0,0376)	0,3654 (0,0623)	1,2054 (0,4328)
SiZer	0,0380 (0,1656)	0,0508 (0,2217)	0,6894 (3,0080)	0,1163 (0,5093)
EM	0,0000 (0,0000)	0,3907 (0,0085)	1,1174 (0,0653)	0,3771 (0,0449)
	0,75	0,60	1,20	4,30
EMVb	0,7293 (0,1093)	0,5733 (0,0678)	2,4174 (1,5986)	2,5639 (1,5062)
EMVf	0,7322 (0,1101)	0,5959 (0,0380)	2,4787 (1,7588)	2,8565 (1,5620)
SiZer	0,8553 (0,0147)	1,0475 (0,0210)	5,7744 (1,0055)	3,0164 (0,5144)
EM	0,1600 (0,3742)	1,2138 (1,9752)	0,8322 (0,3030)	326,86 (1355,1)

Tabela A.0.8:  $p=0,75$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,75			1,10			3,00			2,00		
EMVb	(0,7000)	0,7500	(0,7500)	(0,9900)	1,2100	(1,2100)	(1,4300)	2,2000	(4,2900)	(1,4300)	2,7500	(4,2900)
EMVf	(0,7179)	0,7437	(0,7666)	(1,0551)	1,1501	(1,2891)	(1,6266)	2,3474	(4,3043)	(1,4462)	2,9503	(4,4557)
SiZer	(0,8837)	0,8914	(0,9000)	(1,2260)	1,2660	(1,3060)	(3,7117)	3,8165	(3,9084)	(1,9100)	2,0100	(2,1100)
EM	(1,0000)	1,0000	(1,0000)	(0,5966)	1,4526	(3,9418)	(0,5168)	1,6072	(3,2998)	(0,7266)	3,5402	(8,4023)
	0,75			1,10			5,00			2,00		
EMVb	(0,7500)	0,7500	(0,7500)	(0,9900)	1,2100	(1,2100)	(2,7500)	4,0700	(5,1700)	(1,8700)	2,7500	(4,2900)
EMVf	(0,7386)	0,7466	(0,7541)	(1,0572)	1,1512	(1,2032)	(2,6533)	4,1059	(5,2434)	(1,9253)	2,6354	(4,4158)
SiZer	(0,8062)	0,8156	(0,8249)	(1,1460)	1,1760	(1,2060)	(4,2605)	4,4194	(4,5951)	(2,2100)	2,3100	(2,4100)
EM	(1,0000)	1,0000	(1,0000)	(6,4938)	14,823	(28,269)	(10,310)	17,857	(29,947)	(9,5056)	12,532	(15,912)
	0,75			1,10			5,00			5,00		
EMVb	(0,7500)	0,7500	(0,7500)	(0,9900)	0,9900	(1,2100)	(5,1700)	5,3900	(5,3900)	(4,0700)	5,1700	(5,3900)
EMVf	(0,7453)	0,7506	(0,7557)	(1,0530)	1,0840	(1,1526)	(5,0457)	5,2657	(5,4754)	(4,4011)	4,9740	(5,1538)
SiZer	(0,7498)	0,7593	(0,7672)	(1,0960)	1,1160	(1,1360)	(6,7978)	7,1709	(7,5935)	(2,6100)	2,8100	(3,0100)
EM	(0,0000)	1,0000	(1,0000)	(0,5824)	0,6581	(2,6035)	(0,8198)	0,9350	(2,2121)	(0,9961)	1,4386	(8,2545)
	0,75			1,10			2,00			5,00		
EMVb	(0,7500)	0,7500	(0,7500)	(0,9900)	0,9900	(1,2100)	(1,8700)	2,3100	(3,1900)	(2,9700)	4,5100	(5,1700)
EMVf	(0,7450)	0,7556	(0,7654)	(1,0149)	1,0683	(1,1599)	(1,9655)	2,2866	(3,0987)	(3,0428)	4,6582	(5,3341)
SiZer	(0,8341)	0,8424	(0,8510)	(1,1260)	1,1560	(1,1860)	(4,0731)	4,2228	(4,3678)	(2,3100)	2,5100	(2,6100)
EM	(1,0000)	1,0000	(1,0000)	(0,8283)	11,761	(38,426)	(1,8860)	12,802	(36,830)	(1,5852)	6,8276	(10,139)
	0,75			2,00			1,90			5,00		
EMVb	(0,7000)	0,7500	(0,7500)	(1,6500)	1,8700	(2,0900)	(1,6500)	2,0900	(2,9700)	(3,1900)	4,7300	(5,1700)
EMVf	(0,7186)	0,7540	(0,7771)	(1,6985)	1,8798	(2,1356)	(1,6731)	2,2822	(2,8367)	(3,1901)	4,5640	(5,3344)
SiZer	(0,9282)	0,9371	(0,9465)	(1,3960)	1,4560	(1,5060)	(3,9680)	4,0541	(4,1626)	(1,7100)	1,7100	(1,8100)
EM	(1,0000)	1,0000	(1,0000)	(0,9745)	3,9498	(41,793)	(0,8747)	4,3411	(25,111)	(2,9008)	11,274	(18,221)
	0,75			2,00			2,00			6,10		
EMVb	(0,7500)	0,7500	(0,7500)	(1,6500)	1,8700	(2,0900)	(2,0900)	2,3100	(3,1900)	(3,7400)	4,9500	(5,3900)
EMVf	(0,7298)	0,7613	(0,7777)	(1,7153)	1,8758	(2,0639)	(2,0871)	2,5462	(3,1667)	(3,7944)	5,0056	(5,4059)
SiZer	(0,8884)	0,8967	(0,9073)	(1,3060)	1,3460	(1,3760)	(4,0113)	4,1074	(4,2037)	(1,8100)	1,9100	(2,0100)
EM	(1,0000)	1,0000	(1,0000)	(1,1374)	7,4423	(88,542)	(0,8551)	8,6294	(112,75)	(2,3368)	7,5300	(18,357)
	0,75			2,10			2,00			5,00		
EMVb	(0,7000)	0,7500	(0,7500)	(1,6500)	1,8700	(2,3100)	(1,8700)	2,0900	(3,1900)	(2,7500)	4,7300	(5,3900)
EMVf	(0,7222)	0,7560	(0,7781)	(1,7399)	1,9905	(2,2118)	(1,8144)	2,2964	(3,2133)	(2,7367)	4,6251	(5,4453)
SiZer	(0,9272)	0,9365	(0,9460)	(1,4160)	1,4660	(1,5160)	(4,0124)	4,1103	(4,2000)	(1,6600)	1,7100	(1,8100)
EM	(1,0000)	1,0000	(1,0000)	(0,7336)	1,9196	(20,517)	(0,3233)	0,9857	(29,663)	(1,3794)	2,7794	(5,7116)
	0,75			1,10			4,00			1,30		
EMVb	(0,7000)	0,7500	(0,7500)	(0,9900)	1,2100	(1,2100)	(1,5400)	2,9700	(4,7300)	(1,2100)	1,8700	(3,6300)
EMVf	(0,7130)	0,7421	(0,7639)	(1,0975)	1,1556	(1,2781)	(1,8337)	3,1496	(4,5662)	(1,2339)	2,0666	(3,6139)
SiZer	(0,8976)	0,9056	(0,9149)	(1,2760)	1,3210	(1,3760)	(3,7317)	3,8146	(3,9118)	(1,8100)	1,9100	(2,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,9307)	6,2296	(19,981)	(0,9080)	5,9980	(25,435)	(1,9002)	6,5538	(13,985)
	0,75			1,10			1,20			4,30		
EMVb	(0,7000)	0,7500	(0,8000)	(0,9900)	0,9900	(1,2100)	(1,2100)	1,4300	(2,2000)	(2,5300)	3,6300	(4,7300)
EMVf	(0,7554)	0,7920	(0,8099)	(1,0487)	1,1241	(1,2781)	(1,3946)	1,6266	(2,3129)	(2,7419)	3,8650	(4,9662)
SiZer	(0,9173)	0,9276	(0,9370)	(1,2160)	1,2760	(1,3360)	(3,4828)	3,5520	(3,6255)	(1,7100)	1,9100	(2,2100)
EM	(1,0000)	1,0000	(1,0000)	(0,2233)	0,6797	(1,6155)	(18,172)	33,527	(73,741)	(2,4297)	6,8254	(11,088)
	0,75			1,10			4,20			4,30		
EMVb	(0,7500)	0,7500	(0,7500)	(0,9900)	1,2100	(1,2100)	(3,8500)	4,2900	(5,1700)	(2,9700)	4,2900	(5,3900)
EMVf	(0,7446)	0,7506	(0,7564)	(1,0464)	1,1237	(1,1633)	(3,8677)	4,3867	(5,2049)	(3,2792)	4,3517	(5,0846)
SiZer	(0,7650)	0,7732	(0,7810)	(1,1060)	1,1260	(1,1560)	(5,5225)	5,7556	(6,0391)	(2,5100)	2,7100	(2,9100)
EM	(0,0000)	1,0000	(1,0000)	(0,6487)	0,6974	(7,8650)	(0,8162)	1,2529	(5,1882)	(1,3420)	1,9105	(17,591)
	0,75			0,40			5,00			2,50		
EMVb	(0,3500)	0,4000	(0,4500)	(0,3300)	0,3300	(0,3300)	(0,3300)	0,3300	(0,3300)	(0,7700)	0,9900	(1,2100)
EMVf	(0,1512)	0,2020	(0,2585)	(0,2132)	0,2433	(0,2657)	(0,3285)	0,3465	(0,3740)	(0,9408)	1,1941	(1,4480)
SiZer	(0,0000)	0,0000	(0,0000)	(0,0000)	0,0000	(0,0000)	(0,0000)	0,0000	(0,0000)	(0,0000)	0,0000	(0,0000)
EM	(0,0000)	0,0000	(0,0000)	(0,3849)	0,3906	(0,3970)	(1,0711)	1,1163	(1,1633)	(0,3443)	0,3712	(0,4086)
	0,75			0,60			1,20			4,30		
EMVb	(0,7000)	0,7500	(0,8000)	(0,5500)	0,5500	(0,5500)	(1,2100)	1,8700	(4,0700)	(1,2100)	2,3100	(3,6300)
EMVf	(0,7268)	0,7628	(0,7949)	(0,5728)	0,5875	(0,6064)	(1,1857)	1,5038	(4,0701)	(1,2672)	2,8059	(4,2742)
SiZer	(0,8454)	0,8555	(0,8659)	(1,0360)	1,0460	(1,0560)	(5,0196)	5,7590	(6,2113)	(2,7100)	3,0100	(3,3100)
EM	(0,0000)	0,0000	(0,0000)	(0,6286)	0,6440	(0,6629)	(0,9014)	0,9424	(0,9745)	(0,8894)	0,9311	(1,0008)

Tabela A.0.9:  $p=0,80$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,80	1,10	3,00	2,00
EMVb	0,7244 (0,1122)	1,1365 (0,1467)	2,4002 (1,6583)	3,2965 (1,5968)
EMVf	0,7459 (0,0983)	1,1667 (0,1650)	2,5232 (1,6543)	3,3767 (1,6156)
SiZer	0,9145 (0,0138)	1,2961 (0,1370)	3,6068 (0,1772)	1,9844 (0,4093)
EM	0,9688 (0,1734)	61,259 (278,79)	77,533 (672,29)	259,29 (1228,195)
	0,80	1,10	5,00	2,00
EMVb	0,7894 (0,0368)	1,1414 (0,1254)	3,7622 (1,3626)	3,2846 (1,4109)
EMVf	0,7912 (0,0339)	1,1539 (0,1210)	3,6784 (1,3558)	3,4457 (1,4622)
SiZer	0,8526 (0,0132)	1,1765 (0,0771)	3,9686 (0,3042)	2,4701 (0,3828)
EM	0,6047 (0,4155)	153,58 (786,58)	110,52 (632,72)	44,461 (250,41)
	0,80	1,10	5,00	5,00
EMVb	0,7993 (0,0057)	1,1024 (0,1105)	5,1207 (0,4605)	4,8468 (0,6488)
EMVf	0,8002 (0,0085)	1,1114 (0,0853)	5,1327 (0,4952)	4,9706 (0,5944)
SiZer	0,8068 (0,0130)	1,1072 (0,0292)	6,0740 (0,5371)	3,2435 (0,3981)
EM	0,9794 (0,1421)	15,293 (87,262)	89,969 (648,86)	92,711 (582,16)
	0,80	1,10	2,00	5,00
EMVb	0,7894 (0,0527)	1,0923 (0,1300)	2,7177 (1,3277)	4,0995 (1,3521)
EMVf	0,7964 (0,0455)	1,1080 (0,1280)	2,7531 (1,3276)	4,2331 (1,3757)
SiZer	0,8742 (0,0138)	1,1578 (0,0502)	3,8299 (0,2527)	2,6021 (0,4278)
EM	0,5684 (0,4118)	30,889 (150,53)	259,07 (1157,6)	135,17 (1020,2)
	0,80	2,00	1,90	5,00
EMVb	0,4864 (0,3702)	1,2516 (0,9860)	1,4738 (1,4819)	2,8373 (2,3252)
EMVf	0,5624 (0,2799)	1,5633 (0,7211)	1,8706 (1,2318)	3,2573 (2,0210)
SiZer	0,9498 (0,0124)	1,4895 (0,1540)	3,8426 (0,2470)	1,6366 (0,3511)
EM	1,0000 (0,0000)	0,2778 (0,0240)	1,2733 (0,2069)	0,1960 (0,0667)
	0,80	2,00	2,00	6,10
EMVb	0,7810 (0,0860)	1,8656 (0,3179)	2,6849 (1,2078)	4,4634 (1,1858)
EMVf	0,7874 (0,0794)	1,9413 (0,3659)	2,7924 (1,2692)	4,5992 (1,2509)
SiZer	0,9206 (0,0137)	1,3502 (0,1701)	3,8183 (0,2557)	1,9040 (0,3631)
EM	1,0000 (0,0000)	51,256 (313,81)	40,055 (155,41)	75,965 (486,64)
	0,80	2,10	2,00	5,00
EMVb	0,7499 (0,1223)	1,9677 (0,4467)	2,4614 (1,3892)	4,2020 (1,3642)
EMVf	0,7605 (0,1066)	2,0266 (0,5024)	2,5584 (1,4565)	4,2644 (1,4515)
SiZer	0,9503 (0,0127)	1,4963 (0,1348)	3,8624 (0,2793)	1,6404 (0,3616)
EM	0,9561 (0,2061)	185,48 (634,50)	209,04 (885,25)	231,05 (1149,6)
	0,80	1,10	4,00	1,30
EMVb	0,7141 (0,1205)	1,1598 (0,1461)	2,6140 (1,7500)	2,8494 (1,5864)
EMVf	0,7344 (0,1083)	1,1943 (0,1602)	2,7224 (1,7445)	2,8895 (1,5148)
SiZer	0,9251 (0,0140)	1,3647 (0,1789)	3,6209 (0,2814)	1,8428 (0,4517)
EM	0,9265 (0,2616)	76,692 (122,29)	241,29 (722,43)	37,793 (296,03)
	0,80	1,10	1,20	4,30
EMVb	0,7725 (0,0913)	1,1013 (0,1609)	1,8982 (1,3907)	3,4712 (1,4609)
EMVf	0,8016 (0,0796)	1,1475 (0,1801)	2,0780 (1,3751)	3,6483 (1,5349)
SiZer	0,9324 (0,0309)	1,4312 (0,2802)	3,4030 (0,2527)	1,7186 (0,5970)
EM	0,9360 (0,2456)	96,895 (352,38)	451,04 (1744,9)	150,67 (424,83)
	0,80	1,10	4,20	4,30
EMVb	0,7991 (0,0074)	1,1031 (0,1118)	4,4348 (0,8366)	4,2231 (1,0390)
EMVf	0,8005 (0,0095)	1,1077 (0,0885)	4,4542 (0,8810)	4,2836 (1,0324)
SiZer	0,8176 (0,0130)	1,1218 (0,0497)	5,0195 (0,4035)	2,9626 (0,3415)
EM	0,9384 (0,2368)	42,113 (186,64)	142,09 (721,43)	76,359 (444,74)
	0,80	0,40	5,00	2,50
EMVb	0,5951 (0,1553)	0,3528 (0,0671)	1,6626 (1,8847)	2,5193 (1,6306)
EMVf	0,5121 (0,2201)	0,3554 (0,0536)	1,8207 (2,0122)	2,4323 (1,6283)
SiZer	0,8079 (0,0127)	1,0173 (0,0044)	13,527 (0,8099)	2,3756 (0,3330)
EM	0,9618 (0,1918)	9,8199 (13,659)	173,50 (916,92)	83,602 (648,08)
	0,80	0,60	1,20	4,30
EMVb	0,7520 (0,1137)	0,5764 (0,0716)	2,2370 (1,6419)	2,5128 (1,5498)
EMVf	0,7516 (0,1221)	0,6011 (0,0435)	2,3183 (1,7808)	2,7902 (1,5731)
SiZer	0,8842 (0,0150)	1,0548 (0,0266)	4,7917 (0,9393)	3,2536 (0,5611)
EM	0,7954 (0,4031)	13,253 (111,74)	63,456 (539,78)	81,280 (580,88)

Tabela A.0.10:  $p=0,80$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,80			1,10			3,00			2,00		
EMVb	(0,6500)	0,8000	(0,8000)	(0,9900)	1,2100	(1,2100)	(0,9900)	1,6500	(3,8500)	(1,6500)	3,5200	(4,9500)
EMVf	(0,7112)	0,7889	(0,8101)	(1,0440)	1,1423	(1,2773)	(1,1073)	1,8862	(4,1329)	(1,7001)	3,5371	(4,8369)
SiZer	(0,9046)	0,9146	(0,9241)	(1,2360)	1,2860	(1,3460)	(3,5316)	3,6058	(3,6865)	(1,8100)	1,9100	(2,1100)
EM	(1,0000)	1,0000	(1,0000)	(0,1775)	2,0261	(5,2435)	(0,0839)	2,5137	(6,5783)	(0,0349)	4,6700	(8,1075)
	0,80			1,10			5,00			2,00		
EMVb	(0,8000)	0,8000	(0,8000)	(0,9900)	1,2100	(1,2100)	(2,5300)	3,6300	(5,1700)	(2,0900)	2,9700	(4,7300)
EMVf	(0,7872)	0,7967	(0,8054)	(1,0546)	1,1571	(1,2138)	(2,4760)	3,5879	(4,6861)	(2,0841)	3,0916	(4,8479)
SiZer	(0,8436)	0,8525	(0,8613)	(1,1460)	1,1760	(1,1960)	(3,8447)	3,9690	(4,1156)	(2,2100)	2,4100	(2,5100)
EM	(0,0000)	0,5144	(1,0000)	(0,2302)	0,8273	(0,8924)	(0,8198)	0,8728	(1,7225)	(0,1648)	1,5257	(1,7831)
	0,80			1,10			5,00			5,00		
EMVb	(0,8000)	0,8000	(0,8000)	(0,9900)	1,2100	(1,2100)	(4,9500)	5,3900	(5,3900)	(4,2900)	5,1700	(5,3900)
EMVf	(0,7951)	0,8013	(0,8058)	(1,0495)	1,1198	(1,1742)	(4,9344)	5,2470	(5,4459)	(4,5262)	5,0423	(5,4065)
SiZer	(0,7978)	0,8072	(0,8159)	(1,0860)	1,0960	(1,1260)	(5,7016)	6,0314	(6,4238)	(3,0100)	3,2100	(3,4100)
EM	(1,0000)	1,0000	(1,0000)	(0,1635)	1,1700	(6,4520)	(0,2690)	1,6798	(3,9211)	(0,0014)	1,2191	(5,5810)
	0,80			1,10			2,00			5,00		
EMVb	(0,8000)	0,8000	(0,8000)	(0,9900)	0,9900	(1,2100)	(1,8700)	2,3100	(3,4100)	(2,9700)	4,7300	(5,3900)
EMVf	(0,7929)	0,8036	(0,8142)	(1,0105)	1,0985	(1,1738)	(1,8567)	2,3086	(3,4149)	(2,9953)	4,7926	(5,3328)
SiZer	(0,8647)	0,8742	(0,8840)	(1,1260)	1,1560	(1,1860)	(3,6849)	3,8160	(3,9607)	(2,3100)	2,5100	(2,7100)
EM	(0,0000)	0,5114	(1,0000)	(0,0012)	0,8207	(0,8685)	(0,8165)	0,9042	(2,9250)	(0,0000)	1,2331	(1,4834)
	0,80			2,00			1,90			5,00		
EMVb	(0,0000)	0,7000	(0,8000)	(0,0000)	1,6500	(2,0900)	(0,0000)	1,2100	(2,3100)	(0,0000)	3,4100	(5,1700)
EMVf	(0,2000)	0,7260	(0,8126)	(0,7661)	1,5962	(2,0337)	(0,9998)	1,2857	(2,4218)	(0,9998)	3,3366	(5,2273)
SiZer	(0,9412)	0,9504	(0,9592)	(1,4060)	1,4660	(1,5460)	(3,7806)	3,8599	(3,9305)	(1,5100)	1,6100	(1,7100)
EM	(1,0000)	1,0000	(1,0000)	(0,2609)	0,2772	(0,2937)	(1,1287)	1,2452	(1,4040)	(0,1479)	0,1839	(0,2374)
	0,80			2,00			2,00			6,10		
EMVb	(0,8000)	0,8000	(0,8000)	(1,6500)	1,8700	(2,0900)	(1,9800)	2,3100	(3,1900)	(3,7400)	5,1700	(5,3900)
EMVf	(0,7796)	0,8114	(0,8256)	(1,6743)	1,9282	(2,1782)	(1,9204)	2,5478	(3,3821)	(3,8700)	5,0846	(5,5082)
SiZer	(0,9112)	0,9208	(0,9290)	(1,2960)	1,3360	(1,3760)	(3,7371)	3,8280	(3,9231)	(1,7100)	1,8100	(2,0100)
EM	(1,0000)	1,0000	(1,0000)	(1,8019)	5,0409	(9,2855)	(3,3369)	9,6907	(26,016)	(1,8250)	14,600	(19,138)
	0,80			2,10			2,00			5,00		
EMVb	(0,7500)	0,8000	(0,8000)	(1,6500)	1,8700	(2,3100)	(1,4300)	2,0900	(3,1900)	(3,1900)	4,9500	(5,3900)
EMVf	(0,7472)	0,7941	(0,8279)	(1,6633)	1,9259	(2,2573)	(1,4760)	2,2866	(3,3510)	(3,2046)	4,7168	(5,4056)
SiZer	(0,9418)	0,9510	(0,9592)	(1,4060)	1,4760	(1,5560)	(3,7951)	3,8773	(3,9637)	(1,5100)	1,6100	(1,7100)
EM	(1,0000)	1,0000	(1,0000)	(0,0107)	0,2530	(5,0395)	(0,0001)	0,3917	(1,9861)	(0,0000)	0,0009	(0,6661)
	0,80			1,10			4,00			1,30		
EMVb	(0,6000)	0,7500	(0,8000)	(0,9900)	1,2100	(1,2100)	(0,8800)	1,8700	(4,7300)	(1,2100)	2,7500	(4,2900)
EMVf	(0,6876)	0,7850	(0,8062)	(1,0983)	1,1600	(1,2997)	(0,9693)	2,1356	(4,4665)	(1,4331)	2,8455	(4,2389)
SiZer	(0,9161)	0,9260	(0,9343)	(1,2760)	1,3560	(1,4160)	(3,5674)	3,6621	(3,7267)	(1,6100)	1,7600	(1,9100)
EM	(1,0000)	1,0000	(1,0000)	(0,9851)	14,396	(83,197)	(1,5407)	16,944	(96,905)	(1,6661)	7,0469	(11,956)
	0,80			1,10			1,20			4,30		
EMVb	(0,7500)	0,8000	(0,8500)	(0,9900)	0,9900	(1,2100)	(0,9900)	1,2100	(2,0900)	(2,5300)	3,6300	(4,7300)
EMVf	(0,7839)	0,8339	(0,8537)	(1,0288)	1,1157	(1,2832)	(1,2339)	1,6266	(2,3264)	(2,6359)	3,7390	(4,9283)
SiZer	(0,9290)	0,9410	(0,9485)	(1,2560)	1,3460	(1,4960)	(3,3537)	3,4458	(3,5386)	(1,3100)	1,6100	(1,9100)
EM	(0,9973)	1,0000	(1,0000)	(0,0006)	0,0114	(0,7046)	(0,0000)	0,0023	(0,5818)	(0,0000)	0,0000	(0,1838)
	0,80			1,10			4,20			4,30		
EMVb	(0,8000)	0,8000	(0,8000)	(0,9900)	1,2100	(1,2100)	(3,8500)	4,5100	(5,3900)	(3,1900)	4,2900	(5,3900)
EMVf	(0,7950)	0,8019	(0,8069)	(1,0370)	1,1293	(1,1727)	(3,7257)	4,4216	(5,2658)	(3,3570)	4,4164	(5,2148)
SiZer	(0,8089)	0,8176	(0,8267)	(1,0960)	1,1160	(1,1460)	(4,7399)	4,9447	(5,2085)	(2,7100)	2,9100	(3,1100)
EM	(1,0000)	1,0000	(1,0000)	(0,2266)	0,9024	(5,8899)	(0,8603)	3,6672	(15,867)	(0,6908)	4,1874	(15,832)
	0,80			0,40			5,00			2,50		
EMVb	(0,4500)	0,6000	(0,8000)	(0,3300)	0,3300	(0,3300)	(0,5500)	0,5500	(3,4100)	(1,2100)	1,6500	(3,8500)
EMVf	(0,3445)	0,4445	(0,7759)	(0,3215)	0,3389	(0,3952)	(0,5986)	0,6527	(3,1419)	(1,0553)	2,0741	(3,6821)
SiZer	(0,7996)	0,8078	(0,8157)	(1,0160)	1,0160	(1,0160)	(13,784)	13,785	(13,785)	(2,1100)	2,3100	(2,5100)
EM	(1,0000)	1,0000	(1,0000)	(4,5983)	6,1016	(11,727)	(5,3696)	20,108	(33,615)	(0,0456)	0,1319	(3,5947)
	0,80			0,60			1,20			4,30		
EMVb	(0,7000)	0,8000	(0,8250)	(0,5500)	0,5500	(0,5500)	(0,9900)	1,4300	(3,6300)	(1,2100)	2,2000	(3,7400)
EMVf	(0,7049)	0,8003	(0,8316)	(0,5756)	0,5908	(0,6113)	(0,9593)	1,4062	(3,8911)	(1,2568)	2,6885	(4,1563)
SiZer	(0,8748)	0,8840	(0,8944)	(1,0360)	1,0460	(1,0660)	(4,0400)	4,5638	(5,5534)	(2,9100)	3,2100	(3,5100)
EM	(1,0000)	1,0000	(1,0000)	(0,3605)	0,6705	(4,6028)	(0,9421)	1,5729	(3,4104)	(0,7259)	1,3360	(2,0778)

Tabela A.0.11:  $p=0,85$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,85	1,10	3,00	2,00
EMVb	0,7956 (0,0863)	1,1326 (0,1795)	2,4900 (1,6803)	3,1728 (1,6009)
EMVf	0,8184 (0,0756)	1,1815 (0,1946)	2,6603 (1,6377)	3,3439 (1,6652)
SiZer	0,9320 (0,0231)	1,3907 (0,2133)	3,4353 (0,1976)	1,7090 (0,5351)
EM	1,0000 (0,0000)	20,323 (281,35)	154,73 (857,83)	44,074 (519,46)
	0,85	1,10	5,00	2,00
EMVb	0,8384 (0,0378)	1,1634 (0,1459)	3,6542 (1,4651)	3,4773 (1,4611)
EMVf	0,8446 (0,0331)	1,2084 (0,1659)	3,8119 (1,4801)	3,6136 (1,4616)
SiZer	0,8910 (0,0131)	1,1829 (0,0857)	3,5850 (0,2348)	2,5004 (0,5671)
EM	0,9019 (0,2953)	754,43 (1054,2)	615,35 (903,35)	16,487 (268,63)
	0,85	1,10	5,00	5,00
EMVb	0,8492 (0,0063)	1,1044 (0,1118)	5,0868 (0,5274)	4,8827 (0,6542)
EMVf	0,8486 (0,0083)	1,1090 (0,1028)	5,1457 (0,6357)	4,8973 (0,6189)
SiZer	0,8542 (0,0135)	1,0982 (0,0633)	5,0132 (0,4918)	3,6086 (0,4569)
EM	0,7494 (0,3390)	117,83 (788,03)	124,25 (696,95)	76,724 (638,66)
	0,85	1,10	2,00	5,00
EMVb	0,8409 (0,0446)	1,0916 (0,1539)	2,8230 (1,4129)	4,0594 (1,4316)
EMVf	0,8495 (0,0393)	1,1279 (0,1655)	2,9824 (1,3935)	4,1952 (1,4798)
SiZer	0,9069 (0,0137)	1,1779 (0,1147)	3,4699 (0,2111)	2,6202 (0,7012)
EM	0,7436 (0,4223)	143,92 (554,02)	143,77 (586,06)	12,552 (65,697)
	0,85	2,00	1,90	5,00
EMVb	0,8172 (0,0912)	1,9549 (0,5351)	2,5995 (1,5220)	4,0999 (1,4642)
EMVf	0,8293 (0,0784)	2,0020 (0,4870)	2,6470 (1,4638)	4,2656 (1,6076)
SiZer	0,9413 (0,0432)	1,6831 (0,3464)	3,5568 (0,3200)	1,3274 (0,3343)
EM	1,0000 (0,0000)	45,846 (182,49)	88,629 (482,65)	110,57 (897,40)
	0,85	2,00	2,00	6,10
EMVb	0,8294 (0,0773)	1,8722 (0,4432)	2,8701 (1,4484)	4,3318 (1,3208)
EMVf	0,8382 (0,0682)	1,9507 (0,4340)	2,9640 (1,4108)	4,5611 (1,5012)
SiZer	0,9397 (0,0155)	1,3946 (0,2203)	3,5952 (0,2075)	1,7010 (0,4134)
EM	0,8634 (0,3253)	191,01 (631,04)	173,61 (677,12)	64,280 (362,80)
	0,85	2,10	2,00	5,00
EMVb	0,8077 (0,1074)	2,0636 (0,4850)	2,5542 (1,5075)	4,4220 (1,2707)
EMVf	0,8154 (0,0938)	2,1041 (0,4759)	2,5873 (1,4360)	4,4835 (1,3566)
SiZer	0,9442 (0,0383)	1,6150 (0,2696)	3,6178 (0,2950)	1,3642 (0,2437)
EM	1,0000 (0,0000)	398,19 (749,41)	10,385 (11,938)	70,106 (121,41)
	0,85	1,10	4,00	1,30
EMVb	0,7879 (0,0947)	1,1686 (0,2118)	2,7496 (1,7567)	2,7271 (1,6192)
EMVf	0,8113 (0,0775)	1,2106 (0,2159)	2,9167 (1,7596)	2,8547 (1,6736)
SiZer	0,9340 (0,0308)	1,5253 (0,2736)	3,4576 (0,3053)	1,4984 (0,4940)
EM	1,0000 (0,0000)	0,0000 (0,0000)	0,0013 (0,0000)	0,0000 (0,0000)
	0,85	1,10	1,20	4,30
EMVb	0,8373 (0,0776)	1,1202 (0,2299)	2,1872 (1,5067)	3,3317 (1,6368)
EMVf	0,8546 (0,0600)	1,1415 (0,2206)	2,2337 (1,3962)	3,5089 (1,7255)
SiZer	0,8893 (0,0650)	1,8062 (0,4634)	3,0363 (0,3933)	1,3402 (0,6273)
EM	0,0962 (0,2952)	0,8451 (0,2746)	3,5291 (48,233)	4,6060 (63,947)
	0,85	1,10	4,20	4,30
EMVb	0,8485 (0,0091)	1,1062 (0,1271)	4,4172 (0,9185)	4,2904 (1,0879)
EMVf	0,8493 (0,0107)	1,1090 (0,1210)	4,4998 (1,0059)	4,2664 (1,0582)
SiZer	0,8644 (0,0141)	1,1099 (0,0281)	4,2670 (0,3246)	3,2274 (0,4588)
EM	0,8025 (0,3299)	136,40 (683,30)	132,85 (791,76)	141,51 (810,05)
	0,85	0,40	5,00	2,50
EMVb	0,8105 (0,0919)	0,3621 (0,0778)	3,8887 (1,8156)	2,6158 (1,7251)
EMVf	0,8097 (0,1020)	0,4019 (0,0415)	4,0276 (1,8961)	3,1171 (1,7138)
SiZer	0,0858 (0,2578)	0,1021 (0,3065)	1,2712 (3,8369)	0,2566 (0,7821)
EM	0,9343 (0,2495)	218,76 (680,22)	530,99 (1517,2)	484,53 (1556,4)
	0,85	0,60	1,20	4,30
EMVb	0,7726 (0,1682)	0,5878 (0,0832)	2,1780 (1,5416)	2,3786 (1,6243)
EMVf	0,7826 (0,1472)	0,6052 (0,0514)	2,2022 (1,5840)	2,6141 (1,6478)
SiZer	0,9137 (0,0155)	1,0834 (0,1007)	3,8110 (0,6976)	3,4276 (0,7902)
EM	1,0000 (0,0000)	0,5892 (0,6142)	0,8537 (0,5409)	7,5438 (57,071)

Tabela A.0.12:  $p=0,85$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,85			1,10			3,00			2,00		
EMVb	(0,7500)	0,8500	(0,8500)	(0,9900)	1,2100	(1,2100)	(1,2100)	1,6500	(4,2900)	(1,6500)	3,1900	(4,7300)
EMVf	(0,8118)	0,8468	(0,8616)	(1,0492)	1,1392	(1,3135)	(1,3935)	1,9551	(4,2742)	(1,6305)	3,2936	(4,9482)
SiZer	(0,9271)	0,9356	(0,9437)	(1,2660)	1,3360	(1,4410)	(3,3690)	3,4622	(3,5448)	(1,4100)	1,6100	(1,8100)
EM	(1,0000)	1,0000	(1,0000)	(0,0408)	1,7760	(4,0773)	(2,0015)	3,6476	(5,3431)	(0,0698)	0,8653	(3,0344)
	0,85			1,10			5,00			2,00		
EMVb	(0,8500)	0,8500	(0,8500)	(0,9900)	1,2100	(1,2100)	(2,3100)	3,4100	(5,1700)	(2,0900)	3,1900	(5,1700)
EMVf	(0,8412)	0,8532	(0,8605)	(1,1111)	1,1592	(1,3460)	(2,5066)	3,7873	(5,2582)	(2,2538)	3,2994	(4,9731)
SiZer	(0,8821)	0,8908	(0,8999)	(1,1560)	1,1760	(1,2060)	(3,4696)	3,5727	(3,7093)	(2,1100)	2,4100	(2,6100)
EM	(0,9997)	0,9998	(0,9999)	(590,60)	607,36	(623,50)	(472,82)	486,23	(497,17)	(3,1003)	3,1619	(3,2168)
	0,85			1,10			5,00			5,00		
EMVb	(0,8500)	0,8500	(0,8500)	(0,9900)	1,2100	(1,2100)	(4,9500)	5,3900	(5,3900)	(4,2900)	5,3900	(5,3900)
EMVf	(0,8443)	0,8487	(0,8536)	(1,0632)	1,1058	(1,1448)	(4,7533)	5,1538	(5,6265)	(4,5433)	4,9850	(5,1538)
SiZer	(0,8449)	0,8537	(0,8639)	(1,0760)	1,0860	(1,1060)	(4,7217)	4,9700	(5,2596)	(3,3100)	3,5100	(3,8100)
EM	(0,5165)	1,0000	(1,0000)	(0,6383)	0,6655	(11,644)	(0,9456)	1,0396	(18,633)	(0,3321)	0,8074	(1,0998)
	0,85			1,10			2,00			5,00		
EMVb	(0,8500)	0,8500	(0,8500)	(0,9900)	0,9900	(1,2100)	(1,8700)	2,3100	(4,0700)	(2,7500)	4,7300	(5,3900)
EMVf	(0,8445)	0,8589	(0,8665)	(1,0416)	1,1132	(1,1818)	(1,9743)	2,5458	(4,1260)	(2,8235)	4,7520	(5,4210)
SiZer	(0,8973)	0,9068	(0,9161)	(1,1360)	1,1660	(1,1960)	(3,3528)	3,4548	(3,5626)	(2,1100)	2,4100	(2,9100)
EM	(0,5059)	1,0000	(1,0000)	(0,8150)	1,3254	(17,566)	(0,9191)	2,0731	(20,566)	(0,4402)	1,1365	(3,2631)
	0,85			2,00			1,90			5,00		
EMVb	(0,8000)	0,8500	(0,8500)	(1,6500)	1,8700	(2,0900)	(1,4300)	2,0900	(3,7400)	(3,1900)	4,8400	(5,3900)
EMVf	(0,8217)	0,8578	(0,8680)	(1,6209)	1,9247	(2,2999)	(1,5638)	2,2267	(3,4771)	(3,1738)	4,8614	(5,5536)
SiZer	(0,9463)	0,9572	(0,9652)	(1,4460)	1,5660	(1,7910)	(3,5394)	3,6595	(3,7438)	(1,1100)	1,3100	(1,5100)
EM	(1,0000)	1,0000	(1,0000)	(1,1487)	4,1626	(12,580)	(1,2325)	5,0980	(9,4504)	(2,8192)	6,2278	(12,916)
	0,85			2,00			2,00			6,10		
EMVb	(0,8500)	0,8500	(0,8500)	(1,6500)	1,8700	(2,0900)	(1,8700)	2,5300	(4,0700)	(3,4100)	4,9500	(5,3900)
EMVf	(0,8424)	0,8579	(0,8681)	(1,6488)	1,9241	(2,2459)	(2,0329)	2,5889	(4,0515)	(3,3920)	5,2157	(5,7092)
SiZer	(0,9302)	0,9400	(0,9503)	(1,3060)	1,3560	(1,4260)	(3,5249)	3,6007	(3,6857)	(1,5100)	1,7100	(1,8100)
EM	(0,9788)	1,0000	(1,0000)	(0,0050)	0,7220	(45,377)	(0,0000)	0,0876	(1,4396)	(0,0000)	0,0005	(1,5917)
	0,85			2,10			2,00			5,00		
EMVb	(0,8000)	0,8500	(0,8500)	(1,6500)	2,0900	(2,3100)	(1,4300)	2,3100	(3,6300)	(3,8500)	5,1700	(5,3900)
EMVf	(0,8077)	0,8535	(0,8659)	(1,6769)	2,0376	(2,4087)	(1,4859)	2,3659	(3,2826)	(3,6701)	5,0846	(5,5221)
SiZer	(0,9401)	0,9530	(0,9631)	(1,4260)	1,5460	(1,6660)	(3,5791)	3,6627	(3,7805)	(1,2100)	1,3600	(1,5100)
EM	(1,0000)	1,0000	(1,0000)	(1,5505)	20,713	(446,94)	(2,5807)	6,1152	(12,019)	(0,3996)	5,9796	(77,325)
	0,85			1,10			4,00			1,30		
EMVb	(0,7500)	0,8500	(0,8500)	(0,9900)	1,2100	(1,2100)	(0,9900)	2,0900	(4,7300)	(1,2100)	2,5300	(4,0700)
EMVf	(0,8025)	0,8419	(0,8557)	(1,0587)	1,1525	(1,3129)	(1,3744)	2,3262	(4,9534)	(1,1955)	2,7504	(4,3237)
SiZer	(0,9327)	0,9415	(0,9492)	(1,3360)	1,4560	(1,6260)	(3,4268)	3,5196	(3,6067)	(1,2100)	1,4100	(1,6100)
EM	(1,0000)	1,0000	(1,0000)	(0,0000)	0,0000	(0,0000)	(0,0013)	0,0013	(0,0013)	(0,0000)	0,0000	(0,0000)
	0,85			1,10			1,20			4,30		
EMVb	(0,8000)	0,8500	(0,9000)	(0,9900)	0,9900	(1,2100)	(1,2100)	1,4300	(2,6400)	(1,6500)	3,6300	(4,9500)
EMVf	(0,8462)	0,8651	(0,8920)	(1,0104)	1,1020	(1,2660)	(1,2352)	1,6300	(2,7332)	(1,9566)	3,7394	(5,0402)
SiZer	(0,8335)	0,9284	(0,9507)	(1,3260)	1,8410	(2,2460)	(2,7859)	3,0756	(3,3487)	(1,0100)	1,0100	(1,4600)
EM	(0,0000)	0,0000	(0,0000)	(0,8950)	0,9230	(0,9469)	(0,9357)	0,9620	(0,9904)	(1,0822)	1,1678	(1,2707)
	0,85			1,10			4,20			4,30		
EMVb	(0,8500)	0,8500	(0,8500)	(0,9900)	0,9900	(1,2100)	(3,8500)	4,5100	(5,3900)	(3,1900)	4,7300	(5,3900)
EMVf	(0,8434)	0,8499	(0,8560)	(1,0530)	1,1002	(1,1506)	(3,6856)	4,5964	(5,4322)	(3,2116)	4,5861	(5,1535)
SiZer	(0,8541)	0,8643	(0,8745)	(1,0860)	1,1060	(1,1260)	(4,0347)	4,2223	(4,4478)	(2,9100)	3,1100	(3,4100)
EM	(0,5132)	1,0000	(1,0000)	(0,6915)	0,7495	(5,7045)	(0,9197)	1,2763	(10,090)	(0,4921)	1,0402	(4,2197)
	0,85			0,40			5,00			2,50		
EMVb	(0,8500)	0,8500	(0,8500)	(0,3300)	0,3300	(0,3300)	(1,8700)	4,9500	(5,3900)	(1,4300)	1,6500	(4,6200)
EMVf	(0,8298)	0,8477	(0,8547)	(0,3784)	0,3938	(0,4084)	(1,8497)	4,7139	(5,6420)	(1,9574)	2,3974	(5,0697)
SiZer	(0,0000)	0,0000	(0,0000)	(0,0000)	0,0000	(0,0000)	(0,0000)	0,0000	(0,0000)	(0,0000)	0,0000	(0,0000)
EM	(1,0000)	1,0000	(1,0000)	(0,0013)	0,0018	(0,0033)	(0,0000)	0,0000	(215,30)	(0,0000)	0,0000	(21,829)
	0,85			0,60			1,20			4,30		
EMVb	(0,7500)	0,8500	(0,8500)	(0,5500)	0,5500	(0,5500)	(0,9900)	1,4300	(3,1900)	(0,9900)	1,8700	(3,6300)
EMVf	(0,7400)	0,8546	(0,8681)	(0,5724)	0,5914	(0,6193)	(1,0237)	1,5449	(2,8820)	(1,0375)	2,3963	(3,9482)
SiZer	(0,9035)	0,9140	(0,9238)	(1,0460)	1,0660	(1,0960)	(3,2869)	3,6053	(4,1625)	(2,9100)	3,4100	(4,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,2672)	0,4858	(0,6890)	(0,4903)	0,7513	(1,0348)	(0,0206)	0,1531	(0,5738)

Tabela A.0.13:  $p=0,90$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,90	1,10	3,00	2,00
EMVb	0,8244 (0,1067)	1,1590 (0,2896)	2,1501 (1,6884)	3,3167 (1,5536)
EMVf	0,6248 (0,1045)	0,2236 (0,2456)	1,2998 (1,5775)	2,3654 (1,5167)
SiZer	0,2758 (0,0079)	1,2800 (0,1008)	20,999 (1,1154)	1,2935 (0,1397)
EM	0,9698 (0,1467)	65,621 (316,47)	684,46 (1983,6)	510,0 (1854,2)
	0,90	1,10	5,00	2,00
EMVb	0,8717 (0,0636)	1,1581 (0,2080)	3,2419 (1,6968)	3,7594 (1,4895)
EMVf	0,8801 (0,0546)	1,1893 (0,2175)	3,3886 (1,6772)	3,9046 (1,6068)
SiZer	0,9208 (0,0317)	1,2776 (0,2346)	3,2314 (0,2254)	2,1318 (0,8549)
EM	1,0000 (0,0000)	631,85 (24,612)	465,23 (16,062)	1,6778 (0,0399)
	0,90	1,10	5,00	5,00
EMVb	0,8989 (0,0073)	1,0960 (0,1249)	4,9909 (0,6912)	4,9223 (0,6408)
EMVf	0,9009 (0,0093)	1,1416 (0,1279)	5,1419 (0,7527)	5,3317 (0,7354)
SiZer	0,9031 (0,0134)	1,0861 (0,0219)	3,9766 (0,3364)	3,9006 (0,5884)
EM	1,0000 (0,0000)	804,09 (40,610)	385,46 (14,438)	3,7777 (0,0762)
	0,90	1,10	2,00	5,00
EMVb	0,8712 (0,0689)	1,0934 (0,1930)	2,5975 (1,5724)	4,1408 (1,4662)
EMVf	0,8812 (0,0593)	1,1309 (0,2084)	2,7480 (1,5534)	4,3744 (1,5801)
SiZer	0,9191 (0,0492)	1,3679 (0,3514)	3,1185 (0,3084)	1,9996 (0,9503)
EM	1,0000 (0,0000)	517,44 (22,160)	568,34 (25,483)	1,6548 (0,2342)
	0,90	2,00	1,90	5,00
EMVb	0,8312 (0,1088)	1,9360 (0,8156)	2,3989 (1,7343)	4,0823 (1,4757)
EMVf	0,8446 (0,0956)	1,9026 (0,7789)	2,3614 (1,6473)	4,0913 (1,4917)
SiZer	0,8681 (0,0604)	2,1232 (0,3862)	2,9473 (0,3861)	1,0840 (0,2840)
EM	1,0000 (0,0000)	0,1098 (0,2505)	5759,7 (1885,2)	0,0050 (0,0275)
	0,90	2,00	2,00	6,10
EMVb	0,8495 (0,0967)	1,8157 (0,5886)	2,6492 (1,6716)	4,2722 (1,3701)
EMVf	0,8568 (0,0881)	1,7783 (0,5596)	2,5618 (1,5806)	4,3384 (1,4394)
SiZer	0,8955 (0,0682)	1,8196 (0,4572)	3,0877 (0,4332)	1,2418 (0,4552)
EM	0,7707 (0,4098)	220,02 (955,01)	435,92 (1476,8)	534,24 (1719,0)
	0,90	2,10	2,00	5,00
EMVb	0,8354 (0,0995)	2,0238 (0,8206)	2,3822 (1,6709)	4,2056 (1,4438)
EMVf	0,8475 (0,0884)	1,9869 (0,7667)	2,3348 (1,5945)	4,2625 (1,4926)
SiZer	0,8709 (0,0612)	2,1171 (0,3841)	2,9656 (0,4017)	1,0849 (0,3133)
EM	0,5528 (0,4959)	155,01 (865,85)	407,71 (1639,2)	234,48 (1096,93)
	0,90	1,10	4,00	1,30
EMVb	0,8167 (0,1172)	1,2100 (0,3360)	2,2486 (1,7428)	3,1339 (1,6166)
EMVf	0,8382 (0,0989)	1,2232 (0,3326)	2,2535 (1,5950)	3,2646 (1,6856)
SiZer	0,8680 (0,0644)	1,9927 (0,4246)	2,9193 (0,3825)	1,1564 (0,4414)
EM	0,7505 (0,4254)	9,8778 (32,190)	133,53 (790,07)	17,508 (225,96)
	0,90	1,10	1,20	4,30
EMVb	0,8434 (0,1167)	1,1675 (0,4485)	2,0112 (1,6095)	3,3176 (1,6161)
EMVf	0,8627 (0,0948)	1,1742 (0,4454)	2,0528 (1,5261)	3,3878 (1,6259)
SiZer	0,8396 (0,0556)	2,1731 (0,3706)	2,7081 (0,2853)	1,1001 (0,4348)
EM	0,6667 (0,5774)	0,1848 (0,3199)	232,92 (400,26)	0,0000 (0,0000)
	0,90	1,10	4,20	4,30
EMVb	0,8974 (0,0139)	1,1205 (0,1546)	4,3450 (1,0705)	4,3872 (1,0921)
EMVf	0,9006 (0,0150)	1,1620 (0,1606)	4,4033 (1,0756)	4,7837 (1,1933)
SiZer	0,9105 (0,0137)	1,1215 (0,1447)	3,5348 (0,2993)	3,4054 (0,7242)
EM	1,0000 (0,0000)	702,36 (36,093)	427,68 (17,449)	2,6446 (0,0566)
	0,90	0,40	5,00	2,50
EMVb	0,8405 (0,1132)	0,3773 (0,0905)	3,1255 (2,0096)	2,8065 (1,8988)
EMVf	0,8316 (0,1170)	0,4038 (0,0466)	3,2063 (2,1859)	3,1831 (1,9154)
SiZer	0,9034 (0,0143)	1,0188 (0,0064)	8,2537 (2,3539)	3,8388 (0,9498)
EM	1,0000 (0,0000)	153,33 (242,49)	77,339 (121,71)	1,0524 (1,3181)
	0,90	0,60	1,20	4,30
EMVb	0,8043 (0,1726)	0,6046 (0,0961)	1,8256 (1,5238)	2,5564 (1,7871)
EMVf	0,8091 (0,1405)	0,6140 (0,0684)	1,7514 (1,5522)	2,7837 (1,9194)
SiZer	0,9023 (0,0676)	1,4227 (0,4710)	3,0202 (0,5097)	2,3684 (1,2839)
EM	1,0000 (0,0000)	2014,5 (92,376)	232,74 (4,3433)	6,6147 (0,0947)

Tabela A.0.14:  $p=0,90$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,90			1,10			3,00			2,00		
EMVb	(0,7500)	0,8500	(0,9000)	(0,9900)	0,9900	(1,2100)	(0,7700)	1,4300	(3,1900)	(2,0900)	3,1900	(4,9500)
EMVf	(0,5500)	0,6500	(0,7000)	(0,0780)	0,1149	(0,2101)	(0,1647)	0,4301	(2,1901)	(1,0901)	2,2033	(3,9501)
SiZer	(0,2703)	0,2755	(0,2811)	(1,2060)	1,2610	(1,3360)	(20,342)	21,119	(21,736)	(1,2100)	1,2100	(1,3100)
EM	(1,0000)	1,0000	(1,0000)	(0,0183)	0,2303	(1,4989)	(0,0021)	0,6022	(26,449)	(0,0000)	0,0031	(27,409)
EMVb	(0,8500)	0,9000	(0,9000)	(0,9900)	1,2100	(1,2100)	(1,8700)	2,9700	(5,1700)	(2,3100)	4,0700	(5,3900)
EMVf	(0,8822)	0,8972	(0,9047)	(1,0348)	1,1727	(1,2963)	(1,9725)	3,1731	(5,1544)	(2,3264)	4,2318	(5,4058)
SiZer	(0,9162)	0,9275	(0,9361)	(1,1760)	1,2160	(1,2760)	(3,1701)	3,2542	(3,3468)	(1,5600)	1,9100	(2,5100)
EM	(1,0000)	1,0000	(1,0000)	(618,47)	628,34	(642,73)	(457,93)	466,60	(472,80)	(1,6508)	1,6799	(1,7055)
EMVb	(0,9000)	0,9000	(0,9000)	(0,9900)	0,9900	(1,2100)	(4,7300)	5,3900	(5,3900)	(4,5100)	5,3900	(5,3900)
EMVf	(0,8969)	0,9018	(0,9063)	(1,0451)	1,1201	(1,2342)	(4,8618)	5,2655	(5,6265)	(4,8618)	5,5500	(5,9180)
SiZer	(0,8950)	0,9034	(0,9113)	(1,0660)	1,0860	(1,0960)	(3,7186)	3,9275	(4,1759)	(3,5100)	3,8100	(4,4100)
EM	(1,0000)	1,0000	(1,0000)	(763,44)	811,40	(839,71)	(373,50)	381,36	(399,87)	(3,7242)	3,7757	(3,8265)
EMVb	(0,8500)	0,9000	(0,9000)	(0,9900)	0,9900	(1,2100)	(1,4300)	2,0900	(3,6300)	(2,9700)	4,9500	(5,3900)
EMVf	(0,8742)	0,9004	(0,9143)	(0,9959)	1,0903	(1,2660)	(1,5693)	2,4166	(3,8100)	(3,0469)	4,9733	(5,6262)
SiZer	(0,9236)	0,9354	(0,9448)	(1,1660)	1,2360	(1,3460)	(3,0906)	3,1935	(3,2836)	(1,3100)	1,7100	(2,5100)
EM	(0,9999)	1,0000	(1,0000)	(494,45)	529,17	(536,99)	(546,11)	553,34	(594,87)	(1,4401)	1,5485	(1,8928)
EMVb	(0,7500)	0,9000	(0,9000)	(1,4300)	1,6500	(2,3100)	(0,7700)	1,6500	(3,8500)	(3,1900)	4,7300	(5,3900)
EMVf	(0,7976)	0,8889	(0,9048)	(1,4056)	1,6790	(2,1149)	(0,9310)	1,6510	(3,5423)	(3,1736)	4,6418	(5,2652)
SiZer	(0,8373)	0,8412	(0,9483)	(1,7960)	2,3260	(2,3760)	(2,7703)	2,8045	(3,2695)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,0397)	0,0731	(0,1263)	(4324,4)	5637,7	(7088,8)	(0,0003)	0,0009	(0,0027)
EMVb	(0,8500)	0,9000	(0,9000)	(1,4300)	1,6500	(2,0900)	(1,2100)	2,3100	(4,0700)	(3,1900)	4,9500	(5,3900)
EMVf	(0,8573)	0,8936	(0,9054)	(1,4164)	1,6668	(2,0057)	(1,3297)	2,0735	(3,9253)	(3,2783)	4,9184	(5,4453)
SiZer	(0,8350)	0,9396	(0,9578)	(1,3860)	1,6310	(2,2760)	(2,7774)	3,2545	(3,4604)	(1,0100)	1,0100	(1,4100)
EM	(0,9940)	1,0000	(1,0000)	(0,1847)	1,3517	(21,022)	(0,2152)	1,1037	(21,129)	(0,0779)	1,3537	(7,9476)
EMVb	(0,7500)	0,9000	(0,9000)	(1,4300)	1,6500	(2,3100)	(0,7700)	1,8700	(3,4100)	(3,1900)	4,9500	(5,3900)
EMVf	(0,8052)	0,8891	(0,9044)	(1,4658)	1,7550	(2,2569)	(0,9828)	1,7699	(3,2453)	(3,3246)	4,8941	(5,4057)
SiZer	(0,8378)	0,8417	(0,9520)	(1,7710)	2,3260	(2,3760)	(2,7783)	2,8093	(3,3578)	(1,0100)	1,0100	(1,0100)
EM	(0,0000)	1,0000	(1,0000)	(1,0873)	2,7167	(25,184)	(0,0000)	0,0908	(6,2438)	(0,0001)	0,1023	(1,4484)
EMVb	(0,7500)	0,8500	(0,9000)	(0,9900)	1,2100	(1,2100)	(0,7700)	1,4300	(3,8500)	(1,6500)	2,9700	(4,7300)
EMVf	(0,8152)	0,8692	(0,9004)	(1,0116)	1,1574	(1,3580)	(0,9400)	1,5130	(3,5423)	(1,7452)	3,2059	(4,8124)
SiZer	(0,8327)	0,8377	(0,9465)	(1,5360)	2,2160	(2,3060)	(2,7610)	2,7985	(3,2931)	(1,0100)	1,0100	(1,1100)
EM	(0,5014)	1,0000	(1,0000)	(0,0784)	0,9527	(6,3584)	(0,2664)	0,9519	(9,4884)	(0,0454)	0,4819	(1,1866)
EMVb	(0,8000)	0,9000	(0,9000)	(0,9900)	0,9900	(1,2100)	(0,7700)	1,2100	(2,7500)	(2,0900)	3,1900	(4,9500)
EMVf	(0,8239)	0,8938	(0,9225)	(0,9483)	1,0758	(1,2650)	(0,8146)	1,4291	(2,6620)	(2,1065)	3,2805	(4,9871)
SiZer	(0,7933)	0,8368	(0,8420)	(2,1960)	2,2660	(2,3860)	(2,4517)	2,7508	(2,7903)	(1,0100)	1,0100	(1,0100)
EM	(0,2500)	1,0000	(1,0000)	(0,0001)	0,0003	(0,4157)	(0,9152)	3,6610	(522,241)	(0,0000)	0,0000	(0,0000)
EMVb	(0,9000)	0,9000	(0,9000)	(0,9900)	1,2100	(1,2100)	(3,6300)	4,7300	(5,3900)	(3,4100)	4,9500	(5,3900)
EMVf	(0,8966)	0,9020	(0,9074)	(1,0395)	1,1617	(1,2679)	(3,5747)	4,7144	(5,2655)	(3,6786)	5,1100	(5,9180)
SiZer	(0,9005)	0,9111	(0,9196)	(1,0960)	1,1060	(1,1260)	(3,3662)	3,5186	(3,6907)	(2,8100)	3,3100	(3,9100)
EM	(1,0000)	1,0000	(1,0000)	(670,36)	688,95	(738,09)	(410,51)	432,80	(443,45)	(2,6105)	2,6463	(2,6795)
EMVb	(0,8000)	0,9000	(0,9000)	(0,3300)	0,3300	(0,3300)	(0,9900)	3,6300	(5,1700)	(1,2100)	2,0900	(5,1700)
EMVf	(0,7888)	0,8959	(0,9036)	(0,3685)	0,3936	(0,4284)	(0,8088)	3,1341	(5,5062)	(1,6656)	2,6481	(5,2619)
SiZer	(0,8939)	0,9039	(0,9134)	(1,0160)	1,0160	(1,0160)	(7,3842)	7,3846	(7,9067)	(3,1600)	4,0100	(4,7100)
EM	(1,0000)	1,0000	(1,0000)	(0,6114)	0,6667	(494,51)	(1,5099)	1,7271	(229,21)	(0,2011)	0,2702	(2,4663)
EMVb	(0,8000)	0,8500	(0,9000)	(0,5500)	0,5500	(0,5500)	(0,7700)	1,2100	(1,8700)	(0,9900)	1,8700	(4,4000)
EMVf	(0,7619)	0,8667	(0,8941)	(0,5686)	0,5956	(0,6462)	(0,7623)	1,0522	(1,6392)	(0,9605)	2,1881	(4,7325)
SiZer	(0,8310)	0,9374	(0,9473)	(1,0860)	1,1560	(2,1060)	(2,7489)	3,0335	(3,1840)	(1,0100)	2,2100	(3,4100)
EM	(1,0000)	1,0000	(1,0000)	(1950,6)	2013,7	(2082,2)	(229,36)	232,42	(235,74)	(6,5598)	6,6143	(6,6786)

Tabela A.0.15:  $p=0,95$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,95	1,10	3,00	2,00
EMVb	0,8869 (0,0832)	1,1893 (0,4929)	2,0843 (1,6291)	3,1799 (1,7927)
EMVf	0,9032 (0,0633)	1,1863 (0,4669)	2,0822 (1,6076)	3,2916 (1,8523)
SiZer	0,8256 (0,0406)	2,3282 (0,2689)	2,5931 (0,2080)	1,0848 (0,4273)
EM	1,0000 (0,0000)	509,40 (32,076)	502,00 (63,703)	1,4915 (0,2381)
	0,95	1,10	5,00	2,00
EMVb	0,8993 (0,0968)	1,2448 (0,4674)	2,9858 (1,8661)	3,7765 (1,6462)
EMVf	0,9097 (0,0778)	1,2342 (0,4694)	2,9663 (1,8352)	3,8575 (1,6988)
SiZer	0,8401 (0,0685)	2,0578 (0,4423)	2,6326 (0,3246)	1,1162 (0,5101)
EM	1,0000 (0,0000)	267,57 (19,438)	348,63 (34,896)	0,6205 (0,0344)
	0,95	1,10	5,00	5,00
EMVb	0,9229 (0,1284)	1,0974 (0,1892)	4,5707 (1,3729)	4,8998 (1,0072)
EMVf	0,9260 (0,1104)	1,0991 (0,1941)	4,5360 (1,3643)	4,9938 (1,0775)
SiZer	0,9151 (0,0688)	1,3894 (0,4917)	2,9388 (0,3948)	2,6052 (1,4575)
EM	1,0000 (0,0000)	1389,5 (60,080)	275,32 (6,0179)	6,3384 (0,0952)
	0,95	1,10	2,00	5,00
EMVb	0,8928 (0,1003)	1,1312 (0,3728)	2,4600 (1,6929)	3,6837 (1,7620)
EMVf	0,9059 (0,0785)	1,1158 (0,3655)	2,4505 (1,6803)	3,7771 (1,8090)
SiZer	0,8311 (0,0596)	2,1625 (0,4179)	2,5898 (0,2653)	1,1564 (0,6035)
EM	1,0000 (0,0000)	3,2790 (19,418)	485,31 (1374,6)	53,929 (575,09)
	0,95	2,00	1,90	5,00
EMVb	0,8813 (0,0978)	1,8876 (1,0745)	2,2972 (1,7476)	3,5218 (1,8129)
EMVf	0,8975 (0,0793)	1,8888 (1,0677)	2,3331 (1,7357)	3,6304 (1,9099)
SiZer	0,8333 (0,0313)	2,4272 (0,2197)	2,6560 (0,1946)	1,0260 (0,2049)
EM	1,0000 (0,0000)	1,5016 (2,2012)	3350,6 (3851,7)	0,0338 (0,1027)
	0,95	2,00	2,00	6,10
EMVb	0,7570 (0,3302)	1,5407 (1,0423)	2,2209 (1,9136)	3,2188 (2,0682)
EMVf	0,7985 (0,2619)	1,6685 (0,8740)	2,3971 (1,7583)	3,4278 (1,9421)
SiZer	0,8329 (0,0466)	2,3203 (0,3035)	2,6288 (0,3051)	1,0664 (0,4685)
EM	1,0000 (0,0000)	0,5558 (0,0542)	1,0662 (0,1457)	0,5071 (0,1566)
	0,95	2,10	2,00	5,00
EMVb	0,8747 (0,1001)	1,8225 (0,9772)	2,2497 (1,7405)	3,4984 (1,7617)
EMVf	0,8920 (0,0830)	1,8256 (0,9552)	2,2703 (1,7201)	3,6180 (1,8337)
SiZer	0,8350 (0,0340)	2,4182 (0,2014)	2,6512 (0,2327)	1,0384 (0,3467)
EM	1,0000 (0,0000)	516,94 (37,352)	393,41 (70,236)	1,5515 (0,1058)
	0,95	1,10	4,00	1,30
EMVb	0,8763 (0,1075)	1,3468 (0,7167)	2,1868 (1,7261)	3,1284 (1,7732)
EMVf	0,8927 (0,0852)	1,3313 (0,6989)	2,2128 (1,7390)	3,1996 (1,8544)
SiZer	0,8290 (0,0406)	2,3413 (0,2509)	2,6259 (0,2454)	1,0334 (0,3204)
EM	1,0000 (0,0000)	8,5612 (10,018)	1177,8 (248,27)	70,780 (423,47)
	0,95	1,10	1,20	4,30
EMVb	0,8778 (0,0899)	1,1801 (0,6051)	1,8973 (1,4399)	2,7047 (1,9273)
EMVf	0,8954 (0,0691)	1,1699 (0,5918)	1,8975 (1,4611)	2,8307 (2,0299)
SiZer	0,8351 (0,0409)	2,3997 (0,2657)	2,6053 (0,2579)	1,0630 (0,4623)
EM	1,0000 (0,0000)	4,3544 (10,839)	2,5224 (12,569)	1093,8 (2183,5)
	0,95	1,10	4,20	4,30
EMVb	0,9224 (0,0956)	1,1163 (0,2546)	4,0286 (1,6652)	4,2900 (1,3478)
EMVf	0,9256 (0,0809)	1,1053 (0,2504)	3,9287 (1,6300)	4,3345 (1,3894)
SiZer	0,8836 (0,0798)	1,6614 (0,5534)	2,7641 (0,3425)	1,6840 (1,1201)
EM	1,0000 (0,0000)	1776,8 (177,98)	203,86 (4,1781)	7,2581 (0,1483)
	0,95	0,40	5,00	2,50
EMVb	0,8707 (0,1294)	0,4088 (0,1056)	2,7293 (1,9618)	2,5120 (2,0663)
EMVf	0,8466 (0,1284)	0,4286 (0,0690)	2,4449 (1,8506)	2,4095 (1,9340)
SiZer	0,9353 (0,0515)	1,1616 (0,3563)	4,3145 (1,6002)	3,7032 (1,3566)
EM	1,0000 (0,0000)	368,45 (196,70)	255,88 (73,976)	3,3273 (2,7868)
	0,95	0,60	1,20	4,30
EMVb	0,8191 (0,1688)	0,6481 (0,1275)	1,9633 (1,4931)	1,6685 (1,7046)
EMVf	0,8388 (0,1372)	0,6472 (0,1068)	1,9082 (1,5031)	1,7567 (1,7038)
SiZer	0,8254 (0,0531)	2,2404 (0,3714)	2,5473 (0,2630)	1,2054 (0,7585)
EM	0,1594 (0,3687)	0,8558 (0,7890)	18,836 (142,91)	2,9351 (13,150)

Tabela A.0.16:  $p=0,95$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,95			1,10			3,00			2,00		
EMVb	(0,8500)	0,9000	(0,9500)	(0,9900)	0,9900	(1,2100)	(0,8800)	1,2100	(2,5300)	(1,4300)	3,4100	(5,1700)
EMVf	(0,8833)	0,9280	(0,9481)	(0,9265)	1,0694	(1,3448)	(0,9715)	1,4384	(2,2142)	(1,5571)	3,2456	(5,1859)
SiZer	(0,7949)	0,8370	(0,8425)	(2,2660)	2,3260	(2,4960)	(2,4251)	2,6961	(2,7403)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(489,95)	513,86	(529,03)	(456,84)	520,36	(548,22)	(1,3406)	1,4248	(1,5312)
	0,95			1,10			5,00			2,00		
EMVb	(0,9000)	0,9500	(0,9500)	(0,9900)	1,2100	(1,4300)	(0,9900)	2,7500	(5,1700)	(2,3100)	4,2900	(5,3900)
EMVf	(0,8999)	0,9442	(0,9515)	(0,9469)	1,1234	(1,3950)	(1,2938)	2,6222	(4,9658)	(2,3831)	4,3061	(5,4453)
SiZer	(0,7913)	0,7949	(0,8437)	(2,1760)	2,2360	(2,2860)	(2,4087)	2,4415	(2,7767)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(253,37)	264,86	(279,25)	(324,65)	344,09	(367,98)	(0,5965)	0,6142	(0,6352)
	0,95			1,10			5,00			5,00		
EMVb	(0,9500)	0,9500	(0,9500)	(0,9900)	0,9900	(1,2100)	(4,2900)	5,3900	(5,3900)	(4,9500)	5,3900	(5,3900)
EMVf	(0,9430)	0,9482	(0,9523)	(0,9712)	1,1010	(1,2251)	(4,0333)	5,1538	(5,4456)	(4,8257)	5,4059	(5,6256)
SiZer	(0,9246)	0,9463	(0,9582)	(1,0860)	1,1260	(1,2960)	(2,7745)	2,9966	(3,1562)	(1,0100)	2,5100	(4,0100)
EM	(1,0000)	1,0000	(1,0000)	(1348,8)	1387,5	(1428,5)	(271,27)	275,26	(279,15)	(6,2768)	6,3389	(6,4020)
	0,95			1,10			2,00			5,00		
EMVb	(0,8750)	0,9500	(0,9500)	(0,7700)	0,9900	(1,2100)	(0,9900)	1,8700	(3,8500)	(2,0900)	4,2900	(5,3900)
EMVf	(0,8833)	0,9425	(0,9513)	(0,8602)	1,0233	(1,2348)	(1,0570)	1,8888	(3,6458)	(2,1945)	4,3160	(5,4456)
SiZer	(0,7924)	0,7956	(0,8416)	(2,2160)	2,2660	(2,3810)	(2,4115)	2,4435	(2,7423)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,4463)	0,8195	(1,5372)	(0,5918)	4,3772	(73,155)	(0,0001)	0,0091	(0,0899)
	0,95			2,00			1,90			5,00		
EMVb	(0,8500)	0,9000	(0,9500)	(1,2100)	1,4300	(2,3100)	(0,9900)	1,2100	(3,7400)	(1,8700)	3,7400	(5,3900)
EMVf	(0,8725)	0,9271	(0,9514)	(1,1521)	1,5579	(2,3204)	(0,9557)	1,4462	(3,6679)	(1,8856)	3,8146	(5,5145)
SiZer	(0,7982)	0,8428	(0,8458)	(2,3160)	2,4860	(2,5360)	(2,4430)	2,7209	(2,7580)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,0005)	0,6230	(1,6429)	(26,077)	1065,9	(6866,3)	(0,0000)	0,0001	(0,0004)
	0,95			2,00			2,00			6,10		
EMVb	(0,7500)	0,9000	(0,9500)	(0,9900)	1,4300	(2,0900)	(0,7700)	1,4300	(4,0700)	(1,2100)	3,6300	(5,3900)
EMVf	(0,8182)	0,9229	(0,9511)	(1,0091)	1,4154	(1,9993)	(0,9998)	1,4571	(4,0895)	(1,3971)	3,4821	(5,4453)
SiZer	(0,7960)	0,8399	(0,8442)	(2,2660)	2,3660	(2,4860)	(2,4290)	2,7089	(2,7584)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,5158)	0,5520	(0,5884)	(0,9634)	1,0488	(1,1529)	(0,3891)	0,4875	(0,6001)
	0,95			2,10			2,00			5,00		
EMVb	(0,8000)	0,9000	(0,9500)	(1,2100)	1,4300	(2,0900)	(0,7700)	1,2100	(3,5200)	(1,8700)	3,6300	(5,3900)
EMVf	(0,8537)	0,9192	(0,9522)	(1,1722)	1,5571	(2,1258)	(0,8890)	1,4391	(3,4800)	(2,0046)	3,6422	(5,4453)
SiZer	(0,7984)	0,8430	(0,8458)	(2,3160)	2,4860	(2,5360)	(2,4451)	2,7213	(2,7573)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(488,66)	514,00	(544,59)	(338,35)	382,13	(442,31)	(1,4879)	1,5269	(1,5764)
	0,95			1,10			4,00			1,30		
EMVb	(0,8500)	0,9000	(0,9500)	(0,9900)	1,2100	(1,4300)	(0,7700)	1,2100	(3,6300)	(1,5400)	2,9700	(5,1700)
EMVf	(0,8663)	0,9216	(0,9495)	(0,9459)	1,1411	(1,4325)	(0,8138)	1,4077	(3,2937)	(1,5611)	2,8756	(5,1964)
SiZer	(0,7957)	0,8394	(0,8436)	(2,2760)	2,3810	(2,4860)	(2,4266)	2,7051	(2,7499)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(2,3182)	5,4673	(10,753)	(1014,9)	1170,3	(1355,6)	(29,165)	37,6283	(51,2513)
	0,95			1,10			1,20			4,30		
EMVb	(0,8500)	0,9000	(0,9500)	(0,7700)	0,9900	(1,2100)	(0,7700)	1,2100	(2,3100)	(0,9900)	2,0900	(5,1700)
EMVf	(0,8582)	0,9135	(0,9496)	(0,8561)	1,0260	(1,2505)	(0,8019)	1,2938	(2,3256)	(0,9659)	2,2201	(5,1697)
SiZer	(0,7969)	0,8419	(0,8455)	(2,2960)	2,4860	(2,5460)	(2,4247)	2,6980	(2,7392)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,0000)	0,1348	(1,9105)	(0,0000)	0,0111	(0,3134)	(0,0000)	11,381	(1141,2)
	0,95			1,10			4,20			4,30		
EMVb	(0,9500)	0,9000	(0,9500)	(0,9900)	0,9900	(1,2100)	(2,9700)	4,9500	(5,3900)	(3,1900)	4,9500	(5,3900)
EMVf	(0,9405)	0,9474	(0,9522)	(0,9362)	1,0408	(1,2297)	(2,6909)	4,8168	(5,2240)	(3,2056)	4,9733	(5,4456)
SiZer	(0,7913)	0,9373	(0,9546)	(1,1560)	1,2710	(2,2360)	(2,4166)	2,8832	(3,0200)	(1,0100)	1,0100	(2,1100)
EM	(1,0000)	1,0000	(1,0000)	(1655,4)	1764,2	(1889,6)	(200,92)	203,90	(206,60)	(7,1650)	7,2687	(7,3638)
	0,95			0,40			5,00			2,50		
EMVb	(0,8500)	0,9000	(0,9500)	(0,3300)	0,3300	(0,5500)	(0,9900)	1,8700	(5,1700)	(0,5500)	1,8700	(5,1700)
EMVf	(0,7822)	0,9174	(0,9449)	(0,3686)	0,4248	(0,4754)	(0,8563)	1,3015	(4,6020)	(0,7069)	1,6335	(4,7339)
SiZer	(0,9382)	0,9498	(0,9608)	(1,0260)	1,0360	(1,0660)	(2,9966)	3,9451	(5,1933)	(2,9100)	4,2100	(4,9100)
EM	(1,0000)	1,0000	(1,0000)	(175,58)	332,42	(586,78)	(201,72)	219,98	(314,97)	(0,4804)	3,6129	(6,1091)
	0,95			0,60			1,20			4,30		
EMVb	(0,7500)	0,9000	(0,9500)	(0,5500)	0,5500	(0,7700)	(0,9900)	1,2100	(2,4200)	(0,5500)	0,7700	(1,8700)
EMVf	(0,8018)	0,8814	(0,9435)	(0,5739)	0,6087	(0,7258)	(0,9022)	1,2368	(2,1924)	(0,5792)	0,9621	(2,2042)
SiZer	(0,7922)	0,7946	(0,8405)	(2,2360)	2,2760	(2,4460)	(2,4003)	2,4325	(2,7123)	(1,0100)	1,0100	(1,0100)
EM	(0,0000)	0,0000	(0,0000)	(0,8251)	0,8512	(0,8719)	(0,9882)	1,0349	(1,0778)	(0,7675)	0,8313	(0,9101)

Tabela A.0.17:  $p=0,97$ . Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	0,97	1,10	3,00	2,00
EMVb	0,8889 (0,0959)	1,2564 (0,7557)	1,8586 (1,5225)	2,7940 (1,9827)
EMVf	0,9045 (0,0762)	1,2595 (0,7623)	1,8419 (1,5290)	2,9854 (2,1329)
SiZer	0,8325 (0,0358)	2,4405 (0,2145)	2,6036 (0,2224)	1,0162 (0,2710)
EM	1,0000 (0,0000)	0,9100 (0,1078)	33,440 (4,7589)	0,7238 (0,3804)
	0,97	1,10	5,00	2,00
EMVb	0,8865 (0,0973)	1,0868 (0,4969)	1,9637 (1,5268)	3,0628 (2,0645)
EMVf	0,9034 (0,0740)	1,0801 (0,4983)	1,9387 (1,5291)	3,2618 (2,2108)
SiZer	0,8251 (0,0438)	2,3503 (0,2965)	2,5438 (0,2298)	1,0562 (0,4298)
EM	1,0000 (0,0000)	327,81 (23,363)	231,52 (14,6214)	1,0229 (0,0308)
	0,97	1,10	5,00	5,00
EMVb	0,8176 (0,1933)	0,9297 (0,2387)	2,1745 (1,5033)	3,4536 (2,1912)
EMVf	0,8484 (0,1637)	0,9110 (0,2263)	2,1856 (1,5539)	3,6531 (2,2881)
SiZer	0,8449 (0,0700)	2,1147 (0,5326)	2,5776 (0,2829)	1,3314 (0,9977)
EM	1,0000 (0,0000)	546,59 (70,374)	211,64 (6,2896)	1,5198 (0,0510)
	0,97	1,10	2,00	5,00
EMVb	0,7056 (0,3635)	0,8668 (0,6297)	1,2676 (1,3267)	2,5551 (2,1999)
EMVf	0,7617 (0,2882)	1,0369 (0,4608)	1,4644 (1,1906)	2,9010 (2,1738)
SiZer	0,8245 (0,0384)	2,3842 (0,2586)	2,5644 (0,2350)	1,0462 (0,3719)
EM	1,0000 (0,0000)	0,6348 (0,0616)	1,0123 (0,1252)	0,5624 (0,1720)
	0,97	2,00	1,90	5,00
EMVb	0,6670 (0,3917)	1,2023 (1,1243)	1,3620 (1,5362)	2,2697 (2,1449)
EMVf	0,7290 (0,3105)	1,4345 (0,9147)	1,5953 (1,3394)	2,6197 (2,0458)
SiZer	0,8347 (0,0300)	2,4856 (0,1751)	2,6162 (0,2023)	1,0380 (0,3436)
EM	1,0000 (0,0000)	0,5453 (0,0498)	1,0123 (0,1262)	0,5735 (0,1692)
	0,97	2,00	2,00	6,10
EMVb	0,8895 (0,0937)	1,5431 (0,9329)	1,8634 (1,5365)	3,2991 (1,8959)
EMVf	0,9043 (0,0770)	1,5520 (0,9522)	1,8705 (1,5370)	3,4767 (2,0358)
SiZer	0,8317 (0,0361)	2,4314 (0,2211)	2,6029 (0,2259)	1,0400 (0,3663)
EM	0,9633 (0,1885)	638,44 (398,29)	433,88 (151,43)	19,120 (120,12)
	0,97	2,10	2,00	5,00
EMVb	0,8787 (0,1066)	1,6122 (0,9789)	1,8814 (1,5768)	3,0131 (1,9350)
EMVf	0,8956 (0,0866)	1,6145 (0,9659)	1,8685 (1,5541)	3,1569 (2,0410)
SiZer	0,8351 (0,0340)	2,4730 (0,2015)	2,6132 (0,2463)	1,0256 (0,3028)
EM	1,0000 (0,0000)	10,269 (148,38)	90,593 (404,246)	84,684 (553,66)
	0,97	1,10	4,00	1,30
EMVb	0,8824 (0,0864)	1,2786 (0,6947)	1,7600 (1,4684)	2,6902 (1,9784)
EMVf	0,9006 (0,0633)	1,2763 (0,6820)	1,7489 (1,4565)	2,8628 (2,1492)
SiZer	0,8333 (0,0369)	2,4509 (0,2578)	2,5963 (0,2048)	1,0604 (0,4320)
EM	1,0000 (0,0000)	5,8363 (19,583)	208,42 (1043,5)	106,59 (518,36)
	0,97	1,10	1,20	4,30
EMVb	0,8864 (0,1026)	1,2351 (0,8100)	1,6474 (1,3460)	2,5098 (2,0112)
EMVf	0,9022 (0,0814)	1,2369 (0,7973)	1,5993 (1,3085)	2,7279 (2,2001)
SiZer	0,8360 (0,0345)	2,4824 (0,2291)	2,6023 (0,2075)	1,0372 (0,3821)
EM	0,0000 (0,0000)	0,9819 (0,0400)	0,9953 (0,0400)	1,0320 (0,1140)
	0,97	1,10	4,20	4,30
EMVb	0,8514 (0,1540)	0,9834 (0,3016)	2,0284 (1,4949)	3,4045 (2,0935)
EMVf	0,8768 (0,1256)	0,9672 (0,2940)	2,0315 (1,5475)	3,6146 (2,1949)
SiZer	0,8321 (0,0613)	2,1979 (0,4316)	2,5384 (0,2801)	1,1783 (0,7471)
EM	1,0000 (0,0000)	0,1846 (0,2717)	42,869 (66,383)	0,1546 (0,3099)
	0,97	0,40	5,00	2,50
EMVb	0,8688 (0,1057)	0,4026 (0,1045)	1,4731 (1,1254)	1,1814 (1,4873)
EMVf	0,8365 (0,1081)	0,4437 (0,0817)	1,3603 (0,9969)	1,1991 (1,4474)
SiZer	0,8334 (0,0684)	2,1212 (0,4955)	2,6259 (0,7440)	1,3608 (1,0871)
EM	1,0000 (0,0000)	5,8515 (17,778)	210,78 (743,58)	142,66 (936,23)
	0,97	0,60	1,20	4,30
EMVb	0,8367 (0,1360)	0,6882 (0,1823)	1,4498 (1,0568)	1,4146 (1,5469)
EMVf	0,8633 (0,1033)	0,6781 (0,1687)	1,3814 (1,0056)	1,5023 (1,6086)
SiZer	0,8257 (0,0379)	2,4070 (0,2530)	2,5473 (0,2219)	1,0574 (0,4469)
EM	1,0000 (0,0000)	3,8194 (62,231)	76,862 (233,26)	148,14 (843,35)

Tabela A.0.18:  $p=0,97$ . (Quartil1) Mediana (Quartil3). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Algoritmo EM).

Método	$\hat{p}$			$\hat{\alpha}$			$\hat{\beta}$			$\hat{\lambda}$		
	0,97			1,10			3,00			2,00		
EMVb	(0,8500)	0,9500	(0,9500)	(0,7700)	0,9900	(1,4300)	(0,7700)	1,2100	(2,3100)	(0,7700)	2,3100	(5,1700)
EMVf	(0,8833)	0,9281	(0,9533)	(0,8528)	1,0553	(1,3460)	(0,7852)	1,1935	(2,1056)	(0,9392)	2,3259	(5,4453)
SiZer	(0,7976)	0,8427	(0,8459)	(2,3160)	2,5160	(2,5760)	(2,4161)	2,6855	(2,7257)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,8672)	0,9182	(0,9717)	(30,491)	33,545	(36,253)	(0,4764)	0,6485	(0,8859)
	0,97			1,10			5,00			2,00		
EMVb	(0,8500)	0,9500	(0,9500)	(0,7700)	0,9900	(1,2100)	(0,7700)	1,2100	(2,5300)	(0,7700)	2,9700	(5,3900)
EMVf	(0,8792)	0,9379	(0,9542)	(0,7874)	0,9624	(1,1791)	(0,7793)	1,2484	(2,6241)	(0,9305)	2,9141	(5,6256)
SiZer	(0,7946)	0,7986	(0,8430)	(2,2860)	2,3260	(2,5260)	(2,3977)	2,4372	(2,7077)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(312,90)	325,37	(341,66)	(221,22)	229,07	(238,76)	(1,0016)	1,0215	(1,0433)
	0,97			1,10			5,00			5,00		
EMVb	(0,6500)	0,9500	(0,9500)	(0,7700)	0,7700	(0,9900)	(0,9900)	1,4300	(2,9700)	(0,7700)	5,1700	(5,3900)
EMVf	(0,7730)	0,9497	(0,9584)	(0,7680)	0,8086	(0,9909)	(0,9400)	1,4182	(2,9859)	(0,9251)	5,3900	(5,6256)
SiZer	(0,7929)	0,7979	(0,8443)	(2,2360)	2,2960	(2,4710)	(2,3867)	2,4311	(2,7153)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(494,78)	539,60	(583,10)	(207,39)	210,57	(215,23)	(1,4842)	1,5171	(1,5547)
	0,97			1,10			2,00			5,00		
EMVb	(0,7000)	0,9000	(0,9500)	(0,7700)	0,8800	(0,9900)	(0,5500)	0,9900	(1,4300)	(0,5500)	1,8700	(5,3900)
EMVf	(0,7834)	0,9059	(0,9501)	(0,8190)	0,9105	(1,0806)	(0,8037)	0,9998	(1,4362)	(0,8814)	2,0487	(5,5120)
SiZer	(0,7953)	0,8380	(0,8430)	(2,2960)	2,3560	(2,5410)	(2,3974)	2,6053	(2,7041)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,5944)	0,6321	(0,6725)	(0,9238)	1,0058	(1,0932)	(0,4385)	0,5394	(0,6545)
	0,97			2,00			1,90			5,00		
EMVb	(0,2750)	0,9000	(0,9500)	(0,2750)	0,9900	(1,6500)	(0,2750)	0,9900	(1,4300)	(0,1650)	1,6500	(4,8400)
EMVf	(0,4166)	0,8926	(0,9490)	(0,9105)	1,0919	(1,6311)	(0,8752)	0,9998	(1,4456)	(0,9163)	1,7742	(4,8611)
SiZer	(0,7994)	0,8441	(0,8468)	(2,3560)	2,5460	(2,5860)	(2,4277)	2,6850	(2,7289)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,5084)	0,5406	(0,5793)	(0,9225)	1,0033	(1,1012)	(0,4467)	0,5445	(0,6744)
	0,97			2,00			2,00			6,10		
EMVb	(0,8500)	0,9500	(0,9500)	(0,9900)	1,2100	(1,6500)	(0,7700)	1,2100	(2,3100)	(1,4300)	3,4100	(5,3900)
EMVf	(0,8706)	0,9364	(0,9558)	(0,9949)	1,2622	(1,7056)	(0,7862)	1,1865	(2,2553)	(1,5133)	3,4790	(5,6022)
SiZer	(0,7976)	0,8425	(0,8456)	(2,3260)	2,4960	(2,5610)	(2,4180)	2,6847	(2,7264)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(558,70)	649,08	(735,41)	(396,38)	445,16	(482,97)	(3,8010)	4,1588	(5,2939)
	0,97			2,10			2,00			5,00		
EMVb	(0,8500)	0,9000	(0,9500)	(0,9900)	1,2100	(1,8700)	(0,7700)	1,2100	(2,3100)	(1,2100)	2,7500	(5,1700)
EMVf	(0,8663)	0,9203	(0,9537)	(0,9959)	1,3138	(1,8289)	(0,7907)	1,1491	(2,2576)	(1,2567)	2,8489	(5,3076)
SiZer	(0,7988)	0,8439	(0,8468)	(2,3360)	2,5360	(2,5860)	(2,4301)	2,6876	(2,7295)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,2448)	1,8660	(2,8339)	(0,3082)	62,128	(73,027)	(0,0383)	1,4171	(2,0323)
	0,97			1,10			4,00			1,30		
EMVb	(0,8500)	0,9000	(0,9500)	(0,9900)	0,9900	(1,4300)	(0,7700)	1,2100	(1,8700)	(0,7700)	2,0900	(5,3900)
EMVf	(0,8663)	0,9134	(0,9501)	(0,8692)	1,0795	(1,4149)	(0,7948)	1,1538	(1,9253)	(0,8643)	2,0570	(5,4059)
SiZer	(0,7977)	0,8428	(0,8459)	(2,3260)	2,5160	(2,5760)	(2,4245)	2,6725	(2,7188)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,1781)	0,4047	(0,7338)	(0,1315)	1,1219	(6,8858)	(0,0014)	0,0407	(3,1814)
	0,97			1,10			1,20			4,30		
EMVb	(0,8500)	0,9500	(0,9500)	(0,7700)	0,9900	(1,2100)	(0,7700)	0,9900	(1,8700)	(0,5500)	1,8700	(5,1700)
EMVf	(0,8749)	0,9271	(0,9527)	(0,7911)	1,0086	(1,2678)	(0,7736)	1,1229	(1,6414)	(0,6873)	1,9529	(5,3341)
SiZer	(0,7985)	0,8438	(0,8467)	(2,3460)	2,5560	(2,5960)	(2,4268)	2,6750	(2,7173)	(1,0100)	1,0100	(1,0100)
EM	(0,0000)	0,0000	(0,0000)	(0,9532)	0,9782	(1,0077)	(0,9699)	0,9946	(1,0191)	(0,9541)	1,0181	(1,0978)
	0,97			1,10			4,20			4,30		
EMVb	(0,8000)	0,9000	(0,9500)	(0,7700)	0,9900	(0,9900)	(0,9900)	1,4300	(2,5300)	(0,8800)	4,5100	(5,3900)
EMVf	(0,8417)	0,9438	(0,9573)	(0,7783)	0,8964	(1,0572)	(0,9103)	1,3187	(2,7270)	(1,0001)	4,8213	(5,6256)
SiZer	(0,7927)	0,7956	(0,8423)	(2,2560)	2,2960	(2,4510)	(2,3839)	2,4194	(2,7032)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,0001)	0,0127	(0,3155)	(27,942)	35,422	(45,189)	(0,0033)	0,0163	(0,0690)
	0,97			0,40			5,00			2,50		
EMVb	(0,8000)	0,9000	(0,9500)	(0,3300)	0,3300	(0,5500)	(0,7700)	0,9900	(1,4300)	(0,3300)	0,3300	(1,2100)
EMVf	(0,7541)	0,8416	(0,9341)	(0,3787)	0,4293	(0,5033)	(0,7969)	1,0431	(1,2775)	(0,3672)	0,5143	(1,1941)
SiZer	(0,7915)	0,7937	(0,8405)	(2,2460)	2,2860	(2,3160)	(2,3699)	2,4020	(2,6749)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,3710)	1,1612	(6,7288)	(6,1668)	51,039	(126,84)	(0,0068)	0,0154	(0,0525)
	0,97			0,60			1,20			4,30		
EMVb	(0,7500)	0,9000	(0,9500)	(0,5500)	0,7700	(0,7700)	(0,7700)	0,9900	(1,4300)	(0,3300)	0,7700	(1,6500)
EMVf	(0,8250)	0,8936	(0,9404)	(0,5838)	0,6786	(0,7447)	(0,7693)	1,0665	(1,4119)	(0,5024)	0,8060	(1,7618)
SiZer	(0,7950)	0,8394	(0,8434)	(2,3060)	2,4460	(2,5560)	(2,3929)	2,6203	(2,6892)	(1,0100)	1,0100	(1,0100)
EM	(1,0000)	1,0000	(1,0000)	(0,0000)	0,0025	(0,0972)	(1,3881)	51,236	(78,761)	(0,0004)	0,0065	(0,0425)

Tabela A.0.19: Número de NaN e Inf em EM.

Parâmetros reais ( $\alpha; \beta; \lambda$ )	$p$									
	0,25		0,50		0,70		0,75		0,80	
	#NaN	#Inf								
(1,1;3,0;2,0)	0	0	12	85	63	170	159	453	31	75
(1,1;5,0;2,0)	298	163	0	0	12	36	179	372	0	0
(1,1;5,0;5,0)	0	0	0	0	62	314	132	289	0	0
(1,1;2,0;5,0)	0	0	0	0	84	236	214	231	0	0
(2,0;1,9;5,0)	0	0	110	120	126	280	0	0	0	0
(2,0;2,0;6,1)	0	0	96	218	96	168	132	360	130	194
(2,1;2,0;5,0)	0	0	124	122	102	286	100	530	244	254
(1,1;4,0;1,3)	0	0	52	125	52	226	116	346	59	242
(1,1;1,2;4,3)	0	0	70	226	0	0	88	390	227	505
(1,1;4,2;4,3)	206	96	0	0	60	84	62	174	0	0
(0,4;5,0;2,5)	0	0	24	12	0	0	25	144	65	157
(0,6;1,2;4,3)	0	0	0	0	200	656	102	144	0	0
Parâmetros reais ( $\alpha; \beta; \lambda$ )	$p$									
	0,85		0,90		0,95		0,97			
	#NaN	#Inf	#NaN	#Inf	#NaN	#Inf	#NaN	#Inf		
(1,1;3,0;2,0)	0	0	0	0	54	38	2	0		
(1,1;5,0;2,0)	0	0	0	0	0	0	12	62		
(1,1;5,0;5,0)	0	0	0	0	76	236	88	170		
(1,1;2,0;5,0)	6	14	0	0	134	446	132	290		
(2,0;1,9;5,0)	24	40	0	0	70	172	148	504		
(2,0;2,0;6,1)	0	0	236	106	142	320	58	194		
(2,1;2,0;5,0)	0	0	6	8	142	316	146	350		
(1,1;4,0;1,3)	0	0	12	24	104	350	142	664		
(1,1;1,2;4,3)	6	70	0	2	10	10	130	82		
(1,1;4,2;4,3)	0	0	0	0	50	114	70	152		
(0,4;5,0;2,5)	0	0	16	26	0	0	308	538		
(0,6;1,2;4,3)	432	454	14	60	36	902	0	0		

Tabela A.0.20: Distância Hellinger média.

Param	Método	$p$								
		0,25	0,50	0,70	0,75	0,80	0,85	0,90	0,95	0,97
(1,1;3,0;2,0)	EMVb	0,0512	0,0002	<b>0,0002</b>	0,0002	0,0002	0,0001	<b>0,0001</b>	0,0001	0,0002
	EMVf	0,0511	<b>0,0002</b>	0,0002	<b>0,0002</b>	<b>0,0001</b>	<b>0,0001</b>	0,0709	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0245	0,0096	0,0041	0,0033	0,0025	0,0021	0,1974	0,0023	0,0017
	EM	<b>0,0007</b>	0,0433	0,0213	0,0129	0,0215	0,0044	0,0164	0,0005	0,0002
(1,1;5,0;2,0)	EMVb	0,1281	0,0212	0,0001	0,0001	0,0001	<b>0,0001</b>	0,0001	0,0001	0,0002
	EMVf	0,1352	0,0224	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0097	0,0035	0,0018	0,0016	0,0014	0,0012	0,0015	0,0029	0,0023
	EM	<b>0,0069</b>	<b>0,0024</b>	0,0276	0,0199	0,0139	0,0036	NaN	0,0010	0,0004
(1,1;5,0;5,0)	EMVb	0,2169	0,0993	0,0017	0,0001	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0002
	EMVf	0,2714	0,1094	0,0018	<b>0,0001</b>	<b>0,0001</b>	0,0001	0,0001	0,0001	<b>0,0002</b>
	SiZer	<b>0,0029</b>	<b>0,0011</b>	<b>0,0007</b>	0,0006	0,0007	0,0007	0,0008	0,0020	0,0028
	EM	0,4996	0,0108	0,0431	0,0274	0,0249	0,0108	NaN	NaN	0,0009
(1,1;2,0;5,0)	EMVb	0,0973	0,0071	0,0002	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
	EMVf	0,1029	0,0076	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0002</b>
	SiZer	0,0120	0,0042	0,0020	0,0017	0,0014	0,0013	0,0018	0,0028	0,0022
	EM	<b>0,0004</b>	<b>0,0019</b>	0,0419	0,0324	0,0302	0,0049	NaN	0,0008	0,0003
(2,0;1,9;5,0)	EMVb	0,0187	0,0001	0,0002	0,0002	<b>0,0002</b>	0,0001	0,0002	0,0001	0,0002
	EMVf	0,0182	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0014	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0002
	SiZer	0,0339	0,0145	0,0058	0,0043	0,0030	0,0022	0,0022	0,0018	0,0015
	EM	<b>0,0005</b>	0,0229	0,0134	0,0099	0,0054	0,0031	0,0014	0,0004	<b>0,0001</b>
(2,0;2,0;6,1)	EMVb	0,0466	0,0002	0,0001	0,0001	<b>0,0002</b>	0,0001	0,0002	<b>0,0002</b>	0,0002
	EMVf	0,0463	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0002	<b>0,0001</b>	<b>0,0001</b>	0,1502	<b>0,0001</b>
	SiZer	0,0256	0,0110	0,0047	0,0035	0,0026	0,0018	0,0022	0,1515	0,0017
	EM	<b>0,0006</b>	0,0426	0,0115	0,0101	0,0073	0,0035	0,0156	0,1503	0,1766
(2,1;2,0;5,0)	EMVb	0,0214	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0002	0,0002
	EMVf	0,0209	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0354	0,0151	0,0061	0,0045	0,0031	0,0021	0,0022	0,0018	0,0015
	EM	<b>0,0006</b>	0,0221	0,0208	0,0164	0,0339	0,0032	0,2039	0,0004	0,0001
(1,1;4,0;1,3)	EMVb	0,0383	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001	0,0002
	EMVf	0,0419	<b>0,0002</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0303	0,0121	0,0051	0,0038	0,0028	0,0023	0,0027	0,0021	0,0017
	EM	<b>0,0011</b>	0,0420	0,0164	0,0154	0,0123	0,0040	0,0014	0,0005	0,0002
(1,1;1,2;4,3)	EMVb	<b>0,0014</b>	0,0002	<b>0,0001</b>	0,0002	0,0001	0,0001	0,0002	0,0002	0,0002
	EMVf	0,0015	<b>0,0002</b>	0,0030	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0162	0,0072	0,0033	0,0027	0,0024	0,0028	0,0027	0,0018	0,0015
	EM	0,0574	0,0248	0,0097	0,0069	0,0241	0,0005	0,0012	0,0003	0,0001
(1,1;4,2;4,3)	EMVb	0,1869	0,0699	0,0001	0,0001	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0002
	EMVf	0,2247	0,0759	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0001	0,0001	0,0001	<b>0,0002</b>
	SiZer	<b>0,0040</b>	<b>0,0012</b>	0,0006	0,0007	0,0007	0,0007	0,0008	0,0026	0,0027
	EM	0,0076	0,0066	0,0157	0,0210	0,0227	0,0093	NaN	NaN	0,0007
(0,4;5,0;2,5)	EMVb	0,2227	0,1401	0,0652	0,0339	0,0040	0,0005	0,0004	0,0002	0,0002
	EMVf	0,2342	0,1034	0,0628	0,0360	<b>0,0038</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0002</b>
	SiZer	<b>0,0493</b>	0,0305	0,0176	0,0141	0,0111	0,0081	0,0052	0,0037	0,0045
	EM	0,4914	<b>0,0052</b>	<b>0,0103</b>	<b>0,0006</b>	0,0363	0,0252	0,0142	0,0052	0,0025
(0,6;1,2;4,3)	EMVb	0,1081	0,0174	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
	EMVf	0,1127	0,0181	<b>0,0001</b>	<b>0,0001</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0169	0,0080	0,0039	0,0031	0,0026	0,0023	0,0034	0,0033	0,0023
	EM	<b>0,0005</b>	<b>0,0007</b>	0,0184	0,0032	0,0104	0,0080	NaN	0,0003	0,0005

Tabela A.0.21: Distância Hellinger (desvios).

Param	Método	$p$								
		0,25	0,50	0,70	0,75	0,80	0,85	0,90	0,95	0,97
(1,1;3,0;2,0)	EMVb	0,0067	0,0001	<b>0,0001</b>	0,0001	0,0001	0,0001	<b>0,0001</b>	0,0001	0,0001
	EMVf	0,0039	<b>0,0001</b>	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0418	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0021	0,0016	0,0012	0,0010	0,0009	0,0011	0,0060	0,0005	0,0004
	EM	<b>0,0001</b>	0,0045	0,0653	0,0254	0,0818	0,0000	0,0829	0,0000	0,0000
(1,1;5,0;2,0)	EMVb	0,0034	0,0033	0,0001	0,0001	0,0001	<b>0,0001</b>	0,0002	0,0001	0,0001
	EMVf	0,0097	0,0039	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0022	0,0016	0,0013	0,0011	0,0010	0,0007	0,0015	0,0013	0,0006
	EM	<b>0,0467</b>	<b>0,0001</b>	0,0009	0,0000	0,0564	0,0033	NaN	0,0000	0,0000
(1,1;5,0;5,0)	EMVb	0,0070	0,0034	0,0018	0,0001	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0002</b>	0,0001
	EMVf	0,0053	0,0043	0,0019	<b>0,0001</b>	<b>0,0001</b>	0,0001	0,0001	0,0001	<b>0,0001</b>
	SiZer	<b>0,0016</b>	<b>0,0005</b>	<b>0,0016</b>	0,0004	0,0004	0,0010	0,0005	0,0021	0,0012
	EM	0,0000	0,0002	0,0364	0,0136	0,0030	0,0055	NaN	NaN	0,0000
(1,1;2,0;5,0)	EMVb	0,0071	0,0033	0,0001	0,0001	0,0001	0,0001	0,0002	0,0001	0,0001
	EMVf	0,0068	0,0030	<b>0,0001</b>						
	SiZer	0,0012	0,0007	0,0010	0,0009	0,0008	0,0008	0,0019	0,0011	0,0005
	EM	<b>0,0001</b>	<b>0,0001</b>	0,1070	0,0871	0,1069	0,0026	NaN	0,0000	0,0000
(2,0;1,9;5,0)	EMVb	0,0030	0,0001	0,0001	0,0001	<b>0,0001</b>	0,0001	0,0001	0,0001	0,0001
	EMVf	0,0032	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0017	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0002
	SiZer	0,0024	0,0015	0,0009	0,0008	0,0006	0,0006	0,0009	0,0003	0,0003
	EM	<b>0,0001</b>	0,0134	0,0376	0,0314	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>
(2,0;2,0;6,1)	EMVb	0,0024	0,0001	0,0001	0,0001	<b>0,0001</b>	0,0001	0,0001	<b>0,0001</b>	0,0001
	EMVf	0,0034	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0001	<b>0,0001</b>	<b>0,0001</b>	0,3573	<b>0,0001</b>
	SiZer	0,0020	0,0015	0,0010	0,0008	0,0007	0,0007	0,0014	0,3559	0,0004
	EM	<b>0,0001</b>	0,0363	0,0325	0,0030	0,0000	0,0014	0,0822	0,3571	0,2462
(2,1;2,0;5,0)	EMVb	0,0028	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
	EMVf	0,0035	<b>0,0001</b>							
	SiZer	0,0024	0,0016	0,0010	0,0008	0,0006	0,0008	0,0009	0,0004	0,0003
	EM	<b>0,0001</b>	0,0436	0,0643	0,0678	0,1164	0,0000	0,2457	0,0000	0,0000
(1,1;4,0;1,3)	EMVb	0,0038	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
	EMVf	0,0034	<b>0,0001</b>							
	SiZer	0,0022	0,0018	0,0013	0,0010	0,0009	0,0011	0,0013	0,0005	0,0004
	EM	<b>0,0001</b>	0,0243	0,0327	0,0523	0,0536	0,0000	0,0007	0,0000	0,0000
(1,1;1,2;4,3)	EMVb	<b>0,0016</b>	0,0001	<b>0,0001</b>	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
	EMVf	0,0019	<b>0,0001</b>	0,0020	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>
	SiZer	0,0019	0,0012	0,0010	0,0009	0,0013	0,0016	0,0009	0,0005	0,0003
	EM	0,0000	0,0627	0,0000	0,0000	0,0991	0,0007	0,0000	0,0000	0,0001
(1,1;4,2;4,3)	EMVb	0,0009	0,0040	0,0001	0,0001	0,0001	<b>0,0001</b>	<b>0,0001</b>	<b>0,0002</b>	0,0001
	EMVf	0,0057	0,0044	<b>0,0001</b>	<b>0,0001</b>	<b>0,0001</b>	0,0001	0,0001	0,0001	<b>0,0001</b>
	SiZer	<b>0,0017</b>	<b>0,0018</b>	0,0005	0,0013	0,0009	0,0005	0,0007	0,0020	0,0010
	EM	0,0419	0,0001	0,0284	0,0118	0,0407	0,0043	NaN	NaN	0,0000
(0,4;5,0;2,5)	EMVb	0,0000	0,0000	0,0057	0,0053	0,0036	0,0002	0,0002	0,0001	0,0001
	EMVf	0,0093	0,0100	0,0104	0,0088	<b>0,0037</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0002</b>	<b>0,0001</b>
	SiZer	<b>0,0008</b>	0,0008	0,0008	0,0008	0,0007	0,0006	0,0005	0,0018	0,0012
	EM	0,0000	<b>0,0001</b>	<b>0,0232</b>	<b>0,0001</b>	0,0072	0,0000	0,0000	0,0000	0,0000
(0,6;1,2;4,3)	EMVb	0,0053	0,0051	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001	0,0001
	EMVf	0,0078	0,0058	<b>0,0001</b>						
	SiZer	0,0006	0,0004	0,0004	0,0005	0,0006	0,0010	0,0025	0,0009	0,0005
	EM	<b>0,0001</b>	<b>0,0001</b>	0,0114	0,0069	0,0062	0,0000	NaN	0,0004	0,0000

Tabela A.0.22: FDR - Média (Desvio padrão). A coluna *Método* identifica o método de estimação, sendo EMVb (EMV por busca direta), EMVf (EMV por busca de Fibonacci), SiZer (estimação através do SiZer) e EM (Método EM).

Parâmetros reais	Método usado	$p$								
		0,25	0,50	0,70	0,75	0,80	0,85	0,90	0,95	0,97
(1,1;3,0;2,0)	FDR real	0,0710	0,1845	0,3431	0,4715	0,6671	0,8086	0,8777	0,4013	0,5579
	EMVb	0,0004 (0,0006)	0,1811 (0,0247)	0,3261 (0,0523)	0,3772 (0,0614)	0,4287 (0,0765)	0,5207 (0,0689)	0,6141 (0,0928)	0,7537 (0,0827)	0,8074 (0,1065)
	EMVf	0,0003 (0,0005)	0,1835 (0,0257)	0,3398 (0,0506)	0,3903 (0,0574)	0,4422 (0,0705)	0,5372 (0,0641)	0,0246 (0,0602)	0,7681 (0,0709)	0,8214 (0,0899)
	SiZer	0,2738 (0,0248)	0,3945 (0,0392)	0,5832 (0,0569)	0,6551 (0,0624)	0,7363 (0,0737)	0,8370 (0,0770)	0,0290 (0,0032)	0,9717 (0,0178)	0,9788 (0,0116)
	EM	0,0000 (0,0000)	0,9888 (0,0967)	0,8239 (0,3777)	0,9956 (0,0505)	0,9674 (0,1732)	0,9982 (0,0000)	0,9461 (0,2116)	0,9982 (0,0000)	0,9982 (0,0000)
	FDR real	0,0441	0,1199	0,2394	0,3502	0,5504	0,7237	0,8177	0,2879	0,4337
(1,1;5,0;2,0)	EMVb	0,0001 (0,0001)	0,0013 (0,0034)	0,2394 (0,0203)	0,2863 (0,0232)	0,3459 (0,0300)	0,4314 (0,0363)	0,5350 (0,0551)	0,6880 (0,0875)	0,7554 (0,0994)
	EMVf	0,0000 (0,0000)	0,0001 (0,0018)	0,2401 (0,0176)	0,2872 (0,0209)	0,3470 (0,0282)	0,4356 (0,0363)	0,5401 (0,0515)	0,6973 (0,0763)	0,7715 (0,0826)
	SiZer	0,1111 (0,0106)	0,1967 (0,0176)	0,3406 (0,0360)	0,4040 (0,0433)	0,4820 (0,0548)	0,5934 (0,0682)	0,7639 (0,1043)	0,9547 (0,0372)	0,9739 (0,0178)
	EM	0,0022 (0,0466)	0,0000 (0,0000)	0,9970 (0,0344)	0,9982 (0,0000)	0,5501 (0,4356)	0,8996 (0,2964)	0,9984 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)
	FDR real	0,0182	0,0520	0,1131	0,1810	0,3434	0,5447	0,6831	0,1412	0,2412
	(1,1;5,0;5,0)	EMVb	0,0000 (0,0003)	0,0003 (0,0007)	0,0987 (0,0209)	0,1416 (0,0114)	0,1824 (0,0143)	0,2435 (0,0161)	0,3456 (0,0235)	0,5392 (0,0820)
EMVf		0,0000 (0,0000)	0,0000 (0,0000)	0,0895 (0,0227)	0,1421 (0,0092)	0,1822 (0,0123)	0,2421 (0,0160)	0,3432 (0,0238)	0,5399 (0,0746)	0,6249 (0,1225)
SiZer		0,0289 (0,0031)	0,0703 (0,0051)	0,0755 (0,0735)	0,1782 (0,0122)	0,2299 (0,0178)	0,3137 (0,0434)	0,4608 (0,0466)	0,8117 (0,1208)	0,9537 (0,0560)
EM		0,0000 (0,0000)	0,0000 (0,0000)	0,7730 (0,4175)	0,6994 (0,4574)	0,9751 (0,1503)	0,6707 (0,4064)	0,9984 (0,0000)	0,9984 (0,0000)	0,9982 (0,0000)
FDR real		0,0441	0,1204	0,2416	0,3546	0,5578	0,7311	0,8239	0,2910	0,4395
(1,1;2,0;5,0)		EMVb	0,0003 (0,0005)	0,0091 (0,0189)	0,2438 (0,0242)	0,2905 (0,0283)	0,3543 (0,0342)	0,4361 (0,0392)	0,5436 (0,0554)	0,6909 (0,0903)
	EMVf	0,0000 (0,0000)	0,0115 (0,0174)	0,2459 (0,0206)	0,2943 (0,0253)	0,3573 (0,0325)	0,4405 (0,0371)	0,5493 (0,0504)	0,7032 (0,0767)	0,6544 (0,2607)
	SiZer	0,1291 (0,0118)	0,2165 (0,0162)	0,3671 (0,0400)	0,4348 (0,0454)	0,5165 (0,0585)	0,6303 (0,0829)	0,8136 (0,1057)	0,9577 (0,0407)	0,9751 (0,0141)
	EM	0,0000 (0,0000)	0,0000 (0,0000)	0,7507 (0,4282)	0,8610 (0,3275)	0,5106 (0,4275)	0,7272 (0,4314)	0,9984 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)
	FDR real	0,1686	0,3655	0,5593	0,6771	0,8183	0,9019	0,9388	0,6159	0,7441
	(2,0;1,9;5,0)	EMVb	0,0178 (0,0202)	0,3674 (0,0342)	0,5409 (0,0795)	0,5910 (0,0904)	0,4117 (0,3162)	0,7171 (0,0939)	0,7552 (0,1111)	0,8352 (0,1079)
EMVf		0,0032 (0,0079)	0,3680 (0,0301)	0,5463 (0,0710)	0,5987 (0,0814)	0,4519 (0,2761)	0,7291 (0,0829)	0,7637 (0,0995)	0,8487 (0,0917)	0,6953 (0,3111)
SiZer		0,5457 (0,0330)	0,6515 (0,0393)	0,7879 (0,0478)	0,8332 (0,0469)	0,8821 (0,0476)	0,9363 (0,0373)	0,9653 (0,0219)	0,9776 (0,0107)	0,9809 (0,0072)
EM		0,0000 (0,0000)	0,7593 (0,3757)	0,8922 (0,3064)	0,9478 (0,2171)	0,9982 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)
FDR real		0,1361	0,3049	0,4859	0,6062	0,7648	0,8681	0,9163	0,5426	0,6793
(2,0;2,0;6,1)		EMVb	0,0005 (0,0019)	0,3064 (0,0260)	0,4835 (0,0499)	0,5367 (0,0555)	0,5925 (0,0784)	0,6599 (0,0816)	0,7183 (0,0978)	0,6880 (0,3047)
	EMVf	0,0000 (0,0002)	0,3081 (0,0237)	0,4850 (0,0473)	0,5400 (0,0530)	0,6009 (0,0771)	0,6700 (0,0755)	0,7210 (0,0902)	0,7202 (0,2547)	0,8467 (0,0836)
	SiZer	0,3655 (0,0231)	0,4762 (0,0332)	0,6343 (0,0495)	0,6915 (0,0484)	0,7646 (0,0559)	0,8461 (0,0593)	0,9375 (0,0443)	0,9731 (0,0118)	0,9783 (0,0129)
	EM	0,0000 (0,0000)	0,9452 (0,2243)	0,6427 (0,4763)	0,9183 (0,2693)	0,9982 (0,0000)	0,8620 (0,3302)	0,7679 (0,4101)	0,9982 (0,0000)	0,9617 (0,1882)

Parámetros reais	Método usado	$p$								
		0,25	0,50	0,70	0,75	0,80	0,85	0,90	0,95	0,97
(2,1;2,0;5,0)	FDR real	0,1761	0,3782	0,5729	0,6891	0,8264	0,9067	0,9419	0,6290	0,7545
	EMVb	0,0163 (0,0193)	0,3773 (0,0368)	0,5627 (0,0701)	0,6088 (0,0861)	0,6436 (0,1158)	0,7140 (0,1039)	0,7678 (0,1060)	0,8278 (0,1028)	0,8465 (0,1089)
	EMVf	0,0026 (0,0072)	0,3780 (0,0311)	0,5652 (0,0640)	0,6146 (0,0754)	0,6552 (0,1068)	0,7234 (0,0915)	0,7756 (0,0956)	0,8428 (0,0877)	0,8614 (0,0916)
	SiZer	0,5581 (0,0299)	0,6556 (0,0373)	0,7907 (0,0484)	0,8347 (0,0496)	0,8840 (0,0484)	0,9249 (0,0408)	0,9660 (0,0216)	0,9780 (0,0089)	0,9807 (0,0076)
	EM	0,0000 (0,0000)	0,6445 (0,3484)	0,9762 (0,1472)	0,8172 (0,3839)	0,9548 (0,2058)	0,9982 (0,0000)	0,5518 (0,4950)	0,9982 (0,0000)	0,9982 (0,0000)
(1,1;4,0;1,3)	FDR real	0,0809	0,2065	0,3749	0,5056	0,6957	0,8276	0,8905	0,4347	0,5909
	EMVb	0,0000 (0,0003)	0,2005 (0,0349)	0,3542 (0,0603)	0,4084 (0,0681)	0,4529 (0,0862)	0,5493 (0,0818)	0,6341 (0,1045)	0,7683 (0,1084)	0,8190 (0,0950)
	EMVf	0,0000 (0,0000)	0,2037 (0,0301)	0,3665 (0,0570)	0,4204 (0,0638)	0,4686 (0,0819)	0,5675 (0,0710)	0,6493 (0,0922)	0,7812 (0,0929)	0,8350 (0,0741)
	SiZer	0,3408 (0,0315)	0,4716 (0,0448)	0,2973 (0,3329)	0,7231 (0,0659)	0,7936 (0,0683)	0,8866 (0,0650)	0,9489 (0,0394)	0,9734 (0,0134)	0,9792 (0,0088)
	EM	0,0000 (0,0000)	0,9908 (0,0857)	0,9474 (0,2162)	0,9112 (0,2823)	0,9248 (0,2612)	0,9982 (0,0000)	0,7479 (0,4255)	0,9982 (0,0000)	0,9982 (0,0000)
(1,1;1,2;4,3)	FDR real	0,0815	0,2090	0,3807	0,5131	0,7042	0,8341	0,8951	0,4415	0,5993
	EMVb	0,0481 (0,0473)	0,2174 (0,0365)	0,0952 (0,1668)	0,4456 (0,0589)	0,5004 (0,0687)	0,5950 (0,0691)	0,6640 (0,1035)	0,7769 (0,0976)	0,8258 (0,1115)
	EMVf	0,0578 (0,0440)	0,2344 (0,0319)	0,1462 (0,1510)	0,4625 (0,0526)	0,5204 (0,0640)	0,6069 (0,0594)	0,6781 (0,0897)	0,7924 (0,0823)	0,8396 (0,0942)
	SiZer	0,3573 (0,0361)	0,4984 (0,0472)	0,6923 (0,0691)	0,7621 (0,0824)	0,8502 (0,0822)	0,9235 (0,0625)	0,9613 (0,0318)	0,9768 (0,0152)	0,9812 (0,0086)
	EM	0,9982 (0,0000)	0,6413 (0,4784)	0,9982 (0,0000)	0,9982 (0,0000)	0,9358 (0,2455)	0,0960 (0,2947)	0,6655 (0,5763)	0,9982 (0,0000)	0,0000 (0,0000)
(1,1;4,2;4,3)	FDR real	0,0250	0,0705	0,1497	0,2333	0,4153	0,6124	0,7355	0,1850	0,3038
	EMVb	0,0000 (0,0000)	0,0015 (0,0005)	0,1500 (0,0134)	0,1850 (0,0142)	0,2338 (0,0170)	0,3030 (0,0210)	0,4172 (0,0301)	0,5994 (0,0713)	0,6568 (0,1277)
	EMVf	0,0000 (0,0000)	0,0000 (0,0001)	0,1506 (0,0101)	0,1857 (0,0120)	0,2347 (0,0158)	0,3033 (0,0212)	0,4181 (0,0303)	0,6027 (0,0635)	0,6851 (0,1076)
	SiZer	0,0461 (0,0052)	0,0999 (0,0097)	0,1937 (0,0143)	0,2392 (0,0312)	0,3028 (0,0344)	0,3984 (0,0397)	0,5559 (0,0745)	0,8932 (0,0923)	0,9626 (0,0401)
	EM	0,0033 (0,0571)	0,0000 (0,0000)	0,2446 (0,4299)	0,6620 (0,4724)	0,9372 (0,2382)	0,7558 (0,3782)	0,9984 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)
(0,4;5,0;2,5)	FDR real	0,0004	0,0020	0,0098	0,0297	0,1447	0,3979	0,5919	0,0164	0,0605
	EMVb	0,0000 (0,0000)	0,0000 (0,0000)	0,0233 (0,0306)	0,0224 (0,0121)	0,0219 (0,0116)	0,0506 (0,0144)	0,1322 (0,0255)	0,3650 (0,0608)	0,5408 (0,0905)
	EMVf	0,0000 (0,0000)	0,0000 (0,0000)	0,0001 (0,0002)	0,0003 (0,0003)	0,0149 (0,0134)	0,0586 (0,0101)	0,1367 (0,0221)	0,3489 (0,0627)	0,5044 (0,0946)
	SiZer	0,0132 (0,0006)	0,0357 (0,0015)	0,0808 (0,0040)	0,0051 (0,0223)	0,1352 (0,0080)	0,0190 (0,0571)	0,2952 (0,0261)	0,6669 (0,1393)	0,9497 (0,0690)
	EM	0,0028 (0,0000)	0,0000 (0,0000)	0,1426 (0,3495)	0,0000 (0,0000)	0,9982 (0,0000)	0,8747 (0,3376)	0,9982 (0,0000)	0,9982 (0,0000)	0,9982 (0,0000)
(0,6;1,2;4,3)	FDR real	0,0102	0,0385	0,1123	0,2110	0,4410	0,6616	0,7805	0,1521	0,3007
	EMVb	0,0022 (0,0034)	0,0033 (0,0020)	0,1059 (0,0212)	0,1440 (0,0299)	0,1963 (0,0379)	0,2702 (0,0656)	0,3960 (0,0889)	0,5671 (0,1282)	0,6773 (0,1257)
	EMVf	0,0000 (0,0000)	0,0012 (0,0015)	0,1139 (0,0199)	0,1502 (0,0277)	0,2012 (0,0378)	0,2777 (0,0584)	0,3999 (0,0745)	0,5819 (0,1099)	0,6993 (0,1039)
	SiZer	0,0758 (0,0044)	0,1347 (0,0086)	0,2518 (0,0249)	0,3080 (0,0352)	0,3991 (0,0560)	0,5391 (0,0862)	0,7945 (0,1362)	0,9619 (0,0396)	0,9768 (0,0148)
	EM	0,0000 (0,0000)	0,0000 (0,0000)	0,7378 (0,4482)	0,1597 (0,3735)	0,7928 (0,4041)	0,9982 (0,0000)	0,9984 (0,0000)	0,8136 (0,3882)	0,9955 (0,0521)



Parâmetros reais	Método usado	$p$								
		0,25	0,50	0,70	0,75	0,80	0,85	0,90	0,95	0,97
(2,0;2,0;6,1)	FDR real	0,1361	0,3049	0,4859	0,6062	0,7648	0,8681	0,9163	0,5426	0,6793
	EMVb	(0,0000)	(0,2939)	(0,4586)	(0,5151)	(0,5637)	(0,6288)	(0,6637)	(0,6941)	(0,7931)
		0,0000	0,3034	0,4855	0,5379	0,6052	0,6691	0,7362	0,8050	0,8459
	EMVf	(0,0000)	(0,2916)	(0,4639)	(0,5189)	(0,5742)	(0,6420)	(0,6785)	(0,7176)	(0,8035)
		0,0000	0,3087	0,4912	0,5447	0,6150	0,6817	0,7393	0,8151	0,8560
SiZer	(0,3504)	(0,4573)	(0,6014)	(0,6600)	(0,7282)	(0,8054)	(0,9169)	(0,9666)	(0,9711)	
	0,3659	0,4746	0,6327	0,6921	0,7642	0,8459	0,9562	0,9761	0,9826	
EM	(0,3808)	(0,4923)	(0,6624)	(0,7230)	(0,8017)	(0,8840)	(0,9698)	(0,9817)	(0,9851)	
	(0,0000)	(0,9982)	(0,0000)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	
(2,1;2,0;5,0)	FDR real	0,1761	0,3782	0,5729	0,6891	0,8264	0,9067	0,9419	0,6290	0,7545
	EMVb	(0,0000)	(0,3572)	(0,5423)	(0,5790)	(0,6128)	(0,6758)	(0,7012)	(0,7713)	(0,8036)
		0,0000	0,3800	0,5704	0,6221	0,6746	0,7448	0,7803	0,8499	0,8607
	EMVf	(0,0323)	(0,4002)	(0,6024)	(0,6565)	(0,7181)	(0,7898)	(0,8458)	(0,9012)	(0,9139)
		(0,0000)	(0,3579)	(0,5409)	(0,5897)	(0,6213)	(0,6876)	(0,7142)	(0,7930)	(0,8235)
SiZer	0,0000	0,3781	0,5715	0,6227	0,6789	0,7467	0,7901	0,8564	0,8712	
	(0,0000)	(0,3997)	(0,6013)	(0,6583)	(0,7302)	(0,7972)	(0,8434)	(0,9051)	(0,9208)	
EM	(0,5378)	(0,6285)	(0,7590)	(0,8019)	(0,8558)	(0,8879)	(0,9647)	(0,9705)	(0,9726)	
	0,5578	0,6553	0,7916	0,8350	0,8875	0,9348	0,9732	0,9817	0,9840	
EM	(0,5787)	(0,6801)	(0,8228)	(0,8675)	(0,9174)	(0,9642)	(0,9762)	(0,9839)	(0,9859)	
	(0,0000)	(0,4368)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,0000)	(0,9982)	(0,9982)	
(1,1;4,0;1,3)	FDR real	0,0809	0,2065	0,3749	0,5056	0,6957	0,8276	0,8905	0,4347	0,5909
	EMVb	(0,0000)	(0,1833)	(0,3292)	(0,3870)	(0,3920)	(0,5068)	(0,5826)	(0,7159)	(0,7733)
		0,0000	0,1996	0,3622	0,4230	0,4713	0,5633	0,6470	0,7796	0,8327
	EMVf	(0,0000)	(0,2260)	(0,3924)	(0,4565)	(0,5118)	(0,6005)	(0,7027)	(0,8362)	(0,8814)
		(0,0000)	(0,1937)	(0,3492)	(0,4016)	(0,4227)	(0,5375)	(0,6076)	(0,7321)	(0,7898)
SiZer	0,0000	0,2042	0,3749	0,4314	0,4872	0,5781	0,6631	0,7904	0,8435	
	(0,0000)	(0,2193)	(0,4000)	(0,4612)	(0,5274)	(0,6116)	(0,7081)	(0,8368)	(0,8815)	
EM	(0,3195)	(0,4434)	(0,0000)	(0,6831)	(0,7510)	(0,8487)	(0,9495)	(0,9670)	(0,9712)	
	0,3415	0,4706	0,0000	0,7245	0,7932	0,8978	0,9647	0,9772	0,9834	
EM	(0,3605)	(0,4958)	(0,6544)	(0,7638)	(0,8388)	(0,9392)	(0,9717)	(0,9818)	(0,9855)	
	(0,0000)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,4517)	(0,9982)	(0,9982)	
(1,1;1,2;4,3)	FDR real	0,0815	0,2090	0,3807	0,5131	0,7042	0,8341	0,8951	0,4415	0,5993
	EMVb	(0,0000)	(0,1939)	(0,0000)	(0,4166)	(0,4605)	(0,5562)	(0,6190)	(0,7265)	(0,7770)
		0,0497	0,2244	0,0000	0,4517	0,5148	0,6014	0,6732	0,7905	0,8383
	EMVf	(0,0937)	(0,2442)	(0,1293)	(0,4835)	(0,5479)	(0,6397)	(0,7250)	(0,8443)	(0,8959)
		(0,0211)	(0,2157)	(0,0548)	(0,4346)	(0,4878)	(0,5745)	(0,6322)	(0,7446)	(0,7956)
SiZer	0,0496	0,2374	0,0548	0,4653	0,5335	0,6092	0,6832	0,7974	0,8450	
	(0,0973)	(0,2547)	(0,1816)	(0,4955)	(0,5631)	(0,6446)	(0,7334)	(0,8473)	(0,9003)	
EM	(0,3339)	(0,4682)	(0,6500)	(0,7090)	(0,7950)	(0,8943)	(0,9600)	(0,9695)	(0,9729)	
	0,3548	0,4967	0,6916	0,7653	0,8613	0,9548	0,9672	0,9818	0,9847	
EM	(0,3758)	(0,5257)	(0,7287)	(0,8214)	(0,9149)	(0,9672)	(0,9765)	(0,9845)	(0,9866)	
	(0,9982)	(0,0000)	(0,9982)	(0,9982)	(0,9982)	(0,0000)	(0,2496)	(0,9982)	(0,0000)	
(1,1;4,2;4,3)	FDR real	0,0250	0,0705	0,1497	0,2333	0,4153	0,6124	0,7355	0,1850	0,3038
	EMVb	(0,0000)	(0,0011)	(0,1385)	(0,1733)	(0,2218)	(0,2887)	(0,3969)	(0,5780)	(0,6176)
		0,0000	0,0014	0,1523	0,1862	0,2331	0,3009	0,4163	0,6084	0,6865
	EMVf	(0,0000)	(0,0017)	(0,1596)	(0,1952)	(0,2434)	(0,3165)	(0,4360)	(0,6399)	(0,7394)
		(0,0000)	(0,0000)	(0,1434)	(0,1773)	(0,2234)	(0,2888)	(0,3975)	(0,5757)	(0,6487)
SiZer	0,0000	0,0000	0,1507	0,1855	0,2346	0,3017	0,4169	0,6090	0,7110	
	(0,0000)	(0,0000)	(0,1577)	(0,1928)	(0,2457)	(0,3178)	(0,4377)	(0,6403)	(0,7528)	
EM	(0,0424)	(0,0945)	(0,1835)	(0,2266)	(0,2850)	(0,3703)	(0,5052)	(0,8271)	(0,9660)	
	0,0457	0,0991	0,1938	0,2376	0,3030	0,3982	0,5532	0,9516	0,9690	
EM	(0,0495)	(0,1043)	(0,2040)	(0,2483)	(0,3196)	(0,4244)	(0,6036)	(0,9640)	(0,9811)	
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,9982)	(0,2893)	(0,9984)	(0,9982)	(0,9982)	
EM	0,0000	0,0000	0,0000	0,9982	0,9982	0,9982	0,9984	0,9982	0,9982	
	(0,0000)	(0,0000)	(0,0000)	(0,9982)	(0,9982)	(0,9982)	(0,9984)	(0,9982)	(0,9982)	

Parâmetros reais	Método usado	$p$								
		0,25	0,50	0,70	0,75	0,80	0,85	0,90	0,95	0,97
(0,4;5,0;2,5)	FDR real	0,0004	0,0020	0,0098	0,0297	0,1447	0,3979	0,5919	0,0164	0,0605
	EMVb	(0,0000)	(0,0000)	(0,0056)	(0,0132)	(0,0129)	(0,0426)	(0,1179)	(0,3492)	(0,4787)
		0,0000	0,0000	0,0090	0,0189	0,0210	0,0474	0,1312	0,3750	0,5567
		(0,0000)	(0,0000)	(0,0217)	(0,0287)	(0,0250)	(0,0537)	(0,1498)	(0,3997)	(0,6065)
	EMVf	(0,0000)	(0,0000)	(0,0000)	(0,0001)	(0,0039)	(0,0536)	(0,1247)	(0,3131)	(0,4286)
		0,0000	0,0000	0,0001	0,0002	0,0083	0,0593	0,1384	0,3563	0,5031
		(0,0000)	(0,0000)	(0,0002)	(0,0004)	(0,0307)	(0,0646)	(0,1512)	(0,3914)	(0,5886)
	SiZer	(0,0128)	(0,0346)	(0,0781)	(0,0000)	(0,1297)	(0,0000)	(0,2772)	(0,5648)	(0,9649)
		0,0132	0,0356	0,0809	0,0000	0,1348	0,0000	0,2938	0,6366	0,9680
		(0,0135)	(0,0367)	(0,0834)	(0,0000)	(0,1405)	(0,0000)	(0,3096)	(0,7334)	(0,9718)
EM	(0,0028)	(0,0000)	(0,0000)	(0,0000)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	
	0,0028	0,0000	0,0000	0,0000	0,9982	0,9982	0,9982	0,9982	0,9982	
	(0,0028)	(0,0000)	(0,0000)	(0,0000)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	(0,9982)	
(0,6;1,2;4,3)	FDR real	0,0102	0,0385	0,1123	0,2110	0,4410	0,6616	0,7805	0,1521	0,3007
	EMVb	(0,0000)	(0,0020)	(0,0997)	(0,1314)	(0,1778)	(0,2415)	(0,3814)	(0,5194)	(0,6207)
		0,0008	0,0029	0,1055	0,1443	0,2019	0,2849	0,4155	0,6007	0,7062
		(0,0032)	(0,0042)	(0,1141)	(0,1586)	(0,2192)	(0,3084)	(0,4439)	(0,6494)	(0,7714)
	EMVf	(0,0000)	(0,0002)	(0,1061)	(0,1395)	(0,1877)	(0,2548)	(0,3745)	(0,5406)	(0,6539)
		0,0000	0,0007	0,1159	0,1530	0,2071	0,2895	0,4168	0,6078	0,7159
		(0,0000)	(0,0018)	(0,1262)	(0,1682)	(0,2272)	(0,3148)	(0,4457)	(0,6530)	(0,7724)
	SiZer	(0,0727)	(0,1294)	(0,2342)	(0,2835)	(0,3563)	(0,4806)	(0,6788)	(0,9640)	(0,9696)
		0,0759	0,1347	0,2472	0,3020	0,3907	0,5299	0,7917	0,9675	0,9807
		(0,0787)	(0,1405)	(0,2648)	(0,3284)	(0,4336)	(0,5796)	(0,9487)	(0,9796)	(0,9851)
EM	(0,0000)	(0,0000)	(0,2496)	(0,0000)	(0,9982)	(0,9982)	(0,9984)	(0,9982)	(0,9982)	
	0,0000	0,0000	0,9982	0,0000	0,9982	0,9982	0,9984	0,9982	0,9982	
	(0,0000)	(0,0000)	(0,9982)	(0,0000)	(0,9982)	(0,9982)	(0,9984)	(0,9982)	(0,9982)	

# Referências Bibliográficas

- [1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. John Wiley & Sons, New York, 10th edition, 1972.
- [2] M. Aitkin and D. B. Rubin. Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Series B*, 47(1):67–75, 1985.
- [3] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular Biology of the Cell*. Garland Publishing, New York and London, 4rd edition, 2002.
- [4] D. B. Allison, G. L. Gadbury, G. Heo, J. R. Fernández, C. K. Lee, T. A. Prolla, and R. Weindruch. A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics & Data Analysis*, 39:1–20, 2002.
- [5] F. Bartolucci and L. Scaccia. The use of mixtures for dealing with non-normal regression errors. *Computational Statistics & Data Analysis*, 48:821–834, 2005.
- [6] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1):289–300, 1995.
- [7] L. Birgé. On estimating a density using hellinger distance and some other strange facts. *Probability Theory and Related Fields*, 71(2):271–291, January 1986.
- [8] T. A. Brown. *Genomes*. BIOS Scientific Publishers, Ltd, Oxford, UK, 2002.

- [9] G. Casella and R. L. Berger. *Statistica Inference*. Duxbury - Thompson Learning, California, 2nd edition, 2002.
- [10] R. Chen, R. Goodman, and A. Zame. On the limiting distribution of two random sequences. *Journal of Multivariate Analysis*, 14:221–230, 1984.
- [11] C. Dalmasso, P. Broët, and T. Moreau. A simple procedure for estimating the false discovery rate. *Bioinformatics*, 21(5):660–668, 2005.
- [12] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society B*, 39:1–38, 1977.
- [13] K. A. Do and G. J. McLachlan. Estimation of mixture proportions: a case study. *Applied Statistics*, 33:134–140, 1984.
- [14] R. D. Drummond, A. Pinheiro, C. S. Rocha, and M. Menossi. Iser: Selection of differentially expressed genes from dna array data by nonlinear data transformation and local fitting. *Bioinformatics*, 21:4427–4429, 2005.
- [15] B. S. Everitt. *Introduction to optimization methods and their application in statistics*. Chapman and Hall, New York, 1st edition, 1987.
- [16] H. Exton. *Multiple hypergeometric functions and applications*. Ellis Howard, Chichester, 1st edition, 1976.
- [17] J. Fan and J. S. Marron. Fast implementations of nonparametric curves estimatos. *Journal of Computational an Graphical Statistics*, 3(1):35–36, March 1994.
- [18] J. M. Felix, R. D. Drummond, F. T. S. Nogueira, V. E. R. Junior, R. A. Jorge, P. Arruda, and M. Menossi. Genoma funcional. *Bio Tecnologia - Ciência e Desenvolvimento*, 24:60–67, 2001.

- [19] Z. Feng and C. E. McCulloch. Using bootstrap likelihood ratios in finite mixture models. *Journal of Royal Statistical Society - Series B*, 58(3):609–617, 1996.
- [20] R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *The Computer Journal*, 6:163–168, 1963.
- [21] W. M. Freeman, D. J. Robertson, and K. E. Vrana. Fundamentals of dna hybridization arrays for gene expression analysis. *Biotechniques*, 29:1042:1055, 2000.
- [22] J. K. Ghosh and P. K. Sen. *On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results*. Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (Vol. II), L.M. Le Cam and R.A. Olshen (Eds.), Monterey: Wadsworth, 1985.
- [23] A. J. F. Griffiths, W. M. Gelbart, J. H. Miller, and R. C. Lewontin. *Modern Genetic Analysis*. W. H. Freeman & Co, New York, 1999.
- [24] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, Oxford, 3rd edition, 2004.
- [25] H. L. Harter. Adaptive robust estimation of location and scale parameters of symmetric populations. Technical report, Wright-Patterson Air Force Base, OH, 1978. AFFDL-TR-78-128.
- [26] D. P. Helmbold, R. E. Schapire, Warmuth Singer, Y., and M. K. *A Comparison of New and Old Algorithms for a Mixture Estimation Problem*. Annual Workshop on Computational Learning Theory - Proceedings of the eighth annual conference on Computational learning theory, Santa Cruz, California, United States, 1995.
- [27] L. F. James, C. E. Priebe, and D. J. Marchette. Consistent estimation of mixture complexity. *The Annals of Statistics*, 29(5):1281–96, 2001.
- [28] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions - Volume 2*. John Wiley & Sons, USA, 2nd edition, 1995a.

- [29] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Univariate Discrete Distributions*. John Wiley & Sons, USA, 2nd edition, 1995b.
- [30] S. Jung and W. Jang. How accurately can we control the *fdr* in analyzing microarray data? *Biostatistics*, 22(14), 2006.
- [31] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2nd edition, 2000.
- [32] R. J. Karunamunia, T. N. Sriramb, and J. Wu. Rates of convergence of an adaptive kernel density estimator for finite mixture models. *Statistics & Probability Letters*, 76:221–230, 2006.
- [33] J. Karvanen. Estimation of quantile mixtures via *l*-moments and trimmed *l*-moments. *Computational Statistics & Data Analysis*, 2006. in press.
- [34] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York, 2 edition, 2001.
- [35] S. S. Li, J. Bigler, J. W. Lampe, J. D. Potter, and Z. Feng. *Fdr*-controlling testing procedures and sample size determination for microarrays. *Statistics in Medicine*, 24:2267–2280, 2005.
- [36] D. L. Libby and M. R. Novick. Multivariate generalized beta-distributions with applications to utility assessment. *Journal of Educational Statistics*, 7:271–294, 1982.
- [37] G. J. McLachlan. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 36(3):318–324, 1987.
- [38] G. J. McLachlan and K. E. Basford. *Mixture Models - Inference and Applications to Clustering*. Marcel Dekker, USA, 1988.
- [39] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, New York, 1 edition, 1997.

- [40] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics, USA, 2000.
- [41] M. Miloslavsky and M. J. van der Laan. Fitting of mixtures with unspecified number of components using cross validation distance estimate. *Computational Statistics & Data Analysis*, 41:413–428, 2003.
- [42] G. D. Murray and D. M. Titterton. Estimation problems with data from a mixture. *Applied Statistics*, 27(3):325–334, 1978.
- [43] E. A. Nadaraya. On nonparametric estimates of density functions and regression curves. *Theory of Probability and its Applications*, 10:186–190, 1965.
- [44] M. Neuhauser and F. C. Lam. *Nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments*. Proceedings of the Second Asia-Pacific Bioinformatics Conference (APBC2004), Dunedin, New Zealand; Chen, Y. P. (Ed), *Conferences in Research and Practice in Information Technology*, vol. 29, ACS, 139-143, 2004.
- [45] A. W. Norris and C. R. Kahn. Analysis of gene expression in pathophysiological states: Balancing false discovery and false negative rates. *Proceedings of the National Academy of Sciences of the United States of America*, 103(3):649–653, January 2006.
- [46] R. A. Parker and R. B. Rothenberg. Identifying important results from multiple statistical tests. *Statistics in Medicine*, 7:1031–1043, 1988.
- [47] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962.
- [48] T. Pham-Gia and Q. P. Duong. The generalized beta and f distributions in statistical modelling. *Mathematical and Computer Modelling*, 13:1613–1625, 1989.

- [49] T. Pham-Gia and N. Turkkan. Distributions of the linear combination of two general beta variables and applications. *Communications in Statistics - Theory and Methods*, 27(7):1851–1869, 1998.
- [50] D. Pokhyl’ko. Wavelet estimators of a density constructed from observations of a mixture. *Theory of Probability and Mathematical Statistics*, 70:135–145, 2005.
- [51] A. Polymenis and D. M. Titterington. On the determination of the number of components in a mixture. *Statistics & Probability Letters*, 38:295–298, 1998.
- [52] S. Pounds and S. W. Morrins. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics*, 19(10):1236–1242, 2003.
- [53] B. G. Quinn and G. J. McLachlan. A note on the aitkin-rubin approach to hypothesis testing in mixture models. *Journal of the Royal Statistical Society, Series B*, 49(3):311–314, 1987.
- [54] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26(2):195–239, April 1984.
- [55] A. Reiner, D. Yekutieli, and Y. Benjamini. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 19(3):368–375, 2003.
- [56] S. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 2nd edition, 1996.
- [57] B. W. Silverman. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics*, 6:177–184, 1978.
- [58] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1st edition, 1993.

- [59] C. Steinhilff, T. Müller, U. A. Nuber, and M. Vingron. Gaussian mixture density estimation applied to microarray data. *IDA*, pages 418–429, 2003.
- [60] T. Strachan and A. P. Read. *Human Molecular Genetics*. BIOS Scientific Publishers, Ltd, Oxford, UK, 1999.
- [61] M. W. Thompson, R. R. McInnes, and H. F. Willard. *Thompson & Thompson - Genética Médica*. Guanabara Koogan, Rio de Janeiro, 5<sup>a</sup> edition, 1993.
- [62] D. M. Titterton, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, Great Britain, 1985.
- [63] V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarray applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9):5116–5121, 2001.
- [64] T. Unger, Z. Korade, O. Lazarov, D. Terrano, S. S. Sisodia, and K. Mirnics. True and false discovery in dna microarray experiments: transcriptome changes in the hippocampus of presenilin 1 mutant mice. *Methods*, 37(3):261–73, November 2005.
- [65] N. A. Vlassis, G. Papakonstantinou, and P. Tsanakas. Mixture density estimation based on maximum likelihood and sequential test statistics. *Neural Processing Letters*, 9:63–76, 1999.
- [66] G. R. Walsh. *Methods of Optimization*, volume 6. John Wiley, London, 1975.
- [67] B. Wu, Z. Guan, and H. Zhao. Parametric and nonparametric fdr estimation revisited. *Biometrics*, 62:735–744, September 2006.
- [68] Q. Xiang, J. Edwards, and G. L. Gadbury. Interval estimation in a finite mixture model: Modeling p-values in multiple testing applications. *Computational Statistics & Data Analysis*, 2006. in press.

- [69] J. J. Yang and M. C. K. Yang. An improved procedure for gene selection from microarray experiments using false discovery rate criterion. *BMC Bioinformatics*, 7(15), 2006.

**Índice Onomástico**

- Abramowitz, M., 5  
Aitkin, M., 42  
Alberts, B., 18, 20–22  
Allison, D. B., 37  
Arruda, P., 22, 23, 25  
  
Balakrishnan, N., 5–7, 9, 11  
Bartolucci, F., 42  
Basford, K. E., 35, 39, 42  
Benjamini, Y., 23, 29–31  
Berger, R. L., 43  
Bigler, J., 2  
Birgé, L., 73  
Bray, D., 18, 20–22  
Broët, P., 33  
Brown, T. A., 18, 20–22, 24  
  
Casella, G., 43–45, 53  
Chaudhuri, P., 3, 62, 65  
Chen, R., 2, 8, 11–13  
Chu, G., 3  
  
Dalmaso, C., 33  
Dempster, A. P., 50  
Do, K. A., 38  
Drummond, R. D., 22, 23, 25, 75  
Duong, Q. P., 2, 8, 11, 13, 14, 46, 71  
Edwards, J., 3, 30, 33  
  
Everitt, B. S., 48, 54–56  
Exton, H., 11, 12  
  
Fan, J., 65, 67  
Felix, J. M., 22, 23, 25  
Feng, Z., 2, 36  
Fernández, J. R., 37  
Fletcher, R., 57  
Freeman, W. M., 23  
  
Gadbury, G. L., 3, 30, 33, 37  
Gelbart, W. M., 17, 18  
Ghosh, J. K., 36  
Goodman, R., 2, 8, 11–13  
Griffiths, A. J. F., 17, 18  
Grimmett, G., 68  
Guan, Z., 2  
  
Harter, H. L., 6  
Helmbold, 42  
Heo, G., 37  
Hochberg, Y., 29  
  
James, L. F., 37  
Jang, W., 2  
Johnson, N. L., 5–7, 9, 11  
Jorge, R. A., 22, 23, 25  
Jung, S., 2

- Junior, V. E. R., 22, 23, 25
- Kahn, C. R., 2, 28
- Karatzas, I., 68
- Karunamunia, R. J., 42
- Karvanen, J., 43
- Korade, Z., 2, 28
- Kotz, S., 5–7, 9, 11
- Krishnan, T., 50, 51, 53–56
- Laird, N. M., 50
- Lam, F. C., 3
- Lampe, J. W., 2
- Lazarov, O., 2, 28
- Lee, C. K., 37
- Lehmann, E. L., 44, 45, 53
- Lewis, J., 18, 20–22
- Lewontin, R. C., 17, 18
- Li, S. S., 2
- Libby, D. L., 2, 8
- Müller, T., 37, 81
- Maclachlan, G. J., 38, 39, 41, 52
- Makov, U. E., 34, 35, 42
- Marchette, D. J., 37
- Marron, J. S., 3, 62, 65, 67
- McCulloch, C. E., 36
- McInnes, R. R., 17–22
- McLachlan, G. J., 35, 36, 38, 39, 42, 50,  
51, 53–56
- Menossi, M., 22, 23, 25, 75
- Miller, J. H., 17, 18
- Miloslavsky, M., 35, 36
- Mirnics, K., 2, 28
- Moreau, T., 33
- Morrins, S. W., 31
- Murray, G. D., 41
- Nadaraya, E. A., 61
- Neuhauser, M., 3
- Nogueira, F. T. S., 22, 23, 25
- Norris, A. W., 2, 28
- Novick, M. R., 2, 8
- Nuber, U. A., 37, 81
- Papakonstantinou, G., 37, 41
- Parker, R. A., 2, 28, 33, 36
- Parzen, E., 61
- Peel, D., 35, 38, 39, 41, 52
- Pham-Gia, T., 2, 5, 7, 8, 11, 13, 14, 46,  
71
- Pinheiro, A., 75
- Pokhyl'ko, D., 43
- Polymenis, A., 37
- Potter, J. D., 2
- Pounds, S., 31
- Powell, M. J. D., 57
- Priebe, C. E., 37
- Prolla, T. A., 37

- Quinn, B. G., 36
- Raff, M., 18, 20–22
- Read, A. P., 23, 24
- Redner, R. A., 39
- Reiner, A., 23, 30, 31
- Roberts, K., 18, 20–22
- Robertson, D. J., 23
- Rocha, C. S., 75
- Rosenblatt, M., 68
- Ross, S., 68
- Rothenberg, R. B., 2, 28, 33, 36
- Rubin, D. B., 42, 50
- Scaccia, L., 42
- Schreve, S. E., 68
- Sen, P. K., 36
- Silverman, B. W., 59–61, 69
- Sisodia, S. S., 2, 28
- Smith, A. F. M., 34, 35, 42
- Sriramb, T. N., 42
- Stegun, I. A., 5
- Steinholff, C., 37, 81
- Stirzaker, D., 68
- Strachan, T., 23, 24
- Terrano, D., 2, 28
- Thompson, M. W., 17–22
- Tibshirani, R., 3
- Titterington, D. M., 34, 35, 37, 41, 42
- Tsanakas, P., 37, 41
- Turkkan, N., 5, 7
- Tusher, V. G., 3
- Unger, T., 2, 28
- van der Laan, M. J., 35, 36
- Vingron, M., 37, 81
- Vlassis, N. A., 37, 41
- Vrana, K. E., 23
- Walker, H. F., 39
- Walsh, G. R., 48
- Watson, J. D., 18, 20–22
- Weindruch, R., 37
- Willard, H. F., 17–22
- Wu, B., 2
- Wu, J., 42
- Xiang, Q., 3, 30, 33
- Yang, J. J., 2, 29
- Yang, M. C. K., 2, 29
- Yekutieli, D., 23, 30, 31
- Zame, A., 2, 8, 11–13
- Zhao, H., 2

## Índice Remissivo

- Algoritmo EM, 42, 50–54, 72, 73, 75, 81
- Beta generalizada, 8, 13, 81
- Bonferroni, 30
- Bootstrap, 3, 36, 37, 42, 66
- Browniano
- movimento, 68
  - ponte, 68, 69
- Códon, 21
- Comparações múltiplas, 1, 27, 28
- DNA, 17–20, 22–25
- Arranjo, 22–25
- Erro
- tipo I, 28
- Expansão de Taylor, 54
- Expressão gênica, 18, 22–24, 27, 37
- F generalizada, 14
- FDR, 27–31, 33, 73, 75, 82
- Fibonacci, 48, 72, 74, 82
- Função hipergeométrica, 9, 11, 13
- Função núcleo, 3, 38, 59, 60, 62–64, 67
- FWER, 29–31
- Gene, 1, 2, 18, 20, 22–25
- Informação de Fisher, 15, 37, 46
- Integral de Euler, 11, 12
- Kernel, 37, 41, 42
- Macroarranjo, 23, 25
- Matriz Hessiana, 55
- Microarranjo, 1–3, 23–25, 27, 30–33, 37
- Newton-Raphson, 7, 42, 50, 55, 56
- p-valor, 1, 2, 28, 30–33, 71, 75, 81
- PCER, 29
- Pochhammer, 9
- Processo Gaussiano, 68
- Projeto Genoma, 23
- RNA, 2, 18–20, 22, 23
- SiZer, 3, 62, 64, 65, 67, 72, 74, 75, 81, 82
- Teorema da Soma de Gauss, 11
- Verossimilhança, 7, 37, 42, 48, 50
- equação, 42, 45, 46
  - estimador de máxima, 41, 43, 47, 48, 50, 52, 55, 72, 81, 82
  - função, 42, 43, 51
  - log-verossimilhança, 15, 46, 51, 52, 55, 56, 73
  - máxima, 7, 15, 33, 41–43, 50, 51, 53
  - razão de, 36

Wavelet, 43