

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Classificação de Biopotenciais
via Cadeias de Markov Ocultas

por

Iara Moreira Frondana

Orientador: Prof. George Freitas von Borries, PhD

Fevereiro de 2012

Iara Moreira Frondana

Classificação de Biopotenciais via Cadeias de Markov Ocultas

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

Universidade de Brasília
Brasília, Fevereiro de 2012

TERMO DE APROVAÇÃO

Iara Moreira Frondana

CLASSIFICAÇÃO DE BIOPOTENCIAIS
VIA CADEIAS DE MARKOV OCULTAS

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

Data da defesa: 17 de Fevereiro de 2012

Orientador:

Prof. George Freitas von Borries, PhD
Departamento de Estatística, UnB

Comissão Examinadora:

Profa. Dra. Cira Etheowalda Guevara Otiniano
Departamento de Estatística, UnB

Profa. Dra. Florencia Graciela Leonardi
Instituto de Matemática e Estatística, USP

Brasília, Fevereiro de 2012

Ficha Catalográfica

FRONDANA, IARA MOREIRA

Classificação de biopotenciais via cadeias de Markov ocultas,
(UnB - IE, Mestre em Estatística, 2011).

Dissertação de Mestrado - Universidade de Brasília. Departamento de Estatística
- Instituto de Ciências Exatas.

1. biopotencial 2. EEG 3. cadeia de Markov oculta (*hidden Markov models*
- HMM) 4. estimação 5. classificação 6. R em paralelo

É concedida à Universidade de Brasília a permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de Mestrado pode ser reproduzida sem a autorização por escrito do autor.

Iara Moreira Frondana

Ao meu amado marido Raphael.

Aos meus pais e irmãs.

Agradecimentos

Agradeço ao meu orientador, prof. George, por toda dedicação, entusiasmo e paciência oferecidos ao longo deste projeto. Certamente este trabalho não seria o que é sem seu constante apoio e direcionamento. E não posso deixar de agradecer a todos os meus queridos professores que, com dom e sabedoria, se tornam grandes inspirações.

Também agradeço aos meus pais por me ensinarem que enquanto todos os bens materiais podem ser perdidos, a educação e o saber nunca o são. Por terem me oferecido todo o suporte quando precisei.

E, finalmente, agradeço ao meu marido por ouvir, pacientemente, todos os devaneios, planos e angústias, me encorajando e amando mesmo quando não compreendia. Por sempre apoiar meus projetos, comemorar meus sucessos e me consolar quando algo não acontece como o planejado. Obrigada por estar sempre ao meu lado.

Sumário

Lista de Figuras	5
Lista de Tabelas	8
Resumo	9
1 Introdução	12
2 Conceitos Básicos	15
2.1 Modelo de Mistura	15
2.1.1 Formalização	16
2.1.2 Mistura de Normais	17
2.2 Cadeias de Markov	18
2.2.1 Conceitos Importantes	18
2.2.2 Inferência em Cadeias de Markov	22
3 Modelo de Markov Oculto	29
3.1 Exemplos	30
3.2 Formalização	32
3.2.1 HMM básico	32
3.2.2 Distribuição marginal	33
3.2.3 Função de verossimilhança	34
3.2.4 Probabilidades <i>forward</i> e <i>backward</i>	35
3.3 Considerações	37
4 Estimação e Classificação	39
4.1 Estimação	39

4.1.1	Algoritmo EM	40
4.1.2	Algoritmo EM para o HMM	40
4.1.3	O problema do máximo local	44
4.2	Seleção do modelo	47
4.2.1	AIC e BIC	49
4.3	Classificação	50
4.3.1	Método para a classificação	51
5	Simulação	54
5.1	Programação em Paralelo	54
5.2	Parâmetros do modelo HMM e escolha dos valores de T e S	57
5.3	Avaliação da estimação	59
5.3.1	Ordem $m = 1$	59
5.3.2	Ordem $m = 2$	61
5.3.3	Ordem $m = 3$	65
5.3.4	Ordem $m = 4$	74
5.3.5	Ordem $m = 5$	81
5.3.6	Comentários gerais	90
5.4	Avaliação dos critérios AIC e BIC	90
6	Estimação e Classificação dos Dados Reais	105
6.1	Base de Dados	105
6.2	Estimação	107
6.2.1	Estimativas dos parâmetros: Figura 6 - Eletrodo 69	114
6.3	Classificação	115
7	Conclusões e Trabalhos Futuros	123
7.1	Conclusões	123
7.2	Trabalhos futuros	125
	Referências Bibliográficas	127
A	Programas R	131
A.1	Programa 1: <code>r.normal.hmm</code>	131

A.2	Programa 2: <code>matriz.P</code>	132
A.3	Programa 3: <code>veross</code>	133
A.4	Programa 3: <code>log.alpha.beta.norm.uni</code>	134
A.5	Programa 3: <code>EM.hmm.norm.uni</code>	135
A.6	Programa 3: <code>r.dirichlet</code>	139
A.7	Programa 3: <code>trans.ergotica</code>	139
A.8	Programa 3: <code>mu.sigma2.iniciais</code>	140
A.9	Programa 3: <code>avalia.EM.uni</code>	141
A.10	Programa 3: <code>estima.cluster</code>	144
A.11	Programa 3: <code>avalia.AICeBIC.cluster</code>	145
A.12	Programa 3: <code>avalia.estima</code>	145
A.13	Programa 3: <code>avalia.AICeBIC</code>	148

Lista de Figuras

2.1	Cadeia de Markov.	19
3.1	Esquema para o sorteio das bolas.	30
3.2	Diagrama da relação de dependência condicional em um HMM	33
4.1	Histograma de uma amostra dividida em 3 grupos, com 10 conjuntos de pontos iniciais para as médias μ_i	46
5.1	Exemplo da distribuição de tarefas em diferentes processadores. . . .	55
5.2	Box-plots das estimativas e intervalos para a média, com $m = 1$. . .	60
5.3	Box-plots das estimativas e intervalos para a variância, com $m = 1$. .	61
5.4	Box-plots das estimativas e intervalos para as médias μ_1 e μ_2 , com $m = 2$	63
5.5	Box-plots das estimativas e intervalos para as variâncias σ_1^2 e σ_2^2 , com $m = 2$	64
5.6	Box-plots das estimativas e intervalos para as distribuições iniciais δ_1 e δ_2 , com $m = 2$	65
5.7	Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{1.}$, com $m = 2$	68
5.8	Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{2.}$, com $m = 2$	70
5.9	Box-plots das estimativas e intervalos para as médias μ_1 , μ_2 e μ_3 , com $m = 3$	96
5.10	Box-plots das estimativas e intervalos para as variâncias σ_1^2 , σ_2^2 e σ_3^2 , com $m = 3$	97
5.11	Box-plots das estimativas e intervalos para as distribuições iniciais δ_1 , δ_2 e δ_3 , com $m = 3$	98

5.12	Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{1.}$, com $m = 3$	99
5.13	Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{2.}$, com $m = 3$	100
5.14	Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{3.}$, com $m = 3$	101
5.15	Proporção de acertos dos critérios AIC e BIC para $m = 1$	102
5.16	Proporção de acertos dos critérios AIC e BIC para $m = 2$	102
5.17	Proporção de acertos dos critérios AIC e BIC para $m = 3$	103
5.18	Proporção de acertos dos critérios AIC e BIC para $m = 4$	103
5.19	Proporção de acertos dos critérios AIC e BIC para $m = 5$	104
6.1	Coleta dos dados EEG pelo BIM Lab - UTEP.	105
6.2	Figuras observadas durante o experimento para a coleta dos sinais EEG.106	
6.3	Posicionamento dos eletrodos no escalpo (visto de cima).	108
6.4	Valores dos critérios AIC e BIC para a estimação do modelo da figura 10 no eletrodo 96, com $S = 1$ e $S = 2$. Os pontos em vermelho indicam o valor mínimo obtido em cada critério.	109
6.5	Classes da classificação c.2.	115
6.6	Classes da classificação c.3.	115
6.7	Agrupamento dos eletrodos para nova classificação das sequências. . .	122

Lista de Tabelas

3.1	Resultados obtidos para o Exemplo 3.1.	30
3.2	Resultados obtidos para o Exemplo 3.1, agora com o processo oculto.	31
3.3	Resultados obtidos para o Exemplo 3.2.	31
4.1	Resultados para 5 das 300 sequências geradas no Exemplo 4.1.	52
4.2	Resultados da classificação das 300 sequências do Exemplo 4.1.	52
5.1	Resumo dos valores obtidos para a média com $m = 1$	60
5.2	Resumo dos valores obtidos para a variância com $m = 1$	61
5.3	Resumo dos valores obtidos para as médias com $m = 2$	62
5.4	Resumo dos valores obtidos para a variâncias com $m = 2$	62
5.5	Resumo dos valores obtidos para a distribuição inicial com $m = 2$. . .	66
5.6	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{1,}$, com $m = 2$	66
5.7	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{2,}$, com $m = 2$	67
5.8	Resumo dos valores obtidos para as médias com $m = 3$	67
5.9	Resumo dos valores obtidos para a variâncias com $m = 3$	69
5.10	Resumo dos valores obtidos para a distribuição inicial com $m = 3$. . .	69
5.11	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{1,}$, com $m = 3$	71
5.12	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{2,}$, com $m = 3$	72
5.13	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{3,}$, com $m = 3$	73

5.14	Resumo dos valores obtidos para as médias com $m = 4$	74
5.15	Resumo dos valores obtidos para a variâncias com $m = 4$	75
5.16	Resumo dos valores obtidos para a distribuição inicial com $m = 4$. . .	76
5.17	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{1.}$, com $m = 4$	77
5.18	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{2.}$, com $m = 4$	78
5.19	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{3.}$, com $m = 4$	79
5.20	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{4.}$, com $m = 4$	80
5.21	Resumo dos valores obtidos para as médias com $m = 5$	82
5.22	Resumo dos valores obtidos para a variâncias com $m = 5$	83
5.23	Resumo dos valores obtidos para a distribuição inicial com $m = 5$. . .	84
5.24	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{1.}$, com $m = 5$	85
5.25	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{2.}$, com $m = 5$	86
5.26	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{3.}$, com $m = 5$	87
5.27	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{4.}$, com $m = 5$	88
5.28	Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{5.}$, com $m = 5$	89
5.29	Resultado da avaliação para o critério AIC para $T = 100$ fixo, e $S = 1$ e $S = 2$	91
5.30	Resultado da avaliação para o critério AIC para $T = 250$ fixo, e $S = 1$ e $S = 2$	92
5.31	Resultado da avaliação para o critério BIC para $T = 100$ fixo, e $S = 1$ e $S = 2$	93
5.32	Resultado da avaliação para o critério BIC para $T = 250$ fixo, e $S = 1$ e $S = 2$	94

6.1	Ordens dos modelos, obtidas pelo critério AIC, para os eletrodos 69 ao 74 e 78 ao 79.	110
6.2	Ordens dos modelos, obtidas pelo critério AIC, para os eletrodos 80 ao 73 e 91 ao 94.	111
6.3	Ordens dos modelos, obtidas pelo critério AIC, para os eletrodos 95, 96 e 101 ao 103.	112
6.4	Frequência das ordens dos modelos em cada figura, separadas por $S = 1$ e $S = 2$	113
6.5	Classificação das 30 sequências, eletrodo 70 com $S = 1$. Classes c.1. .	116
6.6	Classificação das 20 sequências, eletrodo 70 com $S = 2$. Classes c.1. .	117
6.7	Classificação das 30 sequências, eletrodo 70 com $S = 1$. Classes c.2. .	117
6.8	Classificação das 20 sequências, eletrodo 70 com $S = 2$. Classes c.2. .	118
6.9	Classificação das 30 sequências, eletrodo 70 com $S = 1$. Classes c.3. .	118
6.10	Classificação das 30 sequências, eletrodo 70 com $S = 2$. Classes c.3. .	118
6.11	Porcentagem de classificação correta dos eletrodos para as diferentes classes c.1, c.2 e c.3, nos casos $S = 1$ e $S = 2$	120
6.12	Porcentagem de classificação correta nos casos $S = 1$ e $S = 2$, utilizando as somas e as somas ponderadas das verossimilhanças.	121

Resumo

Este trabalho concentra-se na investigação do uso de Modelos de Markov Ocultos (HMM) para classificação de sinais de eletroencefalografia (EEG) obtidos por estímulos visuais. EEG é um biopotencial que possui grande aplicação na área médica, tanto para diagnósticos quanto para a investigação de sistemas de interface entre cérebro e computador (BCI - *Brain-Computer Interfaces*).

A bibliografia recente apresenta vários estudos sobre aplicação de modelos HMM para classificação de sinais de EEG e mostram que o HMM é adequado por ser um modelo genérico que abrange diversas complexidades que surgem em dados de séries temporais.

Este trabalho está dividido em três partes. Primeiro é apresentado o modelo HMM, o processo de estimação a ser utilizado e os critérios AIC e BIC para seleção dos modelos. A segunda parte apresenta estudos de simulação que utilizam diferentes condições amostrais para avaliar o procedimento de estimação, qualidade das estimativas e critérios de seleção. Finalmente, é feito o uso de um processo de estimação que inicializa o algoritmo EM com diferentes conjuntos de pontos iniciais aleatórios, para classificar sinais de EEG. Os sinais foram obtidos de estímulos visuais num estudo experimental conduzido por pesquisadores da Universidade do Texas em El Paso.

As simulações realizadas mostram que os resultados obtidos para os estimadores são adequados e o critério AIC é superior ao BIC na indicação da ordem apropriada para o modelo. O estudo com dados experimentais apresenta indicações de que o modelo HMM consegue identificar as diferenças entre os sinais captados quando são comparados grupos de estímulos visuais envolvendo figuras abstradas e expressões aritméticas.

Palavras Chave: *Biopotencial, EEG, Modelo de Markov Oculto, Estimação,*

Classificação, R em paralelo.

Abstract

This work focus in the investigation of Hidden Markov Models (HMM) applied to classification of electroencephalograph (EEG) data obtained by visual stimuli. EEG is a biopotencial signal with large application in medical fiel, as in diagnosis of diseases and investigation of brain-computer interfaces (BCI).

The recent literature presents several studies on application od HMM models for classification of EEG signals and show that the HMM is suitable for being a generic model that includes several complexities arising in time series data.

This work is divided into three main parts. First, it introduces the HMM model, the estimation process to be used and the AIC and BIC criteria for choosing models. Then the work presents simulation studies using different sampling conditions for evaluating the estimation procedure, quality of the estimates and selection criteria. Finally, it is used an estimation process for classification od EEG signals using different sets of random starting points for initialization of an EM algorithm. For the application study, EEG signals were obtained from visual stimuli in an experimental study conducted by researchers at the University of Texas at El Paso.

The simulation study showed that the results obtained are suitable for the estimators and the AIC criterion is superior to the BIC criteria when indicating the proper order for the model. The application in experimental data indicated that the HMM model can identify differences between signals obtained from visual stimuli involving abstract figures and arithmetic expressions.

key words: *Biopotencial, EEG, Hidden Markov Models, Estimation, Classification, Parallel R.*

Capítulo 1

Introdução

Este trabalho faz uma avaliação do uso de cadeias de Markov ocultas para classificação de sinais biopotenciais emitidos pelo cérebro. Biopotencial é a medida da eletricidade emitida por um determinado órgão ao realizar uma certa atividade. Exemplos de biopotenciais são o eletrocardiograma (ECG), eletromiograma (EMG) e eletroencefalograma (EEG). O estudo de biopotenciais gerados pelo cérebro (EEG) tem importância crescente em pesquisas de interfaces cérebro-computador (*brain-computer interfaces* - *BCI*). Uma grande vantagem do EEG é a possibilidade de medição sem a utilização de técnicas invasivas.

A técnica *Steady State Visually Evoked Potential (SSVEP)*, por exemplo, utiliza flashes de luz para obter sinais elétricos emitidos pelo cérebro durante a observação de imagens previamente selecionadas. Os sinais são coletados por meio de eletrodos posicionados no escalpo do indivíduo, sem intervenção cirúrgica. Um exemplo da grande atenção dada a esta área de pesquisa é a recente chamada (primeiro semestre de 2011) da revista *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS)* em que 9 dos 14 tópicos listados estão relacionados a estudos biopotenciais e *brain-computer interface*.

A utilização dos sinais de EEG tem sido explorada para diversas aplicações. Na área médica são utilizados em diagnósticos clínicos, como a identificação de pessoas com epilepsia ou distúrbios do sono, por exemplo. Há também quem vislumbre a implementação dos sinais de EEG na área militar, para o desenvolvimento de máquinas que exijam um número de controles além da capacidade do operador.

Contudo, as aplicações que mais se destacam, por seu óbvio benefício, são as que envolvem pessoas com paralisias em membros e tronco. Ressaltam-se os casos extremos, conhecidos na área médica por *locked in*, em que a paralisia impede até mesmo o processo da fala, cortando toda a comunicação destas pessoas com o mundo a sua volta. Nestes casos a classificação dos sinais EEG pode ser usada para detectar as intenções da pessoa e, assim, definir comandos de controle.

Os estudos sobre a classificação dos sinais EEG tentam então resolver questões que envolvem diagnósticos, e estabelecer a interação homem-máquina nos sistemas *BCI (brain-computer interface)*.

O objetivo deste trabalho é estudar a classificação dos sinais EEG utilizando os modelos de Markov oculto (*hidden Markov models - HMM*). Cappé et al. [4] ressalta o uso de estados ocultos para tornar o modelo mais genérico, capaz de englobar diferentes complexidades que surgem em dados de séries temporais reais. Cappé et al. [4] também observa que a estrutura simples da cadeia de Markov permite eficiência aos processos computacionais. Os artigos de Xu et al. [38], Bengio e Chiappa [2], Zhong e Ghosh [39], Obermaier et al. [21] e Haung et al. [13] são alguns exemplos de trabalhos que utilizam o modelo HMM e suas variações para classificação de sinais EEG.

Aqui serão analisados dados coletados no *Biopotentials Imaging Laboratory* (BIM Lab) da Universidade do Texas em El Paso (UTEP) utilizando a técnica SS-VEP. Os sinais são obtidos por 128 eletrodos através de estímulos visuais envolvendo figuras abstratas e operações aritméticas. O protocolo do estudo pode ser obtido sob consulta¹.

Outros estudos classificam os sinais em duas ou três classes, onde em geral pede-se que o participante da pesquisa pense em alguma atividade ou operação aritmética. Em Xu et al. [38], por exemplo, as atividades eram imaginar movimento com a mão direita e mão esquerda. Este trabalho pretende classificar os sinais em diferentes classes, de maneira a identificar as 10 figuras, ou classes de figuras, observadas durante a captação dos sinais.

Contudo, para que seja possível realizar a classificação dos sinais deve-se pri-

¹O Capítulo 6 apresenta uma breve descrição do experimento. Detalhes podem ser conseguidos através dos e-mails *iara.frondana@gmail.com* e *gborries@unb.br*

meio estimar os modelos que serão utilizados como base para a classificação. No momento da estimação deve-se avaliar a estrutura dos modelos: sua complexidade (número de parâmetros do modelo), a distribuição de probabilidade adequada para os dados e seus parâmetros, e o resultado encontrado para a cadeia de Markov relacionada ao modelo. Os Capítulos 2, 3 e 4 descrevem os detalhes destas etapas.

Neste trabalho é proposto o algoritmo EM para a realização da estimação do modelo, e estuda-se este método por meio de simulações para avaliar seu desempenho considerando diferentes parâmetros amostrais. Também por meio de simulações são estudados os critérios AIC e BIC de seleção dos modelos para que se conheça a performance desses critérios em diferentes condições amostrais.

Os resultados explorados neste trabalho, por meio de simulações, servirão como treinamento para a concepção de estudos de classificação futuros, a serem realizados por pesquisadores do BIM Lab, UTEP. São feitas sugestões para que novos estudos sejam capazes de esclarecer pontos que não puderam ser explorados aqui.

Este trabalho faz parte de um conjunto de trabalhos (Araújo [1], Matos [15], Silva [35] e Rocha [29]) realizados pelo grupo de pesquisa em biopotenciais do departamento de estatística da UnB. Estes trabalhos procuram conhecer melhor os dados dos sinais de EEG, com o intuito de gerar novas soluções para os problemas e incentivar novas aplicações com sinais de EEG e sinais biopotenciais em geral.

O presente trabalho está organizado da seguinte maneira: o Capítulo 2 apresenta conhecimentos básicos sobre modelos de misturas e cadeias de Markov necessários para a melhor compreensão do modelo de Markov oculto. O Capítulo 3 introduz os conceitos do modelo de Markov oculto. O Capítulo 4 apresenta o algoritmo EM como método de estimação a ser utilizado, medidas para a escolha do modelo mais apropriado, e o método usado para a classificação dos sinais. O Capítulo 5 conta a maneira como as simulações foram realizadas e os resultados obtidos. Já o Capítulo 6 descreve o experimento realizado para a obtenção dos dados, a aplicação do HMM nestes, os resultados obtidos na estimação dos modelos e na classificação dos sinais. E por fim, no Capítulo 7 são apresentadas as conclusões e sugestões de estudos futuros.

Capítulo 2

Conceitos Básicos

Este capítulo introduz conceitos básicos sobre modelo de misturas, mistura de distribuições, e cadeias de Markov que são necessários para a melhor compreensão da estrutura e idéia dos modelos de Markov Oculto (*Hidden Markov Models* - HMM). Diferentes autores descrevem a relação entre os modelos de mistura e HMM de diversas maneiras, para Zucchini e MacDonald [40] a distribuição marginal do HMM é uma mistura, Chambaz et al [6] destacam que a mistura com observações independentes é um caso particular do HMM e, McLachlan e Peel [18] citam ainda o HMM como uma alternativa para desenvolver uma mistura com uma estrutura de dependência entre as observações. Essa estrutura de dependência é, por sua vez, oferecida pelas cadeias de Markov, o processo oculto do modelo que gera os valores observados no HMM.

A seguir apresentam-se as definições básicas de modelo de mistura, um exemplo explicativo e a definição de mistura de densidades normais.

2.1 Modelo de Mistura

Um modelo de mistura é um modelo probabilístico para representar a presença de sub-populações dentro de uma população maior sem que se necessite identificar de qual sub-população as observações se originam. Formalmente, um modelo de mistura corresponde a uma mistura de distribuições que descreve a distribuição de probabilidades das observações da população geral.

Os modelos de mistura começaram a ser estudados e utilizados em análises de modelos biológicos há mais de cem anos. Em 1894 Pearson [23] analisou a razão entre os comprimentos da testa e do corpo de carangueijos encontrados na Baía de Nápoles. Para este conjunto de dados foi ajustado um modelo de mistura com duas distribuições normais com médias e variâncias diferentes. Esta análise, que se tornou clássica, indicou a presença de subespécies na população de carangueijos estudada.

O método de estimação utilizado por Pearson [23] em seu estudo foi o método dos momentos. Desde então, com os avanços computacionais, outras técnicas foram desenvolvidas. A mais utilizada e conhecida é o algoritmo EM (*Expectation-Maximization*), também utilizado para estimação em outros modelos, como o HMM. No Capítulo 4 o conceito do algoritmo EM é apresentado, assim como sua aplicação na estimação do modelo HMM.

2.1.1 Formalização

Após situar historicamente o surgimento do modelo de mistura, pode-se apresentar sua definição básica.

Definição 2.1.1. (*Modelo de Mistura*)

Seja X uma variável aleatória, ou vetor aleatório, com função densidade $f(x)$, no caso contínuo. Suponha que, $\forall x \in \Omega$ (Ω espaço amostral de X), $f(x)$ de X pode ser escrita na forma

$$f(x) = \sum_{i=1}^m \pi_i f_i(x) \quad (2.1)$$

onde: $f_i(x)$, $i = 1, \dots, m$, são densidades,

$$0 \leq \pi_i \leq 1, \text{ para } i = 1, \dots, m$$

$$\text{e } \sum_{i=1}^m \pi_i = 1.$$

A função $f(x)$ é assim definida como modelo de mistura.

Todas as considerações são as mesmas para o caso discreto, substituindo as funções densidades $f(x)$ e $f_i(x)$ pelas funções de probabilidade $p(x)$ e $p_i(x)$.

A interpretação deste modelo pode ser simplificada por meio de um exemplo.

Exemplo 2.1. *Suponha que uma população de animais possua 3 subespécies ($m = 3$), A , B e C , e deseja-se estudar uma de suas características, altura por exemplo. Mas, não é possível identificar a qual subespécie (A , B ou C) os animais pertencem, e sabe-se que em relação a altura as subespécies apresentam diferenças. Por exemplo, os animais da subespécie A tendem a ser mais altos, os da C mais baixos, já a altura dos animais da subespécie B tem maior variabilidade. Essas particularidades definem funções densidades $f_i(x)$, $i = 1, 2, 3$, diferentes, uma para cada subespécie.*

A prevalência, ou proporção, π_1 , π_2 e π_3 , de cada subespécie dentro da população também deve, obviamente, ser levada em consideração, pois influencia na distribuição da altura de toda a população. Então, a Equação (2.1) pode ser utilizada para descrever a densidade da altura da população como uma combinação, ou mistura, das densidades das alturas das subespécies, ponderadas por suas proporções.

2.1.2 Mistura de Normais

As misturas podem envolver distribuições de probabilidade da mesma classe, com parâmetros diferentes, ou mistura de diferentes distribuições de probabilidade. A mistura de normais é a mais conhecida das misturas, não somente por ter sido a primeira a ser estudada, mas também por sua versatilidade. Outro tipo de mistura interessante é a de densidades α -estáveis simétricas (Valdés [36]) que permite englobar uma gama variada de densidades no processo, como a Cauchy, a Lévy, a normal e outras distribuições cujas funções densidades não tem forma fechada simples, mas suas funções características são bem conhecidas.

McLachlan e Peel [18] indicam a utilização de mistura de normais para aproximar distribuições contínuas mais complexas, e Priebe [25], por exemplo, mostra uma boa aproximação de uma amostra de tamanho 10.000 originada por uma densidade log-normal utilizando uma mistura de 30 densidades normais.

Definição 2.1.2. *Uma mistura de m densidades normais com médias e variâncias diferentes é dada por*

$$f(x) = \sum_{i=1}^m \pi_i \phi_i(x; \mu_i, \sigma_i^2), \quad (2.2)$$

onde: $\phi_i(x; \mu_i, \sigma_i^2) = (2\pi)^{-\frac{1}{2}} \sigma_i^{-1} \exp \left\{ -\frac{1}{2\sigma_i^2} (x - \mu_i)^2 \right\}$ é a densidade Normal de média μ_i e variância σ_i^2 .

Apresentados os conceitos de misturas, deve-se agora compreender algumas definições básicas em Cadeias de Markov e seu procedimento inferencial para depois introduzir os problemas envolvendo o HMM. A seção seguinte apresenta tais conceitos de forma direta e resumida.

2.2 Cadeias de Markov

Ross [32] e Durrett [7] são um excelente ponto de partida para aqueles que se interessam por Cadeias de Markov (CM) mas não possuem os conhecimentos básicos. Em nível introdutório, particularmente, recomenda-se Ross [32] por suas claras e detalhadas explicações. Os conceitos mais importantes são resumidos a seguir.

2.2.1 Conceitos Importantes

Seja $\{C_t, t = 1, 2, 3, \dots\}$ um processo estocástico que assume valores em um conjunto finito. Quando $C_t = i$ diz-se que o processo encontra-se no estado i no tempo t . Este processo é dito **Cadeia de Markov (CM)** de primeira ordem quando satisfaz, para todo $t \in \mathbb{N}$, a seguinte **propriedade de Markov**

$$\Pr(C_{t+1} | C_t, C_{t-1}, \dots, C_1) = \Pr(C_{t+1} | C_t). \quad (2.3)$$

Ou seja, a probabilidade condicional do próximo estado C_{t+1} dado todos os estados anteriores C_1, C_2, \dots, C_{t-1} e C_t , só depende do estado atual (C_t). A Figura 2.1 apresenta esta relação de dependência entre as variáveis.

A CM definida acima é uma cadeia de primeira ordem, pois depende apenas do estado imediatamente anterior. Há também cadeias de ordens superiores, i.e., $o = 2, 3, \dots$. Nestes casos a probabilidade condicional vai depender, dos o estados imediatamente anteriores. Neste texto será dada ênfase para cadeias de primeira ordem ($o = 1$).

Associada a CM a **probabilidade de transição** $\gamma_{ij}(t)$ mede a probabilidade

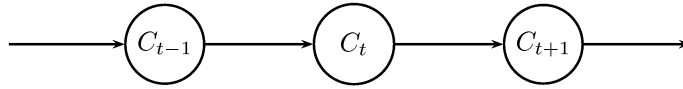


Figura 2.1: Cadeia de Markov.

do processo ir do estado i para o estado j em t passos, ou seja,

$$\gamma_{ij}(t) = \Pr(C_{s+t} = j \mid C_s = i). \quad (2.4)$$

Quando esta probabilidade não se altera para diferentes valores de s dizemos que a CM é **homogênea**.

A partir das probabilidades de transição a um passo, $\gamma_{ij}(1) = \gamma_{ij}$, é possível construir a **matriz de transição** $\Gamma(1) = \mathbf{\Gamma}$, cujo elemento da posição i, j corresponde a transição do estado i para o estado j . Esta matriz é quadrada e os elementos na linha somam 1, uma vez que a i -ésima linha corresponde a função de probabilidade condicional $\Pr(C_{t+1} = j \mid C_t = i)$ para todo j pertencente ao espaço de estados do processo.

Um exemplo básico e bastante conhecido na literatura pode ser utilizado para ilustrar esses conceitos.

Exemplo 2.2. *Suponha que o tempo (sol=1 e chuva=2) do dia de amanhã depende somente do tempo do dia de hoje e que, se hoje faz sol, então a probabilidade de também fazer sol amanhã é 0,7. Por outro lado, se hoje chove a probabilidade de amanhã também chover é 0,4. Assim obtemos:*

$$i) \gamma_{1,1} = \Pr(C_1 = 1 \mid C_0 = 1) = 0,7$$

$$ii) \gamma_{1,2} = \Pr(C_1 = 2 \mid C_0 = 1) = 1 - \Pr(C_1 = 1 \mid C_0 = 1) = 1 - 0,7 = 0,3$$

$$iii) \gamma_{2,2} = \Pr(C_1 = 2 \mid C_0 = 2) = 0,4$$

$$iv) \gamma_{2,1} = \Pr(C_1 = 1 \mid C_0 = 2) = 1 - \Pr(C_1 = 2 \mid C_0 = 2) = 1 - 0,4 = 0,6$$

Com as probabilidades calculadas em (i)-(iv) pode-se construir a matriz de transição:

$$\mathbf{\Gamma} = \begin{pmatrix} 0,7 & 0,3 \\ 0,6 & 0,4 \end{pmatrix}$$

Uma vez obtida $\mathbf{\Gamma} = \mathbf{\Gamma}(1)$, é possível mostrar que para cadeias de Markov homogêneas, as probabilidades de transição $\gamma_{ij}(t)$ da matriz $\mathbf{\Gamma}(t)$, no tempo t , são obtidas elevando $\mathbf{\Gamma}$ a t -ésima potência. Para isto basta utilizar as equações de Chapman-Kolmogorov,

$$\mathbf{\Gamma}(t+u) = \mathbf{\Gamma}(t)\mathbf{\Gamma}(u), \quad \text{para todos os tempos } t \text{ e } u \in \mathbb{N} \quad (2.5)$$

Estas idéias são descritas no seguinte teorema:

Teorema 2.2.1. *Para uma cadeia de Markov homogênea com matriz de transição $\mathbf{\Gamma}$*

$$\mathbf{\Gamma}(t) = \mathbf{\Gamma}(1)^t = \mathbf{\Gamma}^t$$

para todo $t \in \mathbb{N}$.

O Teorema 2.2.1 é importante para simplificar o cálculo das probabilidades de transição com muitos passos.

Exemplo 2.3. *Voltando ao Exemplo 2.2, suponha que o interesse está em saber se vai chover ou fazer sol daqui a dois dias (depois de amanhã). Assim*

$$\mathbf{\Gamma}(2) = \mathbf{\Gamma}^2 = \begin{pmatrix} 0,7 & 0,3 \\ 0,6 & 0,4 \end{pmatrix} \begin{pmatrix} 0,7 & 0,3 \\ 0,6 & 0,4 \end{pmatrix} = \begin{pmatrix} 0,67 & 0,33 \\ 0,68 & 0,32 \end{pmatrix}.$$

Logo, se hoje choveu ($C_0 = 2$) a probabilidade de que depois de amanhã também chova ($C_2 = 2$) é 0,32.

Mas, se o interesse estiver na probabilidade $\Pr(C_t = j)$ ao invés da probabilidade de transição, isso significa saber qual a probabilidade da cadeia estar em um determinado estado, j , no tempo t quando não se tem nenhuma informação sobre qualquer um dos estados anteriores a t . Para isso o vetor

$$\mathbf{u}(t) = \left(\Pr(C_t = 1), \Pr(C_t = 2), \dots, \Pr(C_t = m) \right), \quad t \in \mathbb{N}$$

é definido para uma cadeia com m estados. O vetor $\mathbf{u}(1)$ é conhecido como a **distribuição inicial** da CM.

Usando as probabilidades condicionais é possível mostrar que

$$\mathbf{u}(t) = \mathbf{u}(t-1)\mathbf{\Gamma}. \quad (2.6)$$

Logo, tendo a distribuição inicial $\mathbf{u}(1)$, a matriz de transição $\mathbf{\Gamma}$ e utilizando (2.6) repetidas vezes, é possível obter, para $t > 5$, por exemplo,

$$\begin{aligned} \mathbf{u}(t) &= \mathbf{u}(t-1)\mathbf{\Gamma} \\ \mathbf{u}(t) &= \mathbf{u}(t-2)\mathbf{\Gamma}\mathbf{\Gamma} = \mathbf{u}(t-2)\mathbf{\Gamma}^2 \\ \mathbf{u}(t) &= \mathbf{u}(t-3)\mathbf{\Gamma}\mathbf{\Gamma}^2 = \mathbf{u}(t-3)\mathbf{\Gamma}^3 \\ &\dots \\ \mathbf{u}(t) &= \mathbf{u}(2)\mathbf{\Gamma}\mathbf{\Gamma}^{t-3} = \mathbf{u}(2)\mathbf{\Gamma}^{t-2} \\ \mathbf{u}(t) &= \mathbf{u}(1)\mathbf{\Gamma}\mathbf{\Gamma}^{t-2} = \mathbf{u}(1)\mathbf{\Gamma}^{t-1} \end{aligned} \quad (2.7)$$

e portanto, $u(t) = u(1)\mathbf{\Gamma}^{t-1}$, $t > 1$.

Definição 2.2.1. (Distribuição estacionária) Um vetor $\boldsymbol{\delta}$, de elementos não negativos, é dito **distribuição estacionária** para uma cadeia de Markov com matriz de transição $\mathbf{\Gamma}$, se

$$\boldsymbol{\delta}\mathbf{\Gamma} = \boldsymbol{\delta} \quad e \quad \boldsymbol{\delta}\mathbf{1}' = 1, \quad (2.8)$$

onde $\boldsymbol{\delta}\mathbf{1}' = 1$ garante que $\boldsymbol{\delta}$ seja de fato uma distribuição.

Quando a distribuição inicial $\mathbf{u}(1)$ da CM for sua própria distribuição estacionária, ou seja, $\mathbf{u}(1) = \boldsymbol{\delta}$, então Zucchini e MacDonald [40] chamam o processo de **cadeia de Markov estacionária**.¹ Uma CM **irredutível**² e **aperiódica**³ (homogênea, de tempo discreto, espaço de estados finito) possui uma única, estritamente positiva, distribuição estacionária que pode ser encontrada substituindo uma

¹Neste trabalho segue-se esta definição

²Um conjunto C de estados é **fechado** se não existe estado fora de C que possa ser alcançado por algum estado E_j contido em C . Uma Cadeia de Markov é **irredutível** se o menor conjunto fechado de estados é formado por todos os estados da cadeia.

³Para a definição de cadeia de Markov **periódica** veja Feller [8] pgs 360 a 362.

das equações do sistema $\delta\mathbf{\Gamma} = \delta$ por $\sum_i \delta_i = 1$, e depois resolvendo o novo sistema formado.

Uma vez introduzidos os conceitos básicos sobre cadeias de Markov é preciso apresentar o seu processo inferencial. Pois este é importante para o desenvolvimento e compreensão do método de estimação EM para os modelos HMM. A seguir apresenta-se de forma simplificada este processo.

2.2.2 Inferência em Cadeias de Markov

Esta seção se baseia em Bhat [3], Zucchini e MacDonald [40], e Whittle [37]. A partir deste ponto, a distribuição inicial da Cadeia de Markov será sempre chamada de δ ⁴.

Observe uma Cadeia de Markov finita com m estados $(1, 2, \dots, m)$ até que f transições tenham transcorrido. Seja f_{ij} o número de transições do estado i para o estado j ($i, j = 1, 2, \dots, m$), e $\sum_{j=1}^m f_{ij} = f_i$ o número de transições que saem do estado i ($i = 1, 2, \dots, m$).

Seja $\mathbf{\Gamma}$ a matriz de transição da CM dada por

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{bmatrix}.$$

Deseja-se estimar os elementos γ_{ij} ($i, j = 1, 2, \dots, m$).

Para um dado estado inicial i e um número de experimentos f_i , a amostra $(f_{i1}, f_{i2}, \dots, f_{im})$ dos números de transições observadas pode ser considerada como uma amostra de tamanho f_i de uma distribuição multinomial com probabilidades $(\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{im})$, tal que $\sum_{j=1}^m \gamma_{ij} = 1$. Assim, a probabilidade desta amostra ocorrer é dada por

$$\frac{f_i!}{f_{i1}! f_{i2}! \dots f_{im}!} \gamma_{i1}^{f_{i1}} \gamma_{i2}^{f_{i2}} \cdots \gamma_{im}^{f_{im}}$$

com $\sum_{j=1}^m f_{ij} = f_i$ e $\sum_{j=1}^m \gamma_{ij} = 1$.

Utilizando o mesmo argumento para os m estados iniciais $(1, 2, \dots, m)$, onde o número total de experimentos f é desmembrado em (f_1, f_2, \dots, f_m) conhecidos,

⁴Quando esta também for a distribuição estacionária da cadeia ficará claro no texto.

temos que a probabilidade das realizações f_{ij} ($i, j = 1, 2, \dots, m$) ocorrerem é dada por

$$\prod_{i=1}^m \frac{f_{i.}!}{f_{i1}! f_{i2}! \dots f_{im}!} \gamma_{i1}^{f_{i1}} \gamma_{i2}^{f_{i2}} \dots \gamma_{im}^{f_{im}}.$$

Como as somas $f_{i.}$ ($i = 1, 2, \dots, m$) também são variáveis aleatórias, a função de verossimilhança não-condicional $L(\mathbf{\Gamma})$ da amostra observada também depende da distribuição conjunta destas variáveis aleatórias. É possível mostrar que esta distribuição independe dos elementos γ_{ij} (veja Whittle [37]), e portanto é possível substituí-la, em $L(\mathbf{\Gamma})$, por um termo $A = g(f_{1.}, f_{2.}, \dots, f_{m.})$. Assim, a função de verossimilhança $L(\mathbf{\Gamma})$ e seu logaritmo $l(\mathbf{\Gamma})$ podem ser expressos como

$$L(\mathbf{\Gamma}) = A \prod_{i=1}^m \frac{f_{i.}!}{f_{i1}! f_{i2}! \dots f_{im}!} \gamma_{i1}^{f_{i1}} \gamma_{i2}^{f_{i2}} \dots \gamma_{im}^{f_{im}}$$

$$l(\mathbf{\Gamma}) = \log B \sum_{i=1}^m \sum_{j=1}^m f_{ij} \log \gamma_{ij} \quad (2.9)$$

onde B contém todos os termos independentes de γ_{ij} .

De (2.9) é possível notar que f_{ij} é uma *estatística suficiente* (Casella e Berger [5]) para estimar a probabilidade de transição γ_{ij} ($i, j = 1, 2, \dots, m$).

Observando que $l_i(\mathbf{\Gamma}) = \sum_{j=1}^m f_{ij} \log \gamma_{ij}$ são funções côncavas, para maximizar a função $l(\mathbf{\Gamma})$ em relação a γ_{ij} , basta maximizar cada soma $l_i(\mathbf{\Gamma})$. Assim, tem-se,

$$l_i(\mathbf{\Gamma}) = \sum_{j=1}^{m-1} f_{ij} \log \gamma_{ij} + f_{im} \log(1 - \gamma_{i1} - \gamma_{i2} - \dots - \gamma_{i,m-1}),$$

uma vez que $\sum_{j=1}^m \gamma_{ij} = 1$.

Diferenciando $l_i(\mathbf{\Gamma})$ com respeito a γ_{ij} ($j = 1, 2, \dots, m-1$) e igualando a zero tem-se:

$$\frac{f_{i1}}{\gamma_{i1}} - \frac{f_{im}}{1 - \gamma_{i1} - \gamma_{i2} - \dots - \gamma_{i,m-1}} = 0$$

$$\frac{f_{i2}}{\gamma_{i2}} - \frac{f_{im}}{1 - \gamma_{i1} - \gamma_{i2} - \dots - \gamma_{i,m-1}} = 0$$

$$\vdots$$

$$\frac{f_{i,m-1}}{\gamma_{i,m-1}} - \frac{f_{im}}{1 - \gamma_{i1} - \gamma_{i2} - \dots - \gamma_{i,m-1}} = 0$$

Combinando as equações tem-se

$$\frac{f_{i1}}{\gamma_{i1}} = \frac{f_{i2}}{\gamma_{i2}} = \dots = \frac{f_{i,m-1}}{\gamma_{i,m-1}} = \frac{f_{im}}{1 - \gamma_{i1} - \gamma_{i2} - \dots - \gamma_{i,m-1}}$$

obtendo-se então

$$\begin{aligned}
 \frac{f_{i1}}{f_{i1}}\gamma_{i1} &= \gamma_{i1} \\
 \frac{f_{i2}}{f_{i1}}\gamma_{i1} &= \gamma_{i2} \\
 &\vdots \\
 \frac{f_{i,m}}{f_{i1}}\gamma_{i1} &= 1 - \gamma_{i1} - \gamma_{i2} - \dots - \gamma_{i,m-1}
 \end{aligned} \tag{2.10}$$

Somando cada lado das equações em (2.10), tem-se

$$\begin{aligned}
 \frac{f_{i1} + f_{i2} + \dots + f_{im}}{f_{i1}}\gamma_{i1} &= 1 \quad \text{ou,} \\
 \frac{f_{i.}}{f_{i1}}\gamma_{i1} &= 1
 \end{aligned}$$

e, portanto o estimador de máxima verossimilhança será:

$$\hat{\gamma}_{i1} = \frac{f_{i1}}{f_{i.}}.$$

Analogamente é possível obter os estimadores

$$\hat{\gamma}_{ij} = \frac{f_{ij}}{f_{i.}} \quad (i, j = 1, 2, \dots, m).$$

Sendo a razão entre a frequência observada das transições do estado i para o estado j , pela frequência observada das transições que saem do estado i , este é um estimador natural para as probabilidades de transição de interesse.

Agora, considere a primeira observação (C_1) obtida de uma CM finita com m estados $(1, 2, \dots, m)$ cuja distribuição inicial é $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_m)$. E defina

$$u_i = \begin{cases} 1 & \text{se e somente se } C_1 = i, \\ 0 & \text{caso contrário.} \end{cases} \quad \text{para } i = 1, 2, \dots, m.$$

Note que C_1 é uma amostra de tamanho 1 de uma multinomial com probabilidades $(\delta_1, \delta_2, \dots, \delta_m)$. Assim, a probabilidade desta amostra ocorrer, ou seja, sua função de verossimilhança $L(\boldsymbol{\delta})$, é

$$L(\boldsymbol{\delta}) = \delta_1^{u_1} \delta_2^{u_2} \dots \delta_m^{u_m} = \prod_{i=1}^m \delta_i^{u_i}$$

com $\sum_{i=1}^m \delta_i = 1$ e $\sum_{i=1}^m u_i = 1$.

Assim, o logaritmo de $L(\boldsymbol{\delta})$ é

$$l(\boldsymbol{\delta}) = \sum_{i=1}^m u_i \log \delta_i.$$

Note que como C_1 pode assumir apenas um dos valores $1, 2, \dots, m$, apenas um u_i ($1, 2, \dots, m$) recebe o valor 1, todos os outros recebem o valor 0. Assim, suponha que $C_1 = i$, logo $u_1 = u_2 = \dots = u_{i-1} = 0$, $u_i = 1$, $u_{i+1} = \dots = u_m = 0$ e $l(\boldsymbol{\delta})$ resume-se a

$$l(\boldsymbol{\delta}) = u_i \log \delta_i = 1 \log \delta_i.$$

Como δ_i é uma probabilidade, ou seja, $0 \leq \delta_i \leq 1$, $l(\boldsymbol{\delta})$ é máximo quando $\delta_i = 1$ e portanto, os estimadores de máxima verossimilhança de $\boldsymbol{\delta}$ são dados por

$$\hat{\delta}_i = \begin{cases} 1 & \text{se e somente se } C_1 = i, \\ 0 & \text{caso contrário.} \end{cases} \quad \text{para } i = 1, 2, \dots, m.$$

Assim, obtém-se os estimadores da distribuição inicial e das probabilidades de transição de uma CM quando se observa apenas uma sequência desta cadeia.

Agora, suponha que sejam observadas S sequências **independentes**, que podem ter tamanhos diferentes, de uma CM com m estados ($1, 2, \dots, m$) e matriz de transição $\boldsymbol{\Gamma} = (\gamma_{ij})$, $i, j = 1, 2, \dots, m$.

Seja f_{ij}^s o número de transições do estado i para o estado j na s -ésima sequência observada, $\sum_{s=1}^S f_{ij}^s = f_{ij}$ o número de transições do estado i para o estado j em todas as sequências observadas, $\sum_{j=1}^m f_{ij}^s = f_i^s$ o número de transições que saem do estado i ($i = 1, 2, \dots, m$) na s -ésima sequência, e $\sum_{s=1}^S \sum_{j=1}^m f_{ij}^s = \sum_{j=1}^m \sum_{s=1}^S f_{ij}^s = f_i$ o número de transições que saem do estado i em todas as sequências.

Analogamente ao caso em que se observa uma sequência, tem-se que a função de verossimilhança $L_s(\boldsymbol{\Gamma})$ para a sequência s ($s = 1, 2, \dots, S$) é dada por

$$L_s(\boldsymbol{\Gamma}) = A_s \prod_{i=1}^m \frac{f_i^s!}{f_{i1}^s! f_{i2}^s! \dots f_{im}^s!} \gamma_{i1}^{f_{i1}^s} \gamma_{i2}^{f_{i2}^s} \dots \gamma_{im}^{f_{im}^s}$$

onde $A_s = g(f_1^s, f_2^s, \dots, f_m^s)$.

Como assume-se que as sequências são independentes, tem-se que a função de

verossimilhança $L(\Gamma)$ de todas as sequências e seu logaritmo são dados por

$$\begin{aligned}
L(\Gamma) &= \prod_{s=1}^S L_s(\Gamma) \\
&= \prod_{s=1}^S A_s \prod_{i=1}^m \frac{f_i^s!}{f_{i1}^s! f_{i2}^s! \dots f_{im}^s!} \gamma_{i1}^{f_{i1}^s} \gamma_{i2}^{f_{i2}^s} \dots \gamma_{im}^{f_{im}^s} \\
&= A \prod_{s=1}^S \prod_{i=1}^m \frac{f_i^s!}{f_{i1}^s! f_{i2}^s! \dots f_{im}^s!} \gamma_{i1}^{f_{i1}^s} \gamma_{i2}^{f_{i2}^s} \dots \gamma_{im}^{f_{im}^s} \\
&= A \prod_{s=1}^S \prod_{i=1}^m \frac{f_i^s!}{f_{i1}^s! f_{i2}^s! \dots f_{im}^s!} \prod_{i=1}^m \prod_{s=1}^S \gamma_{i1}^{f_{i1}^s} \gamma_{i2}^{f_{i2}^s} \dots \gamma_{im}^{f_{im}^s} \\
&= A \prod_{s=1}^S \prod_{i=1}^m \frac{f_i^s!}{f_{i1}^s! f_{i2}^s! \dots f_{im}^s!} \prod_{i=1}^m \gamma_{i1}^{\sum_{s=1}^S f_{i1}^s} \gamma_{i2}^{\sum_{s=1}^S f_{i2}^s} \dots \gamma_{im}^{\sum_{s=1}^S f_{im}^s},
\end{aligned}$$

onde $A = \prod_{s=1}^S A_s$. E,

$$l(\Gamma) = \log B \sum_{i=1}^m \sum_{j=1}^m \left(\sum_{s=1}^S f_{ij}^s \right) \log \gamma_{ij}, \quad (2.11)$$

onde B contém todos os termos independentes de γ_{ij} .

De (2.11) é possível perceber que $\left(\sum_{s=1}^S f_{ij}^s = f_{ij} \right)$ é uma estatística suficiente para estimar γ_{ij} ($i, j = 1, 2, \dots, m$).

Note que (2.11) é da mesma forma de (2.9), assim, fazendo as mesmas operações, temos que o estimador de máxima verossimilhança de γ_{ij} ($i, j = 1, 2, \dots, m$) é dado por

$$\hat{\gamma}_{ij} = \frac{f_{ij}}{f_i} = \frac{\sum_{s=1}^S f_{ij}^s}{\sum_{s=1}^S f_i^s}.$$

Do mesmo jeito que no caso em que se tem apenas uma sequência observada, o estimador da probabilidade de transição γ_{ij} é obtido pela razão entre o número observado de transições do estado i para o estado j pelo número de transições que saem do estado i .

Para a estimação da distribuição inicial $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_m)$, considere u_i o número de sequências que começaram no estado i ($i = 1, 2, \dots, m$). Assim, as realizações (u_1, u_2, \dots, u_m) de uma amostra de tamanho S (número de sequências observadas), terão distribuição multinomial com probabilidades $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_m)$. Com isso a função de verossimilhança $L(\boldsymbol{\delta})$ é dada por

$$L(\boldsymbol{\delta}) = \frac{S!}{u_1! u_2! \dots u_m!} \delta_1^{u_1} \delta_2^{u_2} \dots \delta_m^{u_m}$$

e

$$l(\boldsymbol{\delta}) = \log L(\boldsymbol{\delta}) = C + \sum_{i=1}^m u_i \log \delta_i \quad (2.12)$$

onde $C = \log \frac{S}{u_1! u_2! \dots u_m!}$.

Por outro lado, como $\sum_{i=1}^m u_i = S$ e $\sum_{i=1}^m \delta_i = 1$, tem-se que (2.12) pode ser reescrita como

$$l(\boldsymbol{\delta}) = C + \sum_{i=1}^{m-1} u_i \log \delta_i + u_m \log(1 - \delta_1 - \delta_2 - \dots - \delta_{m-1}).$$

Derivando $l(\boldsymbol{\delta})$ em relação a δ_i ($i = 1, 2, \dots, m-1$) e igualando a zero tem-se

$$\begin{aligned} \frac{u_1}{\delta_1} - \frac{u_m}{1 - \delta_1 - \delta_2 - \dots - \delta_{m-1}} &= 0 \\ \frac{u_2}{\delta_2} - \frac{u_m}{1 - \delta_1 - \delta_2 - \dots - \delta_{m-1}} &= 0 \\ &\vdots \\ \frac{u_{m-1}}{\delta_{m-1}} - \frac{u_m}{1 - \delta_1 - \delta_2 - \dots - \delta_{m-1}} &= 0 \end{aligned}$$

Combinando as equações tem-se

$$\frac{u_1}{\delta_1} = \frac{u_2}{\delta_2} = \dots = \frac{u_{m-1}}{\delta_{m-1}} = \frac{u_m}{1 - \delta_1 - \delta_2 - \dots - \delta_{m-1}}.$$

obtendo-se então

$$\begin{aligned} \frac{u_1}{u_1} \delta_1 &= \delta_1 \\ \frac{u_2}{u_1} \delta_1 &= \delta_2 \\ &\vdots \\ \frac{u_m}{u_1} \delta_1 &= 1 - \delta_1 - \delta_2 - \dots - \delta_{m-1} \end{aligned} \quad (2.13)$$

Somando cada lado das equações em (2.13), tem-se

$$\begin{aligned} \frac{u_1 + u_2 + \dots + u_m}{u_1} \delta_1 &= 1 \quad \text{ou,} \\ \frac{S}{u_1} \delta_1 &= 1 \end{aligned}$$

e portanto, o estimador de máxima verossimilhança de δ_1 é

$$\hat{\delta}_1 = \frac{u_1}{S}.$$

Analogamente, é possível obter os estimadores

$$\hat{\delta}_i = \frac{u_i}{S} \quad (i, j = 1, 2, \dots, m).$$

Note que, diferentemente do caso em que se tem apenas uma sequência observada, o vetor $\hat{\boldsymbol{\delta}}$ de estimativas pode ser diferente de vetores em que uma posição recebe o valor 1 e as restantes o valor 0.

Uma vez revistos os conceitos básicos em modelo de misturas e Cadeias de Markov, pode-se introduzir o modelo de Markov oculto através de suas definições, propriedades e demais detalhes necessários para o estudo proposto no Capítulo 1 deste trabalho.

Capítulo 3

Modelo de Markov Oculto

A teoria básica do Modelo de Markov Oculto (*Hidden Markov Model*-HMM) foi desenvolvida no fim dos anos 1960, mas apenas recentemente começou a ser extensivamente aplicada em variadas áreas de interesse. Cappé et al. [4] aponta que a utilização de estados ocultos torna o modelo suficientemente genérico para abranger as diversas complexidades que surgem em dados reais de séries temporais, enquanto a estrutura simples da cadeia de Markov permite uma eficiência nos processos computacionais. Sua aplicação mais conhecida é, provavelmente, em reconhecimento de fala [27], mas também há aplicações em análise de cadeias proteicas e DNA, reconhecimento de escrita manual [9], reconhecimento de gestos, e ultimamente na classificação de sinais biométricos [19], que é o interesse deste trabalho.

Além da estrutura univariada básica do HMM, diversas outras estruturas podem ser consideradas. Entre elas, o HMM multivariado (Zucchini e MacDonald [40], e Zhong e Ghosh [39]), *Input - Output* HMM (Bengio e Chiappa [2]), HMM Fatorial (Ghahramani e Jordan [12]) e HMM autorregressivo (Rabiner [27]).

Neste capítulo apresentaremos, segundo Zucchini e MacDonald [40], as definições básicas de um HMM univariado; a estrutura da sua função de verossimilhança; e as probabilidades *forward* e *backward*, necessárias para a estimação usando o algoritmo EM.

3.1 Exemplos

Antes de apresentar os conceitos e definições teóricas do HMM, considere os seguintes exemplos para a melhor compreensão da idéia do modelo.

Exemplo 3.1. *Imagine que à sua frente estejam dispostas 2 urnas (A e B), cada uma delas contém bolinhas coloridas. A urna A possui 2 bolas amarelas e 8 bolas azuis e a urna B possui 6 amarelas e 4 azuis. A cada jogada é retirada uma única bola da urna A ou B, anota-se 1 se a bola retirada é amarela e 0 se azul, e devolve-se a bola à sua urna.*

Todavia, a escolha da urna para a retirada da bola é feita lançando-se uma moeda. Quando obtêm-se cara (1) retira-se a bola da urna A e coroa (2) retira-se a bola da urna B. A Figura 3.1 ilustra o esquema para o sorteio das bolas. E a Tabela 3.1 contém uma possível realização para 10 primeiras retiradas.

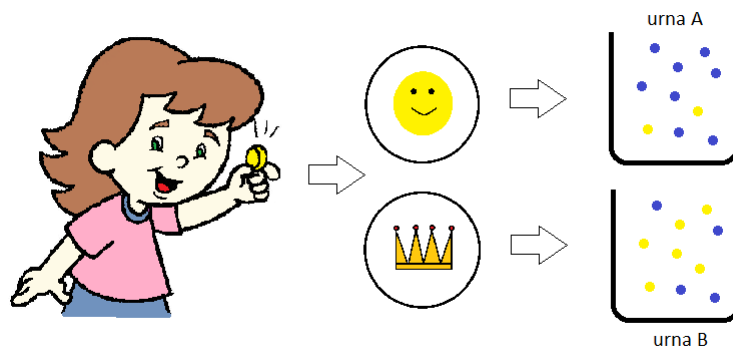


Figura 3.1: Esquema para o sorteio das bolas.

retirada	1	2	3	4	5	6	7	8	9	10
moeda	1	2	2	2	1	1	2	1	1	2
bola	0	1	1	0	0	1	0	0	0	1

Tabela 3.1: Resultados obtidos para o Exemplo 3.1.

Neste experimento sabe-se os resultados dos lançamentos da moeda, e portanto de qual urna a bola retirada se originou e qual função de probabilidade está associada à retirada da bola. A urna A representa uma Bernoulli com probabilidade de sucesso $p_A = 0,2$ e a urna B uma Bernoulli com probabilidade $p_B = 0,6$. Note que o resultado

do lançamento da moeda também segue uma distribuição Bernoulli e, se a moeda for honesta, tem-se uma Bernoulli(0,5).

Agora, imagine que as urnas e a moeda estão atrás de uma cortina e que a única informação visível é a cor da bola retirada. Assim no lugar da Tabela 3.1 tem-se a Tabela 3.2.

retirada	1	2	3	4	5	6	7	8	9	10
bola	0	1	1	0	0	1	0	0	0	1

Tabela 3.2: Resultados obtidos para o Exemplo 3.1, agora com o processo oculto.

Nesta situação sabe-se que as retiradas das bolas podem vir de distribuições diferentes, mas não de qual. Aqui, o processo que escolhe a distribuição está **oculto** atrás da cortina.

Exemplo 3.2. Suponha uma cadeia de Markov com distribuição inicial e matriz de transição dadas por:

$$\delta = (0,4 \quad 0,6) \quad e \quad \Gamma = \begin{pmatrix} 0,4 & 0,6 \\ 0,7 & 0,3 \end{pmatrix}.$$

Suponha ainda que associada ao estado 1 da cadeia de Markov temos uma distribuição, chamada **distribuição estado-dependente**, Normal (ϕ_1) de média $\mu_1 = -1$ e variância $\sigma_1^2 = 1$, e ao estado 2 uma distribuição Normal (ϕ_2) de média $\mu_2 = 2$ e variância $\sigma_2^2 = 2$.

O software livre R (www.r-project.org) foi utilizado para implementar a função `r.normal.hmm`, disponível no Apêndice A (Programa A.1). Por meio desta função foram gerados 10 pares (C_t, X_t) , onde C_t é o estado observado no tempo t e X_t é o valor gerado pela densidade ϕ_1 , se $C_t = 1$, e ϕ_2 , se $C_t = 2$. O resultado pode ser observado na Tabela 3.3.

t	1	2	3	4	5	6	7	8	9	10
C_t	1	1	2	1	1	2	1	1	2	1
X_t	-1,24	-2,09	2,71	-0,32	-2,03	3,33	0,54	1,30	3,00	-0,51

Tabela 3.3: Resultados obtidos para o Exemplo 3.2.

Como no Exemplo 3.1, sabe-se de qual distribuição (ϕ_1 ou ϕ_2) as observações X_t , $t = 1, 2, \dots, 10$, vieram. Mas o interesse se concentra nos casos onde os valores C_t , $t = 1, 2, \dots, 10$, da cadeia de Markov estão **ocultos**, caracterizando assim um modelo de Markov oculto (Hidden Markov Model - HMM).

Agora que já foram apresentados dois exemplos de HMM, tem-se melhores condições para introduzir, na seção seguinte, os conceitos teóricos do modelo.

3.2 Formalização

3.2.1 HMM básico

Definição 3.2.1. Um *modelo de Markov oculto (HMM)* $\{X_t : t \in \mathbb{N}\}$ é definido como uma mistura de observações dependentes. Sejam os registros históricos do tempo 1 ao t , $\mathbf{X}^{(t)}$ e $\mathbf{C}^{(t)}$. Onde $\mathbf{X}^{(t)}$ representa a sequência X_1, X_2, \dots, X_t de valores observados, e $\mathbf{C}^{(t)}$ representa a sequência C_1, C_2, \dots, C_t de estados ocultos. A estrutura do modelo é definida por

$$\Pr(C_t | \mathbf{C}^{(t-1)}) = \Pr(C_t | C_{t-1}), \quad t = 2, 3, \dots \quad e \quad (3.1)$$

$$\Pr(X_t | \mathbf{X}^{(t-1)}, \mathbf{C}^{(t)}) = \Pr(X_t | C_t), \quad t \in \mathbb{N}. \quad (3.2)$$

O modelo então é caracterizado por duas partes:

- i) o processo oculto $\{\mathbf{C}^{(t)} : t = 1, 2, \dots\}$ que satisfaz a propriedade de Markov (2.3), e
- ii) o processo estado-dependente $\{\mathbf{X}^{(t)} : t = 1, 2, \dots\}$, cuja probabilidade condicionada a $\mathbf{C}^{(t)}$ e $\mathbf{X}^{(t-1)}$ depende apenas do estado atual C_t .

A Figura 3.2 descreve a relação de dependência condicional entre os estados e observações. É esta relação, oculta, que diferencia o HMM da mistura de distribuições com observações independentes. É por sua existência que o HMM é capaz de tratar a sequência de dados levando em consideração a estrutura temporal existente, tornando-se útil em análises de dados de séries temporais. O mesmo não é possível quando opta-se por um modelo de mistura.

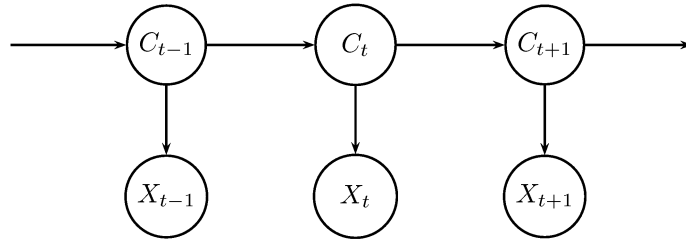


Figura 3.2: Diagrama da relação de dependência condicional em um HMM

Agora, para um HMM cuja cadeia de Markov $\{C_t\}$ possui m estados, define-se, quando X_t é discreto, para $i = 1, 2, \dots, m$

$$p_i(x) = \Pr(X_t = x \mid C_t = i). \quad (3.3)$$

Ou seja, p_i é a *função de probabilidade* de X_t quando a CM está no estado i no tempo t . Quando X_t é contínuo p_i é interpretado como a *função densidade*. As m distribuições p_i são chamadas de **distribuições estado-dependente** do modelo. Note que p_i não depende do valor de t .

3.2.2 Distribuição marginal

Seja $u_i(t) = \Pr(C_t = i)$, para $t = 1, 2, \dots, T$. Para X_t discreto, tem-se

$$\begin{aligned} \Pr(X_t = x) &= \sum_{i=1}^m \Pr(X_t = x, C_t = i) \\ &= \sum_{i=1}^m \Pr(C_t = i) \Pr(X_t = x \mid C_t = i) \\ &= \sum_{i=1}^m u_i(t) p_i(x) \end{aligned} \quad (3.4)$$

Para o caso em que X_t é contínuo pode-se obter (3.4) de maneira análoga.

Reescrevendo (3.4) em notação matricial obtem-se

$$\begin{aligned} \Pr(X_t = x) &= (u_1(t) \ \dots \ u_m(t)) \begin{pmatrix} p_1(x) & & 0 \\ & \ddots & \\ 0 & & p_m(x) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \mathbf{u}(t)\mathbf{P}(x)\mathbf{1}' \end{aligned} \quad (3.5)$$

onde $\mathbf{P}(x)$ é a matriz diagonal cujo i -ésimo elemento da diagonal é $p_i(x)$. Utilizando (2.7) ($\mathbf{u}(t) = \mathbf{u}(1)\mathbf{\Gamma}^{t-1}$), obtem-se

$$\Pr(X_t = x) = \mathbf{u}(1)\mathbf{\Gamma}^{t-1}\mathbf{P}(x)\mathbf{1}'. \quad (3.6)$$

Quando a CM $\{C_t\}$ for estacionária, com distribuição estacionária $\boldsymbol{\delta}$, (3.6) resume-se a

$$\Pr(X_t = x) = \boldsymbol{\delta}\mathbf{P}(x)\mathbf{1}', \quad (3.7)$$

pois, pela definição (2.8), $\boldsymbol{\delta}\mathbf{\Gamma}^{t-1} = \boldsymbol{\delta}$.

3.2.3 Função de verossimilhança

Suponha uma sequência de observações $\mathbf{x}^{(T)} = (x_1, x_2, \dots, x_T)$ originada de um modelo HMM. A CM associada a tal modelo possui m estados, distribuição inicial $\boldsymbol{\delta}$ e matriz de transição $\mathbf{\Gamma}$, com p_i , $i = 1, 2, \dots, m$, as funções de probabilidade (ou densidade) estado-dependentes. A probabilidade L_T de observar tal sequência sob o modelo HMM descrito é dita **função de verossimilhança**.

Proposição 3.1. *A função de verossimilhança L_T é dada por*

$$L_T = L_T(\boldsymbol{\delta}, \mathbf{\Gamma}, p_1, \dots, p_m | \mathbf{x}^{(T)}) = \boldsymbol{\delta}\mathbf{P}(x_1)\mathbf{\Gamma}\mathbf{P}(x_2) \dots \mathbf{\Gamma}\mathbf{P}(x_T)\mathbf{1}'. \quad (3.8)$$

Se $\boldsymbol{\delta}$, a distribuição de C_1 , é a distribuição estacionária da CM, então

$$L_T = \boldsymbol{\delta}\mathbf{\Gamma}\mathbf{P}(x_1)\mathbf{\Gamma}\mathbf{P}(x_2) \dots \mathbf{\Gamma}\mathbf{P}(x_T)\mathbf{1}'. \quad (3.9)$$

Exemplo 3.3. *Voltando ao Exemplo 3.2, pode-se calcular a verossimilhança da sequência observada*

$$\mathbf{x}^{(10)} = \left(-1,24 \quad -2,09 \quad 2,71 \quad -0,32 \quad -2,03 \quad 3,33 \quad 0,54 \quad 1,30 \quad 3,00 \quad -0,51 \right).$$

Lembre que os parâmetros do modelo são: $\mathbf{u}(1) = \boldsymbol{\delta} = (0,4 \ 0,6)$ e matriz de transição

$$\mathbf{\Gamma} = \begin{pmatrix} 0,4 & 0,6 \\ 0,7 & 0,3 \end{pmatrix}.$$

E as densidades estado-dependentes são normais $\phi_1(-1,1)$ e $\phi_2(2,2)$. Com isso, a verossimilhança é

$$L_T = \boldsymbol{\delta}\mathbf{P}(-1,24)\mathbf{\Gamma}\mathbf{P}(-2,09) \dots \mathbf{\Gamma}\mathbf{P}(-0,51) = 2,262364 \times 10^{-9}.$$

O cálculo da verossimilhança requer um número de operações $O(Tm^T)$, mas a utilização de formas recursivas para o cálculo faz o número de operações cair para $O(Tm^2)$ ¹. Uma maneira de alcançar tal objetivo é fazer uso das **probabilidades forward e backward**². Estes procedimentos são descritos na seção seguinte.

3.2.4 Probabilidades *forward* e *backward*

Considere uma sequência $t = 1, 2, \dots, T$, e defina os vetores (linhas) $\boldsymbol{\alpha}_t$ e $\boldsymbol{\beta}_t$ como

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta}\mathbf{P}(x_1)\mathbf{\Gamma}\mathbf{P}(x_2) \dots \mathbf{\Gamma}\mathbf{P}(x_t) = \boldsymbol{\delta}\mathbf{P}(x_1) \prod_{s=2}^t \mathbf{\Gamma}\mathbf{P}(x_s) \quad (3.10)$$

e

$$\boldsymbol{\beta}'_t = \mathbf{\Gamma}\mathbf{P}(x_{t+1})\mathbf{\Gamma}\mathbf{P}(x_{t+2}) \dots \mathbf{\Gamma}\mathbf{P}(x_T)\mathbf{1}' = \left(\prod_{s=t+1}^T \mathbf{\Gamma}\mathbf{P}(x_s) \right) \mathbf{1}' \quad (3.11)$$

Os elementos de $\boldsymbol{\alpha}_t$ são chamados de **probabilidades forward**, e os de $\boldsymbol{\beta}_t$ de **probabilidades backward**³.

As seguintes proposições⁴ estão relacionadas as probabilidades *forward* e *backward*.

As Proposições 3.2 e 3.3 seguem direto das definições (3.10) e (3.11).

¹ $f(x) = O(g(x))$, se e somente se, existem M , real positivo, e x_0 , real, tais que

$$|f(x)| = M|g(x)|, \quad \text{para todo } x > x_0.$$

²Aqui preferiu-se utilizar os termos *forward* e *backward* em inglês. Uma possível tradução para o português dos termos é probabilidades **progressivas** e **regressivas**.

³A função implementada `log.alphabeta.norm.uni`, Programa A.4 do Apêndice A, calcula o log das probabilidades *forward* e *backward*.

⁴Provas e detalhes estão em Zucchini e MacDonald [40], págs. 60 a 63.

Proposição 3.2. Para $t = 1, 2, \dots, T$ e $j = 1, 2, \dots, m$,

$$\alpha_t(j) = \Pr(\mathbf{X}^{(t)} = \mathbf{x}^{(t)}, C_t = j), \quad (3.12)$$

isto é, $\alpha_t(j)$ é a probabilidade conjunta de se observar a sequência $\mathbf{x}^{(t)}$ e o processo estar no estado j no tempo t ($C_t = j$).

Proposição 3.3. Para $t = 1, 2, \dots, T - 1$ e $i = 1, 2, \dots, m$,

$$\beta_t(i) = \Pr(X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, \dots, X_T = x_T \mid C_t = j), \quad (3.13)$$

sabendo que $\Pr(C_t = i) > 0$. Em notação compacta,

$$\beta_t(i) = \Pr(\mathbf{X}_{t+1}^T = \mathbf{x}_{t+1}^T \mid C_t = i), \quad (3.14)$$

onde \mathbf{X}_a^b é o vetor $\begin{pmatrix} X_a & X_{a+1} & \dots & X_b \end{pmatrix}$.

Isso significa que $\beta_t(i)$ é a probabilidade de se observar a sequência \mathbf{x}_{t+1}^T dado que o processo está no estado i no tempo t ($C_t = i$).

Agora, as Proposições 3.4 e 3.5 seguintes serão utilizadas no algoritmo EM (Capítulo 4).

Proposição 3.4. Para $t = 1, 2, \dots, T$ e $i = 1, 2, \dots, m$,

$$\alpha_t(i)\beta_t(i) = \Pr(\mathbf{X}^{(T)} = \mathbf{x}^{(T)}, C_t = i), \quad (3.15)$$

e conseqüentemente, para cada t

$$\boldsymbol{\alpha}_t \boldsymbol{\beta}_t' = \Pr(\mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = L_T. \quad (3.16)$$

Esta proposição define T caminhos para o cálculo de L_T , um para cada valor de t . O caminho $t = T$ que define $L_T = \boldsymbol{\alpha}_T \mathbf{1}'$, parece o mais conveniente pois usa apenas as probabilidades forward.

Proposição 3.5. Para $t = 1, 2, \dots, T$,

$$\Pr(C_t = j \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \frac{\alpha_t(j)\beta_t(j)}{L_T}; \quad (3.17)$$

e, para $t = 2, 3, \dots, T$

$$\Pr(C_{t-1} = j, C_t = k \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \frac{\alpha_{t-1}(j)\gamma_{jk}p_k(x_t)\beta_t(k)}{L_T}, \quad (3.18)$$

onde $\gamma_{jk} = \Pr(C_t = k \mid C_{t-1} = j)$ é a probabilidade de transição de ir do estado j para o estado k a um passo, e $p_k(x_t) = \Pr(X_t = x_t \mid C_t = k)$ é o valor da k -ésima densidade estado-dependente no ponto x_t .

A expressão (3.17) define a probabilidade do processo estar no estado j no tempo t , dado que se observou a sequência $\mathbf{x}^{(T)}$, em função das probabilidades forward e backward. E a expressão (3.18) define a probabilidade do processo ir do estado j para o estado k nos tempos $t - 1$ e t , dado que se observou a sequência $\mathbf{x}^{(T)}$, em função das probabilidades forward e backward, da probabilidade de transição γ_{jk} e da densidade estado-dependente $p_k(x_t)$.

Após apresentar os conceitos teóricos sobre HMM apresenta-se na seção a seguir algumas considerações sobre este modelo.

3.3 Considerações

Antes de prosseguir com o conteúdo deste trabalho deve-se fazer algumas considerações sobre o HMM. A primeira delas é sobre as distribuições estado-dependentes que, apesar de nenhum exemplo ter sido apresentado, não precisam ser, necessariamente, do mesmo tipo ou classe. Ou seja, em um estado é possível ter-se uma distribuição Gamma, em outro Poisson ou Normal, e assim por diante. Isso tudo contanto que faça sentido para o estudo.

Outro ponto a ser considerado é a estrutura de dependência do modelo. Em uma avaliação mais geral é possível ver que, na verdade, o modelo HMM trata-se de uma **rede Bayesiana**, que é definida como um modelo gráfico para representar independências condicionais entre os conjuntos de variáveis aleatórias. A Figura 3.2 é um exemplo da definição de rede Bayesiana. Obviamente as redes Bayesinas abrangem outras estruturas de dependências e independências além da descrita pelo HMM.

Gharhramani [11], coloca o HMM dentro de uma subclasse das redes Bayesianas, as **redes Bayesianas Dinâmicas** (*Dinamic Bayesian networks*). As redes Bayesianas Dinâmicas são simplesmente redes Bayesianas utilizadas para modelar séries temporais. No modelo de séries temporais assume-se que um evento pode implicar outro evento no futuro, mas não o contrário, simplificando a estrutura da rede

Bayesiana.

Nesta mesma subclasse das redes Bayesianas Dinâmicas estão os modelos de Markov e os modelos de estado-espço linear-Gaussianos, também conhecidos como *Kalman filter*. Às vezes pode-se fazer confusão entre o HMM e o *Kalman filter*. Para que se entenda melhor, o *Kalman filter* é considerado uma versão do HMM com espaço de estados contínuo.

No HMM o espaço de estados é discreto e finito. Este espaço de estados é gerado por uma cadeia de Markov, que é um processo estocástico com suas características e propriedades particulares (veja a Seção 2.2). Já no *Kalman filter*, o espaço de estados é não-enumerável, e por tanto infinito, e considera-se ter uma distribuição Normal. Para este a probabilidade de transição do estado C_{t-1} para o estado C_t ($\Pr(C_t | C_{t-1})$) é decomposta em componentes determinística e estocástica:

$$C_t = f_t(C_{t-1}) + w_t$$

onde f_t é a componente determinística e w_t é um ruído aleatório com distribuição Normal de média zero. Essa estrutura linear e a distribuição Normal do ruído é que caracterizam o nome do modelo espaço-estado linear-Gaussiano.

Feitas as considerações necessárias o capítulo seguinte trata da estimação e da escolha do modelo, e do problema de classificação.

Capítulo 4

Estimação e Classificação

Este capítulo apresenta o algoritmo EM (*Expectation-Maximization*) como um método de estimação para o modelo HMM; duas medidas para a seleção do modelo: o AIC (*Akaike information criterion*) e o BIC (*Bayesian information criterion*). Uma vez estimado, o modelo é usado para a classificação das sequências seguintes.

4.1 Estimação

Zucchini e MacDonald [40] apresentam três técnicas para estimação: a maximização direta da verossimilhança, o algoritmo EM (*Expectation-Maximization*) e usando inferência Bayesiana. A maximização direta pode ser feita por meio de algoritmos de otimização não linear, já que não existem formas fechadas para a maximização. Vários pacotes estatísticos apresentam funções implementadas para a otimização não linear. No pacote R, por exemplo, pode-se usar as funções `nlm`, `optimize` e `optim` (Rizzo [28]), entre outras.

Quando opta-se pela maximização direta é importante fazer transformações nos parâmetros naturais, quando necessárias. Por exemplo, se a distribuição estado-dependente for uma $\text{Poisson}(\lambda)$, o espaço paramétrico de λ é $\{\lambda : \lambda > 0\}$. Tomando $\eta = \log \lambda$, tem-se $\eta \in \mathbb{R}$. No caso da distribuição $\text{Normal}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ e $\sigma^2 \geq 0$, e deve-se fazer uma transformação em σ^2 .

Rydén [34] compara a utilização das combinações bootstrap - EM com amostrador de Gibbs - MCMC (*Markov chain Monte Carlo*) para a estimação do HMM

por uma perspectiva computacional. Rydén [34] chega à conclusão de que, quando o interesse está apenas na estimação pontual, o algoritmo EM é o mais simples e rápido. Quando a estimação pontual não é suficiente a comparação não é tão simples. Nos exemplos dados em [34], no contexto Bayesiano, o amostrador de Gibbs e algoritmos MCMC tendem a consumir menos tempo computacional.

Neste trabalho optou-se pela utilização do algoritmo EM, que aplicado à estimação dos parâmetros do HMM é também conhecido por algoritmo de Baum-Welch.

4.1.1 Algoritmo EM

O algoritmo EM é um método de otimização frequentemente utilizado para encontrar estimadores de máxima verossimilhança quando se tem dados incompletos, considerando o fato de que a verossimilhança (ou log-verossimilhança) dos dados completos é provavelmente mais simples de maximizar do que avaliando apenas os dados (incompletos) observados. O algoritmo é composto por dois passos (**E** e **M**). Aplicado à estimação dos parâmetros do modelo HMM estes passos são:

- No **passo E** (*Expectation*) é calculada a **esperança** condicional dos dados não observados dadas as observações e as estimativas dos parâmetros do modelo naquele instante.
- No **passo M** (*Maximization*) substitui-se os dados faltantes pelas suas esperanças condicionais calculadas no passo E e então **maximiza**-se a verossimilhança (ou log-verossimilhança) em função dos parâmetros.

Os dois passos são repetidos até que se atinja um critério de convergência estipulado.

4.1.2 Algoritmo EM para o HMM

Apresenta-se a seguir um resumo do algoritmo EM descrito por Zucchini e MacDonald [40], Capítulo 4.

Considere que uma sequência de tamanho T de um HMM foi observada. Defina

c_1, c_2, \dots, c_T como a sequência de estados da CM,

$$u_j(t) = \begin{cases} 1 & \text{se e somente se } c_t = j, \\ 0 & \text{caso contrário.} \end{cases} \quad \text{para } t = 1, 2, \dots, T$$

e

$$v_{jk}(t) = \begin{cases} 1 & \text{se e somente se } c_{t-1} = j \text{ e } c_t = k, \\ 0 & \text{caso contrário.} \end{cases} \quad \text{para } t = 2, 3, \dots, T.$$

Assim, a log-verossimilhança dos dados completos ℓ_c é dada por

$$\begin{aligned} \ell_c &= \log(\Pr(\mathbf{x}^{(T)}, \mathbf{c}^{(T)})) \\ &= \log\left(\delta_{c_1} \prod_{t=2}^T \gamma_{c_{t-1}, c_t} \prod_{t=1}^T p_{c_t}(x_t)\right) \\ &= \sum_{j=1}^m u_j(1) \log \delta_j + \sum_{j=1}^m \sum_{k=1}^m \left(\sum_{t=2}^T v_{jk}(t)\right) \log \gamma_{jk} \\ &\quad + \sum_{j=1}^m \sum_{t=1}^T u_j(t) \log p_j(x_t) \\ &= \text{termo 1} + \text{termo 2} + \text{termo 3} \end{aligned} \tag{4.1}$$

onde $\boldsymbol{\delta}$ é a distribuição inicial da CM, γ_{jk} são as probabilidades de transição que compõem $\boldsymbol{\Gamma}$, matriz de transição da CM, e p_j são as distribuições estado-dependentes.

Com isso, para o algoritmo EM tem-se:

- Passo E: Calcula-se as esperanças condicionais de $u_j(t)$ e $v_{jk}(t)$ obtidas pelas Proposições 3.4 e 3.5:

$$\hat{u}_j(t) = \Pr(C_t = j \mid \mathbf{x}^{(T)}) = \frac{\alpha_t(j) \beta_t(j)}{L_T}$$

e

$$\hat{v}_{jk}(t) = \Pr(C_{t-1} = j, C_t = k \mid \mathbf{x}^{(T)}) = \frac{\alpha_{t-1}(j) \gamma_{jk} p_k(x_t) \beta_t(k)}{L_T}$$

Assim, $\hat{u}_j(t)$ é a estimativa para a probabilidade de o processo estar no estado j no tempo t ($C_t = j$), dado que se observou a sequência $\mathbf{x}^{(T)}$. E, $\hat{v}_{jk}(t)$ é a estimativa da probabilidade do processo transitar do estado j para o k nos tempos $t-1$ e t ($C_{t-1} = j, C_t = k$), dado que se observou a sequência $\mathbf{x}^{(T)}$.

- Passo M: Após substituir $u_j(t)$ e $v_{jk}(t)$ por $\hat{u}_j(t)$ e $\hat{v}_{jk}(t)$ deve-se maximizar (4.1), mas maximizar (4.1) significa maximizar cada termo de (4.1), pois o termo 1 depende apenas de δ , o termo 2 de Γ e o termo 3 dos parâmetros da distribuição estado-dependente. Assim, os estimadores dos parâmetros são:

(i) $\hat{\delta}_j = \frac{\hat{u}_j(1)}{\sum_{j=1}^m \hat{u}_j(1)} = \hat{u}_j(1)$, a estimativa para a probabilidade inicial do estado j é a estimativa da probabilidade do processo estar no estado j no tempo 1 ($C_1 = j$), dado que se observou a sequência $\mathbf{x}^{(T)}$.

(ii) $\hat{\gamma}_{jk} = \frac{f_{jk}}{\sum_{k=1}^m f_{jk}}$, onde $f_{jk} = \sum_{t=2}^T \hat{v}_{jk}(t)$. Note que f_{jk} é a frequência esperada de transições do estado j para o estado k . (Veja Seção 2.2.2.)

(iii) para $p_j(x)$ normal univariada com média μ_j e σ_j^2 , tem-se

$$\hat{\mu}_j = \frac{\sum_{t=1}^T \hat{u}_j(t)x_t}{\sum_{t=1}^T \hat{u}_j(t)} \quad \text{e} \quad \hat{\sigma}_j^2 = \frac{\sum_{t=1}^T \hat{u}_j(t)(x_t - \hat{\mu}_j)^2}{\sum_{t=1}^T \hat{u}_j(t)}.$$

A média e a variância da j -ésima densidade estado-dependente normal são obtidas ponderando cada ponto x_t observado pela probabilidade do processo estar no estado j no tempo t ($C_t = j$), dada a sequência $\mathbf{x}^{(T)}$.

Um processo genérico é obtido considerando que S sequências independentes de tamanho T_s ($s = 1, 2, \dots, S$) de um HMM foram observadas. Defina $c_1^s, c_2^s, \dots, c_{T_s}^s$ como a s -ésima sequência de estados da CM,

$$u_j^s(t) = \begin{cases} 1 & \text{se e somente se } c_t^s = j, \\ 0 & \text{caso contrário.} \end{cases} \quad \text{para } t = 1, 2, \dots, T_s$$

e

$$v_{jk}^s(t) = \begin{cases} 1 & \text{se e somente se } c_{t-1}^s = j \text{ e } c_t^s = k, \\ 0 & \text{caso contrário.} \end{cases} \quad \text{para } t = 2, 3, \dots, T_s.$$

Assim, a log-verossimilhança dos dados completos ℓ_c é dada por

$$\begin{aligned}
\ell_c &= \sum_{s=1}^S \log(\Pr(\mathbf{x}_s^{(T)}, \mathbf{c}_s^{(T)})) \\
&= \sum_{s=1}^S \log\left(\delta_{c_1^s} \prod_{t=2}^T \gamma_{c_{t-1}^s, c_t^s} \prod_{t=1}^T p_{c_t^s}(x_t^s)\right) \\
&= \sum_{j=1}^m \left(\sum_{s=1}^S u_j^s(1)\right) \log \delta_j + \sum_{j=1}^m \sum_{k=1}^m \left(\sum_{s=1}^S \sum_{t=2}^{T_s} v_{jk}^s(t)\right) \log \gamma_{jk} \\
&+ \sum_{j=1}^m \sum_{s=1}^S \sum_{t=1}^{T_s} u_j^s(t) \log p_j(x_t^s) \\
&= \text{termo 1} + \text{termo 2} + \text{termo 3}
\end{aligned} \tag{4.2}$$

onde $\boldsymbol{\delta}$ é a distribuição inicial da CM, γ_{jk} são as probabilidades de transição que compõem $\boldsymbol{\Gamma}$, matriz de transição da CM, e p_j são as distribuições estado-dependentes.

Com isso, para o algoritmo EM tem-se:

- Passo E: Calcula-se as esperanças condicionais de $u_j^s(t)$ e $v_{jk}^s(t)$ obtidas pelas Proposições 3.4 e 3.5:

$$\hat{u}_j^s(t) = \Pr(C_t^s = j \mid \mathbf{x}_s^{(T_s)}) = \frac{\alpha_t(j)\beta_t(j)}{L_{T_s}}$$

e

$$\hat{v}_{jk}^s(t) = \Pr(C_{t-1}^s = j, C_t^s = k \mid \mathbf{x}_s^{(T_s)}) = \frac{\alpha_{t-1}(j)\gamma_{jk}p_k(x_t^s)\beta_t(k)}{L_{T_s}}$$

Assim, $\hat{u}_j^s(t)$ é a estimativa para a probabilidade do s -ésimo processo estar no estado j no tempo t ($C_t^s = j$), dado que se observou a sequência $\mathbf{x}_s^{(T_s)}$. E, $\hat{v}_{jk}^s(t)$ é a estimativa da probabilidade do s -ésimo processo transitar do estado j para o k nos tempos $t-1$ e t ($C_{t-1}^s = j, C_t^s = k$), dado que se observou a sequência $\mathbf{x}_s^{(T_s)}$.

- Passo M: Após substituir $u_j^s(t)$ e $v_{jk}^s(t)$ por $\hat{u}_j^s(t)$ e $\hat{v}_{jk}^s(t)$ deve-se maximizar (4.2), mas maximizar (4.2) significa maximizar cada termo de (4.2), pois o termo 1 depende apenas de $\boldsymbol{\delta}$, o termo 2 de $\boldsymbol{\Gamma}$ e o termo 3 dos parâmetros da distribuição estado-dependente. Assim, os estimadores dos parâmetros são:

$$(i) \hat{\delta}_j = \frac{\sum_{s=1}^S \hat{u}_j^s(1)}{\sum_{s=1}^S \sum_{j=1}^m \hat{u}_j^s(1)} = \frac{\sum_{s=1}^S \hat{u}_j^s(1)}{S}. \text{ Note que } \sum_{s=1}^S \hat{u}_j^s(1) \text{ é a frequência esperada de sequências cujo primeiro estado } (t=1) \text{ está em } j.$$

(ii) $\hat{\gamma}_{jk} = \frac{\sum_{s=1}^S f_{jk}^s}{\sum_{s=1}^S \sum_{k=1}^m f_{jk}^s}$, onde $f_{jk}^s = \sum_{t=2}^{T_s} \hat{v}_{jk}^s(t)$. Note que f_{jk}^s é a frequência esperada de transições do estado j para o estado k na s -ésima sequência.

(Caso tenha dúvidas nos itens anteriores (i) e (ii) volte para 2.2.2 (Inferência em Cadeias de Markov)).

(iii) para $p_j(x)$ normal univariada com média μ_j e σ_j^2 , tem-se

$$\hat{\mu}_j = \frac{\sum_{s=1}^S \sum_{t=1}^{T_s} \hat{u}_j^s(t) x_t^s}{\sum_{s=1}^S \sum_{t=1}^{T_s} \hat{u}_j^s(t)} \quad \text{e} \quad \hat{\sigma}_j^2 = \frac{\sum_{s=1}^S \sum_{t=1}^{T_s} \hat{u}_j^s(t) (x_t^s - \hat{\mu}_j)^2}{\sum_{s=1}^S \sum_{t=1}^{T_s} \hat{u}_j^s(t)}.$$

A média e variância agora são ponderadas pela probabilidade do s -ésimo processo estar no estado j no tempo t dado que se observou $\mathbf{x}_s^{(T_s)}$.

Agora que se apresentou o algoritmo EM¹ para estimar modelos HMM deve-se fazer outras considerações que afetam a qualidade desse algoritmo. A seguir, o problema do máximo local é apresentado.

4.1.3 O problema do máximo local

Um problema do algoritmo EM (assim como de outros algoritmos de otimização) é que, dependendo do ponto inicial escolhido, o algoritmo pode convergir para um ponto de máximo local da função objetivo ao invés de alcançar o ponto de máximo global desejado. Uma solução possível é inicializar o algoritmo com diferentes pontos iniciais e verificar as verossimilhanças obtidas pelos conjuntos de estimativas e ver qual dos conjuntos gera a maior verossimilhança.

Obviamente esta solução não garante que se obtenha o máximo global, mas é fato que quanto maior o número de pontos iniciais diferentes maior é a chance de evitar os pontos de máximo locais.

Rydén [34] sugere que se inicie o EM por diferentes pontos aleatórios e sugere

$$\boldsymbol{\delta} \sim \text{Dirichlet}(1, 1, \dots, 1) \tag{4.3}$$

$$\boldsymbol{\gamma}_j \sim \text{Dirichlet}(1, 1, \dots, 1) \tag{4.4}$$

¹Implementado na função `EM.hmm.norm.uni`, Programa A.5 do Apêndice A

para os parâmetros que compõem a CM do modelo; onde $\boldsymbol{\delta}$ é a distribuição inicial da cadeia, γ_j é a j -ésima linha de $\boldsymbol{\Gamma}$, matriz de transição e, a densidade Dirichlet de parâmetros $\theta_1, \theta_2, \dots, \theta_k > 0$ é dada por

$$f(y_1, y_2, \dots, y_k \mid \theta_1, \theta_2, \dots, \theta_k) = \frac{1}{B(\boldsymbol{\theta})} \prod_{i=1}^k y_i^{\theta_i-1}$$

com $B(\boldsymbol{\theta})$ função Beta multinomial expressa em termos da função Gama, isto é,

$$B(\boldsymbol{\theta}) = \frac{\prod_{i=1}^k \Gamma(\theta_i)}{\Gamma\left(\sum_{i=1}^k \theta_i\right)}, \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_k).$$

Neste trabalho geram-se os vetores γ_j ($j = 1, 2, \dots, m$) como em 4.4 ², mas é acrescido um passo no processo para garantir que a matriz $\boldsymbol{\Gamma}$ gerada seja ergódica. Para isso, basta que o sistema $\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}$ tenha solução, ou seja, que $\det(\boldsymbol{\Gamma} - \mathbf{I}_m) = 0$, \mathbf{I}_m matriz identidade de ordem m (veja Feller [8], página 394)³. E também $\boldsymbol{\delta}$ é determinada sendo a distribuição estacionária da matriz $\boldsymbol{\Gamma}$ gerada, ou seja, $\boldsymbol{\delta}$ é a solução do sistema $\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}$ ⁴.

Rydén [34] também sugere a inicialização para os parâmetros da distribuição estado-dependente quando esta é normal univariada. Assim, para os parâmetros μ_i e σ_i da i -ésima densidade normal (ϕ_i) tem-se:

$$\mu_i \sim \text{Normal}\left(\frac{\min x_t + \max x_t}{2}, (\max x_t - \min x_t)^2\right)$$

$$\text{e } \sigma_i \sim \text{Uni}\left(0, \frac{\max |x_t|}{2}\right).$$

Aqui, sugere-se distribuições Uniformes para as médias e variâncias, os motivos serão explicados mais à frente. Seja $x_{(n)}$ a n -ésima estatística de ordem observada na amostra. Para gerar o valor inicial μ_i ($i = 1, 2, \dots, m$), que corresponde a média da i -ésima densidade normal (ϕ_i), utiliza-se

$$\mu_i \sim \text{Uni}\left[x_{((i-1)\frac{T}{m}+1)}, x_{(i\frac{T}{m})}\right].$$

²A função implementada `r.dirichlet`, Programa A.6 do Apêndice A, gera amostras de uma distribuição Dirichlet.

³“An irreducible aperiodic chain possesses an invariant probability distribution $\boldsymbol{\delta}$ if, and only if, it is ergodic. . . .” (Feller [8])

⁴A função implementada `trans.ergotica`, Programa A.7 do Apêndice A, gera a matriz de transição e a distribuição inicial descritas aqui.

Ou seja, a amostra primeiramente é ordenada, depois divide-se esta em m conjuntos com a mesma quantidade de elementos. Assim, tem-se m intervalos, não necessariamente do mesmo tamanho. As médias são obtidas de uma distribuição uniforme para cada intervalo obtido e, com isso, tem-se que conjuntos com menor variabilidade, ou regiões mais concentradas, geram intervalos menores, e o contrário ocorrendo em regiões cuja amostra é mais dispersa.

Assim garante-se duas coisas:

- i) o conjunto das m médias pode percorrer todo o intervalo $[x_{(1)}, x_{(T)}]$ da amostra observada;
- ii) regiões mais concentradas terão probabilidade maior de conter uma média.

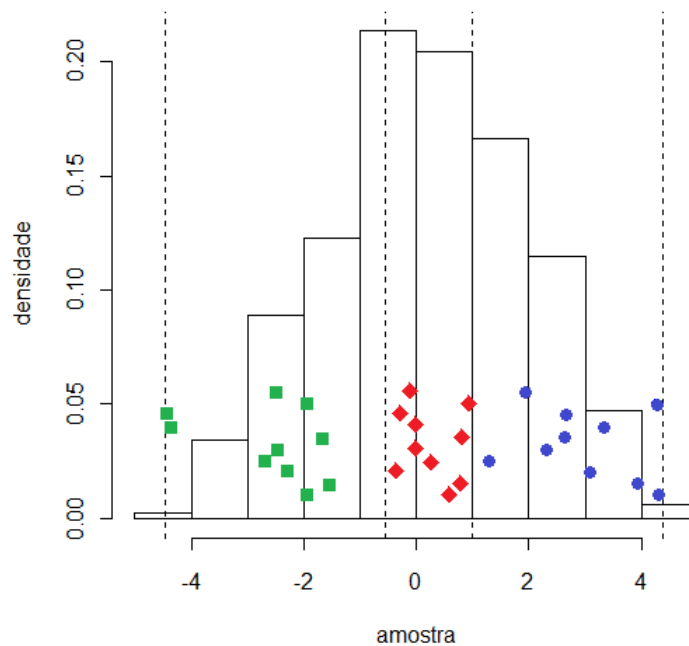


Figura 4.1: Histograma de uma amostra dividida em 3 grupos, com 10 conjuntos de pontos iniciais para as médias μ_i .

A Figura 4.1 exemplifica a divisão da amostra em diferentes grupos. Nele tem-se o histograma de uma amostra em que se deseja encontrar 3 distribuições. As linhas pontilhadas definem os 3 intervalos que irão gerar as médias, e 10 conjuntos com 3

médias foram geradas a partir de distribuições Uniformes para cada intervalo. Os pontos em verde (quadrado), vermelho (losango) e azul (círculo), marcam as médias iniciais para os intervalos 1, 2 e 3, respectivamente.

Agora, note que no segundo intervalo, que contém a região de maior concentração dos dados, os pontos vermelhos estão menos dispersos em comparação com os pontos verdes e azuis que estão nos primeiro e terceiro intervalos, respectivamente.

Este método para a escolha das médias iniciais tem duas vantagens principais em relação ao método proposto por Rydén [34]. A primeira é que no método de Rydén [34] existe a possibilidade de alguma média não pertencer ao intervalo dos valores observados na amostra, o que não faz muito sentido.

A segunda vantagem é que no método de Rydén [34] todas as m médias geradas estão mais concentradas no ponto médio do intervalo formado pelos pontos de mínimo e máximo da amostra, o que pode trazer um prejuízo para a estimação. Pois quanto mais diversificados os conjuntos de pontos iniciais forem, mais resultados diferentes podem ser obtidos para a função de verossimilhança. Isso diminui as chances de se obter pontos de máximo locais no lugar do ponto de máximo global que se almeja.

Agora, para σ_i^2 ($i = 1, 2, \dots, m$), variância da densidade normal, define-se $S_{\mathbf{x}^{(T)}}$ a variância amostral de $\mathbf{x}^{(T)}$ observada. Assim σ_i^2 é obtida por

$$e \sigma_i \sim \text{Uni} \left[\frac{1}{4} S_{\mathbf{x}^{(T)}}, 4 S_{\mathbf{x}^{(T)}} \right].$$

evitando que a variância seja zero ou muito próxima de zero, reduzindo os erros nos cálculos das probabilidades das densidades normais ⁵.

A Função `avalia.EM.uni`, Programa A.9 do Apêndice A, inicializa o algoritmo EM por um número N de conjuntos de pontos iniciais, como descrito nesta seção.

4.2 Seleção do modelo

Quando se trabalha com dados reais o mais comum é que não se conheça a complexidade do modelo que gerou os dados observados. Ou seja, além de não se saber o valor dos parâmetros do modelo, também não se tem informação sobre a dimensão

⁵A função implementada `mu.sigma2.iniciais`, Programa A.8 do Apêndice A, gera as médias e variâncias iniciais como foi sugerido aqui.

do espaço paramétrico, neste caso, determinado pelo número m de estado ocultos. O número m de estados é chamado de **ordem** do modelo.

Suponha que, para uma determinada série temporal univariada, estima-se um modelo HMM de ordem m com densidade estado-dependente normal. Assim, o número de parâmetros (p_m) a serem estimados é dado por

$$p_m = [(m - 1) + (m - 1)m] + (m + m) = m^2 + 2m - 1$$

ou seja, $(m-1)$ parâmetros da distribuição inicial $\boldsymbol{\delta}$, $(m-1)m$ parâmetros da matriz de transição $\boldsymbol{\Gamma}$, m médias e m variâncias das m densidades estado-dependentes normais.

Acrescentando uma unidade à ordem do modelo, o número de parâmetros a serem estimados é

$$p_{m+1} = [(m + 1 - 1) + (m + 1 - 1)(m + 1)] + (m + 1 + m + 1) = m^2 + 4m + 2.$$

Assim, o aumento de uma unidade na ordem do modelo gera um aumento de $p_{m+1} - p_m = 2m + 3$ unidades no número de parâmetros a serem estimados.

Fica claro que quanto maior a complexidade do modelo, maior é a informação dos dados que se consegue incorporar ao modelo. Contudo, deve-se fazer uso de parcimônia para determinar a complexidade do modelo, caso contrário estimaria-se um modelo de ordem $m \rightarrow \infty$, o que não é nada razoável e nem informativo. De fato, o que se deseja em um modelo é o balanceamento entre a qualidade do seu ajuste (informação dos dados que se capta com o modelo) e a sua complexidade (número de parâmetros envolvidos).

Outra situação que leva à escolha entre modelos concorrentes é o não conhecimento da distribuição estado-dependente dos dados. Pode-se, por exemplo, ter um modelo com distribuição Poisson e outro com distribuição binomial negativa. Neste caso, também deve-se ponderar o ajuste e complexidade de cada modelo.

Os critérios de seleção AIC (*Akaike information criterion*) e BIC (*Bayesian information criterion*), apresentados a seguir, levam em consideração ambos os fatores, ajuste e complexidade, para criar uma medida em que se possa basear a escolha de um modelo em detrimento de outros.

4.2.1 AIC e BIC

Suponha que a sequência $\mathbf{x}^{(T)} = (x_1, x_2, \dots, x_T)$ de observações seja gerada por um modelo f , desconhecido, e que esta sequência foi ajustada por duas famílias de modelos diferentes, $\{g_1 \in G_1\}$ e $\{g_2 \in G_2\}$. O que se almeja é escolher, entre os dois modelos, o que melhor representa as observações.

Este trabalho foca-se em duas formas de enfrentar a questão, uma com visão frequentista e outra Bayesiana. A visão frequentista tenta quantificar a discrepância entre o modelo verdadeiro f e os modelos G_1 e G_2 , aquele cuja diferença com o modelo verdadeiro for menor é o escolhido. O critério AIC (*Akaike information criterion*) quantifica esta diferença da seguinte maneira:

$$\begin{aligned} \text{AIC} &= -2\log L(G) + 2p \\ &= -2\log L + 2p \end{aligned} \tag{4.5}$$

onde $L = L(G)$ é a verossimilhança do modelo G ajustado e p o número de parâmetros do modelo.

A visão Bayesiana considera as probabilidades a priori do modelo f ser originado por G_1 e G_2 , os modelos aproximados e depois avalia as probabilidades a posteriori dada as observações $\mathbf{x}^{(T)}$. Este procedimento é feito pelo BIC (*Bayesian information criterion*) dado por

$$\begin{aligned} \text{BIC} &= -2\log L(G) + p \log T \\ &= -2\log L + p \log T \end{aligned} \tag{4.6}$$

onde L e p são os mesmos que em (4.5) e T é o número de observações.

Tanto o AIC quanto o BIC são compostos por dois termos, o primeiro, igual em ambos, mede o ajuste do modelo e decresce a medida que se aumenta a ordem m do modelo. O segundo termo, $2p$ em (4.5) e $p \log T$ em (4.6), penaliza o aumento do número de parâmetros, aumentando com a ordem do modelo. Para o BIC o aumento do número de parâmetros tem mais peso que no AIC quando $T > e^2 = 7,389$. Nestes casos, que são mais comuns, o BIC favorece os modelos de ordem menor que o AIC.

Apesar do AIC e BIC serem uma forma de seleção do modelo, Rydén [34] ressalta que ambos não fornecem informação sobre a confiança do modelo escolhido

em relação aos outros. Rydén apresenta o teste da razão de verossimilhança como alternativa, mas para o HMM não existe uma distribuição fechada para a estatística do teste, sendo assim necessária a utilização de técnicas de reamostragem, como o bootstrap, para a estimação desta distribuição. Entretanto, Rydén mostra que o gasto de tempo computacional pode chegar a 100h ao comparar modelos de ordem 3 e 4 usando 200 reamostragens, o que torna o processo exaustivo quando se tem um número grande de modelos para comparar. Por isso este trabalho considera apenas o AIC e o BIC para a seleção dos modelos.

4.3 Classificação

O segundo objetivo deste trabalho, sendo o primeiro a estimação dos sinais, é a classificação dos sinais EEG para identificar qual figura estava sendo observada enquanto os sinais eram captados. Em sistemas BCI (*Brain-Computer Interface*), baseados em sinais EEG, a utilização do HMM tem sido explorada por suas qualidades como classificador. Lotte et al [14] faz uma extensa revisão de algoritmos de classificação para BCI e destaca o HMM como um classificador **dinâmico** e **não-discriminativo**, o que lhe permite captar as alterações temporais da sequência, e o torna mais eficiente para rejeitar amostras duvidosas. O HMM faz parte de um conjunto de classificadores Bayesianos não-lineares.

Min e Luo [19], ao apresentarem aplicações médicas da classificação dos sinais EEG, citam o HMM com um modelo possível para modelar e classificar os sinais EEG. Pfurtscheller e Neuper [24] e Obermaier et al. [21] utilizam o HMM para a classificação do EEG em sistemas BCI, Obermaier et al. ainda ressalta que o HMM diminui o erro de classificação comparado com discriminante linear. Zhong e Ghosh [39] também fazem uso do HMM para classificar os sinais EEG, só que neste caso apresentam a classificação para dois eletrodos.

Além de sua utilização para a classificação, Xu et al. [38] e Oates et al. [20] aplicam o modelo em análise de agrupamento para séries temporais.

4.3.1 Método para a classificação

A classificação das sequências de observações é simples. Cada classe terá um modelo com conjunto de parâmetros λ , que foi previamente estimado. Uma sequência $\mathbf{x}^{(T)}$ será classificada como de uma determinada classe k se a verossimilhança desta sequência para o modelo desta classe ($L(\mathbf{x}^{(T)} | \lambda_k) = L_k$) for maior que as verossimilhanças de todos os outros modelos das outras classes. Ou seja, para K classes, a sequência $\mathbf{x}^{(T)}$ é da classe k se

$$L_k > L_l, \quad l \neq k = 1, 2, \dots, K. \quad (4.7)$$

Este trabalho pretende comparar a classificação de duas maneiras:

- i) as sequências serão classificadas segundo os modelos univariados estimados para cada eletrodo. Isso permite avaliar a qualidade da classificação dos eletrodos e assim reconhecer aqueles que podem ser mais significativos para a classificação.
- ii) as sequências serão classificadas baseadas na soma das verossimilhanças dos modelos univariados, e com isso notar se esse método melhora a qualidade da classificação comparando com os modelos univariados.

Outro ponto que será avaliado é se a quantidade de sequências utilizadas na estimação afeta a qualidade da classificação.

O exemplo a seguir exemplifica melhor o procedimento da classificação.

Exemplo 4.1. *Suponha três modelos: A, B e C, todos HMM univariados. Os modelos A e C são de ordem 2 ($m = 2$) e o modelo B de ordem 3 ($m = 3$). As densidades estado-dependente de todos os modelos são normais ($\phi(x; \mu, \sigma^2)$). E os parâmetros dos modelos são:*

a) modelo A: densidades estado-dependente $\phi_{1.A}(x; 0, 1)$ e $\phi_{2.A}(x; 1, 1)$,

$$\boldsymbol{\delta}_A = \begin{pmatrix} 0, 2 & 0, 8 \end{pmatrix} \quad e \quad \boldsymbol{\Gamma}_A = \begin{pmatrix} 0, 6 & 0, 4 \\ 0, 3 & 0, 7 \end{pmatrix}.$$

b) modelo B: densidades estado-dependente $\phi_{1.B}(x; 0, 1)$, $\phi_{2.B}(x; 1, 1)$ e $\phi_{3.B}(x; -1, 1)$,

$$\boldsymbol{\delta}_B = \begin{pmatrix} 0, 2 & 0, 2 & 0, 6 \end{pmatrix} \quad e \quad \boldsymbol{\Gamma}_B = \begin{pmatrix} 0, 1 & 0, 7 & 0, 2 \\ 0, 3 & 0, 2 & 0, 5 \\ 0, 6 & 0, 1 & 0, 3 \end{pmatrix}.$$

c) modelo C: densidades estado-dependente $\phi_{1.C}(x; -1, 1)$ e $\phi_{2.C}(x; 1, 1)$,

$$\delta_C = \begin{pmatrix} 0,5 & 0,5 \end{pmatrix} \quad e \quad \Gamma_C = \begin{pmatrix} 0,7 & 0,3 \\ 0,3 & 0,7 \end{pmatrix}.$$

sequência	origem	L_A	L_B	L_C	classificação
1	A	$6,75 \times 10^{-68}$	$1,63 \times 10^{-76}$	$4,65 \times 10^{-72}$	A
14	A	$2,04 \times 10^{-73}$	$2,79 \times 10^{-73}$	$1,19 \times 10^{-73}$	B
104	B	$2,37 \times 10^{-83}$	$1,77 \times 10^{-76}$	$5,99 \times 10^{-78}$	B
299	C	$9,30 \times 10^{-86}$	$1,55 \times 10^{-81}$	$2,88 \times 10^{-78}$	C
300	C	$1,67 \times 10^{-86}$	$1,01 \times 10^{-81}$	$8,67 \times 10^{-78}$	C

Tabela 4.1: Resultados para 5 das 300 sequências geradas no Exemplo 4.1.

Foram geradas ⁶ 300 sequências de tamanho 100 ($T = 100$), 100 de cada modelo, e calculadas ⁷ as verossimilhanças das sequências para cada modelo. A Tabela 4.1 mostra os resultados para as sequências 1, 14, 104, 299 e 300. Para exemplificar o procedimento da classificação observe os resultados da sequência 14: esta sequência foi originada pelo modelo A, mas a maior verossimilhança foi obtida com o modelo B, assim ela é classificada como do modelo B.

Modelo de origem	Classificação		
	A	B	C
A	99	1	0
B	1	95	4
C	1	7	92

Tabela 4.2: Resultados da classificação das 300 sequências do Exemplo 4.1.

A Tabela 4.2 apresenta o resumo das classificações. Note que, usando as verossimilhanças dos modelos HMM, foi possível classificar corretamente 286 ($99 + 95 + 92$) das 300 sequências, ou seja, aproximadamente 95,33% das sequências.

Feitas todas as considerações teóricas e apresentado um exemplo do processo de classificação, o Capítulo 5 faz um estudo por simulação para avaliação do método

⁶Utilizando a função `r.normal.hmm`, Programa A.1 do Apêndice A.

⁷Utilizando a função `veross`, Programa A.3 do Apêndice A.

de estimação proposto. No Capítulo 6 é realizado um estudo com dados reais obtidos pelo Laboratório de imagens biopotenciais (BIM-Lab) da Universidade do Texas em El Paso.

Capítulo 5

Simulação

Este capítulo apresenta a avaliação do método de estimação proposto: algoritmo EM inicializado por N diferentes conjuntos de pontos iniciais; e a performance dos critérios de seleção dos modelos: AIC e BIC. Neste estudo tenta-se apreender o comportamento destes diante da variação de três parâmetros, a saber: a ordem dos modelos (m), o número de pontos das sequências (T), e o número de sequências a serem utilizadas na estimação (S).

A cada combinação de parâmetros são geradas $R = 200$ amostras que então são estimadas sob as condições do estudo. Para tornar o processo mais ágil usou-se o conceito de programação em paralelo, aqui focada no software livre R. A seguir apresenta-se a idéia central deste tipo de programação, o pacote R que permite separar o processo em diferentes processadores e algumas dicas para o leitor que estiver interessado em reproduzir estes conceitos neste ou em outro tipo de problema similar.

5.1 Programação em Paralelo

Com o surgimento de computadores com múltiplos processadores a programação em paralelo tem se tornado foco de interesse de pesquisadores, estudantes e até empresas que precisam lidar com volumes grandes de dados e operações. Muitos podem ter a ingenuidade de pensar que basta possuir um destes computadores com 2, 3, ..., processadores que todos os seus problemas terão acabado, mas a solução pode não ser tão simples, pois programas como o R e o SAS não são capazes de sozinhos dividir

entre processadores as tarefas a serem realizadas.

Neste trabalho todas as funções utilizadas para a estimação foram implementadas no R, e neste programa o pacote `snow` possui as funções necessárias para a programação em paralelo. Para quem tiver interesse Rossini et al. [33], e mais recentemente McCallum e Weston [16], são boas referências sobre o assunto.

Dentro da programação, o ponto em que se divide as tarefas entre os múltiplos processadores pode ser diferente, e vai depender, basicamente, das características e dos objetivos do estudo. No mundo estatístico, os exemplos mais comuns de programação em paralelo são os que envolvem reamostragem, especificamente o bootstrap, procedimento que pode consumir muito tempo de processamento e em que são facilmente aplicados os conceitos de programação em paralelo.

Suponha que deseja-se simular uma amostra com 100.000 elementos independentes de uma determinada distribuição. Este problema pode ser resolvido de duas maneiras:

- A) sem dividir em diferentes processadores; ou
- B) dividindo as 100.000 tarefas entre os processadores.

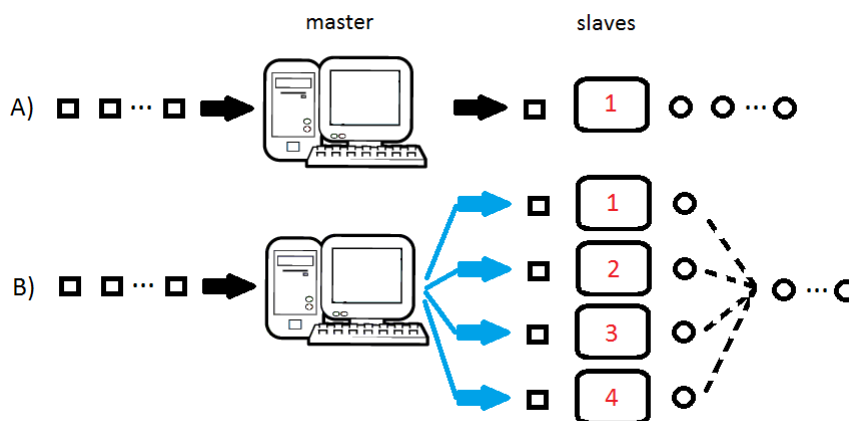


Figura 5.1: Exemplo da distribuição de tarefas em diferentes processadores.

Imagine que a máquina em que se está trabalhando tenha 4 processadores. A Figura 5.1 descreve visualmente as soluções A e B para o problema em questão. A máquina central é chamada de *master* e os processadores em que se divide as tarefas são chamados de *slaves*. O mecanismo que divide as tarefas (indicado pelas setas

azuis) e faz a comunicação master-slaves não está visível aos olhos do usuário e pode ser de três tipos: sockets, MPI e PVM.

Neste exemplo ao invés de gerar 1 ponto por vez geram-se 4 pontos. Assim no lugar de 1 amostra de tamanho 100.000, geram-se 4 amostras de tamanho 25.000 simultaneamente. Isso é possível pois temos uma amostra de elementos independentes e pode-se dividi-la e depois uní-la sem problemas. Mas, cuidado! Os processadores “virtuais” são apenas cópias de um verdadeiro. Se este for o caso, o processador “virtual” apenas gera cópias dos pontos gerados pelo processador verdadeiro. Uma maneira de resolver este problema é usando sementes (*seeds*) diferentes para cada processador.

Por outro lado, para problemas que envolvem operações matemáticas, por exemplo, não é preciso ter a mesma precaução. Para que se entenda melhor, suponha um polinômio $p(x) = ax^3 + bx + c$ e dois pontos y_1 e y_2 , como o objetivo é obter os valores $p(y_1)$ e $p(y_2)$ é possível “enviar” cada ponto para processadores diferentes onde o polinômio será calculado.

Finalmente, em alguns problemas, como o HMM, os elementos da sequência não são independentes, e por isso, não é possível desmembrá-la em diversos processadores. Como a ordem dos elementos é importante, a sequência inteira deve ser gerada de uma vez. E, para evitar sequências repetidas, no caso de haver processadores virtuais, as sequências são geradas antes de dividir o procedimento em diferentes processadores. A divisão é feita para o processo de estimação, que é o procedimento mais dispendioso, pois para cada amostra são realizadas $N = 50$ estimações pelo algoritmo EM com o propósito, já discutido, de evitar os pontos de máximo locais.

Neste estudo foram utilizados os computadores do sistema Virgo2¹ da UTEP. Neste sistema cada computador possui 8 processadores. As funções `avalia.estima` e `avalia.AICeBIC` (Programas A.12 e A.13, do Apêndice A) são organizadas nos

¹O Virgo2 é um *Beowulf cluster* (clusters de desempenho escaláveis, baseados numa infraestrutura de hardware comum, rede privada e software *open source* como o Linux) construído em 2007-2008. O Virgo2 é constituído por um computador *frontend*, que faz contato com o usuário, e 23 computadores que são os nós. Cada um destes computadores possui 2 processadores Intel Xeon E5430 2.66 GHz, 8 GB de memória principal e 917 GB de espaço em disco. Os computadores nós são contactados com o frontend por uma conexão Gigabit Ethernet. E o sistema operacional é o Rocks 5.

seguintes passos:

- Passo 1) Gera-se $nc = 8$ amostras, cada amostra tem a seguinte configuração: S sequências de tamanho T de um modelo HMM com m estados ocultos e densidades estado-dependentes normais. Para cada valor de m foi escolhido um único conjunto de parâmetros $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, μ_i e σ_i^2 ($i = 1, 2, \dots, m$).
- Passo 2) Cada amostra é estimada em um processador e seus resultados são guardados.
- Passo 3) Os passos 1 e 2 são repetidos até que se alcance $R = 200$ amostras estimadas, ou seja, 1 e 2 são repetidos 25 vezes.

Ao término das simulações, as $R = 200$ estimativas obtidas para os parâmetros do modelo HMM são analisadas para verificar quais configurações da amostra, ou seja, quantas sequências e quais os tamanhos destas, geram os melhores resultados para a estimação de modelos HMM de diferentes ordens.

Diante dos recursos disponíveis (tempo e máquina), este procedimento se mostrou mais eficiente. Em problemas diferentes e com recursos diferentes, outros procedimentos podem ser recomendáveis.

5.2 Parâmetros do modelo HMM e escolha dos valores de T e S

Em Zhong e Ghosh [39] os sinais utilizados no estudo tinham aproximadamente 250 pontos e em Xu et al. [38] os sinais tinham 200 pontos. Por isso um dos valores do parâmetro T escolhido foi 250. Mas, deseja-se saber o que acontece com a estimação dos parâmetros nos casos em que o número de pontos é menor (100) do que vem sendo utilizado. Dessa forma os valores escolhidos para o parâmetro T foram iguais a: 100 e 250.

Já para o número de sequências foram utilizadas S iguais a 1 e 2. A principal razão para esta escolha é que no experimento a ser estudado no Capítulo 6, onde para cada estímulo existem 4 sequências, é preciso ao menos uma sequência do sinal

para estimar cada um dos 10 estímulos estudados, e deseja-se ao menos 20 sinais para avaliar a classificação.

Para avaliar os critérios de seleção dos modelos AIC e BIC, não basta estimar o modelo para a ordem que se conhece verdadeira, deve-se também estimá-lo para ordens diferentes. Com isso é possível saber não somente se o critério erra na escolha da ordem, mas ao errar este escolhe ordens maiores ou menores do que a correta. Assim, no momento da avaliação desses critérios para cada valor verdadeiro de m as amostras são estimadas para modelos de ordens $m-1$, m e $m+1$, quando possível, já que a ordem deve ser no mínimo 1. E, para cada ordem m do HMM foram definidos valores para os parâmetros $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, μ_i e σ_i^2 ($i = 1, 2, \dots, m$). Estes são:

- $m = 1$: No caso de $m = 1$ a distribuição inicial e a matriz de transição são, obviamente, o escalar 1. Assim temos,

$$\boldsymbol{\delta} = 1, \quad \boldsymbol{\Gamma} = 1, \quad \text{e} \quad \mu_1 = 0, \quad \sigma_1^2 = 1 .$$

E o intervalo dos valores de m para os quais as amostras são estimadas é $m = \{1, 2\}$.

- $m = 2$: Para $m = 2$ temos os valores

$$\boldsymbol{\delta} = \begin{pmatrix} 0,3 & 0,7 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} 0,4 & 0,6 \\ 0,7 & 0,3 \end{pmatrix}, \quad \text{e} \quad \begin{matrix} \mu_1 = 0, & \sigma_1^2 = 1 \\ \mu_2 = 1, & \sigma_2^2 = 2 \end{matrix} .$$

E o intervalo dos valores de m para os quais as amostras são estimadas é $m = \{1, 2, 3\}$.

- $m = 3$: Para $m = 3$ temos os valores

$$\boldsymbol{\delta} = \begin{pmatrix} 0,3 & 0,5 & 0,2 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} 0,6 & 0,2 & 0,2 \\ 0,4 & 0,4 & 0,2 \\ 0,1 & 0,3 & 0,6 \end{pmatrix}, \quad \text{e} \quad \begin{matrix} \mu_1 = -1, & \sigma_1^2 = 2 \\ \mu_2 = 0, & \sigma_2^2 = 1 \\ \mu_3 = 1, & \sigma_3^2 = 1 \end{matrix} .$$

E o intervalo dos valores de m para os quais as amostras são estimadas é $m = \{2, 3, 4\}$.

- $m = 4$: Para $m = 4$ temos os valores

$$\boldsymbol{\delta} = \begin{pmatrix} 0,2 & 0,3 & 0,1 & 0,4 \end{pmatrix},$$

$$\mathbf{\Gamma} = \begin{pmatrix} 0,2 & 0,2 & 0,2 & 0,4 \\ 0,3 & 0,1 & 0,5 & 0,1 \\ 0,1 & 0,4 & 0,2 & 0,3 \\ 0,1 & 0,3 & 0,4 & 0,2 \end{pmatrix}, \quad \text{e} \quad \begin{matrix} \mu_1 = -2, & \sigma_1^2 = 0,5 \\ \mu_2 = -1, & \sigma_2^2 = 1 \\ \mu_3 = 0, & \sigma_3^2 = 1 \\ \mu_4 = 1, & \sigma_4^2 = 2 \end{matrix}.$$

E o intervalo dos valores de m para os quais as amostras são estimadas é $m = \{3, 4, 5\}$.

- $m = 5$: Para $m = 5$ temos os valores

$$\boldsymbol{\delta} = \begin{pmatrix} 0,2 & 0,1 & 0,1 & 0,3 & 0,3 \end{pmatrix},$$

$$\mathbf{\Gamma} = \begin{pmatrix} 0,3 & 0,2 & 0,2 & 0,1 & 0,2 \\ 0,1 & 0,3 & 0,2 & 0,1 & 0,3 \\ 0,2 & 0,1 & 0,2 & 0,3 & 0,2 \\ 0,1 & 0,1 & 0,4 & 0,2 & 0,2 \\ 0,2 & 0,2 & 0,2 & 0,3 & 0,1 \end{pmatrix}, \quad \text{e} \quad \begin{matrix} \mu_1 = -2, & \sigma_1^2 = 0,5 \\ \mu_2 = -1, & \sigma_2^2 = 1 \\ \mu_3 = 0, & \sigma_3^2 = 1 \\ \mu_4 = 1, & \sigma_4^2 = 0,5 \\ \mu_5 = 2, & \sigma_5^2 = 1 \end{matrix}.$$

E o intervalo dos valores de m para os quais as amostras são estimadas é $m = \{4, 5, 6\}$.

As seções seguintes apresentam os resultados obtidos nas simulações.

5.3 Avaliação da estimação

Para cada modelo de ordem m , e combinação dos parâmetros S e T , foram estimadas $R = 200$ amostras dos modelos. A estimação é realizada utilizando o algoritmo EM inicializado por $N = 50$ diferentes conjuntos de pontos iniciais $\boldsymbol{\delta}^{(0)}$, $\mathbf{\Gamma}^{(0)}$, $\mu_i^{(0)}$ e $(\sigma_i^2)^{(0)}$, $i = 1, 2, \dots, m$.

A seguir apresentam-se os resultados obtidos para cada modelo de ordem m .

5.3.1 Ordem $m = 1$

Para um modelo de ordem $m = 1$ não faz sentido estudar os parâmetros $\boldsymbol{\delta}$ e $\mathbf{\Gamma}$, já que ambos são iguais a 1, ou seja, tem-se uma amostra i.i.d.. Aqui apenas a média e a variância da distribuição normal são avaliadas.

Média

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
μ_1	0	100	1	0,003	0,105	-0,206	0,213
		100	2	0	0,066	-0,132	0,132
		250	1	-0,002	0,068	-0,138	0,134
		250	2	0,001	0,046	-0,092	0,093

Tabela 5.1: Resumo dos valores obtidos para a média com $m = 1$.

As Tabelas 5.1 e 5.2 mostram o comportamento das estimativas dos parâmetros avaliados, contendo o valor verdadeiro do parâmetro, a média e erro padrão das estimativas, e os limites inferior e superior do intervalo para o parâmetro. Os limites são construídos tomando a média das estimativas mais (e menos) duas vezes o erro padrão.

Note que tanto para a média quanto para a variância as médias das estimativas estão próximas dos valores verdadeiros e os intervalos construídos contêm estes valores. Perceba também que ao fixar o parâmetro T e aumentar o S o erro padrão das estimativas diminui. O mesmo ocorre ao fixar o S e aumentar o T . Isso mostra que estes aumentos melhoram a precisão dos parâmetros.

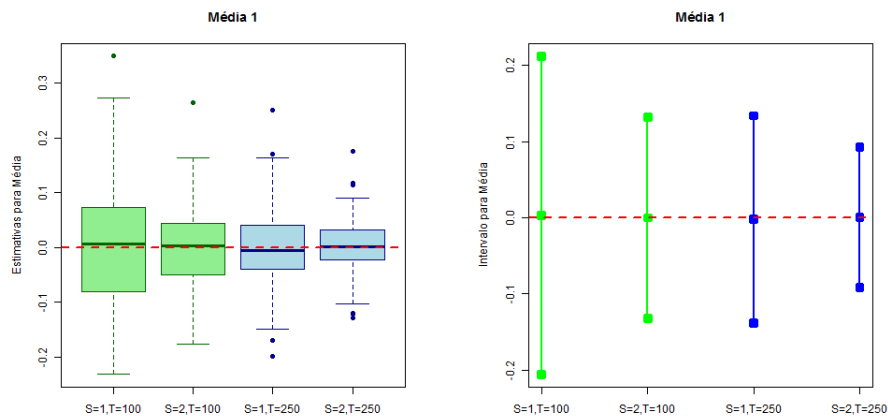


Figura 5.2: Box-plots das estimativas e intervalos para a média, com $m = 1$

Os gráficos das Figuras 5.2 e 5.3 possibilitam ver a distribuição das estimativas dos dois parâmetros por meio de box-plots e a posição dos valores verdadeiros dos

Variância

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
σ_1^2	0	100	1	0,988	0,151	0,687	1,289
		100	2	1,004	0,109	0,786	1,223
		250	1	1,003	0,084	0,835	1,172
		250	2	0,999	0,068	0,864	1,134

Tabela 5.2: Resumo dos valores obtidos para a variância com $m = 1$.

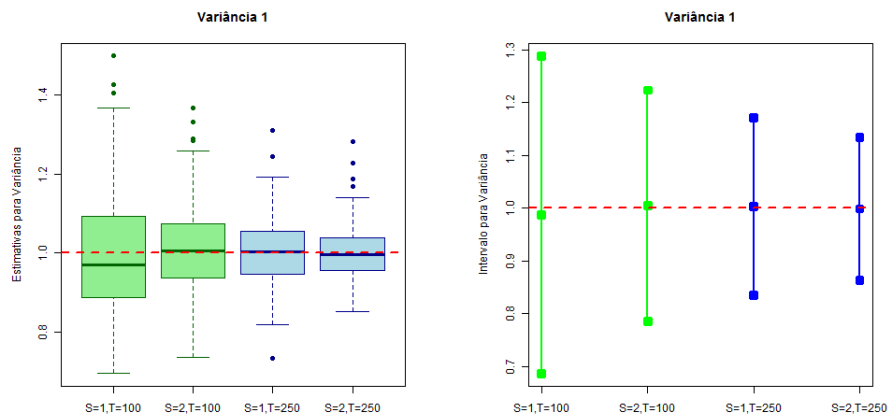


Figura 5.3: Box-plots das estimativas e intervalos para a variância, com $m = 1$

parâmetros nos intervalos construídos e nos box-plots.

5.3.2 Ordem $m = 2$

Para os parâmetros relacionados às densidades estado-dependentes normais note, pelas Tabelas 5.3 e 5.4, que ao fixar o parâmetros T e aumentar o parâmetros S o erro padrão das estimativas diminui em todos os casos. Agora, quando fixa-se o parâmetro S e aumenta-se o parâmetro T o erro padrão só não diminui para a média μ_2 .

As Figuras 5.4 e 5.5 contêm os gráficos box-plots das estimativas e o intervalo para os parâmetros μ_i e σ_i^2 , $i = 1, 2$. Note, pelo box-plot para a média μ_2 que o erro padrão para o caso $S = 2$ e $T = 250$ aumentou por causa dos dois *outliers* destacados no gráfico.

Médias

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
μ_1	0	100	1	-0,143	0,53	-1,204	0,918
		100	2	-0,053	0,42	-0,892	0,786
		250	1	-0,038	0,406	-0,85	0,773
		250	2	0,009	0,249	-0,489	0,508
μ_2	1	100	1	1,569	0,963	-0,357	3,495
		100	2	1,489	0,92	-0,35	3,329
		250	1	1,583	1,023	-0,464	3,63
		250	2	1,58	0,909	-0,238	3,398

Tabela 5.3: Resumo dos valores obtidos para as médias com $m = 2$.

Variâncias

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
σ_1^2	1	100	1	0,823	0,52	-0,218	1,863
		100	2	0,901	0,427	0,048	1,755
		250	1	0,926	0,475	-0,023	1,876
		250	2	0,943	0,355	0,234	1,653
σ_2^2	2	100	1	1,452	1,243	-1,033	3,937
		100	2	1,49	0,769	-0,048	3,028
		250	1	1,507	1,079	-0,651	3,664
		250	2	1,544	0,649	0,246	2,841

Tabela 5.4: Resumo dos valores obtidos para a variâncias com $m = 2$.

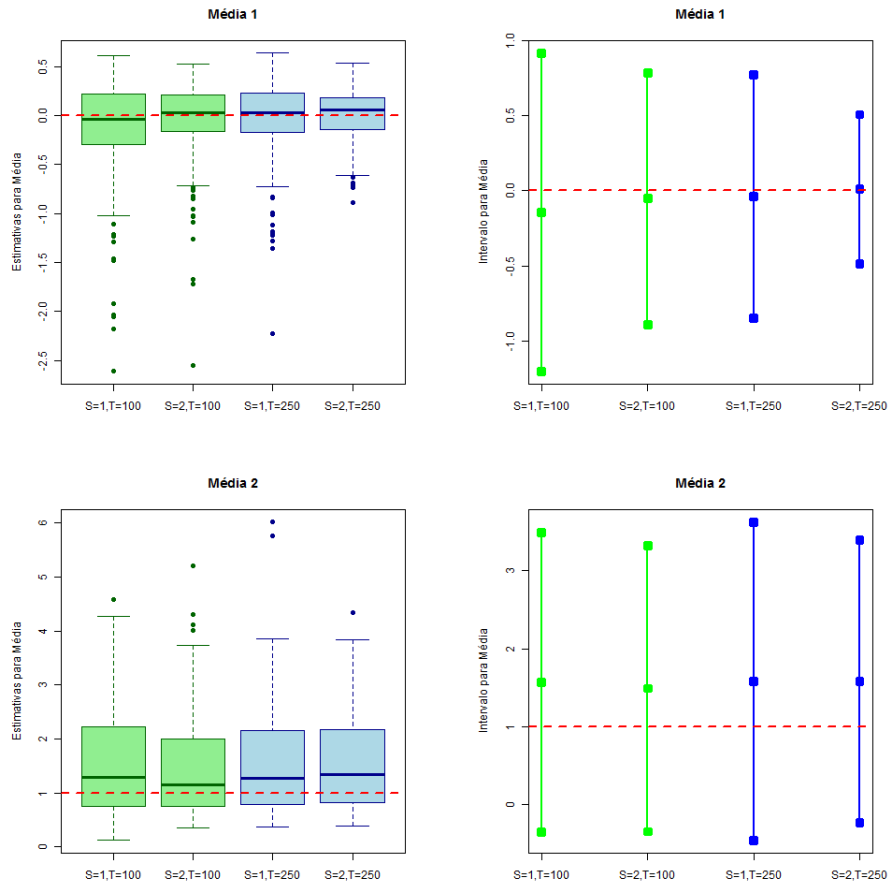


Figura 5.4: Box-plots das estimativas e intervalos para as médias μ_1 e μ_2 , com $m = 2$

Pela Tabela 5.5, que contém os resultados obtidos para a distribuição inicial δ , note que as médias das estimativas não estão próximas dos valores reais. Levando em consideração que estes são valores que variam no intervalo $[0, 1]$, os erros padrões observados são elevados, e por isso, mesmo que os intervalos contenham os valores verdadeiros dos parâmetros, não é sensato dizer que estas são boas estimativas. Isso faz sentido pois, para os casos $S = 1$ e $S = 2$, apenas 1 e 2 observações, respectivamente, são utilizadas para a estimação destes parâmetros.

A Figura 5.6 contém os gráficos box-plots das estimativas e intervalos para os parâmetros δ_i , $i = 1, 2$, note por eles que a caixa dos box-plots compreendem todo o intervalo $[0, 1]$.

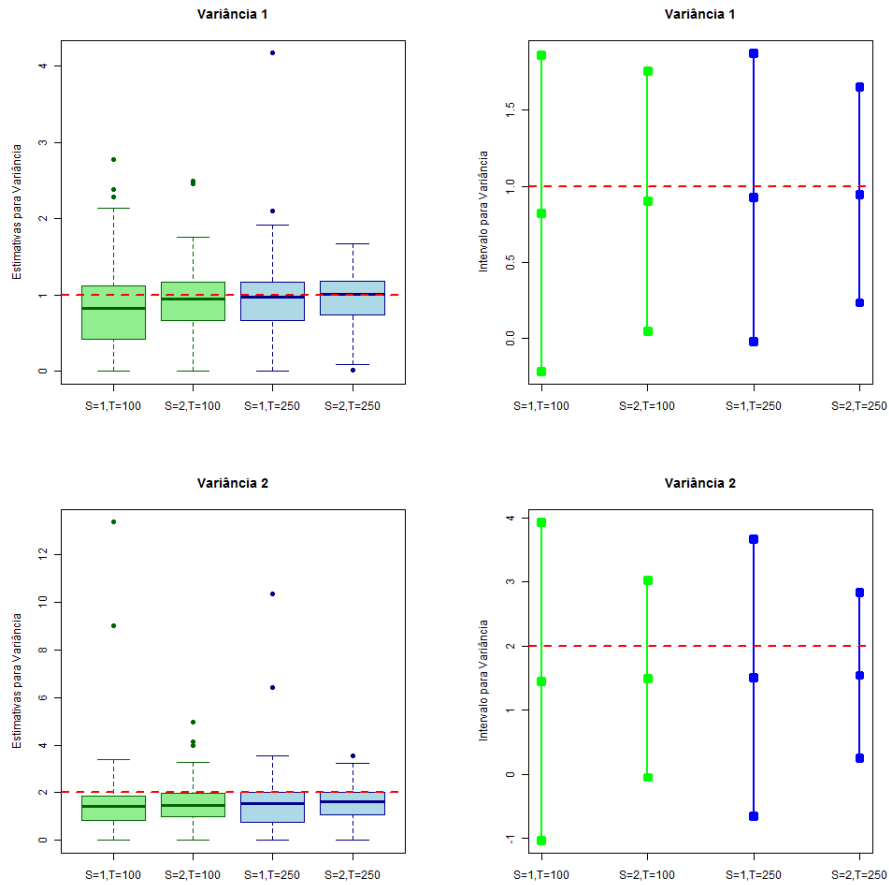


Figura 5.5: Box-plots das estimativas e intervalos para as variâncias σ_1^2 e σ_2^2 , com $m = 2$

As Tabelas 5.6 e 5.7 possuem os resultados obtidos para as probabilidades de transição γ_{ij} . Note por elas, que as médias das estimativas para as probabilidades $\gamma_{1.}$, da Tabela 5.6, não estão próximas dos valores verdadeiros. Já as médias das estimativas das probabilidades $\gamma_{2.}$ estão mais próximas dos valores verdadeiros.

Para as probabilidades de transição o erro padrão diminui com o aumento dos parâmetros T e S , com excessão do parâmetro γ_{12} cujo erro padrão aumenta quando $S = 1$ é fixo e o parâmetros T aumenta. As Figuras 5.7 e 5.8 contêm os gráficos box-plots para as estimativas e intervalos para os parâmetros γ_{ij} , $i, j = 1, 2$.

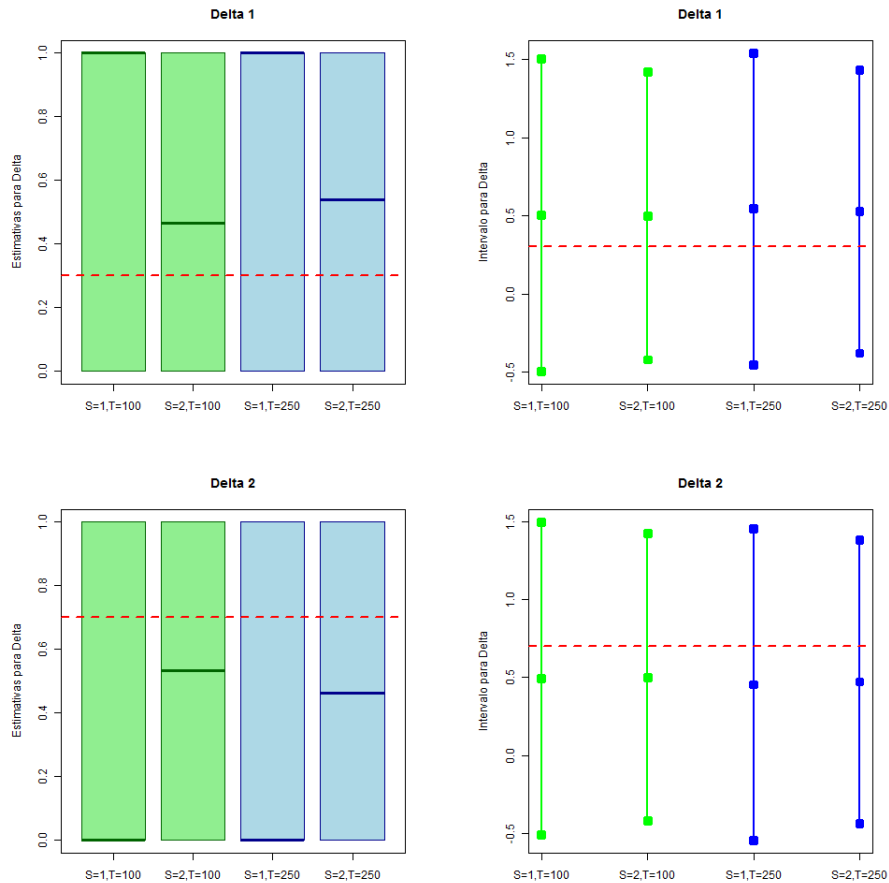


Figura 5.6: Box-plots das estimativas e intervalos para as distribuições iniciais δ_1 e δ_2 , com $m = 2$

5.3.3 Ordem $m = 3$

A Tabela 5.8 resume os resultados obtidos para os parâmetros μ_i , $i = 1, 2, 3$, e a Figura 5.9 contém os box-plots para as estimativas e os intervalos para as médias das três densidades normais. Pela Tabela nota-se que a média das estimativas para μ_3 está longe do valor verdadeiro do parâmetro.

Pela Tabela 5.8, também percebe-se que para os valores: $(\mu_1, S = 2 \text{ e } T = 250)$, $(\mu_2, S = 1 \text{ e } T = 100)$ e $(\mu_3, S = 1 \text{ e } T = 250)$, os erros padrões não diminuem com o aumento dos parâmetros S e T . Nestes casos é possível notar pelos gráficos box-plots que os erros padrões foram maiores por causa da presença de pontos *outliers*.

Distribuição Inicial (δ)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
δ_1	0,3	100	1	0,505	0,501	-0,497	1,507
		100	2	0,498	0,461	-0,423	1,42
		250	1	0,545	0,499	-0,453	1,543
		250	2	0,527	0,454	-0,38	1,435
δ_2	0,7	100	1	0,495	0,501	-0,507	1,497
		100	2	0,502	0,461	-0,42	1,423
		250	1	0,455	0,499	-0,543	1,453
		250	2	0,473	0,454	-0,435	1,38

Tabela 5.5: Resumo dos valores obtidos para a distribuição inicial com $m = 2$.Probabilidades de Transição ($\gamma_{1.}$)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{11}	0,4	100	1	0,53	0,346	-0,161	1,222
		100	2	0,513	0,339	-0,164	1,191
		250	1	0,534	0,358	-0,182	1,249
		250	2	0,549	0,321	-0,093	1,192
γ_{12}	0,6	100	1	0,47	0,346	-0,222	1,161
		100	2	0,487	0,339	-0,191	1,164
		250	1	0,466	0,358	-0,249	1,182
		250	2	0,451	0,321	-0,192	1,093

Tabela 5.6: Resumo dos valores obtidos para a as probabilidades de transição $\gamma_{1.}$, com $m = 2$.

Probabilidades de Transição (γ_2 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{21}	0,7	100	1	0,639	0,355	-0,072	1,35
		100	2	0,685	0,345	-0,005	1,375
		250	1	0,698	0,328	0,041	1,354
		250	2	0,723	0,306	0,111	1,335
γ_{22}	0,3	100	1	0,361	0,355	-0,35	1,072
		100	2	0,315	0,345	-0,375	1,005
	0,3	250	1	0,302	0,328	-0,354	0,959
		250	2	0,277	0,306	-0,335	0,889

Tabela 5.7: Resumo dos valores obtidos para as probabilidades de transição γ_2 , com $m = 2$.

Médias

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
μ_1	-1	100	1	-1,709	0,924	-3,557	0,138
		100	2	-1,632	0,88	-3,393	0,128
		250	1	-1,68	0,921	-3,521	0,162
		250	2	-1,715	1,069	-3,853	0,424
μ_2	0	100	1	-0,18	0,6	-1,38	1,021
		100	2	-0,253	0,647	-1,548	1,041
		250	1	-0,123	0,575	-1,273	1,028
		250	2	-0,223	0,562	-1,347	0,9
μ_3	1	100	1	1,141	0,571	-0,001	2,282
		100	2	1,095	0,553	-0,01	2,201
		250	1	1,074	0,572	-0,07	2,218
		250	2	1,063	0,491	0,082	2,045

Tabela 5.8: Resumo dos valores obtidos para as médias com $m = 3$.

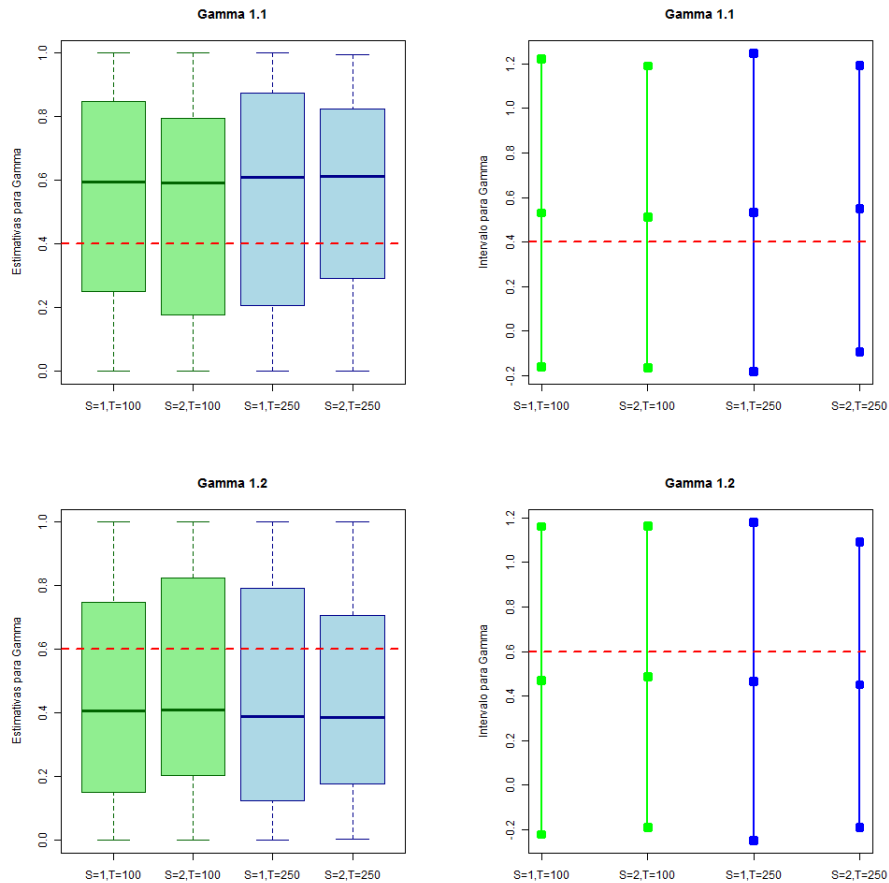


Figura 5.7: Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{1.}$, com $m = 2$

Para as variâncias das três densidades normais, nota-se pela Tabela 5.9 que as médias das estimativas para σ_1^2 estão mais longe do valor verdadeiro. O erro padrão para o caso em que σ_1^2 , $T = 250$ e $S = 2$ é o maior, novamente pela presença de *outlier*, que pode ser observado na Figura 5.10.

Pela Tabela 5.10 nota-se que as médias das estimativas estão mais próximas dos valores verdadeiros, e que os erros padrões diminuem com o aumento de T e S , com exceção do caso δ_1 quando fixa-se $S = 1$ e aumenta o valor de T . O mesmo comportamento pode ser observado na Figura 5.11, que contém os gráficos box-plots e intervalos para os parâmetros.

Variâncias

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
σ_1^2	2	100	1	1,084	0,76	-0,436	2,603
		100	2	1,248	0,761	-0,275	2,771
		250	1	1,373	0,694	-0,014	2,76
		250	2	1,454	0,873	-0,292	3,2
σ_2^2	1	100	1	0,817	0,829	-0,84	2,475
		100	2	0,891	0,822	-0,754	2,535
		250	1	0,823	0,64	-0,458	2,103
		250	2	0,892	0,587	-0,282	2,065
σ_3^2	1	100	1	0,748	0,525	-0,302	1,797
		100	2	0,781	0,465	-0,149	1,711
		250	1	0,859	0,468	-0,077	1,794
		250	2	0,856	0,387	0,082	1,629

Tabela 5.9: Resumo dos valores obtidos para a variâncias com $m = 3$.Distribuição Inicial (δ)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
δ_1	0,3	100	1	0,255	0,437	-0,619	1,129
		100	2	0,235	0,373	-0,511	0,981
		250	1	0,275	0,448	-0,62	1,17
		250	2	0,265	0,401	-0,537	1,067
δ_2	0,5	100	1	0,39	0,489	-0,588	1,368
		100	2	0,426	0,452	-0,478	1,329
		250	1	0,41	0,493	-0,576	1,396
		250	2	0,384	0,45	-0,515	1,283
δ_3	0,2	100	1	0,355	0,48	-0,604	1,314
		100	2	0,339	0,417	-0,494	1,173
		250	1	0,315	0,466	-0,616	1,246
		250	2	0,351	0,434	-0,517	1,219

Tabela 5.10: Resumo dos valores obtidos para a distribuição inicial com $m = 3$.

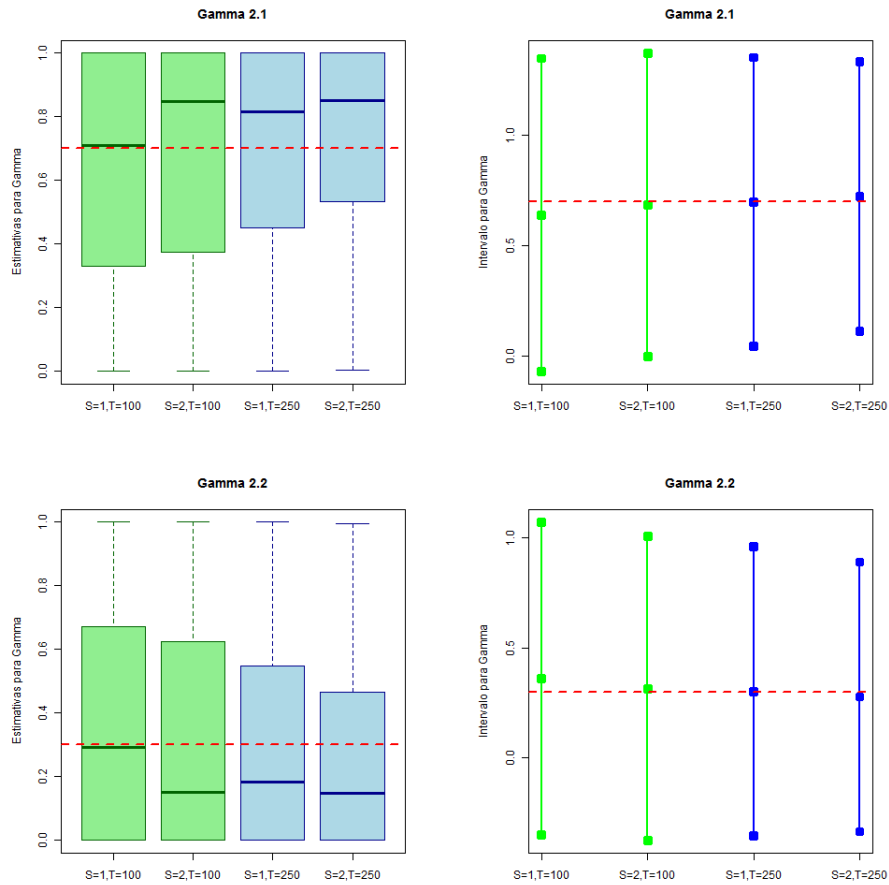


Figura 5.8: Box-plots das estimativas e intervalos para as probabilidades de transição γ_2 , com $m = 2$

Pelas Tabelas 5.11 e 5.12, e Figuras 5.12, 5.13 e 5.14 percebe-se que entre os três vetores de probabilidades de transição γ_1 , γ_2 e γ_3 , o que possui as melhores estimativas é o γ_3 . Para este, a média das estimativas estão mais próximas dos valores verdadeiros dos parâmetros e o comportamento do erro padrão é o esperado, ou seja, diminui com o aumento dos parâmetros T e S .

Probabilidades de Transição (γ_1 .)

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{11}	0,6	100	1	0,39	0,306	-0,222	1,002
		100	2	0,427	0,307	-0,186	1,041
		250	1	0,448	0,311	-0,175	1,071
		250	2	0,446	0,3	-0,154	1,045
γ_{12}	0,2	100	1	0,354	0,34	-0,327	1,035
		100	2	0,267	0,28	-0,293	0,828
		250	1	0,297	0,297	-0,296	0,891
		250	2	0,3	0,297	-0,294	0,894
γ_{13}	0,2	100	1	0,256	0,282	-0,308	0,819
		100	2	0,305	0,294	-0,284	0,894
		250	1	0,255	0,277	-0,298	0,808
		250	2	0,254	0,261	-0,267	0,775

Tabela 5.11: Resumo dos valores obtidos para a as probabilidades de transição γ_1 ., com $m = 3$.

Probabilidades de Transição (γ_2 .)

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{21}	0,4	100	1	0,295	0,309	-0,323	0,914
		100	2	0,331	0,302	-0,272	0,935
		250	1	0,333	0,294	-0,255	0,921
		250	2	0,287	0,26	-0,233	0,807
γ_{22}	0,4	100	1	0,364	0,334	-0,303	1,031
		100	2	0,416	0,335	-0,254	1,085
		250	1	0,394	0,317	-0,239	1,027
		250	2	0,468	0,305	-0,142	1,078
γ_{23}	0,2	100	1	0,341	0,321	-0,3	0,982
		100	2	0,253	0,289	-0,326	0,832
		250	1	0,273	0,29	-0,306	0,853
		250	2	0,245	0,235	-0,224	0,715

Tabela 5.12: Resumo dos valores obtidos para a as probabilidades de transição γ_2 ., com $m = 3$.

Probabilidades de Transição (γ_3 .)

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{31}	0,1	100	1	0,163	0,214	-0,266	0,591
		100	2	0,113	0,169	-0,224	0,45
		250	1	0,118	0,179	-0,239	0,475
		250	2	0,096	0,152	-0,207	0,4
γ_{32}	0,3	100	1	0,375	0,345	-0,314	1,064
		100	2	0,344	0,287	-0,229	0,917
		250	1	0,367	0,306	-0,246	0,98
		250	2	0,33	0,264	-0,197	0,857
γ_{33}	0,6	100	1	0,463	0,321	-0,18	1,105
		100	2	0,542	0,291	-0,04	1,125
		250	1	0,515	0,307	-0,098	1,129
		250	2	0,574	0,266	0,042	1,105

Tabela 5.13: Resumo dos valores obtidos para a as probabilidades de transição γ_3 ., com $m = 3$.

Médias							
parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
μ_1	-2	100	1	-1,937	0,517	-2,971	-0,903
		100	2	-1,976	0,45	-2,876	-1,077
		250	1	-1,956	0,41	-2,775	-1,137
		250	2	-1,969	0,428	-2,824	-1,113
μ_2	-1	100	1	-0,849	0,515	-1,879	0,181
		100	2	-0,798	0,484	-1,766	0,169
		250	1	-0,824	0,523	-1,87	0,223
		250	2	-0,772	0,496	-1,763	0,219
μ_3	0	100	1	0,311	0,5	-0,689	1,312
		100	2	0,236	0,455	-0,675	1,146
		250	1	0,171	0,497	-0,823	1,166
		250	2	0,221	0,445	-0,669	1,11
μ_4	1	100	1	1,67	0,859	-0,048	3,388
		100	2	1,785	0,934	-0,084	3,654
		250	1	1,75	0,85	0,051	3,449
		250	2	1,684	0,883	-0,083	3,45

Tabela 5.14: Resumo dos valores obtidos para as médias com $m = 4$.

5.3.4 Ordem $m = 4$

Pela Tabela 5.14 é possível perceber que as médias das estimativas estão próximas dos valores verdadeiros dos parâmetros μ_i , $i = 1, 2, 3, 4$. Os casos em que o erro padrão deveria diminuir com o aumento dos parâmetros T e S e isso não ocorre, são justificados pelos *outliers* existentes².

Pelos resultados obtidos para as variâncias σ_i^2 , $i = 1, 2, 3, 4$, contidos na Tabela 5.15, parece-se que para as três primeiras variâncias os resultados parecem adequados: as médias das estimativas estão próximas dos valores reais dos parâmetros e os erros padrões diminuem com o aumento de T e S . Contudo, para a quarta variância, σ_4^2 , as médias das estimativas estão bem abaixo do valor verdadeiro (2).

²Os gráficos para a distribuição das estimativas podem ser obtidos sob consulta.

Variâncias

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
σ_1^2	0,5	100	1	0,408	0,389	-0,37	1,186
		100	2	0,439	0,389	-0,339	1,217
		250	1	0,445	0,36	-0,275	1,166
		250	2	0,427	0,313	-0,199	1,053
σ_2^2	1	100	1	0,754	0,737	-0,719	2,227
		100	2	0,738	0,575	-0,412	1,888
		250	1	0,767	0,643	-0,519	2,053
		250	2	0,864	0,631	-0,398	2,127
σ_3^2	1	100	1	0,971	1,049	-1,126	3,068
		100	2	0,883	0,809	-0,734	2,5
		250	1	0,975	0,806	-0,638	2,588
		250	2	0,969	0,655	-0,341	2,279
σ_4^2	2	100	1	1,092	1,084	-1,076	3,261
		100	2	1,042	0,746	-0,451	2,534
		250	1	1,066	0,729	-0,392	2,525
		250	2	1,261	0,73	-0,199	2,722

Tabela 5.15: Resumo dos valores obtidos para a variâncias com $m = 4$.

Distribuição Inicial (δ)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
δ_1	0,2	100	1	0,325	0,47	-0,614	1,264
		100	2	0,29	0,381	-0,473	1,052
		250	1	0,305	0,462	-0,618	1,228
		250	2	0,271	0,37	-0,469	1,01
δ_2	0,3	100	1	0,235	0,425	-0,615	1,085
		100	2	0,243	0,373	-0,504	0,989
		250	1	0,175	0,381	-0,587	0,937
		250	2	0,205	0,349	-0,493	0,903
δ_3	0,1	100	1	0,185	0,389	-0,594	0,964
		100	2	0,23	0,358	-0,487	0,947
		250	1	0,26	0,44	-0,619	1,139
		250	2	0,24	0,374	-0,508	0,988
δ_4	0,4	100	1	0,255	0,437	-0,619	1,129
		100	2	0,238	0,334	-0,431	0,906
		250	1	0,26	0,44	-0,619	1,139
		250	2	0,284	0,393	-0,503	1,071

Tabela 5.16: Resumo dos valores obtidos para a distribuição inicial com $m = 4$.

Para a distribuição inicial, a Tabela 5.16 revela o comportamento esperado para os erros padrões, mas as médias das estimativas não parecem muito adequadas para estimar os parâmetros.

Pelas Tabelas 5.17, 5.18, 5.19 e 5.20 nota-se que, apesar dos erros padrões diminuírem com o aumento de T e S , as médias das estimativas não parecem adequadas para estimar as probabilidades de transição.

Probabilidades de Transição (γ_1 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{11}	0,2	100	1	0,219	0,232	-0,246	0,683
		100	2	0,224	0,207	-0,191	0,639
		250	1	0,223	0,218	-0,213	0,659
		250	2	0,226	0,199	-0,172	0,624
γ_{12}	0,2	100	1	0,267	0,304	-0,342	0,876
		100	2	0,27	0,288	-0,306	0,845
		250	1	0,252	0,289	-0,326	0,831
		250	2	0,244	0,287	-0,329	0,818
γ_{13}	0,2	100	1	0,309	0,308	-0,308	0,926
		100	2	0,275	0,291	-0,307	0,858
		250	1	0,29	0,281	-0,272	0,853
		250	2	0,278	0,289	-0,301	0,857
γ_{14}	0,4	100	1	0,205	0,261	-0,317	0,727
		100	2	0,231	0,242	-0,254	0,715
		250	1	0,234	0,251	-0,268	0,736
		250	2	0,252	0,254	-0,256	0,759

Tabela 5.17: Resumo dos valores obtidos para a as probabilidades de transição γ_1 ., com $m = 4$.

Probabilidades de Transição (γ_2 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{21}	0,3	100	1	0,271	0,302	-0,332	0,874
		100	2	0,249	0,27	-0,292	0,789
		250	1	0,275	0,278	-0,282	0,831
		250	2	0,305	0,265	-0,224	0,835
γ_{22}	0,1	100	1	0,248	0,296	-0,344	0,84
		100	2	0,279	0,299	-0,32	0,878
		250	1	0,257	0,292	-0,326	0,841
		250	2	0,289	0,297	-0,305	0,883
γ_{23}	0,5	100	1	0,216	0,284	-0,352	0,784
		100	2	0,276	0,309	-0,341	0,893
		250	1	0,282	0,305	-0,327	0,892
		250	2	0,252	0,273	-0,294	0,799
γ_{24}	0,1	100	1	0,265	0,328	-0,391	0,921
		100	2	0,196	0,279	-0,361	0,753
		250	1	0,185	0,263	-0,341	0,712
		250	2	0,153	0,205	-0,257	0,563

Tabela 5.18: Resumo dos valores obtidos para a as probabilidades de transição γ_2 ., com $m = 4$.

Probabilidades de Transição (γ_3 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{31}	0,1	100	1	0,223	0,293	-0,364	0,81
		100	2	0,249	0,273	-0,296	0,794
		250	1	0,208	0,241	-0,273	0,69
		250	2	0,165	0,207	-0,249	0,578
γ_{32}	0,4	100	1	0,3	0,335	-0,37	0,969
		100	2	0,263	0,3	-0,336	0,863
		250	1	0,315	0,333	-0,351	0,981
		250	2	0,323	0,316	-0,31	0,956
γ_{33}	0,2	100	1	0,259	0,31	-0,361	0,878
		100	2	0,285	0,283	-0,281	0,852
		250	1	0,29	0,304	-0,317	0,897
		250	2	0,321	0,309	-0,298	0,94
γ_{34}	0,3	100	1	0,218	0,286	-0,353	0,79
		100	2	0,203	0,268	-0,333	0,738
		250	1	0,186	0,236	-0,287	0,659
		250	2	0,191	0,247	-0,304	0,686

Tabela 5.19: Resumo dos valores obtidos para a as probabilidades de transição γ_3 ., com $m = 4$.

Probabilidades de Transição (γ_4 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{41}	0,1	100	1	0,214	0,291	-0,368	0,796
		100	2	0,179	0,25	-0,32	0,679
		250	1	0,198	0,256	-0,313	0,71
		250	2	0,171	0,236	-0,302	0,643
γ_{42}	0,3	100	1	0,332	0,344	-0,356	1,021
		100	2	0,36	0,341	-0,322	1,042
		250	1	0,324	0,33	-0,336	0,984
		250	2	0,357	0,319	-0,281	0,994
γ_{43}	0,4	100	1	0,319	0,329	-0,339	0,976
		100	2	0,302	0,325	-0,347	0,951
		250	1	0,322	0,308	-0,294	0,938
		250	2	0,316	0,308	-0,3	0,932
γ_{44}	0,2	100	1	0,135	0,201	-0,266	0,537
		100	2	0,158	0,198	-0,237	0,554
		250	1	0,156	0,19	-0,225	0,536
		250	2	0,156	0,19	-0,224	0,536

Tabela 5.20: Resumo dos valores obtidos para a as probabilidades de transição γ_4 ., com $m = 4$.

5.3.5 Ordem $m = 5$

Para as médias das cinco densidades normais percebe-se, pela Tabela 5.21 que as médias das estimativas estão próximas dos valores verdadeiros e que, de modo geral, os erros padrões diminuem com o aumento de T e S .

Ao observar a Tabela 5.22, nota-se que o comportamento dos erros padrões segue o esperado, e que as médias das estimativas estão próximas dos valores verdadeiros das variâncias σ_i^2 , $i = 1, 2, 3, 4, 5$.

Já para a distribuição inicial, pela Tabela 5.23, nota-se que as estimativas não são adequadas, mesmo que os erros padrões tenham o comportamento esperado.

Agora, pelas Tabelas 5.24, 5.25, 5.26, 5.27 e 5.28, nota-se que as médias das estimativas estão relativamente próximas dos valores verdadeiros das probabilidades de transição, e os erros padrões se comportam de maneira adequada.

Médias

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
μ_1	-2	100	1	-2,029	0,434	-2,897	-1,162
		100	2	-1,977	0,355	-2,687	-1,266
		250	1	-2,017	0,372	-2,76	-1,274
		250	2	-2,013	0,34	-2,694	-1,332
μ_2	-1	100	1	-0,894	0,616	-2,126	0,337
		100	2	-0,872	0,565	-2,001	0,257
		250	1	-0,875	0,585	-2,045	0,295
		250	2	-0,798	0,669	-2,137	0,54
μ_3	0	100	1	0,204	0,555	-0,906	1,314
		100	2	0,263	0,567	-0,872	1,397
		250	1	0,214	0,535	-0,855	1,284
		250	2	0,299	0,548	-0,797	1,396
μ_4	1	100	1	1,117	0,406	0,306	1,929
		100	2	1,151	0,413	0,326	1,976
		250	1	1,13	0,391	0,348	1,912
		250	2	1,154	0,42	0,314	1,995
μ_5	2	100	1	2,167	0,602	0,963	3,371
		100	2	2,146	0,595	0,955	3,336
		250	1	2,188	0,643	0,903	3,473
		250	2	2,269	0,646	0,977	3,562

Tabela 5.21: Resumo dos valores obtidos para as médias com $m = 5$.

Variâncias

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
σ_1^2	0,5	100	1	0,413	0,355	-0,297	1,123
		100	2	0,458	0,38	-0,301	1,217
		250	1	0,452	0,28	-0,108	1,012
		250	2	0,477	0,327	-0,177	1,131
σ_2^2	1	100	1	0,73	0,81	-0,889	2,35
		100	2	0,711	0,558	-0,405	1,827
		250	1	0,761	0,791	-0,822	2,344
		250	2	0,668	0,565	-0,461	1,798
σ_3^2	1	100	1	0,668	0,888	-1,108	2,443
		100	2	0,822	0,926	-1,029	2,674
		250	1	0,786	0,735	-0,684	2,257
		250	2	0,958	0,822	-0,687	2,603
σ_4^2	0,5	100	1	0,72	0,755	-0,79	2,229
		100	2	0,786	0,834	-0,881	2,453
		250	1	0,806	0,779	-0,752	2,364
		250	2	0,722	0,624	-0,526	1,969
σ_5^2	1	100	1	0,634	0,598	-0,562	1,83
		100	2	0,741	0,623	-0,505	1,987
		250	1	0,644	0,559	-0,474	1,761
		250	2	0,719	0,452	-0,185	1,624

Tabela 5.22: Resumo dos valores obtidos para a variâncias com $m = 5$.

Distribuição Inicial (δ)

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
δ_1	0,2	100	1	0,19	0,393	-0,597	0,977
		100	2	0,21	0,307	-0,405	0,824
		250	1	0,195	0,397	-0,599	0,989
		250	2	0,138	0,259	-0,38	0,656
δ_2	0,1	100	1	0,19	0,393	-0,597	0,977
		100	2	0,15	0,3	-0,449	0,75
		250	1	0,195	0,397	-0,599	0,989
		250	2	0,158	0,297	-0,435	0,752
δ_3	0,1	100	1	0,185	0,389	-0,594	0,964
		100	2	0,159	0,315	-0,471	0,789
		250	1	0,2	0,401	-0,602	1,002
		250	2	0,211	0,347	-0,483	0,904
δ_4	0,3	100	1	0,225	0,419	-0,612	1,062
		100	2	0,226	0,358	-0,489	0,942
		250	1	0,17	0,377	-0,583	0,923
		250	2	0,217	0,354	-0,492	0,925
δ_5	0,3	100	1	0,21	0,408	-0,607	1,027
		100	2	0,255	0,369	-0,482	0,992
		250	1	0,24	0,428	-0,616	1,096
		250	2	0,276	0,363	-0,45	1,003

Tabela 5.23: Resumo dos valores obtidos para a distribuição inicial com $m = 5$.

Probabilidades de Transição (γ_1 .)

parâmetro	valor verdadeiro			estimativas		limites	
				T	S	média	erro padrão
γ_{11}	0,3	100	1	0,238	0,216	-0,194	0,67
		100	2	0,255	0,207	-0,159	0,669
		250	1	0,243	0,186	-0,13	0,615
		250	2	0,281	0,193	-0,105	0,667
γ_{12}	0,2	100	1	0,236	0,273	-0,31	0,781
		100	2	0,23	0,259	-0,287	0,748
		250	1	0,256	0,268	-0,281	0,793
		250	2	0,221	0,229	-0,237	0,678
γ_{13}	0,2	100	1	0,196	0,264	-0,331	0,724
		100	2	0,165	0,22	-0,275	0,605
		250	1	0,184	0,23	-0,276	0,643
		250	2	0,175	0,231	-0,286	0,636
γ_{14}	0,1	100	1	0,163	0,218	-0,273	0,6
		100	2	0,159	0,233	-0,306	0,624
		250	1	0,16	0,214	-0,267	0,588
		250	2	0,166	0,217	-0,269	0,6
γ_{15}	0,2	100	1	0,166	0,208	-0,25	0,583
		100	2	0,191	0,22	-0,249	0,631
		250	1	0,157	0,183	-0,208	0,523
		250	2	0,158	0,181	-0,204	0,519

Tabela 5.24: Resumo dos valores obtidos para as probabilidades de transição γ_1 , com $m = 5$.

Probabilidades de Transição (γ_2 .)

parâmetro	valor ver- dadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{21}	0,1	100	1	0,194	0,248	-0,302	0,69
		100	2	0,213	0,256	-0,299	0,725
		250	1	0,212	0,224	-0,237	0,66
		250	2	0,193	0,211	-0,23	0,615
γ_{22}	0,3	100	1	0,186	0,25	-0,314	0,686
		100	2	0,17	0,22	-0,269	0,61
		250	1	0,144	0,191	-0,238	0,525
		250	2	0,168	0,205	-0,242	0,578
γ_{23}	0,2	100	1	0,212	0,291	-0,37	0,794
		100	2	0,208	0,273	-0,338	0,755
		250	1	0,208	0,249	-0,29	0,707
		250	2	0,196	0,241	-0,287	0,679
γ_{24}	0,1	100	1	0,201	0,282	-0,362	0,765
		100	2	0,197	0,272	-0,347	0,74
		250	1	0,241	0,289	-0,338	0,82
		250	2	0,248	0,277	-0,305	0,802
γ_{25}	0,3	100	1	0,206	0,263	-0,319	0,732
		100	2	0,211	0,261	-0,311	0,734
		250	1	0,196	0,234	-0,273	0,664
		250	2	0,195	0,247	-0,3	0,689

Tabela 5.25: Resumo dos valores obtidos para as probabilidades de transição γ_2 , com $m = 5$.

Probabilidades de Transição (γ_3 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{31}	0,2	100	1	0,208	0,274	-0,341	0,757
		100	2	0,154	0,214	-0,274	0,581
		250	1	0,145	0,217	-0,288	0,578
		250	2	0,172	0,211	-0,249	0,594
γ_{32}	0,1	100	1	0,192	0,274	-0,357	0,741
		100	2	0,176	0,25	-0,324	0,676
		250	1	0,176	0,262	-0,348	0,701
		250	2	0,179	0,234	-0,289	0,648
γ_{33}	0,2	100	1	0,187	0,249	-0,31	0,684
		100	2	0,241	0,287	-0,333	0,814
		250	1	0,263	0,298	-0,333	0,859
		250	2	0,212	0,258	-0,303	0,728
γ_{34}	0,3	100	1	0,256	0,312	-0,368	0,88
		100	2	0,237	0,285	-0,332	0,806
		250	1	0,224	0,286	-0,348	0,796
		250	2	0,231	0,268	-0,305	0,767
γ_{35}	0,2	100	1	0,157	0,237	-0,317	0,631
		100	2	0,192	0,252	-0,312	0,696
		250	1	0,192	0,246	-0,299	0,684
		250	2	0,205	0,242	-0,278	0,688

Tabela 5.26: Resumo dos valores obtidos para a as probabilidades de transição γ_3 ., com $m = 5$.

Probabilidades de Transição (γ_4 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{41}	0,1	100	1	0,179	0,263	-0,346	0,704
		100	2	0,16	0,216	-0,271	0,591
		250	1	0,173	0,213	-0,253	0,6
		250	2	0,18	0,199	-0,217	0,577
γ_{42}	0,1	100	1	0,2	0,272	-0,345	0,745
		100	2	0,223	0,264	-0,304	0,75
		250	1	0,205	0,263	-0,321	0,732
		250	2	0,21	0,248	-0,287	0,706
γ_{43}	0,4	100	1	0,198	0,27	-0,342	0,738
		100	2	0,216	0,27	-0,324	0,756
		250	1	0,228	0,28	-0,331	0,788
		250	2	0,203	0,26	-0,316	0,722
γ_{44}	0,2	100	1	0,201	0,259	-0,318	0,719
		100	2	0,223	0,248	-0,272	0,718
		250	1	0,225	0,255	-0,284	0,735
		250	2	0,241	0,253	-0,264	0,747
γ_{45}	0,2	100	1	0,222	0,299	-0,375	0,82
		100	2	0,178	0,254	-0,33	0,685
		250	1	0,196	0,234	-0,273	0,664
		250	2	0,166	0,221	-0,276	0,607

Tabela 5.27: Resumo dos valores obtidos para a as probabilidades de transição γ_4 , com $m = 5$.

Probabilidades de Transição (γ_5 .)

parâmetro	valor verdadeiro			estimativas		limites	
		T	S	média	erro padrão	inferior	superior
γ_{51}	0,2	100	1	0,199	0,252	-0,305	0,704
		100	2	0,206	0,241	-0,276	0,688
		250	1	0,228	0,235	-0,242	0,698
		250	2	0,199	0,209	-0,219	0,617
γ_{52}	0,2	100	1	0,216	0,277	-0,338	0,771
		100	2	0,237	0,287	-0,338	0,811
		250	1	0,22	0,256	-0,292	0,733
		250	2	0,196	0,233	-0,271	0,663
γ_{53}	0,2	100	1	0,216	0,292	-0,368	0,799
		100	2	0,219	0,284	-0,349	0,787
		250	1	0,206	0,255	-0,305	0,716
		250	2	0,27	0,289	-0,309	0,849
γ_{54}	0,3	100	1	0,25	0,296	-0,342	0,842
		100	2	0,248	0,27	-0,292	0,788
		250	1	0,235	0,26	-0,285	0,754
		250	2	0,231	0,248	-0,265	0,726
γ_{55}	0,1	100	1	0,119	0,195	-0,271	0,508
		100	2	0,091	0,165	-0,24	0,421
		250	1	0,111	0,166	-0,221	0,442
		250	2	0,104	0,181	-0,259	0,466

Tabela 5.28: Resumo dos valores obtidos para as probabilidades de transição γ_5 , com $m = 5$.

5.3.6 Comentários gerais

De maneira geral percebe-se para todas as ordens, $m = 1, 2, 3, 4, 5$, que aumentar o tamanho das sequências (T) e o número de sequências (S) no processo de estimação, ou seja, aumentar o número de observações, melhora a precisão dos estimadores. Pode-se notar também que as distribuições iniciais são os parâmetros com os piores resultados. Mas isso faz sentido, pois, como dito antes, apenas 1 ou 2 ($S = 1$ ou $S = 2$) observações são utilizadas na estimação.

Outro ponto que deve ser levado em consideração é o número de parâmetros a serem estimados: para $m = 1$, são apenas 2 parâmetros; para $m = 2$, já são 7 parâmetros; para $m = 3$, são 14 parâmetros; para $m = 4$, são 23 parâmetros; e para $m = 5$, são 34 parâmetros. E o número crescente de parâmetros a serem estimados influencia de maneira negativa a qualidade da estimação, uma vez que o número de observações utilizadas não aumenta na mesma proporção.

Note que o processo de estimação implementado inicializa o algoritmo EM com 50 conjuntos de pontos iniciais diferentes, e que isso também pode influenciar a qualidade das estimativas obtidas. É natural pensar que talvez o aumento desse número possa melhorar o processo. Contudo, deve-se levar em conta que o aumento do número de conjuntos iniciais para o algoritmo EM faz com que o tempo de processamento também aumente.

5.4 Avaliação dos critérios AIC e BIC

As Tabelas 5.29, 5.30, 5.31 e 5.32 contêm a frequência e a porcentagem dos acertos e erros cometidos pelos critérios de seleção AIC e BIC. Para os erros, elas também indicam se o critério sugere um modelo de ordem menor (erro $<$) ou ordem maior (erro $>$). Por estas Tabelas é possível perceber que o aumento de observações, tanto pelo aumento do tamanho das sequências (T) quanto pelo aumento do número de sequências (S), melhora a porcentagem de acertos do critério AIC para todas as ordens, $m = 1, 2, 3, 4, 5$.

Já o critério BIC só piora com o aumento das observações, na verdade, nem para os casos em que $T = 100$ e $S = 1$ este critério é adequado. Perceba que para

AIC $(T = 100, S = 1)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	147	73,5%	53	26,5%	200	100%
2	139	69,5%	45	22,5%	16	8%	200	100%
3	172	86%	23	11,5%	5	2,5%	200	100%
4	184	92%	16	8%	0	0%	200	100%
5	185	92,5%	14	7%	1	0,5%	200	100%

 $(T = 100, S = 2)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	168	84%	32	16%	200	100%
2	100	50%	76	38%	24	12%	200	100%
3	145	72,5%	48	24%	7	3,5%	200	100%
4	170	85%	28	14%	2	1%	200	100%
5	182	91%	18	9%	0	0%	200	100%

Tabela 5.29: Resultado da avaliação para o critério AIC para $T = 100$ fixo, e $S = 1$ e $S = 2$.

$m = 3, 4, 5$ o critério erra para uma ordem menor 100% das vezes. De fato, como foi discutido na Seção 4.2, este critério, que leva em consideração o tamanho amostral, penaliza grandes amostras (veja expressão 4.6).

Para ambos os critérios a maior proporção dos erros está concentrada nos modelos de ordem menor que a ordem verdadeira. E, pelos gráficos das Figuras 5.15, 5.16, 5.17, 5.18 e 5.19, fica clara a superioridade do critério AIC para a seleção da ordem dos modelos, com excessão do caso $m = 1$, onde o critério BIC o superou. Essa diferença para a ordem $m = 1$ se deve pela tendência do critério BIC em indicar ordens menores, sendo assim, o critério AIC ainda é o mais indicado.

AIC

$(T = 250, S = 1)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	164	82%	36	18%	200	100%
2	74	37%	117	58,5%	9	4,5%	200	100%
3	163	81,5%	35	17,5%	2	1%	200	100%
4	177	88,5%	21	10,5%	2	1%	200	100%
5	186	93%	14	7%	0	0%	200	100%

$(T = 250, S = 2)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	167	83,5%	33	16,5%	200	100%
2	52	26%	120	60%	28	14%	200	100%
3	142	71%	50	25%	8	4%	200	100%
4	159	79,5%	38	19%	3	1,5%	200	100%
5	170	85%	23	11,5%	7	3,5%	200	100%

Tabela 5.30: Resultado da avaliação para o critério AIC para $T = 250$ fixo, e $S = 1$ e $S = 2$.

BIC $(T = 100, S = 1)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	177	88,5%	23	11,5%	200	100%
2	186	93%	14	7%	0	0%	200	100%
3	200	100%	0	0%	0	0%	200	100%
4	200	100%	0	0%	0	0%	200	100%
5	200	100%	0	0%	0	0%	200	100%

 $(T = 100, S = 2)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	200	100,0%	0	0%	200	100%
2	197	98,5%	3	1,5%	0	0%	200	100%
3	200	100%	0	0%	0	0%	200	100%
4	200	100%	0	0%	0	0%	200	100%
5	200	100%	0	0%	0	0%	200	100%

Tabela 5.31: Resultado da avaliação para o critério BIC para $T = 100$ fixo, e $S = 1$ e $S = 2$.

BIC $(T = 250, S = 1)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	196	98%	4	2%	200	100%
2	198	99%	2	1%	0	0%	200	100%
3	200	100%	0	0%	0	0%	200	100%
4	200	100%	0	0%	0	0%	200	100%
5	200	100%	0	0%	0	0%	200	100%

 $(T = 250, S = 2)$

m	erro (<)		acerto		erro (>)		total	
	freq.	%	freq.	%	freq.	%	freq.	%
1	—	—	200	100%	0	0%	200	100%
2	199	99,5%	1	0,5%	0	0%	200	100%
3	200	100%	0	0%	0	0%	200	100%
4	200	100%	0	0%	0	0%	200	100%
5	200	100%	0	0%	0	0%	200	100%

Tabela 5.32: Resultado da avaliação para o critério BIC para $T = 250$ fixo, e $S = 1$ e $S = 2$.

Feitas todas as considerações teóricas e discutidos os resultados das simulações, o capítulo seguinte apresenta as informações sobre o experimento, a coleta dados EEG, e a estimação e classificação destes.

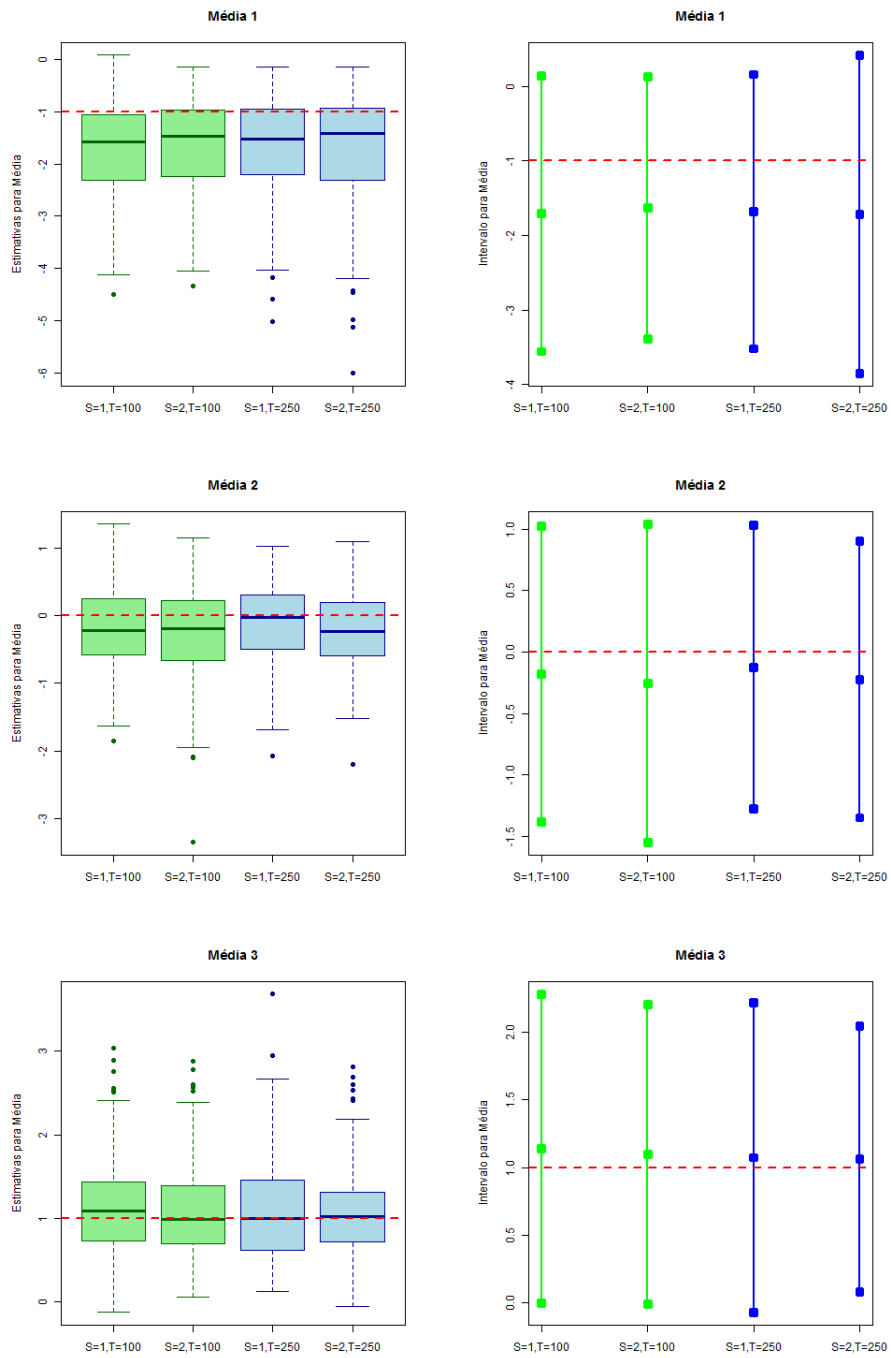


Figura 5.9: Box-plots das estimativas e intervalos para as médias μ_1 , μ_2 e μ_3 , com $m = 3$

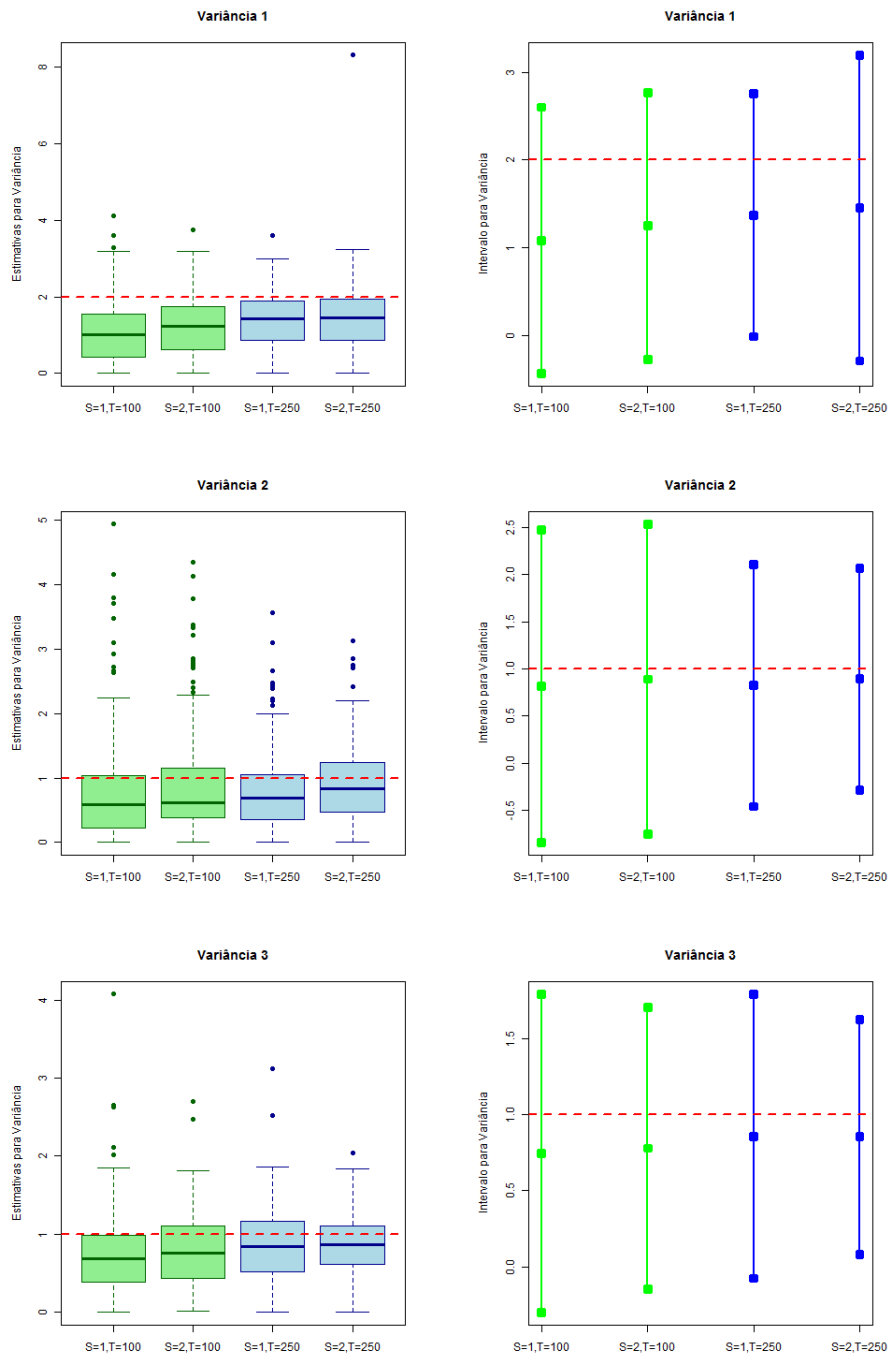


Figura 5.10: Box-plots das estimativas e intervalos para as variâncias σ_1^2 , σ_2^2 e σ_3^2 , com $m = 3$

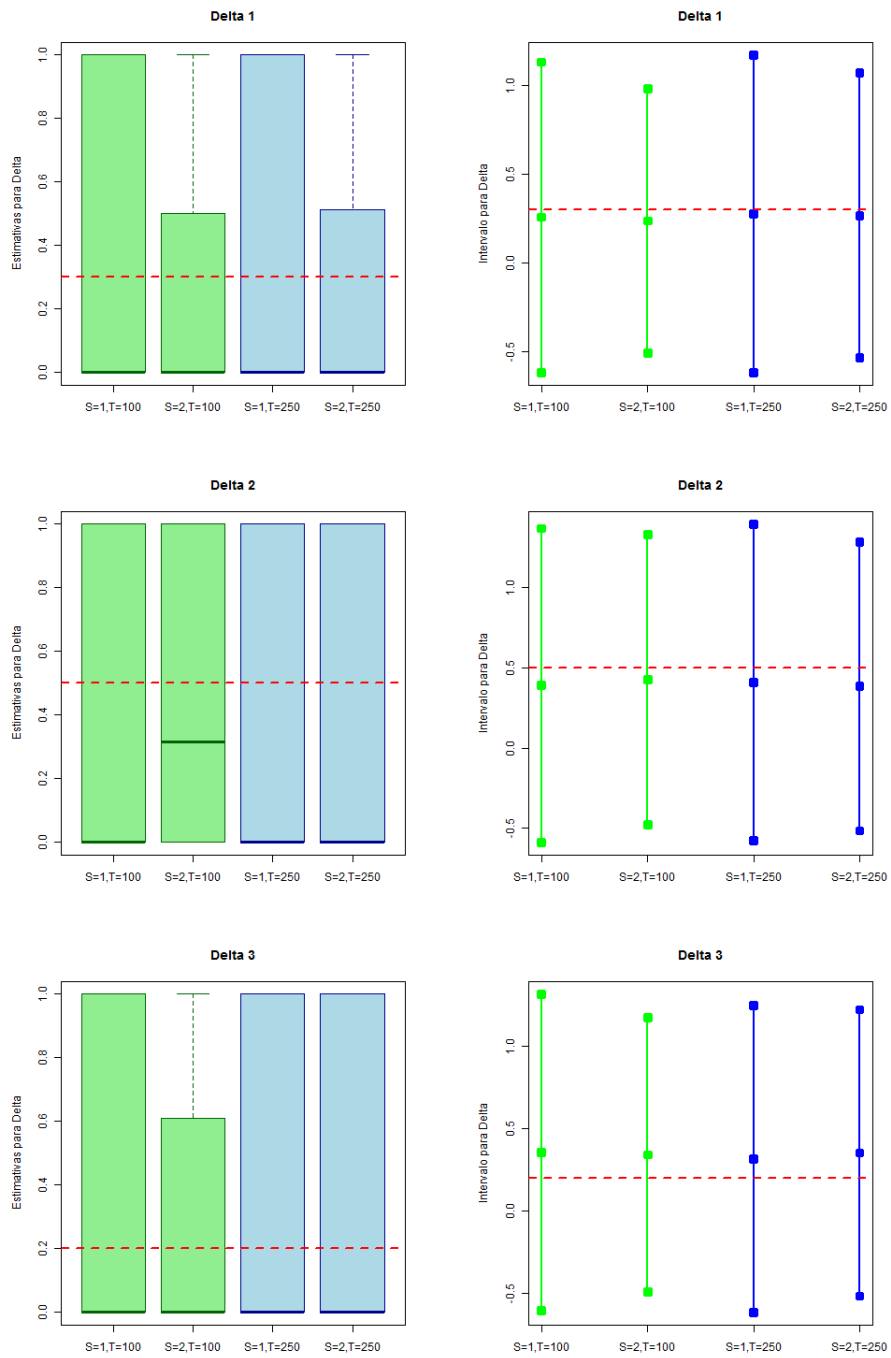


Figura 5.11: Box-plots das estimativas e intervalos para as distribuições iniciais δ_1 , δ_2 e δ_3 , com $m = 3$

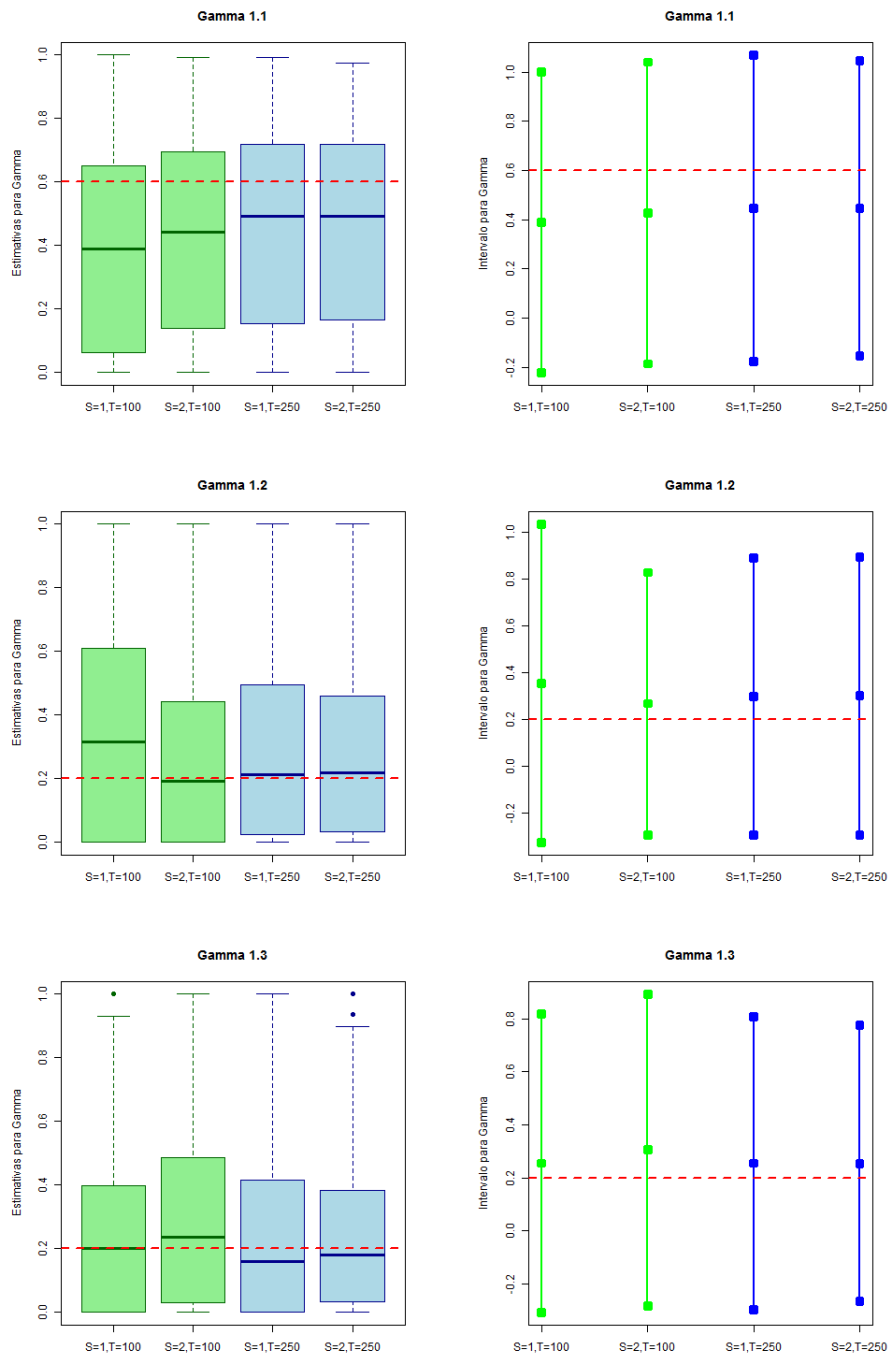


Figura 5.12: Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{1.}$, com $m = 3$

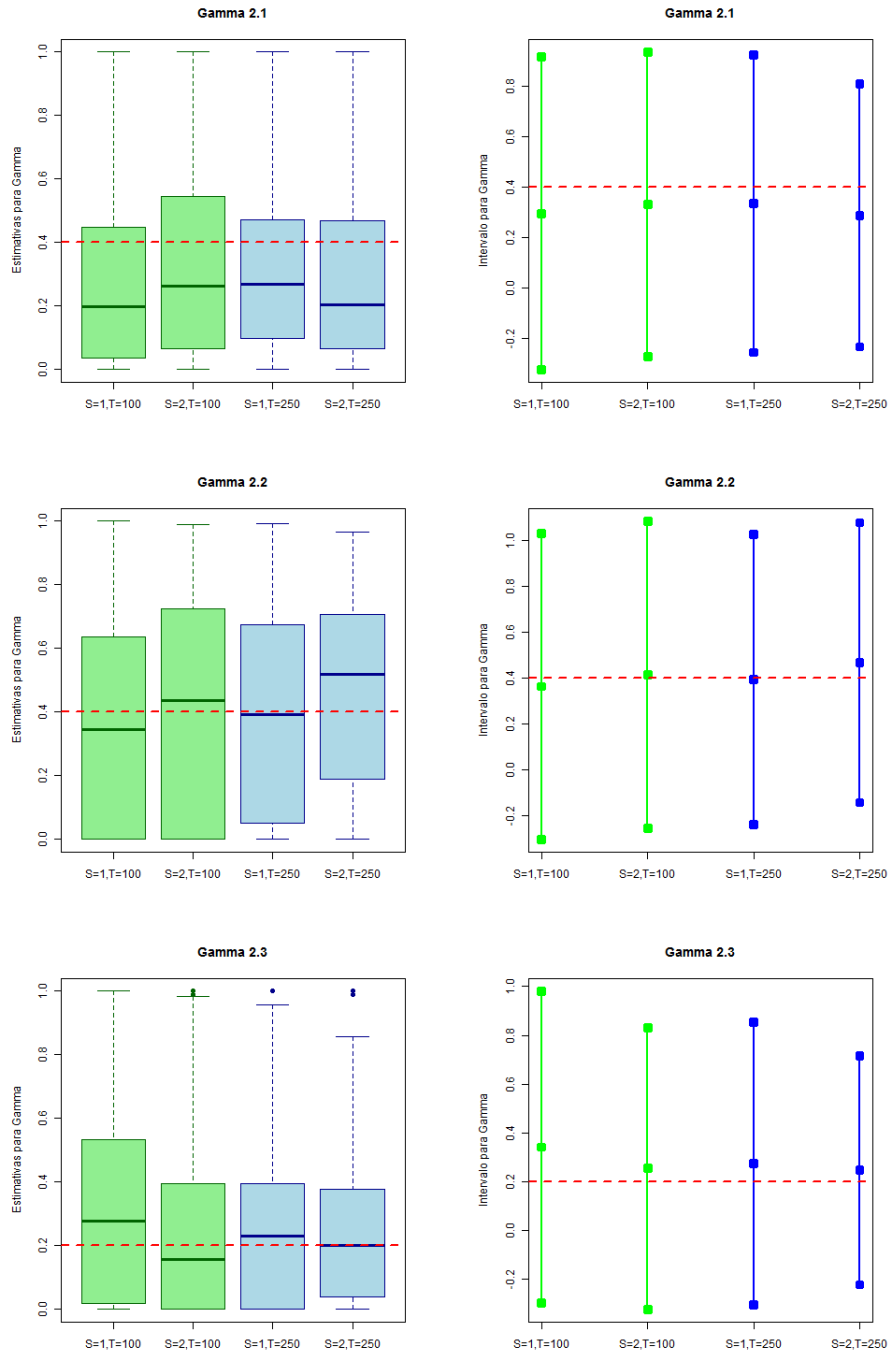


Figura 5.13: Box-plots das estimativas e intervalos para as probabilidades de transiç o $\gamma_{2,\cdot}$, com $m = 3$

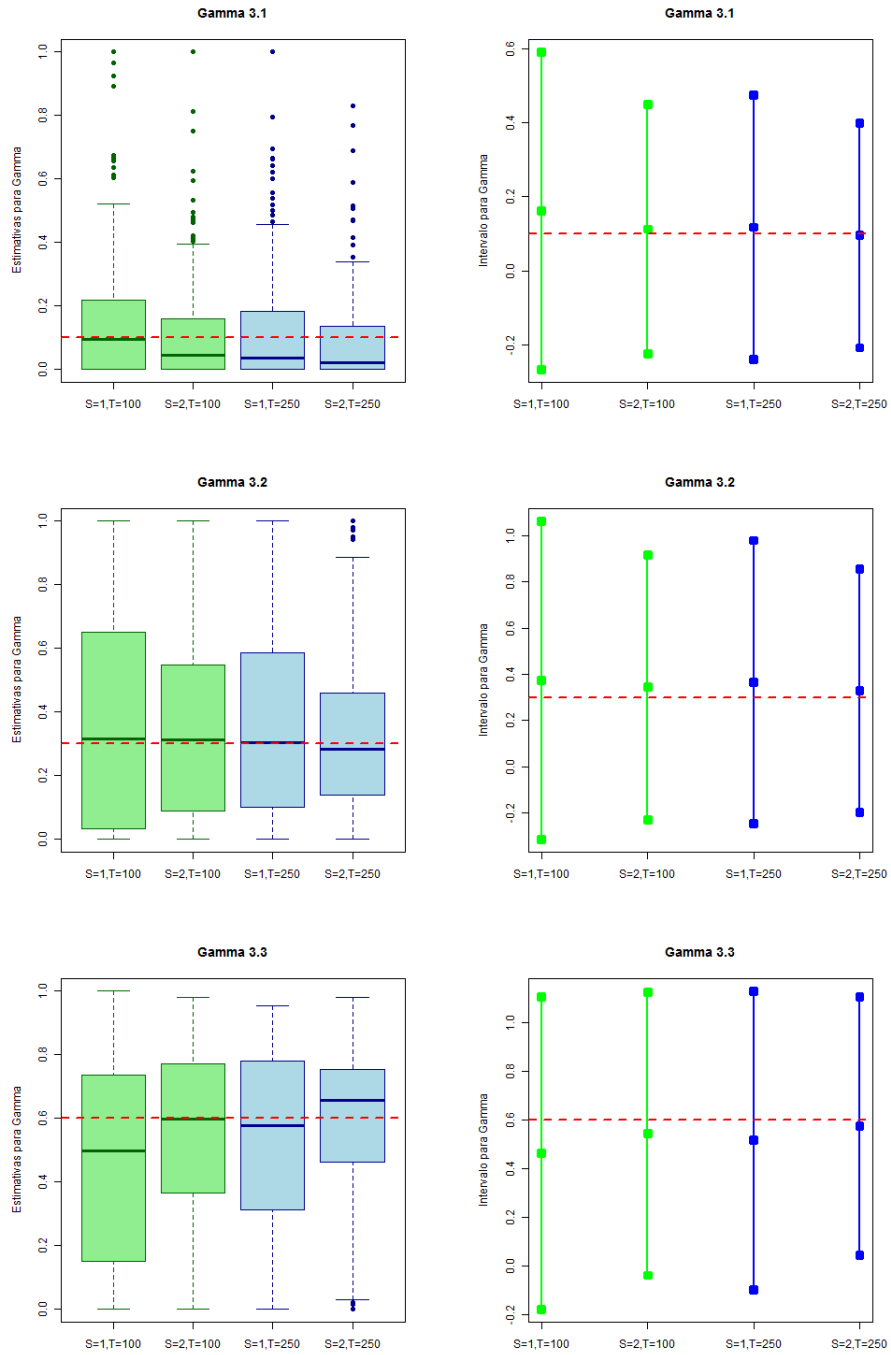


Figura 5.14: Box-plots das estimativas e intervalos para as probabilidades de transição $\gamma_{3.}$, com $m = 3$

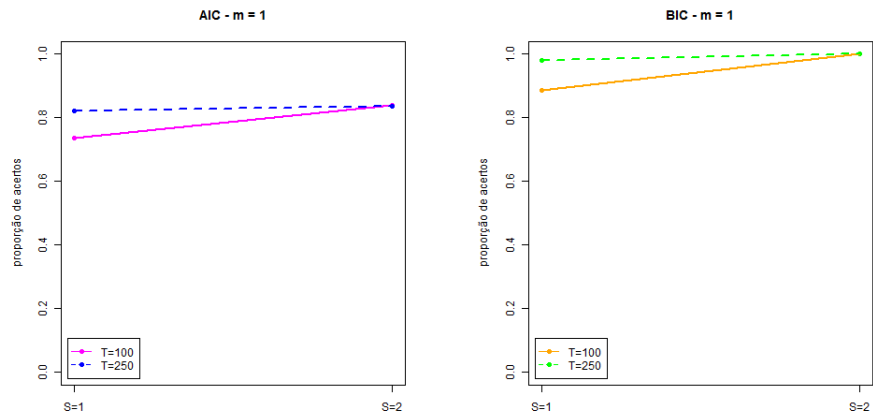


Figura 5.15: Proporção de acertos dos critérios AIC e BIC para $m = 1$

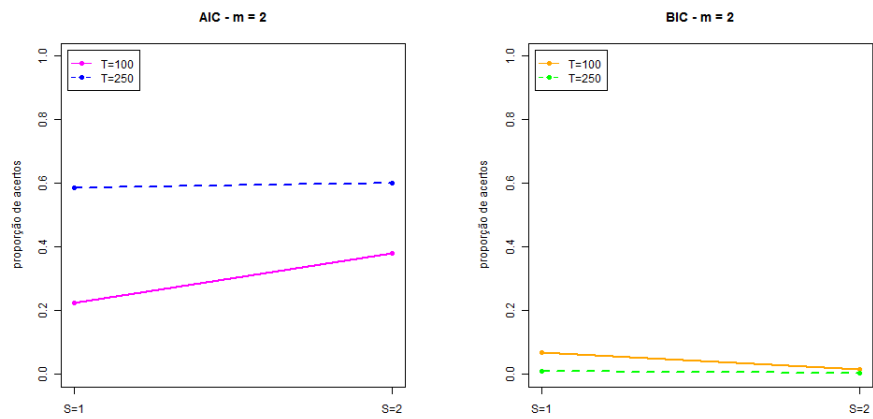


Figura 5.16: Proporção de acertos dos critérios AIC e BIC para $m = 2$

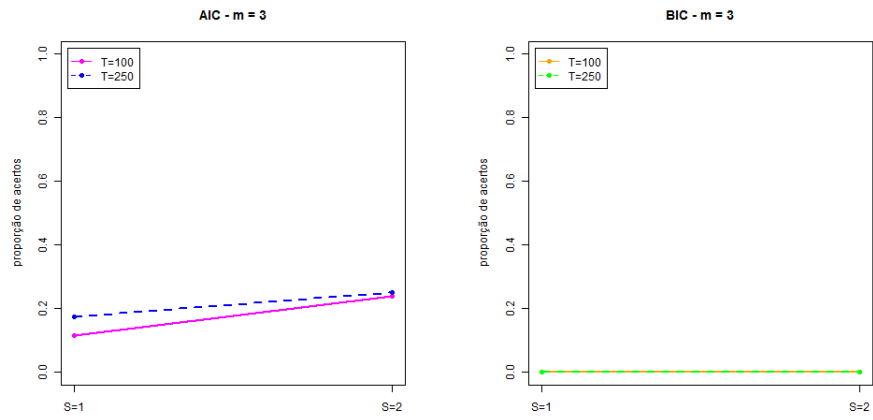


Figura 5.17: Proporção de acertos dos critérios AIC e BIC para $m = 3$

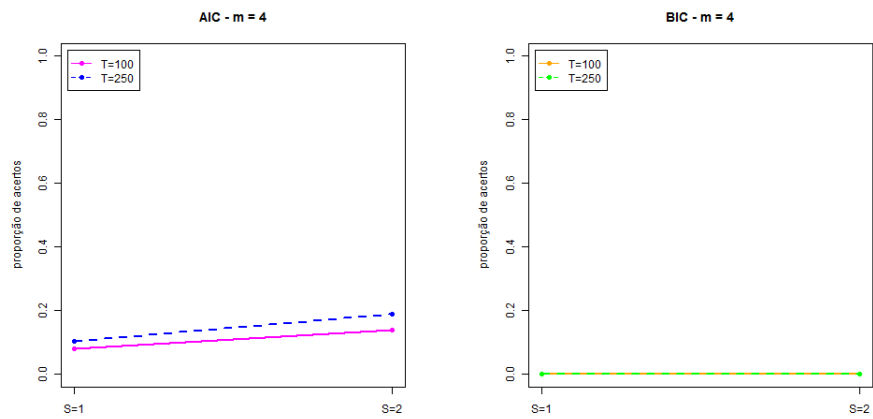


Figura 5.18: Proporção de acertos dos critérios AIC e BIC para $m = 4$

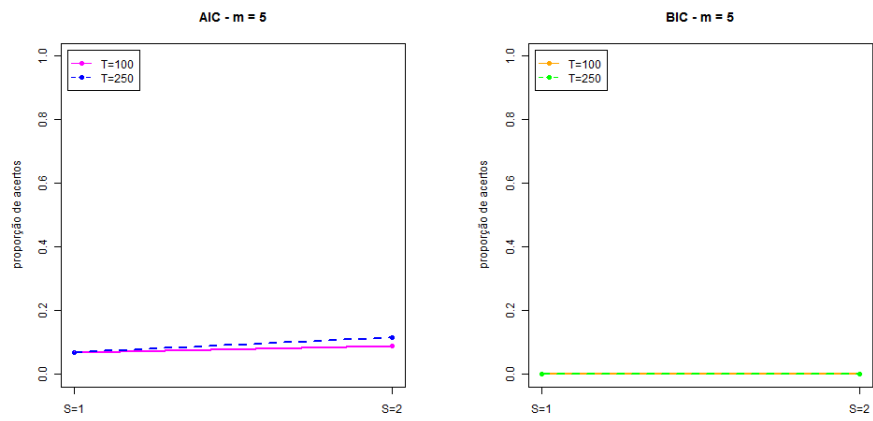


Figura 5.19: Proporção de acertos dos critérios AIC e BIC para $m = 5$

Capítulo 6

Estimação e Classificação dos Dados Reais

Neste capítulo aplicam-se os conceitos estudados e avaliados até o presente momento na base de dados reais. A seguir apresenta-se os procedimentos da coleta dos dados

6.1 Base de Dados

A base de dados desta pesquisa foi coletada no *Biopotentials Imaging Laboratory* (BIM Lab) da Universidade do Texas em El Passo, este laboratório tem se dedicado à coleta e processamento de bipotenciais. O eletroencefalograma é só um dos sinais com o qual o laboratório vem trabalhando.



Figura 6.1: Coleta dos dados EEG pelo BIM Lab - UTEP.

Para esta pesquisa foram selecionados voluntários, que deviam estar em plenas condições físicas e mentais. Qualquer problema de saúde detectado eliminaria o participante. O voluntário também estaria livre para abandonar o projeto quando desejasse. Como os dados são coletados de maneira não invasiva (veja a Figura 6.1) o procedimento não oferece riscos aos participantes.

Após responder a um questionário inicial e verificar as condições físicas do voluntário, são tomadas medidas da cabeça do participante para que os 128 eletrodos sejam posicionados nos lugares corretos. Começa então a coleta dos sinais EEG.

O voluntário, já usando a touca, irá observar 10 figuras, apresentadas na Figura 6.2, cada uma por quatro vezes, e em cada repetição as figuras aparecem por 4 segundos. A ordem em que as figuras aparecem é aleatória. Uma vez coletados os sinais os voluntários estão liberados.

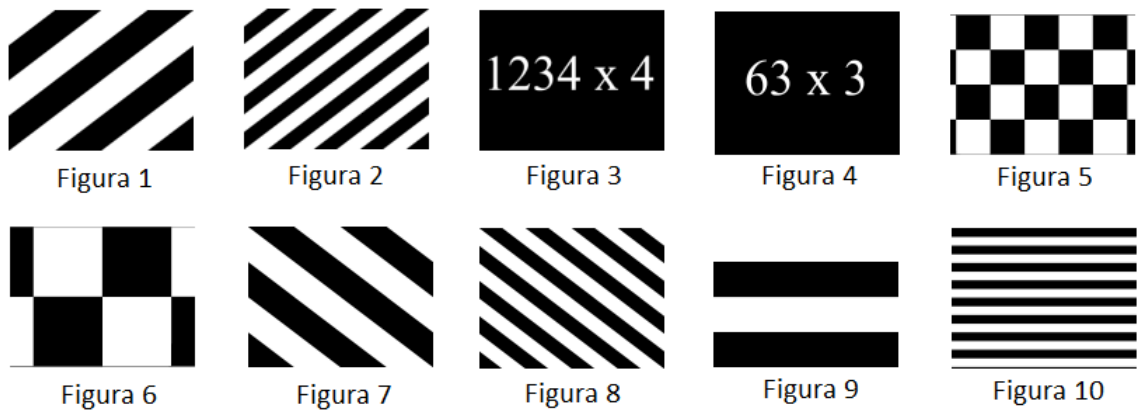


Figura 6.2: Figuras observadas durante o experimento para a coleta dos sinais EEG.

Para cada repetição das figuras é registrado o sinal elétrico por 4 segundos. Neste momento é utilizada a transformada de Fourier para determinar a frequência máxima (f_{\max}) do sinal registrado. Esta frequência encontrada determina a taxa máxima de reamostragem (T_x):

$$T_x \leq \frac{1}{2f_{\max}}.$$

A taxa de reamostragem em processamento de sinais é o intervalo de tempo usado para registrar os pontos do sinal, assim um sinal contínuo passa a ser um conjunto discreto de pontos. Por tanto, a frequência máxima obtida pela transformada

de Fourier vai determinar o número mínimo de pontos usados para representar todo o sinal sem perder informações essenciais.

Alguns trabalhos tentam reduzir o número de pontos das sequências dividindo a sequência em intervalos e utilizando os pontos médios ou medianas desses intervalos, mas esse tipo de procedimento pode eliminar as variações que serão importantes para o estudo.

Neste estudo utilizaremos os sinais EEG coletados para apenas um voluntário. As quatro repetições, sequências dos pontos registrados, das figuras serão utilizadas das seguintes maneiras:

- i) a primeira sequência de cada figura é utilizada para a estimação dos 10 modelos, e as três sequências seguintes serão utilizadas para avaliar a qualidade da classificação, tendo assim 30 sequências para a classificação; e
- ii) as duas primeiras sequências são utilizadas na estimação e as duas seguintes para avaliar a classificação, tendo 20 sequências para a classificação.

Com isso é possível apreender como e se o número de sequências na estimação altera os resultados da classificação dos sinais.

6.2 Estimação

Aqui estuda-se o comportamento de 21 dos 128 eletrodos presentes no estudo. A escolha destes 21 eletrodos foi baseada nos resultados obtidos por Silva [35] que, utilizando a técnica *Support Vector Machine*, identificou-os como sendo os mais significantes no processo de classificação dos sinais obtidos pelo estímulo visual aplicado no estudo.

A Figura 6.3 indica o posicionamento e o número do eletrodo correspondente. Estão em destaque os 21 eletrodos avaliados: 69 ao 74, 78 ao 80, 91 ao 96 e 101 ao 103. Note que estes estão localizados na parte frontal da cabeça. Deve-se também ressaltar que para este estudo o tamanho (T) das sequências estudadas é de 164 pontos, e os dados foram multiplicados por 10^6 .

Como este trabalho utiliza os dados de apenas um indivíduo, essa avaliação não permite encontrar indícios de um comportamento médio de todo o procedimento, mas apenas uma primeira idéia de seu comportamento com dados reais.

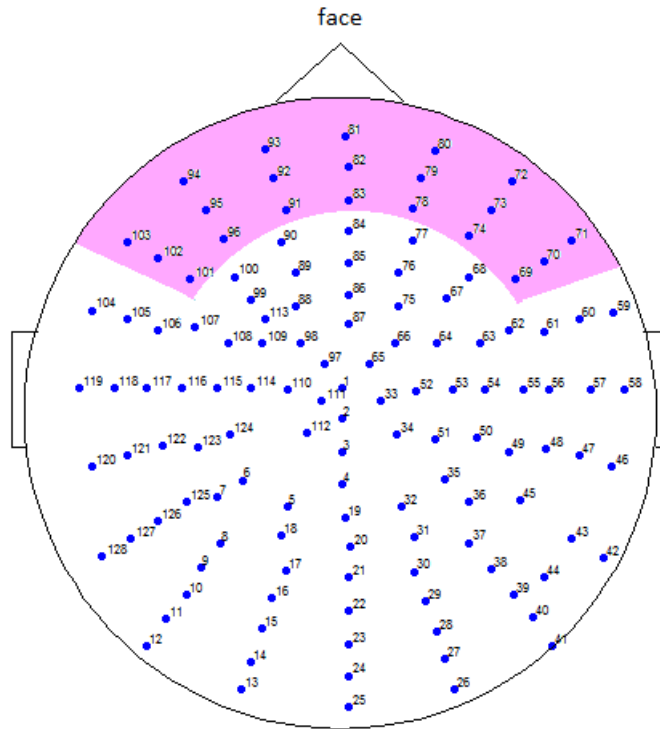


Figura 6.3: Posicionamento dos eletrodos no escalpo (visto de cima).

Para cada uma das 10 figuras, em cada um dos 21 eletrodos, é preciso encontrar o modelo que representará esta figura no momento da classificação, em duas situações diferentes:

- i) utilizando a primeira repetição para a estimação, tomando assim $S = 1$ (S é o número de sequências presentes na estimação do modelo em questão); e
- ii) utilizando as duas primeiras repetições, tomando assim $S = 2$.

Para encontrar o modelo desejado é preciso, não somente determinar as estimativas para o conjunto de parâmetros $\lambda = (\Gamma, \delta, \mu, \sigma^2)$, mas também encontrar a ordem m do modelo. Por isso, para cada figura foram estimados 6 modelos diferentes ($m = \{1, 2, 3, 4, 5, 6\}$). E, observando os valores obtidos no critério AIC (veja as Seções 4.2 e A.13), determina-se a ordem adequada do modelo.

Para exemplificar este processo envolvido na estimação dos modelos de cada figura observe o gráfico da Figura 6.4. Este mostra os valores obtidos nos critérios AIC e BIC para a figura 10 (veja Figura 6.2) no eletrodo 96. Note que para a estimação com uma sequência, $S = 1$, a ordem indicada pelo critério AIC foi $m = 4$, o mesmo

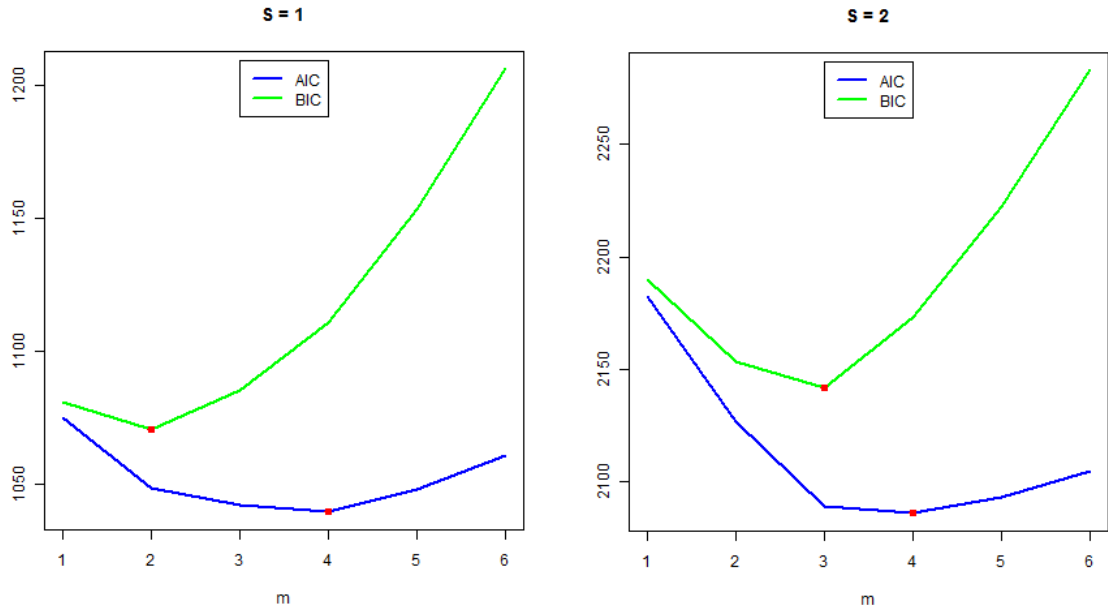


Figura 6.4: Valores dos critérios AIC e BIC para a estimação do modelo da figura 10 no eletrodo 96, com $S = 1$ e $S = 2$. Os pontos em vermelho indicam o valor mínimo obtido em cada critério.

ocorre para a estimação com duas sequências, $S = 2$. Observe também que o critério BIC indicou, nos dois casos, ordens menores para os modelos, $m = 2$ e $m = 3$, respectivamente.

A seguir, nas Tabelas 6.1, 6.2 e 6.3, estão relacionadas as ordens, obtidas pelo critério AIC, dos modelos de cada figura em cada eletrodo, para as configurações $S = 1$ e $S = 2$.

É possível resumir as informações contidas nas Tabelas 6.1, 6.2 e 6.3, como na Tabela 6.4. Por ela nota-se que, para ambos os casos $S = 1$ e $S = 2$, a ordem dos modelos varia de 2 a 5, mas no primeiro caso a maioria deles (47,6%) possui ordem 4, já no segundo caso a maioria (49,5%) possui ordem 5. E, dos 210 casos (10 figuras \times 21 eletrodos), 59 deles o critério AIC indicou a mesma ordem para as estimacões com uma ($S = 1$) ou duas sequências ($S = 2$).

Uma vez que definem-se as ordens dos modelos deve-se usar as estimativas dos seus parâmetros para obter os modelos que farão, posteriormente, parte do processo de classificação. Para exemplificar, apresentam-se a seguir os resultados obtidos das

	<u>Eletrodo 69</u>		<u>Eletrodo 70</u>		<u>Eletrodo 71</u>		<u>Eletrodo 72</u>	
Figura	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$
1	4	5	4	5	3	5	3	5
2	3	5	4	4	3	4	4	5
3	4	4	4	5	4	4	4	5
4	2	4	2	4	2	4	4	5
5	4	4	3	4	4	5	5	5
6	2	3	4	5	3	5	4	4
7	4	4	4	5	3	4	5	5
8	4	3	4	4	3	4	3	3
9	3	5	2	5	3	5	3	4
10	4	4	4	5	4	4	5	5

	<u>Eletrodo 73</u>		<u>Eletrodo 74</u>		<u>Eletrodo 78</u>		<u>Eletrodo 79</u>	
Figura	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$
1	2	3	2	4	2	3	3	3
2	5	5	3	4	4	5	4	4
3	4	5	4	4	4	4	4	5
4	5	4	4	5	4	4	4	5
5	5	5	4	5	4	5	4	5
6	3	5	4	4	4	4	4	4
7	4	5	4	4	4	5	4	5
8	3	4	4	5	2	3	3	5
9	3	4	3	3	3	5	3	4
10	4	5	4	4	4	4	3	4

Tabela 6.1: Ordens dos modelos, obtidas pelo critério AIC, para os eletrodos 69 ao 74 e 78 ao 79.

	<u>Eletrodo 80</u>		<u>Eletrodo 81</u>		<u>Eletrodo 82</u>		<u>Eletrodo 83</u>	
Figura	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$
1	3	5	4	4	3	5	4	3
2	5	5	4	5	4	5	4	5
3	5	5	5	5	4	4	3	4
4	4	5	3	5	3	5	3	5
5	5	5	5	5	4	4	4	5
6	5	5	4	5	5	5	4	5
7	4	5	4	5	4	5	4	5
8	4	5	3	4	4	3	3	3
9	3	5	3	4	3	5	3	2
10	4	5	5	5	3	4	4	3

	<u>Eletrodo 91</u>		<u>Eletrodo 92</u>		<u>Eletrodo 93</u>		<u>Eletrodo 94</u>	
Figura	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$
1	2	3	3	4	4	4	3	3
2	4	5	3	4	4	5	5	5
3	5	5	5	5	5	5	5	5
4	4	4	4	5	5	5	4	4
5	5	5	4	5	4	5	4	5
6	5	4	4	4	4	5	3	5
7	5	5	4	5	4	5	3	5
8	4	3	4	3	3	3	2	3
9	2	3	2	3	3	4	3	3
10	3	4	3	4	5	5	5	4

Tabela 6.2: Ordens dos modelos, obtidas pelo critério AIC, para os eletrodos 80 ao 73 e 91 ao 94.

	<u>Eletrodo 95</u>		<u>Eletrodo 96</u>		<u>Eletrodo 101</u>	
Figura	$S = 1$	$S = 2$	$S = 1$	$S = 2$	$S = 1$	$S = 2$
1	4	5	3	3	3	3
2	4	5	4	3	4	5
3	5	5	3	5	4	5
4	4	4	4	4	3	4
5	5	5	4	5	4	5
6	3	5	3	3	3	3
7	4	4	5	5	2	4
8	4	3	3	3	2	3
9	3	4	3	3	3	3
10	4	4	4	4	5	4

	<u>Eletrodo 102</u>		<u>Eletrodo 103</u>	
Figura	$S = 1$	$S = 2$	$S = 1$	$S = 2$
1	4	5	4	5
2	4	5	4	5
3	4	4	4	4
4	3	4	3	4
5	4	5	4	5
6	3	3	3	3
7	3	4	3	5
8	3	4	3	4
9	3	4	3	3
10	4	4	4	4

Tabela 6.3: Ordens dos modelos, obtidas pelo critério AIC, para os eletrodos 95, 96 e 101 ao 103.

$S = 1$							
Figura	m						Total
	1	2	3	4	5	6	
1	0	4 (19,0%)	9 (42,9%)	8 (38,1%)	0	0	21 (100%)
2	0	0	4 (19,0%)	14 (66,7%)	3 (14,3%)	0	21 (100%)
3	0	0	2 (9,5%)	12 (57,1%)	7 (33,3%)	0	21 (100%)
4	0	3 (14,3%)	6 (28,6%)	10 (47,6%)	2 (9,5%)	0	21 (100%)
5	0	0	1 (4,8%)	14 (66,7%)	6 (28,6%)	0	21 (100%)
6	0	1 (4,8%)	8 (38,1%)	9 (42,9%)	3 (14,3%)	0	21 (100%)
7	0	1 (4,8%)	4 (19,0%)	13 (61,9%)	3 (14,3%)	0	21 (100%)
8	0	3 (14,3%)	10 (47,6%)	8 (38,1%)	0	0	21 (100%)
9	0	3 (14,3%)	18 (85,7%)	0	0	0	21 (100%)
10	0	0	4 (19,0%)	12 (57,1%)	5 (23,8%)	0	21 (100%)
Total	0	15 (7,1%)	66 (31,4%)	100 (47,6%)	29 (13,8%)	0	210 (100%)

$S = 2$							
Figura	m						Total
	1	2	3	4	5	6	
1	0	0	8 (38,1%)	4 (19,0%)	9 (42,9%)	0	21 (100%)
2	0	0	1 (4,8%)	5 (23,8%)	15 (71,4%)	0	21 (100%)
3	0	0	0	8 (38,1%)	13 (61,9%)	0	21 (100%)
4	0	0	0	12 (57,1%)	9 (42,9%)	0	21 (100%)
5	0	0	0	3 (14,3%)	18 (85,7%)	0	21 (100%)
6	0	0	5 (23,8%)	6 (28,6%)	10 (47,6%)	0	21 (100%)
7	0	0	0	6 (28,6%)	15 (71,4%)	0	21 (100%)
8	0	0	12 (57,1%)	6 (28,6%)	3 (14,3%)	0	21 (100%)
9	0	1 (4,8%)	7 (33,3%)	7 (33,3%)	6 (28,6%)	0	21 (100%)
10	0	0	1 (4,8%)	14 (66,7%)	6 (28,6%)	0	21 (100%)
Total	0	1 (0%)	34 (16,2%)	71 (33,8%)	104 (49,5%)	0	210 (100%)

Tabela 6.4: Frequência das ordens dos modelos em cada figura, separadas por $S = 1$ e $S = 2$.

estimativas dos modelos para a figura 6 no eletrodo 69¹.

6.2.1 Estimativas dos parâmetros: Figura 6 - Eletrodo 69

Aqui apresenta-se as estimativas dos parâmetros para a figura 6 do estudo, no eletrodo 69, quando a estimação é baseada em uma sequência ($S = 1$):

i) ordem do modelo: $m = 2$

ii) vetor de distribuição inicial e matriz de transição da CM associada ao modelo:

$$\hat{\delta} = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{e} \quad \hat{\Gamma} = \begin{pmatrix} 0.9093 & 0.0907 \\ 0.0584 & 0.9416 \end{pmatrix}$$

ii) parâmetros $(\mu, \sigma^2)_i$, $i = 1, \dots, m$, das densidades normais:

$$(\hat{\mu}, \hat{\sigma}^2)_1 = (-7.0133, 22.1058) \quad \text{e} \quad (\hat{\mu}, \hat{\sigma}^2)_2 = (1.9819, 31.1458).$$

Para o caso em que a estimação foi baseada em duas sequências ($S = 2$), os valores obtidas para as estimativas dos parâmetros foram;

i) ordem do modelo: $m = 3$

ii) vetor de distribuição inicial e matriz de transição da CM associada ao modelo:

$$\hat{\delta} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \quad \text{e} \quad \hat{\Gamma} = \begin{pmatrix} 0.4813 & 0.2601 & 0.2586 \\ 0.4278 & 0.5722 & 0 \\ 0.0597 & 0.0363 & 0.9040 \end{pmatrix}$$

ii) parâmetros $(\mu, \sigma^2)_i$, $i = 1, \dots, m$, das densidades normais:

$$(\hat{\mu}, \hat{\sigma}^2)_1 = (-3.3063, 6.0482), \quad (\hat{\mu}, \hat{\sigma}^2)_2 = (-9.5448, 16.5163)$$

$$\text{e} \quad (\hat{\mu}, \hat{\sigma}^2)_3 = (2.1121, 33.0458).$$

Tendo estimado todos os 10 modelos, um para cada figura, o passo seguinte é a classificação das sequências que não foram usadas para a estimação. Com isso pode-se avaliar a qualidade da classificação baseada nos modelos HMM.

¹Os resultados para as demais figuras e eletrodos podem ser conseguidos mediante contato com a autora ou o orientador desta dissertação

6.3 Classificação

As classes para a classificação serão determinadas de 3 maneira (c.1, c.2 e c.3):

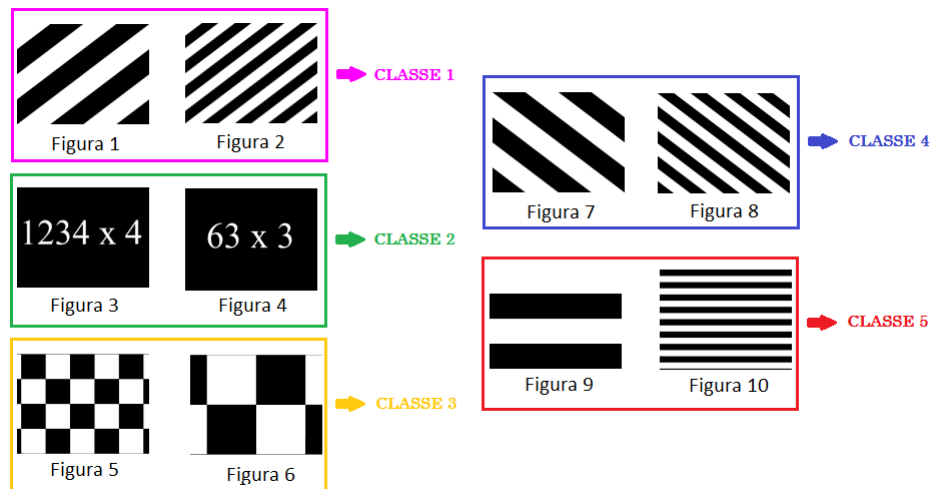


Figura 6.5: Classes da classificação c.2.

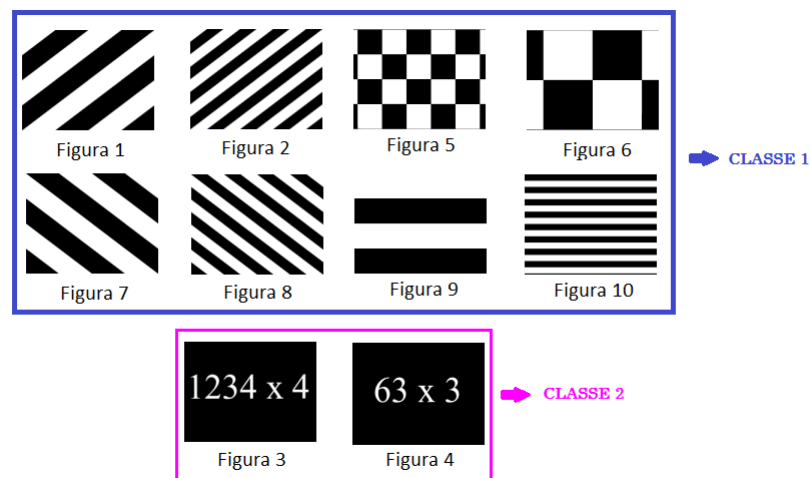


Figura 6.6: Classes da classificação c.3.

- c.1) neste caso cada classe representa uma das figuras do estudo, tendo-se assim 10 classes;
- c.2) uni-se as figuras parecidas em uma mesma classe, como mostrado na Figura 6.5, configurando-se 5 classes para a classificação;

		Classificação									
		1	2	3	4	5	6	7	8	9	10
	1	1	0	1	1	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	2	0	0
	3	0	0	0	0	0	0	1	0	0	2
Classe	4	0	1	1	0	0	0	1	0	0	0
de	5	0	1	0	1	0	0	0	1	0	0
Origem	6	0	0	2	0	0	1	0	0	0	0
	7	0	0	1	0	0	1	0	1	0	0
	8	0	0	0	0	1	0	0	0	1	1
	9	1	0	0	1	0	1	0	0	0	0
	10	0	0	1	1	0	1	0	0	0	0

Tabela 6.5: Classificação das 30 seqüências, eletrodo 70 com $S = 1$. Classes c.1.

c.3) separa-se as expressões aritméticas das outras figuras, como mostrado na Figura 6.6, gerando-se 2 classes para a classificação.

Para cada eletrodo construiu-se as matrizes de classificação nos três casos descritos em c.1, c.2 e c.3. A seguir apresenta-se como exemplo este procedimento para o eletrodo 70:

- c.1) Das Tabelas 6.5 e 6.6 pode-se notar que 2 de 30 (6,67%) e 4 de 20 (20%) seqüências foram classificadas corretamente para os casos $S = 1$ e $S = 2$, respectivamente.
- c.2) Das Tabelas 6.7 e 6.8 nota-se que para $S = 1$, 5 de 30 (16,67%) seqüências foram classificadas corretamente, já para $S = 2$ foram classificadas corretamente 5 de 20 (25%) seqüências.
- c.3) Com estas classes foram classificadas corretamente 15 de 30 (50%) seqüências, para $S = 1$, e 14 de 20 (70%) para $S = 2$. Estes resultados são observados nas Tabelas 6.9 e 6.10.

		Classificação									
		1	2	3	4	5	6	7	8	9	10
Classe de Origem	1	1	0	0	0	0	0	0	0	0	1
	2	1	0	0	1	0	0	0	0	0	0
	3	0	0	0	0	0	0	1	0	0	1
	4	0	1	0	1	0	0	0	0	0	0
	5	1	0	0	0	0	0	1	0	0	0
	6	0	0	0	0	0	1	1	0	0	0
	7	0	0	0	1	0	0	1	0	0	0
	8	0	0	0	0	1	0	0	0	0	1
	9	2	0	0	0	0	0	0	0	0	0
	10	1	0	0	0	0	1	0	0	0	0

Tabela 6.6: Classificação das 20 seqüências, eletrodo 70 com $S = 2$. Classes c.1.

		Classificação				
		1	2	3	4	5
Classe de Origem	1	2	2	0	2	0
	2	2	1	0	2	2
	3	0	3	1	1	0
	4	0	1	2	1	2
	5	1	3	2	0	0

Tabela 6.7: Classificação das 30 seqüências, eletrodo 70 com $S = 1$. Classes c.2.

		Classificação				
		1	2	3	4	5
	1	2	1	0	0	1
Classe	2	2	1	0	1	1
de	3	0	0	1	2	0
Origem	4	0	1	1	1	1
	5	3	0	1	0	0

Tabela 6.8: Classificação das 20 seqüências, eletrodo 70 com $S = 2$. Classes c.2.

		Classificação	
		1	2
Classes de	1	14	9
Origem	2	6	1

Tabela 6.9: Classificação das 30 seqüências, eletrodo 70 com $S = 1$. Classes c.3.

		Classificação	
		1	2
Classes de	1	13	2
Origem	2	4	1

Tabela 6.10: Classificação das 30 seqüências, eletrodo 70 com $S = 2$. Classes c.3.

Feito o mesmo procedimento para os 20 outros eletrodos, ou seja, classificando as sequências restantes que não foram utilizadas para a estimação dos modelos, pode-se resumir estas informações para que seja visível a comparação entre os eletrodos e os casos $S = 1$ e $S = 2$ (segundo a proporção de acertos obtidos). Na Tabela 6.11 estão em destaque as duas melhores porcentagens de classificação correta para cada combinação de valores de S e conjunto de classes.

Pela observação da Tabela 6.11 pode-se perceber os seguintes pontos:

- i) A medida que o número de classes para a classificação diminui a qualidade da classificação melhora. Note ainda que para as classes c.3, que separam as figuras das expressões aritméticas, as porcentagens de acertos variam entre 50% e 90% para $S = 1$, e entre 45% e 85% para $S = 2$.
- ii) Não parece haver relação entre a melhora ou piora da classificação com o aumento do número de sequências usadas na estimação dos modelos, pois para aproximadamente 43% e 56% dos casos o aumento do número de sequência melhora e piora a classificação, respectivamente.
- iii) Para $S = 1$, os eletrodos 94 e 79 se destacam como os melhores classificadores. Já para $S = 2$, os que se sobressaem são os eletrodos 69 e 95.

Como dito na Seção 4.3, outra maneira de classificar é utilizando conjuntos de eletrodos, para isso determina-se a classificação por meio da soma das verossimilhanças e também pela soma ponderada pela porcentagem de acertos obtidas em cada eletrodo para c.1. Em um primeiro momento os eletrodos são agrupados 3 a 3, como mostra a Figura 6.7, formando 7 novos arranjos de eletrodos: E-1 (69, 70 e 71), E-2 (72, 73 e 74), E-3 (78, 79 e 80), E-4 (81, 82 e 83), E-5 (91, 92 e 93), E-6(94, 95 e 96) e E-7(101, 102 e 103). Posteriormente, estes são agrupados pelos seus posicionamentos (esquerda, direita e centro), como na Figura 6.7. E por último a classificação é feita para a soma de todos os eletrodos.

A Tabela 6.12 indica que a utilização das somas para a classificação melhora pouco a qualidade comparando com os eletrodos separadamente, e ponderar pela porcentagem de acerto parece não alterar a qualidade. Nota-se também que o conjunto E-6, formado pelos eletrodos 94, 95 e 96, é o que melhor classifica. Entre os conjun-

Porcentagem de Classificação Correta						
Eletrodo	$S = 1$			$S = 2$		
	c.1	c.2	c.3	c.1	c.2	c.3
69	10%	20%	63.33%	20%	45%	80%
70	6.67%	16.67%	50%	20%	25%	70%
71	16.67%	30%	83.33%	10%	15%	85%
72	13.33%	16.67%	73.33%	15%	20%	70%
73	10%	20%	80%	15%	25%	50%
74	13.33%	16.67%	63.33%	10%	25%	75%
78	30%	30%	70%	5%	25%	75%
79	20%	33.33%	80%	5%	25%	70%
80	6.67%	13.33%	70%	5%	20%	75%
81	16.67%	26.67%	76.67%	10%	25%	55%
82	10%	20%	56.67%	15%	25%	70%
83	20%	26.67%	73.33%	5%	20%	70%
91	16.67%	20%	70%	10%	10%	70%
92	10%	30%	50%	5%	25%	70%
93	6.67%	20%	70%	10%	15%	45%
94	20%	33.33%	90%	15%	30%	60%
95	16.67%	36.67%	80%	15%	55%	85%
96	20%	26.67%	76.67%	15%	20%	75%
101	10%	23.33%	76.67%	15%	20%	80%
102	13.33%	23.33%	76.67%	10%	15%	65%
103	6.67%	13.33%	50%	10%	10%	55%

Tabela 6.11: Porcentagem de classificação correta dos eletrodos para as diferentes classes c.1, c.2 e c.3, nos casos $S = 1$ e $S = 2$.

Porcentagem de Classificação Correta

Soma						
Eletrodo	$S = 1$			$S = 2$		
	c.1	c.2	c.3	c.1	c.2	c.3
E-1	10%	20%	73.44%	15%	35%	95%
E-2	13.33%	29%	73.33%	15%	35%	89%
E-3	16.67%	20%	76.67%	15%	35%	80%
E-4	10%	26.67%	80%	5%	20%	70%
E-5	13.33%	20%	53.33%	10%	25%	80%
E-6	26.67%	33.33%	93.33%	20%	45%	75%
E-7	13.33%	30%	56.67%	10%	20%	70%
esquerda	16.67%	23.33%	63.33%	5%	25%	85%
direita	16.67%	26.67%	66.67%	10%	35%	75%
centro	10%	26.67%	80%	5%	20%	70%
todos	16.67%	20%	53.33%	10%	40%	80%
Soma Ponderada						
Eletrodo	$S = 1$			$S = 2$		
	c.1	c.2	c.3	c.1	c.2	c.3
E-1	10%	23.33%	80%	15%	30%	85%
E-2	13.33%	20%	73.33%	15%	30%	75%
E-3	26.67%	30%	80%	15%	35%	80%
E-4	10%	30%	73.33%	5%	10%	60%
E-5	13.33%	13.33%	53.33%	5%	20%	75%
E-6	26.67%	33.33%	93.33%	20%	45%	75%
E-7	10%	30%	53.33%	15%	30%	80%
esquerda	13.33%	23.33%	63.33%	10%	35%	75%
direita	20%	26.67%	76.67%	10%	40%	80%
centro	10%	30%	76.67%	5%	10%	60%
todos	23.33%	26.67%	60%	10%	40%	80%

Tabela 6.12: Porcentagem de classificação correta nos casos $S = 1$ e $S = 2$, utilizando as somas e as somas ponderadas das verossimilhanças.

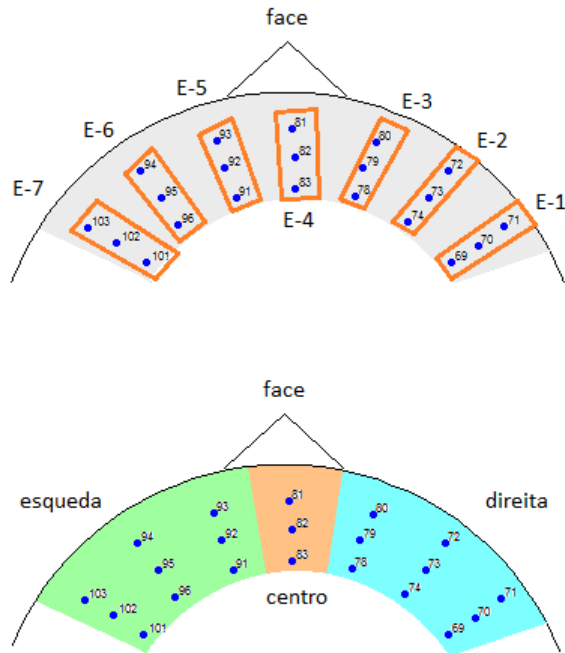


Figura 6.7: Agrupamento dos eletrodos para nova classificação das sequências.

tos de eletrodos esquerda, direita e centro, o conjunto da direita é o que se destaca em todas as condições. É possível perceber que é melhor utilizar os conjuntos de 3 eletrodos do que a soma de todos eles para classificar.

De maneira geral pode-se apreender que independentemente de se usar os eletrodos separadamente ou a soma das verossimilhanças de conjuntos destes, a qualidade da classificação que separa as 10 figuras não é muito boa. Mas quando se classifica as sequências para as classes c.2 e c.3 a qualidade tem uma boa melhora. Uma razão para isto é que de fato as 10 figuras (com exceção das expressões aritméticas) não apresentam diferenças expressivas. Isso fica ainda mais claro quando se classifica as sequências para as classes c.3, que separa as expressões aritméticas das outras, pois neste caso a qualidade da classificação é muito alta, chegando a 90% de acerto.

Capítulo 7

Conclusões e Trabalhos Futuros

Neste capítulo serão apresentadas as conclusões obtidas, projetos e intenções para trabalhos futuros.

7.1 Conclusões

Este trabalho fez um resumo sobre a teoria envolvendo modelos de Markov oculto (HMM), e um levantamento das bibliografias que apresentam a aplicação deste modelo no estudo de sinais eletroencefalogramas (EEG). Por essas bibliografias pode-se perceber a crescente utilização, não somente do modelo HMM para estudar os sinais EEG, mas do próprio interesse nesse tipo de sinal, visando sua aplicação, principalmente, na área médica.

Aqui também foi apresentado o processo de estimação para o HMM que inicializa o algoritmo EM por diferentes conjuntos de pontos iniciais aleatórios. Este processo foi baseado em Rydén [34], contudo as distribuições de probabilidade usadas neste trabalho são diferentes das sugeridas por Rydén em seu artigo.

Uma vez sugerido o processo de estimação foram realizadas simulações para avaliar a qualidade do processo e dos critérios AIC e BIC de seleção dos modelos. Essa avaliação levou em conta a variação da ordem m do modelo, o tamanho T das sequências e o número S de sequências utilizadas na estimação. De maneira geral, o processo de estimação se mostrou adequado para os parâmetros avaliados.

As simulações mostraram que há indícios de que o aumento do número de

observações, seja pelo aumento de T ou pelo aumento de S , melhora a precisão dos estimadores propostos. As simulações também indicaram possíveis problemas com o processo. O primeiro deles é o elevado número de parâmetros a serem estimados quando a ordem m do modelo aumenta. Por exemplo, para um modelo de ordem 5, 34 parâmetros devem ser estimados. O segundo problema revelado é o estimador para a distribuição inicial δ , que não se mostrou tão adequado. Isso ocorre porque o número de observações utilizadas para estimar este vetor de probabilidades é pequeno, uma e duas observações, quando $S = 1$ e $S = 2$, respectivamente.

O elevado número de parâmetros também interfere na precisão dos critérios de seleção AIC e BIC (veja as expressões 4.5 e 4.6). Entre estes critérios o AIC se mostrou mais eficiente para indicar a ordem do modelo. Este critério melhora com o aumento do número de observações.

Este trabalho também aplicou o modelo HMM estudado em um conjunto de dados reais, coletados pelo BIM Lab - UTEP. Os dados foram coletados, de maneira não invasiva, durante a observação de 10 diferentes figuras. A intenção foi estimar os modelos e usá-los na classificação dos sinais e aqui foram utilizados os dados obtidos para apenas um indivíduo. Por isso todas as conclusões obtidas não podem ser, necessariamente, extrapoladas para outros indivíduos.

Aparentemente alguns eletrodos (69, 79, 94 e 95) se mostraram mais eficazes na classificação das figuras. O mesmo ocorre com o conjunto E-6, formado pelos eletrodos 94, 95 e 96, em comparação com os outros conjuntos de 3 eletrodos. Para os eletrodos agrupados pelos seus posicionamentos, o lado direito, formado pelos eletrodos 69, 70, 71, 72, 73, 74, 78, 79 e 80, foi o que se destacou no processo de classificação.

De maneira geral, a tentativa de identificar cada 1 das 10 figuras não produz bons resultados. Estes apresentam uma melhora significativa quando os pares de figuras parecidas são classificadas conjuntamente. Já para separar as figuras abstratas das expressões matemáticas o método mostrou-se muito bom, chegando a alcançar 90% de acerto na classificação.

Os resultados obtidos nas classificações dos sinais EEG indicam a fragilidade do método para identificar e separar essas figuras escolhidas para o estudo. Uma razão para isso pode ser o próprio modelo proposto, que pode não ser o mais apropriado para a tarefa em questão, ou pelas figuras. Talvez figuras não abstratas (como pessoas,

animais, objetos,...) possam gerar resultados melhores.

Há também de se considerar a possibilidade de que todo o processo de coleta dos dados, feitos de maneira não invasiva, não seja capaz de captar as sutis diferenças entre os sinais produzidos pelo cérebro, ou ainda, o fato de que realmente não existem diferenças nos sinais produzidos quando se vê uma ou outra figura muito parecida. Um problema conhecido com sinais de EEG é sua pequena amplitude, tornando estudos de classificação naturalmente mais difíceis.

Outra observação que deve ser feita é o fato de que o estudo foi realizado apenas para um sujeito participante da pesquisa. Para outros indivíduos os resultados podem ser bem diferentes dos obtidos neste trabalho. Aqui temos apenas indicações do comportamento do modelo estimado e utilizado na classificação. Por isso, não se pode concluir que os resultados obtidos aqui se repetirão em experimentos com pessoas diferentes nas mesmas condições do experimento, ou com a mesma pessoas em experimentos diferentes.

Todos estes questionamentos só poderão ser respondidos com o avanço e implementação de novos estudos. Não se deve perder o ânimo quando os resultados obtidos não são tão bons quando os desejados, pois, para um *cientista*, até o resultado negativo é uma informação útil que pode levar a mais e melhores pesquisas. Assim, a seção seguinte apresenta idéias para trabalhos futuros relacionados.

7.2 Trabalhos futuros

Apesar de algumas questões terem sido respondidas com este trabalho outras tantas foram levantadas: umas em relação a qualidade e sensibilidade do processo de estimação, e outras sobre o experimento feito para a coleta dos dados.

Em relação a qualidade e sensibilidade do processo de estimação sugere-se alguns pontos de estudo:

- i) variar o número N de conjuntos de pontos que inicializam o algoritmo EM. Neste estudo foram utilizados 50 conjuntos de pontos iniciais. Seria interessante avaliar o comportamento das estimativas quando este número aumenta e diminui.
- ii) avaliar outros valores para o tamanho T das sequências observadas, tanto para

a estimação do modelo quanto para a classificação dos dados reais.

- iii) avaliar também outros valores para o número de sequências S utilizadas na estimação.
- iv) considerar variações do modelo HMM, como utilização de distribuições normais multivariadas para as densidades estado-dependentes ou outras distribuições como as α -estáveis [36]. Além disso a escolha de outros modelos de Markov podem ser mais indicadas para o estudo.

Para o experimento de coleta dos dados sugere-se a inclusão de figuras diferentes e até mesmo outros estímulos além dos visuais, como estímulos auditivos. Outro ponto a ser considerado é replicar todo procedimento de estimação e classificação para outros indivíduos para melhor avaliar os procedimentos propostos.

Este trabalho visou, acima de tudo, estimular o desenvolvimento de novos estudos envolvendo a estimação e classificação dos sinais EEG e outros sinais biopotenciais através do uso de Cadeias de Markov Ocultas. É evidente que o desenvolvimento de novas e melhores técnicas envolvendo estes tipos de dados podem melhorar a qualidade de vida de muitas pessoas.

Referências Bibliográficas

- [1] T. C. Araújo, *Agrupamento de Sinais de Eletroencefalografia Utilizando Mistura de Distribuições Normais.*, Trabalho de Conclusão de Curso. Departamento de Estatística - UnB, 2009.
- [2] S. Bengio and S. Chiappa, *HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems.*, In European Symposium on Artificial Neural Networks, ESANN, 2004.
- [3] U. N. Bhat, *Elements of Applied Stochastic Processes*, John Wiley & Sons, 1984.
- [4] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models.*, Springer, 2007.
- [5] G. Casella and R. L. Berger, *Statistical Inference.*, Duxbury, 2002.
- [6] A. Chambaz, A. Garivier, and E. Gassiat, *A minimum description length approach to hidden Markov models with Poisson and Gaussian emissions. Application to order identification.*, Journal of Statistical Planning and Inference **139** (2009), 962 – 977.
- [7] R. Durrett, *Elementary Probability for Applications.*, Cambridge University Press, 2009.
- [8] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. I, John Wiley & Sons, 1968.
- [9] G. A. Fink, *Markov Models for Pattern Recognition: from Theory to Applications.*, Springer, 2008.

- [10] S. Geman and C. R. Hwang, *Nonparametric maximum likelihood estimation by the method of sieves.*, *Annals of Statistics* **10** (1982), 401 – 414.
- [11] Z. Ghahramani, *An Introduction to Hidden Markov Models and Bayesian Networks.*, *International Journal of Pattern Recognition and Artificial Intelligence* **15** (2001), 9 – 42.
- [12] Z. Ghahramani and M. I. Jordan, *Factorial Hidden Markov Models*, *Machine Learning* **29** (1997), 245 – 273.
- [13] R. S. Haung, C. J. Kuo, L. Tsai, and O. T. C. Chen, *EEG Pattern Recognition - Arousal States Detection and Classification.*, *Proc. IEEE International Conference on Neural Network (ICNN)* (1996).
- [14] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, *A review of classification algorithms for EEG-based brain-computer interfaces.*, *Journal of Neural Engineering* **4** (2007), R1.
- [15] G. G. Matos, *Agrupamento de Sinais de Eletroencefalografia Utilizando Estimaco de Densidades*, Trabalho de Concluso de Curso. Departamento de Estatística - UnB, 2009.
- [16] Q. E. McCallum and S. Weston, *Parallel R*, O’Reilly Media, Inc., 2011.
- [17] C. A. McGrory and D. M. Titterington, *Variational Bayesian Analysis for Hidden Markov Models.*, *Australian & New Zealand Journal of Statistics* **51** (2009), 227 – 244.
- [18] G. McLachlan and D. Peel, *Finite Mixture Models.*, Wiley, 2000.
- [19] W. Min and G. Luo, *Medical Applications of EEG Wave Classification.*, *Chance* **4** (2009).
- [20] T. Oates, L. Firoiu, and P. R. Cohen, *Clustering Time Series with Hidden Markov Models and Dynamic Time Warping.*, In *Proceedings of the IJCAI-99 Workshop on Neural, Symbolic and Reinforcement. Learning Methods for Sequence Learning*, 1999, pp. 17 – 21.

- [21] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller, *Hidden Markov models for online classification of single trial EEG data.*, Pattern Recognition Letters **22** (2001), 1299 – 1309.
- [22] A. Panuccio, M. Bicego, and V. Murino, *A Hidden Markov Model-based approach to sequential data clustering.*, Springer, 2002, pp. 734–742.
- [23] K. Pearson, *Contributions to the theory of mathematical evolution.*, Philosophical Transactions of the Royal Society of London **A 185** (1984), 71 – 110.
- [24] G. Pfurtscheller and C. Neuper, *Motor Imagery and Direct Brain-Computer communication.*, Proceedings of the IEEE **89** (2001), 1123 – 1134.
- [25] C. E. Priebe, *Adaptive mixtures.*, Journal of the American Statistical Association **89** (1994), 796 – 806.
- [26] C. E. Priebe and D. J. Marchette, *Adaptive mixture density estimation.*, Pattern Recognition **26** (1993), 771 – 785.
- [27] L. R. Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.*, Proceedings of the IEEE **77** (1989), 257 – 286.
- [28] M. L. Rizzo, *Statistical Computing with R.*, Chapman and Hall, 2007.
- [29] L. C. S. Rocha, *Análise de Componentes Independentes em Sinais de Eletroencefalografia.*, Trabalho de Conclusão de Curso. Departamento de Estatística - UnB, 2010.
- [30] K. Roeder, *Semiparametric estimation of normal mixture densities.*, Annals of Statistics **20** (1992), 929 – 943.
- [31] K. Roeder and L. Wasserman, *Practical density estimation using mixture of normals.*, Journal of the American Statistical Association **92** (1997), 894 – 902.
- [32] S. M. Ross, *Introduction to Probability Models.*, Academic Press, 1997.
- [33] A. Rossini, L. Tierney, and N. Li, *Simple Parallel Statistical Computing in R*, UW Biostatistics Working Paper Series. (2003).

- [34] T. Rydén, *EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective.*, Bayesian Analysis **3** (2008), 659 – 688.
- [35] M. C. Silva, *Um novo método para a classificação de dados EEG: Uma aplicação de Máquinas de Suporte Vetorial e Análise de Fourier*, Trabalho de Conclusão de Curso. Departamento de Estatística - UnB, 2010.
- [36] R. C. C. Valdés, *Estimação dos Parâmetros da Mistura de Densidades Estáveis Simétricas*, Dissertação de Mestrado. Departamento de Estatística - UnB, 2011.
- [37] P. Whittle, *Some Distribution and Moment Formulae for the Markov Chain*, J. Roy Stat. Soc. **B 17** (1955), 235–242.
- [38] W. Xu, J. Wu, Z. Huang, and C. Guan, *Kernel Based Hidden Markov Models with Applications to EEG Signal Classification.*, 2005.
- [39] S. Zhong and J. Ghosh, *HMMs and Coupled HMMs for Multi-channel EEG Classification.*, 2002.
- [40] W. Zucchini and I. L. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R.*, CRC Press, 2009.

Apêndice A

Programas R

A.1 Programa 1: `r.normal.hmm`

Aqui apresenta-se a função `r.normal.hmm` que gera uma sequência de tamanho T de um HMM univariado de dimensão m cujas distribuições estado-dependetes são Normais. Os parâmetros da função são:

- `T`: tamanho da sequência que se deseja gerar.
- `u`: vetor de tamanho m . Probabilidade inicial da cadeia de Markov.
- `Gamma`: matriz quadrada $m \times m$. Matriz de transição da cadeia de Markov.
- `mu`: vetor de tamanho m , que guarda as médias das distribuições normais.
- `sigma2`: vetor de tamanho m , que guarda as variâncias das distribuições normais.

O código em R da função é:

```
r.normal.hmm<-function(T,u,Gamma,mu,sigma2){  
  obs.estados<-c(rep(0,times=T))  
  obs<-c(rep(0,times=T))  
  m<-length(u)  
  estados<-seq(1:m)  
  estado.aux<-0
```

```

sigma<-sqrt(sigma2)
if(m==1){
  obs.estados<-rep(1,times=t)
  obs<-rnorm(T,mean=mu,sd=sigma)
  return(list(estados=obs.estados,x=obs))
}
estado.aux<-rmultinom(1,1,u)
obs.estados[1]<-estados%*%estado.aux
obs[1]<-rnorm(1,mean=mu[obs.estados[1]],sd=sigma[obs.estados[1]])
for(t in 2:T){
  estado.aux<-rmultinom(1,1,Gamma[obs.estados[t-1],])
  obs.estados[t]<-estados%*%estado.aux
  i<-obs.estados[t]
  obs[t]<-rnorm(1,mean=mu[[i]],sd=sigma[[i]])
  t<-t+1
}
return(list(estados=obs.estados,x=obs))
}

```

A.2 Programa 2: matriz.P

A função `matriz.P` calcula a matriz $P(x)$, matriz diagonal cujo i -ésimo elemento é o valor da i -ésima densidade estado-dependente p_i normal, no ponto x , $i = 1, 2, \dots, m$.

Os parâmetros da função são:

- `x`: valor do ponto que se deseja calcular as probabilidades.
- `mu`: vetor com as médias das densidades normais.
- `sigma2`: vetor com as variâncias das densidades normais.

A seguir está o programa em R da função `matriz.P`:

```

matriz.P<-function(x, mu, sigma2){
  m<-length(mu)

```

```

P<-matrix(rep(0, times=m*m),ncol=m,nrow=m)
for(i in 1:m){
  P[i,i]<-dnorm(x, mean=mu[i], sd=sqrt(sigma2[i]))
}
return(P)
}

```

Este programa é utilizado em A.3

A.3 Programa 3: veross

A função `veross` calcula a verossimilhança de um vetor x de tamanho T para um modelo HMM univariado, cujas densidades estado-dependente são normais. Os parâmetros da função são;

- x : amostra para a qual se calcula a verossimilhança.
- u : vetor das probabilidades iniciais da CM associada ao modelo.
- γ : matriz de transição da CM associada ao modelo.
- μ : vetor com as médias das densidades normais.
- σ^2 : vetor com as variâncias das densidades normais.

A seguir o programa em R da função `veross`:

```

veross<-function(x, u, gamma, mu, sigma2){
  T<-length(x)
  m<-length(u)
  uns<-matrix(rep(1, times=m),ncol=1,nrow=m)
  Lt<-u%*%matriz.P(x[1], mu, sigma2)
  for(t in 2:T){
    Lt<-Lt%*%gamma%*%matriz.P(x[t], mu, sigma2)
  }
  Lt<-Lt%*%uns
}

```

```

    return(Lt)
}

```

A.4 Programa 3: `log.alphabeta.norm.uni`

Aqui apresenta-se a função `log.alphabeta.norm.uni` que calcula o log das probabilidades *forward* e *backward* para uma sequência observada, e parâmetros conhecidos do modelo HMM univariado de dimensão m , cujas distribuições estado-dependetes são Normais. Os parâmetros da função são:

- x: sequência observada de um HMM.
- m: ordem do modelo.
- gamma: matriz quadrada $m \times m$. Matriz de transição da cadeia de Markov.
- delta: vetor de tamanho m . Distribuição inicial da cadeia de Markov.
- mu: vetor de tamanho m , que guarda as médias das distribuições normais.
- sigma2: vetor de tamanho m , que guarda as variâncias das distribuições normais.

O código em R da função é:

```

log.alphabeta.norm.uni<-function(x,m,gamma,delta,mu,sigma2){
  T<-length(x)
  lalpha<-lbeta<-matrix(NA,m,T)
  allprobs<-matrix(NA,T,m)
  for(t in 1:T){
    P<-matriz.P(x[t],mu,sigma2)
    allprobs[t,]<-diag(P)
  }
  foo <-delta*allprobs[1,]
  sumfoo<-sum(foo)
  lscale<-log(sumfoo)
}

```

```

foo <-foo/sumfoo
lalpha[,1]<-log(foo)+lscale
for(t in 2:T){
  foo <-foo%%gamma*allprobs[t,]
  sumfoo<-sum(foo)
  lscale<-lscale+log(sumfoo)
  foo <-foo/sumfoo
  lalpha[,t]<-log(foo)+lscale
}
lbeta[,T]<-rep(0,m)
foo <-rep(1/m,m)
lscale<-log(m)
for(t in (T-1):1){
  foo <-gamma%%(allprobs[t+1,]*foo)
  lbeta[,t]<-log(foo)+lscale
  sumfoo<-sum(foo)
  foo <-foo/sumfoo
  lscale<-lscale+log(sumfoo)
}
list(la=lalpha,lb=lbeta)
}

```

A.5 Programa 3: EM.hmm.norm.uni

Aqui apresenta-se a função `EM.hmm.norm.uni` que calcula o algoritmo EM para S seqüências x observadas. Retornando as estimativas dos parâmetros do modelo, a verossimilhança máxima obtida, os valores dos critérios AIC e BIC e o número de iterações necessárias para a convergência.

Os parâmetros da função são:

- x : é uma lista (objeto do tipo `list`) de S seqüências. Cada elemento da lista é uma seqüência observada de um HMM.

- m: ordem do modelo.
- delta: vetor de tamanho m . Parâmetro inicial para a distribuição inicial da cadeia de Markov.
- gamma: matriz quadrada $m \times m$. Parâmetro inicial para a matriz de transição da cadeia de Markov.
- mu: vetor de tamanho m , que guarda os parâmetros iniciais das médias das distribuições normais.
- sigma2: vetor de tamanho m , que guarda os parâmetros iniciais das variâncias das distribuições normais.
- I: número máximo de iterações, aqui sugere-se $I = 1000$.
- critério: critério que indica a convergência do algoritmo, aqui sugere-se critério = 10^{-6} .

O código em R da função é:

```
EM.hmm.norm.uni<-function(x,S,m,delta,gamma,mu,sigma2,I=1000,
                           critério=10^(-6)){
  mu.next <-mu
  sigma2.next <-sigma2
  delta.next <-delta
  gamma.next <-gamma
  T<-rep(0,times=S)
  for(s in 1:S){
    T[s]<-length(x[[s]])
  }
  for(i in 1:I){
    probs<-matrix(NA,T[s],m)
    lallprobs<-vector("list",S)
    la<-vector("list",S)
    lb<-vector("list",S)
  }
}
```

```

llk<-rep(0,times=S)
for(s in 1:S){
  for(t in 1:T[s]){
    P<-matriz.P(x[[s]][t],mu,sigma2)
    probs[t,]<-log(diag(P))
  }
  lallprobs[[s]]<-probs
  fb <-log.alphabeta.norm.uni(x[[s]],m,gamma,delta,
                              mu,sigma2)

  la[[s]] <-fb$la
  lb[[s]] <-fb$lb
  c <-max(la[[s]][,T[s]])
  llk[s]<-c+log(sum(exp(la[[s]][,T[s]]-c)))
}

u.1 <- rep (0, times=m)
F <-matrix(0,ncol=m,nrow=m)
for(s in 1:S){
  for (j in 1:m){
    for (k in 1:m){
      F[j,k] <- F[j,k] + gamma[j,k]*
        sum(exp(la[[s]][j,1:(T[s]-1)]+
                lallprobs[[s]][2:T[s],k]+
                lb[[s]][k,2:T[s]]-llk[s]))
    }
  }
  u.1 <- u.1 + exp(la[[s]][,1]+lb[[s]][,1]-llk[s])
}

for (j in 1:m){
  aux.div<-0
  aux.mu<-0
  aux.sigma2<-0
  for(s in 1:S){

```

```

        aux.div<- aux.div + sum(exp(la[[s]][j,]
                                +lb[[s]][j,]-llk[s]))
        aux.mu<- aux.mu + sum(exp(la[[s]][j,]+
                                lb[[s]][j,]-llk[s]) * x[[s]] )
    }
    mu.next[j] <- aux.mu / aux.div
    for(s in 1:S){
        aux.sigma2<- aux.sigma2 + sum(exp(la[[s]][j,]+
                                lb[[s]][j,]-llk[s])*(x[[s]]-
                                mu.next[j])^2 )
    }
    sigma2.next[j]<- aux.sigma2 / aux.div
}
gamma.next <- F/apply(F,1,sum)
delta.next <- u.1/sum(u.1)
crit <- sum(abs(mu-mu.next))+sum(abs(sigma2-sigma2.next))+
        sum(abs(gamma-gamma.next))+
        sum(abs(delta-delta.next))
if(is.nan(crit)==TRUE){
    return(NA)
}
else{
    if(crit<criterio){
        sum.T<-sum(T)
        sum.llk<-sum(llk)
        np <- (m-1)+(m-1)*m + 2*m
        AIC <- -2*(sum.llk-np)
        BIC <- -2*sum.llk+np*log(sum.T)
        return(list(mu=mu,sigma2=sigma2,gamma=gamma,
                    delta=delta,mllk=sum.llk,AIC=AIC,
                    BIC=BIC,iter=i))
    }
}

```



```

        mu <- mu.next
        sigma2 <- sigma2.next
        gamma <- gamma.next
        delta <- delta.next
    }
}
print(paste("O algoritmo não convergiu depois de ",I," iterações."))
return(NA)
}

```

A.6 Programa 3: r.dirichlet

Aqui apresenta-se a função `r.dirichlet` gera um vetor aleatória da distribuição Dirichlet com parâmetro `tetha`. O parâmetro da função é:

- `tetha`: parâmetro da distribuição Dirichlet.

O código em R da função é:

```

r.dirichlet<-function(tetha){
  dim<-length(tetha)
  y<-rep(0,times=dim)
  for(i in 1:dim){
    y[i]<-rgamma(1,shape=tetha[i],scale=1)
  }
  x<-y/sum(y)
  return(x)
}

```

A.7 Programa 3: trans.ergotica

Esta é a função `trans.ergotica` que gera uma matriz de transição ergótica e seu vetor de distribuição estacionária associado. O parâmetro da função é:

- m: ordem da matriz de transição e do vetor de distribuição estacionária.

O código em R da função é:

```
trans.ergotica<-function(m){
  gamma<-matrix(rep(0,times=m^2),ncol=m,nrow=m)
  I<-diag(m)
  verifica<-0
  while(verifica==0){
    for(i in 1:m){
      gamma[i,]<-r.dirichlet(diag(I))
    }
    if(det(gamma-I)==0){
      verifica<-1
    }
  }
  delta<-solve(t(I-gamma+1),rep(1,m))
  return(list(gamma=gamma,delta=delta))
}
```

A.8 Programa 3: mu.sigma2.iniciais

Esta é a função `mu.sigma2.iniciais` que gera um vetor de médias iniciais e outro para as variâncias. Os parâmetros da função são:

- x: é uma lista (objeto do tipo `list`) de S sequências. Cada elemento da lista é uma sequência observada de um HMM.
- S: número de sequências.
- m: ordem do modelo.

O código em R da função é:

```
mu.sigma2.iniciais<-function(x,S,m){
  if(S==1){
```

```

        y<-t(x[[1]])
    }
    else{
        y<-t(x[[1]])
        for(s in 2:S){
            y<-cbind(y,t(x[[s]]))
        }
    }
    y<-sort(y)
    T<-length(y)
    mu<-rep(0,times=m)
    sigma2<-rep(0,times=m)
    for(i in 1:m){
        inicio<-round( (i-1)*T/m + 1 )
        fim<-round( i*T/m )
        mu[i]<-runif(1,y[inicio],y[fim])
    }
    sigma2<-(runif(m, min=(1/4)*sd(y)^2,max=4*sd(y)^2))
    return(list(mu=mu,sigma2=sigma2))
}

```

A.9 Programa 3: `avalia.EM.uni`

Aqui apresenta-se a função `avalia.EM.uni` que inicializa o algoritmo EM com N diferentes conjuntos de pontos iniciais, retornando a estimativa que apresenta a maior das verossimilhanças.

Os parâmetros da função são:

- x : é uma lista (objeto do tipo `list`) de S sequências. Cada elemento da lista é uma sequência observada de um HMM.
- S : número de sequências observadas.
- m : ordem do modelo.

- N: número de conjuntos iniciais para o algoritmo Em, aqui sugere-se 50.
- I: número máximo de iterações, aqui sugere-se $I = 1000$.
- tol.EM: critério que indica a convergência do algoritmo EM, aqui sugere-se $\text{critério} = 10^{(-6)}$.

O código em R da função é:

```

avalia.EM.uni<-function(x,S,m,N=50,I=1000,tol.EM=10^(-6)){
  estima.mu<-numeric(m)
  estima.sigma2<-numeric(m)
  estima.delta<-numeric(m)
  estima.gamma<-matrix(rep(0,times=m^2),ncol=m,nrow=m)
  AIC<-0
  BIC<-0
  historico.mllk<-numeric(N)
  itera.necessarias<-numeric(N)
  melhor.inicial.mu<-numeric(m)
  melhor.inicial.sigma2<-numeric(m)
  melhor.inicial.delta<-numeric(m)
  melhor.inicial.gamma<-matrix(rep(0,times=m^2),ncol=m,nrow=m)
  n<-1
  while(n<N+1){
    print(paste("iteração da avaliação", n))
    m.s2<-mu.sigma2.iniciais(x,S,m)
    mu<-m.s2$mu
    sigma2<-m.s2$sigma2
    gamma.delta<-trans.ergotica(m)
    gamma<-gamma.delta$gamma
    delta<-gamma.delta$delta
    resultado<-EM.hmm.norm.uni(x,S,m,delta,gamma,mu,
                               sigma2,I=I,critério=tol.EM)
    if(length(resultado)<2){

```

```

n<-n
}
else{
  historico.mllk[n]<-resultado$mllk
  itera.necessarias[n]<-resultado$iter
  if(n==1){
    estima.mu<-resultado$mu
    estima.sigma2<-resultado$sigma2
    estima.delta<-resultado$delta
    estima.gamma<-resultado$gamma
    AIC<-resultado$AIC
    BIC<-resultado$BIC
    melhor.inicial.mu<-mu
    melhor.inicial.sigma2<-sigma2
    melhor.inicial.delta<-delta
    melhor.inicial.gamma<-gamma
  }
  else{
    if(historico.mllk[n]>historico.mllk[n-1]){
      estima.mu<-resultado$mu
      estima.sigma2<-resultado$sigma2
      estima.delta<-resultado$delta
      estima.gamma<-resultado$gamma
      AIC<-resultado$AIC
      BIC<-resultado$BIC
      melhor.inicial.mu<-mu
      melhor.inicial.sigma2<-sigma2
      melhor.inicial.delta<-delta
      melhor.inicialgamma<-gamma
    }
  }
n<-n+1

```

```

    }
}
return(list(estimativa.mu=estima.mu,
           estimativa.sigma2=estima.sigma2,
           estimativa.delta=estima.delta,
           estimativa.gamma=estima.gamma,
           AIC=AIC, BIC=BIC,
           mllk=max(historico.mllk)))
}

```

A.10 Programa 3: `estima.cluster`

A função `estima.cluster`, aqui apresentada, é apenas um meio de chamar a função `avalia.EM.uni` usando como parâmetro um objeto tipo lista no lugar de vários objetos, isso é feito somente por causa da estrutura exigida pelo pacote `snow` usado na programação em paralelo. Assim o parâmetro `param` da função condensa os parâmetros `x`, `S`, `m` e `N` da função `avalia.EM.uni`. E os valores retornados são apenas as estimativas dos parâmetros do modelo HMM.

O código em R da função é:

```

estima.cluster<-function(param){
  x<-param$x
  S<-param$S
  m<-param$m
  N<-param$N
  resul<-avalia.EM.uni(x,S,m,N,I=1000,tol.EM=10^(-6))
  return(list(mu=resul$estimativa.mu,
            sigma2=resul$estimativa.sigma2,
            delta=resul$estimativa.delta,
            gamma=resul$estimativa.gamma))
}

```

A.11 Programa 3: `avalia.AICeBIC.cluster`

A função `avalia.AICeBIC.cluster`, aqui apresentada, é apenas um meio de chamar a função `avalia.EM.uni` usando como parâmetro um objeto tipo lista no lugar de vários objetos, isso é feito somente por causa da estrutura exigida pelo pacote `snow` usado na programação em paralelo. Assim o parâmetro `param` da função condensa os parâmetros `x`, `S`, `m` e `N` da função `avalia.EM.uni`. E os valores retornados são apenas os resultados dos critérios AIC e BIC, e a verossimilhança máxima obtida.

O código em R da função é:

```
estima.AICeBIC.cluster<-function(param){
  x<-param$x
  S<-param$S
  m<-param$m
  N<-param$N
  resul<-avalia.EM.uni(x,S,m,N,I=1000,tol.EM=10^(-6))
  return(list(AIC=resul$AIC, BIC=resul$BIC,
             mllk=resul$mllk))
}
```

A.12 Programa 3: `avalia.estima`

A função `avalia.estima` é a que realiza a simulação com o intuito de avaliar a qualidade das estimativas para os parâmetros do modelo HMM, utilizando programação em paralelo. Antes de rodar esta função é preciso realizar alguns passos:

- 1- Rode as funções, como se faz normalmente, dos Programas: A.1, A.2, A.4, A.5, A.6, A.7, A.8, A.9 e A.10.
- 2- Instale e carregue o pacote `snow`.
- 3- Indique o número de clusters (`num.cluster`) que serão utilizados fazendo o comando:

```
cl <- makeCluster(num.cluster)
```

- 4- Rode as mesmas funções citadas em (1) como o seguinte exemplo para a função A.6:

```
clusterEvalQ(cl,r.dirichlet<-function(tetha){
  dim<-length(tetha)
  y<-rep(0,times=dim)
  for(i in 1:dim){
    y[i]<-rgamma(1,shape=tetha[i],scale=1)
  }
  x<-y/sum(y)
  return(x)
})
```

Isso faz com que cada um dos clusters indicados em (3) reconheçam todas as funções que serão utilizadas.

Agora, apresentam-se a seguir os parâmetros da função `avalia.estima`:

- R: número de amostras que serão simuladas e analisadas.
- S: número de sequências para compor as amostras.
- T: tamanho das sequências geradas para as amostras.
- m: ordem do modelo HMM.
- mu: vetor com as médias das densidades normais do modelo HMM.
- sigma2: vetor com as variâncias das densidades normais do modelo HMM.
- delta: distribuição inicial da CM do modelo HMM.
- gamma: matriz de transição da CM do modelo HMM.
- num.cl: número de clusters utilizados.

O código em R da função é:


```

avalia.estima<-function(R,S,T,N,m,mu,sigma2,delta,
                        gamma,num.cl){
  medias<-matrix(rep(0,times=m*R),ncol=m,nrow=R)
  var<-matrix(rep(0,times=m*R),ncol=m,nrow=R)
  d<-matrix(rep(0,times=m*R),ncol=m,nrow=R)
  g<-matrix(rep(0,times=m*m*R),ncol=m*m,nrow=R)
  r<-1
  while(r < (R/num.cl + 1) ){
    param<-vector("list",num.cl)
    y<-vector("list",S)
    for(i in 1:num.cl){
      for(s in 1:S){
        aux<-r.normal.hmm(T,delta,gamma,mu,sigma2)
        y[[s]]<-aux$x
      }
      param[[i]]<-list(x=y,S=S,m=m,N=N)
    }
    resp<-clusterApply(cl, param , estima.cluster)
    for(j in 1:num.cl){
      aux.ordem<-rank(resp[[j]]$mu)
      medias[num.cl*(r-1)+j,]<-sort(resp[[j]]$mu)
      for(i in 1:m){
        var[num.cl*(r-1)+j,i]<-
          resp[[j]]$sigma2[aux.ordem[i]]
        d[num.cl*(r-1)+j,i]<-
          resp[[j]]$delta[aux.ordem[i]]
      }
      aux.g.linha<-matrix(rep(0,times=m*m),ncol=m)
      for(i in 1:m){
        aux.g.linha[i,]<-
          resp[[j]]$gamma[aux.ordem[i],]
      }
    }
  }
}

```

```

    aux.g.coluna<-matrix(rep(0,times=m*m),ncol=m)
    for(i in 1:m){
        aux.g.coluna[,i]<-aux.g.linha[,aux.ordem[i]]
    }
    for(i in 1:m){
        g[num.cl*(r-1)+j,(m*(i-1)+1):(m*(i-1)+m)]<-
            aux.g.coluna[i,]
    }
}
r<-r+1
}
return(list(est.medias=medias,
            est.var=var,
            est.delta=d,
            est.gamma=g))
}

```

A.13 Programa 3: `avalia.AICeBIC`

A função `avalia.AICeBIC` é a que realiza a simulação com o intuito de avaliar a qualidade dos critérios AIC e BIC, utilizando programação em paralelo. Antes de rodar esta função é preciso realizar alguns passos:

- 1- Rode as funções, como se faz normalmente, dos Programas: A.1, A.2, A.4, A.5, A.6, A.7, A.8, A.9 e A.11.
- 2- Instale e carregue o pacote `snow`.
- 3- Indique o número de clusters (`num.cluster`) que serão utilizados fazendo o comando:

```
cl <- makeCluster(num.cluster)
```
- 4- Rode as mesmas funções citadas em (1) como foi explicado no passo (4) do Programa A.12.

Agora, apresentam-se a seguir os parâmetros da função `avalia.estima`:

- R: número de amostras que serão simuladas e analisadas.
- S: número de sequências para compor as amostras.
- T: tamanho das sequências geradas para as amostras.
- m: ordem do modelo HMM.
- mu: vetor com as médias das densidades normais do modelo HMM.
- sigma2: vetor com as variâncias das densidades normais do modelo HMM.
- delta: distribuição inicial da CM do modelo HMM.
- gamma: matriz de transição da CM do modelo HMM.
- num.cl: número de clusters utilizados.

O código em R da função é:

```
avalia.AICeBIC<-function(R,S,T,N,m,mu,sigma2,delta,
                        gamma,inter.m,num.cl){
  AIC<-matrix(rep(0,times=length(inter.m)*R),
             ncol=length(inter.m),nrow=R)
  BIC<-matrix(rep(0,times=length(inter.m)*R),
             ncol=length(inter.m),nrow=R)
  mllk<-matrix(rep(0,times=length(inter.m)*R),
             ncol=length(inter.m),nrow=R)
  r<-1
  while(r < (R/num.cl + 1) ){
    param<-vector("list",num.cl)
    y<-vector("list",S)
    Y<-vector("list",num.cl)
    for(i in 1:num.cl){
      for(s in 1:S){
        aux<-r.normal.hmm(T,delta,gamma,mu,sigma2)
```

```

        y[[s]]<-aux$x
    }
    Y[[i]]<-y
}
for(k in 1:length(inter.m)){
    for( i in 1:num.cl){
        param[[i]]<-list(x=Y[[i]],S=S,
                        m=inter.m[k],N=N)
    }
    resp<-clusterApply(cl, param ,
                      estima.AICeBIC.cluster)
    for(j in 1:num.cl){
        AIC[num.cl*(r-1)+j,k]<-resp[[j]]$AIC
        BIC[num.cl*(r-1)+j,k]<-resp[[j]]$BIC
        mllk[num.cl*(r-1)+j,k]<-resp[[j]]$mllk
    }
}
r<-r+1
}
return(list(AIC=AIC, BIC=BIC, mllk=mllk))
}

```