

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Algoritmos de estimação para modelos Markovianos  
não-homogêneos**

**Gustavo Alexis Sabillón Lee**

Dissertação de Mestrado do Programa Interinstitucional de  
Pós-Graduação em Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Gustavo Alexis Sabillón Lee**

## Algoritmos de estimação para modelos Markovianos não-homogêneos

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Profa. Dra. Daiane Aparecida Zuanetti

**USP/UFSCar – São Carlos**  
**Fevereiro de 2020**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

S116a Sabillón Lee, Gustavo Alexis  
Algoritmos de estimação para modelos Markovianos  
não-homogêneos / Gustavo Alexis Sabillón Lee;  
orientadora Daiane Aparecida Zuanetti. -- São  
Carlos, 2020.  
104 p.

Dissertação (Mestrado - Programa  
Interinstitucional de Pós-graduação em Estatística) --  
Instituto de Ciências Matemáticas e de Computação,  
Universidade de São Paulo, 2020.

1. Modelos Markovianos ocultos. 2. Matriz de  
transição. 3. Estados ocultos. 4. Cadeias de Markov.  
5. Modelagem estatística. I. Zuanetti, Daiane  
Aparecida, orient. II. Título.

**Gustavo Alexis Sabillón Lee**

## Estimation algorithms for non-homogeneous Markov models

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics.  
*FINAL VERSION*

Concentration Area: Statistics

Advisor: Profa. Dra. Daiane Aparecida Zuanetti

**USP/UFSCar – São Carlos**  
**February 2020**



# UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

## Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Gustavo Alexis Sabillón Lee, realizada em 27/02/2020:

*Daiane Aparecida Zuanetti*

---

Profa. Dra. Daiane Aparecida Zuanetti  
UFSCar

*Erlandson*

---

Prof. Dr. Erlandson Ferreira Saraiva  
UFMS

*Ricardo Felipe Ferreira*

---

Prof. Dr. Ricardo Felipe Ferreira  
UFSCar



*Para ustedes, Abuela Gloria y Abuela Mercedes.*





# AGRADECIMENTOS

---

---

Primeiro agradeço a Deus, nosso Padre e criador. Ele me deu a sabedoria e a perseverança para completar esse projeto e ilumina meu caminho em todo momento.

Aos meus pais, Norman e Gioconda, pelo seus exemplos de amor, bondade e superação. Em todo momento, mesmo estando longe, tive todo o apoio deles fazendo todo o possível para me ajudar nos momentos difíceis.

A meu irmão Wilberto, por ser o melhor exemplo do bem e humildade que tenho na minha vida. A minha irmã Gloria, por todo o carinho e a dedicação à família. Sem o exemplo deles, não teria sido possível realizar este sonho.

A minha orientadora, a professora Daiane Aparecida Zuanetti, pela confiança que depositou em mim, pela sua enorme paciência e por sempre ter a melhor disposição para ajudar nos momentos difíceis. Muito obrigado por ser um referência profissional e de excelência para mim.

Aos professores da USP e da UFSCar pelos valiosos ensinamentos, por fornecer uma excelente base para meu encaminhamento na área da Estatística.

Ao meu amigo Alex de la Cruz, por sua amizade, e disponibilidade para ajudar em todo momento. Aos meus colegas do PIPGEs, pelos bons tempos compartilhados e pelas muitas horas de estudo.

A todas as pessoas que de alguma forma ajudaram, quero deixar um agradecimento especial, porque sem eles isto não teria sido possível.

Finalmente, quero agradecer a minha companheira de aventuras e meu amor eterno, Jeny. Sem o apoio dela, sem a sua ajuda e sem a força que ela me da todos os dias, nunca teria nem pensado em estar onde eu estou e conseguir os objetivos que tenho atingido. Te amo para sempre. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



*“One of the things we do not know is the number of things we do not know.”*

*Peter J. Green*



# RESUMO

GUSTAVO, S. L. **Algoritmos de estimação para modelos Markovianos não-homogêneos.** 2020. 104 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Os modelos Markovianos ocultos são um paradigma estatístico que podem ser utilizados para modelar processos estocásticos onde valores observáveis dependem diretamente de uma sequência de variáveis aleatórias não observáveis. No modelo Markoviano oculto o sistema que está sendo modelado é considerado um processo de Markov com estados não observáveis (isto é, ocultos) e em cada estado oculto temos a emissão de um valor observável. Os modelos Markovianos ocultos podem ser homogêneos ou não-homogêneos. O foco principal deste trabalho, serão os modelos Markovianos não-homogêneos.

Neste trabalho, apresentamos alguns procedimentos de estimação utilizados com modelos Markovianos. A estimação dos parâmetros é realizada sob abordagem Bayesiana e frequentista, fazendo uma comparação da performance dos mesmos usando algumas métricas como o erro quadrático médio e o viés dos estimadores. A seleção dos modelos é desenvolvida utilizando diferentes critérios para seleção de modelos como o *Bayes Information Criterion* e o *Deviance Information Criterion*. Os menores erros quadraticos medios são obtidos sob a abordagem Bayesiana. Sob a abordagem frequentista, o algoritmo EM Estocástico obtém resultados similares ao algoritmo Bayesiano. O algoritmo EM apresenta problemas de no processo de estimação em todas as situações.

**Palavras-chave:** Cadeia de Markov, matriz de transição, estados ocultos, modelo Markoviano oculto.



# ABSTRACT

GUSTAVO, S. L. **Estimation algorithms for non-homogeneous Markov models**. 2020. 104 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Hidden Markov models are a statistical paradigm which can be used to model stochastic processes where the observable values are directly dependent on a sequence of hidden random variables. In the context of the hidden Markov model, the system being modeled is considered a Markov process with non-observable hidden states, and for each hidden state we have the emission of an observable value. Hidden Markov models can be homogeneous or non-homogeneous.

In this investigation, we present estimation procedures used with Markov models. Parameters estimation is done under Bayesian and frequentist perspectives, comparing the performance of these methods using metrics such as mean squared error and bias. Model selection is carried out using different criteria such as the Bayes Information Criterion and the Deviance Information Criterion. The smallest mean squared errors and biases were obtained using the Bayesian estimation algorithm. In the frequentist perspective, the Stochastic EM algorithm obtained results which were similar to the Bayesian algorithm. The EM algorithm presented problems in the estimation procedure in all situations studied.

**Keywords:** Markov chain, transition matrix, hidden states, hidden Markov model.





# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Urnas associadas ao experimento. . . . .	26
Figura 2 – Representação de uma sequência de bolas retiradas. . . . .	26
Figura 3 – Sequência de bolas retiradas e as urnas escolhidas. . . . .	27
Figura 4 – Exemplo de retirada de bolas com os elementos do modelo Markoviano oculto homogêneo. . . . .	29
Figura 5 – Ilustração da parte <i>forward</i> do algoritmo <i>forward-backward</i> . . . . .	39
Figura 6 – Ilustração da parte <i>backward</i> do algoritmo <i>forward-backward</i> . . . . .	41
Figura 7 – Funções de transição do estado 1 com $K = 2$ na situação A. . . . .	55
Figura 8 – Funções de transição do estado 2 com $K = 2$ na situação A. . . . .	55
Figura 9 – Funções de transição do estado 1 com $K = 2$ na Situação B. . . . .	60
Figura 10 – Funções de transição do estado 2 com $K = 2$ na Situação B. . . . .	60
Figura 11 – Funções de Transição do estado 1 com $K = 3$ na Situação A. . . . .	65
Figura 12 – Funções de Transição do estado 2 com $K = 3$ na Situação A. . . . .	65
Figura 13 – Funções de Transição do estado 3 com $K = 3$ na Situação A. . . . .	66
Figura 14 – Funções de Transição do estado 1 com $K = 3$ na situação B. . . . .	71
Figura 15 – Funções de Transição do estado 2 com $K = 3$ na Situação B. . . . .	71
Figura 16 – Funções de Transição do estado 3 com $K = 3$ na Situação B. . . . .	71
Figura 17 – Histograma e densidade dos dados de dias com chuva no mês. . . . .	78
Figura 18 – Média de dias com chuva mensais no período 1969-2019. . . . .	78
Figura 19 – Gráficos de convergência para os parâmetros do modelo DiasChuva-Temperatura. . . . .	81
Figura 20 – Gráficos de densidade <i>a posteriori</i> para os parâmetros do modelo DiasChuva- Temperatura. . . . .	82
Figura 21 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva- Temperatura. . . . .	82
Figura 22 – Gráficos de autocorrelação para os parâmetros do modelo DiasChuva-Temperatura. . . . .	82
Figura 23 – Gráficos de convergência para os parâmetros do modelo DiasChuva-UmidadeRelativa. . . . .	83
Figura 24 – Gráficos de densidade <i>a posteriori</i> para os parâmetros do modelo DiasChuva- UmidadeRelativa. . . . .	83
Figura 25 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva- UmidadeRelativa. . . . .	84
Figura 26 – Gráficos de autocorrelação para os parâmetros do modelo DiasChuva-UmidadeRelativa. . . . .	84
Figura 27 – Gráficos de convergência para os parâmetros do modelo DiasChuva-Temperatura. . . . .	88

Figura 28 – Gráficos de densidade <i>a posteriori</i> para os parâmetros do modelo DiasChuva-Temperatura. . . . .	88
Figura 29 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva-Temperatura. . . . .	89
Figura 30 – Gráficos de autocorrelação para os parâmetros do modelo DiasChuva-Temperatura.	89
Figura 31 – Gráficos de convergência para os parâmetros do modelo DiasChuva-UmidadeRelativa.	91
Figura 32 – Gráficos de densidade <i>a posteriori</i> para os parâmetros do modelo DiasChuva-UmidadeRelativa. . . . .	91
Figura 33 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva-UmidadeRelativa. . . . .	92
Figura 34 – Gráficos de autocorrelação para os parâmetros do DiasChuva-Temperatura. .	92
Figura 35 – Funções de transição do estado 1 . . . . .	93
Figura 36 – Funções de transição do estado 2 . . . . .	94
Figura 37 – Funções de transição do estado 3 . . . . .	94
Figura 38 – Sequência de estados ocultos e valores observados no período 1969-1975. .	96
Figura 39 – Sequência de estados ocultos e valores observados no período 1976-1982. .	96
Figura 40 – Sequência de estados ocultos e valores observados no período 1983-1989. .	96
Figura 41 – Sequência de estados ocultos e valores observados no período 1990-1996. .	97
Figura 42 – Sequência de estados ocultos e valores observados no período 1997-2003. .	97
Figura 43 – Sequência de estados ocultos e valores observados no período 2004-2010. .	97
Figura 44 – Sequência de estados ocultos e valores observados no período 2011-2019. .	98

# LISTA DE TABELAS

---

---

Tabela 1 – Resultados do algoritmo EM para a Situação A e $K = 2$ com $T = 300$ . . . .	55
Tabela 2 – Resultados do algoritmo EM Estocástico para a Situação A e $K = 2$ com $T = 300$ . . . . .	56
Tabela 3 – Resultados do algoritmo Bayesiano para a Situação A e $K = 2$ com $T = 300$ .	56
Tabela 4 – Indicadores para os 3 algoritmos para $K = 2$ e $T = 300$ . . . . .	56
Tabela 5 – Resultados do algoritmo EM para a Situação A e $K = 2$ com $T = 600$ . . . .	57
Tabela 6 – Resultados do algoritmo EM Estocástico para a Situação A e $K = 2$ com $T = 600$ . . . . .	57
Tabela 7 – Resultados do algoritmo Bayesiano para a Situação A e $K = 2$ com $T = 600$ .	57
Tabela 8 – Indicadores para os 3 algoritmos para $K = 2$ e $T = 600$ . . . . .	58
Tabela 9 – Resultados do algoritmo EM para a Situação A e $K = 2$ com $T = 1200$ . . . .	58
Tabela 10 – Resultados do algoritmo EM Estocástico para a Situação A e $K = 2$ com $T = 1200$ . . . . .	58
Tabela 11 – Resultados do algoritmo Bayesiano para a Situação A e $K = 2$ com $T = 1200$ .	58
Tabela 12 – Indicadores para os 3 algoritmos para $K = 2$ e $T = 1200$ na situação A. . . .	59
Tabela 13 – Resultados do algoritmo EM para a Situação B e $K = 2$ com $T = 300$ . . . .	61
Tabela 14 – Resultados do algoritmo EM Estocástico para a Situação B e $K = 2$ com $T = 300$ . . . . .	61
Tabela 15 – Resultados do algoritmo Bayesiano para a Situação B e $K = 2$ com $T = 300$ .	61
Tabela 16 – Indicadores para os 3 algoritmos para $K = 2$ e $T = 300$ na situação B. . . .	61
Tabela 17 – Resultados do algoritmo EM para a Situação B e $K = 2$ com $T = 600$ . . . .	62
Tabela 18 – Resultados do algoritmo EM Estocástico para a Situação B e $K = 2$ com $T = 600$ . . . . .	62
Tabela 19 – Resultados do algoritmo Bayesiano para a Situação B e $K = 2$ com $T = 600$ .	62
Tabela 20 – Indicadores para os 3 algoritmos para $K = 2$ e $T = 600$ na situação B. . . .	63
Tabela 21 – Resultados do algoritmo EM para a Situação B e $K = 2$ com $T = 1200$ . . . .	63
Tabela 22 – Resultados do algoritmo EM Estocástico para a Situação B e $K = 2$ com $T = 1200$ . . . . .	63
Tabela 23 – Resultados do algoritmo Bayesiano para a Situação B e $K = 2$ com $T = 1200$ .	63
Tabela 24 – Indicadores para os 3 algoritmos para $K = 2$ e $T = 1200$ na situação B. . . .	64
Tabela 25 – Resultados do algoritmo EM Estocástico para a Situação A e $K = 3$ com $T = 300$ . . . . .	66
Tabela 26 – Resultados do algoritmo Bayesiano para a Situação A e $K = 3$ com $T = 300$ .	67

Tabela 27 – Indicadores para os 2 algoritmos para $K = 3$ e $T = 300$ . . . . .	67
Tabela 28 – Resultados do algoritmo EM Estocástico para a Situação A e $K = 3$ com $T = 600$ . . . . .	67
Tabela 29 – Resultados do algoritmo Bayesiano para a Situação A e $K = 3$ com $T = 600$ .	68
Tabela 30 – Indicadores para os 2 algoritmos para $K = 3$ e $T = 600$ . . . . .	68
Tabela 31 – Resultados do algoritmo EM Estocástico para a Situação A e $K = 3$ com $T = 1200$ . . . . .	69
Tabela 32 – Resultados do algoritmo Bayesiano para a Situação A e $K = 3$ com $T = 1200$ .	69
Tabela 33 – Indicadores para os 2 algoritmos para $K = 3$ e $T = 600$ . . . . .	70
Tabela 34 – Resultados do algoritmo EM Estocástico para a Situação B e $K = 3$ com $T = 300$ . . . . .	72
Tabela 35 – Resultados do algoritmo Bayesiano para a Situação B e $K = 3$ com $T = 300$ .	72
Tabela 36 – Indicadores para os 2 algoritmos para $K = 3$ e $T = 300$ na Situação B. . . .	73
Tabela 37 – Resultados do algoritmo EM Estocástico para a Situação B e $K = 3$ com $T = 600$ . . . . .	73
Tabela 38 – Resultados do algoritmo Bayesiano para a Situação B e $K = 3$ com $T = 600$ .	74
Tabela 39 – Indicadores para os 2 algoritmos para $K = 3$ e $T = 600$ na situação B. . . .	74
Tabela 40 – Resultados do algoritmo EM Estocástico para a Situação B e $K = 3$ com $T = 1200$ . . . . .	75
Tabela 41 – Resultados do algoritmo Bayesiano para a Situação B e $K = 3$ com $T = 1200$ .	75
Tabela 42 – Indicadores para os 2 algoritmos para $K = 3$ e $T = 1200$ na Situação B. . .	75
Tabela 43 – Estatística descritiva das variáveis utilizadas. . . . .	78
Tabela 44 – Resultados para o ajuste DiasChuva-Temperatura. . . . .	79
Tabela 45 – Resultados para o ajuste DiasChuva-UmidadeRelativa. . . . .	80
Tabela 46 – Resultados para o ajuste DiasChuva-Temperatura. . . . .	80
Tabela 47 – Resultados para o ajuste DiasChuva-UmidadeRelativa. . . . .	80
Tabela 48 – Resultados para o ajuste DiasChuva-Temperatura. . . . .	81
Tabela 49 – Resultados para os ajuste DiasChuva-UmidadeRelativa. . . . .	83
Tabela 50 – Resultados para o ajuste DiasChuva-Temperatura. . . . .	85
Tabela 51 – Resultados para o ajuste DiasChuva-UmidadeRelativa. . . . .	85
Tabela 52 – Resultados para o ajuste DiasChuva-Temperatura. . . . .	86
Tabela 53 – Resultados para o ajuste DiasChuva-UmidadeRelativa. . . . .	86
Tabela 54 – Resultados para o ajuste DiasChuva-Temperatura. . . . .	87
Tabela 55 – Resultados para o ajuste DiasChuva-UmidadeRelativa. . . . .	90
Tabela 56 – Comparação de BIC entre modelos ajustados. . . . .	93

# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	21
2	MODELOS MARKOVIANOS OCULTOS . . . . .	25
2.1	Exemplo ilustrativo de HMM . . . . .	25
2.2	Modelo Markoviano oculto homogêneo . . . . .	27
2.3	Definição do HHMM . . . . .	27
2.4	Estimações de interesse . . . . .	30
2.5	Modelo Markoviano oculto não-homogêneo . . . . .	32
2.6	Definição do NHMM . . . . .	32
2.6.1	<i>Matriz de transição não-homogênea</i> . . . . .	34
3	PROCESSOS DE ESTIMAÇÃO . . . . .	37
3.1	Algoritmo <i>Expectation-Maximization</i> Tradicional . . . . .	37
3.1.1	<i>Algoritmo Forward-Backward</i> . . . . .	38
3.1.1.1	<i>Procedimento Forward</i> . . . . .	38
3.1.1.2	<i>Procedimento Backward</i> . . . . .	40
3.1.2	<i>Algoritmo de Viterbi</i> . . . . .	40
3.1.3	<i>Algoritmo de Baum-Welch</i> . . . . .	42
3.2	Algoritmo EM Estocástico . . . . .	44
3.2.1	<i>Estimação dos Parâmetros</i> . . . . .	45
3.3	Estimação Bayesiana para o número de estados ocultos $K$ fixo . . . . .	48
3.3.1	<i>Algoritmo de estimação</i> . . . . .	49
3.4	Troca de Rótulos - <i>Label switching</i> . . . . .	50
4	SIMULAÇÃO . . . . .	51
4.1	Escolha da função de ligação . . . . .	51
4.1.1	<i>Escolha dos coeficientes da matriz de transição</i> . . . . .	52
4.2	Estudo de simulação . . . . .	52
4.2.1	<i>Situação A com <math>K = 2</math></i> . . . . .	54
4.2.2	<i>Situação B com <math>K = 2</math></i> . . . . .	59
4.2.3	<i>Situação A com <math>K = 3</math></i> . . . . .	64
4.2.4	<i>Situação B para <math>K = 3</math></i> . . . . .	70
5	APLICAÇÃO EM DADOS REAIS . . . . .	77

<b>5.1</b>	<b>Descrição dos dados</b> . . . . .	<b>77</b>
<b>5.2</b>	<b>Estimação com NHMM</b> . . . . .	<b>79</b>
<b>5.2.1</b>	<b>Ajuste com dois estados ocultos, <math>K = 2</math></b> . . . . .	<b>79</b>
5.2.1.1	<i>Algoritmo EM</i> . . . . .	79
5.2.1.2	<i>Algoritmo EM Estocástico</i> . . . . .	80
5.2.1.3	<i>Estimação Bayesiana</i> . . . . .	81
<b>5.2.2</b>	<b>Ajuste com três estados ocultos, <math>K=3</math></b> . . . . .	<b>84</b>
5.2.2.1	<i>Algoritmo EM</i> . . . . .	84
5.2.2.2	<i>Algoritmo EM Estocástico</i> . . . . .	85
5.2.2.3	<i>Estimação Bayesiana</i> . . . . .	86
<b>6</b>	<b>CONSIDERAÇÕES FINAIS E PROPOSTAS FUTURAS</b> . . . . .	<b>101</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>103</b>

---

# INTRODUÇÃO

---

Modelos Markovianos Ocultos são modelos estatísticos nos quais o sistema que está sendo modelado é considerado um processo de Markov com estados não observáveis (isto é, ocultos) e em cada estado oculto temos a emissão de um valor observável.

[Espindola \(2009\)](#) menciona que o formalismo de Modelos Ocultos de Markov (*Hidden Markov Models - HMM*) pode ser encontrado sob os mais diversos nomes, dentre eles: Processos Ocultos de Markov (*Hidden Markov Processes*), Fontes Markovianas (*Markov Sources*), Cadeias de Markov Ocultas (*Hidden Markov Chains*), Funções Probabilísticas de Cadeias de Markov (*Probabilistic Functions of Markov Chains*).

Neste trabalho é estudado predominantemente o modelo Markoviano oculto (*Hidden Markov Model*) ou HMM por sua sigla em inglês. Este paradigma refere-se aos modelos com espaço de estados ocultos discreto e desenvolvendo-se em tempos discretos. Os modelos Markovianos ocultos podem ser homogêneos ou não-homogêneos. A diferença entre os HMM homogêneos e não-homogêneos é que a matriz de transição de estados do HMM homogêneo é constante ao longo do tempo, enquanto a matriz de transição do HMM não-homogêneo não é constante.

Muitos dos processos que podem ser representados teoricamente por modelos Markovianos ocultos homogêneos na prática apresentam condições externas que modificam as probabilidades de transição entre os estados ocultos. Essas modificações ocorrem em cada instante do tempo. Por esse motivo, o HMM homogêneo pode não ser adequado para a modelagem de alguns processos ou sistemas e é necessário mudar a sua estrutura, para que ele se ajuste a essas condições externas que influem nas probabilidades de transição entre os estados ocultos. Dito isso, a ideia do modelo Markoviano não-homogêneo é introduzida como uma opção interessante para modelar processos e sistemas que apresentam essas mudanças nas probabilidades de transição ao longo do tempo.

Os modelos Markovianos ocultos tem uma ampla gama de aplicações. Sua aplicação



pode ser encontrada em diversas áreas do conhecimento. Pennoni e Genge (2019) propõem um método estatístico baseado no modelo Markoviano oculto não-homogêneo para analisar a confiança pública em entidades políticas. Elas concebem a confiança como um atributo mental não observável que está relacionado com covariáveis que variam no tempo e propõem um modelo Markoviano para modelar o fenômeno. Holsclaw *et al.* (2017) os aplicam em dados de quantidades de chuva na Índia ao longo de 30 anos, chegando assim em múltiplas conclusões sobre vários fenômenos relacionados com a chuva na região. Lagona, Maruotti e Picone (2011) utilizam os modelos Markovianos não-homogêneos para estimar as probabilidades de excedências de poluentes múltiplos, condicionando sobre fatores que variam no tempo e podem influenciar a ocorrência e persistência de episódios de poluição. Zuanetti e Milan (2017) propõe um algoritmo *Data-Driven Reversible Jump* para fazer inferência sobre modelos Markovianos homogêneos e analisar dados de incidência de diabetes nos Estados Unidos. Zamanzad, Claesen e Burzykowski (2015) propõem o uso do modelo não-homogêneo para analisar dados de marcadores genéticos. O modelo proposto por eles inclui vários estados ocultos, cada um associado à probabilidade de observar algum nucleótido numa sequência genética.

É evidente que as aplicações dos modelos Markovianos ocultos são inúmeras e o estudo destes modelos e sua estimação é de grande importância para todas as áreas do conhecimento. Portanto, este trabalho tem como objetivo definir, apresentar e comparar alguns métodos de estimação dos HMMs não homogêneos, aplicando-os a dados simulados e reais. O Capítulo 2 aborda os aspectos teóricos dos modelos Markovianos ocultos homogêneos e não-homogêneos. Entre eles: definição, elementos e função de verosimilhança. O Capítulo 3 apresenta algumas ferramentas computacionais para a estimação dos HMM. Estas ferramentas são apresentadas sob duas perspectivas diferentes de estimação. Na metodologia frequentista, são apresentados o Algoritmo *Expectation-Maximization* (EM) e o Algoritmo EM Estocástico. Sob a perspectiva Bayesiana é apresentado um método de estimação via Monte Carlo com Cadeias de Markov, com número de estados ocultos  $K$  fixo. O Capítulo 4 traz um estudo de simulação sob diferentes condições que foram propostas. Neste estudo de simulação, a performance dos algoritmos é comparada usando métricas como o erro quadrático médio e viés dos estimadores. Para seleção de modelos são utilizados critérios de seleção de modelos como o *Bayes Information Criteria* e *Deviance Information Criteria*. O método de estimação via Monte Carlo com Cadeias de Markov apresenta os menores vieses e erros quadráticos médios. Além disto, ao aumentar o comprimento da cadeia observamos uma melhora considerável nas estimativas obtidas por esse método. O algoritmo EM Estocástico apresenta resultados similares ao algoritmo Bayesiano. O algoritmo EM apresenta problemas no processo de estimação em todas as situações propostas na simulação. No Capítulo 5 apresenta-se uma aplicação para análise de padrões de chuva usando dados climatológicos da cidade de Tegucigalpa, Honduras. Na aplicação em dados reais, o modelo que apresenta o menor BIC foi o ajustado pelo algoritmo Bayesiano, com 3 estados ocultos identificados. No estudo de simulação e na aplicação, todos os algoritmos são programados

utilizando a linguagem R, desenvolvida por [R Core Team \(2013\)](#). Na abordagem Bayesiana, utilizamos o RJAGS. Finalmente algumas ideias e propostas futuras são mencionadas no [Capítulo 6](#) para o desenvolvimento deste trabalho.



---

# MODELOS MARKOVIANOS OCULTOS

---

Os modelos Markovianos ocultos são uma classe de modelos estatísticos na qual uma sequência de valores observáveis são gerados por uma sequência de estados que não são observáveis. Diversos autores tem perspectivas similares em quanto à definição do que são modelos Markovianos ocultos.

[Rabiner \(1989\)](#) comenta que um modelo Markoviano no qual cada estado corresponde a um evento observável é um modelo que se torna restritivo na aplicação a problemas de interesse. Quando se estende o conceito de cadeia de Markov para incluir a situação na qual o valor observável é uma função de um estado não observável, o modelo resultante é um processo estocástico observável embutido em um processo estocástico que não se observa (é oculto) e somente pode ser predito através da sequência de valores observáveis. [Rabiner \(1989\)](#) chama o sistema anteriormente descrito de Modelo Markoviano Oculto. Eles são conhecidos por sua sigla em inglês, HMM. Os HMM, de acordo com [Eddy \(2004\)](#), são uma base formal para a construção de modelos probabilísticos aplicáveis para problemas de rotulação de sequências lineares. [Zuanetti \(2006\)](#) os descreve como um caso particular das redes probabilísticas dinâmicas e os define como um processo estocástico finito no qual cada estado está associado a uma distribuição de probabilidades que gera as observações. [Ross \(2006\)](#) chama de modelo Markoviano oculto um modelo no qual uma sequência de sinais é observada, e uma sequência de estados de uma cadeia de Markov subjacente não é observada.

## 2.1 Exemplo ilustrativo de HMM

Como ilustração simples de um HMM, apresentamos o seguinte exemplo. Vamos supor que uma pessoa, denominada de “jogador” está em uma sala fechada. O jogador tem na sua frente 3 urnas, que chamaremos de Urna1, Urna 2 e Urna 3 com diferentes quantidades de bolas pretas e brancas. Para uma ilustração ver [Figura 1](#).

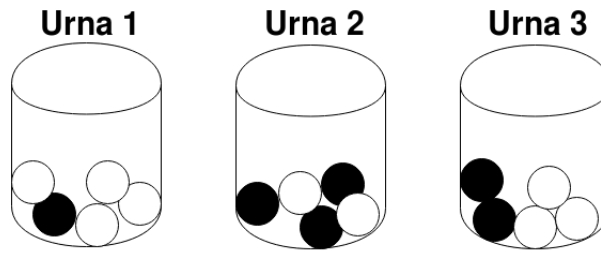


Figura 1 – Urnas associadas ao experimento.

Inicialmente, a probabilidade de escolher qualquer urna pode ser igual a  $1/3$ , mas as probabilidades iniciais delas poderiam ser diferentes, somando um. Cada urna contém quantidades diferentes de bolas brancas e pretas. Denotaremos por “B” o evento “retirar uma bola branca” e denotaremos por “P” o evento “retirar uma bola preta”. O jogador escolhe uma urna e retira uma bola dessa urna. Ele registra a cor da bola e a recoloca na urna. Depois ele escolhe uma segunda urna, que pode ser igual a urna anterior ou outra, para a seguinte retirada de bola, retira uma bola e registra o resultado. Os resultados das retiradas de bolas formam uma sequência de observações que podem ser “P” ou “B”. Uma possível sequência é, por exemplo, BPPBPBPB. Ver Figura 2.



Figura 2 – Representação de uma sequência de bolas retiradas.

Esta sequência, depois de finalizar o experimento, é reportada para uma outra pessoa que não presenciou o experimento, e será chamada de “observador”. Claramente, o observador não conhece a sequência em que foram escolhidas as urnas, nem a quantidade de bolas brancas e pretas que cada urna contém. Em outras palavras o fenômeno que gera as retiradas de bolas está oculto ao observador. Uma ilustração do processo completo da retirada de bolas das urnas é dado na Figura 3. As urnas apresentam probabilidade de bolas pretas e brancas diferente e a sua sequência não é observada. Dessa maneira, as urnas representam os estados ocultos, mencionados anteriormente. Os três conjuntos de probabilidade, isto é, o conjunto das probabilidades iniciais das urnas, o conjunto de probabilidades de transição de uma urna para a próxima, e conjunto de probabilidades condicionais de observar uma cor de bola específica em cada urna, caracterizam o que é chamado de HMM para o experimento de retirada de bolas de urnas.

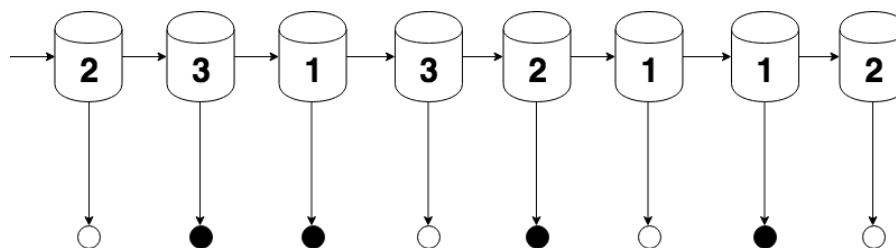


Figura 3 – Sequência de bolas retiradas e as urnas escolhidas.

No exemplo anterior e como podemos observar na Figura 3, a sequência de urnas pode ser representada por uma cadeia de Markov, na qual a probabilidade da  $t$ -ésima urna ser escolhida depende só da escolha da  $(t - 1)$ -ésima urna. Cada urna gera uma observação (a cor da bola retirada da urna).

Os HMMs podem ser classificados de acordo com suas várias características. Aqui os classificaremos nos baseando na sua matriz de transição. Vários autores mencionam duas possíveis classificações: HMM homogêneo e HMM não-homogêneo.

Zucchini, MacDonald e Langrock (2017) chamam de cadeia de Markov homogênea aquela na qual as probabilidades de transição não dependem do momento de tempo, caso contrário elas são chamadas de não-homogêneas. Estes conceitos podem ser estendidos à cadeia de Markov subjacente de um HMM. Stewart (2009) diz que as probabilidades de transição podem depender da quantidade de tempo que transcorreu desde o início da cadeia até o momento atual e de covariáveis que se modificam ao longo do tempo e, quando essas transições, de fato, dependem do tempo, a cadeia é chamada de não-homogênea. Aqui trataremos como HMM homogêneo o modelo cuja cadeia de Markov subjacente é homogênea e como HMM não homogêneo o modelo cuja cadeia de Markov subjacente é não homogênea.

## 2.2 Modelo Markoviano oculto homogêneo

Como foi dito anteriormente, um HMM homogêneo (HHMM, sua sigla em inglês) é aquele no qual a matriz de probabilidades de transição é constante no tempo. Sendo assim, os elementos do modelo para o caso em que a variável observável é discreta são os seguintes:

### 2.3 Definição do HHMM

1. Um espaço discreto de estados não observáveis  $\varphi = \{1, 2, \dots, K\}$ ;
2. Um conjunto de valores observáveis  $\omega = \{1, 2, \dots, M\}$ ;
3. Uma variável aleatória  $S$  que assume os valores do espaço de estados ocultos  $\varphi$ , em diferentes instantes do tempo;

4. Uma variável aleatória  $Y$  que assume os valores do conjunto de observáveis  $\omega$ , em diferentes instantes de tempo;
5. Uma distribuição de probabilidade inicial para os estados não observáveis  $\mathbf{p} = \{p_i\}$ , tal que  $p_i = P(S_0 = i)$ ;
6. Uma matriz com as probabilidades de transição entre estados  $\mathbf{A} = \{a_{ij}\}$ , tal que  $a_{ij} = P(S_t = j \mid S_{t-1} = i)$ ;
7. Uma distribuição de probabilidade de observação  $\mathbf{B} = \{b_j(k)\}$ , tal que  $b_j(k) = P(Y_t = k \mid S_t = j)$ .

Esse modelo pode ser facilmente adaptado para quando a variável observável é contínua e não discreta. Aqui optamos por considerar apenas a variável observável como discreta porque é o caso de nosso interesse na aplicação em dados reais que será mostrada no Capítulo 5.

Tratando-se de mais uma variante de processos Markovianos, na qual trabalhamos com uma cadeia de Markov que é não-observável ao observador, todas as propriedades que valem para cadeias de Markov também valem para modelos Markovianos ocultos. Esta ideia é mencionada por [Espindola \(2009\)](#).

Para conseguir entender claramente a ideia do modelo, vamos voltar no exemplo das urnas que tínhamos discutido anteriormente. Definindo cada um dos elementos no contexto do exemplo das urnas, temos que:

1.  $\varphi = \{1, 2, 3\}$ ;
2.  $\omega = \{P, B\}$ ;
3. Uma variável aleatória  $S$  que assume os valores 1, 2 ou 3 representando a urna atual, em diferentes instantes do tempo;
4. Uma variável aleatória  $Y$  que assume os valores 1 se a bola retirada for preta ou 0 se a bola retirada for branca, em diferentes instantes de tempo;
5. A probabilidade inicial para cada urna:  $p_1 = P(S_0 = 1)$ ;  $p_2 = P(S_0 = 2)$  e  $p_3 = P(S_0 = 3)$  sendo que  $p_1 + p_2 + p_3 = 1$ .

6. A matriz com as probabilidades de transição entre urnas  $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$  tal que  $a_{ij} = P(S_t = j \mid S_{t-1} = i)$ , ou a probabilidade de estar na  $j$ -ésima urna no instante do tempo  $t$ , dado que estivemos na  $i$ -ésima urna no instante de tempo  $t - 1$ , sendo que  $a_{i1} + a_{i2} + a_{i3} = 1$ .

7. A distribuição de probabilidade de observar uma bola preta ou uma bola branca  $\mathbf{B} = \{b_j(1), b_j(0)\}$ , tal que  $b_j(k) = P(Y_t = k | S_t = j)$ , é a probabilidade de observar uma bola preta ou branca, respetivamente, no instante de tempo  $t$  dado que estivemos na  $j$ -ésima urna no instante de tempo  $t$ , sendo que  $b_j(1) + b_j(0) = 1$ .

Tendo definido os elementos do Modelo Markoviano Oculto representado no exemplo inicial, podemos observar claramente o funcionamento do mesmo na Figura 4.

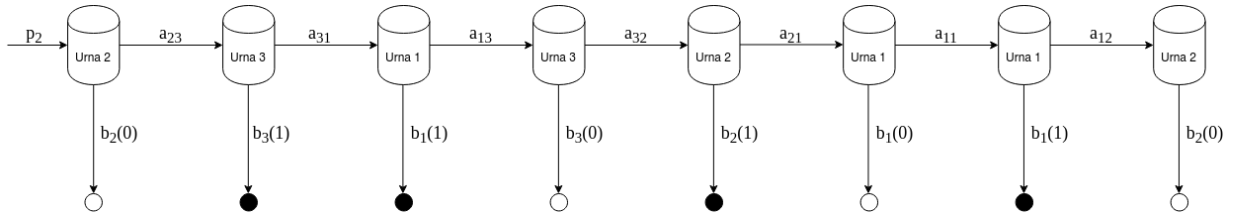


Figura 4 – Exemplo de retirada de bolas com os elementos do modelo Markoviano oculto homogêneo.

Para ter uma compreensão mais clara do cálculo da probabilidade de observar uma sequência de valores observáveis quando conhecemos a sequência de estados ocultos, vamos utilizar a Figura 3 como referência para construir um exemplo. As equações utilizadas neste exemplo serão explicadas com detalhe nas próximas seções. Vamos supor que a matriz de transição que gerou a sequência não observável na Figura 3 é a seguinte:

$$\mathbf{A} = \begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.20 & 0.60 & 0.20 \\ 0.24 & 0.24 & 0.52 \end{bmatrix}. \quad (2.1)$$

Também temos que as probabilidades de observar uma bola preta ou uma bola branca são dadas nas Equações (2.2):

$$\begin{aligned} b_1(1) &= P(Y_t = 1 | S_t = 1) = 0.20 \\ b_1(0) &= P(Y_t = 0 | S_t = 1) = 0.80 \\ b_2(1) &= P(Y_t = 1 | S_t = 2) = 0.60 \\ b_2(0) &= P(Y_t = 0 | S_t = 2) = 0.40 \\ b_3(1) &= P(Y_t = 1 | S_t = 3) = 0.40 \\ b_3(0) &= P(Y_t = 0 | S_t = 3) = 0.60 \end{aligned} \quad (2.2)$$

Além disto, temos que o comprimento da sequência de urnas é de  $T = 8$ . Também podemos assumir que no primeiro instante de tempo, as 3 urnas tem a mesma probabilidade de



ser escolhidas. Como isso, temos o cálculo da probabilidade de observar aquela sequência de bolas, dada a sequência de urnas escolhidas:

$$\begin{aligned}
 P(\mathbf{Y} | \mathbf{S}, \eta) &= \prod_{t=1}^T P(Y_t | S_t, \eta) \\
 &= P(Y_1 | S_1)P(Y_2 | S_2)P(Y_3 | S_3)P(Y_4 | S_4)P(Y_5 | S_5)P(Y_6 | S_6)P(Y_7 | S_7)P(Y_8 | S_8) \\
 &= P(Y_1 = 0 | S_1 = 2)P(Y_2 = 1 | S_2 = 3)P(Y_3 = 1 | S_3 = 1)P(Y_4 = 0 | S_4 = 3) \\
 &\quad P(Y_5 = 1 | S_5 = 2)P(Y_6 = 0 | S_6 = 1)P(Y_7 = 1 | S_7 = 1)P(Y_8 = 0 | S_8 = 2) \\
 &= (0.40)(0.40)(0.20)(0.60)(0.60)(0.80)(0.20)(0.40) \\
 &= 0.00073728
 \end{aligned} \tag{2.3}$$

Porém, numa situação real a sequência de estados ocultos geralmente é desconhecida. Para estimarmos os parâmetros desconhecidos do modelo a partir dos dados observáveis precisamos afrontar o problema da sequência de estados ocultos desconhecida. Nas próximas seções apresentamos estratégias para fazer o cálculo desta e outras quantidades de interesse para o HMM.

## 2.4 Estimações de interesse

Rabiner (1989) menciona 3 problemas de interesse para um HMM. Para simplificar a notação temos que  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$ . Esses problemas são:

1. Dado o modelo  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$  e a sequência de variáveis aleatórias observáveis  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ . Como calcular de forma eficiente a probabilidade de essa sequência ser gerada pelo modelo  $\eta$ , ou seja,  $P(\mathbf{Y} | \eta)$ ?
2. Sejam o modelo  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$  e a sequência de valores observados  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ . Dentre as diversas sequências de estados não observáveis que poderiam ter gerado essa sequência de valores observáveis, qual é a mais provável?
3. Como estimamos os parâmetros  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$  do modelo?

Vários autores citam estes problemas como sendo o foco principal do interesse de estimação dos HMM. Zuanetti (2006) define os três problemas da seguinte forma: o primeiro problema é o cálculo da probabilidade de ocorrência de sequência de observações dado que a estrutura e os parâmetros das distribuições envolvidas no modelos são conhecidos. Em outras palavras, queremos encontrar a distribuição marginal de  $\mathbf{Y}$  necessária para a estimação dos parâmetros do modelo de uma maneira computacionalmente viável. O segundo problema é

definido como a identificacão da seqüência de estados ocultos mais provável dada a seqüência de observacões e o modelo. Esse problema se trata da predicão de  $\mathbf{S}$ , muito útil em vários problemas práticos. O terceiro problema é estimacão dos parâmetros envolvidos no modelo. A soluçao destes três problemas será abordada com detalhes no Capítulo 3, mas a seguir apresenta-se uma soluçao para o primeiro problema de interesse. Para facilitar a notaçao, seja:

$$\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B}) \quad (2.4)$$

onde  $\mathbf{p}$  refere-se ao vetor de probabilidades iniciais para a cadeia oculta,  $\mathbf{A}$  refere-se à matriz de probabilidades de transiçao e  $\mathbf{B}$  refere-se à distribuicão de probabilidade das observacões. Com isso, temos que a probabilidade de uma certa seqüência  $\mathbf{S}$  é dada por:

$$\begin{aligned} P(\mathbf{S} | \eta) &= P(S_0, S_1, \dots, S_T | \eta) \\ &= P(S_0 | \eta)P(S_1 | S_0, \eta)P(S_2 | S_1, \eta) \dots P(S_T | S_{T-1}, \eta) \\ &= p_{s_0} a_{s_0 s_1} a_{s_1 s_2} \dots a_{s_{T-1} s_T} \\ &= p_{s_0} \prod_{t=1}^T a_{s_{t-1} s_t}. \end{aligned} \quad (2.5)$$

Segundo [Jelinek \(1997\)](#), podemos assumir que as observacões são condicionalmente independentes dado a seqüência  $\mathbf{S}$ , com isso podemos escrever:

$$\begin{aligned} P(\mathbf{Y} | \mathbf{S}, \eta) &= \prod_{t=1}^T P(Y_t | S_t, \eta) \\ &= \prod_{t=1}^T b_{s_t}(y_t). \end{aligned} \quad (2.6)$$

Agora das Equaçoes (2.5) e (2.6) podemos escrever a probabilidade conjunta de  $\mathbf{Y}$  e  $\mathbf{S}$  como sendo

$$P(\mathbf{Y}, \mathbf{S} | \eta) = P(\mathbf{Y} | \mathbf{S}, \eta)P(\mathbf{S} | \eta). \quad (2.7)$$

Finalmente, fazendo o somatório da Equaçao (2.7) sobre todas as seqüências de estados possíveis, obtemos a distribuicão marginal de  $\mathbf{Y}$  que é dada por:

$$\begin{aligned} P(\mathbf{Y} | \eta) &= \sum_{s_0, s_1, s_2, \dots, s_T} P(\mathbf{Y} | \mathbf{S}, \eta)P(\mathbf{S} | \eta) \\ &= \sum_{s_0, s_1, s_2, \dots, s_T} p_{s_0} \prod_{t=1}^T a_{s_{t-1} s_t} b_{s_t}(y_t). \end{aligned} \quad (2.8)$$

[Espindola \(2009\)](#) explica a Equaçao (2.8) da seguinte forma: considere uma única seqüência de estados  $\mathbf{s}$ . A probabilidade da cadeia de Markov ocupar um dos  $K$  possíveis estados no instante  $t = 0$  é dada por  $p_{s_0}$ . Em  $t = 1$ , o sistema sofre transiçao do estado  $s_0$  para o estado  $s_1$ , gerando a observacão  $y_1$ , de acordo com as probabilidades de transiçao e de observacão,  $a_{s_0 s_1}$  e  $b_{s_1}(y_1)$ , respectivamente. Esse procedimento se repete até  $t = T$ . Tendo calculado a probabilidade para uma dada seqüência  $\mathbf{s}$ , passa-se à próxima, dentre as seqüências restantes.

A soma sobre todas as sequências resulta na probabilidade do modelo gerar a sequência  $\mathbf{y}$  de observações. Observe, que temos  $K^T$  diferentes e possíveis sequências  $\mathbf{s}$  e, desta forma, o cálculo da Equação (2.8) é uma solução rudimentar para o primeiro problema de interesse mesmo para valores não muito grandes de  $K$  e  $T$ .

## 2.5 Modelo Markoviano oculto não-homogêneo

O foco central deste trabalho são os modelos Markovianos ocultos não-homogêneos (NHMM, sigla em inglês) e muitas definições e ideias envolvidas nele são extensões do que foi apresentado sobre o HHMM no capítulo anterior. Similarmente ao modelo Markoviano oculto homogêneo, um modelo Markoviano oculto não-homogêneo pode ser definido como um conjunto finito de estados, cada um associado a uma distribuição de probabilidade. Transições entre os estados são governadas por um conjunto de probabilidades contidas numa matriz de probabilidades de transição, comumente só chamada de matriz de transição, e em cada estado um resultado ou observação pode ser gerada de acordo com a distribuição de probabilidade associada. Somente o resultado, não o estado, é visível a um “observador” e portanto, os estados são “ocultos” ou não-observáveis ao exterior.

A grande maioria dos fenômenos que podem ser modelados por modelos Markovianos ocultos são caracterizados pelo fato de ter probabilidades de transição entre estados ocultos que são dinâmicas no tempo. Por este motivo, a matriz de transição constante no tempo considerada no HHMM pode não ser um elemento adequado para a modelagem de alguns fenômenos. A flexibilidade dos NHMM, por sua vez, permite que a matriz de transição mude em cada instante do tempo em função de outras variáveis que não são constantes no tempo.

Segundo [Holsclaw et al. \(2017\)](#), a homogeneidade do HHMM é um fator que o torna menos flexível na prática e um enfoque para melhorar está limitação é fazer que as probabilidades de transição sejam dependentes de uma série de co-variáveis temporais resultando assim um NHMM. [Lagona, Maruotti e Picone \(2011\)](#) mencionam que os NHMM generalizam a classe de modelos Markovianos homogêneos ao permitir que as probabilidades de transição entre estados variem com o tempo. As covariáveis de entrada não influenciam diretamente o valor observável sendo gerado em cada instante do tempo. [Sarkar, Bhadra e Mallick \(2012\)](#) também apresentam uma situação onde as covariáveis só afetam a dinâmica da sequência de estados ocultos.

## 2.6 Definição do NHMM

A definição do NHMM é similar à do homogêneo, mas a grande diferença entre eles é a matriz de transição mudando em cada instante do tempo. Essas probabilidades de transição são influenciadas por variáveis externas observadas que estão relacionadas ao processo. Dito isso, os elementos do NHMM considerando como variável observável uma variável aleatória discreta,

são os seguintes:

1. Um espaço discreto de estados não observáveis  $\varphi = \{1, 2, \dots, K\}$ ;
2. Um conjunto de valores observáveis  $\omega = \{1, 2, \dots, M\}$ ;
3. Uma variável aleatória  $S$  que assume os valores do espaço de estados ocultos  $\varphi$ , em diferentes instantes do tempo;
4. Uma variável aleatória  $Y$  que assume os valores do conjunto de observáveis  $\omega$ , em diferentes instantes de tempo;
5. Uma distribuição de probabilidade inicial para os estados não observáveis  $\mathbf{p} = \{p_i\}$ , tal que  $p_i = P(S_0 = i)$ ;
6. Um vetor linha  $\mathbf{X}_t = (X_{t1}, \dots, X_{tD})$ , para  $t = 1, \dots, T$  que representam os valores das  $D$  co-variáveis observadas que influenciam as probabilidades de transição entre o estado oculto no tempo  $t - 1$  e  $t$ ;

7. Uma matriz de coeficientes  $\boldsymbol{\beta}$  que tem entradas  $\beta_{ij} = \begin{pmatrix} \beta_{ij1} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{ijD} \end{pmatrix}$  que são vetores de  $D$

coeficientes, cada um associado a uma covariável observada, sendo que  $\beta_{ij1}$  pode ser considerado o intercepto, acompanhado da covariável  $X_{t1} = 1$  para todo  $t$ ;

8. Uma distribuição de probabilidade de transição entre estados  $\mathbf{A}_t = \{a_{ijt}\}$ , tal que  $a_{ijt} = h(\mathbf{X}_t \boldsymbol{\beta}_{ij})$  onde  $h(\cdot)$  é tal que  $0 \leq h(\cdot) \leq 1$  e  $\sum_{j=1}^K h(\mathbf{X}_t \boldsymbol{\beta}_{ij}) = 1$ ;
9. Uma distribuição de probabilidade para as observações associada a cada estado não-observável  $\mathbf{B} = \{b_j(k)\}$ , tal que  $b_j(k) = P(Y_t = k | S_t = j)$ .

Os problemas de interesse associados ao NHMM são os mesmos associados ao HHMM já citados na Seção 2.4 e, por isso, não serão discutidos aqui novamente. Como dito, uma cadeia de Markov é caracterizada pela probabilidade do estado inicial da cadeia  $\mathbf{p} = \{p_i\}$ , tal que  $p_i = P(S_0 = i)$  e pelas probabilidades de transição  $P(S_t = j | S_{t-1} = i) = a_{ijt}$ . Usando probabilidade condicional, tem-se:

$$\begin{aligned}
 P(\mathbf{S} | \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}) &= P(S_0, S_1, \dots, S_T | \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}) \\
 &= P(S_0 | \mathbf{p}) P(S_1 | S_0, \boldsymbol{\beta}, \mathbf{X}) P(S_2 | S_1, \boldsymbol{\beta}, \mathbf{X}) \dots P(S_T | S_{T-1}, \boldsymbol{\beta}, \mathbf{X}) \\
 &= p_{s_0} \cdot a_{s_0 s_1 1} \cdot a_{s_1 s_2 2} \cdot \dots \cdot a_{s_T s_{T-1} T} \\
 &= p_{s_0} \prod_{t=1}^T a_{s_{t-1} s_t t}.
 \end{aligned} \tag{2.9}$$

Assumindo que as observações são condicionalmente independentes entre si dado a sequência  $\mathbf{S}$ , podemos escrever:

$$\begin{aligned} P(\mathbf{Y} | \mathbf{S}, \mathbf{B}) &= \prod_{t=1}^T P(Y_t | S_t, \mathbf{B}) \\ &= \prod_{t=1}^T b_{s_t}(y_t). \end{aligned} \quad (2.10)$$

Dessas equações, podemos escrever a probabilidade conjunta de  $\mathbf{Y}$  e  $\mathbf{S}$

$$P(\mathbf{Y}, \mathbf{S} | \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}, \mathbf{B}) = P(\mathbf{Y} | \mathbf{S}, \mathbf{B})P(\mathbf{S} | \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}). \quad (2.11)$$

Agora, fazendo o somatório da Equação (2.11) sobre todas as sequências de estados  $\mathbf{s} = (s_0, s_1, s_2, \dots, s_T)$ , tem-se que:

$$\begin{aligned} P(\mathbf{Y} | \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}, \mathbf{B}) &= \sum_{s_0, s_1, s_2, \dots, s_T} P(\mathbf{Y} | \mathbf{S}, \mathbf{B})P(\mathbf{S} | \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}) \\ &= \sum_{s_0, s_1, s_2, \dots, s_T} p_{s_0} \prod_{t=1}^T a_{s_{t-1}s_t} b_{s_t}(y_t), \end{aligned} \quad (2.12)$$

que é a distribuição marginal de  $\mathbf{Y}$ .

### 2.6.1 Matriz de transição não-homogênea

Como é a estrutura da matriz de transição que muda entre o modelo HHMM e NHMM, essa seção se dedica a oferecer uma percepção clara da sua estrutura, do seu funcionamento e o papel que ela tem nas transições entre estados ocultos. Para tal propósito e sem perder a generalidade, consideremos um modelo Markoviano oculto não homogêneo com dois estados ocultos. Para dito modelo, a matriz de coeficientes de transição seria dada pela seguinte expressão:

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{11} & \boldsymbol{\beta}_{12} \\ \boldsymbol{\beta}_{21} & \boldsymbol{\beta}_{22} \end{bmatrix}, \quad (2.13)$$

onde cada elemento da matriz de transição é um vetor com  $D$  posições. Se considerarmos apenas  $D = 2$  covariáveis, uma delas associada ao intercepto, cada elemento tem 2 posições, no caso  $\boldsymbol{\beta}_{21}$  é dado pela seguinte expressão:

$$\boldsymbol{\beta}_{21} = \begin{pmatrix} \beta_{211} \\ \beta_{212} \end{pmatrix}, \quad (2.14)$$

onde  $\beta_{211}$  é o intercepto e  $\beta_{212}$  multiplica a covariável  $X$  no tempo  $t$ . Então as entradas no tempo  $t$  são dadas pela seguinte expressão:

$$\mathbf{X}_t = \begin{pmatrix} 1 & x_t \end{pmatrix}. \quad (2.15)$$

A probabilidade de transição entre o estado  $i$  e  $j$  é então calculada pela função de ligação  $h(\mathbf{X}_t \boldsymbol{\beta}_{ij})$  descrita no elemento número 8 da Seção 2.6.

Agora consideremos que a função de ligação para o modelo é a função *SoftMax*, definida por Gao e Pavel (2017). A escolha da função de ligação será discutida com maior detalhe na Seção 4.1. A função *SoftMax* é dada pela seguinte expressão:

$$a_{ijt} = \frac{\exp(\mathbf{X}_t \boldsymbol{\beta}_{ij})}{\sum_{l=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{il})}, \quad 1 \leq i, j \leq K \quad (2.16)$$

lembrando que  $a_{ijt}$  refere-se a transição do estado  $i$  ao estado  $j$  no instante de tempo  $t$ .

Para ter uma perspectiva clara do funcionamento da matriz de transição não-homogênea, agora vamos supor que a cadeia oculta do modelo se encontra no estado oculto 2, no instante de tempo  $t$ . Pelo fato de ser um modelo com dois estados ocultos, então só existem duas possíveis transições:

1. Sair do estado 2 no tempo  $t$  e ir para o estado 1 no tempo  $t + 1$ ;
2. continuar no estado 2 no tempo  $t + 1$ .

Para saber quais são as probabilidades de ir para o estado 1, ou continuar no estado 2 no instante de tempo  $t + 1$ , então utilizamos os elementos previamente mencionados nesta seção para calcular as quantidades  $a_{21t}$  e  $a_{22t}$ .

Com isso temos que  $a_{21t}$  é dado por:

$$a_{21t} = \frac{\exp(\mathbf{X}_t \boldsymbol{\beta}_{21})}{\sum_{j=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{2j})} = \frac{\exp(\beta_{211} + \beta_{212} \cdot x_t)}{\exp(\beta_{211} + \beta_{212} \cdot x_t) + \exp(\beta_{221} + \beta_{222} \cdot x_t)}. \quad (2.17)$$

Além disso, temos que  $a_{22t}$  é dado por:

$$a_{22t} = \frac{\exp(\mathbf{X}_t \boldsymbol{\beta}_{22})}{\sum_{j=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{2j})} = \frac{\exp(\beta_{221} + \beta_{222} \cdot x_t)}{\exp(\beta_{211} + \beta_{212} \cdot x_t) + \exp(\beta_{221} + \beta_{222} \cdot x_t)}. \quad (2.18)$$

No caso dessa ilustração, podemos afirmar que  $S_t \mid S_{t-1} = 2$  tem uma distribuição *Discreta* com duas classes cujas probabilidades são, respectivamente dadas por  $a_{21t}$  e  $a_{22t}$ , ou:

$$S_t \mid S_{t-1} = 2 \sim \text{Discreta}(a_{21t}, a_{22t}), \quad a_{21t} + a_{22t} = 1. \quad (2.19)$$

Este processo acontece em cada instante do tempo para calcularmos a probabilidade de cada estado oculto ter acontecido.



---

## PROCESSOS DE ESTIMAÇÃO

---

Segundo [Espindola \(2009\)](#), a modelagem de um sistema físico ou realidade qualquer é uma versão bastante simplificada da realidade propriamente dita e, dessa forma, não há modelo absoluto, existem apenas modelos mais ou menos adequados para um dado sistema. Dessa maneira, poderíamos tratar a modelagem como tendo duas etapas: a definição de um modelo que seja adequado para a situação em questão e seus parâmetros e o ajuste do modelo pela resolução de uma série de problemas de interesse. Existem vários enfoques que tem sido usados para realizar inferência sobre os HMM, mas os interesses inferenciais relacionados aos HMMs independem do enfoque sendo utilizado para a realização da mesma. Neste trabalho são explorados dois métodos de estimação. O método frequentista, para o qual apresentaremos dois algoritmos, o algoritmo EM e o algoritmo EM Estocástico, e o método de estimação Bayesiano.

Neste trabalho apresentamos e comparamos 3 algoritmos de estimação diferentes para a resolução dos "*Problemas Canônicos*" do HMM como são chamados por [Espindola \(2009\)](#). A performance desses algoritmos de estimação é comparada a traves do erro quadrático médio e o viés. Além disso, alguns critérios de informação como o DIC e o BIC são utilizados para seleção de modelos. Os algoritmos de estimação são:

- Algoritmo *Expectation-Maximization* (EM);
- Algoritmo EM Estocástico; e
- Algoritmo Bayesiano com número de estados ocultos  $K$  fixo.

### 3.1 Algoritmo *Expectation-Maximization* Tradicional

Nessa seção vamos descrever o algoritmo EM mais tradicional para a estimação de um HHMM, cuja a extensão para o NHMM é muito similar e é melhor detalhada no [Visser e Speekenbrink \(2010\)](#). Sejam os parâmetros do modelo e a sequência de variáveis aleatórias



observáveis respectivamente denotados por:

$$\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B}), \quad (3.1)$$

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_{T-1}, Y_T). \quad (3.2)$$

O nosso primeiro interesse é calcular  $P(\mathbf{Y} | \eta)$ , a probabilidade de gerar a sequência  $\mathbf{Y}$  a partir desse modelo. A forma mais direta de calcular essa probabilidade envolve a identificação de cada sequência de estados  $\mathbf{S}$  que possa gerar  $\mathbf{Y}$  como descrito na Equação (2.12). Espindola (2009) chama esta estratégia da resolução à “força bruta” e comenta que por consequência esta tende a ser onerosa, pois dispende mais tempo e poder computacional. Para simplificar os cálculos, consideramos que cada transição entre estados  $s_{t-1}$  e  $s_t$  gera um valor observável  $Y_t$ . Também suponha que  $\mathbf{Y}$  tenha sido gerada pela seguinte sequência de estados não observáveis

$$\mathbf{s} = (s_0, s_1, \dots, s_{T-1}, s_T), \quad (3.3)$$

na qual o índice numérico é um inteiro,  $0 \leq t \leq T$ , que indica um instante no tempo. Assim,  $s_0$  significa o estado da cadeia de Markov no instante  $t = 0$ , ou simplesmente o estado inicial.

### 3.1.1 Algoritmo Forward-Backward

#### 3.1.1.1 Procedimento Forward

Considere a variável *forward*, definida como

$$\alpha_t(i) = P(Y_1, Y_2, \dots, Y_t, S_t = i | \eta), \quad (3.4)$$

isto é, a probabilidade da observação parcial da sequência de variáveis observáveis, de  $Y_1$  até  $Y_t$ , conjunta com a probabilidade de ocupação do estado  $i$  da cadeia de Markov oculta no instante  $t$ . Como estamos trabalhando com sequências em função do tempo, podemos dizer que trabalhamos com conjuntos ordenados de eventos, o que nos permite assumir, por indução, que  $\alpha_t(i)$  vale para qualquer instante de tempo dentro dos limites do problema,  $0 \leq t \leq T$ . Assim, resolvemos o Problema 1 pela aplicação do seguinte procedimento:

1.) Inicialização

$$\alpha_0(i) = p_i, \quad 1 \leq i \leq K \quad (3.5)$$

2.) Indução:

$$\alpha_{t+1}(j) = \left( \sum_{i=1}^K \alpha_t(i) a_{ij} \right) b_j(y_{t+1}) \quad 1 \leq j \leq K \quad e \quad 0 \leq t \leq T-1 \quad (3.6)$$

3.) Finalização:

$$P(\mathbf{Y} | \eta) = \sum_{i=1}^K \alpha_T(i). \quad (3.7)$$

A Figura 5 ilustra a situação do passo 2.

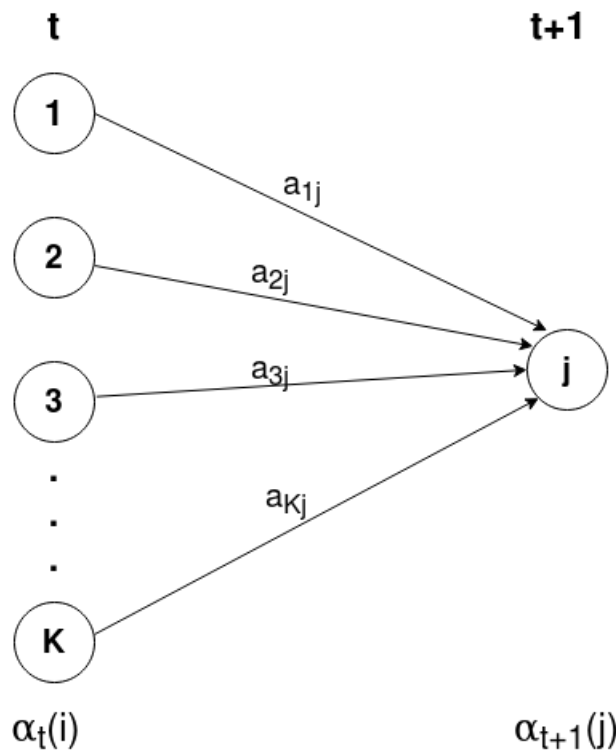


Figura 5 – Ilustração da parte *forward* do algoritmo *forward-backward*.

No procedimento anterior temos que o termo  $\alpha_t(j)$  é a probabilidade conjunta da:

- observação parcial  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_t)$ ; e
- ocupação do estado  $S_t = j$ .

Ao multiplicar  $a_{ij}$  por  $b_j(y_{t+1})$  estamos calculando a probabilidade conjunta de:

- transição do estado  $S_t = i$  para o estado  $S_{t+1} = j$ ; e
- emissão da variável  $Y_{t+1}$  em consequência da transição  $a_{ij}$ .

Multiplicando, então, os termos  $\alpha_t(i)$ ,  $a_{ij}$  e  $b_j(y_{t+1})$ , e somando sobre todos os estados  $1 \leq i \leq K$ , obtemos a probabilidade conjunta de:

- observação parcial  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_t)$ ;
- ocupação do estado  $S_{t+1} = j$ , qualquer que tenha sido o estado no instante anterior; e
- emissão do observável  $Y_{t+1}$  em consequência de todas as transições com destino a  $S_{t+1} = j$ , que nada mais é do que o valor de  $\alpha_{t+1}(j)$ , Equação (3.6).

Para finalizar o procedimento, faz-se o somatório de  $\alpha_T(i)$  sobre todos os estados  $1 \leq i \leq K$ . Isso faz todo o sentido quando analisamos a definição da variável *forward* por ocasião do último

instante de observação  $T$ :

$$\alpha_T(i) = P(Y_1, Y_2, \dots, Y_T, S_T = i | \eta). \quad (3.8)$$

Essa equação representa a probabilidade conjunta da sequência completa de observação com a probabilidade de ocupar o estado  $i$  no instante  $T$ . Dessa forma, ao somarmos a Equação (3.8) sobre todos os estados, obtemos a probabilidade de que um dado modelo  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$  gere a sequência de observáveis  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)$ , ou seja, obtemos a Equação (3.7).

### 3.1.1.2 Procedimento Backward

Considere agora a variável *backward*, definida como

$$\beta_t(i) = P(Y_{t+1}, Y_{t+2}, \dots, Y_T, S_t = i | \eta), \quad (3.9)$$

ou seja, a probabilidade conjunta da cadeia de Markov estar no estado  $i$  no instante  $t$  com a probabilidade da observação parcial,  $Y_{t+1}, Y_{t+2}, \dots, Y_T$ , nos instantes subsequentes a  $t$ . A parte *backward* do procedimento *forward-backward* é muito semelhante ao que acabamos de ver para a parte *forward*. Logo, por analogia, segue:

1. Inicialização:

$$\beta_T(i) = 1, \quad 1 \leq i \leq K \quad (3.10)$$

2. Indução:

$$\beta_t(j) = \left( \sum_{i=1}^K a_{ij} b_j(y_{t+1}) \right) \beta_{t+1}(i), \quad t = T-1, T-2, \dots, 0; 1 \leq j \leq K. \quad (3.11)$$

A Figura 6 ilustra a situação do passo 2.

Se assumirmos que o modelo tem um determinado estado inicial,  $i$ , com probabilidade  $P(S_0 = i)$ , podemos dizer que o que buscamos calcular é justamente  $\beta_0$ , que é a probabilidade da sequência completa de observação, dado que o estado inicial foi  $S_0 = i$ .

## 3.1.2 Algoritmo de Viterbi

Nessa seção vamos apresentar um dos métodos para resolver o segundo problema, identificar a sequência de estados ocultos mais provável que poderia ter gerado a sequência de valores observáveis. [Espindola \(2009\)](#) diz que o algoritmo de Viterbi, proposto em 1967 e desde então usado em uma grande gama de aplicações, é uma solução ótima recursiva para o problema de estimar a sequência de estados de um processo Markoviano de estado finito e tempo discreto.

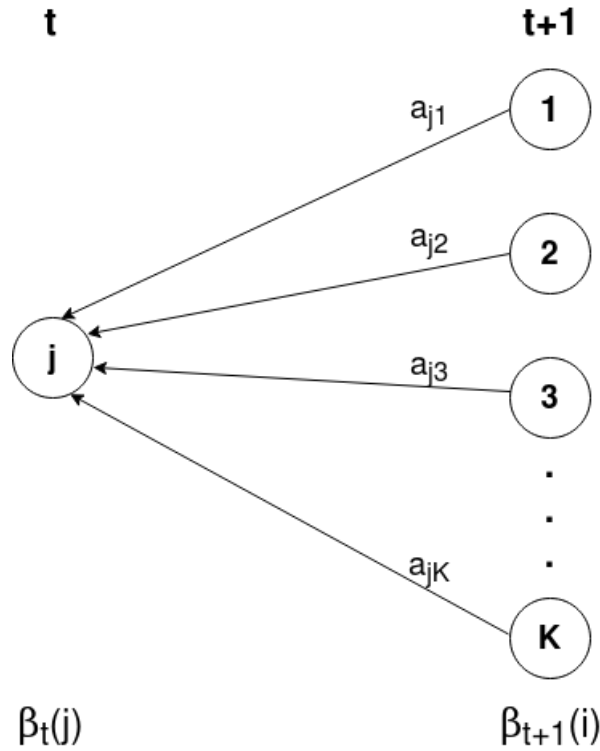


Figura 6 – Ilustração da parte *backward* do algoritmo *forward-backward*.

Ainda, identificar a sequência de estados mais provável,  $\mathbf{s} = (s_1, s_2, \dots, s_T)$ , dada a sequência de observação  $\mathbf{y} = (y_1, y_2, \dots, y_T)$ , ou mais formalmente, maximizar  $P(\mathbf{S}|\mathbf{Y}, \eta)$  é equivalente a maximizar  $P(\mathbf{S}, \mathbf{Y}|\eta)$ , pois ambas as operações de maximização vão devolver a sequência de estados mais provável.

Seja, então, a seguinte definição:

$$\delta_t(j) = \max_{s_1, s_2, \dots, s_{t-1}} P(S_1, S_2, \dots, S_t = j, Y_1, \dots, Y_t | S_0 = i, \eta) \quad 1 \leq j \leq K, \quad (3.12)$$

ou seja,  $\delta_t(j)$  guarda a probabilidade do caminho (ou sequência de estados) mais provável que leva ao estado  $j$  no instante  $t$ , gerando os primeiros  $t$  valores observáveis. Por indução, temos:

$$\delta_{t+1}(i) = \max_i (\delta_t(i) a_{ij} b_j(y_{t+1})). \quad (3.13)$$

Para guardar a sequência de estados, usamos um vetor auxiliar  $\Psi_t(k)$  que guarda em cada posição  $t$  o índice do estado que maximiza a sequência até o estado  $S_t = k$ . O procedimento completo é mostrado a seguir:

#### 1. Inicialização

$$\delta_1(j) = a_{ij} b_j(y_1), 1 \leq j \leq K \quad (3.14)$$

$$\psi_1(j) = 0; \quad (3.15)$$

## 2. Indução

$$\delta_t(k) = \max_{1 \leq j \leq K} [\delta_{t-1}(j) a_{ij} b_j(y_t)], \quad 2 \leq t \leq T$$

$$1 \leq k \leq K \quad (3.16)$$

$$\psi_t(k) = \operatorname{argmax}_{1 \leq j \leq K} [\delta_{t-1}(j) a_{ij}], \quad 2 \leq t \leq T$$

$$1 \leq k \leq K \quad (3.17)$$

## 3. Finalização

$$P^* = \max_{1 \leq j \leq K} [\delta_T(j)]$$

$$q_T^* = \operatorname{argmax}_{1 \leq j \leq K} [\delta_T(j)]; \quad (3.18)$$

## 4. Recriação do caminho (sequência de estados):

$$q_t^* = \{\psi_{t+1}, \{q_{t+1}^*\}\} \quad t = T-1, T-2, \dots, 1. \quad (3.19)$$

### 3.1.3 Algoritmo de Baum-Welch

Dentre os três problemas de interesse, o terceiro problema é o mais difícil de resolver. Isto é devido a que não existe um método analítico que permita estimar os parâmetros  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$  que maximizam a probabilidade de um modelo gerar a sequência de observáveis  $P(\mathbf{Y} | \eta)$ , ou seja, encontrar

$$\hat{\eta} = \operatorname{argmax}_{\eta} P(\mathbf{Y} | \eta). \quad (3.20)$$

No entanto existe um algoritmo capaz de maximizar a probabilidade local. Esse algoritmo, de acordo com [Jelinek \(1997\)](#) é citado na literatura sob diferentes nomes, tais como *algoritmo de Baum*, *Baum-Welch* ou algoritmo *forward-backward*. O critério de parada para este algoritmo é a diferença entre o valor da função de verossimilhança no passo atual e no passo seguinte. Quando essa diferença atinge um valor menor do que uma tolerância pré-estabelecida então o algoritmo chega ao fim. Os problemas que o algoritmo Baum-Welch apresenta são, geralmente, sensibilidade aos valores iniciais do algoritmo e convergência lenta. Tendo dito isto, considere a seguinte definição:

$$\xi_t(i, j) = P(S_t = i, S_{t+1} = j | \mathbf{y}, \eta), \quad (3.21)$$

ou seja,  $\xi_t(i, j)$  é a probabilidade conjunta de estar no estado  $i$  no instante  $t$  e no estado  $j$  no instante  $t+1$ , dado o modelo inicial  $\eta = (\mathbf{p}, \mathbf{A}, \mathbf{B})$  e a sequência observada  $\mathbf{y}$ . Podemos expressar

essa variável em função das variáveis *forward* e *backward* que foram definidas no Algoritmo *Forward-Backward*, tomando a seguinte forma:

$$\begin{aligned}\xi_t(i, j) &= P(S_t = i, S_{t+1} = j | \mathbf{Y}, \eta) = \frac{P(S_t = i, S_{t+1} = j, \mathbf{Y} | \eta)}{P(\mathbf{Y} | \eta)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}{P(\mathbf{Y} | \eta)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^K \sum_{j=1}^K \alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}.\end{aligned}\quad (3.22)$$

Agora fazendo o somatório da Equação (3.22) sobre o índice  $j, 1 \leq j \leq K$ :

$$\begin{aligned}\sum_{j=1}^K \xi_t(i, j) &= \sum_{j=1}^K \frac{\alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}{P(\mathbf{Y} | \eta)} \\ &= \frac{\alpha_t(i) \left[ \sum_{j=1}^K a_{ij} b_j(y_{t+1}) \beta_{t+1}(j) \right]}{P(\mathbf{Y} | \eta)} \\ &= \frac{\alpha_t(i) \beta_t(i)}{P(\mathbf{Y} | \eta)}.\end{aligned}\quad (3.23)$$

A parte entre colchetes na Equação (3.23) é exatamente a Equação (3.13) relativa a variável *backward* no instante  $t$ . Logo, a Equação (3.23) se iguala à Equação (3.21) apresentada durante a resolução do segundo problema. Agora definimos  $\gamma_t(i)$  em função das variáveis *forward* e *backward*. Portanto,

$$\gamma_t(i) = \sum_{j=1}^K \xi_t(i, j). \quad (3.24)$$

Fazendo o somatório de  $\gamma_t(i)$  sobre o instante de tempo de observação  $t$ , obtemos a estimativa da quantidade de vezes que a cadeia transitou no estado  $i$  nesse período. Para saber o número de transições que partem de  $i$ , só precisaríamos estender o somatório até chegar no instante de tempo  $T - 1$ . Ao fazer o somatório de  $\xi_t(i, j)$  até  $T - 1$ , obtemos uma estimativa para a quantidade de vezes que aconteceram transições entre os estados  $S_{t-1} = i$  e  $S_t = j$ . Formalmente:

$$\sum_{t=0}^{T-1} \gamma_t(i) = \text{número esperado de transições a partir do estado } i, \text{ e} \quad (3.25)$$

$$\sum_{t=0}^{T-1} \xi_t(i, j) = \text{número esperado de transições do estado } i \text{ para o estado } j. \quad (3.26)$$

Usando essas fórmulas, podemos usar o seguinte método para estimar os parâmetros de um HHMM:

$$\begin{aligned}
 \hat{p}_i &= \text{número esperado de vezes no estado } (S_0 = i) = \gamma_1(i) \\
 \hat{a}_{ij} &= \frac{\text{número esperado de transições do estado } i \text{ para o estado } j}{\text{número esperado de transições a partir do estado } i} \\
 &= \frac{\sum_{t=0}^{T-1} \xi_t(i, j)}{\sum_{t=0}^{T-1} \gamma_t(i)} e; \\
 \hat{b}_{ij}(k) &= \frac{\text{número esperado de transições entre os estados } (i, j) \text{ e observações } y_k}{\text{número esperado de transições entre os estados } (i, j)} \\
 &= \frac{\sum_{t=0}^T \gamma_t(j)}{\sum_{t=0}^T \gamma_t(j)} \Big|_{s.t. Y_t=k}.
 \end{aligned} \tag{3.27}$$

Se definirmos o modelo atual com parâmetro  $\eta^* = (\mathbf{p}^*, \mathbf{A}^*, \mathbf{B}^*)$  e usarmos esses parâmetros para calcular os parâmetros do modelo atualizado,  $\bar{\eta} = (\bar{\mathbf{p}}, \bar{\mathbf{A}}, \bar{\mathbf{B}})$ , foi provado por Baum e seus colegas que:

1.  $\bar{\eta}$  é mais provável que  $\eta^*$ , pois  $P(\mathbf{Y}|\bar{\eta}) > P(\mathbf{Y}|\eta^*)$ , significando que encontramos um novo modelo com parâmetro  $\bar{\eta}$ , de onde é mais provável que a sequência de observação  $\mathbf{Y}$  tenha sido gerada;
2. A partir de um número de iterações que garante a convergência,  $\eta^* = \bar{\eta}$ , significa que  $\eta^*$  define um ponto crítico da função de probabilidade e, portanto, o modelo com parâmetro  $\bar{\eta}$  é aquele que maximiza a probabilidade da sequência de observações.

Uma extensão do algoritmo EM para modelos Markovianos não-homogêneos está disponível no pacote *depmixS4*. Este pacote foi desenvolvido por [Visser e Speekenbrink \(2010\)](#) e é parte do aplicativo estatístico R desenvolvido por [R Core Team \(2013\)](#).

## 3.2 Algoritmo EM Estocástico

A partir desse momento, consideramos que as variáveis observáveis seguem uma distribuição Binomial porque nossa aplicação em dados reais a ser apresentada no Capítulo 4 considera variáveis dessa distribuição. No entanto, essa distribuição pode ser facilmente alterada para qualquer outra, inclusive para distribuições de variáveis contínuas. Além disto, para o processo de estimação dos parâmetros do NHMM usando o algoritmo EM Estocástico incluímos alguns detalhes específicos relacionados com as variáveis observáveis.

- Assuma que a distribuição de  $Y_t | S_t = j$  é Binomial( $n_t, \theta_j$ );

- Seja  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{K-1}, \theta_K)$  onde  $\theta_k$  é a probabilidade de sucesso relacionada ao  $k$ -ésimo estado oculto e;
- Seja  $\mathbf{n} = (n_1, n_2, \dots, n_T)$  onde  $n_t$  representa o tamanho da amostra relacionada ao instante do tempo  $t$ .

Com isto, função de verossimilhança completa ou aumentada para o modelo NHMM pode ser escrita como:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \mid \mathbf{y}, \mathbf{s}, \mathbf{n}, \mathbf{X}) &= P(\mathbf{Y} = \mathbf{y} \mid \mathbf{S} = \mathbf{s}, \boldsymbol{\theta}, \mathbf{n}) P(\mathbf{S} = \mathbf{s} \mid \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}) \\
&= \prod_{t=1}^T P(Y_t = y_t \mid S_t = s_t, \boldsymbol{\theta}, \mathbf{n}) P(S_1 = s_1 \mid \mathbf{p}) \prod_{t=2}^T P(S_t = s_t \mid S_{t-1} = s_{t-1}, \boldsymbol{\beta}, \mathbf{X}) \\
&= \left\{ \prod_{t=1}^T \binom{y_t}{n_t} \right\} \left\{ \prod_{t=1}^T \prod_{j=1}^K (\theta_j^{y_t} (1 - \theta_j)^{n_t - y_t})^{\mathbb{I}_{s_t}(j)} \right\} \left\{ \prod_{j=1}^K p_j^{\mathbb{I}_{s_1}(j)} \right\} \\
&\quad \times \left\{ \prod_{t=2}^T \prod_{i=1}^K \prod_{j=1}^K a_{ijt}^{\mathbb{I}_{s_{t-1}}(i) \mathbb{I}_{s_t}(j)} \right\} \\
&= \left\{ \prod_{t=1}^T \binom{y_t}{n_t} \right\} \left\{ \prod_{j=1}^K (\theta_j^{\sum_{t:s_t=j} y_t} (1 - \theta_j)^{\sum_{t:s_t=j} (n_t - y_t)}) \right\} \left\{ \prod_{j=1}^K p_j^{\mathbb{I}_{s_1}(j)} \right\} \\
&\quad \times \left\{ \prod_{t=2}^T \prod_{i=1}^K \prod_{j=1}^K \left( \frac{\exp(\mathbf{X}_t \boldsymbol{\beta}_{ij})}{\sum_{j=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{ij})} \right)^{\mathbb{I}_{s_{t-1}}(i) \mathbb{I}_{s_t}(j)} \right\} \\
&= \left\{ \prod_{t=1}^T \binom{y_t}{n_t} \right\} \\
&\quad \times \prod_{j=1}^K \left\{ p_j^{\mathbb{I}_{s_1}(j)} \theta_j^{\sum_{t:s_t=j} y_t} (1 - \theta_j)^{\sum_{t:s_t=j} (n_t - y_t)} \prod_{i=1}^K \prod_{t:t \geq 2, (s_{t-1}, s_t) = (i, j)} \frac{\exp(\mathbf{X}_t \boldsymbol{\beta}_{ij})}{\sum_{j=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{ij})} \right\} \tag{3.28}
\end{aligned}$$

Aplicando o logaritmo na função de verossimilhança, obtém-se:

$$\begin{aligned}
\log \mathcal{L}(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \mid \mathbf{y}, \mathbf{s}, \mathbf{n}, \mathbf{X}) &= \sum_{t=1}^T \log \binom{n_t}{y_t} + \sum_{j=1}^K \{ \mathbb{I}_{s_1}(j) \log(p_j) \\
&\quad + \sum_{t:s_t=j} y_t \log(\theta_j) + \sum_{t:s_t=j} (n_t - y_t) \log(1 - \theta_j) \\
&\quad + \sum_{i=1}^K \sum_{t:t \geq 2, (s_{t-1}, s_t) = (i, j)} \left[ \mathbf{X}_t \boldsymbol{\beta}_{ij} - \log \sum_{j=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{ij}) \right] \}. \tag{3.29}
\end{aligned}$$

### 3.2.1 Estimação dos Parâmetros

Como no caso apresentado na Equação (3.29) existe a restrição  $\sum_{j=1}^K p_j = 1$ , utilizamos os multiplicadores de Lagrange para sua maximização. Então temos que a função de log-



verossimilhança juntamente com o multiplicador de Lagrange é dada por:

$$\begin{aligned}
\ell(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \mid \mathbf{y}, \mathbf{s}, \mathbf{n}, \mathbf{X}) &= \sum_{t=1}^T \log \binom{n_t}{y_t} + \sum_{j=1}^K \{ \mathbb{I}_{s_1}(j) \log(p_j) \\
&+ \sum_{t:s_t=j} y_t \log(\theta_j) + \sum_{t:s_t=j} (n_t - y_t) \log(1 - \theta_j) \\
&+ \sum_{i=1}^K \sum_{(t:t \geq 2, (s_{t-1}, s_t) = (i, j))}^T \left[ \boldsymbol{\beta}_{ij}^T \mathbf{X}_t - \log \sum_{j=1}^T \exp(\boldsymbol{\beta}_{ij}^T \mathbf{X}_t) \right] \} \\
&+ \lambda_0 \left( 1 - \sum_{j=1}^K p_j \right).
\end{aligned} \tag{3.30}$$

Onde  $\lambda_0$  é o multiplicador de Lagrange usado na maximização dessa equação. Derivando a Equação (3.30) em cada um dos parâmetros e igualando a sua derivada a zero, são obtidos os estimadores de máxima verossimilhança. Para obter o estimador para  $\theta_j$  tem-se:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \mid \mathbf{y}, \mathbf{s}, \mathbf{n}, \mathbf{X})}{\partial \theta_j} &= 0 \\
\sum_{t:s_t=j} \frac{y_t}{\theta_j} - \sum_{t:s_t=j} \frac{n_t - y_t}{1 - \theta} &= 0. \\
\hat{\theta} &= \frac{\sum_{t:s_t=j} y_t}{\sum_{t:s_t=j} n_t}
\end{aligned} \tag{3.31}$$

Agora, para obter o estimador para o parâmetro  $\mathbf{p}$  tem-se que:

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \mid \mathbf{y}, \mathbf{s}, \mathbf{n}, \mathbf{X})}{\partial p_j} &= 0 \\
\frac{\mathbb{I}_{s_1}(j)}{p_j} - \lambda_0 &= 0
\end{aligned} \tag{3.32}$$

Como na derivada em termos de  $p_j$  existe a constante  $\lambda_0$ , é necessário encontrar seu valor para a obtenção do estimador. Igualando a derivada a zero tem-se:

$$\hat{p}_j = \frac{\mathbb{I}_{s_1}(j)}{\lambda_0}. \tag{3.33}$$

Além disso, é sabido que

$$p_1 + p_2 + \dots + p_K = 1 \Rightarrow \sum_{j=1}^K \hat{p}_j = 1. \tag{3.34}$$

Assim,

$$\sum_{j=1}^K \hat{p}_j = 1 \Rightarrow \sum_{j=1}^K \frac{\mathbb{I}_{s_1}(j)}{\lambda_0} = 1 \Rightarrow \sum_{j=1}^K \mathbb{I}_{s_1}(j) = \lambda_0. \tag{3.35}$$

Portanto, o valor de  $p_j$  que maximiza a verossimilhança é  $\hat{p}_j = \mathbb{I}_{s_1}(j)$ .

Além disto os estimadores para  $\beta_{ij}$  são obtidos utilizando métodos numéricos como o método **Nelder-Mead**. Isto pelo fato de que não é possível chegar numa solução analítica para os estimadores desses parâmetros.

Se assim como  $\mathbf{Y}$ , a sequência  $\mathbf{S}$  também fosse observável, as estimativas de máxima verossimilhança para  $\boldsymbol{\theta}$ ,  $\mathbf{p}$  e  $\beta_{ij}$  seriam encontradas apenas usando os dados observados nas equações derivadas acima. Porém, como isso não acontece, recorreremos a um método iterativo que prediz a realização de  $s$  em cada iteração e baseado nessa predição encontra os valores dos parâmetros que maximizam a log-verossimilhança. Esse algoritmo é o EM estocástico, que segundo a literatura apresenta convergência mais rápida do que o EM tradicional em vários cenários, além de simplificar significativamente as contas matemáticas necessárias para sua implementação e fácil interpretação. O algoritmo utilizado na obtenção dos estimadores pode ser descrito da seguinte forma:

1. Inicialmente, atribuem-se valores arbitrários a  $\mathbf{S}$  e  $\beta_{ij}$ ;
2. Com base de  $\mathbf{y}$  observado e  $s$  atribuído, estimam-se os parâmetros  $\hat{\theta}_j$  e  $\hat{p}_j$  por meio dos estimadores de máxima verossimilhança encontrados acima e  $\hat{\beta}_{ij}$  por métodos numéricos;
3. Com base nos valores atualizados dos parâmetros, atualizam-se os valores de  $s$  por meio do Teorema de Bayes:

$$\begin{aligned}
 P(S_t = j | S_{t-1} = i, Y_t = y_t, \mathbf{X}, \mathbf{n}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta}) &= \frac{P(S_t = j, Y_t = y_t | S_{t-1} = i, \mathbf{X}, \mathbf{n}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})}{P(Y_t = y_t | S_{t-1} = i, \mathbf{X}, \mathbf{n}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})} \\
 &= \frac{P(S_t = j, Y_t = y_t | S_{t-1} = i, \mathbf{X}, \mathbf{n}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})}{\sum_{l=1}^K P(S_t = l, Y_t = y_t | S_{t-1} = i, \mathbf{X}, \mathbf{n}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})} \\
 &= \frac{P(S_t = j | S_{t-1} = i, \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}) P(Y_t = y_t | S_t = j, \boldsymbol{\theta}, \mathbf{n})}{\sum_{l=1}^K P(S_t = l | S_{t-1} = i, \mathbf{p}, \boldsymbol{\beta}, \mathbf{X}) P(Y_t = y_t | S_t = l, \boldsymbol{\theta}, \mathbf{n})} \\
 &= \frac{a_{ijt} \binom{n_t}{y_t} \theta_j^{y_t} (1 - \theta_j)^{n_t - y_t}}{\sum_{l=1}^K a_{ilt} \binom{n_t}{y_t} \theta_l^{y_t} (1 - \theta_l)^{n_t - y_t}}
 \end{aligned} \tag{3.36}$$

e,

$$S_t | S_{t-1} = i, Y_t = y_t, \mathbf{X}, \mathbf{n}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \sim \text{Discreta}(a_{i1t}, a_{i2t}, \dots, a_{iKt}); \tag{3.37}$$

4. Repetir os passos 2. e 3. até a obtenção da convergência. O critério de convergência utilizado geralmente é comparar o quanto a log-verossimilhança se altera de uma interação para a outra. A partir do momento em que a diferença é menor que um nível de tolerância pré-fixado, finalizamos o processo iterativo.

Uma desvantagem do algoritmo EM Estocástico é que em todas as iterações do algoritmo precisamos observar transições entre todos os possíveis estados ocultos. No caso que alguma transição não seja observada em uma das iterações, o algoritmo se encontra com um erro e não continua iterando. Para evitar esse tipo de problema, se em alguma iteração uma transição não

fosse observada, escolhemos aleatoriamente 4 pontos na cadeia oculta para atribuímos essa transição específica. O algoritmo EM Estocástico foi implementado completamente usando a ferramenta estatística R. Esta ferramenta foi desenvolvida pelo [R Core Team \(2013\)](#).

### 3.3 Estimação Bayesiana para o número de estados ocultos $K$ fixo

Os algoritmos EM e EM estocástico são ferramentas de estimação muito úteis no momento de realizar inferência sobre os NHMM. Porém, eles podem apresentar problemas como falta de convergência ou convergência para uma solução local. O algoritmo EM também pode apresentar sensibilidade aos valores iniciais sendo utilizados. Por estes motivos, uma alternativa interessante para a estimação dos NHMM pode ser o uso de ferramentas Bayesianas como, por exemplo, os algoritmos Monte Carlo via Cadeias de Markov.

Para a aplicação de algoritmos Bayesianos para fazer inferência sobre o modelo Markoviano oculto não-homogêneo, primeiro estabelecemos algumas definições. Lembrando que os nossos interesses de inferência são estimar:

- Os coeficientes de transição  $\boldsymbol{\beta}$ ;
- As probabilidades do estado inicial  $\mathbf{p}$ ;
- $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{K-1}, \theta_K)$  onde  $\theta_k$  é a probabilidade de sucesso relacionada ao  $k$ -ésimo estado oculto; e
- Predizer a sequência  $\mathbf{S}$  que gerou a sequência observável  $\mathbf{Y}$ .

Com isso, temos que a distribuição *a posteriori* dos parâmetros a serem estimados é dada por

$$P(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta} \mid \mathbf{y}, \mathbf{s}, \mathbf{n}, \mathbf{X}) \propto P(\mathbf{y} \mid \mathbf{s}, \mathbf{n}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})P(\mathbf{s} \mid \mathbf{n}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})P(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta}) \quad (3.38)$$

onde  $P(\mathbf{y} \mid \mathbf{s}, \mathbf{n}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})$  e  $P(\mathbf{s} \mid \mathbf{n}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})$  compõem a função de verossimilhança definida na Equação (3.28) e  $P(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta})$  é a distribuição *a priori* conjunta dos parâmetros. Se assumimos independência condicional entre os parâmetros, então temos que:

$$P(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\beta}) = P(\mathbf{p}) \prod_{j=1}^K P(\theta_j) \prod_{i=1}^K \prod_{j=2}^K \prod_{d=1}^D P(\beta_{ijd}). \quad (3.39)$$

Para o NHMM, as distribuições *a priori* que foram consideradas para os parâmetros são as seguintes:

- $\mathbf{p} \sim \text{Dirichlet}(1, 1, 1)$ ;
- $\theta_j \sim \text{Beta}(1, 1) \quad 1 \leq j \leq K$ ;

- $\beta_{ijd} \sim \text{Normal}(0, 100) \quad 1 \leq i \leq K, 2 \leq j \leq K, 1 \leq d \leq D.$

As distribuições *a posteriori* condicionais completas dos parâmetros são dadas pelas seguintes expressões:

- $\mathbf{p} \mid \dots \sim \text{Dirichlet}(1 + \mathbb{I}_{S_1}(1), \dots, 1 + \mathbb{I}_{S_1}(K));$
- $\theta_j \mid \dots \sim \text{Beta}(1 + \sum_{t:s_t=j} y_t, 1 + \sum_{t:s_t=j} (n_t - y_t));$
- $\beta_{ijd} \mid \dots$  não tem uma forma conhecida.

Além disso, como queremos prever a sequência  $\mathbf{s}$  que gera a sequência observável, temos que  $S_t$  tem uma distribuição *a posteriori* Discreta, dada pela Equação (3.36).

### 3.3.1 Algoritmo de estimação

Utilizando uma metodologia MCMC, o algoritmo de estimação Bayesiano é definido pelos seguintes passos::

1. Atribuir valores iniciais aos parâmetros  $\mathbf{p}, \theta_j, \beta_{ijd}$ ;
2. Atualizar  $\mathbf{p}$  através de passo de *Gibbs-Sampler*;
3. Atualizar  $\theta_j$  através de passo de *Gibbs-Sampler*;
4. Atualizar  $\beta_{ijd}$  através de passo de Metropolis-Hastings usando a sua distribuição *a posteriori* condicional. O pacote RJAGS que é utilizado na implementação desse algoritmo usa o método *Slice Sampling* Neal (2003) para essa simulação. O método *Slice Sampling* não será detalhado nesse trabalho;
5. Atualizar o valor de  $\mathbf{s}$ , onde  $s_t$  é atualizada de uma distribuição *Discreta* com as probabilidades de cada categoria dadas pela Equação (3.36) para  $1 \leq j \leq K$ ;
6. Repetir passos 2, 3, 4 e 5 até atingir o ponto de equilíbrio da cadeia MCMC.

Executamos o algoritmo de estimação Bayesiana considerando um total de 55,000 iterações, do qual descartamos as 5,000 primeiras iterações como *burn-in*. Além disso, consideramos saltos de tamanho 5, ou seja, apenas os valores de uma iteração a cada 5 iterações foram extraídos da sequência original para obter uma subsequência de tamanho 10,000 para realizar as inferências. Como habitualmente é feito na análise Bayesiana, consideramos como estimativas pontuais para os parâmetros de interesse as médias *a posteriori* que são dadas pela média dos valores da subsequência.

O desenvolvimento do processo de estimação com enfoque Bayesiano foi feito utilizando a ferramenta RJAGS. RJAGS foi desenvolvido por Martyn Plummer como uma ferramenta

de modelagem gráfica Bayesiana que tem como objetivo criar uma compatibilidade entre o R e o BUGS clássico. Segundo [Plummer et al. \(2003\)](#), a ferramenta *JAGS* tem duas grandes motivações. A primeira é implementar uma ferramenta que pode ser modificada pelos seus usuários. Este é um lema do *software livre*. A segunda motivação é criar uma plataforma para explorar novas ideias metodológicas e, em particular, ferramentas para análise de modelos de grafos. Segundo [Coro \(2013\)](#), o foco general do *JAGS* é produzir amostras compatíveis com a teoria do amostragem *Gibbs-Sampler*, mas o *JAGS* pode utilizar outros algoritmos de simulação como ser *Adaptative Rejection sampling* ou *Slice sampling*. Se a função condicional completa *a posteriori* do parâmetro a ser amostrado é conhecida o *JAGS* utiliza o *Gibbs-Sampler*, caso contrário ele escolhe o algoritmo mais eficiente disponível para fazer amostragem. O usuário do *JAGS* não tem a opção de escolher o amostrador a ser utilizado.

### 3.4 Troca de Rótulos - *Label switching*

Na estimação de modelos de mistura, cujo HHMM e NHMM são casos especiais que consideram dependência entre as observações, encontramos um problema chamado de *label switching* ou troca de rótulos. Segundo [Papastamoulis \(2014\)](#) a raiz do problema do *label switching* é o fato que a verossimilhança observada é invariante sob permutações dos rótulos dos estados ocultos ou componentes. Ele menciona que na análise Bayesiano, se as distribuições *a priori* não fazem distinção entre os parâmetros, então a distribuição *a posteriori* também é invariante. Então, se uma amostra é simulada da distribuição *a posteriori* e o amostrador MCMC explora varias permutações dos parâmetros, o método comum de calcularmos as estimativas Bayesianas como sendo a média a posteriori da amostra MCMC nos leva a estimativas insensíveis que serão muito similares para todos os estados ocultos.

[Papastamoulis \(2014\)](#) também propõe uma solução para o problema do *label switching*. Ele propõe o algoritmo *Equivalence Classes Representatives*, que é um algoritmo que funciona utilizando uma permutação dos rótulos da sequência  $s$  predita pelo modelo em uma das iterações, escolhida pelo usuário. O algoritmo utiliza esta permutação como “pivô” para comparar com os rótulos da sequência  $s$  em todas as iterações do MCMC. Tendo feita esta comparação, o algoritmo nos dará uma matriz com as permutações dos rotulos que precisam ser usadas para reordenar as estimativas na saída do MCMC. Este processo é feito com o pacote *label.switching* disponível para R. Este pacote foi utilizado para solucionar o problema neste estudo de simulação.

## SIMULAÇÃO

Para estudar e comparar a performance dos processos de estimação mencionados no capítulo anterior, um estudo de simulação foi feito. Algumas condições gerais para este estudo de simulação são mencionadas a seguir:

- Foram escolhidos três comprimentos de cadeia:  $T = 300, 600, 1200$ ;
- O número de estados ocultos foi fixado em  $K = 2$  e  $3$ ;
- Sem perder a generalidade, mas para simplificar a programação e a comparação entre os diferentes algoritmos, apenas uma covariável foi considerada no estudo. Ela foi simulada como  $X \sim N(0, 1)$  em todos os instantes do tempo;
- Para cada situação diferente de simulação,  $M = 30$  réplicas do conjunto de dados foram geradas aleatoriamente;
- A escolha dos valores dos coeficientes de transição e as probabilidades de sucesso serão descritas com detalhe em cada uma das próximas seções.

### 4.1 Escolha da função de ligação

Para o modelo Markoviano oculto não-homogêneo, a função de ligação tem o propósito de gerar probabilidades de transição entre os estados ocultos nos diferentes instantes da cadeia em função de covariáveis que também variam no tempo. Inicialmente a função de ligação que foi escolhida foi a função *Softmax*. A função *Softmax* é dada pela seguinte expressão:

$$a_{ijt} = h(\mathbf{X}_t \boldsymbol{\beta}_{ij}) = \frac{\exp(\mathbf{X}_t \boldsymbol{\beta}_{ij})}{\sum_{l=1}^K \exp(\mathbf{X}_t \boldsymbol{\beta}_{il})}. \quad (4.1)$$

De acordo com [Gao e Pavel \(2017\)](#), na perspectiva da regressão logística multiclasse, a função *Softmax* mapeia um vetor de covariáveis à uma distribuição de probabilidades *a posteriori*. [Wolfe](#)

*et al.* (2017) reforça essa ideia afirmando que numa regressão *Softmax*, uma classificação é feita entre múltiplas classes baseando-se nas probabilidades geradas por ela.

Para o caso do NHMM, a função *Softmax* introduz um problema de não-identificabilidade nos processos de estimação. Souza (2010) define uma família de densidades como sendo identificáveis se valores de parâmetros distintos determinam membros distintos de uma família de densidades. No caso dos modelos NHMM, diferentes e infinitos valores de  $\beta$  podem calcular o mesmo valor para as probabilidades de transição através da função *Softmax*, o que torna sua estimação inviável. Por este motivo, decidiu-se mudar para função de ligação multinomial-logística, *mlogit*. Ela é um caso específico da função *Softmax*, na qual os coeficientes relacionados com a transição para o primeiro estado oculto são fixados com valor igual a zero. A função *mlogit* é dada pela seguinte expressão:

$$a_{ijt} = h(\mathbf{X}_t \beta_{ij}) = \begin{cases} \frac{1}{1 + \sum_{l=2}^K \exp(\mathbf{X}_t \beta_{il})}, & \text{se } j = 1 \\ \frac{\exp(\mathbf{X}_t \beta_{ij})}{1 + \sum_{l=2}^K \exp(\mathbf{X}_t \beta_{il})}, & \text{se } j > 1. \end{cases} \quad (4.2)$$

#### 4.1.1 Escolha dos coeficientes da matriz de transição

Inicialmente, a matriz de coeficientes  $\beta$  foi escolhida aleatoriamente. Esta escolha aleatória ocasionou incerteza em relação aos primeiros resultados que foram obtidos dos processos de estimação. Era difícil determinar se as estimativas que estavam sendo obtidas para os parâmetros eram ruins devido à performance dos algoritmos ou se devia à uma escolha de parâmetros de transição pouco diferentes entre os estados ocultos. Para resolver este problema decidiu-se fazer uma análise gráfica das funções de transição que são geradas pela matriz de coeficientes  $\beta$ , e a de covariáveis de entrada,  $\mathbf{X}$ . Esta análise gráfica será apresentada e descrita para cada uma das situações propostas para o estudo de simulação.

## 4.2 Estudo de simulação

Várias situações foram analisadas para escolher condições nas quais as características dos algoritmos poderiam ser avaliadas e comparadas nesse trabalho. Decidiu-se por duas situações que chamaremos de Situação A e Situação B.

As métricas que escolhemos para a comparação e avaliação dos resultados dos algoritmos são o desvio padrão, viés e erro quadrático médio dos estimadores. Para definir estas métricas, seja  $\theta$  um parâmetro de interesse, e seja  $\hat{\theta}$  o estimador para esse parâmetro.

O viés de  $\hat{\theta}$  é definido na Equação (4.3) como

$$\text{Viés}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta. \quad (4.3)$$

O erro quadrático médio de  $\hat{\theta}$  é definido na Equação (4.4) como

$$\text{EQM}(\hat{\theta}) = [\text{Viés}(\hat{\theta})]^2 + \text{Var}(\hat{\theta}). \quad (4.4)$$

Além de definir as métricas para avaliar a performance dos algoritmos de estimação, utilizamos dois critérios de seleção de modelos, o BIC e o DIC. De acordo com [Bhat e Kumar \(2010\)](#), para um modelos com  $d$  parâmetros e tendo um tamanho de amostra  $n$ , o BIC é definido na Equação (4.5) como

$$\text{BIC} = -2 \cdot \text{LL} + d \cdot \log(n), \quad (4.5)$$

onde LL é a log-verossimilhança do modelo.

Para o cálculo do DIC, [Spiegelhalter et al. \(2014\)](#) define primeiro a *Posterior Mean Deviance*, ou desvio da média *a posteriori*  $D(\theta)$  como sendo  $-2 \cdot \text{LL}$ . Com isso, a definição do DIC é dada na Equação (4.6)

$$\text{DIC} = D(\theta) + \text{Var}(D(\theta)). \quad (4.6)$$

Além das métricas e critérios de informação mencionados, utilizamos outros indicadores como a porcentagem de acertos na predição da sequência  $\mathbf{S}$  e tempo de processamento de cada algoritmo para compara-los. Além das métricas e indicadores de performance mencionados, também construímos um intervalo de confiança de 95% (nas visões frequentistas) para os parâmetros com base nas 30 réplicas que foram geradas para cada algoritmo e o intervalo de credibilidade 95% na visão Bayesiana.

Para os algoritmos EM e EM Estocástico, as seguintes condições iniciais foram usadas nas simulações:

- Os valores iniciais para os parâmetros de transição foram gerados aleatoriamente de uma distribuição  $N(0, 1)$ ;
- Os valores iniciais para a sequência de estados oculta,  $\mathbf{S}$ , é gerada de uma distribuição Discreta com probabilidades iguais de gerar qualquer um dos  $K$  estados ocultos para qualquer instante do tempo;
- O critério de parada para o algoritmo EM Estocástico é a diferença entre a log-verossimilhança entre o passo  $q$  e o passo  $q + 1$ . A tolerância para este critério foi fixada em  $10^{-5}$ .

Para o algoritmo Bayesiano apresentamos algumas condições de simulação que foram utilizadas:

- 55,000 iterações do algoritmo;
- 5,000 iterações de *burn-in*;



- Intervalo de espaçamento entre amostras coletadas igual a 5.

Todas as simulações apresentadas a partir deste ponto são comparadas utilizando as métricas, critérios de informação e indicadores anteriormente descritos. Todos os algoritmos foram replicados em  $M = 30$  amostras. Os intervalos de confiança apresentados em todas as situações são compostos pelo quantil 2.5 e 97.5 das 30 amostras.

#### 4.2.1 Situação A com $K = 2$

A primeira situação estudada com dois estados ocultos, situação A, apresenta condições pouco favoráveis para os algoritmos de estimação. No âmbito dos NHMM, uma situação pode-se considerar pouco favorável quando os parâmetros associados à distribuição das variáveis observáveis são muito próximos entre si. Esse caso indica que, independente do estado oculto no qual a cadeia esteja, as observações são provenientes de uma única e comum distribuição e podem ser analisadas por um modelo mais simples que o NHMM. Além disto uma matriz de coeficientes de transição composta por valores que geram probabilidades de transição similares entre os estados ocultos, independente do valor da covariável, também caracteriza uma situação em que o modelo NHMM é mais complexo (grande número de parâmetros) ao que seria necessário para analisar os dados em questão.

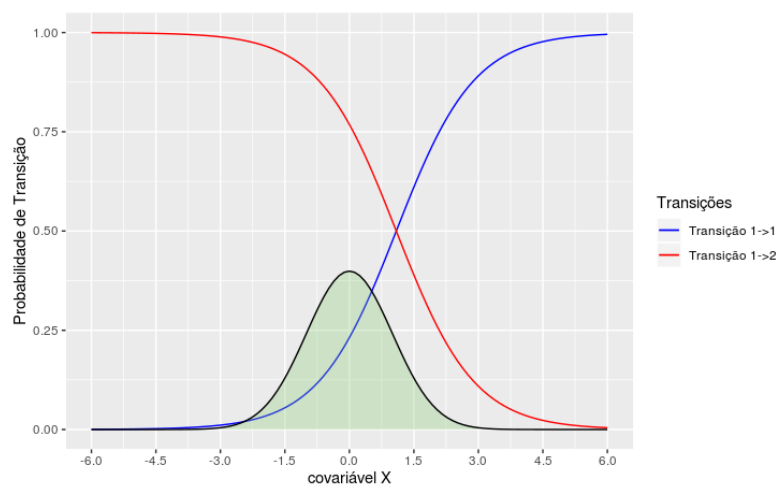
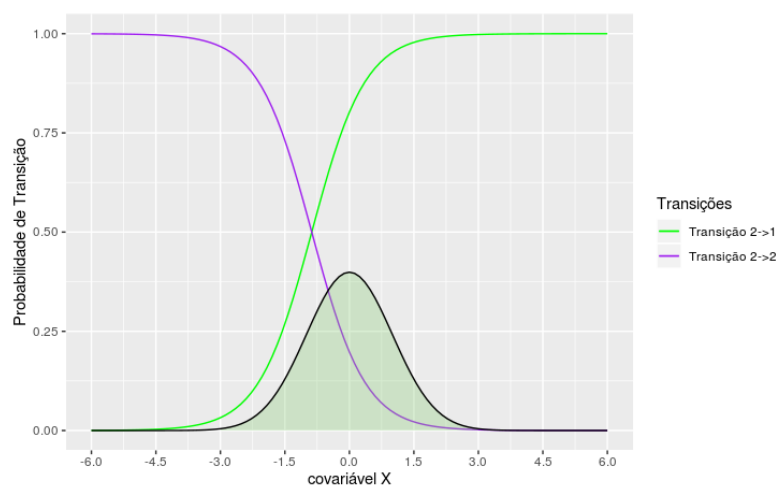
Para a situação A com dois estados ocultos se fixaram as seguintes condições de simulação:

- A distribuição das variáveis observáveis em cada instante do tempo é  $Y_t | S_t = j \sim Binomial(n_t, \theta_j)$ ;
- vetor de probabilidades de sucesso da distribuição Binomial associada a cada estado oculto,  $\theta = (0.45, 0.55)$ ;
- Número de ensaios de Bernoulli em cada instante da cadeia  $n_t = 200$ , para  $t = 1, 2, \dots, T$ .

Observe que apesar do número de 200 ensaios em cada instante de tempo ser relativamente grande, as probabilidades de sucesso fixadas das distribuições binomiais são próximas. Além destas condições e depois de fazer análises gráficas das funções de transição, fixamos a seguinte matriz de coeficientes de transição:

$$\beta = \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 1.2 \\ -1.1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} -1.3 \\ -1.4 \end{pmatrix} \end{bmatrix}. \quad (4.7)$$

As Figuras 7 e 8 mostram as funções de transição obtidas com os valores fixos dos coeficientes de transição para cada valor da covariável X. Nelas observamos que as funções de transição para o estado 1 e para o estado 2 são relativamente diferentes entre os estados ocultos.

Figura 7 – Funções de transição do estado 1 com  $K = 2$  na situação A.Figura 8 – Funções de transição do estado 2 com  $K = 2$  na situação A.Tabela 1 – Resultados do algoritmo EM para a Situação A e  $K = 2$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.027	0.297	0.174	0.118	0.532, 1.670
$\beta_{122}$	-1.1	0.056	0.219	-1.156	1.386	-0.384, 0.317
$\beta_{221}$	-1.3	-0.944	0.236	-0.356	0.182	-1.364, -0.539
$\beta_{222}$	-1.4	-0.001	0.185	-1.400	1.994	-0.367, 0.339
$\theta_1$	0.45	0.449	0.004	0.001	0.001	0.446, 0.457
$\theta_2$	0.55	0.550	0.005	0.001	0.000	0.540, 0.558

Trinta diferentes conjuntos de dados foram simulados nessas mesmas condições e cada um foi analisado pelos três algoritmos de estimação apresentados nesse trabalho. As Tabelas 1, 2 e 3 mostram métricas de performance de estimação dos algoritmos para os três algoritmos estudados na situação A para comprimento de cadeia  $T = 300$ . Os algoritmos apresentam resultados razoavelmente diferentes, com especial atenção ao EM Estocástico que geralmente apresenta os menores vieses e menores erros quadráticos médios observados. Os resultados

Tabela 2 – Resultados do algoritmo EM Estocástico para a Situação A e  $K = 2$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.151	0.544	0.049	0.299	-0.157, 1.989
$\beta_{122}$	-1.1	-0.975	0.380	-0.125	0.160	-1.435, 0.069
$\beta_{221}$	-1.3	-0.558	0.548	-0.742	0.851	-1.666, 0.297
$\beta_{222}$	-1.4	-0.705	0.331	-0.695	0.593	-1.275, -0.051
$\theta_1$	0.45	0.449	0.009	0.001	0.000	0.441, 0.474
$\theta_2$	0.55	0.535	0.023	0.015	0.001	0.471, 0.556

Tabela 3 – Resultados do algoritmo Bayesiano para a Situação A e  $K = 2$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.407	0.420	-0.207	0.219	0.795, 2.345
$\beta_{122}$	-1.1	-1.204	0.444	0.104	0.208	-2.088, -0.598
$\beta_{221}$	-1.3	-4.790	18.342	3.490	348.605	-30.059, -0.931
$\beta_{222}$	-1.4	-3.811	12.595	2.411	164.436	-21.352, -0.719
$\theta_1$	0.45	0.450	0.004	0.000	0.001	0.442, 0.457
$\theta_2$	0.55	0.549	0.004	0.001	0.000	0.542, 0.556

do EM Estocástico apresentam, para a maioria dos parâmetros, vieses menores do que são observados nos outros dois algoritmos estudados. Entre os detalhes que podemos mencionar está o baixo rendimento do procedimento Bayesiano para estimar os coeficientes  $\beta_{221}$  e  $\beta_{222}$ . Isto pode ser devido a que as probabilidades de transição geradas pelos coeficientes de transição escolhidos podem ter gerado uma sequência oculta na qual a permanência no estado dois não ocorreu uma quantidade de vezes suficiente para que o procedimento de estimação Bayesiano resultasse numa estimação mais precisa.

Também pode-se observar estimativas com vieses relativamente altos nos resultados da simulação usando o algoritmo EM. Isto pode ser devido ao algoritmo ter chegado numa solução local para o conjunto de parâmetros em algumas replicas. Apesar das probabilidades de sucesso das distribuições Binomiais serem relativamente parecidas, elas foram bem estimadas por todos os algoritmos, com destaque para o algoritmo Bayesiano, cujos intervalos de credibilidade são bem pequenos.

Tabela 4 – Indicadores para os 3 algoritmos para  $K = 2$  e  $T = 300$ .

Indicadores	EM		EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95	Valor	IC95
Acertos em S	0.97	0.91, 1	0.92	0.56, 1	0.95	0.92, 0.97
BIC	2315.9	2270.5, 2346.4	2515.1	2393.8, 2917.5	2295.9	2036.3, 2391.3
DIC	-	-	-	-	2114	2231.8, 2336.5
Tempo (Mins)	0.09	-	0.07	-	260.76	-

A Tabela 4 mostra outros indicadores de performance dos algoritmos como ser a porcentagem de acertos na sequência de estados ocultos  $S$ , valores calculados do BIC, DIC e tempo de

processamento do algoritmo em minutos, para  $T = 300$  na situação A com 2 estados ocultos. Observamos que em relação ao tempo de processamento, o algoritmo EM Estocástico é mais rápido que os outros dois algoritmos. Neste caso, o algoritmo com o menor BIC foi o algoritmo Bayesiano.

Tabela 5 – Resultados do algoritmo EM para a Situação A e  $K = 2$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	0.944	0.279	0.256	0.144	0.329, 1.312
$\beta_{122}$	-1.1	0.088	0.151	-1.188	1.434	-0.235, 0.309
$\beta_{221}$	-1.3	-0.951	0.263	-0.349	0.191	-1.269, -0.323
$\beta_{222}$	-1.4	0.054	0.171	-1.454	2.143	-0.225, 0.297
$\theta_1$	0.45	0.451	0.010	-0.001	0.000	0.446, 0.467
$\theta_2$	0.55	0.548	0.009	0.002	0.000	0.533, 0.554

Tabela 6 – Resultados do algoritmo EM Estocástico para a Situação A e  $K = 2$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.335	0.447	-0.135	0.218	0.507, 1.965
$\beta_{122}$	-1.1	-1.058	0.269	-0.042	0.074	-1.365, -0.498
$\beta_{221}$	-1.3	-0.438	0.698	-0.862	1.230	-1.776, 0.873
$\beta_{222}$	-1.4	-0.769	0.400	-0.631	0.558	-1.518, -0.188
$\theta_1$	0.45	0.447	0.006	0.003	0.000	0.439, 0.457
$\theta_2$	0.55	0.537	0.015	0.013	0.000	0.504, 0.554

Tabela 7 – Resultados do algoritmo Bayesiano para a Situação A e  $K = 2$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.272	0.316	-0.072	0.105	0.868, 1.972
$\beta_{122}$	-1.1	-1.167	0.298	0.067	0.093	-1.834, -0.863
$\beta_{221}$	-1.3	-1.342	0.208	0.042	0.045	-1.691, -0.985
$\beta_{222}$	-1.4	-1.430	0.293	0.031	0.087	-1.916, -1.007
$\theta_1$	0.45	0.449	0.002	0.001	0.000	0.445, 0.454
$\theta_2$	0.55	0.549	0.002	0.001	0.000	0.544, 0.552

Outro estudo de simulação parecido foi feito para uma sequência maior de observações com  $T = 600$ . As Tabelas 5, 6 e 7 mostram os resultados de simulação para os três algoritmos estudados na situação A para comprimento de cadeia  $T = 600$ . Para este comprimento de cadeia, pode-se observar que o algoritmo EM continua tendo uma performance baixa em termos de vieses e erro quadrático médio, possivelmente devido à identificação de um máximo local como solução para o conjunto de parâmetros sendo estimado. Além disto vemos uma melhora considerável nas estimativas obtidas pelo algoritmo Bayesiano. Em geral a performance do algoritmo Bayesiano neste comprimento de cadeia foi o melhor dentre os três algoritmos. Também podemos observar que a estimativa para  $\beta_{221}$  no algoritmo EM Estocástico teve um erro quadrático médio maior que o das estimativas para os outros parâmetros e o seu intervalo de credibilidade inclui o valor zero.

Tabela 8 – Indicadores para os 3 algoritmos para  $K = 2$  e  $T = 600$ .

Indicadores	EM		EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95	Valor	IC95
Acertos em S	0.955	0.918, 1	0.969	0.884, 1	0.948	0.931, 0.964
BIC	4624.8	4555.5, 4831.9	4921.2	4776.9, 5290.4	4271.4	4156.4, 4302.3
DIC	-	-	-	-	4229.6	4496.2, 4638.6
Tempo (Mins)	0.513	-	0.174	-	540.6	-

Na Tabela 8 podem-se observar outros indicadores de performance dos 3 algoritmos para comprimento de cadeia  $T = 600$ . Novamente o algoritmo com a melhor performance em relação ao tempo de processamento é o EM Estocástico, mas o algoritmo que teve os menores valores de BIC foi o Bayesiano, assim como porcentagem de acerto na predição de  $S$  mais estável.

Tabela 9 – Resultados do algoritmo EM para a Situação A e  $K = 2$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	0.711	0.342	0.489	0.356	-0.001, 1.003
$\beta_{122}$	-1.1	0.104	0.108	-1.204	1.461	-0.084, 0.285
$\beta_{221}$	-1.3	-0.776	0.367	-0.524	0.409	-1.111, -0.001
$\beta_{222}$	-1.4	0.080	0.094	-1.480	2.200	-0.067, 0.227
$\theta_1$	0.45	0.458	0.019	-0.008	0.000	0.447, 0.500
$\theta_2$	0.55	0.542	0.019	0.008	0.000	0.498, 0.553

Tabela 10 – Resultados do algoritmo EM Estocástico para a Situação A e  $K = 2$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.151	0.544	0.049	0.299	-0.001, 1.003
$\beta_{122}$	-1.1	-0.975	0.380	-0.125	0.160	-0.084, 0.285
$\beta_{221}$	-1.3	-0.558	0.548	-0.742	0.851	-1.111, -0.001
$\beta_{222}$	-1.4	-0.705	0.331	-0.695	0.593	-0.067, 0.227
$\theta_1$	0.45	0.449	0.009	0.001	0.000	0.447, 0.500
$\theta_2$	0.55	0.535	0.023	0.015	0.001	0.498, 0.553

Tabela 11 – Resultados do algoritmo Bayesiano para a Situação A e  $K = 2$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	1.2	1.165	0.138	0.035	0.020	0.964, 1.412
$\beta_{122}$	-1.1	-1.046	0.118	-0.054	0.017	-1.232, -0.811
$\beta_{221}$	-1.3	-1.372	0.174	0.072	0.035	-1.628, -1.054
$\beta_{222}$	-1.4	-1.499	0.202	0.099	0.051	-1.877, -1.207
$\theta_1$	0.45	0.450	0.001	0.000	0.000	0.448, 0.453
$\theta_2$	0.55	0.550	0.001	0.000	0.000	0.547, 0.552

O estudo de simulação também foi repetido aumentando o número de observações para  $T = 1200$ . As Tabelas 9, 10 e 11 mostram os resultados de simulação para os três algoritmos estudados na situação A para comprimento de cadeia  $T = 1200$ . Para este comprimento de

cadeia, pode-se observar que o algoritmo EM continua tendo uma performance baixa. Isto pode ser devido a uma convergência a um máximo local como solução para o conjunto de coeficientes de transição. As melhores estimativas foram obtidas pelo algoritmo Bayesiano, com os erros quadráticos médios e vieses menores e intervalos de credibilidade bem precisos. Em geral a performance do algoritmo Bayesiano neste comprimento de cadeia foi o melhor dentre os três algoritmos.

Tabela 12 – Indicadores para os 3 algoritmos para  $K = 2$  e  $T = 1200$  na situação A.

Indicadores	EM		EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95	Valor	IC95
Acertos em S	0.966	0.912, 1	0.941	0.746, 1	0.947	0.937, 0.958
BIC	9415.9	9127.1, 10554.0	9876.5	9517.1, 11526.9	9124.3	8337.7, 8558.7
DIC	-	-	-	-	8445.1	9046.3, 9247.6
Tempo (Mins)	0.423		0.322	-	1467.782	-

Na Tabela 12 podem-se observar os outros indicadores de performance para os 3 algoritmos no comprimento de cadeia  $T = 1200$ . O algoritmo no qual observou-se a melhor performance em termos de tempo de processamento é o EM Estocástico, mas o algoritmo que teve o menor BIC e porcentagem de acertos na predição de S mais estável foi o Bayesiano.

Para a situação A com 2 estados ocultos, em geral, o algoritmo Bayesiano apresentou os melhores resultados apesar de ter o maior tempo de processamento. Esse procedimento apresenta os menores erros quadráticos médios e vieses, intervalos de credibilidade mais precisos, porcentagem de acertos na predição de S mais estáveis e menores valores de BIC. A metodologia Bayesiana também tem a vantagem de oferecer facilmente estimadores intervalares mesmo para quando temos apenas uma amostra. Os intervalos de confiança para os outros métodos foram calculados apenas por possuímos réplicas, que geralmente não é um caso prático real.

#### 4.2.2 Situação B com $K = 2$

Uma outra situação também com dois estados ocultos foi estudada, Situação B, que apresenta condições que favorecem resultados bons para os algoritmos de estimação. Uma situação pode ser chamada de favorável no âmbito dos NHMM se os coeficientes de transição geram probabilidades que façam que a sequência de estados ocultos não seja excessivamente dinâmica, alternando demais entre estados ocultos, nem excessivamente estável, ficando presa num estado oculto por grandes intervalos de tempo. Além disto, uma situação cujos parâmetros das distribuições que geram os valores observáveis são bem diferentes favorece a identificação pelos algoritmos das mudanças ao longo do tempo.

Para a situação B com dois estados ocultos se fixaram as seguintes condições de simulação:

- A distribuição das variáveis observáveis em cada instante do tempo é  $Y_t \mid S_t = j \sim \text{Binomial}(n_t, \theta_j)$ ;
- vetor de probabilidades de sucesso da distribuição Binomial associada a cada estado oculto,  $\theta = (0.25, 0.75)$ ;
- Número de ensaios de Bernoulli em cada instante da cadeia  $n_t = 500$ , para  $t = 1, 2, \dots, T$ .

Além destas condições e depois de fazer uma análise gráfica das funções de transição, fixamos para a simulação dos dados a seguinte matriz de transição:

$$\beta = \begin{bmatrix} (0) & (2.3 \\ 1.9) \\ (0) & (-2.1 \\ -1.8) \end{bmatrix}. \quad (4.8)$$

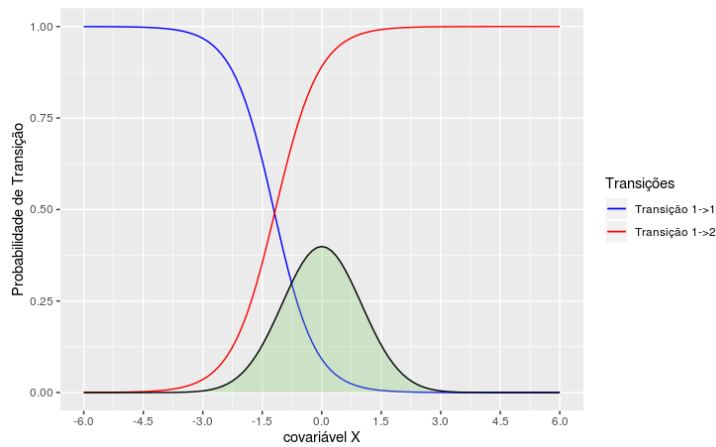


Figura 9 – Funções de transição do estado 1 com  $K = 2$  na Situação B.

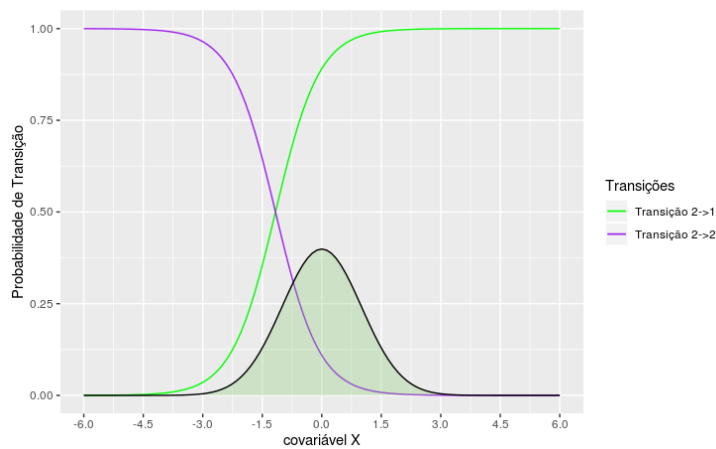


Figura 10 – Funções de transição do estado 2 com  $K = 2$  na Situação B.

Na Figura 9 e 10 observamos os gráficos das funções da transição.

Tabela 13 – Resultados do algoritmo EM para a Situação B e  $K = 2$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	1.528	0.297	0.772	0.684	1.041, 2.099
$\beta_{122}$	1.9	0.035	0.261	1.865	3.548	-0.372, 0.507
$\beta_{221}$	-2.1	-1.459	0.223	-0.641	0.461	-1.778, -1.109
$\beta_{222}$	-1.8	-0.016	0.213	-1.784	3.229	-0.383, 0.346
$\theta_1$	0.25	0.250	0.002	0.000	0.000	0.247, 0.253
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.748, 0.752

Tabela 14 – Resultados do algoritmo EM Estocástico para a Situação B e  $K = 2$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	2.354	0.702	-0.054	0.496	1.162, 3.770
$\beta_{122}$	1.9	1.970	0.969	-0.070	0.944	0.323, 4.166
$\beta_{221}$	-2.1	-6.987	26.186	4.887	709.579	-44.314, -1.139
$\beta_{222}$	-1.8	-21.482	105.461	19.682	11509.398	-162.601, -0.292
$\theta_1$	0.25	0.250	0.002	0.000	0.000	0.248, 0.253
$\theta_2$	0.75	0.750	0.002	0.000	0.000	0.746, 0.752

Tabela 15 – Resultados do algoritmo Bayesiano para a Situação B e  $K = 2$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	2.498	0.397	-0.198	0.197	1.96, 3.247
$\beta_{122}$	1.9	2.221	0.523	-0.321	0.376	1.633, 3.628
$\beta_{221}$	-2.1	-2.194	0.361	0.094	0.139	-2.957, -1.630
$\beta_{222}$	-1.8	-1.852	0.431	0.052	0.189	-2.416, -1.064
$\theta_1$	0.25	0.250	0.002	0.000	0.000	0.247, 0.252
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.748, 0.752

As Tabelas 13, 14 e 15 mostram os resultados de simulação para os três algoritmos estudados na situação B para comprimento de cadeia  $T = 300$ . Podemos observar que os algoritmos EM e EM Estocástico podem ter identificado diferentes valores máximos locais como soluções nas diferentes replicas. O algoritmo EM Estocástico, particularmente apresenta problemas na estimação dos parâmetros  $\beta_{221}$  e  $\beta_{222}$ . Isto pode ser devido à comprimento curto da cadeia. Sendo que nos casos posteriores ( $T = 600$  e  $T = 1200$ ) apresenta resultados bons.

Tabela 16 – Indicadores para os 3 algoritmos para  $K = 2$  e  $T = 300$  na situação B.

Indicadores	EM		EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95	Valor	IC95
Acertos em S	1	1, 1	0.998	0.987, 1	1	1, 1
BIC	2534.1	2465.8, 2584.4	3298.7	2266.9, 4831.1	2444.3	2172.6, 2252.8
DIC	-	-	-	-	2211.8	2391.7, 2480.4
Tempo (Mins)	0.082		0.051	-	237.8	-

Na Tabela 16 podemos observar outros indicadores de performance para os 3 algoritmos no comprimento de cadeia  $T = 300$ . O algoritmo no qual observou-se a melhor performance em



termos de tempo de processamento é o EM Estocástico, mas a tendência continua mostrando que os menores valores de BIC são obtidos pelo algoritmo Bayesiano.

Tabela 17 – Resultados do algoritmo EM para a Situação B e  $K = 2$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	1.484	0.137	0.816	0.684	1.211, 1.681
$\beta_{122}$	1.9	0.045	0.130	1.855	3.457	-0.186, 0.276
$\beta_{221}$	-2.1	-1.426	0.139	-0.674	0.474	-1.707, -1.232
$\beta_{222}$	-1.8	0.000	0.108	-1.800	3.253	-0.181, 0.177
$\theta_1$	0.25	0.250	0.001	0.000	0.000	0.248, 0.252
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.747, 0.752

Tabela 18 – Resultados do algoritmo EM Estocástico para a Situação B e  $K = 2$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	2.354	0.702	-0.054	0.496	1.162, 3.770
$\beta_{122}$	1.9	1.970	0.969	-0.070	0.944	0.323, 4.167
$\beta_{221}$	-2.1	-1.966	0.769	0.133	0.609	-3.107, -1.232
$\beta_{222}$	-1.8	-1.857	0.549	0.057	0.305	-2.682, -1.154
$\theta_1$	0.25	0.250	0.001	0.000	0.000	0.248, 0.253
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.746, 0.752

Tabela 19 – Resultados do algoritmo Bayesiano para a Situação B e  $K = 2$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	2.381	0.225	-0.081	0.057	2.111, 2.916
$\beta_{122}$	1.9	1.973	0.236	-0.073	0.061	1.501, 2.346
$\beta_{221}$	-2.1	-2.204	0.273	0.104	0.086	-2.698, -1.737
$\beta_{222}$	-1.8	-1.860	0.248	0.060	0.065	-2.471, -1.479
$\theta_1$	0.25	0.250	0.001	0.000	0.000	0.248, 0.252
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.748, 0.752

As Tabelas 17, 18 e 19 mostram os resultados de simulação para os três algoritmos estudados na situação B para comprimento de cadeia  $T = 600$ . O algoritmo Bayesiano apresenta resultados ótimos, com erros quadráticos médios muito pequenos. Além disto apresenta vieses e desvios padrões menores do que os outros dois algoritmos. O algoritmo EM Estocástico apresenta vieses pequenos, porém para as estimativas dos coeficientes de transição apresenta desvios padrões relativamente grandes. O algoritmo EM continua apresentando problemas de estimação, possivelmente devido à identificação de máximos locais como soluções para o conjunto de parâmetros.

Na Tabela 20 podemos observar outros indicadores de performance para os 3 algoritmos para comprimento de cadeia  $T = 600$ . O algoritmo no qual observou-se o menor valor calculado para o BIC foi o algoritmo Bayesiano, mas o algoritmo com o menor tempo de processamento foi o algoritmo EM Estocástico.

Tabela 20 – Indicadores para os 3 algoritmos para  $K = 2$  e  $T = 600$  na situação B.

Indicadores	EM		EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95	Valor	IC95
Acertos em S	1	1, 1	0.999	0.998, 1	1	1,1
BIC	5043.4	4992.3, 5110.3	4785.4	4464.1, 7081.3	4423.8	4350.8, 4489,4
DIC	-	-	-	-	4859.2	4804.9, 4916.1
Tempo (Mins)	0.137		0.073	-	606.3	-

Tabela 21 – Resultados do algoritmo EM para a Situação B e  $K = 2$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	1.539	0.114	0.761	0.593	1.324, 1.732
$\beta_{122}$	1.9	0.015	0.120	1.885	3.569	-0.167, 0.272
$\beta_{221}$	-2.1	-1.383	0.103	-0.717	0.525	-1.547, -1.210
$\beta_{222}$	-1.8	-0.010	0.095	-1.790	3.212	-0.211, 0.141
$\theta_1$	0.25	0.250	0.001	0.000	0.000	0.248, 0.251
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.749, 0.751

Tabela 22 – Resultados do algoritmo EM Estocástico para a Situação B e  $K = 2$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	2.332	0.478	-0.032	0.230	1.564, 3.214
$\beta_{122}$	1.9	2.036	0.732	-0.136	0.554	1.084, 3.841
$\beta_{221}$	-2.1	-2.131	0.395	0.031	0.157	-2.808, -1.292
$\beta_{222}$	-1.8	-1.829	0.354	0.029	0.126	-2.510, -1.379
$\theta_1$	0.25	0.250	0.001	0.000	0.000	0.248, 0.251
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.748, 0.751

Tabela 23 – Resultados do algoritmo Bayesiano para a Situação B e  $K = 2$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	2.3	2.423	0.196	-0.123	0.054	2.154, 2.773
$\beta_{122}$	1.9	2.050	0.223	-0.150	0.072	1.635, 2.510
$\beta_{221}$	-2.1	-2.104	0.154	0.004	0.024	-2.304, -1.803
$\beta_{222}$	-1.8	-1.784	0.112	-0.017	0.013	-1.920, -1.498
$\theta_1$	0.25	0.250	0.001	0.000	0.000	0.248, 0.251
$\theta_2$	0.75	0.750	0.001	0.000	0.000	0.749, 0.751

As Tabelas 21, 22 e 23 apresentam os resultados de simulação para os três algoritmos estudados na situação B para comprimento de cadeia  $T = 1200$ . O algoritmo Bayesiano apresenta os menores erros quadráticos médios, porém, o algoritmo EM Estocástico também apresenta estimativas ótimas para todos parâmetros tendo vieses baixos e desvios padrões relativamente baixos mas obteve intervalos de confiança mais largos do que os intervalos de credibilidade do algoritmo Bayesiano. O algoritmo EM, apesar de ter convergido, continua apresentando estimações não muito precisas para os coeficientes de transição.

Na Tabela 24 podemos observar outros indicadores de performance para os 3 algoritmos

Tabela 24 – Indicadores para os 3 algoritmos para  $K = 2$  e  $T = 1200$  na situação B.

Indicadores	EM		EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95	Valor	IC95
Acertos em S	1	1, 1	1	1, 1	1	1,1
BIC	10054.7	9941.8, 10189.7	9054.6	8946.1, 9182.4	9691	9589.1, 9816.1
DIC	-	-	-	-	8848	8740.6, 8946.8
Tempo (Mins)	0.207		0.138	-	1525.32	-

no comprimento de cadeia  $T = 1200$ . O algoritmo Bayesiano obteve os valores de BIC mais baixos para os 2 dos 3 comprimentos estudados na situação B com  $K = 3$ . Os melhor tempos de processamento sempre são obtidos pelo algoritmo EM Estocástico.

Para a situação B com  $K = 2$  estados ocultos, o algoritmo Bayesiano apresenta melhores resultados em termos de performance, apresentando valores do erro quadrático médio e viés menores do que os outros dois algoritmos, além de intervalos de credibilidade mais precisos. Vale ressaltar que só foi possível calcular intervalos de confiança para os estimadores EM e EM Estocástico porque temos réplicas dos conjuntos de dados, o que não acontece num caso prático real. O algoritmo EM Estocástico continua apresentando os melhores tempos de processamento. O algoritmo EM, apesar de ter convergido em cada réplica separadamente, apresenta as piores estimativas principalmente para os coeficientes de transição entre os métodos comparados na situação B com  $K = 2$ .

### 4.2.3 Situação A com $K = 3$

Para as situações que consideram um modelo com 3 estados ocultos, decidiu-se não incluir o algoritmo EM na comparação entre os procedimentos. Isto devido a que para  $K = 3$  estados ocultos, o pacote usado não tem flexibilidade para resolver o problema do *labelswitching* entre as diferentes réplicas e, portanto, não conseguimos sumarizar as estimativas entre todas elas.

Para a situação A com três estados ocultos fixamos as seguintes condições de simulação:

- A distribuição das variáveis observáveis em cada instante do tempo é  $Y_t | S_t = j \sim \text{Binomial}(n_t, \theta_j)$ ;
- vetor de probabilidades de sucesso da distribuição Binomial associada a cada estado oculto,  $\theta = (0.35, 0.50, 0.65)$ ;
- Número de ensaios de Bernoulli em cada instante da cadeia  $n_t = 200$ , para  $t = 1, 2, \dots, T$ .

Além das condições mencionadas anteriormente e tendo feito a análise gráfica das funções de

transição, fixamos a matriz de transição apresentada a seguir:

$$\boldsymbol{\beta} = \begin{bmatrix} (0) & (-1.8) & (-1.9) \\ (0) & (3.7) & (3.1) \\ (0) & (2.4) & (2.9) \end{bmatrix}. \quad (4.9)$$

Ressaltamos que a Situação A é considerada pouco favorável para o processo de estimação, devido a que os coeficientes de transição fixados para esta situação geram probabilidades de transição similares, e dificultam a estimação. Além disto, as probabilidades de sucesso relacionadas a cada estado oculto são próximas entre si, incrementando a dificuldade no momento de identificar o estado oculto em cada posição da cadeia de Markov.

Nas Figuras 11, 12 e 13 observamos a análise gráfica para  $K = 3$  considerando os coeficientes de transição para a situação A.

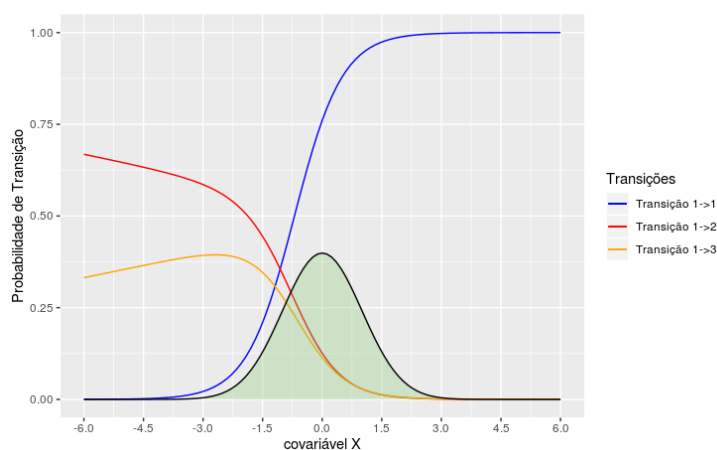


Figura 11 – Funções de Transição do estado 1 com  $K = 3$  na Situação A.

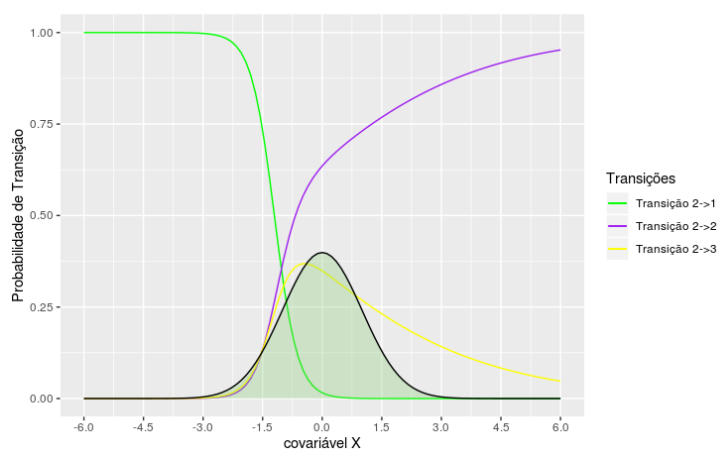


Figura 12 – Funções de Transição do estado 2 com  $K = 3$  na Situação A.

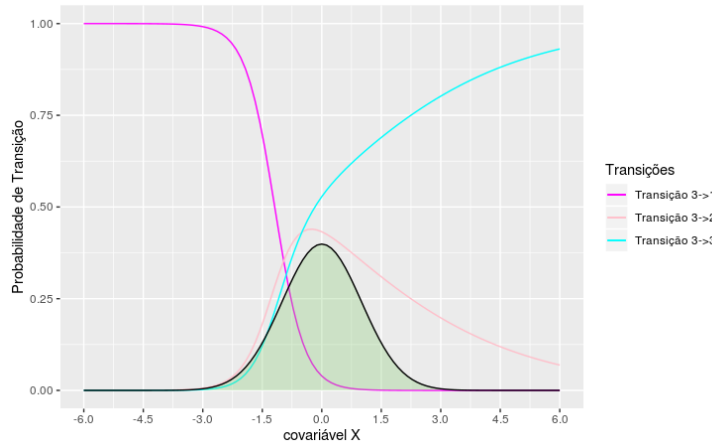


Figura 13 – Funções de Transição do estado 3 com  $K = 3$  na Situação A.

Tabela 25 – Resultados do algoritmo EM Estocástico para a Situação A e  $K = 3$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.8	-1.754	0.418	-0.046	0.177	-2.559, -1.097
$\beta_{122}$	-1.7	-1.599	0.510	-0.101	0.271	-2.641, -0.761
$\beta_{131}$	-1.9	-1.956	0.396	0.056	0.160	-2.659, -1.329
$\beta_{132}$	-1.6	-1.550	0.429	-0.050	0.187	-2.285, -0.882
$\beta_{221}$	3.7	3.105	1.047	0.595	1.451	1.349, 4.917
$\beta_{222}$	3.6	3.002	0.917	0.598	1.197	1.401, 4.670
$\beta_{231}$	3.1	2.619	1.040	0.481	1.312	0.953, 4.627
$\beta_{232}$	3.2	2.346	1.006	0.854	1.743	0.765, 4.200
$\beta_{321}$	2.4	2.612	0.806	-0.212	0.694	1.692, 4.381
$\beta_{322}$	2.5	2.868	0.668	-0.368	0.581	1.959, 4.275
$\beta_{331}$	2.6	2.860	0.814	-0.260	0.730	1.722, 4.644
$\beta_{332}$	2.9	2.991	0.785	-0.091	0.625	1.776, 4.707
$\theta_1$	0.35	0.351	0.005	-0.001	0.000	0.345, 0.360
$\theta_2$	0.5	0.501	0.004	-0.001	0.000	0.495, 0.508
$\theta_3$	0.65	0.649	0.005	0.001	0.000	0.638, 0.657

As Tabelas 25 e 26 mostram os resultados de simulação para os dois algoritmos estudados na situação A para comprimento de cadeia  $T = 300$ . Para este caso, o algoritmo EM Estocástico apresenta uma melhor performance, tendo vieses e erros quadráticos médios menores que o algoritmo Bayesiano para a maioria das estimativas dos parâmetros. Na Tabela 26 observamos que para as transições saindo do estado 2, os erros quadráticos médios ficaram relativamente altos. Isto pode ser devido ao fato desse tipo de transição não acontecer num número suficiente para uma boa estimação.

Na Tabela 27 podemos observar outros indicadores de performance dos 2 algoritmos para comprimento de cadeia  $T = 300$ . O EM Estocástico apresenta um tempo de processamento menor, porém o algoritmo Bayesiano apresenta um valor do BIC menor e mais estável e uma porcentagem de acerto na predição de  $\mathbf{S}$  melhor que o EM Estocástico.

Tabela 26 – Resultados do algoritmo Bayesiano para a Situação A e  $K = 3$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.8	-2.073	0.389	0.273	0.226	-2.736, -1.518
$\beta_{122}$	-1.7	-2.011	0.569	0.311	0.420	-2.958, -1.189
$\beta_{131}$	-1.9	-2.085	0.410	0.185	0.203	-2.855, -1.388
$\beta_{132}$	-1.6	-1.831	0.493	0.231	0.296	2.771, -1.052
$\beta_{221}$	3.7	5.076	2.560	-1.376	8.446	2.597, 12.423
$\beta_{222}$	3.6	4.828	2.436	-1.228	7.444	2.555, 11.788
$\beta_{231}$	3.1	4.405	2.593	-1.305	8.424	2.049, 11.762
$\beta_{232}$	3.2	4.458	2.485	-1.258	7.755	2.093, 11.753
$\beta_{321}$	2.4	3.214	1.140	-0.814	1.963	2.043, 5.827
$\beta_{322}$	2.5	3.410	1.140	-0.910	2.128	1.967, 5.504
$\beta_{331}$	2.6	3.455	1.176	-0.855	2.116	2.167, 6.019
$\beta_{332}$	2.9	3.826	1.119	-0.926	2.109	2.430, 6.110
$\theta_1$	0.35	0.349	0.003	0.001	0.000	0.344, 0.357
$\theta_2$	0.5	0.500	0.004	0.000	0.000	0.494, 0.506
$\theta_3$	0.65	0.650	0.003	0.000	0.000	0.642, 0.656

Tabela 27 – Indicadores para os 2 algoritmos para  $K = 3$  e  $T = 300$ .

Indicadores	EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95
Acertos em S	0.987	0.952, 1	0.994	0.978, 1
BIC	10603.8	2841.2, 18709.4	2060.3	2007.8, 2117.9
DIC	-	-	2780.9	2685.3, 2878.3
Tempo (Mins)	0.184	-	1017.9	-

Tabela 28 – Resultados do algoritmo EM Estocástico para a Situação A e  $K = 3$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.8	-1.665	0.385	-0.135	0.166	-2.324, -1.034
$\beta_{122}$	-1.7	-1.673	0.395	-0.027	0.157	-2.288, -0.996
$\beta_{131}$	-1.9	-1.885	0.332	-0.015	0.110	-2.614, -1.271
$\beta_{132}$	-1.6	-1.567	0.310	-0.033	0.097	-2.329, -1.156
$\beta_{221}$	3.7	3.248	1.077	0.452	1.365	1.410, 4.645
$\beta_{222}$	3.6	3.245	1.092	0.355	1.319	1.405, 4.834
$\beta_{231}$	3.1	2.723	1.041	0.377	1.226	0.988, 4.129
$\beta_{232}$	3.2	2.579	1.063	0.621	1.517	0.604, 4.120
$\beta_{321}$	2.4	2.350	0.547	0.050	0.301	1.265, 3.145
$\beta_{322}$	2.5	2.517	0.524	-0.017	0.275	1.543, 3.361
$\beta_{331}$	2.6	2.514	0.518	0.086	0.275	1.627, 3.303
$\beta_{332}$	2.9	2.673	0.642	0.227	0.463	1.566, 3.544
$\theta_1$	0.35	0.350	0.003	0.000	0.000	0.346, 0.355
$\theta_2$	0.5	0.500	0.003	0.000	0.000	0.495, 0.505
$\theta_3$	0.65	0.650	0.004	0.000	0.000	0.643, 0.656

Tabela 29 – Resultados do algoritmo Bayesiano para a Situação A e  $K = 3$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.8	-1.923	0.279	0.123	0.093	-2.429, -1.498
$\beta_{122}$	-1.7	-1.949	0.430	0.249	0.246	-2.900, -1.417
$\beta_{131}$	-1.9	-1.995	0.309	0.095	0.105	-2.549, -1.425
$\beta_{132}$	-1.6	-1.778	0.488	0.178	0.270	-2.799, -1.114
$\beta_{221}$	3.7	4.391	1.099	-0.691	1.684	3.009, 6.992
$\beta_{222}$	3.6	4.164	0.978	-0.564	1.274	2.764, 6.317
$\beta_{231}$	3.1	3.778	1.041	-0.678	1.543	2.369, 6.189
$\beta_{232}$	3.2	3.770	0.961	-0.570	1.247	2.465, 5.995
$\beta_{321}$	2.4	2.566	0.508	-0.166	0.286	1.775, 3.672
$\beta_{322}$	2.5	2.534	0.484	-0.034	0.235	1.918, 3.453
$\beta_{331}$	2.6	2.695	0.466	-0.095	0.226	2.076, 3.682
$\beta_{332}$	2.9	2.969	0.528	-0.069	0.283	2.269, 4.049
$\theta_1$	0.35	0.350	0.003	0.000	0.000	0.347, 0.355
$\theta_2$	0.5	0.500	0.003	0.000	0.000	0.494, 0.505
$\theta_3$	0.65	0.650	0.004	0.000	0.000	0.644, 0.655

As Tabelas 28 e 29 apresentam os resultados de simulação para os dois algoritmos estudados na situação A para comprimento de cadeia  $T = 600$ . Observamos que para este comprimento de cadeia, os resultados das estimativas para os dois algoritmos são muito similares, mas para alguns parâmetros de transição, o algoritmo EM Estocástico apresenta estimativas menos viesadas.

Tabela 30 – Indicadores para os 2 algoritmos para  $K = 3$  e  $T = 600$ .

Indicadores	EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95
Acertos em S	0.989	0.958, 1	0.994	0.979, 1
BIC	20091.1	5817.4, 33812.3	4131.9	4039.4, 4241.6
DIC	-	-	5468.2	5354.8, 5681.4
Tempo (Mins)	0.611	-	2410.028	-

Na Tabela 30 observamos outros indicadores de performance para os 2 algoritmos no comprimento de cadeia  $T = 600$ . O EM Estocástico apresenta um tempo de processamento consideravelmente mais baixo, mas o valor do BIC para o EM Estocástico é muito maior do que o valor do BIC para o algoritmo Bayesiano.

As Tabelas 31 e 32 apresentam os resultados de simulação para os dois algoritmos estudados na situação A para comprimento de cadeia  $T = 1200$ . Este é o caso de maior comprimento estudado para a situação A. Podemos observar que para o maior comprimento estudado, as estimativas dos dois algoritmos são muito boas. A performance deles em termos de erro quadrático médio e viés é muito parecida, com erros quadráticos médios de algumas estimativas do algoritmo EM Estocástico sendo levemente maiores.

Tabela 31 – Resultados do algoritmo EM Estocástico para a Situação A e  $K = 3$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.8	-1.724	0.254	-0.076	0.070	-2.108, -1.212
$\beta_{122}$	-1.7	-1.711	0.279	0.011	0.078	-2.153, -1.110
$\beta_{131}$	-1.9	-1.914	0.223	0.014	0.050	-2.414, -1.569
$\beta_{132}$	-1.6	-1.625	0.281	0.025	0.080	-2.135, -1.023
$\beta_{221}$	3.7	3.343	0.775	0.357	0.729	1.694, 4.290
$\beta_{222}$	3.6	3.271	0.770	0.329	0.702	1.588, 4.196
$\beta_{231}$	3,1	2.819	0.746	0.281	0.635	1.226, 3.761
$\beta_{232}$	3.2	2.658	0.757	0.542	0.867	1.007, 3.715
$\beta_{321}$	2.4	2.329	0.565	0.071	0.324	1.430, 3.560
$\beta_{322}$	2.5	2.575	0.521	-0.075	0.277	1.791, 3.599
$\beta_{331}$	2.6	2.538	0.555	0.062	0.312	1.713, 3.715
$\beta_{332}$	2.9	2.754	0.510	0.146	0.281	2.005, 3.768
$\theta_1$	0.35	0.351	0.002	-0.001	0.000	0.348, 0.356
$\theta_2$	0.5	0.500	0.002	0.000	0.000	0.496, 0.504
$\theta_3$	0.65	0.649	0.002	0.001	0.000	0.646, 0.652

Tabela 32 – Resultados do algoritmo Bayesiano para a Situação A e  $K = 3$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.8	-1.803	0.201	0.003	0.040	-2.255, -1.502
$\beta_{122}$	-1.7	-1.782	0.278	0.082	0.084	-2.239, -1.292
$\beta_{131}$	-1.9	-1.951	0.202	0.051	0.043	-2.331, -1.632
$\beta_{132}$	-1.6	-1.637	0.270	0.037	0.074	-2.121, -1.223
$\beta_{221}$	3.7	3.957	0.550	-0.257	0.369	3.039, 4.887
$\beta_{222}$	3.6	3.814	0.540	-0.214	0.338	2.887, 4.728
$\beta_{231}$	3.1	3.312	0.543	-0.212	0.340	2.399, 4.246
$\beta_{232}$	3.2	3.428	0.593	-0.228	0.404	2.549, 4.464
$\beta_{321}$	2.4	2.512	0.407	-0.112	0.178	1.913, 3.335
$\beta_{322}$	2.5	2.601	0.330	-0.101	0.119	2.025, 3.313
$\beta_{331}$	2.6	2.651	0.397	-0.051	0.160	1.990, 3.498
$\beta_{332}$	2.9	3.000	0.347	-0.100	0.131	2.385, 3.715
$\theta_1$	0.35	0.350	0.003	0.000	0.000	0.346, 0.355
$\theta_2$	0.5	0.500	0.003	0.000	0.000	0.495, 0.505
$\theta_3$	0.65	0.650	0.004	0.000	0.000	0.643, 0.656

Na Tabela 33 observamos outros indicadores de performance para os 2 algoritmos para comprimento de cadeia  $T = 1200$ . O EM Estocástico apresenta um tempo de processamento consideravelmente mais baixo, Mas o valor do BIC para o EM Estocástico é muito maior do que o valor do BIC para o algoritmo Bayesiano.

Na situação A para três estados ocultos, percebemos que o algoritmo Bayesiano e EM Estocástico tem uma performance similar em relação aos resultados obtidos, exceto no caso  $T = 300$  para o algoritmo Bayesiano, onde os parâmetros de transição do segundo estado oculto apresentaram vieses e desvios padrões relativamente altos. Para alguns parâmetros o



Tabela 33 – Indicadores para os 2 algoritmos para  $K = 3$  e  $T = 600$ .

Indicadores	EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95
Acertos em S	0.994	0.963, 1	0.9953	0.983, 1
BIC	39179.5	11655.7, 67314.8	8263.6	8157.8, 8429.7
DIC	-	-	10923.0	10562.1, 11521.2
Tempo (Mins)	0.541	-	6274.574	-

algoritmo EM Estocástico produz estimativas com viés e desvio padrão menor, porém para outros parâmetros o algoritmo Bayesiano apresenta melhores resultados. O critério no qual o algoritmo Bayesiano claramente tem melhores resultados é o BIC. Para os 3 comprimentos de cadeia, obteve valores menores para o BIC.

#### 4.2.4 Situação B para $K = 3$

Para a situação B com três estados ocultos fixamos as seguintes condições de simulação:

- A distribuição das variáveis observáveis em cada instante do tempo é  $Y_t | S_t = j \sim \text{Binomial}(n_t, \theta_j)$ ;
- vetor de probabilidades de sucesso da distribuição Binomial associada a cada estado oculto,  $\theta = (0.20, 0.50, 0.80)$ ;
- Número de ensaios de Bernoulli em cada instante da cadeia  $n_t = 500$ , para  $t = 1, 2, \dots, T$ .

Além das condições mencionadas anteriormente e fazendo a análise gráfica das funções de transição, fixamos a matriz de transição apresentada a seguir:

$$\beta = \begin{bmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} -1.5 \\ -1.3 \end{pmatrix} & \begin{pmatrix} 1.4 \\ 1.6 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 2.4 \\ 2.5 \end{pmatrix} & \begin{pmatrix} -2.2 \\ -2.8 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} -1.1 \\ -1.5 \end{pmatrix} & \begin{pmatrix} 1.3 \\ 1.7 \end{pmatrix} \end{bmatrix}. \quad (4.10)$$

Ressaltamos que a Situação B é favorável para o processo de estimação, devido a que os coeficientes de transição fixados para esta situação geram probabilidades de transição com diferenças maiores das que são geradas na situação A. Além disto, as probabilidades de sucesso relacionadas a cada estado oculto são mais distantes entre si que na situação A, incrementando a possibilidade do algoritmo identificar o estado oculto em cada posição da cadeia de Markov.

Nas Figuras 14, 15 e 16 observamos a análise gráfica para  $K = 3$  considerando os coeficientes de transição para a situação B. Para esta situação, fixamos coeficientes de transição que geraram probabilidades de transição diferentes no intervalo da covariável com distribuição  $N(0, 1)$ .

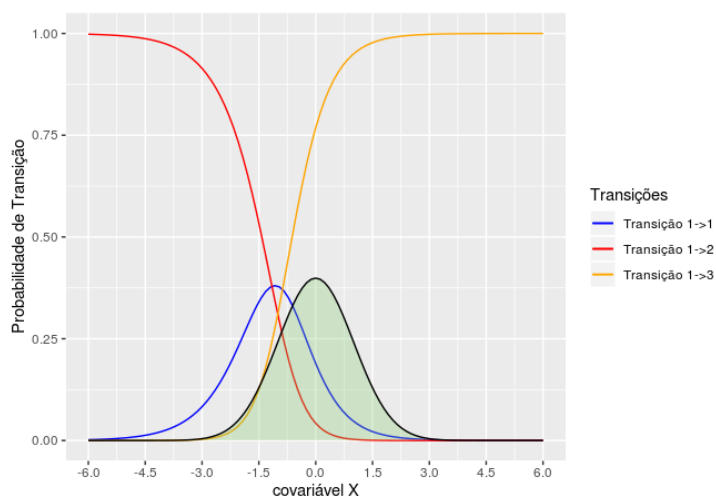


Figura 14 – Funções de Transição do estado 1 com  $K = 3$  na situação B.

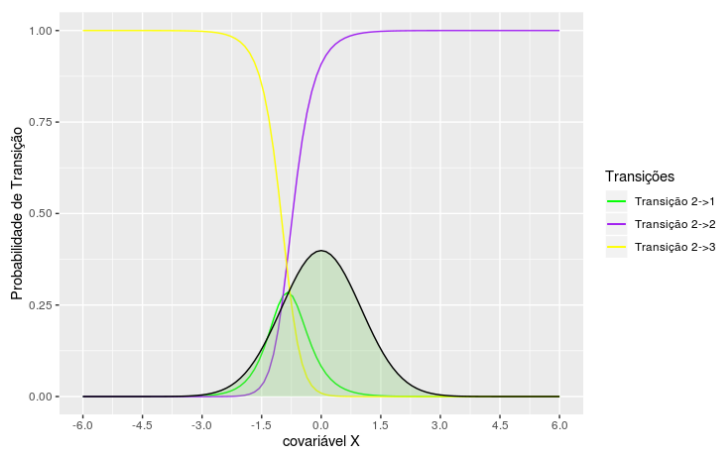


Figura 15 – Funções de Transição do estado 2 com  $K = 3$  na Situação B.

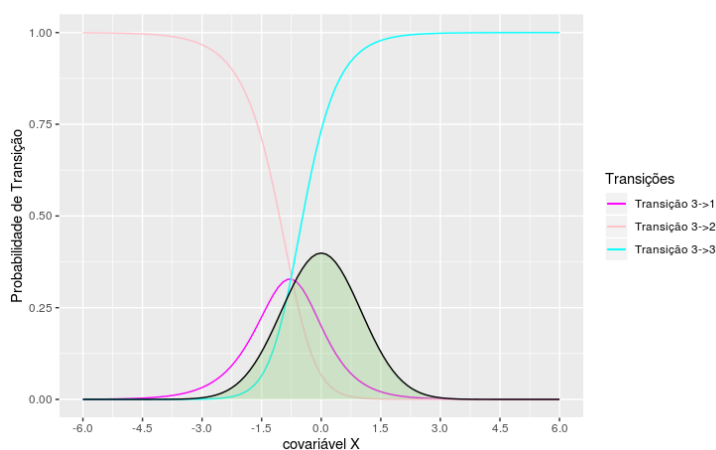


Figura 16 – Funções de Transição do estado 3 com  $K = 3$  na Situação B.

As Tabelas 34 e 35 apresentam os resultados de simulação para os dois algoritmos estudados na situação B para comprimento de cadeia  $T = 300$ . No algoritmo EM Estocástico os parâmetros relacionados a transição do estado 2 ao estado 3 tem um desvio padrão elevado

Tabela 34 – Resultados do algoritmo EM Estocástico para a Situação B e  $K = 3$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.5	-1.575	1.485	0.075	2.210	-4.827, 0.592
$\beta_{122}$	-1.3	-1.548	1.174	0.248	1.441	-3.790, 0.329
$\beta_{131}$	1.4	1.461	0.542	-0.061	0.298	0.403, 2.465
$\beta_{132}$	1.6	1.573	0.924	0.027	0.855	-0.034, 3.109
$\beta_{221}$	2.4	2.503	0.559	-0.103	0.323	1.569, 3.596
$\beta_{222}$	2.5	2.320	1.150	0.180	1.354	-0.195, 3.983
$\beta_{231}$	-2.2	-4.118	6.866	1.918	50.820	-18.509, 2.017
$\beta_{232}$	-2.8	-6.050	8.322	3.250	79.818	-23.886, -0.105
$\beta_{321}$	-1.1	-0.028	0.629	-1.072	1.544	-0.685, 1.692
$\beta_{322}$	-1.5	-0.653	0.375	-0.847	0.857	-1.369, -0.155
$\beta_{331}$	1.3	1.343	0.516	-0.043	0.269	0.584, 2.428
$\beta_{332}$	1.7	1.622	0.489	0.078	0.245	0.562, 2.386
$\theta_1$	0.2	0.237	0.084	-0.037	0.008	0.194, 0.433
$\theta_2$	0.5	0.538	0.089	-0.038	0.009	0.497, 0.796
$\theta_3$	0.8	0.800	0.003	0.000	0.000	0.794, 0.803

Tabela 35 – Resultados do algoritmo Bayesiano para a Situação B e  $K = 3$  com  $T = 300$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.5	-1.499	0.650	-0.001	0.423	-2.560, -0.2701
$\beta_{122}$	-1.3	-1.265	0.550	-0.035	0.304	-2.204, -0.192
$\beta_{131}$	1.4	1.520	0.335	-0.120	0.127	0.939, 2.203
$\beta_{132}$	1.6	1.729	0.342	-0.129	0.134	1.148, 2.345
$\beta_{221}$	2.4	2.213	1.372	0.187	1.918	0.017, 2.969
$\beta_{222}$	2.5	2.428	0.483	0.072	0.238	1.665, 3.220
$\beta_{231}$	-2.2	-2.324	0.705	0.124	0.512	-4.077, -1.296
$\beta_{232}$	-2.8	-2.959	0.569	0.159	0.349	-4.150, -2.187
$\beta_{321}$	-1.1	-1.087	0.249	-0.013	0.062	-1.429, -0.654
$\beta_{322}$	-1.5	-1.479	0.267	-0.021	0.072	-1.953, -0.911
$\beta_{331}$	1.3	1.308	0.180	-0.008	0.032	1.022, 1.601
$\beta_{332}$	1.7	1.755	0.248	-0.055	0.065	1.357, 2.165
$\theta_1$	0.2	0.223	0.110	-0.023	0.013	0.198, 0.432
$\theta_2$	0.5	0.497	0.017	0.003	0.000	0.473, 0.501
$\theta_3$	0.8	0.800	0.001	0.000	0.000	0.798, 0.801

e apresentam um viés alto, e como resultado apresentou um erro quadrático médio alto. Além disto, os intervalos de confiança para alguns dos parâmetros de transição incluem o valor 0. No algoritmo Bayesiano obtivemos estimativas com vieses baixos e desvios padrões baixos, o que nos indica uma melhor performance do algoritmo Bayesiano em termos do erro quadrático médio.

Na Tabela 36 observamos outros indicadores de performance para os dois algoritmos no comprimento de cadeia  $T = 300$ . Os tempos de processamento continuam com a tendência dos casos anteriores, sendo que o algoritmo EM Estocástico sempre tem um tempo de processamento

Tabela 36 – Indicadores para os 2 algoritmos para  $K = 3$  e  $T = 300$  na Situação B.

Indicadores	EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95
Acertos em S	0.970	0.765, 1	0.9941	0.952, 1
BIC	82839.4	3304.3, 170129.3	12215.1	8792.7, 35545.5
DIC	-	-	18232.6	11341.8, 75490.3
Tempo (Mins)	0.138	-	849.2	-

consideravelmente menor. O algoritmo Bayesiano apresenta valores menores do BIC que o EM Estocástico.

Tabela 37 – Resultados do algoritmo EM Estocástico para a Situação B e  $K = 3$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.5	-1.681	1.135	0.181	1.321	-4.525, 0.105
$\beta_{122}$	-1.3	-1.591	0.984	0.291	1.052	-4.036, -0.677
$\beta_{131}$	1.4	1.432	0.518	-0.032	0.269	0.197, 2.105
$\beta_{132}$	1.6	1.635	0.618	-0.035	0.383	0.286, 2.597
$\beta_{221}$	2.4	2.541	0.490	-0.141	0.260	1.777, 3.548
$\beta_{222}$	2.5	2.485	0.686	0.015	0.471	0.719, 3.440
$\beta_{231}$	-2.2	-2.484	1.518	0.284	2.386	-5.146, 0.823
$\beta_{232}$	-2.8	-3.976	1.655	1.176	4.122	-7.456, 0.274
$\beta_{321}$	-1.1	-0.190	0.470	-0.910	1.049	-0.570, 0.599
$\beta_{322}$	-1.5	-0.725	0.342	-0.775	0.718	-1.513, -0.256
$\beta_{331}$	1.3	1.276	0.438	0.024	0.192	0.538, 1.918
$\beta_{332}$	1.7	1.704	0.366	-0.004	0.134	0.938, 2.275
$\theta_1$	0.2	0.236	0.081	-0.036	0.008	0.196, 0.426
$\theta_2$	0.5	0.538	0.089	-0.038	0.009	0.497, 0.766
$\theta_3$	0.8	0.800	0.001	0.000	0.000	0.798, 0.802

As Tabelas 37 e 38 apresentam os resultados de simulação para os dois algoritmos estudados na situação B para comprimento de cadeia  $T = 600$ . Observamos que os resultados para este caso são similares. Para alguns parâmetros, o algoritmo EM teve melhor performance em termos de erro quadrático médio, mas para alguns outros parâmetros o algoritmo Bayesiano teve melhor performance. Um detalhe importante é que os intervalos de confiança para algumas das estimativas no algoritmo EM Estocástico incluem o valor 0.

A Tabela 39 mostra outros indicadores de performance para os 2 algoritmos no comprimento de cadeia  $T = 600$ . Os tempos de processamento seguem a mesma tendência dos casos anteriores, com o algoritmo EM Estocástico tendo um tempo de processamento menor. O algoritmo Bayesiano apresenta valores menores do BIC que o EM Estocástico.

As Tabelas 40 e 41 apresentam os resultados de simulação para os dois algoritmos estudados na situação B para comprimento de cadeia  $T = 1200$ . Para a maioria das estimativas dos parâmetros o algoritmo Bayesiano apresenta erros quadráticos médios menores, principalmente

Tabela 38 – Resultados do algoritmo Bayesiano para a Situação B e  $K = 3$  com  $T = 600$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.5	-2.252	1.881	0.752	4.106	-6.431, -0.266
$\beta_{122}$	-1.3	-1.856	1.496	0.556	2.546	-5.249, -0.108
$\beta_{131}$	1.4	1.595	0.446	-0.195	0.237	0.847, 2.572
$\beta_{132}$	1.6	1.823	0.662	-0.223	0.488	0.973, 3.326
$\beta_{221}$	2.4	2.551	0.315	-0.151	0.122	2.007, 3.150
$\beta_{222}$	2.5	2.663	0.548	-0.163	0.327	1.843, 3.795
$\beta_{231}$	-2.2	-2.683	0.860	0.483	0.973	-4.137, -1.526
$\beta_{232}$	-2.8	-3.388	0.941	0.588	1.230	-5.028, -1.912
$\beta_{321}$	-1.1	-1.063	0.432	-0.037	0.188	-1.964, -0.530
$\beta_{322}$	-1.5	-1.472	0.393	-0.028	0.155	-2.293, -0.836
$\beta_{331}$	1.3	1.323	0.185	-0.023	0.035	1.071, 1.659
$\beta_{332}$	1.7	1.692	0.204	0.008	0.042	1.295, 2.041
$\theta_1$	0.2	0.213	0.073	-0.013	0.006	0.196, 0.312
$\theta_2$	0.5	0.498	0.011	0.002	0.000	0.482, 0.503
$\theta_3$	0.8	0.800	0.001	0.000	0.000	0.798, 0.802

Tabela 39 – Indicadores para os 2 algoritmos para  $K = 3$  e  $T = 600$  na situação B.

Indicadores	EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95
Acertos em S	0.981	0.771, 1	1	1,1
BIC	182658.8	6880.8, 320662.5	4635	4366.3, 6129.8
DIC	-	-	7069	5678.4, 15617.1
Tempo (Mins)	0.209	-	246	-

peço fato de ter vieses muito menores para a maioria das estimativas dos parâmetros.

A Tabela 42 mostra os indicadores para os 2 algoritmos no comprimento de cadeia  $T = 1200$ . Os tempos de processamento seguem a mesma tendência dos casos anteriores, com o algoritmo EM Estocástico tendo um tempo de processamento menor. O algoritmo Bayesiano apresenta valores menores do BIC que o EM Estocástico.

Para a situação B com  $K = 3$  estados ocultos, o algoritmo Bayesiano apresenta melhores resultados porém, o algoritmo EM Estocástico apresenta resultados aceitáveis para todos os comprimentos de cadeia estudados nesta situação excetuando  $T = 300$ , onde houve um problema de estimação. Em termos de tempo de processamento, o algoritmo EM sempre apresenta tempos de processamento menores.

Tabela 40 – Resultados do algoritmo EM Estocástico para a Situação B e  $K = 3$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.5	-1.334	0.924	-0.166	0.881	-2.597, 1.162
$\beta_{122}$	-1.3	-1.423	0.527	0.123	0.293	-2.502, -0.574
$\beta_{131}$	1.4	1.366	0.312	0.034	0.099	0.821, 1.953
$\beta_{132}$	1.6	1.599	0.420	0.001	0.176	0.803, 2.416
$\beta_{221}$	2.4	2.566	0.396	-0.166	0.185	2.140, 3.311
$\beta_{222}$	2.5	2.464	0.425	0.036	0.182	1.280, 2.980
$\beta_{231}$	-2.2	-1.910	1.489	-0.290	2.302	-3.219, 2.015
$\beta_{232}$	-2.8	-3.248	1.647	0.448	2.914	-4.824, 1.075
$\beta_{321}$	-1.1	-0.155	0.516	-0.945	1.159	-0.496, 0.977
$\beta_{322}$	-1.5	-0.650	0.323	-0.850	0.827	-1.411, -0.229
$\beta_{331}$	1.3	1.264	0.359	0.036	0.130	0.993, 1.892
$\beta_{332}$	1.7	1.704	0.301	-0.004	0.090	1.182, 2.068
$\theta_1$	0.2	0.237	0.084	-0.037	0.008	0.198, 0.425
$\theta_2$	0.5	0.504	0.102	-0.004	0.012	0.498, 0.792
$\theta_3$	0.8	0.800	0.001	0.000	0.000	0.798, 0.802

Tabela 41 – Resultados do algoritmo Bayesiano para a Situação B e  $K = 3$  com  $T = 1200$ .

Parâmetro	Real	Estimado	SD	Viés	EQM	IC95
$\beta_{121}$	-1.5	-1.499	0.650	-0.001	0.423	-2.560, -0.271
$\beta_{122}$	-1.3	-1.265	0.550	-0.035	0.304	-2.204, -0.192
$\beta_{131}$	1.4	1.520	0.335	-0.120	0.127	0.939, 2.204
$\beta_{132}$	1.6	1.729	0.342	-0.129	0.134	1.148, 2.344
$\beta_{221}$	2.4	2.213	1.372	0.187	1.918	0.017, 2.969
$\beta_{222}$	2.5	2.428	0.483	0.072	0.238	1.665, 3.219
$\beta_{231}$	-2.2	-2.324	0.705	0.124	0.512	-4.077, -1.296
$\beta_{232}$	-2.8	-2.959	0.569	0.159	0.349	-4.151, -2.187
$\beta_{321}$	-1.1	-1.087	0.249	-0.013	0.062	-1.429, -0.654
$\beta_{322}$	-1.5	-1.479	0.267	-0.021	0.072	-1.953, -0.911
$\beta_{331}$	1.3	1.308	0.180	-0.008	0.032	1.022, 1.601
$\beta_{332}$	1.7	1.755	0.248	-0.055	0.065	1.357, 2.165
$\theta_1$	0.2	0.223	0.110	-0.023	0.013	0.198, 0.432
$\theta_2$	0.5	0.497	0.017	0.003	0.000	0.474, 0.502
$\theta_3$	0.8	0.800	0.001	0.000	0.000	0.798, 0.801

Tabela 42 – Indicadores para os 2 algoritmos para  $K = 3$  e  $T = 1200$  na Situação B.

Indicadores	EM Estocástico		Bayesiano	
	Valor	IC95	Valor	IC95
Acertos em S	0.978	0.764, 1	0.994	0.95, 1
BIC	253380.8	12886.4, 627443.2	12215.1	8792.7, 35545.4
DIC	-	-	18232.6	11341.8, 75490.3
Tempo (Mins)	0.353	-	6660.433	-



---

## APLICAÇÃO EM DADOS REAIS

---

Os modelos Markovianos ocultos em tempo discreto são uma classe de modelos amplamente úteis em áreas como o reconhecimento de linguagem natural, genética e economia. Uma outra área na qual os modelos Markovianos ocultos apresentam muita versatilidade para modelar situações com heterogeneidade nos dados é a Climatologia. [Holsclaw \*et al.\* \(2017\)](#) utilizam um NHMM para modelar quantidades de chuva em Índia, num período de tempo de 30 anos iniciando em 1981 e terminado em 2010. A utilização de um NHMM possibilitou discernimento sobre a duração e acontecimento de períodos de chuva em todo o país.

Nessa seção, usando os algoritmos estudados neste trabalho, apresentamos uma modelagem de períodos de chuva na cidade de Tegucigalpa, capital de Honduras, na América Central. Nosso objetivo é ter uma melhor compreensão do comportamento dos períodos de chuva no lugar mencionado.

### 5.1 Descrição dos dados

Os dados foram coletados na estação meteorológica de Tegucigalpa, Honduras, entre os anos de 1969 e 2019. A estação meteorológica coletava dados de quantidades de precipitação diariamente, mas para esta aplicação teremos como variável repostada a quantidade de dias com chuva registrados mensalmente. Consideramos como dia com chuva, um dia onde o registro de precipitação seja maior que 0.5 mm. A base de dados tem 610 observações, cada uma representando a quantidade de dias com chuva observados mensalmente, considerando meses com duração de 30 dias. Duas covariáveis serão utilizadas separadamente, umidade relativa média, e temperatura média. Estas médias foram calculadas mensalmente e os modelos serão ajustados utilizando uma covariável à vez.



Tabela 43 – Estatística descritiva das variáveis utilizadas.

Variável	Min.	Q1	Média	Q3	Max	Desv. Padrão	Assimetria
DiasChuva (Dias/Mês)	1	7	13	19	29	6.38	0.14
Temperatura (° F)	18.20	21.00	22.03	23.00	25.90	1.44	-0.30
Umidade Relativa (%)	43.50	65.25	70.24	75.78	90.80	7.80	-0.52

Na Tabela 43 podemos observar as estatísticas descritivas dos dados que serão utilizados na aplicação.

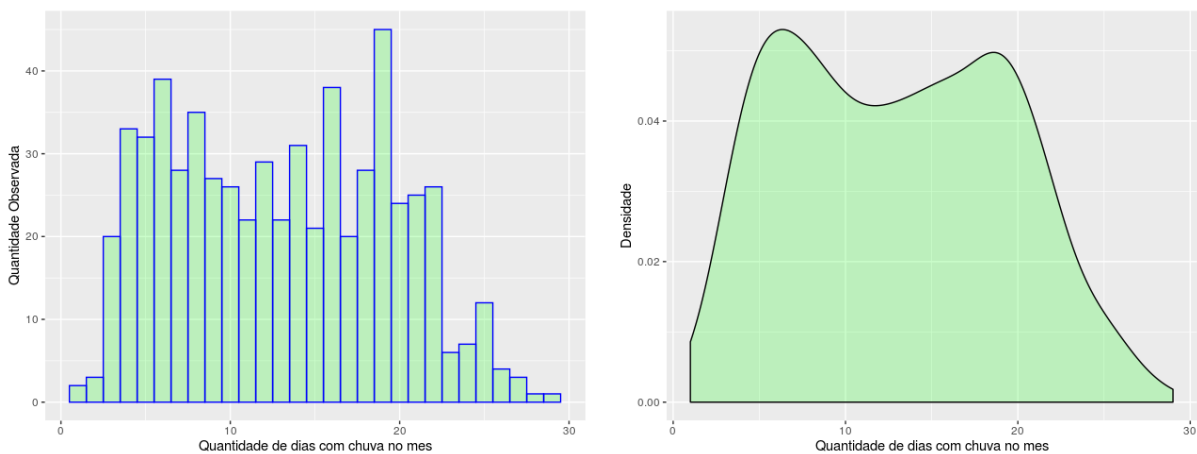


Figura 17 – Histograma e densidade dos dados de dias com chuva no mês.

Na Figura 17, no histograma podemos perceber que as quantidades observadas de dias com chuva parecem estar agrupadas ao redor de 3 valores, porém esta tendência não está completamente clara. Além disso, a densidade dos dados mostra duas áreas com maior concentração de densidade.

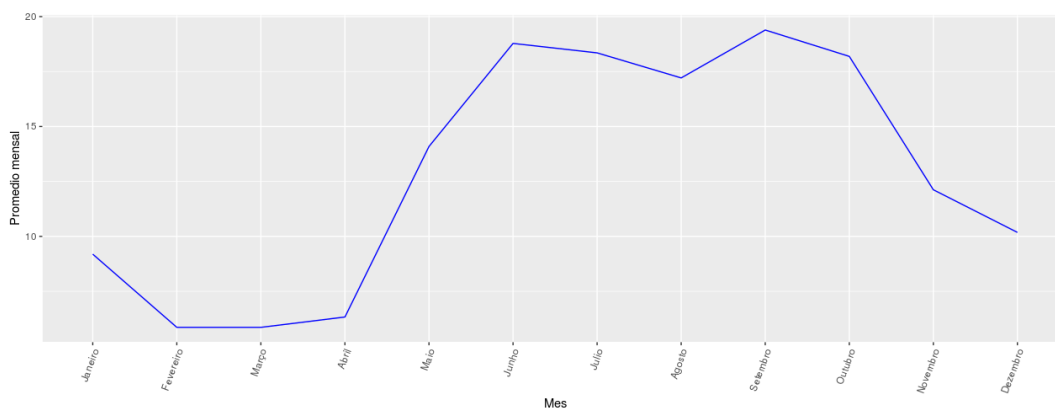


Figura 18 – Média de dias com chuva mensais no período 1969-2019.

A Figura 18 apresenta um gráfico interessante, na qual pode-se ver as médias de dias com chuva para cada um dos doze meses do ano em todo o período no qual foram coletados os dados. Esse gráfico nos dá uma ideia *a priori* do comportamento dos períodos de chuva na

cidade de Tegucigalpa, sendo que existe um período onde acontece uma menor quantidade de dias com chuva por mês que compreende os meses de abril a dezembro. Este período é chamado de estação seca. Além disso existe um período onde acontece uma maior quantidade de dias com chuva por mês entre junho e outubro, e este período é chamado de estação chuvosa.

## 5.2 Estimação com NHMM

Os três algoritmos estudados neste trabalho foram aplicados no conjunto de dados para realizar o processo de estimação. Este processo foi realizado usando um modelo Markoviano oculto não-homogêneo. A função de ligação utilizada para realizar os ajustes foi a função *mlogit*, dada pela Equação (4.2).

Foram realizados dois ajustes considerando uma distribuição *Binomial*( $N, \theta_i$ ) para os dados e fixamos  $N = 30$  e  $i \in \{1, 2, \dots, K - 1, K\}$ . O primeiro ajuste realizado tem como variável resposta dias com chuva (DiasChuva), e a covariável de transição é temperatura média no mês (Temperatura). O segundo ajuste realizado tem como variável resposta dias com chuva, a covariável de transição é umidade relativa média no mês (UmidadeRelativa).

### 5.2.1 Ajuste com dois estados ocultos, $K = 2$

#### 5.2.1.1 Algoritmo EM

Tabela 44 – Resultados para o ajuste DiasChuva-Temperatura.

Parâmetro	Estimado
$\beta_{121}$	-1.349
$\beta_{122}$	-2.853
$\beta_{221}$	1.301
$\beta_{222}$	-1.059
$\theta_1$	0.605
$\theta_2$	0.245

Na Tabela 44 pode-se observar as estimativas para os parâmetros do modelo ajustado utilizando a covariável temperatura média mensal. O tempo de processamento para este ajuste foi menor a 1 segundo e o valor para o BIC foi de 3739.82.

Tabela 45 – Resultados para o ajuste DiasChuva-UmidadeRelativa.

<b>Parâmetro</b>	<b>Estimado</b>
$\beta_{121}$	-1.842
$\beta_{122}$	2.786
$\beta_{221}$	1.187
$\beta_{222}$	-0.090
$\theta_1$	0.608
$\theta_2$	0.244

A Tabela 45 mostra as estimativas para os parâmetros do ajuste utilizando a covariável umidade relativa média mensal. O tempo de processamento para este ajuste foi menor a 1 segundo, o valor calculado para o BIC foi de 3887.42.

Utilizando o menor BIC calculado como critério para escolher um melhor ajuste, então escolhemos o modelo no qual a covariável é a temperatura média mensal.

#### 5.2.1.2 Algoritmo EM Estocástico

Tabela 46 – Resultados para o ajuste DiasChuva-Temperatura.

<b>Parâmetro</b>	<b>Estimado</b>
$\beta_{121}$	-2.129
$\beta_{122}$	-2.297
$\beta_{221}$	1.103
$\beta_{222}$	-0.289
$\theta_1$	0.612
$\theta_2$	0.248

As estimativas obtidas para os parâmetros do modelo ajustado utilizando a covariável temperatura média mensal são mostradas na Tabela 46. Observamos um tempo de processamento menor a 1 segundo para este modelo, e o valor calculado do BIC foi de 4344.21.

Tabela 47 – Resultados para o ajuste DiasChuva-UmidadeRelativa.

<b>Parâmetro</b>	<b>Estimado</b>
$\beta_{211}$	-0.436
$\beta_{221}$	-4.807
$\beta_{212}$	0.948
$\beta_{222}$	-0.184
$\theta_1$	0.616
$\theta_2$	0.252

A Tabela 47 mostra as estimativas para os parâmetros do ajuste utilizando a covariável umidade relativa média mensal. O tempo de processamento para este ajuste foi menor a 1 segundo, o valor calculado para o BIC foi de 4553.06.

Entre os modelos ajustados utilizando o algoritmo EM estocástico, escolhemos o modelo no qual a covariável é a temperatura média mensal, que apresenta o menor valor do BIC.

### 5.2.1.3 Estimação Bayesiana

Para o processo de estimação usamos o RJAGS. A seguir mencionamos algumas condições importantes sob as quais foram feitos os processos de estimação usando esta ferramenta:

- 55,000 iterações do algoritmo.
- 5,000 iterações de *burn-in*.
- Intervalo de espaçamento entre amostras coletadas igual a 5.

Tabela 48 – Resultados para o ajuste DiasChuva-Temperatura.

Parâmetro	Estimado	SD	IC95%	Assimetria
$\beta_{121}$	-2.331	0.299	-2.848, -1.879	-0.248
$\beta_{122}$	-2.387	0.373	-2.983, -1.858	0.129
$\beta_{221}$	1.346	0.171	1.054, 1.659	-0.213
$\beta_{222}$	-0.408	0.154	-0.664, -0.148	-0.088
$\theta_1$	0.607	0.006	0.598, 0.617	-0.019
$\theta_2$	0.244	0.006	0.235, 0.253	0.022

Na Tabela 48 podemos observar as estimativas para os parâmetros do modelo ajustado utilizando a covariável temperatura média mensal. O tempo de processamento para este ajuste foi de 22.99 minutos e os valores para o BIC e DIC foram de 3726.82 e 3527.57 respectivamente. Também podemos observar valores relativamente pequenos do desvio padrão, junto com intervalos de credibilidade de comprimento pequeno para os estimadores.

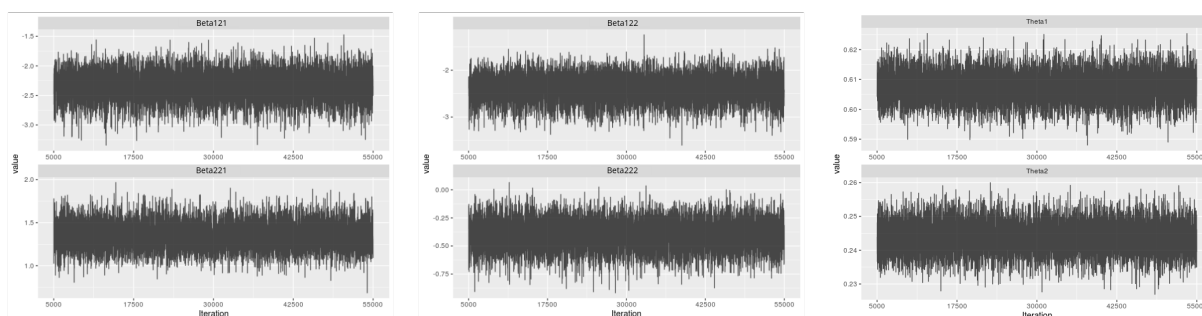


Figura 19 – Gráficos de convergência para os parâmetros do modelo DiasChuva-Temperatura.

Na Figura 19 observamos os gráficos de convergência para os parâmetros. Estes gráficos mostram uma série temporal do processo de amostragem e neles observamos que o amostrador visita o espaço paramétrico de forma aleatória e eficiente, como esperado.

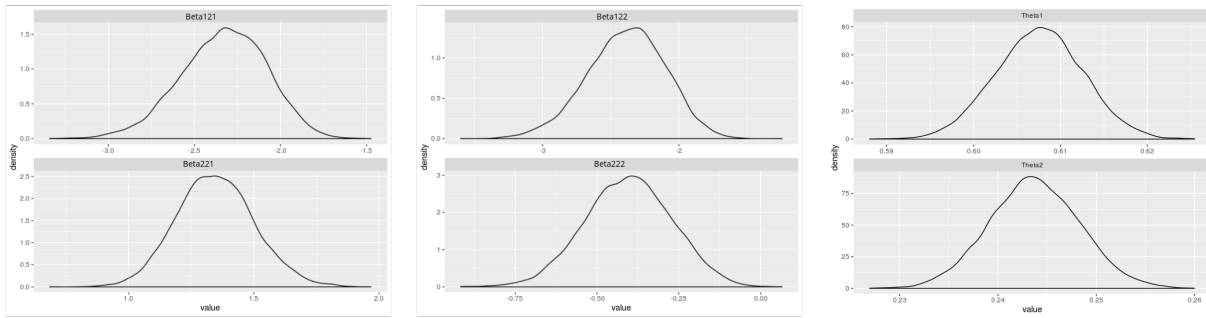


Figura 20 – Gráficos de densidade *a posteriori* para os parâmetros do modelo DiasChuva-Temperatura.

Na Figura 20 observamos os gráficos das densidades *a posteriori* dos parâmetros para este ajuste. Os valores de assimetria para estes parâmetros estão disponíveis na Tabela 48.

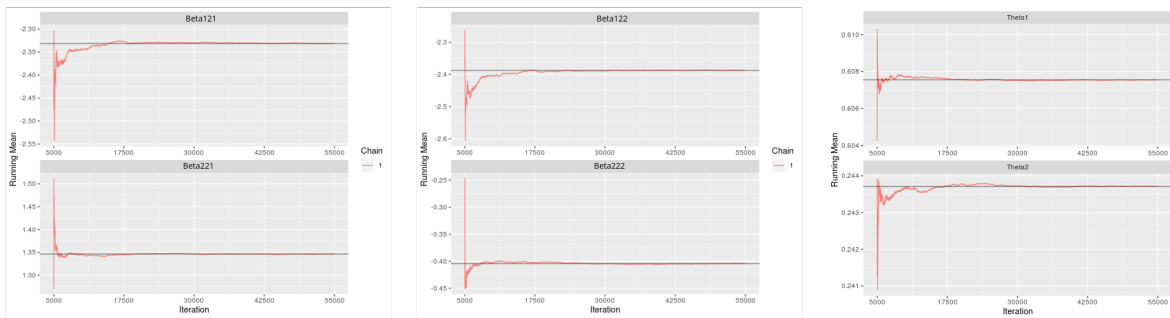


Figura 21 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva-Temperatura.

O gráfico denominado *running mean plot* ou gráfico de média móvel é uma ferramenta útil para diagnosticar convergência do amostrador MCMC. Eles mostram a média de todos os valores amostrados até uma certa iteração do algoritmo MCMC. Além disto, mostram o quão rápido a cadeia vai se aproximando à sua distribuição alvo. Na Figura 21 observamos que as médias dos parâmetros vão se aproximando à média *a posteriori* na medida que o algoritmo MCMC avança no processo de amostragem.

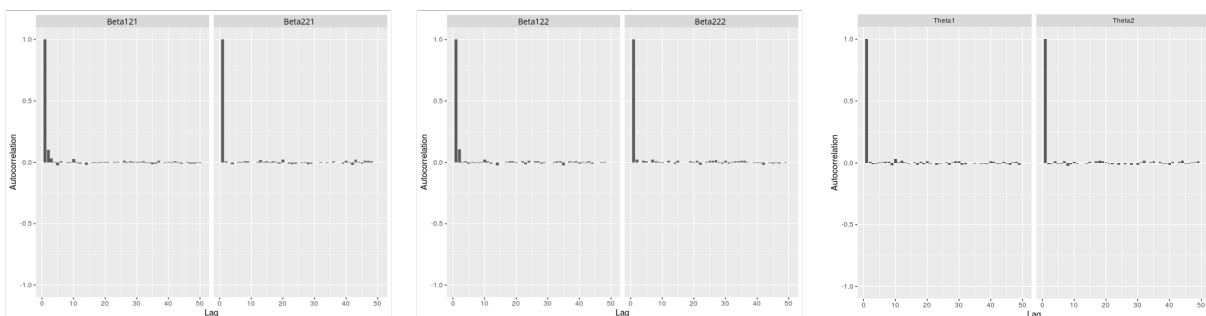


Figura 22 – Gráficos de autocorrelação para os parâmetros do modelo DiasChuva-Temperatura.

Na Figura 22 observamos os gráficos de autocorrelação para os parâmetros. Estes gráficos mostram a correlação entre cada amostra, e os valores amostrados  $l$  iterações atrás. Nos gráficos notamos que a correlação diminui notavelmente quando  $l > 2$ .

Tabela 49 – Resultados para os ajuste DiasChuva-UmidadeRelativa.

Parâmetro	Estimado	SD	IC95%	Assimetria
$\beta_{121}$	-0.626	0.174	-0.963, -0.282	0.026
$\beta_{122}$	0.113	0.210	-0.303, 0.517	-0.084
$\beta_{221}$	-4.539	0.815	-6.169, -2.994	-0.162
$\beta_{222}$	-6.488	0.925	-8.388, -4.816	-0.266
$\theta_1$	0.606	0.006	0.594, 0.617	0.016
$\theta_2$	0.243	0.005	0.232, 0.254	0.005

Na Tabela 49 observamos as estimativas para os parâmetros do modelo ajustado utilizando a covariável umidade relativa média mensal. O tempo de processamento para este ajuste foi de 25.11 minutos e os valores para o BIC e DIC foram de 3833.84 e 3534.26 respectivamente. Aqui ressaltamos desvios padrões maiores do que os calculados na Tabela 48 para os parâmetros  $\beta_{221}$  e  $\beta_{222}$ .

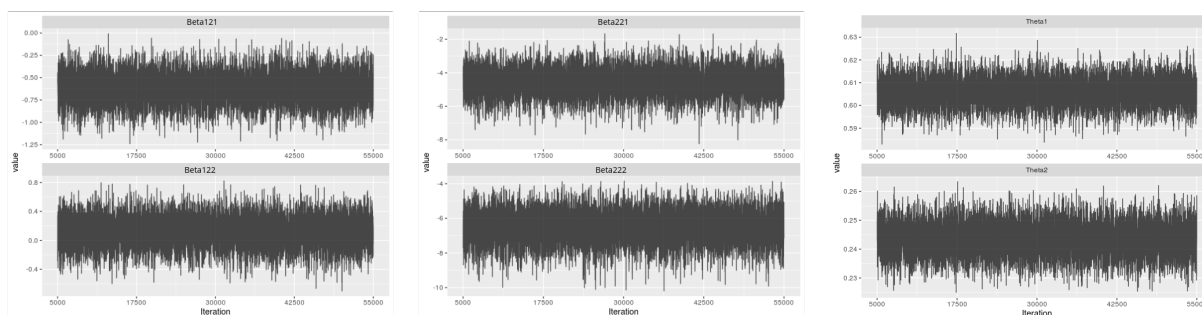


Figura 23 – Gráficos de convergência para os parâmetros do modelo DiasChuva-UmidadeRelativa.

Na Figura 23 observamos os gráficos de convergência para os parâmetros. Observamos que o espaço paramétrico está sendo visitado de forma aleatória.

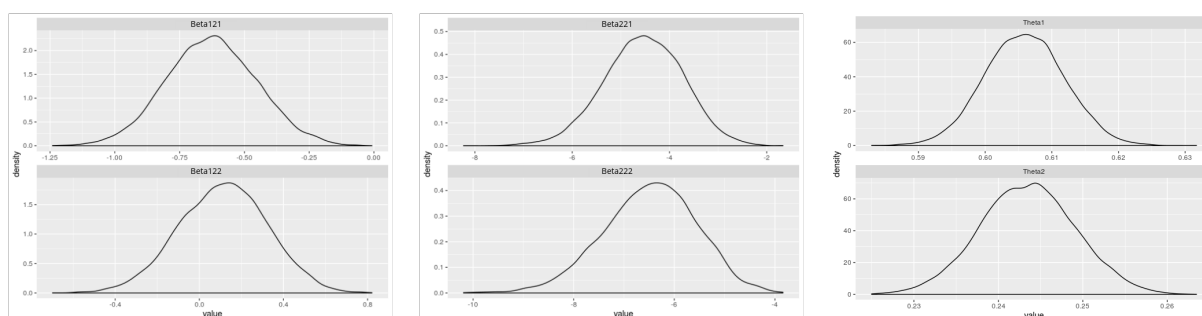


Figura 24 – Gráficos de densidade *a posteriori* para os parâmetros do modelo DiasChuva-UmidadeRelativa.

Na Figura 24 pode-se ver os gráficos das densidades *a posteriori* dos parâmetros. Os valores de assimetria para estes parâmetros estão disponíveis na Tabela 49.

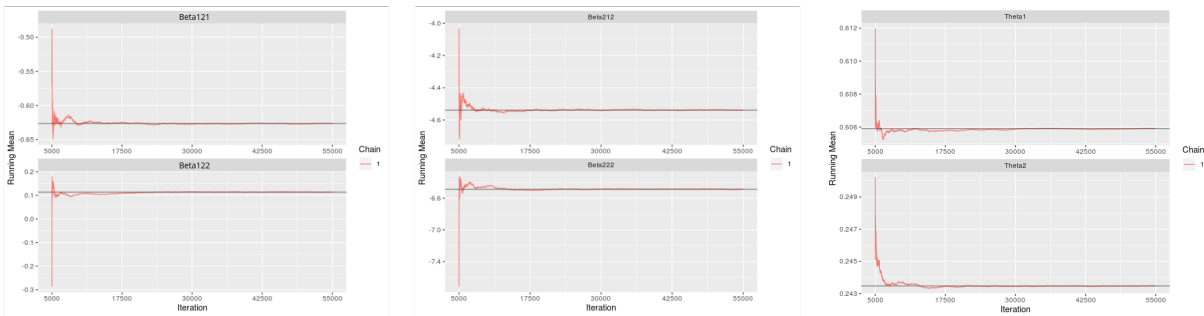


Figura 25 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva-UmidadeRelativa.

Na Figura 25 pode-se ver que as médias dos parâmetros vão se aproximando à média *a posteriori* na medida que o algoritmo MCMC avança no processo de amostragem.

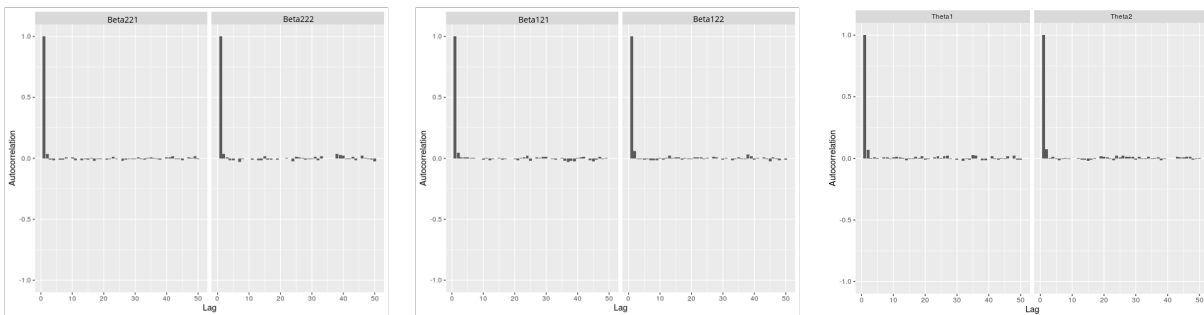


Figura 26 – Gráficos de autocorrelação para os parâmetros do modelo DiasChuva-UmidadeRelativa.

Na Figura 26 pode-se observar os gráficos de autocorrelação para os parâmetros. Nos gráficos pode-se ver que a correlação diminui para  $l > 2$ .

## 5.2.2 Ajuste com três estados ocultos, $K=3$

### 5.2.2.1 Algoritmo EM

A Tabela 50 mostra os resultados do modelo ajustado utilizando a covariável temperatura média mensal. Para este ajuste, o tempo de processamento foi menor a 1 segundo. O valor calculado para o BIC foi de 3583.10.

Na Tabela 51 observamos os resultados do modelo ajustado utilizando a covariável umidade relativa média mensal. Para este ajuste, o tempo de processamento foi menor a 1 segundo. O valor calculado para o BIC foi de 3733.80.

Tabela 50 – Resultados para o ajuste DiasChuva-Temperatura.

<b>Parâmetro</b>	<b>Estimado</b>
$\beta_{121}$	-0.971
$\beta_{122}$	-0.824
$\beta_{131}$	-0.655
$\beta_{132}$	1.809
$\beta_{221}$	1.916
$\beta_{222}$	-2.021
$\beta_{231}$	-0.855
$\beta_{232}$	-0.517
$\beta_{321}$	-2.833
$\beta_{322}$	-2.713
$\beta_{331}$	0.255
$\beta_{332}$	2.088
$\theta_1$	0.428
$\theta_2$	0.207
$\theta_3$	0.664

Tabela 51 – Resultados para o ajuste DiasChuva-UmidadeRelativa.

<b>Parâmetro</b>	<b>Estimado</b>
$\beta_{121}$	-1.302
$\beta_{122}$	0.545
$\beta_{131}$	-5.951
$\beta_{132}$	2.696
$\beta_{221}$	0.031
$\beta_{222}$	4.359
$\beta_{231}$	-0.209
$\beta_{232}$	4.449
$\beta_{321}$	1.103
$\beta_{322}$	-0.120
$\beta_{331}$	3.359
$\beta_{332}$	0.857
$\theta_1$	0.659
$\theta_2$	0.437
$\theta_3$	0.211

### 5.2.2.2 Algoritmo EM Estocástico

As estimativas obtidas para os parâmetros do modelo ajustado utilizando a covariável temperatura média mensal são mostradas na Tabela 52. Observamos um tempo de processamento menor a 1 segundo para este modelo, e o valor calculado do BIC foi de 4119.71.

Na Tabela 53 observamos os resultados do modelo ajustado utilizando a covariável umidade relativa média mensal. Para este ajuste, o tempo de processamento foi menor a 1 segundo. O valor calculado para o BIC foi de 3986.95.



Tabela 52 – Resultados para o ajuste DiasChuva-Temperatura.

Parâmetro	Estimado
$\beta_{121}$	-3.253
$\beta_{122}$	-3.849
$\beta_{131}$	-0.640
$\beta_{132}$	-0.351
$\beta_{221}$	2.082
$\beta_{222}$	-0.254
$\beta_{231}$	0.751
$\beta_{232}$	-0.198
$\beta_{321}$	-0.642
$\beta_{322}$	-1.818
$\beta_{331}$	0.043
$\beta_{332}$	1.124
$\theta_1$	0.670
$\theta_2$	0.236
$\theta_3$	0.497

Tabela 53 – Resultados para o ajuste DiasChuva-UmidadeRelativa.

Parâmetro	Estimado
$\beta_{121}$	0.041
$\beta_{122}$	-1.180
$\beta_{131}$	-0.911
$\beta_{132}$	-2.222
$\beta_{221}$	0.148
$\beta_{222}$	-2.294
$\beta_{231}$	0.269
$\beta_{232}$	-3.120
$\beta_{321}$	0.720
$\beta_{322}$	-1.022
$\beta_{331}$	-0.332
$\beta_{332}$	-3.341
$\theta_1$	0.645
$\theta_2$	0.403
$\theta_3$	0.203

### 5.2.2.3 Estimação Bayesiana

As condições do processo de estimação sob o enfoque Bayesiano foram mudadas para a situação com 3 estados ocultos. Isto para poder melhorar as probabilidades do algoritmo convergir. As condições são apresentadas a seguir:

- 110,000 iterações do algoritmo.
- 10,000 iterações de *burn-in*.

Tabela 54 – Resultados para o ajuste DiasChuva-Temperatura.

Parâmetro	Estimado	SD	IC95	Assimetria
$\beta_{121}$	3.366	0.760	1.992, 4.972	0.320
$\beta_{122}$	2.670	0.581	1.620, 3.903	0.338
$\beta_{131}$	1.844	1.709	-1.824, 4.839	-0.462
$\beta_{132}$	31.602	15.283	12.027, 69.896	1.400
$\beta_{221}$	2.096	0.326	1.500, 2.777	0.289
$\beta_{222}$	-1.102	0.269	-1.656, -0.601	-0.220
$\beta_{231}$	-0.045	0.425	-0.875, 0.782	-0.011
$\beta_{232}$	-0.512	0.344	-1.185, 0.158	-0.011
$\beta_{321}$	-3.317	0.992	-5.519, -1.626	-0.576
$\beta_{322}$	5.419	1.158	3.284, 7.831	0.254
$\beta_{331}$	2.091	0.240	1.646, 2.583	0.201
$\beta_{332}$	2.589	0.335	1.958, 3.273	0.192
$\theta_1$	0.369	0.007	0.354, 0.383	-0.034
$\theta_2$	0.195	0.005	0.185, 0.206	0.052
$\theta_3$	0.625	0.005	0.614, 0.635	0.003

- Intervalo de espaçamento entre amostras coletadas igual a 10.

Na Tabela 54 observamos as estimativas para os parâmetros do modelo ajustado utilizando a covariável temperatura média mensal. O tempo de processamento para este ajuste foi de 74.99 minutos e os valores para o BIC e DIC foram de 3376.82 e 3774.55 respectivamente. Também pode-se observar que alguns dos parâmetros tem intervalos de credibilidade mais largos, e desvios padrões relativamente grandes. Isto pode ser devido à pouca ocorrência da transição relacionada a esses parâmetros. Além disto, os intervalos de credibilidade para o parâmetros  $\beta_{232}$  e  $\beta_{331}$  incluem o valor 0. Isto nos indica que esses parâmetros possuem estimativas próximas do valor 0.

A Figura 27 mostra os gráficos de convergência para os parâmetros. Aparentemente, o amostrador visita o espaço paramétrico de forma aleatória, sem mostrar padrões ou tendências.

Na Figura 28 observamos os gráficos das densidades dos parâmetros. Alguns dos parâmetros mostram valores de assimetria elevados. Esse valores estão disponíveis na Tabela 54.

A Figura 29 mostra que as médias móveis dos parâmetros vão se aproximando à média *a posteriori* na medida que o algoritmo MCMC avança no processo de amostragem. Para alguns parâmetros esta aproximação à média *a posteriori* acontece mais lentamente do esperado, porém a cadeia de todos os parâmetros parece convergir.

Na Figura 30 observamos os gráficos de autocorrelação para o modelo ajustado com a covariável Temperatura. Notamos que as amostras de alguns dos parâmetros mostram correlação elevada com valores que se encontram ainda há 5 posições de distância no processo. Isto indica que os valores dos parâmetros correspondentes não estão sendo atualizados com boa dinâmica.

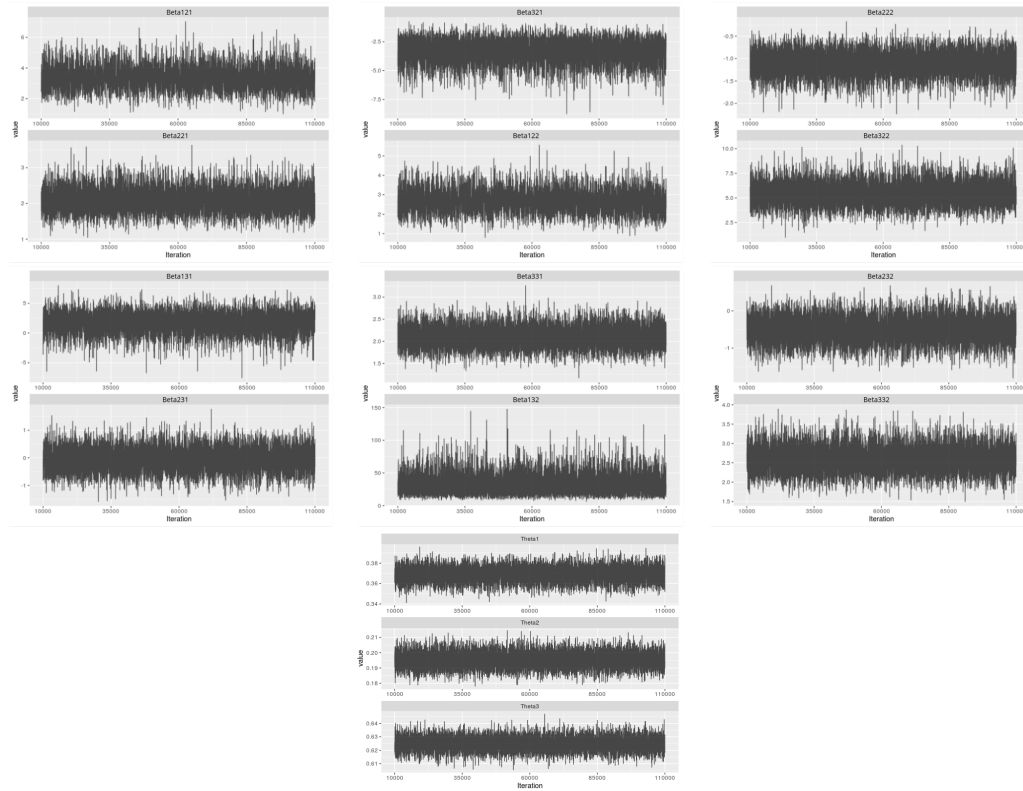


Figura 27 – Gráficos de convergência para os parâmetros do modelo DiasChuva-Temperatura.

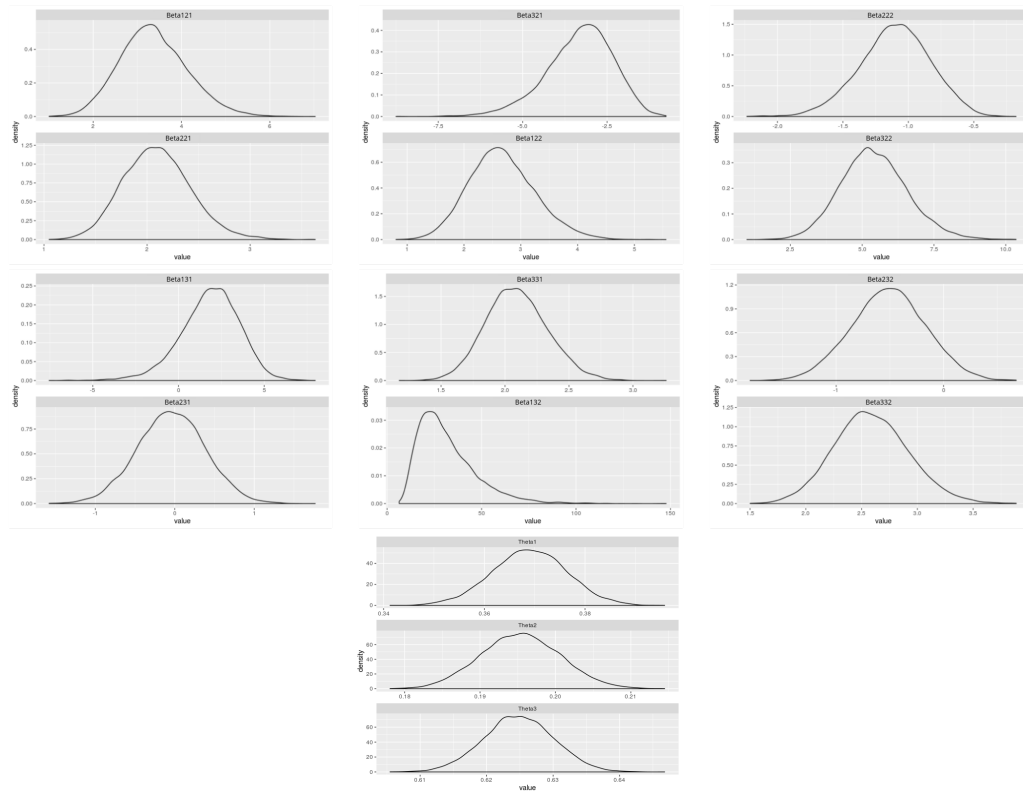


Figura 28 – Gráficos de densidade *a posteriori* para os parâmetros do modelo DiasChuva-Temperatura.

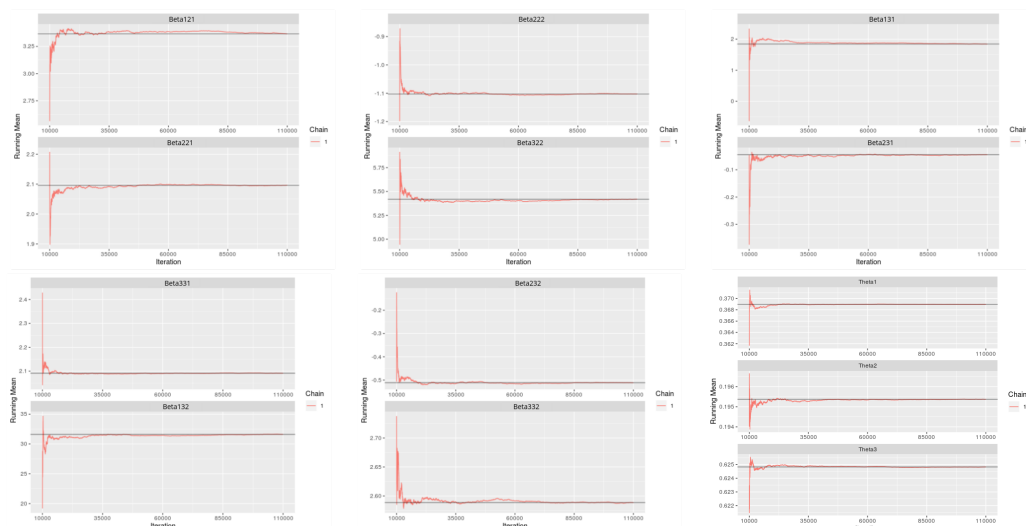


Figura 29 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva-Temperatura.

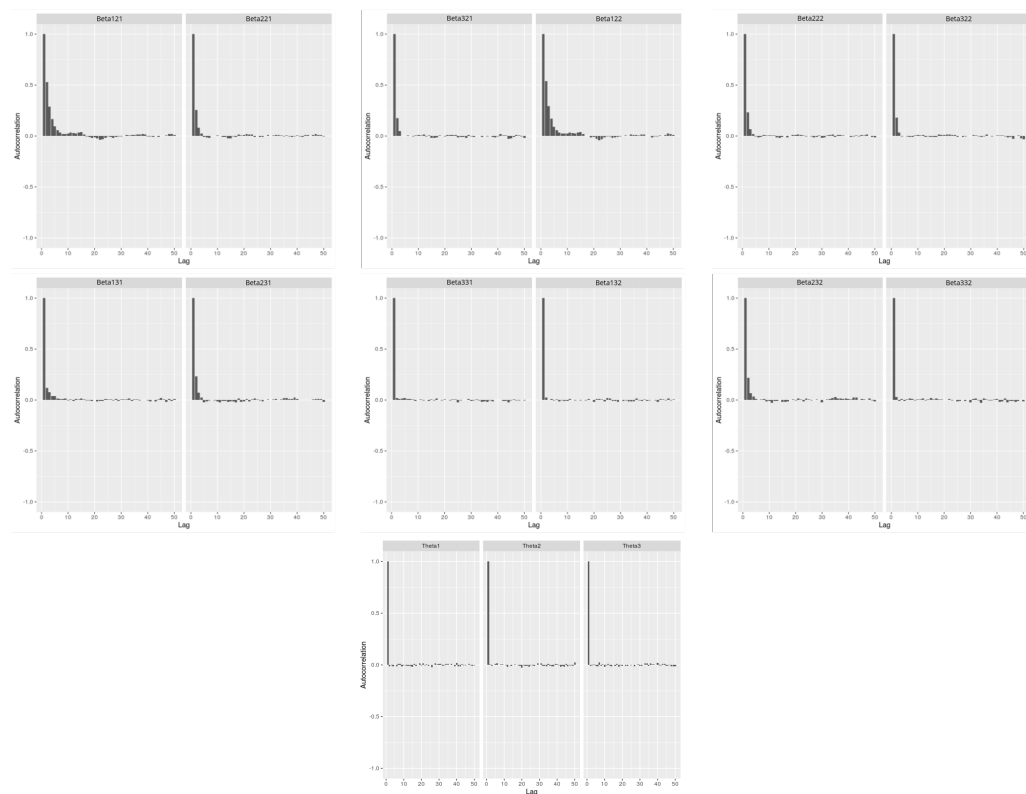


Figura 30 – Gráficos de autocorrelação para os parâmetros do modelo DiasChuva-Temperatura.

A Tabela 55 mostra as estimativas para os parâmetros do modelo ajustado utilizando a covariável umidade relativa média mensal. O tempo de processamento para este ajuste foi de 78.95 minutos e os valores para o BIC e DIC foram de 3392.27 e 3701.17 respectivamente. Também pode-se observar que alguns dos parâmetros relacionados com a transição ao terceiro estado oculto tem intervalos de credibilidade muito cumpridos e desvios padrões relativamente grandes. Isto pode ser devido à pouca ocorrência da transição relacionada a esses parâmetros.

Tabela 55 – Resultados para o ajuste DiasChuva-UmidadeRelativa.

Parâmetro	Estimado	SD	IC95	Assimetria
$\beta_{121}$	-0.301	0.200	-0.689, 0.092	0.001
$\beta_{122}$	1.463	0.263	0.960, 1.997	0.151
$\beta_{131}$	-1.705	0.613	-3.015, -0.608	-0.388
$\beta_{132}$	-1.055	0.259	-1.582, -0.560	-0.092
$\beta_{221}$	-1.951	0.454	-2.887, -1.092	-0.162
$\beta_{222}$	-4.699	1.161	-7.273, -2.688	-0.490
$\beta_{231}$	-84.861	41.194	-175.553, -20.286	-0.684
$\beta_{232}$	-2.614	0.876	-4.547, -1.098	-0.495
$\beta_{321}$	-14.146	3.174	-21.107, -8.840	-0.616
$\beta_{322}$	-93.991	44.791	-193.584, -23.836	-0.698
$\beta_{331}$	-10.841	1.811	-14.769, -7.747	-0.546
$\beta_{332}$	-16.070	3.011	-22.742, -10.941	-0.564
$\theta_1$	0.643	0.005	0.633, 0.654	0.011
$\theta_2$	0.339	0.006	0.327, 0.351	0.052
$\theta_3$	0.180	0.006	0.168, 0.192	0.057

Na Figura 31 pode-se ver os gráficos de convergência para os parâmetros. Aparentemente, o amostrador visita o espaço paramétrico de forma aleatória para a maioria dos parâmetros, porém, para alguns parâmetros os gráficos parecem sugerir que a cadeia não é devidamente dinâmica.

A Figura 32 mostra os gráficos das densidades *a posteriori* dos parâmetros. Alguns dos parâmetros mostram valores de assimetria elevados e estes estão disponíveis na Tabela 55.

Na Figura 33 observamos que as médias dos parâmetros vão se aproximando à média *a posteriori* na medida que o algoritmo MCMC avança no processo de amostragem. Para alguns parâmetros esta aproximação à média *a posteriori* acontece mais lentamente do que esperado, porém todos os parâmetros apresentam convergência.

Na Figura 34 mostra os gráficos de autocorrelação dos parâmetros para o modelo ajustado com a covariável Temperatura. Observamos que as amostras de alguns parâmetros apresentam correlação elevada com amostras que se encontram há 12 posições de distância no processo de amostragem.

A Tabela 56 apresenta os resultados dos valores do BIC para todos os modelos que foram ajustados. Para escolher o modelo mais adequado para o caso de aplicação, utilizamos o valor do BIC como critério. O modelo ajustado que apresentou o menor BIC foi o modelo estimado pelo algoritmo Bayesiano, utilizando a covariável temperatura média mensal com 3 estados ocultos. Sendo que este é o modelo escolhido, então a matriz de coeficientes de transição estimada é dada

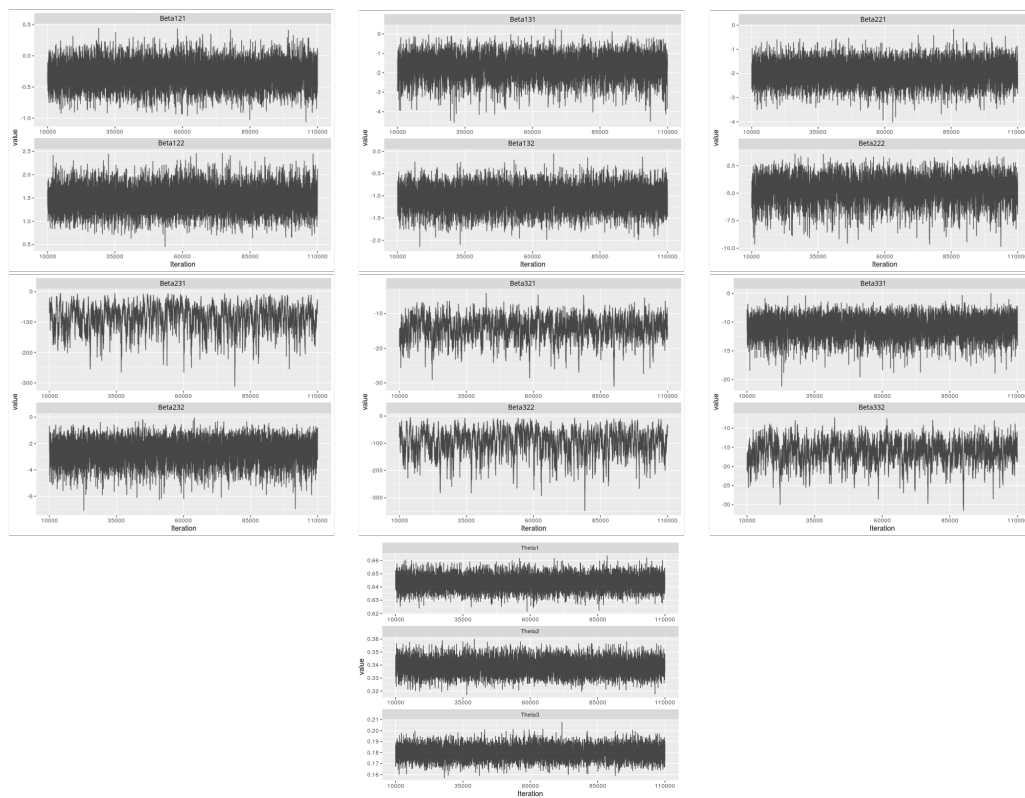


Figura 31 – Gráficos de convergência para os parâmetros do modelo DiasChuva-UmididadeRelativa.

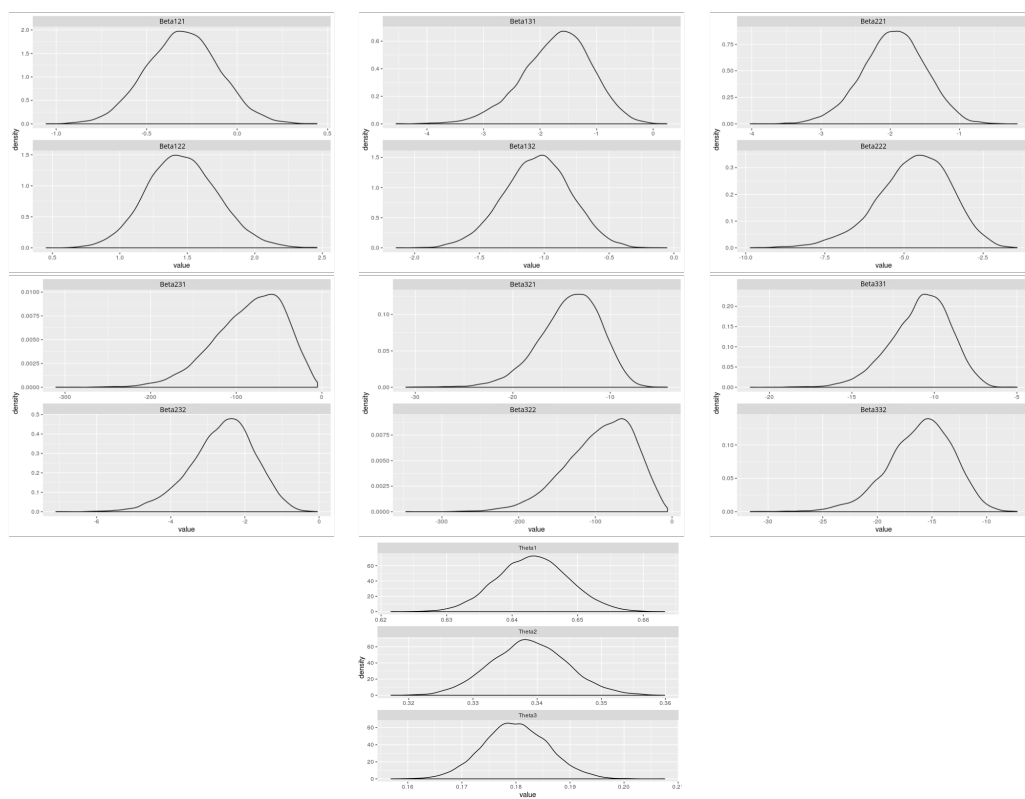


Figura 32 – Gráficos de densidade *a posteriori* para os parâmetros do modelo DiasChuva-UmididadeRelativa.

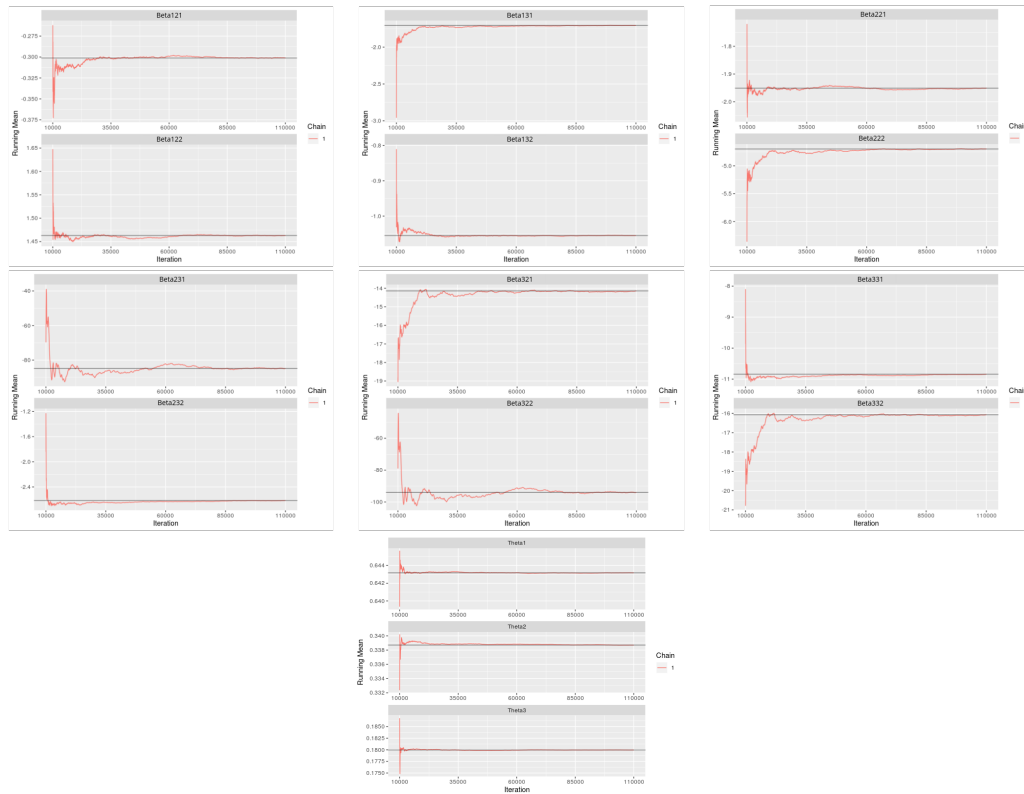


Figura 33 – Gráficos de médias móveis dos valores dos parâmetros para o modelo DiasChuva- UmidadeRelativa.

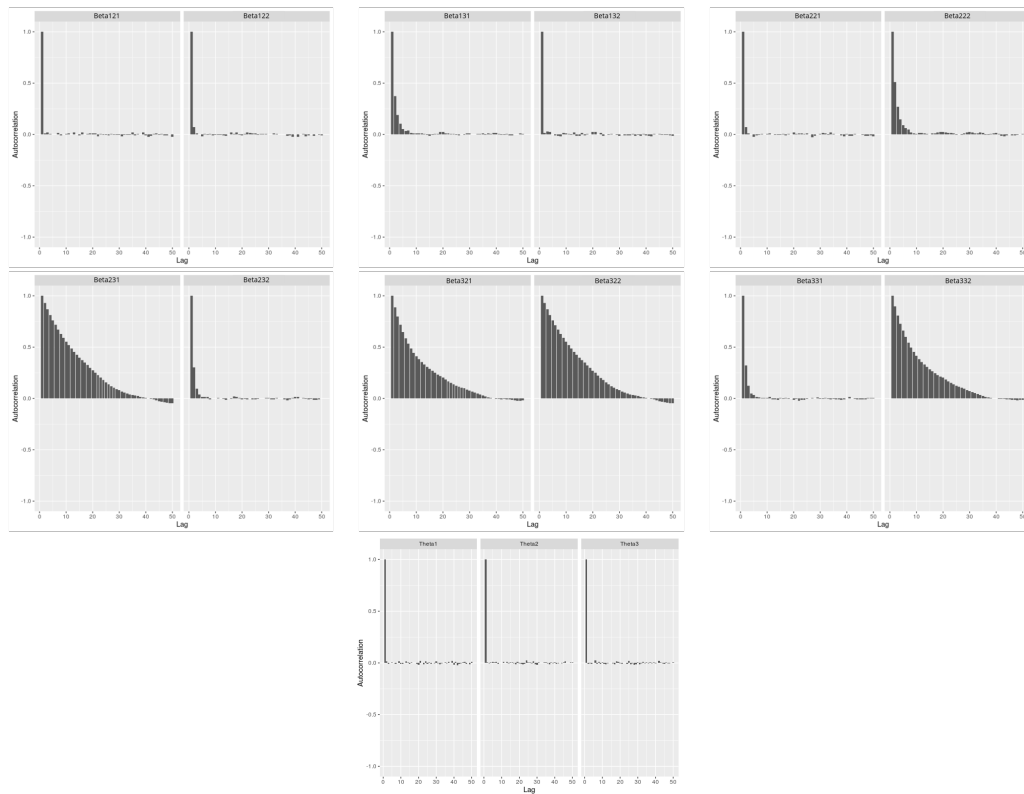


Figura 34 – Gráficos de autocorrelação para os parâmetros do DiasChuva-Temperatura.

Tabela 56 – Comparação de BIC entre modelos ajustados.

Covariável	Algoritmo	K	BIC
Temperatura	EM	2	3739.82
Temperatura	EM Estocástico	2	4344.21
Temperatura	Bayesiano	2	3726.82
Umidade Relativa	EM	2	3887.42
Umidade Relativa	EM Estocástico	2	4553.06
Umidade Relativa	Bayesiano	2	3833.84
Temperatura	EM	3	3583.10
Temperatura	EM Estocástico	3	4119.71
Temperatura	Bayesiano	3	3376.82
Umidade Relativa	EM	3	3733.80
Umidade Relativa	EM Estocástico	3	3986.95
Umidade Relativa	Bayesiano	3	3392.27

pela Equação (5.1)

$$\boldsymbol{\beta} = \begin{bmatrix} (0) & (3.366) & (1.844) \\ (0) & (2.096) & (-0.045) \\ (0) & (-3.317) & (2.091) \end{bmatrix} \cdot \begin{bmatrix} (0) \\ (0) \\ (0) \end{bmatrix} \cdot \begin{bmatrix} (31.602) \\ (-0.512) \\ (2.589) \end{bmatrix}. \quad (5.1)$$

Para verificar se a matriz de coeficientes de transição estimada é coerente com o fenômeno sendo estudado, fazemos uma análise gráfica similar à que foi feita no Capítulo 4. No Capítulo 4, a análise gráfica foi feita para escolher coeficientes de transição adequados para cada situação de simulação. Porém, no estudo de aplicação, estes gráficos são usados para corroborar a coerência da matriz de transição estimada.

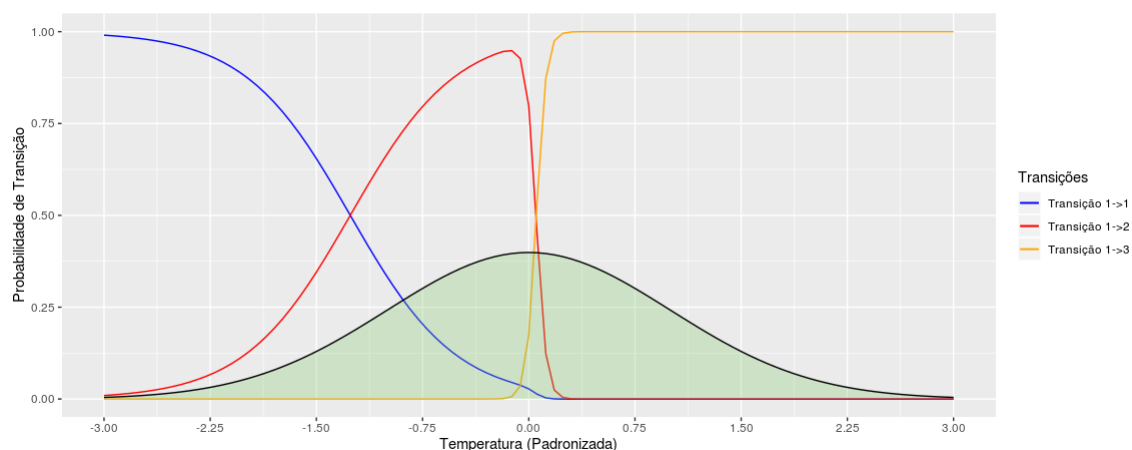


Figura 35 – Funções de transição do estado 1

Na Figura 35 observamos as funções de transição para o estado 1. Este estado gera observações com distribuição Binomial(30,0.369). O gráfico mostra que para valores muito menores que a média das temperaturas mensais observadas há uma probabilidade grande de



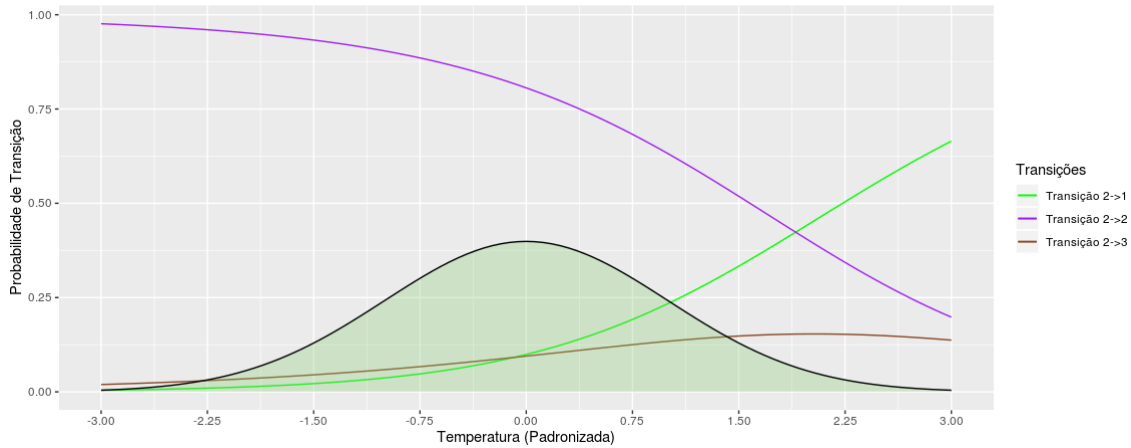


Figura 36 – Funções de transição do estado 2

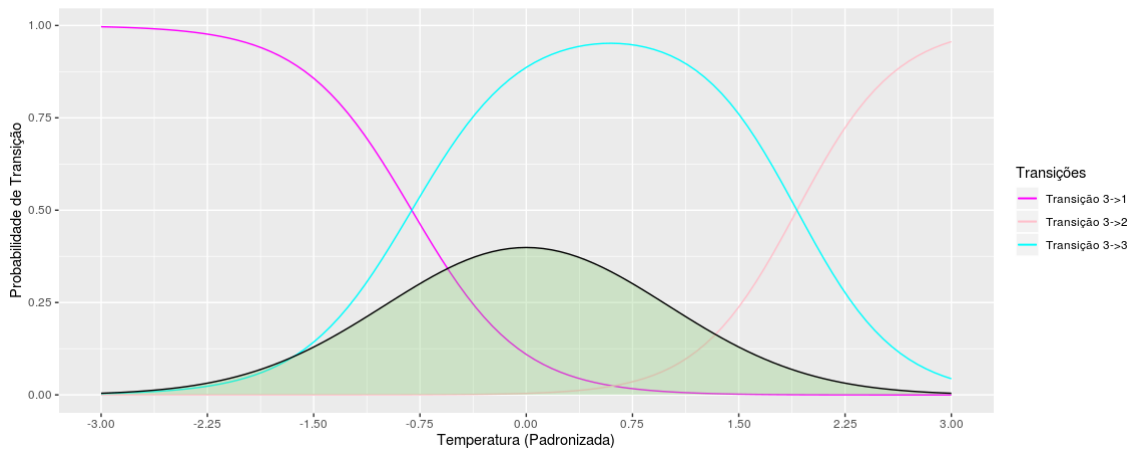


Figura 37 – Funções de transição do estado 3

continuar no estado 1. Na medida que os valores da temperatura se aproximam à media observada das temperaturas mensais desde o extremo negativo da reta real, então as probabilidades de transicionar do estado 1 para o estado 2 incrementam. O estado 2 gera observações com distribuição Binomial(30,0.195). Isto indica que com temperaturas menores que a media das temperaturas mensais, é mais provável ir de um estado com probabilidade de chuva intermediária a um estado com probabilidade de pouca chuva. Quando as temperaturas atingem valores maiores que a media observada das temperaturas mensais, então a probabilidade de transicionar do estado 1 para o estado 3 aumenta. O estado 3 gera observações com distribuição Binomial(30,0.625). Isto indica que com temperaturas maiores que a media das temperaturas mensais, é mais provável ir de um estado com probabilidade de chuva intermediária para um estado com alta probabilidade de chuva.

Na Figura 36 observamos as funções de transição do estado 2. Este estado gera observações com distribuição Binomial(30,0.195). O gráfico mostra que para a maioria dos valores da temperatura mensal há uma probabilidade grande de continuar no estado 2. Isto indica que ao estar num estado com pouca probabilidade de chuva, o mais provável é continuar nesse estado

com pouca probabilidade de chuva. Para valores de temperatura muito maiores que a média observada das temperaturas mensais, as probabilidades de transicionar do estado 2 para o estado 1 incrementam. O estado 1 gera observações com distribuição Binomial(30,0.369). Isto indica que com temperaturas muito altas, é mais provável ir de um estado com pouca probabilidade de chuva para um estado com probabilidade de chuva intermediária. Além disto, para todos os valores possíveis de temperaturas mensais sempre há probabilidades baixas de transicionar do estado 2 ao estado 3. O estado 3 gera observações com distribuição Binomial(30,0.625). Isto nos indica que para qualquer valor da temperatura mensal, é pouco provável transicionar de um estado com pouca probabilidade de chuva para um estado com alta probabilidade de chuva.

Na Figura 37 observamos as funções de transição do estado 3. Este estado gera observações com distribuição Binomial(30,0.625). O gráfico mostra que para valores próximos à média das temperaturas mensais observadas há uma probabilidade alta de continuar no estado 3. Isto indica que ao estar num estado com probabilidade alta de chuva, o mais provável é continuar nesse estado com probabilidade alta de chuva se não aconteceram valores extremos de temperaturas mensais. Para valores de temperatura muito menores que a média observada das temperaturas mensais, as probabilidades de transicionar do estado 3 para o estado 1 incrementam. O estado 1 gera observações com distribuição Binomial(30,0.369). Isto indica que com temperaturas muito baixas, é mais provável ir de um estado com probabilidade alta de chuva para um estado com probabilidade de chuva intermediária. Para valores muito altos da temperatura mensal, há uma probabilidade alta de transicionar do estado 3 para o estado 2. O estado 2 gera observações com distribuição Binomial(30,0.195). Isto indica que com temperaturas muito altas, é mais provável ir de um estado com probabilidade alta de chuva para um estado com probabilidade baixa de chuva. O comportamento das funções de transição para os 3 estados ocultos descrito anteriormente é coerente com o fenômeno observado na natureza.

Um gráfico interessante para analisar o modelo selecionado é o gráfico que mostra a sequência de estados ocultos predita pelo algoritmo junto com os valores observados de dias com chuva em cada mês ao longo do tempo. Esse gráfico nos mostra o comportamento das transições entre estados ocultos e, assim, podemos verificar se ela tem coerência com os valores observados.

Os gráficos 38 até 44 mostram a sequência de estados ocultos que foi prevista junto com os valores observáveis. O valor da quantidade de dias com chuva observados num certo mês é mostrado por linhas contínuas com cores diferentes para cada ano. O estado oculto previsto em cada mês é mostrado como um círculo que representa em geral os meses com probabilidade moderada de chuva, triângulo que representa o período com menor probabilidade de chuva ou quadrado que em geral representa o período com maior probabilidade de chuva. Nos gráficos observamos que para a maioria dos meses no intervalo de tempo estudado, o valor do estado oculto tem coerência com o valor de dias com chuva observado. Pelas estimativas, o estado oculto 1 gera observações com distribuição *Binomial*(30,0.369), o estado oculto 2 gera observações

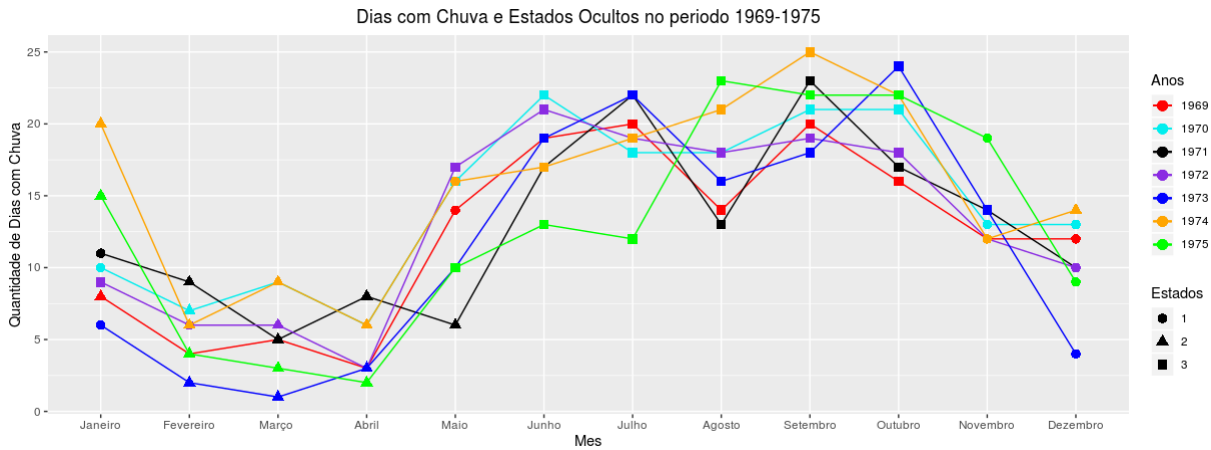


Figura 38 – Sequência de estados ocultos e valores observados no período 1969-1975.

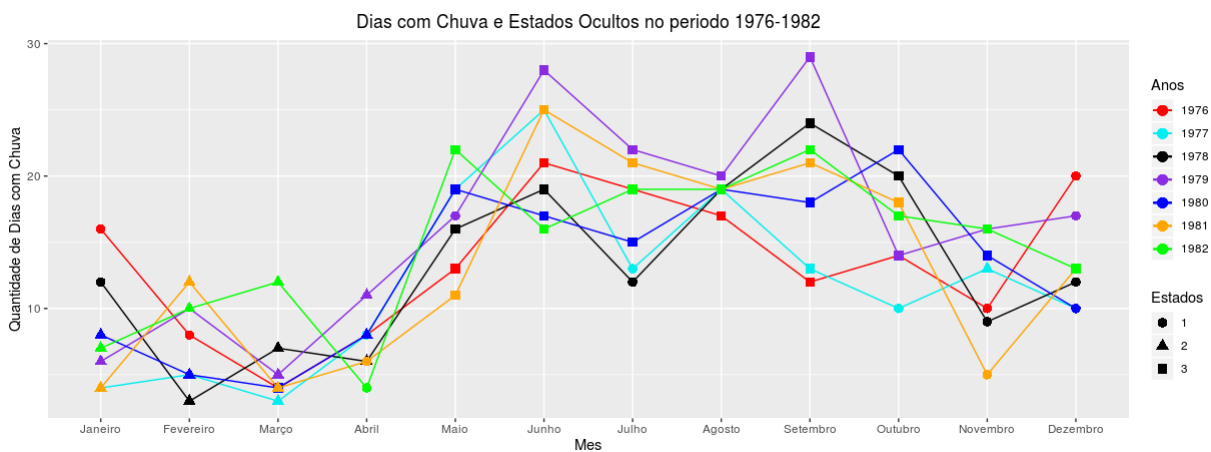


Figura 39 – Sequência de estados ocultos e valores observados no período 1976-1982.

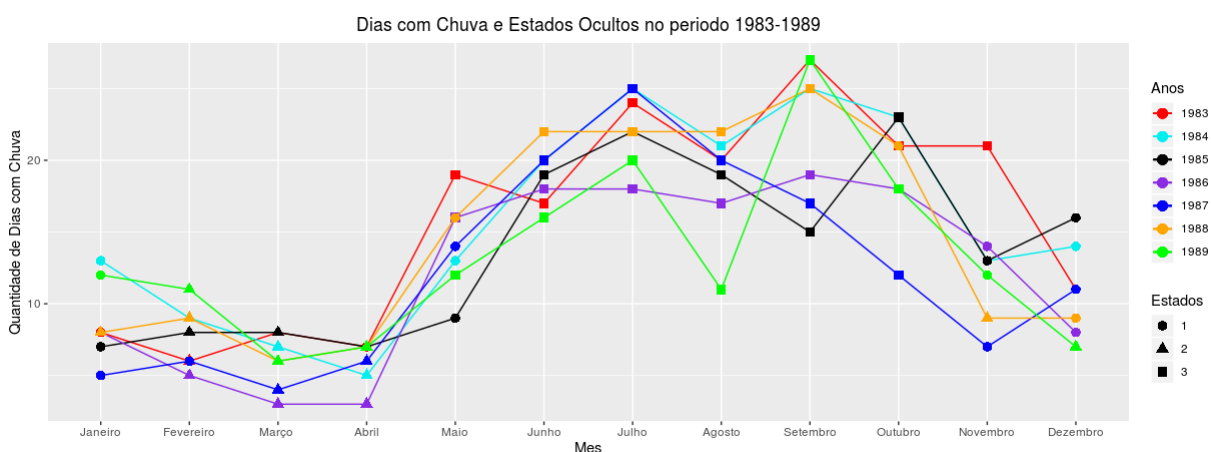


Figura 40 – Sequência de estados ocultos e valores observados no período 1983-1989.

com distribuição  $Binomial(30, 0.195)$  e o estado oculto 3 gera observações com distribuição  $Binomial(30, 0.625)$ .

Podemos utilizar esta sequência prevista de estados ocultos para avaliar os motivos pelos

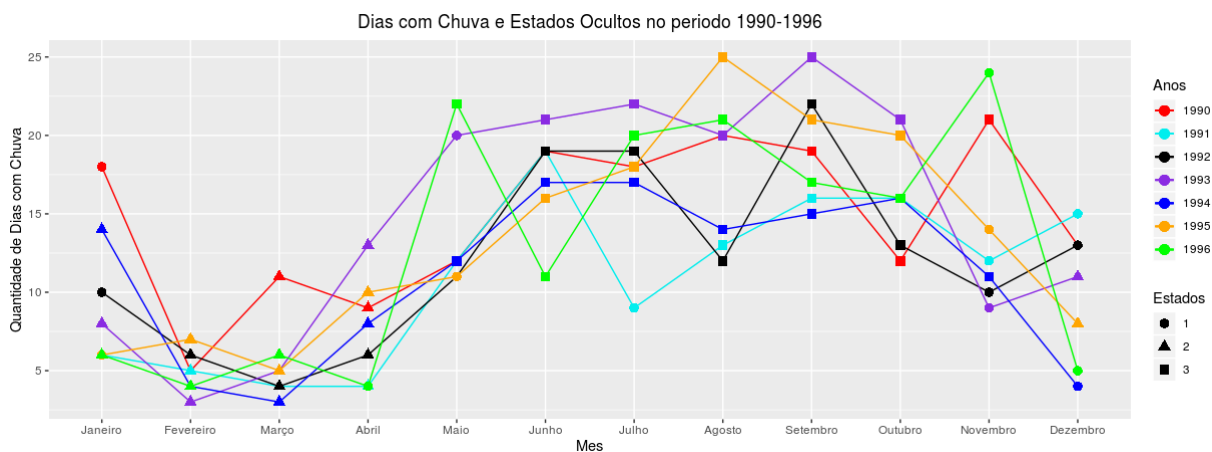


Figura 41 – Sequência de estados ocultos e valores observados no período 1990-1996.

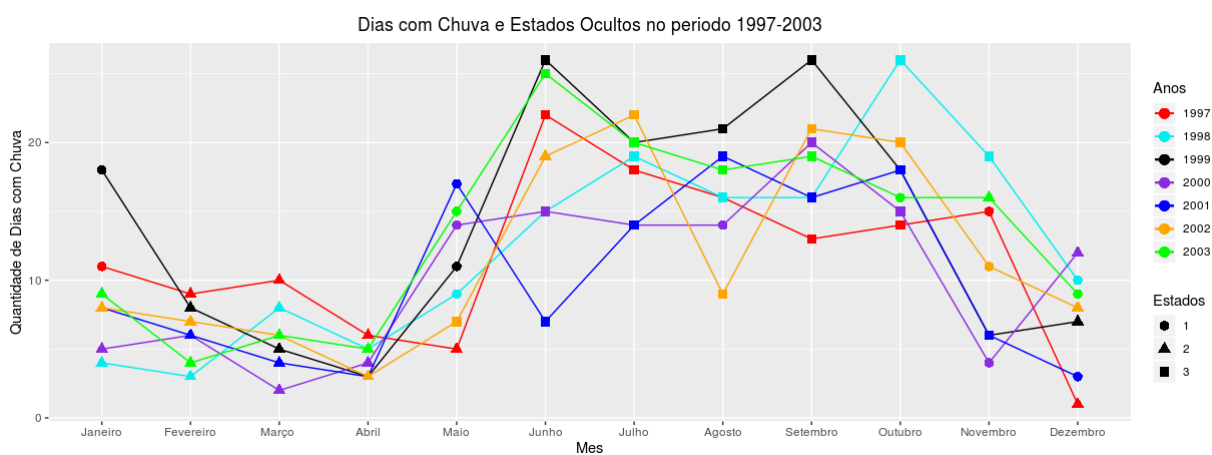


Figura 42 – Sequência de estados ocultos e valores observados no período 1997-2003.

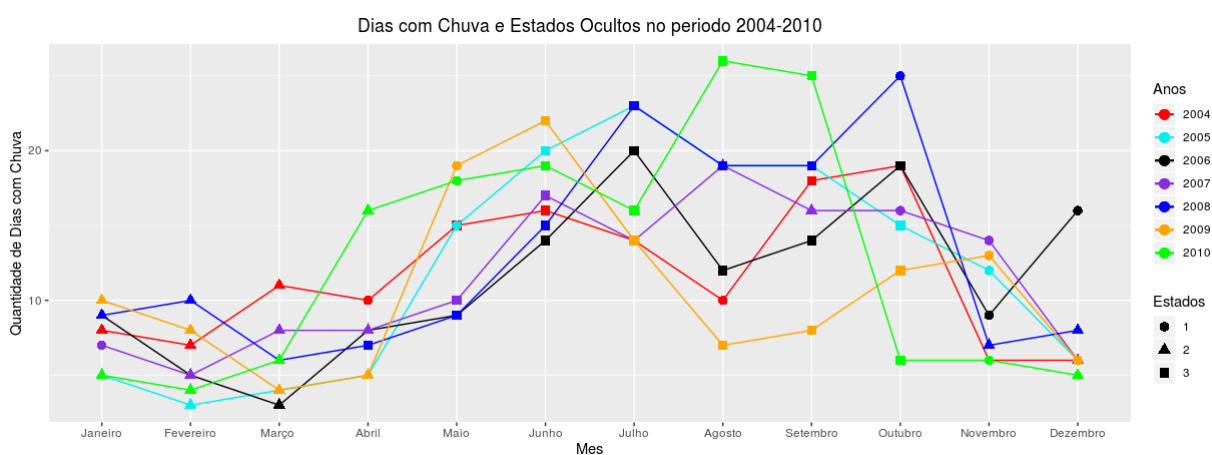


Figura 43 – Sequência de estados ocultos e valores observados no período 2004-2010.

quais observamos desvios padrões e intervalos de credibilidade de tamanho notavelmente maior do que os outros parâmetros no parâmetro  $\beta_{322}$ . Aparentemente, seguindo a sequência de estados ocultos, essa transição não aconteceu a quantidade de vezes suficiente para que o processo

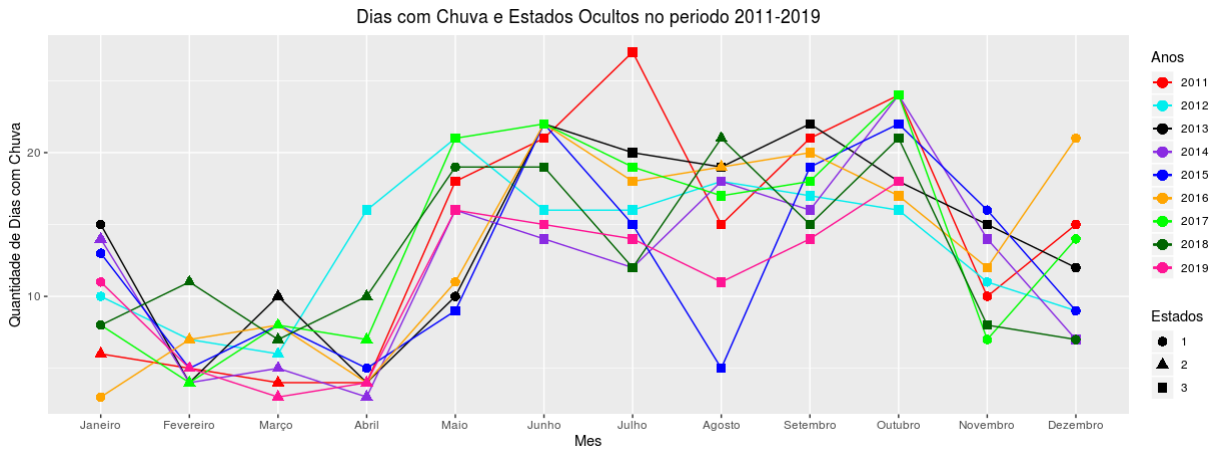


Figura 44 – Sequência de estados ocultos e valores observados no período 2011-2019.

estimasse esse coeficiente de transição de forma precisa. A matriz

$$\begin{bmatrix} 49 & 37 & 55 \\ 53 & 129 & 5 \\ 39 & 22 & 220 \end{bmatrix} \quad (5.2)$$

mostra a quantidade de vezes que acontece cada uma das transições possíveis a partir do modelo estimado. Para ilustrar, o valor que encontramos na linha 1 e coluna 1 corresponde à quantidade de vezes que observamos a permanência no estado 1 do instante  $t - 1$  para o  $t$ . Apenas 5 transições do estado 2 para 3 consecutivamente foram observadas. Isso faz sentido pois dificilmente migramos tão abruptamente de um mês com baixa probabilidade de chuva para um com alta probabilidade. Uma observação interessante é o fato de que as estimativas de alguns parâmetros de transição para os quais observamos desvio padrão relativamente grande apresentam contagens pequenas da transição à que estão relacionados. A Matriz 5.2 também ajuda a verificar a coerência das interpretações feitas anteriormente sobre os coeficientes de transição. De acordo com os coeficientes de transição estimados, é pouco provável transicionar abruptamente entre estados como probabilidades de chuva extremas. Em outras palavras, é pouco provável transicionar de um estado com pouca probabilidade de chuva a um estado com alta probabilidade de chuva. O oposto também é pouco provável.

Percebemos nas análises que os meses de fevereiro, março e abril geralmente apresentam probabilidade baixa de chuva, pois seus estados preditos são na maioria o 2. Os meses de junho a outubro são predominantemente do estado 3, com mais alta probabilidade de chuva. Os meses de novembro, dezembro e maio geralmente são mais heterogêneos ao decorrer dos anos, com diferentes probabilidades de chuva. Nesses meses observamos a ocorrência dos 3 estados ocultos, porém, parece que o estado que acontece mais é o estado 1, que apresenta moderada probabilidade de chuva.

De forma geral, o modelo ajustado usando o algoritmo Bayesiano como  $K = 3$  estados ocultos e a covariável temperatura média mensal é o modelo que melhor descreve o fenômeno

de dias com chuva em cada mês na cidade de Tegucigalpa, Honduras.



---

## CONSIDERAÇÕES FINAIS E PROPOSTAS FUTURAS

---

Nesse trabalho apresentamos os modelos Markovianos ocultos não-homogêneos. Explícamos suas características, elementos e funcionamento. Estudamos e apresentamos 3 diferentes algoritmos de estimação para os modelos Markovianos não-homogêneos: dois destes algoritmos foram apresentados sob a perspectiva de estimação frequentista e um sob a perspectiva Bayesiana. Na visão frequentista discutimos o algoritmo EM e o EM estocástico. O algoritmo apresentado sob a perspectiva de estimação Bayesiano foi estimação Monte Carlo via cadeias de Markov.

Os três algoritmos foram aplicados em dados simulados e, nas situações consideradas, os resultados foram favoráveis ao método Bayesiano, que apesar de apresentar um tempo maior de processamento, apresentou menores valores de erro quadrático médio, viés, desvio padrão e predição da sequência de estados ocultos mais estável. O método Bayesiano também tem a vantagem de oferecer facilmente estimativas intervalares para os parâmetros, que seriam possíveis nos algoritmos frequentistas através apenas de teoria assintótica.

O algoritmo EM estocástico também apresentou bons resultados e estimativas próximas às estimativas Bayesianas. Tem a vantagem de ser de mais fácil programação e exigir pouco esforço computacional.

O algoritmo EM apresenta a vantagem de já estar disponível em pacote no R, mas apresentou resultados bem inconsistentes entre as réplicas utilizadas no estudo de simulação, apesar de apresentar convergência de acordo com o critério implementado no pacote. As diferentes réplicas apresentam estimações bem diferentes para o mesmo modelo.

Os três algoritmos de estimação apresentados nesse trabalho consideram que o número de estados ocultos é conhecido, mas na prática ele dificilmente será. Aqui sugerimos a utilização do BIC e DIC para seleção do número mais adequado de estados.

As metodologias foram aplicadas em dados climatológicos da cidade capital de Honduras,



Tegucigalpa, com o objetivo principal de identificar diferentes padrões no regimento de chuvas da cidade. Modelos foram ajustados utilizando 2 covariáveis de forma separada, considerando 2 e 3 estados ocultos. O modelo que foi considerado como o que melhor descreve o fenômeno de dias com chuva por mês na cidade de Tegucigalpa foi o modelo ajustado com o algoritmo Bayesiano, com  $K = 3$  estados ocultos usando a covariável temperatura média mensal como preditora na transição entre os estados ocultos.

Para estudos futuros, algumas propostas são a extensão do modelo NHMM para incluir mais covariáveis no cálculo das probabilidades de transição entre estados ocultos. Nesse trabalho, as metodologias já foram descritas permitindo um número maior de covariáveis, mas suas implementações (exceto o EM que já está disponibilizado na sua versão estendida) consideram apenas uma covariável. Isto permitirá que as transições entre estados ocultos sejam explicadas por mais fatores externos.

Uma outra alternativa aos critérios de seleção para escolha do número de estados ocultos é o método *Reversible Jump* que inclui essa característica como outro parâmetro no modelo e o estima simultaneamente à estimação do modelo.

Outro aspecto que pode ser explorado é a capacidade preditiva desse modelo e dos algoritmos apresentados. Em muitas situações práticas, a predição de observações futuras, como por exemplo, a quantidade de dias chuvosos nos próximos meses, é de maior interesse do que a estimação dos parâmetros.

## REFERÊNCIAS

---

---

BHAT, H. S.; KUMAR, N. On the derivation of the bayesian information criterion. **School of Natural Sciences, University of California**, 2010. Citado na página 53.

CORO, G. **A lightweight guide on Gibbs sampling and JAGS**. [S.l.], 2013. Citado na página 50.

EDDY, S. R. What is a hidden Markov model? **Nat Biotech**, Nature Publishing Group, v. 22, n. 10, p. 1315–1316, out. 2004. ISSN 10870156. Disponível em: <<http://dx.doi.org/10.1038/nbt1004-1315>>. Citado na página 25.

ESPINDOLA, L. **Um Estudo sobre Modelos Ocultos de Markov**. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2009. Citado nas páginas 21, 28, 31, 37, 38 e 40.

GAO, B.; PAVEL, L. On the properties of the Softmax function with application in game theory and reinforcement learning. 04 2017. Citado nas páginas 35 e 51.

HOLSCLAW, T.; GREENE, A.; ROBERTSON, A.; SMYTH, P. Bayesian non-homogeneous markov models via polya-gamma data augmentation with applications to rainfall modeling. **The Annals of Applied Statistics**, v. 11, 01 2017. Citado nas páginas 22, 32 e 77.

JELINEK, F. **Statistical Methods for Speech Recognition**. Cambridge, MA, USA: MIT Press, 1997. ISBN 0-262-10066-5. Citado nas páginas 31 e 42.

LAGONA, F.; MARUOTTI, A.; PICONE, M. A non-homogeneous hidden Markov model for the analysis of multi-pollutant exceedances data. **Hidden Markov Models, Theory and Applications**, InTech Rijeka, Croatia, p. 207–222, 2011. Citado nas páginas 22 e 32.

NEAL, R. M. Slice sampling. **Annals of statistics**, JSTOR, p. 705–741, 2003. Citado na página 49.

PAPASTAMOULIS, P. Handling the label switching problem in latent class models via the ECR algorithm. **Communications in Statistics: Simulation and Computation**, v. 43, 01 2014. Citado na página 50.

PENNONI, F.; GENGE, E. Analysing the course of trust towards public and financial institutions via hidden Markov models. **SSRN Electronic Journal**, 01 2019. Citado na página 22.

PLUMMER, M. *et al.* JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: VIENNA, AUSTRIA. **Proceedings of the 3rd international workshop on distributed statistical computing**. [S.l.], 2003. v. 124, n. 125.10, p. 1–10. Citado na página 50.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>. Citado nas páginas 23, 44 e 48.

- RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. In: **Proceedings of the IEEE**. [S.l.: s.n.], 1989. p. 257–286. Citado nas páginas 25 e 30.
- ROSS, S. M. **Introduction to Probability Models, Ninth Edition**. Orlando, FL, USA: Academic Press, Inc., 2006. ISBN 0125980620. Citado na página 25.
- SARKAR, A.; BHADRA, A.; MALLICK, B. K. Nonparametric Bayesian approaches to non-homogeneous hidden Markov models. **arXiv preprint arXiv:1205.1839**, 05 2012. Citado na página 32.
- SOUZA, F. L. de. **Mistura de Distribuições Extremas**. Dissertação (Mestrado) — Universidade de Brasília, 2010. Citado na página 52.
- SPIEGELHALTER, D.; BEST, N.; CARLIN, B.; LINDE, A. The deviance information criterion: 12 years on. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 76, 03 2014. Citado na página 53.
- STEWART, W. J. **Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling**. Princeton, NJ, USA: Princeton University Press, 2009. ISBN 0691140626, 9780691140629. Citado na página 27.
- VISSER, I.; SPEEKENBRINK, M. depmixS4: An R package for hidden Markov models. **Journal of Statistical Software**, v. 36, n. 7, p. 1–21, 2010. Disponível em: <<http://www.jstatsoft.org/v36/i07/>>. Citado nas páginas 37 e 44.
- WOLFE, J.; JIN, X.; BAHR, T.; HOLZER, N. Application of Softmax regression and its validation for spectral-based land cover mapping. **ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, XLII-1/W1, p. 455–459, 05 2017. Citado na página 52.
- ZAMANZAD, F.; CLAESEN, J.; BURZYKOWSKI, T. A nonhomogeneous hidden Markov model for gene mapping based on next-generation sequencing data. **Journal of Computational Biology : a Journal of Computational Molecular Cell Biology**, v. 22, 01 2015. Citado na página 22.
- ZUANETTI, D. A. **Modelos HMM com dependência de segunda ordem: aplicação em genética**. Dissertação (Mestrado) — Universidade Federal de São Carlos, São Carlos, 2006. Citado nas páginas 25 e 30.
- ZUANETTI, D. A.; MILAN, L. A. A generalized mixture model applied to diabetes incidence data. **Biometrical Journal**, v. 59, n. 4, p. 826–842, 2017. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201600086>>. Citado na página 22.
- ZUCCHINI, W.; MACDONALD, I.; LANGROCK, R. **Hidden Markov Models for Time Series: An Introduction Using R, Second Edition**. CRC Press, 2017. (Chapman & Hall/CRC Monographs on Statistics and Applied Probability). ISBN 9781482253849. Disponível em: <<https://books.google.com.br/books?id=KIWzDAAAQBAJ>>. Citado na página 27.

