

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

**Inferência da conectividade neuronal via estimação de  
medidas da teoria da informação**

**Victor Augusto Girardi**

**Trabalho de Conclusão de Curso**



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Inferência da conectividade neuronal via estimação de medidas  
da teoria da informação

**Victor Augusto Girardi**

**Orientador: Prof. Dr. Ricardo Felipe Ferreira**

**Coorientador: Dr. Rodrigo Felipe de Oliveira Pena**

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade Federal de São Carlos - DEs-UFSCar, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

**São Carlos**  
**Junho de 2021**



FEDERAL UNIVERSITY OF SÃO CARLOS  
EXACT AND TECHNOLOGY SCIENCES CENTER  
DEPARTMENT OF STATISTICS

Neural connectivity inference from the estimation of  
information theory measures

**Victor Augusto Girardi**

**Advisor: Prof. Dr. Ricardo Felipe Ferreira**

**Co-advisor: Dr. Rodrigo Felipe de Oliveira Pena**

Bachelors dissertation submitted to the Department of Statistics, Federal University of São Carlos - DEs-UFSCar, in partial fulfillment of the requirements for the degree of Bachelor in Statistics.

**São Carlos**

**June 2021**



Victor Augusto Girardi

Inferência da conectividade neuronal via estimação de medidas  
da teoria da informação

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Victor Augusto Girardi e aprovado pela banca examinadora.

Aprovado em 8 de junho de 2021.

Banca Examinadora:

- Prof. Dr. Ricardo Felipe Ferreira
- Prof. Dr. Alessandro Giacomo Grimbert Gallo
- Profa. Dra. Vera Lucia Damasceno Tomazella



*Dedico este trabalho aos meus pais que me deram o dom da vida.*



# Agradecimentos

Como todos os desafios que enfrentamos em nossa vida, somente conseguimos superá-los com o auxílio da nossa família e de nossos amigos. Por conta disso, trago aqui os agradecimentos para todos aqueles que fazem parte da minha vida.

Primeiramente, gostaria de agradecer ao meu orientador Ricardo Ferreira que compartilhou comigo sua pesquisa e conhecimento. Além disso, é um excelente professor, mentor, pesquisador, amigo e uma referência pessoal minha. Posteriormente, o mais justo é o agradecimento ao meu coorientador Rodrigo Pena que me iluminou com seu vasto conhecimento de neurociência, uma paciência imensa para ensinar e conselhos incríveis. Com certeza, o trabalho não existiria sem a presença de Ricardo e Rodrigo.

É de suma importância agradecer minha família. À minha mãe Cristiane e ao meu Pai Marcelo, os quais esse trabalho é dedicado, por serem minhas inspirações diárias e detentores de um vasto amor. Ao meu irmão mais velho Pedro, pelas conversas profundas e conselhos sobre decisões de vida. Ao meu irmão mais novo Thiago, pela parceria de anos e o constante lembrete de que a alma não vive só de trabalho. À minha avó Suely e seu marido Klaus, pelos ensinamentos profundos. À minha avó Maria Aparecida, que é um exemplo de vida para mim. Aos meus tios Paulo e Sandro, que me ajudaram e apoiaram desde que nasci. Às minhas tias Ana Paula, Jane, Fernanda e Patrícia que me acolheram e socorreram sempre. Por fim, meus primos e primas que são responsáveis por diversos momentos de diversão juntos.

Agradeço a minha segunda família: República Gato Mia. Principalmente, Bruno, Danilo, Eric, Fábio, Fernando, Lucas, Matheus Palumbo, Matheus Travaglini, Matheus Von, Murilo e Rafael, que viveram todos os dias dos meus 4 anos de graduação.

Aos grandes amigos que alimentaram minha vida com muita alegria e felicidades. Seria injusto não citar meus amigos de longa data Giovanna Dutra, Pedro Paulo Spagnolo, Vitor Potiens, Pedro Mizusaki e Julie Delfortrie, que estão ao meu lado há anos e espero que continuem por mais.

Agradeço a todos do Departamento de Estatística da Universidade Federal de São Carlos, principalmente Rafaela Marques, Lais Sebastiany, Angelo Chigiarelli, Júlia Bel-

tramini e Daiane Zuanetti. Como também, tenho que agradecer Júlia Bellucci e Pedro Marques.

Em especial, quero agradecer Letícia Aparecida da Silva, pelas longas conversas, incrível paciência, imenso incentivo e companhia essencial.

Reconheço também todas as pessoas que passaram pela minha trajetória e contribuíram para me tornar quem eu sou hoje.

*“Só os loucos sabem.”*

(Charlie Brown)



# Resumo

A Teoria da Informação estuda o processamento, o armazenamento e a transmissão de informação através de sistemas de comunicação. O sistema nervoso é um exemplo de sistema de comunicação no qual neurônios respondem a estímulos emitindo uma sequência de impulsos elétricos, os quais provocam um fluxo da informação neuronal ao longo do sistema. Neste sentido, um dos propósitos dessa monografia é estudar algumas medidas da teoria da informação, com especial atenção para sua aplicação no estudo da conectividade neuronal. O segundo objetivo é compreender o processo de modelagem matemática das sequências de disparos neuronais via cadeias estocásticas de alcance variável e, então, propor estimadores para as medidas da informação obtidos a partir de amostras dessas cadeias. Por fim, verificaremos a performance de tais estimadores via simulação.

**Palavras-chave:** *Cadeias estocásticas de memória com alcance variável, Teoria da informação, Inferência em processos estocásticos, Neurociência.*



# Abstract

Information theory studies how communication systems integrate, encode, compute, and transmit information. The neural system is an example where communication takes place with neurons responding to stimuli and conducting electrical impulses by means of action potentials thereby causing a flow of neuronal information throughout the system. In this sense, one of the goals of this work is to study some information theory measures, with special attention to their application in the study of neural connectivity. The second goal is to understand the process of mathematical modeling of neuronal spike-trains prescribed from stochastic chains with memory of variable length as well as proposing estimators for the information metrics based on samples of these chains. Lastly, we study the performance of the estimators via computational simulation.

**Keywords:** *Chains with memory of variable length, Information theory, Statistical inference in stochastic processes, Neuroscience.*



# Lista de Figuras

2.1	<b>Divisões morfológicas do neurônio.</b> Da esquerda para direita: Dendritos, corpo celular, axônio e terminações do axônio. Adaptado de <a href="https://www.sobiologia.com.br/conteudos/FisiologiaAnimal/nervoso2.php">https://www.sobiologia.com.br/conteudos/FisiologiaAnimal/nervoso2.php</a> . . . . .	8
2.2	<b>Divisões morfológicas da membrana plasmática de um neurônio.</b> Da esquerda para direita: Canais iônicos, bomba de íons, esferas azuis representam íons de potássio e as esferas vermelhas representam os íons de sódio. Adaptado de Gerstner et al. (2014). . . . .	11
2.3	<b>Adaptação da membrana para um sistema elétrico.</b> Da esquerda para direita: A membrana plasmática de forma mais sucinta com um canal fechado e um aberto, ao lado, tem a adaptação em forma de sistema separado pelos diferentes canais. Adaptado de Gerstner et al. (2014) . . .	12
2.4	<b>Comportamento dos canais iônicos.</b> Da esquerda para direita: Velocidades de abertura e fechamento dos canais, ao lado, variação do valor de cada variável dos portões. Adaptado de notas de aula de Roque (n.d.). . .	13
2.5	<b>Sinapse química.</b> Da esquerda para direita: Fusão dos vesículos de neurotransmissores com a membrana do neurônio. Após atravessar a fenda sináptica os neurotransmissores (representado pelo Glutamato na imagem) são capturados pelos canais iônicos permitindo a transição de íons com o meio externo. Adaptado de Gerstner et al. (2014) . . . . .	15
3.1	Representação pictorial das árvores definidas no Exemplo 3.23. . . . .	35
3.2	Representação pictorial da árvore de contexto probabilística $\tau$ . . . . .	36
3.3	Representação pictorial da árvore $\tau = \{1,10,100,1000,\dots\}$ . . . . .	39
4.1	Árvore de contextos obtidas após uma interação do algoritmo. . . . .	58

5.1	$H(X)$ em função de $p$ quando $X$ é uma variável aleatória binária. Adaptado de Cover & Thomas (1991). . . . .	63
5.2	Relação entre entropia e informação mútua. . . . .	68
6.1	Representação pictorial da árvore $\tau_1 = \{1,10,100,1000,0000\}$ . . . . .	80
6.2	Representação no plano cartesiano do gráfico da função logística padrão. . . . .	82
6.3	Representação pictorial da árvore $\tau_2 = \{1,10,100,1000,0000\}$ . . . . .	84
6.4	Representação no plano cartesiano do gráfico da função inversa da função de ligação complemento log-log. . . . .	85
6.5	Representação pictorial da árvore $\tau_3$ . . . . .	86
6.6	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	94
6.7	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	95
6.8	Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	96
6.9	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	98
6.10	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	99
6.11	Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	100
6.12	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(0   \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	108

6.13	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(1 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	109
6.14	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(2 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	110
6.15	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(3 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	111
6.16	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(0 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	118
6.17	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(1 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	119
6.18	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(2 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	120
6.19	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição $\hat{Q}_3(2 \mid \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta. . . . .	121
6.20	Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia de transferência, considerando quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	123
7.1	Registro eletrofisiológico de neurônio piramidal no hipocampo do rato. Traçamos o registro da membrana de um roedor por aproximadamente 1s. Discretizamos o tempo em janelas temporais de $\Delta t = 0.03s$ (linhas tracejadas verticais). A cada janela temporal será atribuído o valor de 1 caso menos um disparo seja observado e 0 caso contrário. . . . .	128
7.2	Representação pictorial da árvore $\tau_{pre} = \{10,01,11,100,000\}$ . . . . .	129
7.3	Representação pictorial da árvore $\tau_{pos} = \{0,1\}$ . . . . .	130
7.4	Representação pictorial da árvore $\tau = \{2,3,00,10,20,30,01,11,21,31\}$ . . . . .	131



# Lista de Tabelas

2.1	Tabela de relação da condutância com potencial de reversão. . . . .	13
4.1	Número de aparições das sequências de tamanhos 1, 2 e 3 na amostra. . . . .	45
4.2	Número de aparições das sequências de tamanhos 1, 2 e 3 na amostra. . . . .	49
4.3	Número de aparições das sequências de tamanhos 1, 2 e 3 na amostra. . . . .	55
4.4	Número de aparições das sequências de tamanho 4 na amostra. . . . .	56
4.5	Número de aparições das sequências de tamanhos 5 na amostra. . . . .	57
6.1	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_1$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	93
6.2	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_1$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	95
6.3	Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando $\hat{Q}_1$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	96
6.4	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_2$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	98
6.5	Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_2$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	99

6.6	Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando $\hat{Q}_2$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	100
6.7	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(0   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	102
6.8	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(0   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	103
6.9	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(1   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	103
6.10	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(1   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	104
6.11	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(2   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	105
6.12	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(2   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	106
6.13	Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando $\hat{Q}_3(3   \text{Estado})$ e quatro diferentes tamanhos de amostra: $n = 13.500$ , $n = 27.000$ , $n = 54.000$ e $n = 108.000$ . . . . .	106

- 6.14 Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 107
- 6.15 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(0 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 112
- 6.16 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(0 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 113
- 6.17 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(1 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 114
- 6.18 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(1 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 115
- 6.19 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(2 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 115
- 6.20 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(2 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 116
- 6.21 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 117

- 6.22 Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 118
- 6.23 Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando  $\hat{Q}_3$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . . . . . 122

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Neurobiologia</b>	<b>7</b>
2.1	O sistema nervoso . . . . .	7
2.2	Neurônios . . . . .	8
2.2.1	Comportamento de uma célula passiva e o modelo integra-dispara . . . . .	9
2.3	Potencial da Membrana . . . . .	10
2.4	Modelo Hodgkin-Huxley . . . . .	12
2.5	Sinapses . . . . .	13
2.5.1	Sinapses Elétricas . . . . .	14
2.5.2	Sinapses Químicas . . . . .	14
2.6	Flutuações e ruídos . . . . .	15
2.7	Estocasticidade . . . . .	16
<b>3</b>	<b>Cadeias estocásticas</b>	<b>19</b>
3.1	Homogeneidade e estacionariedade . . . . .	19
3.2	Cadeias de Markov de ordem 1 . . . . .	22
3.3	Cadeias de Markov de ordem $k$ . . . . .	27
3.4	Cadeias de memória com alcance variável . . . . .	32
<b>4</b>	<b>Inferência em processos estocásticos</b>	<b>41</b>
4.1	Inferência em cadeias de Markov de ordem $k$ . . . . .	41
4.1.1	Estimação pelo método da máxima verossimilhança . . . . .	41
4.1.2	Seleção da ordem da cadeia . . . . .	47
4.2	Inferência em cadeias de memória com alcance variável . . . . .	51
4.2.1	Algoritmo contexto . . . . .	51

<b>5</b>	<b>Medidas da Teoria da Informação</b>	<b>59</b>
5.1	Informação . . . . .	60
5.2	Entropia, entropia conjunta e entropia condicional . . . . .	61
5.3	Entropia de uma cadeia estacionária . . . . .	65
5.4	Medidas da informação . . . . .	67
5.5	Estimação das medidas da informação . . . . .	70
<b>6</b>	<b>Aplicação computacional</b>	<b>75</b>
6.1	Hipóteses e Cenários . . . . .	76
6.2	Simulação . . . . .	78
6.2.1	Amostras dos disparos neuronais . . . . .	78
6.2.2	Estimação da conectividade neuronal . . . . .	89
6.3	Resultados . . . . .	92
6.3.1	Cenário 1 . . . . .	93
6.3.2	Cenário 2 . . . . .	97
6.3.3	Cenário 3 . . . . .	101
<b>7</b>	<b>Aplicação em dados reais</b>	<b>125</b>
7.1	Conjunto de dados . . . . .	125
7.2	Discretização . . . . .	127
7.3	Resultados . . . . .	128
7.3.1	Entropia condicional . . . . .	129
7.3.2	Entropia de transferência . . . . .	131
<b>8</b>	<b>Considerações Finais</b>	<b>133</b>
	<b>Glossário</b>	<b>135</b>
	<b>Referências Bibliográficas</b>	<b>137</b>

# Capítulo 1

## Introdução

Neurociência é o estudo científico do sistema nervoso (Anderson 1987). Os avanços tecnológicos em biologia molecular e neurociência computacional impulsionaram as pesquisas nesta área, uma vez que uma massiva quantidade de dados foi produzida e, conseqüentemente, novos problemas foram revelados. Como exemplo, temos o estudo da capacidade do cérebro transmitir e processar informação. Neste contexto, diferentes abordagens matemáticas para inferir a conectividade neuronal tem sido propostas. Neste trabalho, estamos interessados, em particular, em realizar essa inferência de um ponto de vista livre de um modelo probabilístico, via medidas da Teoria da Informação.

O cérebro é o principal e mais complexo órgão desse sistema e se subdivide entre muitas outras subregiões, cada uma com uma função diferente. Dentre essas funções, se destacam as chamadas atribuições superiores no cérebro humano onde são processados pensamentos, memórias e demais tarefas ligadas aos sentidos e à cognição humana (Kandel et al. 2000). O cérebro é composto por bilhões de neurônios (Herculano-Houzel 2009), que são células eletricamente excitáveis especializadas na troca de informação. A comunicação entre os neurônios ocorre por meio das sinapses, as quais podem ser elétricas ou químicas. Nas sinapses elétricas ocorre um fluxo direto de íons entre as células, enquanto que nas sinapses químicas, após o disparar de um neurônio (pré-sináptico) um neurotransmissor é liberado e se liga a uma proteína em uma célula receptora (pós-sináptica) fazendo com que uma despolarização localizada no seu potencial de membrana ocorra.

O potencial de membrana é a diferença de potencial elétrico entre o meio intracelular e o meio extracelular onde se destacam dois estados: o potencial de repouso no qual essa diferença se mantém aproximadamente constante devido a um estabilização do fluxo iônico e o potencial de ação no qual um estímulo (externo ou sináptico) provoca uma

variação brusca nessa diferença de potencial elétrico levando a uma descarga elétrica. Em neurociência, os potenciais de ação são conhecidos como impulsos nervosos ou disparos neuronais e a sequência de potenciais de ação geradas por um neurônio ao longo do tempo é chamada de trem de disparos neuronais. A atividade neuronal é, então, manifestada pela emissão de sequências de disparos neuronais de milhões de neurônios e o processamento cerebral é decorrente da interação dessas sequências (Gerstner et al. 2014). Conectar a atividade neural e funções desempenhadas pelo cérebro é um dos principais problemas da neurociência moderna.

Em neurobiologia, estudos das diferentes funções e como cada região desempenha suas atribuição no sistema nervoso são de suma importância. Entender, a forma como tal sistema integra, codifica e interpreta informações ao longo de um intervalo de tempo finito é fundamental para compreender a função do cérebro. De acordo com Buesing et al. (2011) dados experimentais sugerem que neurônios, sinapses e sistemas neurais são inerentemente estocásticos. Além disso, durante um dado período de tempo diferentes sequências de disparos poderiam ter ocorrido e o fato de uma única sequência em particular ocorrer em meio a todas as sequências possíveis sugere uma estrutura probabilística para descrição matemática e o tratamento de fenômenos neuronais (London et al. 2002). Portanto, é natural o uso de metodologias relacionadas a processos estocásticos e à teoria da informação para, respectivamente, modelar os componentes neuronais que evoluem ao longo do tempo e analisar a interação entre eles.

Neste contexto, após uma propícia discretização de um intervalo de tempo finito, a atividade de um neurônio é modelada através de uma sequência de zeros e uns, em que 1 significa que o neurônio disparou e 0 significa que o neurônio está em repouso. Essas atividades neuronais são avaliadas nos instantes de tempo que compõem a discretização e as sequências de zeros e uns são os trens de disparos neuronais. Portanto, a atividade de cada neurônio pode ser modelada através de um processo estocástico assumindo valor em  $\{0,1\}$ . Na literatura, esse processo tem sido utilizado envolvendo uma descrição Markoviana, embora alguns trabalhos tem mostrado que tal descrição é inadequada (ver, por exemplo, Friston et al. 2010; Truccolo et al. 2010; Cessac 2011b). Cessac e seus co-autores tem desenvolvido um estudo sistemático sobre a modelagem de trens de disparos neuronais usando cadeias estocásticas de alcance variável (ver, por exemplo, Cessac 2008; Cessac et al. 2009; Cessac 2011a; Cessac 2011b; Cessac & Cofre 2013). Neste contexto, Galves & Löcherbach (2013) e Galves & Löcherbach (2015) consideram um modelo base-

ado em sistemas possivelmente infinitos associados a processos pontuais com memória de tamanho variável. Devido as capacidades matemáticas, práticas e computacionais desse modelo, atualmente, ele vem sendo amplamente utilizado na literatura (ver, por exemplo, Duarte & Ost 2014; Duarte et al. 2015).

A classe das cadeias estocásticas com memória de alcance variável foi introduzida por Rissanen (1983). Essa classe ficou popular na comunidade estatística e probabilística com o trabalho de Bühlmann et al. (1999) e o termo cadeias estocásticas com memória de alcance variável deve-se a Galves & Löcherbach (2008). As probabilidades de transição desse tipo de processo estocástico não dependem de uma quantidade fixa de estados passados, levando em conta a estrutura de dependência presente nos dados. Assim, podemos representar o conjunto das sequências de estados passados como uma árvore, denominada árvore de contexto. Considerando uma memória de tamanho variável temos modelos mais informativos, flexíveis e parcimoniosos do que as cadeias de Markov.

Dada uma trajetória da cadeia de memória com alcance variável podemos realizar a estimação das probabilidade de transição e selecionar o modelo mais simples através do Algoritmo Contexto, que foi proposto por Rissanen (1983). Um dos interesses da neurociência é inferir a informação que flui de uma sequência de disparo neuronal a outra. Para isso podemos utilizar medidas da teoria da informação, as quais são funções das probabilidades de transição. Assim, a estimação de tais medidas está, essencialmente, atrelada à estimação das probabilidades de transição. Neste trabalho, consideramos a modelagem da sequência de disparos neuronais via cadeias estocásticas de alcance variável e utilizamos medidas da informação para entender a transmissão de informação entre neurônios ao longo de um intervalo de tempo finito.

Nesta monografia, estudamos todos os aspectos metodológicos necessários para realizar um estudo de simulação e com dados reais a respeito da conectividade neuronal. Vamos considerar o problema de estimação das medidas da informação em três cenários distintos:

1. Um neurônio, de uma área específica do cérebro, terá sua resposta temporal, disparos e não disparos, observada ao longo de um período de tempo finito e, então, iremos estimar, para diferentes tamanhos de amostras, a incerteza média esperada com relação à atividade do neurônio quando condicionada ao conhecimento do histórico passado de suas respostas temporais. Neste caso, vamos gerar as amostras de disparos neuronais a partir de uma família de probabilidades de transição, em que a probabilidade do neurônio disparar em um dado instante de tempo condicionada

ao seu passado é modelada a partir de uma função de ativação logística  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  dada por

$$\phi(v) = \frac{1}{1 + \exp(-v)}.$$

2. Novamente, vamos considerar o problema de estimar a incerteza média esperada com relação à atividade de um neurônio quando condicionada ao conhecimento do histórico passado de suas respostas. Entretanto, neste caso, as amostras são geradas a partir de uma família de probabilidades de transição, em que a probabilidade do neurônio disparar em um dado instante de tempo condicionada ao seu passado é modelada a partir de uma função de ativação  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  tal que

$$\phi(v) = 1 - \exp(-\exp(v)),$$

cuja inversa  $\phi^{-1}$  é popularmente conhecida na literatura estatística como a função de ligação complemento log-log de uma regressão logística.

3. Dois neurônios  $N_1$  e  $N_2$ , de uma área específica do cérebro, tem sua resposta temporal, disparos e não disparos, observadas ao longo de um período de tempo finito e, então, iremos estimar, para diferentes tamanhos de amostra, a informação que flui de um neurônio ao outro. Em outras palavras, vamos verificar o quanto diminui a incerteza média esperada com relação à atividade do neurônio  $N_i$  quando além do conhecimento do histórico passado das suas respostas temporais, sabemos também o histórico passado da atividade do neurônio  $N_j$ ,  $i, j = 1, 2$  com  $i \neq j$ . Neste caso, vamos gerar as amostras de disparos neuronais a partir de uma cadeia estocástica de alcance variável bivariada, cujas probabilidades de transição são fixadas levando em consideração a Teoria Hebbiana (Hebb, et al. 1949). Esta teoria parte do pressuposto que “células que disparam juntas, permanecem conectadas”. Em outras palavras, quando o neurônio  $N_i$  disparar, o neurônio  $N_j$ , que está conectado a  $N_i$ , possuirá uma probabilidade maior de disparar em sequência,  $i, j = 1, 2$  com  $i \neq j$ . Além disso, quando ambos ficam um tempo suficientemente longo sem disparar, é mais provável que venham a disparar juntos.

O objetivo de cada um desses cenários dentro do escopo deste estudo de simulação é diferente. Nos cenários 1 e 2, estamos interessados em estimar a entropia condicional associada a blocos com um tamanho finito fixado e utilizar réplicas de simulação Monte

Carlo para avaliar o desempenho dos métodos de estimação propostos nesta monografia. Em ambos os cenários, as probabilidades de transição que geram as amostras são modeladas a partir de funções de ativação distintas. Assim, pretendemos também verificar se a escolha dessas funções impactam na performance das metodologias de inferência que estão sendo propostas. Por outro lado, no cenário 3, estamos interessados em estimar a entropia de transferência associada a blocos com um tamanho finito fixado a fim de estudar a existência de conectividade entre dois neurônios. Neste último caso, usamos novamente réplicas de simulação Monte Carlo para avaliar os métodos de estimação propostos. Até onde vai nosso conhecimento, não há na literatura trabalho algum que trate da estimação da entropia de transferência utilizando o Algoritmo Contexto neste cenário mais geral de cadeias de alcance variável. Desta forma, este trabalho complementa, em um certo sentido, os estudos sobre estimação da entropia de transferência.

Este trabalho está organizado da seguinte maneira. No próximo capítulo, apresentamos definições e noções básicas sobre neurociência. No capítulo 3, revisamos alguns conceitos básicos da teoria de processos estocásticos, em particular, das cadeias de Markov e introduzimos a classe das cadeias com memória de alcance variável. No capítulo 4, tratamos da estimação das probabilidades de transição da cadeia de alcance variável através do método da máxima verossimilhança juntamente com o algoritmo contexto. No capítulo 5, apresentamos a definição e algumas propriedades das medidas da informação que são utilizadas neste trabalho. No capítulo 6, apresentamos os resultados via simulação nos cenários descritos. Por fim, no capítulo 7, apresentamos a aplicação em dados reais. O capítulo 8 encerra essa monografia com algumas considerações finais.



# Capítulo 2

## Neurobiologia

Sequências de disparos neuronais são o objeto principal de estudo desta monografia. Uma sequência de disparos somente ocorre por intermédio de mecanismos neurobiológicos que determinam temporalmente onde cada disparo ocorre. Assim, neste capítulo, será feita uma revisão bibliográfica sobre os mecanismos básicos da neurobiologia que determinam o potencial de ação, bem como os mecanismos que influenciam uma sequência de disparos.

### 2.1 O sistema nervoso

O sistema nervoso representa uma rede de comunicações do organismo. O cérebro é o principal e mais complexo órgão desse sistema e se subdivide entre muitas outras subregiões, cada uma com uma função diferente. Dentre essas funções, se destacam as chamadas atribuições superiores do cérebro humano, onde são processados pensamentos, memórias e demais tarefas ligadas aos sentidos e à cognição humana (Kandel et al. 2000). O cérebro é composto por bilhões de neurônios (Herculano-Houzel 2009), que são células eletricamente excitáveis especializados na troca de informação. O termo neurobiologia é utilizado para se referir ao estudo científico das células e da organização do sistema nervoso. A comunicação entre os neurônios ocorre por meio das sinapses, as quais podem ser elétricas ou químicas. Ao conectar nosso estudo com a neurobiologia, é importante a definição de nomenclaturas e funcionalidades do sistema nervoso. Salientamos que embora muitos termos são explicados ao longo do texto, alguns são utilizados sem a devida definição. Neste sentido, sugerimos ao leitor ou leitora a leitura do glossário no final desta monografia (ver, página 77).

## 2.2 Neurônios

Os neurônios são as células do sistema nervoso estudadas neste trabalho. Estes são divididos em três regiões: corpo celular, dendritos e axônio (ver Figura 2.1). O corpo celular é responsável por armazenar o núcleo da célula com suas proteínas características. Os dendritos, em conjunto com o corpo celular, recebem os estímulos elétricos de outros neurônios, em contraste com o axônio que é o responsável por passar um sinal elétrico, chamado potencial de ação (ver, Kandel et al. 2000). Em suma, um potencial de ação é uma alteração rápida na polaridade da tensão elétrica (com amplitude de aproximadamente 100mV). Ao chegar na última região, terminações do axônio, na qual a informação criada pelo potencial de ação deste neurônio, que será denominado pré-sináptico, estimula o neurônio (pós-sináptico) que está em contato com as terminações do neurônio por meio de uma sinapse.

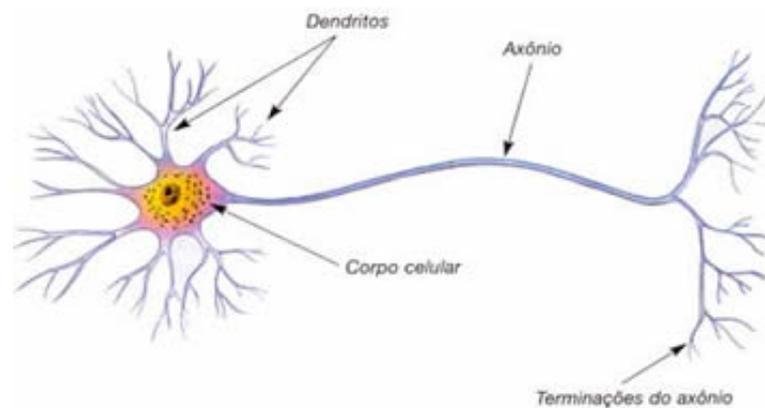


Figura 2.1: **Divisões morfológicas do neurônio.** Da esquerda para direita: Dendritos, corpo celular, axônio e terminações do axônio. Adaptado de <https://www.sobiologia.com.br/conteudos/FisiologiaAnimal/nervoso2.php>.

Contudo, é importante destacar que a ocorrência de um disparo não é comum, tendo em vista que um neurônio recebe diversos estímulos a todos os momentos e somente após integrar uma quantidade necessária em um curto período de tempo que a membrana atinge um potencial elétrico suficiente para disparar. Apresentamos abaixo como os diferentes componentes e mecanismos que compõem um neurônio podem ser explicados matematicamente por meio de modelos usualmente utilizados em neurociência. O objetivo desta apresentação é discutir as funções subjacentes ao potencial de ação.

## 2.2.1 Comportamento de uma célula passiva e o modelo integrador-dispara

A princípio, pode-se considerar o comportamento somente da membrana de uma célula sem propriedades mais complexas como canais iônicos. Assim, assumindo que uma célula neural com membrana passiva, que possui ausência de canais iônicos, receba um estímulo sináptico de curta duração, seu potencial da membrana varia e posteriormente retorna ao estado de repouso, caso o potencial não atinja o limite para ocorrer um potencial de ação. Neste contexto, a interação entre estímulos sinápticos e o potencial de repouso da membrana pode ser descrito como,

$$u_i(t) = u_{rest} + \epsilon_{ij}(t),$$

em que  $u_i(t)$  é o potencial elétrico pós-sináptico do neurônio  $i$  no instante de tempo  $t$ ,  $u_{rest}$  é o potencial de repouso da membrana e  $\epsilon_{ij}(t)$  é o estímulo sináptico recebido em  $i$ , vindo do neurônio  $j$ , no instante de tempo  $t$ .

Portanto, nesta simples configuração de membrana passiva, estímulos recebidos se acumulam de forma linear, ou seja, os mesmos somam suas influências no neurônio pós-sináptico e em conjunto são capazes de modificar o potencial deste, até que um determinado potencial limiar seja superado e assim gerar um novo disparo (Gerstner et al. 2014). Dessa forma, ao considerar que mais de um neurônio pré-sináptico estimule o neurônio em estudo, têm-se:

$$u_i(t) = \sum_j \sum_f \epsilon_{ij}(t - t_j^{(f)}) + u_{rest}.$$

em que  $(t_j^{(f)})$  indica o tempo em que o neurônio  $j$  dispara no instante  $f$ .

Entretanto, a geração do disparo não ocorre de forma espontânea pois há ausência de canais iônicos. Desta forma, é dito que o disparo, neste modelo, ocorre quando o potencial de membrana atinge um determinado limiar, ou seja, não levando em conta todos os fatores biológicos. Sendo assim, é possível tratar os disparos como eventos, considerar um limite constante e o potencial é reiniciado após os disparos. Em resumo, este modelo relata superficialmente o que é ocorrido com a membrana de um neurônio em estado de repouso.

Devido ao fato da membrana celular que cobre o neurônio ser uma boa condutora de eletricidade e as sinapses que chegam ao neurônio, geralmente, entregam carga em

um curto espaço de tempo, podemos descrever este comportamento como um modelo um circuito elétrico RC (resistor e capacitor). Podendo ser escrito como:

$$\begin{aligned} I(t) &= I_R(t) + I_C(t), \\ &= \frac{u - u_{rest}}{R} + C \frac{du}{dt}, \end{aligned} \tag{2.1}$$

em que,  $I_R$  representa a intensidade da corrente que passa pelo resistor  $R$ ,  $I_C$  representa a intensidade da corrente que passa pelo capacitor  $C$  e  $I(t)$  é a intensidade da corrente total. A equação 2.1 implica em

$$\tau \frac{du}{dt} = RI(t) - (u - u_{rest}), \tag{2.2}$$

sendo que  $\tau$  é a constante de tempo da membrana. A solução da equação (2.2) segue um comportamento exponencial, o que faz sentido, já que é utilizado um circuito RC e a fonte perde energia no decorrer do tempo, somado a outros fatores (ver Gerstner et al. 2014). Assim, o neurônio sempre tende a retornar ao seu estado inicial com velocidade exponencial exceto quando é excitado. Portanto, concluímos que não é somente a intensidade da corrente elétrica que importa para criar um novo disparo, mas também, a quantidade de carga que é aplicada que pode ser maior se a corrente for aplicada. Ou seja, correntes podem causar um disparo se aplicadas por um intervalo de tempo maior ou mesmo se diversos disparos com tempo curto são aplicados de forma sucessiva.

## 2.3 Potencial da Membrana

As discussões anteriores se tratavam de modelos mais simplificados. Entretanto, note que a geração de um potencial de ação em um modelo de membrana passiva ocorre de forma artificial e não seguindo uma regra biofísica. Com o intuito de apresentar modelos com maior precisão, nesta seção, é mostrado como é gerado um potencial de ação e os fatores biofísicos necessários para isso. O potencial de membrana é a diferença de potencial elétrico entre o meio intracelular e o meio extracelular. No potencial de membrana será destacado dois estados: o potencial de repouso e o potencial de ação. Para entender ambos temos que entender um pouco mais sobre o funcionamento de um neurônio.

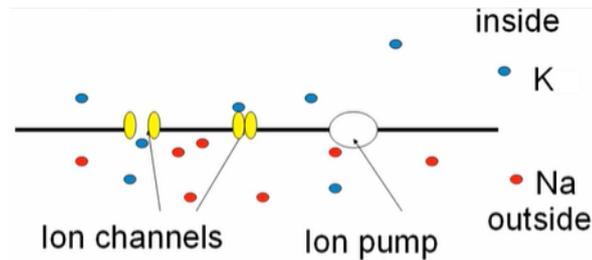


Figura 2.2: **Divisões morfológicas da membrana plasmática de um neurônio.** Da esquerda para direita: Canais iônicos, bomba de íons, esferas azuis representam íons de potássio e as esferas vermelhas representam os íons de sódio. Adaptado de Gerstner et al. (2014).

Como é ilustrado na Figura 2.2, as membranas plasmáticas dos neurônios possuem canais que permitem a troca de íons com o meio extracelular, sendo que ao se abrirem permitem o fluxo de íons. A bomba de íons é responsável por manter um gradiente de concentração, ao atrair potássio para o interior da célula e retirar sódio, há maior concentração de esferas azuis no interior da célula. Como consequência, temos concentrações diferentes dos íons (sódio e potássio) entre o interior e o exterior, esta diferença de concentração, faz com que a membrana assuma diferentes voltagens. Para cada íon, pode-se obter o chamado potencial de reversão (ver, equação (2.3)), que determina o potencial gerado quando o íon possui diferentes concentrações entre o meio intra- e extracelular,

$$(u_1 - u_2) = -\frac{kT}{q} \ln\left(\frac{n_1}{n_2}\right), \quad (2.3)$$

em que  $u_1$  e  $u_2$  são os potenciais elétricos dos neurônios 1 e 2,  $n_1$  e  $n_2$  a concentração de íons dos neurônios 1 e 2,  $T$  é a temperatura,  $k$  é a constante de Boltzmann e  $q$  é a carga dos íons. A combinação dos diversos íons gera um potencial de repouso negativo de aproximadamente -65 mV.

Ao receber uma sinapse excitatória o neurônio pós-sináptico têm uma mudança no fluxo dos seus canais iônicos, o que despolariza (ato de aumentar o potencial em direção a valores positivos) o potencial da membrana, ao liberar um fluxo de íons de sódio para o interior da célula. Dessa forma, o fluxo de sódio muda o potencial do neurônio o que pode vir a gerar um disparo; posteriormente o canal de potássio é acionado, hiperpolarizando a membrana em um estado de potencial inferior ao de repouso, até que, ao decorrer do tempo, a membrana retorne ao potencial de repouso. Portanto, cada potencial de ação, possui uma fase ascendente, uma fase descendente e, ainda, uma curva de tensão elétrica inferior a do potencial de repouso de membrana (importante notar que o ciclo completo

dura em geral 5ms).

## 2.4 Modelo Hodgkin-Huxley

O descrito anteriormente são os fatores para a ocorrência de um potencial de ação, os potenciais de ação são conhecidos como impulsos nervosos ou disparos neuronais e a sequência de potenciais de ação geradas por um neurônio ao longo do tempo é chamada de trem de disparos neuronais. De forma a explicar o funcionamento de um potencial de ação, nesta seção, focaremos no modelo Hodgkin-Huxley. No século XX, os pesquisadores Hodgkin e Huxley desenvolveram um modelo a partir dos axônios de lulas que pode ser descrito com sistemas elétricos do tipo RC (resistor e capacitor), veja Figura 2.3.

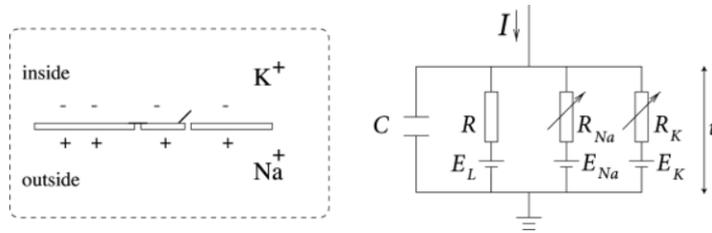


Figura 2.3: **Adaptação da membrana para um sistema elétrico.** Da esquerda para direita: A membrana plasmática de forma mais sucinta com um canal fechado e um aberto, ao lado, tem a adaptação em forma de sistema separado pelos diferentes canais. Adaptado de Gerstner et al. (2014) .

Neste modelo, foram considerados três canais iônicos (sódio, potássio e vazamento; este último se refere aos demais íons) e uma corrente estimulante que inicia as diferentes modificações no sistema e começa o ciclo do potencial de ação. Algumas das possíveis correntes estimulantes são as sinapses, que podem ser químicas ou elétricas. Talvez o fato mais importante deste modelo tenha sido a criação das variáveis de portões Hodgkin-Huxley, que são essenciais para o entendimento de como ocorre o fluxo de íons em neurônios, e são modeladas da seguinte forma:

$$I(t) = I_c(t) + I_K + I_{Na} + I_L,$$

$$C \frac{du}{dt} = I(t) - g_K n^4 (u - E_K) - g_{Na} m^3 h (u - E_{Na}) - g_{R_L} (u - E_L), \quad (2.4)$$

em que  $n^4$  representa a variável de abertura do canal de potássio,  $m^3$  a variável de abertura do canal de sódio e  $h$  a variável de fechamento do canal de sódio;  $g_K, g_{Na}, g_L$  são as condutâncias de cada um dos respectivos íons, tratadas como constantes no estudo; e

$E_K, E_{Na}, E_L$  são os potenciais de reversão de cada um dos respectivos íons, apresentados na Tabela 2.1.

Tabela 2.1: Tabela de relação da condutância com potencial de reversão.

x	$E_x [mV]$	$g_x [mS/cm^2]$
Na	55	40
K	-77	35
L	-65	0,3

O comportamento dos canais iônicos, descrito pelas variáveis de portões, pode ser demonstrado por meio de gráficos, veja Figura 2.4.

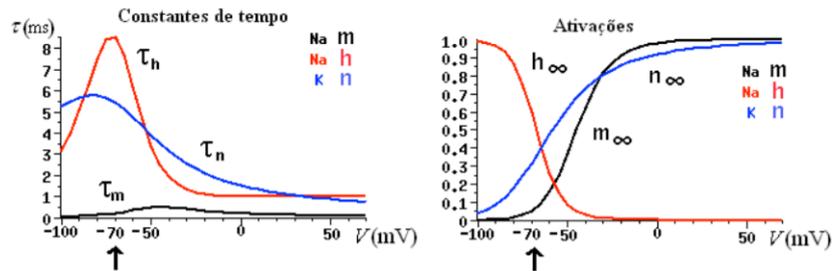


Figura 2.4: **Comportamento dos canais iônicos.** Da esquerda para direita: Velocidades de abertura e fechamento dos canais, ao lado, variação do valor de cada variável dos portões. Adaptado de notas de aula de Roque (n.d.).

Diante da Figura 2.4, observa-se que a velocidade de abertura do canal de sódio é evidentemente maior que as velocidades de abertura do potássio e da inativação do sódio. Portanto, ao receber uma sinapse excitatória,  $m$  alcança valores maiores do que  $n$  e  $h$ , em um menor período de tempo. Como resultado, por um breve período de tempo, o canal iônico de sódio se torna soberano, forçando assim o potencial da membrana para o potencial de Nernst do sódio, o que pode vir a resultar em um disparo. Em seguida, o mesmo canal de sódio inativa e a entrada de íons de potássio se torna massiva dentro da célula, hiperpolarizando-a. Dessa forma, a potencial intra-celular tende para um valor ainda menor que o potencial de repouso, até que, aos poucos, retorne ao seu estado natural.

## 2.5 Sinapses

As conexões entre neurônios são denominadas sinapses, e podem ser divididas em elétricas e químicas, sendo que as sinapses químicas são mais comuns de serem observadas.

### 2.5.1 Sinapses Elétricas

As sinapses elétricas são caracterizadas pela conexão direta entre dois neurônios vizinhos, efetuada por meio de poros que atravessam as membranas de ambos, permitindo a troca de íons entre eles. A região que entra em contato é denominada junção comunicante, nas quais, canais constituídos por conexinas criam a ponte entre os neurônios pré e pós-sinápticos. Dessa forma, temos uma corrente sináptica bidirecional, gerada pela diferença de potencial entre os neurônios. Esta sinapse é extremamente rápida e geralmente é efetuada para reflexos humanos (veja Viana (n.d.)). Para modelar uma sinapse elétrica entre dois neurônios, introduzimos uma corrente sináptica  $I_{syn}$ :

$$I_{syn} = g_c(V_{pre} - V_{pos}),$$

em que  $g_c$  é a condutância da sinapse elétrica e  $V_{pre}$  e  $V_{pos}$  são os potenciais das membranas dos neurônios pré e pós-sinápticos. Entretanto, novamente, sabemos que a rede neural não ocorre somente com os fenômenos de um neurônio, logo, dado um neurônio pós-sináptico  $i$ , ele pode receber mais de uma conexão elétrica, vindo de  $N$  neurônios pré-sinápticos com potenciais de membrana  $V_j (j = 1, 2, \dots, N)$ . Neste caso, a corrente sináptica total sobre o  $i$ -ésimo neurônio será a soma de todas as contribuições:

$$I_{syn,i} = g_c \sum_{j=1}^N (V_j - V_i).$$

### 2.5.2 Sinapses Químicas

Nas sinapses químicas o neurônio pré-sináptico, instigado por um potencial de ação, libera moléculas de um neurotransmissor que difunde-se através da fenda sináptica e liga-se a receptores no neurônio pós-sinápticos. As sinapses químicas podem ser de dois tipos: excitatórias (neurotransmissor Glutamato é o mais comum) ou inibitórias (neurotransmissor GABA é o mais comum). A Figura 2.5 representa melhor a chegada dos neurotransmissores no neurônio pós-sináptico.

Ao analisar as sinapses, podemos verificar que a mesma pode vir a ser tratada como um canal, já que a corrente que esta gera, resulta na abertura de um canal e os íons envolvidos modificam o potencial da membrana pós-sináptica. Portanto, as sinapses químicas podem

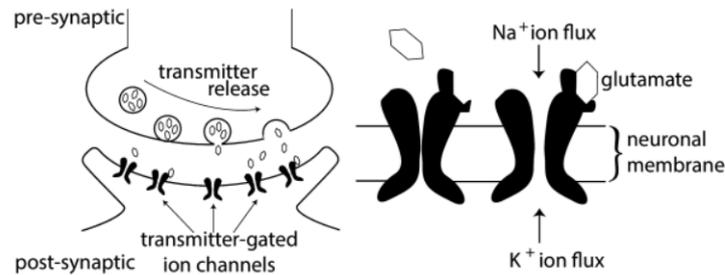


Figura 2.5: **Sinapse química.** Da esquerda para direita: Fusão dos vesículos de neurotransmissores com a membrana do neurônio. Após atravessar a fenda sináptica os neurotransmissores (representado pelo Glutamato na imagem) são capturados pelos canais iônicos permitindo a transição de íons com o meio externo. Adaptado de Gerstner et al. (2014) .

ser escritas matematicamente como:

$$I_{syn}(t) = g_{syn}(t)(u - E_{syn}).$$

Assim, o canal sináptico depende de uma condutância sináptica  $g_{syn}$  e, como nos demais canais (Potássio e Sódio), a corrente que passa pelo canal depende da diferença do potencial sináptico ( $E_{syn}$ ) e do potencial da membrana. Contudo, nesse caso, a condutância sináptica  $g_{syn}$  depende primeiramente do tipo de sinapse química, excitatória ou inibitória. Caso seja a primeira, os canais podem ser AMPA ou NMDA, cada um com uma forma diferente de se calcular a condutância; caso contrário, podem ser os canais  $GABA_A$  ou  $GABA_B$ .

## 2.6 Flutuações e ruídos

Variabilidade são as discrepâncias que ocorrem ao observar um mesmo experimento realizado em regiões neurais semelhantes, neurônios similares, ou até no mesmo neurônio. Sendo que, as flutuações são variações que podem ou não ter um padrão. Dessa forma, denominaremos ruídos como sendo flutuações aleatórias que não fazem parte do sinal de um disparo. Neste contexto, a rede cerebral é influenciada por diversas formas de ruídos no sistema, além disso, estes mesmos são de suma importância para o desempenho de funções da rede neural (ver, por exemplo, Faisal et al. 2008).

Dentre todas as diferentes fontes de ruídos existentes, as apresentadas a seguir são aquelas que possuímos um maior interesse, já que, estas possuem influências estocásticas no ato de disparo do neurônio.

- **Ruído celular:** Este é representado pela variabilidade entre dois neurônios, de uma região semelhante, que são estimulados pela mesma corrente e mesmo assim possuem um tempo de resposta diferente. Logo, entendemos que cada neurônio possui aspectos únicos;
- **Ruído de canal:** Ao se trabalhar com canais iônicos realizando observações diversas ao longo do tempo, é possível notar que a ocorrência de aberturas ou fechamentos dos canais possui um certo grau de estocasticidade o que pode resultar na geração de ruídos;
- **Ruído sináptico:** Este ruído pode vir a ser gerado por meio de sinais sinápticos que falharam ou mesmo pela liberação espontânea de neurotransmissores que são dispersos eventualmente;
- **Ruído da rede:** A rede neural é composta por bilhões de neurônios que geram aproximadamente trilhões de sinapses. O enorme número de sinapses que um neurônio recebe faz com que o sinal do neurônio possua uma característica estocástica que pode ser descrita como um ruído.

## 2.7 Estocasticidade

Levando em conta as flutuações e ruídos mencionados na seção anterior podemos citar o trabalho de Buesing et al. (2011), que trabalha com dados experimentais, sugerindo que neurônios, sinapses e sistemas neurais são inerentemente estocásticos, haja vista que flutuações e ruídos interferem na atividade altamente estocástica das sinapses de neurônios. Ademais, no decorrer de um período específico de tempo distintas sequências de disparos poderiam ocorrer e o fato de uma única sequência em particular ocorrer sugere uma estrutura probabilística para a descrição matemática e o tratamento de fenômenos neuronais.

Dessa forma, após verificar as complexidades do comportamento cerebral, somado as intensas e diversas fontes de ruídos é possível concluir que a inferência e a previsão de um disparo não é algo trivial. Ou seja, os modelos supracitados possuem forte embasamento biofísico, mas enfrentam dificuldades para inferir conexões neurais. Todavia, uma forma de superar essa dificuldade é modelar somente os trens de disparos de forma fidedigna. Logo, nos capítulos seguintes, propomos modelar os trens de disparos neuronais via ca-

deias estocásticas de memória com alcance variável e apresentamos uma metodologia para inferência da conectividade neuronal baseada em medidas da teoria da informação. Assim, vamos trabalhar em um ambiente probabilístico que leva em conta as considerações neurobiológicas discutidas neste capítulo.



# Capítulo 3

## Cadeias estocásticas

A atividade neural é manifestada pela emissão ao longo do tempo de sequências de disparos neuronais de milhões de neurônios e o processamento cerebral é recorrente da interação dessas sequências (Gerstner et al. 2014). De acordo com Buesing et al. (2011), dados experimentais sugerem que neurônios, sinapses e sistemas neurais são inerentemente estocásticos. Portanto, é natural considerar uma estrutura probabilística para descrição matemática e o tratamento de fenômenos neuronais. Neste sentido, utilizamos processos estocásticos para modelar as sequências de disparos neuronais.

### 3.1 Homogeneidade e estacionariedade

Suponha que monitoramos a atividade de um neurônio ao longo de um intervalo de tempo finito e seja  $\mathbb{T}$  o conjunto de tempos sobre os quais observamos a atividade de disparo desse neurônio. Ao invés de considerarmos que dada uma condição inicial a evolução desse sistema é determinística; consideramos que existem diversas trajetórias possíveis pelas quais o sistema pode evoluir a partir de leis probabilísticas, ou seja, a evolução é aleatória. Neste contexto, consideramos uma variável aleatória  $X_t$  definida em um espaço de probabilidade adequado tal que

$$X_t = \begin{cases} 1, & \text{se o neurônio disparar no tempo } t \in \mathbb{T}, \\ 0, & \text{caso contrário.} \end{cases}$$

Logo, o estado do sistema no instante  $t \in \mathbb{T}$  é determinado pela variável aleatória  $X_t$  e, desta forma,  $\mathcal{S} = \{0,1\}$  é o conjunto dos possíveis estados desse sistema. Portanto,

a sequência aleatória  $X := \{X_t : t \in \mathbb{T}\}$  com valores em  $\mathcal{S}$  representa a sequência de disparos do neurônio ao longo do conjunto  $\mathbb{T}$ . Em teoria das probabilidades, a sequência aleatória  $X$  é denominada processo estocástico.

**Definição 3.1** *Um processo estocástico é uma sequência de variáveis aleatórias  $X := \{X_t : t \in \mathbb{T}\}$  definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados  $\mathcal{S}$ .*

Do ponto de vista da neurociência aplicada, é útil considerar que atividade neuronal ocorre ao longo de um conjunto de tempos  $\mathbb{T}$  discreto (MacKay & McCulloch 1952; Rieke 1997). Assim, levando em conta a natureza aplicada deste trabalho, vamos considerar  $\mathbb{T} = \mathbb{Z}$  ou  $\mathbb{T} = \mathbb{N} := \{0, 1, 2, \dots\}$  e o espaço de estados  $\mathcal{S} = \{0, 1\}$ . Deste modo, os processos estocásticos considerados nessa monografia são sempre a tempo discreto e com espaço de estados finito.

Como  $\mathbb{T}$  é discreto a lei ou distribuição de um processo estocástico  $X$  que evolui ao longo do conjunto de tempos  $\mathbb{T}$  é o conjunto de todas as distribuições conjuntas finitas-dimensionais  $\mu$  tais que

$$\mu \left( \bigcap_{t=m}^n \{a_t\} \right) := P \left( \bigcap_{t=m}^n \{X_t = a_t\} \right)$$

quaisquer que sejam  $a_t \in \mathcal{S}$  e  $t \in [m, n] \cap \mathbb{Z}$ . A distribuição do processo  $X$  pode ser também definida especificando-se uma condição inicial e uma dinâmica para a evolução do sistema. Como estamos interessados em sistemas cuja evolução é aleatória, vamos considerar como condição inicial uma distribuição de probabilidade, a qual denominamos distribuição inicial, e a dinâmica do sistema sendo regida pelo conjunto de probabilidades condicionais

$$\left\{ P \left( X_t = a_t \middle| \bigcap_{i=1}^{\infty} \{X_{t-i} = a_{t-i}\} \right) : a_t \in \mathcal{S}, \quad \forall t \in \mathbb{Z} \right\}, \quad (3.2)$$

quando  $\mathbb{T} = \mathbb{Z}$ , e

$$\left\{ P \left( X_t = a_t \middle| \bigcap_{i=1}^t \{X_{t-i} = a_{t-i}\} \right) : a_t \in \mathcal{S}, \quad \forall t \in \mathbb{N} \right\}, \quad (3.3)$$

quando  $\mathbb{T} = \mathbb{N}$ . De forma geral, estas probabilidades condicionais são regras que especificam a transição entre os estados do sistema.

No contexto de inferência estatística sobre modelos neurobiológicos, é interessante

considerar probabilidades de transição que são invariantes ao longo do tempo (homogeneidade) e analisar os resultados estocásticos de um experimento sem uma dependência explícita da escolha do tempo inicial do sistema (estacionariedade). Nesta perspectiva, consideramos, ao longo desse texto, processos estocásticos a tempo discreto com espaço de estados finitos que são homogêneos e estacionários.

**Definição 3.4** *Um processo estocástico a tempo discreto  $X := \{X_t : t \in \mathbb{T}\}$  definido em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados  $\mathcal{S}$  finito é dito ser homogêneo quando as probabilidades de transição são invariantes ao longo do tempo, ou seja, as probabilidades condicionais pertencentes aos conjuntos (3.2) e/ou (3.3) são independentes do valor de  $t \in \mathbb{T}$ .*

**Definição 3.5** *Um processo estocástico a tempo discreto  $X := \{X_t : t \in \mathbb{T}\}$  definido em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados  $\mathcal{S}$  finito é dito ser estacionário quando as distribuições conjuntas não dependem da escolha do tempo inicial, ou seja,*

$$P\left(\bigcap_{t=m}^n \{X_t = a_t\}\right) = P\left(\bigcap_{t=m}^n \{X_{t+1} = a_t\}\right),$$

quaisquer que sejam  $m, n \in \mathbb{T}$  e  $(a_m, \dots, a_n) \in \mathcal{S}^{\setminus \mathbb{T} + \infty}$ .

**Exemplo 3.6** *Processos estocásticos a tempo discreto com espaço de estados finito cujas variáveis aleatórias são independentes e identicamente distribuídas (i.i.d.) são processos homogêneos e estacionários. De fato, se  $X := \{X_t : t \in \mathbb{Z}\}$  é um processo i.i.d. definido sobre um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no espaço de estados  $\mathcal{S}$ , então, para todo  $a_t \in \mathcal{S}$ ,  $t \in \mathbb{Z}$  e  $s \in \mathbb{Z}$ ,*

$$P\left(X_t = a_t \left| \bigcap_{i=1}^{\infty} \{X_{t-i} = a_{t-i}\} \right.\right) \stackrel{(a)}{=} P(X_t = a_t) \stackrel{(b)}{=} P(X_{t+s} = a_t),$$

em que (a) segue da independência das variáveis aleatórias e (b) vem do fato de serem identicamente distribuídas. Concluimos, dessa forma, que  $X$  é homogênea. Além disso,

para todo  $m, n \in \mathbb{Z}$  com  $m \leq n$ ,

$$\begin{aligned} P\left(\bigcap_{t=m}^n \{X_t = a_t\}\right) &\stackrel{(c)}{=} \prod_{t=m}^n P(X_t = a_t) \\ &\stackrel{(d)}{=} \prod_{t=m}^n P(X_{t+1} = a_t) \\ &\stackrel{(e)}{=} P\left(\bigcap_{t=m}^n \{X_{t+1} = a_t\}\right), \end{aligned}$$

em que (c) e (e) seguem da independência e (d) da distribuição das variáveis serem idênticas. Portanto,  $X$  é estacionária.

A classe dos processos estocásticos i.i.d. é a mais simples possível. No entanto, essa nem sempre é uma classe interessante para modelar fenômenos aleatórios com dinâmica temporal, pois a hipótese de independência se mostra muito restritiva. Em sistemas neurobiológicos é observado empiricamente uma influência dos estados passados do sistema sobre a probabilidade do seu estado presente (Kandel et al. 2000). Neste sentido, introduzimos na próxima seção uma classe de processos estocásticos que permite certa dependência probabilística com o passado.

## 3.2 Cadeias de Markov de ordem 1

As cadeias de Markov são modelos matemáticos para sistemas estocásticos cuja transição para um novo estado é dada de acordo com regras probabilísticas que dependem dos estados da cadeia em um passado de tamanho fixo e finito. Esses modelos foram introduzidos no início do século XX pelo matemático russo Andrei Andreyevich Markov que buscava um modelo probabilístico para aparição de vogais em textos literários. Atualmente, as cadeias de Markov vem sendo amplamente utilizadas na modelagem de diversos fenômenos estocásticos que apresentam dinâmica temporal. Na neurociência, alguns autores têm considerado modelos Markovianos no processo de modelagem dos trens de disparos neuronais (ver, por exemplo, Linderman et al. 2016). Desta forma, antes de introduzirmos o modelo que utilizaremos na modelagem da conectividade neuronal, faremos um breve resumo sobre os modelos Markovianos. Neste sentido, iniciamos nossa revisão considerando as cadeias de Markov de ordem 1.

**Definição 3.7** *Um processo estocástico a tempo discreto  $X := \{X_t : t \in \mathbb{N}\}$ , definido em*

um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , é dito ser uma cadeia de Markov quando satisfazer a propriedade de Markov, ou seja,

$$P\left(X_t = a_t \mid \bigcap_{i=1}^t \{X_{t-i} = a_{t-i}\}\right) = P(X_t = a_t | X_{t-1} = a_{t-1}),$$

quaisquer que sejam  $t \in \mathbb{N}$  e  $(a_1, \dots, a_t) \in \mathcal{S}^{\wedge t}$ .

Portanto, nas cadeias de Markov de ordem 1, a probabilidade do estado atual do sistema depende apenas do estado anterior. Neste contexto, as probabilidades condicionais especificam a transição do sistema entre os possíveis estados levando em conta apenas o seu estado passado. Assumindo uma cadeia homogênea, as probabilidades de transição não dependem do valor do instante de tempo  $t \in \mathbb{N}$ . Logo, quando o espaço de estados é finito, é possível descrever a dinâmica do sistema a partir de uma matriz, a qual denominamos matriz de transição.

**Definição 3.8** Uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  de ordem 1 definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , possui matriz de transição  $Q$  tal que

$$Q = \begin{pmatrix} Q(a_1|a_1) & Q(a_2|a_1) & \dots & Q(a_n|a_1) \\ Q(a_1|a_2) & Q(a_2|a_2) & \dots & Q(a_n|a_2) \\ \vdots & \vdots & \ddots & \vdots \\ Q(a_1|a_n) & Q(a_2|a_n) & \dots & Q(a_n|a_n) \end{pmatrix},$$

em que assumimos, sem perda de generalidade,  $\mathcal{S} := \{a_1, a_2, \dots, a_n\}$  e as entradas da matriz  $Q$  são tais que

$$Q(b|a) := P(X_t = b | X_{t-1} = a),$$

quaisquer que sejam  $a, b \in \mathcal{S}$  e  $t \in \mathbb{N}$ .

Na prática, dizemos que a matriz de transição representa a evolução do sistema do estado passado (nas linhas) para o estado atual (nas colunas). Neste sentido, cada entrada da matriz é uma probabilidade de transição. Logo, a matriz de transição  $Q$  é uma matriz

estocástica, ou seja,

$$0 \leq Q(\cdot|a) \leq 1, \quad \forall a \in \mathcal{S},$$

$$\sum_{b \in \mathcal{S}} Q(b|a) = 1.$$

A condição inicial de um sistema Markoviano de ordem 1 é representada pelo valor que a variável aleatória  $X_0$  assume. Denominamos este valor de estado inicial e, então, atribuímos uma probabilidade à ocorrência de cada um dos possíveis estados iniciais. Portanto, o sistema é iniciado de maneira estocástica.

**Definição 3.9** *Uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  de ordem 1 definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , possui distribuição inicial  $\mu_0$  sobre  $\mathcal{S}$  quando*

$$\mu_0(a) := P(X_0 = a),$$

qualquer que seja  $a \in \mathcal{S}$ .

Desta forma,  $\mu_0$  é a distribuição de probabilidade de  $X_0$ . Além disso, se assumirmos, sem perda de generalidade,  $\mathcal{S} = \{a_1, a_2, \dots, a_n\}$ , podemos representar a distribuição inicial pelo vetor coluna

$$\mu_0 = \begin{pmatrix} \mu_0(a_1) \\ \mu_0(a_2) \\ \vdots \\ \mu_0(a_n) \end{pmatrix}.$$

Como vimos, a distribuição de uma cadeia é definida a partir da especificação de uma condição inicial e de uma dinâmica. No caso das cadeias de Markov de ordem 1, a condição inicial é dada pela distribuição inicial e a dinâmica é fixada a partir de uma matriz de transição. Assim, se  $X$  é uma cadeia de Markov de ordem 1 com espaço de estados finito  $\mathcal{S}$  segue, pela regra do produto de Bayes (ver, por exemplo, Teorema 31, página 37, Mood 1974), que

$$P(X_0 = a_0, \dots, X_t = a_t) = P(X_0 = a_0) \prod_{i=1}^t P(X_i = a_i | X_{i-1} = a_{i-1}),$$

quaisquer que seja  $a_t \in \mathcal{S}$  e  $t \in \mathbb{N}$ . Sendo assim, uma vez definidas a matriz de transição

$Q$  e a distribuição inicial  $\mu_0$  da cadeia  $X$ , temos

$$P(X_0 = a_0, \dots, X_t = a_t) = \mu_0(a_0) Q(a_1|a_0) \dots Q(a_t|a_{t-1}), \quad (3.10)$$

quaisquer que sejam  $a_t \in \mathcal{S}$  e  $t \in \mathbb{N}$ . A equação (3.10) é a distribuição da cadeia de Markov homogênea de ordem 1 com matriz de transição  $Q$  e distribuição inicial  $\mu_0$ . Portanto, a distribuição da cadeia de Markov  $X$  é determinada pela sua distribuição inicial  $\mu_0$  e pela sua matriz de transição  $Q$  (ver Teorema 2.1.4, página 66, Brémaud 2013).

Para cada instante de tempo  $t \in \mathbb{N}$  de uma cadeia de Markov homogênea  $X$  de ordem 1 com espaço de estados finito  $\mathcal{S}$ , definimos a distribuição de  $X_t$  através de uma medida de probabilidade  $\mu_t$  sobre o espaço de estados  $\mathcal{S}$  tal que

$$\mu_t(a) := P(X_t = a), \quad a \in \mathcal{S}.$$

Logo,  $\mu_t$  é a distribuição da cadeia de Markov de ordem 1 no  $t$ -ésimo passo. Além disso, podemos escrever tal distribuição em forma vetorial como foi feito com a distribuição inicial. Assim, pelo teorema da probabilidade total (ver, por exemplo, Teorema 29, página 35, Mood 1974) é possível escrever

$$\mu_t^\top = \mu_{t-1}^\top Q,$$

em que  $\top$  denota a transposta de um vetor ou de uma matriz. Por recursividade,

$$\mu_t^\top = \mu_0^\top Q^t,$$

sendo  $Q^t$  ( $t$ -vezes a matriz  $Q$ ) denominada matriz de transição em  $t$  passos.

Note que para todo  $a \in \mathcal{S}$ ,  $\mu_t(a)$  depende do instante de tempo  $t \in \mathbb{N}$ . No entanto, do ponto de vista neurobiológico, estamos interessados em sistemas sem dependência explícita da escolha do tempo inicial do sistema, logo vamos assumir  $\mu_t(a) = \mu_{t-1}(a)$  para todo  $t \in \mathbb{N}$ , o que nos leva à definição de distribuição estacionária.

**Definição 3.11** *Uma distribuição de probabilidade  $\pi$  sobre  $\mathcal{S}$  é dita ser a distribuição estacionária da matriz de transição  $Q$ , quando*

$$\pi^\top = \pi^\top Q. \quad (3.12)$$

Em particular,

$$\pi(a) = \sum_{b \in \mathcal{S}} \pi(b)Q(a|b), \quad \forall a \in \mathcal{S},$$

a qual é denominada equação de balanço do estado  $a$ .

**Definição 3.13** *Seja  $X := \{X_t : t \in \mathbb{N}\}$  uma cadeia de Markov homogênea de ordem 1 definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , matriz de transição  $Q$  e  $\mu_0$  como distribuição inicial. Dizemos que  $X$  é uma cadeia estacionária se  $\mu_0 = \pi$ , em que  $\pi$  é uma distribuição estacionária de  $Q$ .*

Desta forma, se  $X$  é uma cadeia estacionária, então iterando  $t$  vezes a equação (3.12) obtemos, para todo  $t \geq 0$ ,

$$\pi^\top = \pi^\top Q^t,$$

em que  $Q$  é matriz de transição de  $X$  e  $\pi$  é uma distribuição estacionária de  $Q$ . Portanto, o sistema, neste caso, é invariante quanto a escolha do tempo inicial  $t \in \mathbb{N}$ .

A fim de explorar o conceito de estacionariedade em um contexto neurobiológico, consideramos, no Exemplo 3.14, um modelo Markoviano homogêneo a tempo discreto de ordem 1 para a sequência de disparos neuronal.

**Exemplo 3.14** *Suponha que estejamos interessados em monitorar a atividade de um neurônio ao longo dos  $\mathbb{N}$  e que modelamos a evolução desse sistema através de uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  adequado, com valores no espaço de estados  $\mathcal{S} = \{0,1\}$  e com uma matriz de transição  $Q$  dada por*

$$Q := \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

em que  $p, q \in (0,1)$ .

Com o intuito de encontrar a cadeia estacionária  $\pi$  associada à matriz de transição  $Q$ , devemos encontrar  $\pi$  tal que as equações de equilíbrio sejam satisfeitas, i.e.,

$$\begin{aligned} \pi(0) &= \pi(0)(1-p) + \pi(1)q, \\ \pi(1) &= \pi(0)p + \pi(1)(1-q). \end{aligned}$$

Portanto,

$$\begin{aligned}\pi(0) &= \frac{q}{p+q}, \\ \pi(1) &= \frac{p}{p+q}.\end{aligned}$$

Desta forma, se  $\mu_0 = \pi := (\pi(0), \pi(1))^\top$ , a cadeia de Markov de ordem 1 com matriz de transição  $Q$  é estacionária.

Uma cadeia de Markov de ordem 1 homogênea com matriz de transição  $Q$  assumindo valores em um espaço de estados  $\mathcal{S}$  finito sempre possui pelo menos uma distribuição estacionária  $\pi$  sobre  $\mathcal{S}$  (ver, por exemplo, Ferrari & Galves 2000). A distribuição estacionária  $\pi$  também pode ser vista como o comportamento assintótico da distribuição  $\mu_t$ , assim, embora a sequência  $\{X_t : t \in \mathbb{N}\}$  não convirja para um estado fixo, a sequência de distribuições associadas  $\{\mu_t : t \in \mathbb{N}\}$  converge para a distribuição estacionária  $\pi$ . Na Seção 1.1 de Port & Stone 1972, por exemplo; os autores deduzem a distribuição estacionária que obtivemos no Exemplo 3.14 a partir do comportamento assintótico da cadeia.

As cadeias de Markov de ordem 1 também são restritivas do ponto de vista da modelagem de trens de disparos neuronais, pois a evolução do sistema leva em consideração apenas o seu estado no tempo anterior e há evidências empíricas que essa dependência é de ordem maior (ver, Karnup & Zhadin 1983; González-Montoro et al. 2011; Foik et al. 2014), isto decorre do fato de disparos neuronais apresentarem autocorrelações temporais, o que nos leva a considerar, por exemplo, cadeias de Markov de ordem  $k$  com  $k > 1$ .

### 3.3 Cadeias de Markov de ordem $k$

As cadeias de Markov de ordem  $k$ ,  $k > 1$ , são semelhantes aos modelos de ordem 1 no que diz respeito à dinâmica da cadeia, a qual também é dada por regras probabilísticas. Contudo, a probabilidade do sistema assumir um determinado estado no instante presente depende dos estados que ele assumiu em  $k$  instantes passados. Com o intuito de incorporar o impacto das autocorrelações temporais dos disparos neuronais, autores como Cessac (2008) utilizam cadeias de Markov de ordem  $k$  no processo de modelagem.

Antes de formalizarmos o conceito de cadeias de Markov de ordem  $k$ , vamos fixar algumas notações que serão úteis ao decorrer do texto. Dados dois inteiros  $m$  e  $n$  com  $m < n$ , denotamos por  $a_m^n$  a sequência finita  $(a_m, a_{m+1}, \dots, a_n)$  de estados de  $\mathcal{S}$  e por  $X_m^n$  o

vetor aleatório  $(X_m, X_{m+1}, \dots, X_n)$ . Além disso, o conjunto de todas as sequências finitas  $a_m^n$  é denotado por  $\mathcal{S}_m^n$ , exceto quando  $m = 1$ , caso em que utilizamos  $\mathcal{S}^n$ .

**Definição 3.15** *Um processo estocástico a tempo discreto  $X := \{X_t : t \in \mathbb{N}\}$ , definido em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , é dito ser uma cadeia de Markov de ordem  $k$ ,  $k$  um inteiro positivo, quando*

$$P(X_t = a_t | X_0^{t-1} = a_0^{t-1}) = P(X_t = a_t | X_{t-k}^{t-1} = a_{t-k}^{t-1}), \quad (3.16)$$

quaisquer que sejam  $t \in \mathbb{N}$  com  $t \geq k$ ,  $a_t \in \mathcal{S}$  e  $a_0^{t-1} \in \mathcal{S}^t$ .

Isto quer dizer que, ao utilizar cadeias de Markov de ordem finita  $k$ , a probabilidade de transição do sistema para um determinado estado depende somente dos últimos  $k$  estados. Semelhante às cadeias apresentadas anteriormente, ao assumir homogeneidade, as probabilidades de transição não dependem do valor do instante de tempo  $t \in \mathbb{N}$ . Portanto, quando o espaço de estados é finito, as cadeias de Markov de ordem  $k$  possuem sua dinâmica descrita também a partir de uma matriz de transição.

**Definição 3.17** *Uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  de ordem finita  $k$  definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , possui matriz de transição  $Q$  quando as entradas da matriz são tais que para cada  $a_0^{k-1} \in \mathcal{S}^k$*

$$Q(b | a_0^{k-1}) := P(X_t = b | X_{t-k}^{t-1} = a_0^{k-1}),$$

quaisquer que sejam  $b \in \mathcal{S}$  e  $t \in \mathbb{N}$ .

Neste contexto, a matriz de transição, para uma cadeia de Markov de ordem  $k$ , representa a evolução do sistema dos  $k$  estados passados (nas linhas) para o estado presente (nas colunas). Neste contexto, cada entrada da matriz é uma probabilidade de transição. Portanto, a matriz de transição  $Q$  é uma matriz estocástica, ou seja,

$$0 \leq Q(\cdot | a_0^k) \leq 1, \quad \forall a_0^k \in \mathcal{S}^k, \\ \sum_{b \in \mathcal{S}} Q(b | a_0^k) = 1.$$

Diferente dos modelos Markovianos de ordem 1, no contexto das cadeias de Markov

de ordem  $k$ , a condição inicial é um vetor de dimensão  $k$ . Novamente, a cada um dos possíveis vetores de estados iniciais é atribuída uma probabilidade. Logo, os modelos de ordem  $k$  também podem ser iniciados de maneira estocástica.

**Definição 3.18** *Uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  de ordem finita  $k$  definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$ , possui distribuição inicial  $\mu_0$  sobre  $\mathcal{S}^k$  quando*

$$\mu_0(a_0^{k-1}) := P(X_0^{k-1} = a_0^{k-1}),$$

qualquer que seja  $a_0^{k-1} \in \mathcal{S}^k$ .

Portanto,  $\mu_0$  é a distribuição de probabilidade de  $X_0^{k-1}$ . Ademais, podemos representar a distribuição inicial por um vetor coluna, em que cada entrada é cada uma das combinações de ordem  $k$  dos elementos de  $\mathcal{S}$ .

Definidas ambas condição inicial e a dinâmica, possuímos então a distribuição da cadeia. Assim como nas cadeias de Markov de ordem 1, a condição inicial das cadeias de Markov de ordem  $k$  é dada pela distribuição inicial e a dinâmica é fixada a partir de uma matriz de transição.

De acordo com as definições apresentadas anteriormente, é possível notar semelhanças entre as cadeias de Markov de ordens 1 e  $k$ . Na verdade, toda cadeia de Markov de ordem  $k$  no espaço de estados  $\mathcal{S}$ , pode ser reescrita como uma cadeia de Markov de ordem 1 em  $\mathcal{S}^k$  (ver, por exemplo, Proposição 2.1, página 15, Gripp 2019). Portanto, a existência de cadeias estacionárias de Markov de ordem  $k$  pode ser analisada a partir do estudo da existência de cadeias estacionárias de Markov de ordem 1. No Exemplo 3.19 exploramos como interpretar cadeias de Markov de ordem  $k$  a partir de cadeias de ordem 1, ao mesmo tempo que abordamos o conceito de estacionariedade.

**Exemplo 3.19** *Suponha que estejamos interessados em monitorar a atividade de um neurônio ao longo dos  $\mathbb{N}$  e que modelamos a evolução desse sistema através de uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  de ordem 2 definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  adequado, com valores no espaço de estados  $\mathcal{S} = \{0,1\}$  e com uma*

matriz de transição  $Q$  dada por

$$Q := \begin{array}{c} 0 \quad 1 \\ \begin{array}{l} 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{pmatrix} 1-p & p \\ q & 1-q \\ 1-s & s \\ r & 1-r \end{pmatrix} \end{array}$$

em que  $p, q, s, r \in (0,1)$ .

Com o intuito de encontrar a cadeia estacionária  $\pi$  associada à matriz de transição  $Q$ , devemos encontrar  $\pi$  tal que as equações de equilíbrio sejam satisfeitas, i.e.,

$$\pi(00) = (1-p)\pi(00) + (1-s)\pi(01),$$

$$\pi(01) = q\pi(10) + r\pi(11),$$

$$\pi(10) = p\pi(00) + s\pi(01),$$

$$\pi(11) = (1-q)\pi(10) + (1-r)\pi(11).$$

Portanto,

$$\begin{aligned} \pi(00) &= \frac{r(s-1)}{r(s-1) + p(-2r+q-1)}, \\ \pi(01) &= -\frac{pr}{r(s-1) + p(-2r+q-1)}, \\ \pi(10) &= -\frac{pr}{r(s-1) + p(-2r+q-1)}, \\ \pi(11) &= \frac{p(q-1)}{r(s-1) + p(-2r+q-1)}. \end{aligned}$$

Desta forma, se  $\mu_0 = \pi := (\pi(00), \pi(01), \pi(10), \pi(11))^\top$ , a cadeia de Markov  $X$  de ordem 2 com matriz de transição  $Q$  é estacionária.

Tome  $Y_t = (X_{t-1}, X_t)$ , para todo  $t \in \mathbb{N}$ , tal que

$$P(X_t = a_t | X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}) = P(Y_t = u_t | Y_{t-1} = u_{t-1}),$$

em que  $u_t = (a_{t-1}, a_t) \in \mathcal{S}^2$ ,  $t \in \mathbb{N}$ . Assim, a matriz de transição  $Q^*$  de  $Y$  pode ser

definida como

$$Q^* := \begin{array}{c} \begin{array}{cccc} & 00 & 01 & 10 & 11 \end{array} \\ \begin{array}{l} 00 \\ 01 \\ 10 \\ 11 \end{array} \left( \begin{array}{cccc} 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 1-s & s & 0 & 0 \\ 0 & 0 & r & 1-r \end{array} \right) \end{array}$$

em que  $p, q, s, r \in (0, 1)$ .

Logo,  $Y$  é uma cadeia de Markov homogênea de ordem 1 sobre  $\mathcal{S}^2$ .

Com o intuito de encontrar a cadeia estacionária  $\pi^*$  associada à matriz de transição  $Q^*$ , devemos encontrar  $\pi^*$  tal que as equações de equilíbrio sejam satisfeitas, i.e.,

$$\pi^*(00) = (1-p)\pi^*(00) + (1-s)\pi^*(10),$$

$$\pi^*(01) = p\pi^*(00) + s\pi^*(10),$$

$$\pi^*(10) = q\pi^*(01) + r\pi^*(11),$$

$$\pi^*(11) = (1-q)\pi^*(01) + (1-r)\pi^*(11).$$

Portanto,

$$\begin{aligned} \pi^*(00) &= \frac{r(s-1)}{r(s-1) + p(-2r + q - 1)}, \\ \pi^*(01) &= -\frac{pr}{r(s-1) + p(-2r + q - 1)}, \\ \pi^*(10) &= -\frac{pr}{r(s-1) + p(-2r + q - 1)}, \\ \pi^*(11) &= \frac{p(q-1)}{r(s-1) + p(-2r + q - 1)}. \end{aligned}$$

Desta forma, se  $\mu_0^* = \pi^* := (\pi^*(00), \pi^*(01), \pi^*(10), \pi^*(11))^\top$ , a cadeia de Markov  $Y$  de ordem 1, com matriz de transição  $Q^*$  é estacionária.

Ao compararmos  $\pi$  e  $\pi^*$ , vemos que,  $\pi = \pi^*$ . Logo, é possível encontrar tal distribuição tanto considerando a cadeia de Markov de ordem  $k$  quanto com a sua correspondente de ordem 1.

As cadeias de Markov de ordem finita  $k$  enfrentam dificuldades ao modelar trens de disparos neuronais, pois os neurônios possuem características, regiões e funções distintas entre si (veja, por exemplo, Karnup & Zhadin 1983). Logo, considerar uma mesma ordem

$k$  para estimar a atividade de diferentes neurônios pode prejudicar os resultados desejados. Trabalhos como Friston et al. (2010), Truccolo et al. (2010) e Cessac (2011b) tem mostrado que uma descrição Markoviana dos trens de disparos neuronais pode ser inadequada. Neste contexto, apresentamos, na próxima seção, um algoritmo para simular a sequência de disparos de um neurônio com base em um modelo neuronal simples. Veremos que, mesmo neste caso, a cadeia estocástica associada não é Markoviana, o que nos leva a considerar as cadeias de memória com alcance variável.

### 3.4 Cadeias de memória com alcance variável

As cadeias de alcance variável foram introduzidas por Rissanen (1983) como uma classe de processos estocásticos na qual as probabilidades de transição não dependem de uma quantidade fixa de estados passados. Desta forma, a dinâmica dos modelos com memória de alcance variável levam em conta a estrutura de dependência presente nos dados. Por conta disso, é possível apresentar uma representação pictórica do conjunto das sequências de estados passados (memória) como uma árvore, denominada árvore de contexto. Assim, considerando a memória da cadeia de tamanho variável obtemos modelos mais informativos, flexíveis e parcimoniosos do que as Cadeias de Markov.

Antes da formalização matemática das cadeias de alcance variável, é necessário fixar algumas notações que serão úteis ao decorrer do texto. Lembramos que, para todo  $n \geq 1$ ,  $\mathcal{S}^n$  é o conjunto das sequências finitas  $a_1^n$ . Quando  $n = 0$  a sequência  $a_1^n$  é a sequência vazia e a denotamos por  $\emptyset$ . Assim, o conjunto formado somente pela sequência vazia é denotado por  $\mathcal{S}^0$ , ou seja,  $\mathcal{S}^0 := \{\emptyset\}$ . Dessa forma, daqui em diante o conjunto de todas as sequências finitas será denotado como  $\mathcal{S}^*$ , isto é,  $\mathcal{S}^* := \cup_{j=0}^{\infty} \mathcal{S}^j$ . Além disso, o conjunto de todas as sequências infinitas à esquerda será denotado como  $\mathcal{S}^{\mathbb{Z}_-} := \mathcal{S}^{\{\dots, -2, -1\}}$ , em que,  $\mathbb{Z}_-$  denota o conjunto dos inteiros negativos. Por fim, a notação  $|\cdot|$  indicará a quantidade de elementos de uma dada sequência ou conjunto.

Árvore de contexto é um conjunto de sequências finitas ou infinitas à esquerda, que são sufixos de sequências passadas infinitas. Neste sentido, os sufixos são suficientes para definir a dinâmica da cadeia estocástica que é representada pela árvore de contexto. Antes de apresentarmos uma definição formal para árvore de contexto, vamos explicar em mais detalhes o que queremos dizer com “sufixo” e “árvore”.

**Definição 3.20** *Uma sequência de estados  $a_{-|a|}^{-1} \in \mathcal{S}^{|a|}$  é um sufixo de outra sequência*

(finita ou infinita)  $b_{-|b|}^{-1}$  quando  $|a| \leq |b|$  e existe uma seqüência  $l_{-|b|}^{-|a|-1}$  tal que  $b_{-|b|}^{-1} = l_{-|b|}^{-|a|-1} a_{-|a|}^{-1}$ . Neste caso, escrevemos  $a_{-|a|}^{-1} \prec b_{-|b|}^{-1}$ . Agora, quando  $a_{-|a|}^{-1} \prec b_{-|b|}^{-1}$  e  $a_{-|a|}^{-1} = b_{-|b|}^{-1}$ , escrevemos que  $a_{-|a|}^{-1} \preceq b_{-|b|}^{-1}$ .

**Definição 3.21** Um subconjunto  $\tau$  de  $\mathcal{S}^* \cup \mathcal{S}^{\mathbb{Z}^-}$  é uma árvore se nenhuma seqüência  $a_{-|a|}^{-1} \in \tau$  é um sufixo de outra seqüência  $b_{-|b|}^{-1} \in \tau$ ,  $|a| \leq |b|$ . Esta propriedade é chamada propriedade do sufixo.

A altura da árvore  $\tau$  é dada como

$$h(\tau) = \sup\{|a| : a \in \tau\}.$$

Informalmente falando, a altura da árvore  $\tau$  é a maior distância entre a raiz da árvore (desenhada no topo) e uma de suas folhas (desenhada no fim). Caso  $h(\tau) < +\infty$ , dizemos que  $\tau$  é limitada, isto é, o número de seqüências da árvore é finito e o denotamos por  $|\tau|$ . Por outro lado, se  $h(\tau) = +\infty$ , dizemos que  $\tau$  é ilimitada, ou seja,  $\tau$  é um conjunto infinito de seqüências.

**Definição 3.22** A árvore  $\tau$  é completa quando qualquer seqüência  $a_{-\infty}^{-1} \in \mathcal{S}^{\mathbb{Z}^-}$  possui um sufixo pertencente a  $\tau$ . Pela propriedade do sufixo segue que esse sufixo é único. Tal sufixo é chamado de contexto e o denotamos por  $c_\tau(a_{-\infty}^{-1})$ . Uma árvore completa é chamada de árvore de contexto.

**Exemplo 3.23** Considere  $\mathcal{S} = \{0,1\}$  como espaço de estados e considere as quatro árvores seguintes definidas sobre  $\mathcal{S}$ .

- (a) A primeira árvore  $\tau$  é  $\tau_1 = \{0,1,01\}$ . Logo,  $\tau_1$  é limitada, com  $h(\tau_1) = 2$ . Além disso,  $\tau_1$  não é árvore de contexto, uma vez que a seqüência 1 é um sufixo da seqüência 01, violando a propriedade do sufixo.
- (b) A árvore seguinte é  $\tau_2 = \{000,10,01,11\}$ . Logo,  $\tau_2$  é limitada, com  $h(\tau_2) = 3$ . Contudo,  $\tau_2$  não é uma árvore completa pois podemos substituir a seqüência 000 por 00 de tal forma que a propriedade do sufixo ainda seja satisfeita. Portanto,  $\tau_2$  não é árvore de contexto.
- (c) A próxima árvore é  $\tau_3 = \{00,10,1\}$ . Logo,  $\tau_3$  é limitada, com  $h(\tau_3) = 2$ . Além disso,  $\tau_3$  é uma árvore de contexto, pois é uma árvore completa e cada uma de suas seqüências satisfazem a propriedade do sufixo.

(d) Por fim a árvore  $\tau_4 = 0^{\mathbb{Z}^-} \cup \cup_{i \geq 1} 10_{-i}^1$ . Logo,  $\tau_4$  é ilimitada, com  $h(\tau_4) = +\infty$ . Além disso,  $\tau_4$  é uma árvore de contexto, pois é uma árvore completa e cada uma de suas seqüências satisfazem a propriedade do sufixo.

Cada uma das seqüências  $a_{-|a|}^{-1} \in \tau$  pode ser vista como um ramo que conecta a raiz da árvore (desenhada no topo) até uma folha. Por exemplo,  $d_{-k}^{-1}$  é uma seqüência de  $k$  arestas começando no topo e rotuladas, de cima pra baixo, por  $d_{-1}, d_{-2}, \dots, d_{-k}$ . Por conta disso,  $d_{-k}^{-1}$  é denominada de folha da árvore. No caso de árvores ilimitadas, temos seqüências infinitas à esquerda, isto é, folhas infinitas. Todas as seqüências finitas ou infinitas  $l_{-|l|}^{-1}$  tais que  $l_{-|l|}^{-1} \prec a_{-|a|}^{-1}$  qualquer que seja  $a_{-|a|}^{-1} \in \tau$ , com  $|a| \leq |l|$ , são chamadas de nós. Quando os nós são finitos eles são denominados nós internos. Os filhos do nó  $l_{-|l|}^{-1}$  são as seqüências  $bl_{-|l|}^{-1}, b \in \mathcal{S}$ . A raiz é identificada como a seqüência vazia.

Neste contexto, apresentamos na Figura 3.1 a representação pictorial das árvores  $\tau_1, \tau_2, \tau_3$  e  $\tau_4$  discutidas no Exemplo 3.23. Ao analisar as diferentes representações de árvores, é possível perceber que aquelas que são de contexto não possuem estados em seu nós e sim deixam suas seqüências terminarem em uma folha.

Podemos considerar sistemas que evoluem de um dado contexto de uma árvore  $\tau$  para um estado em  $\mathcal{S}$  de acordo com probabilidades de transição, o que nos leva a definição de árvores de contexto probabilística.

**Definição 3.24** *Uma árvore de contexto probabilística com estados em  $\mathcal{S}$  é um par ordenado  $(\tau, Q)$  tal que*

1.  $\tau$  é uma árvore de contexto;
2.  $Q := \left\{ Q(b|a_{-|a|}^{-1}) : b \in \mathcal{S} \text{ e } a_{-|a|}^{-1} \in \tau \right\}$  é uma família de probabilidades de transição sobre  $\mathcal{S}$ .

Neste contexto, a árvore de contexto probabilística pode ser vista como a representação de uma família de probabilidades de transição tais que, para cada passado, precisamos considerar apenas um contexto desse passado para definir a transição do sistema para o próximo estado. No Exemplo 3.25 apresentamos um exemplo de árvore de contexto probabilística.

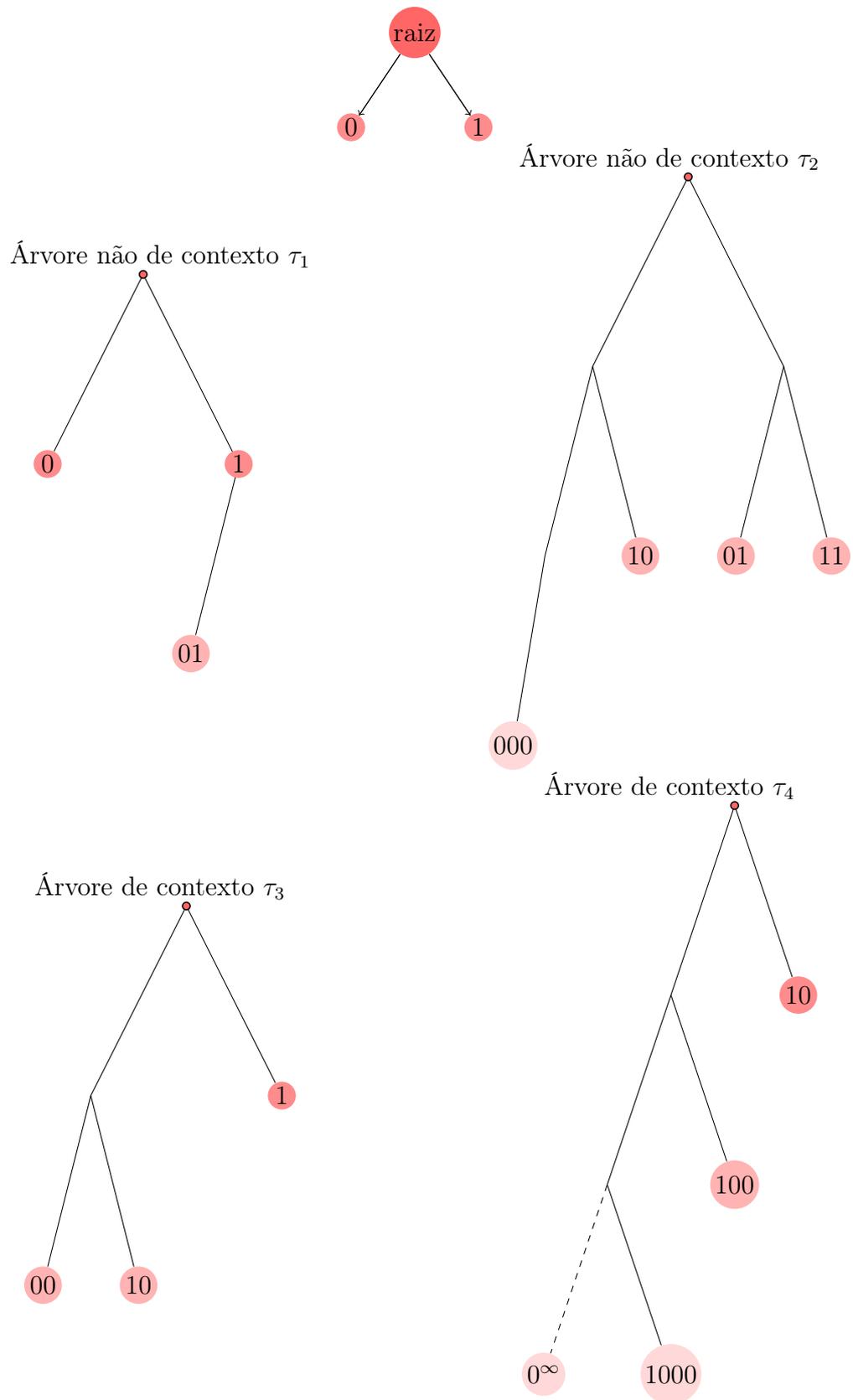


Figura 3.1: Representação pictorial das árvores definidas no Exemplo 3.23.

**Exemplo 3.25** Seja  $\tau = \{00,10,1\}$  uma árvore de contexto probabilística dada pelo par ordenado  $(\tau, Q)$ , tal que  $Q := \{Q(b|a_{-|a|}^{-1}) : b \in \mathcal{S} \text{ e } a_{-|a|}^{-1} \in \tau\}$  é uma família de probabilidades de transição sobre  $\mathcal{S}$ . Observe que podemos escrever  $Q$  em forma matricial, da seguinte maneira

$$Q := \begin{matrix} & 0 & 1 \\ \begin{matrix} 00 \\ 01 \\ 1 \end{matrix} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \\ 1-\gamma & \gamma \end{pmatrix} \end{matrix},$$

em que  $\alpha, \beta, \gamma \in (0,1)$ . Além disso, a representação pictorial de  $(\tau, Q)$  é dada por

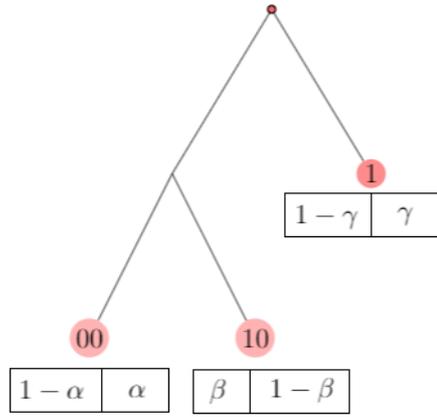


Figura 3.2: Representação pictorial da árvore de contexto probabilística  $\tau$ .

Podemos considerar cadeias estocásticas homogêneas que evoluem de acordo com árvores de contexto probabilísticas, as quais denominadas cadeias de memória com alcance variável.

**Definição 3.26** Dizemos que uma cadeia estocástica  $X := \{X_t : t \in \mathbb{Z}\}$  definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  é compatível com uma árvore de contexto probabilístico  $(\tau, Q)$  quando para  $P$ -quase toda  $a_{-\infty}^{-1} \in \mathcal{S}^{\mathbb{Z}^-}$  e qualquer  $b \in \mathcal{S}$  temos

$$P\{X_0 = b | X_{-\infty}^{-1} = a_{-\infty}^{-1}\} = Q(b|c_{\tau}(a_{-\infty}^{-1})).$$

Neste caso, dizemos que  $X$  é uma cadeia de memória com alcance variável.

Como já mencionamos, no processo de modelagem, estamos interessados em sistemas invariantes ao longo do tempo e, portanto, estamos sempre assumindo que as cadeias

estocásticas consideradas são estacionárias. Se o tamanho dos contextos em  $\tau$  é quase-certamente finito, então a distribuição estacionária compatível com  $(\tau, Q)$  existe (ver, por exemplo, Teorema 2, página 9, Ferreira et al. 2020). Portanto, este sempre será o caso neste trabalho.

Para justificar o uso das cadeias de memória de alcance variável no processo de modelagem dos trens de disparos neuronais, consideramos no Exemplo 3.27 um modelo neuronal simples e verificamos que, mesmo neste caso, a cadeia não é Markoviana.

**Exemplo 3.27** *Ao se trabalhar com a modelagem da sequências de disparos de um neurônio, consideramos uma cadeia estocástica  $V := \{V_t : t \in \mathbb{Z}\}$ , em que  $V_t$  denota o potencial de membrana do neurônio no tempo  $t \in \mathbb{Z}$ . Para  $t \in \mathbb{Z}$*

$$V_t = \begin{cases} v_R, & \text{se o neurônio dispara no tempo } t, \\ \rho V_{t-1} + \mu, & \text{caso contrário,} \end{cases}$$

em que  $0 \leq \rho < 1$  é o efeito da corrente de vazamento e  $\mu$  é a corrente de entrada.

Logo, ao disparar o neurônio retorna instantaneamente para o potencial de reinício, caso contrário, um estímulo constante  $\mu$  é adicionado ao seu potencial e uma variável  $0 \leq \rho \leq 1$  controla o efeito da corrente de vazamento recebida. Assim sendo, o espaço de estados do potencial da membrana é

$$\mathcal{V} := \{v_R, \rho v_R + \mu, \rho^2 v_R + \mu(1 + \rho), \rho^3 v_R + \mu(1 + \rho + \rho^2), \dots\}.$$

Além disso, podemos definir a probabilidade de um neurônio disparar no instante de tempo de  $t \in \mathbb{Z}$  dado o seu potencial de membrana no instante  $t - 1$  por

$$P(\{\text{o neurônio disparar no tempo } t\} | V_{t-1} = v_{t-1}) = \varphi(v_{t-1}),$$

em que  $v_{t-1} \in \mathcal{V}$  e  $\varphi : \mathbb{R} \rightarrow [0, 1]$  é uma função crescente denominada taxa de disparo.

Dessa maneira, seja a cadeia estocástica  $X := \{X_t : t \in \mathbb{Z}\}$  definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no espaço de estados  $\mathcal{S} = \{0, 1\}$  tal que para qualquer  $t \in \mathbb{Z}$

$$X_t = \begin{cases} 1, & \text{se } V_t = v_R, \\ 0, & \text{caso contrário.} \end{cases}$$

A cadeia  $X$  é de Markov? Para responder essa pergunta, considere, para cada  $t \in \mathbb{Z}$ , o valor observado  $u_t$  da variável aleatória  $U_t$  com distribuição uniforme em  $[0,1]$ . A priori, seja  $X_{t-1} = 1$ , então  $V_{t-1} = v_R$  e

$$X_t = \begin{cases} 1, & \text{caso } u_t \leq \varphi(v_R), \\ 0, & \text{caso contrário.} \end{cases}$$

Entretanto, se  $X_{t-1} = 0$ , então  $V_{t-1} \neq v_R$  e será necessário conhecer o valor de  $X_{t-2}$ . Logo, se  $X_{t-2} = 1$ , então  $V_{t-2} = v_R$  e

$$X_t = \begin{cases} 1, & \text{caso } u_t \leq \varphi(\rho v_R + \mu), \\ 0, & \text{caso contrário.} \end{cases}$$

Entretanto, se  $X_{t-2} = 0$ , então  $V_{t-2} \neq v_R$  e seria necessário observar  $X_{t-3}$  e assim sucessivamente. Portanto, é perceptível que o importante para o modelo é descobrir o último disparo ocorrido, ou seja, não existe uma dependência fixa em relação ao passado e sim uma dependência variável. Em outras palavras,  $X_t$  é uma função de  $X_{t-1}, X_{t-2}, X_{t-L_t}$  e de  $U_t$ , em que

$$L_t := \sup\{n \leq t : X_n = 1\}.$$

Concluimos que  $X$  não é uma cadeia de Markov e sim uma cadeia de alcance variável.

Neste contexto, podemos escrever a sequência de trens de disparos passados definida até o último disparo ocorrido, com a árvore de contexto  $\tau = \{1,10,100,1000, \dots\}$ , a qual possui a seguinte representação pictorial:

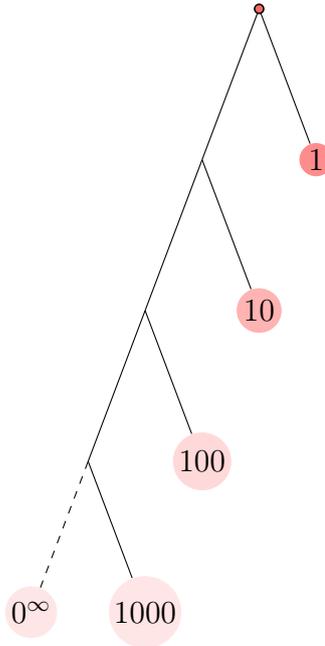


Figura 3.3: Representação pictorial da árvore  $\tau = \{1,10,100,1000, \dots\}$ .

Assim, a cadeia estocástica  $X$  é compatível com a árvore de contexto probabilística  $(\tau, Q)$ , em que  $Q := \{Q(b | a_{-|a|}^{-1}) : b \in \{0,1\} \text{ e } a_{-|a|}^{-1} \in \tau\}$  é tal que

$$Q(1 | a_{-|a|}^{-1}) = \begin{cases} \varphi(v_R), & \text{se } |a| = 1, \\ \varphi(\rho^{|a|-1} v_R + \mu \sum_{i=0}^{|a|-2} \rho^i), & \text{se } 2 \leq |a| < +\infty, \\ \varphi(\frac{\mu\rho}{1-\rho}), & \text{se } |a| = +\infty. \end{cases}$$

Em um contexto de inferência estatística, as cadeias de Markov de ordem finita exigem muitos parâmetros para serem inferidos. Por exemplo, se  $X$  é uma cadeia de Markov de ordem  $k$ , então existem  $(|\mathcal{S}| - 1) |\mathcal{S}|^k$  probabilidades de transição a serem estimadas. Por outro lado, se  $X$  é uma cadeia de memória com alcance variável compatível com um árvore de contexto  $\tau$  tal que  $h(\tau) = k < +\infty$ , então uma coleção de  $(|\mathcal{S}| - 1) |\tau|$  probabilidades de transição é suficiente para descrever a cadeia. Portanto, em um processo de estimação a quantidade de parâmetros a serem estimados na abordagem Markoviana é superior ao da abordagem via cadeias de memória com alcance variável. Dentro desta perspectiva, notamos uma vantagem inferencial ao se considerar cadeias compatíveis com árvores de contexto probabilística, justificando, mais uma vez, o seu uso na modelagem dos trens de disparos neuronais.



# Capítulo 4

## Inferência em processos estocásticos

Inferir é o ato de realizar conclusões a partir de um conjunto de dados observados. Uma vez definido o modelo probabilístico a ser utilizado para modelar a sequência de disparos neuronal, podemos estudar métodos estatísticos para realizar inferência a cerca da dinâmica desses disparos a partir de uma amostra observada. Na prática, em geral, além de não conhecermos as probabilidades de transição que regem a dinâmica do sistema, também desconhecemos qual a árvore de contexto compatível com ele. Portanto, este capítulo tem como objetivo apresentar métodos para estimar as probabilidades de transição e selecionar o modelo mais parcimonioso, isto é, aquele que envolva uma quantidade mínima de parâmetros capazes de explicar bem o comportamento da variável de interesse.

### 4.1 Inferência em cadeias de Markov de ordem $k$

Com o intuito de facilitar o entendimento dos métodos de estimação e seleção de modelo, a primeira abordagem considerada neste trabalho é em um contexto Markoviano. A extensão para as cadeias com memória de alcance variável é bem natural e direta. Primeiramente vamos realizar a estimação das probabilidades de transição pelo método da máxima verossimilhança e, em seguida, vamos estudar uma técnica para a seleção da ordem da cadeia.

#### 4.1.1 Estimação pelo método da máxima verossimilhança

O método da máxima verossimilhança é uma das técnicas de estimação mais populares. Dada uma amostra observada e um modelo estatístico, este método consiste em encontrar

os valores dos parâmetros do modelo que maximizam a probabilidade de ocorrência dos dados observados.

**Definição 4.1** *Sejam  $n$  um número inteiro positivo e  $X_0^{n-1}$  uma sequência de variáveis aleatórias definidas em um mesmo espaço de probabilidade  $(\Omega, \mathcal{F}, P_\theta)$  com valores em um conjunto finito  $\mathcal{S}$ , em que  $\theta \in \mathbb{R}^d$ ,  $d$  um inteiro positivo. Para cada  $\theta \in \mathbb{R}^d$  usamos  $f(\cdot|\theta)$  para denotar a função de distribuição de probabilidade conjunta de  $X_0^{n-1}$ , i.e.,*

$$f(x_0^{n-1}|\theta) := P_\theta \{X_0^{n-1} = x_0^{n-1}\}, \quad \forall x_0^{n-1} \in \mathcal{S}^n,$$

em que  $\theta$  é dito ser um vetor de parâmetros e  $x_0^{n-1}$  são os valores observados. Neste contexto, a função de verossimilhança  $L(\cdot|x_0^{n-1})$ , para uma amostra  $x_0^{n-1} \in \mathcal{S}^n$  fixada, é definida por

$$L(\theta|x_0^{n-1}) := f(x_0^{n-1}|\theta), \quad \theta \in \mathbb{R}^d.$$

Dessa forma,  $\hat{\theta} \in \mathbb{R}^d$  é dito ser um estimador de máxima verossimilhança de  $\theta$  quando

$$\hat{\theta} = \hat{\theta}(x_0^{n-1}) = \arg \max_{\theta \in \mathbb{R}^d} L(\theta|x_0^{n-1}),$$

com  $x_0^{n-1} \in \mathcal{S}^n$  fixada.

Seja  $X := \{X_t : t \in \mathbb{N}\}$  uma cadeia de Markov homogênea de ordem finita  $k$  definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P_Q)$  com valores em um espaço de estados finito  $\mathcal{S}$  e com probabilidades de transição dadas por

$$Q := \{Q(b|a_0^{k-1}) : b \in \mathcal{S} \text{ e } a_0^{k-1} \in \mathcal{S}^k\},$$

Dada uma amostra  $x_0^{n-1} \in \mathcal{S}^n$  da cadeia de Markov  $X$  de ordem  $k$ ,  $n > k$ , podemos escrever a função de verossimilhança associada à matriz de transição  $Q$  da seguinte forma

$$\begin{aligned} L(Q|x_0^{n-1}) &= P_Q\{X_0^{n-1} = x_0^{n-1}\} \\ &= P_Q\{X_0^{k-1} = x_0^{k-1}\} \prod_{t=k}^{n-1} P_Q\{X_t = x_t | X_{t-k}^{t-1} = x_{t-k}^{t-1}\} \\ &= P_Q\{X_0^{k-1} = x_0^{k-1}\} \prod_{b \in \mathcal{S}} \prod_{a_0^{k-1} \in \mathcal{S}^k} Q(b|a_0^{k-1})^{\mathcal{N}_n(a_0^{k-1}b)}, \end{aligned}$$

em que  $\mathcal{N}_n(a_0^{k-1}b)$  é o número de vezes que a sequência  $a_0^{k-1}$  seguida pelo estado  $b$  foi

identificada na amostra  $x_0^{n-1}$ , i.e.,

$$\mathcal{N}_n(a_0^{k-1}b) = \sum_{t=k}^{n-1} \mathbb{I}\{x_{t-k}^{t-1} = a_0^{k-1}, x_t = b\}.$$

Assumindo  $P_Q\{X_0^{k-1} = x_0^{k-1}\} = 1$ , queremos encontrar

$$\hat{Q}_n := \left\{ \hat{Q}_n(b|a_0^{k-1}) : b \in \mathcal{S} \text{ e } a_0^{k-1} \in \mathcal{S}^k \right\}$$

que maximiza a função de verossimilhança  $L(Q|x_0^n)$ ,  $x_0^{n-1} \in \mathcal{S}^n$  fixada, isto é,

$$\hat{Q}_n = \arg \max_{Q \in \mathcal{M}_k(\mathcal{S})} L(Q|x_0^{n-1}),$$

em que  $\mathcal{M}_k(\mathcal{S})$  é o conjunto de todas as matrizes de transição definidas sobre  $\mathcal{S}^k \times \mathcal{S}$ .

Defina  $\ell(Q|x_0^{n-1})$  a função de log da verossimilhança associada à matriz de transição  $Q$ , i.e.,

$$\ell(Q|x_0^{n-1}) := \log L(Q|x_0^{n-1}),$$

em que  $x_0^{n-1} \in \mathcal{S}^n$  está fixada e tomamos log na base  $e$ . Assim,

$$\ell(Q|x_0^{n-1}) = \sum_{b \in \mathcal{S}} \sum_{a_0^{k-1} \in \mathcal{S}^k} \mathcal{N}_n(a_0^{k-1}b) \log Q(b|a_0^{k-1}).$$

Neste contexto, queremos encontrar  $Q \in \mathcal{M}_k(\mathcal{S})$  que maximiza  $\ell(Q|x_0^{n-1})$  quando  $x_0^{n-1} \in \mathcal{S}^n$  está fixado, sujeito à restrição

$$\sum_{b \in \mathcal{S}} Q(b|a_0^{k-1}) = 1, \quad \forall a_0^{k-1} \in \mathcal{S}^k.$$

Por conta disso, utilizamos o método dos multiplicadores de Lagrange. Sejam, então,  $\lambda := \{\lambda_{a_0^{k-1}} \in \mathbb{R} : a_0^{k-1} \in \mathcal{S}^k\}$  e  $F(\lambda, Q)$  definida por

$$F(\lambda, Q) := \sum_{a_0^{k-1} \in \mathcal{S}^k} \left\{ \sum_{b \in \mathcal{S}} \mathcal{N}_n(a_0^{k-1}b) \log Q(b|a_0^{k-1}) + \lambda_{a_0^{k-1}} \left[ 1 - \sum_{b \in \mathcal{S}} Q(b|a_0^{k-1}) \right] \right\}.$$

Como  $F(\lambda, Q)$  é diferenciável em  $\lambda$  temos

$$\frac{d}{d\lambda_{a_0^{k-1}}} F(\lambda, Q) = 1 - \sum_{b \in \mathcal{S}} Q(b|a_0^{k-1}); \quad \forall a_0^{k-1} \in \mathcal{S}^k.$$

Ao igualarmos a zero, segue que

$$\sum_{b \in \mathcal{S}} Q(b|a_0^{k-1}) = 1, \quad \forall a_0^{k-1} \in \mathcal{S}^k. \quad (4.2)$$

Além disso,  $F(\lambda, Q)$  é também diferenciável em relação à  $Q$ , logo

$$\frac{d}{dQ(b|a_0^{k-1})} F(\lambda, Q) = \mathcal{N}_n(a_0^{k-1}b) \frac{1}{Q(b|a_0^{k-1})} - \lambda_{a_0^{k-1}}; \quad \forall a_0^{k-1} \in \mathcal{S}^k \text{ e } \forall b \in \mathcal{S}.$$

Igualando à zero, obtemos

$$\hat{Q}_n(b|a_0^{k-1}) = \frac{\mathcal{N}_n(a_0^{k-1}b)}{\lambda_{a_0^{k-1}}}, \quad \forall a_0^{k-1} \in \mathcal{S}^k \text{ e } \forall b \in \mathcal{S}. \quad (4.3)$$

Das equações (4.2) e (4.3), segue que

$$\sum_{b \in \mathcal{S}} \frac{\mathcal{N}_n(a_0^{k-1}b)}{\lambda_{a_0^{k-1}}} = 1 \Rightarrow \lambda_{a_0^{k-1}} = \sum_{b \in \mathcal{S}} \mathcal{N}_n(a_0^{k-1}b), \quad \forall a_0^{k-1} \in \mathcal{S}^k.$$

Portanto, o estimador de máximo verossimilhança  $\hat{Q}_n := \{ \hat{Q}_n(b|a_0^{k-1}) : b \in \mathcal{S} \text{ e } a_0^{k-1} \in \mathcal{S}^k \}$  é

$$\hat{Q}_n(b|a_0^{k-1}) = \frac{\mathcal{N}_n(a_0^{k-1}b)}{\sum_{b \in \mathcal{S}} \mathcal{N}_n(a_0^{k-1}b)}; \quad \forall a_0^{k-1} \in \mathcal{S}^k \text{ e } \forall b \in \mathcal{S}, \quad (4.4)$$

e quando  $\sum_{b \in \mathcal{S}} \mathcal{N}_n(a_0^{k-1}b) = 0$ , definimos  $\hat{Q}_n(b|a_0^{k-1}) = 0$ .

É importante lembrar que estamos interessados em sistemas que são modelados por cadeias estacionárias. Quando  $X$  é estacionária é possível mostrar que o estimador de máxima verossimilhança  $\hat{Q}_n$  é fortemente consistente (Galves & Leonardi (2008)), ou seja,

$$\hat{Q}_n(b|a_0^{k-1}) = \frac{\frac{\mathcal{N}_n(a_0^{k-1}b)}{n}}{\frac{\mathcal{N}_{n-1}(a_0^{k-1})}{n}} \xrightarrow{q.c.} \frac{\pi(a_0^{k-1})Q(b|a_0^{k-1})}{\pi(a_0^{k-1})} = Q(b|a_0^{k-1}).$$

No Exemplo 4.5 consideramos uma sequência de disparos neuronais observada e utilizamos a metodologia de estimação supracitada para estimar a matriz de transição da cadeia de Markov de ordem  $k$  utilizada como modelo para dinâmica neuronal em questão.

**Exemplo 4.5** *Suponha que estejamos interessados em monitorar a atividade de um neurônio ao longo dos  $\mathbb{N}$  e que modelamos a evolução desse sistema através de uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P_Q)$ ,*

com valores no espaço de estados  $\mathcal{S} = \{0,1\}$  e com matriz de transição  $Q$ . Seja  $(x_0, \dots, x_{999})$  uma amostra da cadeia  $X$ , em que o número de ocorrências das sequências de tamanhos 1, 2 e 3 foi registrado, respectivamente, conforme a Tabela 4.1.

Tabela 4.1: Número de aparições das sequências de tamanhos 1, 2 e 3 na amostra.

$a_0$	$a_1$	$a_2$	$\mathcal{N}_{1000}(a_0a_1a_2)$
0	0	0	118
0	0	1	114
0	1	0	274
1	0	0	113
0	1	1	69
1	0	1	229
1	1	0	68
1	1	1	14
$a_0$	$a_1$	$\mathcal{N}_{1000}(a_0a_1)$	$\mathcal{N}_{999}(a_0a_1)$
0	0	232	232
0	1	343	343
1	0	342	342
1	1	83	82
$a_0$	$\mathcal{N}_{1000}(a_0)$	$\mathcal{N}_{999}(a_0)$	
0	575	575	
1	425	424	

Estimamos, pelo método da máxima verossimilhança, as matrizes de probabilidades de transição considerando a ordem da cadeia como 0 (caso independente), 1 e 2.

Primeiramente, no caso de ordem 0, temos  $\hat{Q}_{1000} := \{\hat{Q}_{1000}(b) : b \in \mathcal{S}\}$  tal que

$$\hat{Q}_{1000}(b) = \frac{\mathcal{N}_{1000}(b)}{\sum_{b \in \mathcal{S}} \mathcal{N}_{1000}(b)} = \frac{\mathcal{N}_{1000}(b)}{\mathcal{N}_{1000}(0) + \mathcal{N}_{1000}(1)}.$$

Logo,

$$\hat{Q}_{1000}(0) = \frac{575}{1000} \quad e \quad \hat{Q}_{1000}(1) = \frac{425}{1000}.$$

Portanto, a matriz de probabilidades de transição estimada é dada por  $\hat{Q}_{1000}$

$$\hat{Q}_{1000} := \begin{pmatrix} 0 & 1 \\ 0.575 & 0.425 \end{pmatrix}$$

No caso de ordem 1, temos

$$\hat{Q}_{1000} := \{\hat{Q}_{1000}(b|a_0) : b \in \mathcal{S} \text{ e } a_0 \in \mathcal{S}\}$$

tal que

$$\hat{Q}_{1000}(b|a_0^{k-1}) = \frac{\mathcal{N}_{1000}(a_0^{k-1}b)}{\sum_{b \in \mathcal{S}} \mathcal{N}_{1000}(a_0^{k-1}b)} = \frac{\mathcal{N}_{1000}(a_0^{k-1}b)}{\mathcal{N}_{999}(a_0^{k-1})}.$$

Logo,

$$\begin{aligned} \hat{Q}_{1000}(0|0) &= \frac{232}{575} & e & \hat{Q}_{1000}(1|0) = \frac{343}{575} \\ \hat{Q}_{1000}(0|1) &= \frac{342}{424} & e & \hat{Q}_{1000}(1|1) = \frac{83}{424} \end{aligned}$$

Portanto, a matriz de probabilidades de transição estimada é dada por  $\hat{Q}_{1000}$

$$\hat{Q}_{1000} := \begin{matrix} & 0 & 1 \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} .403 & .597 \\ .804 & .196 \end{pmatrix} \end{matrix}$$

Por fim, consideramos o caso de ordem 2, sabemos que

$$\hat{Q}_{1000} := \left\{ \hat{Q}_{1000}(b|a_0^1) : b \in \mathcal{S} \text{ e } a_0^1 \in \mathcal{S}^2 \right\}$$

tal que

$$\hat{Q}_{1000}(b|a_0^{k-1}) = \frac{\mathcal{N}_{1000}(a_0^{k-1}b)}{\sum_{b \in \mathcal{S}} \mathcal{N}_{1000}(a_0^{k-1}b)} = \frac{\mathcal{N}_{1000}(a_0^{k-1}b)}{\mathcal{N}_{999}(a_0^{k-1})}$$

Logo,

$$\begin{aligned} \hat{Q}_{1000}(0|00) &= \frac{118}{232} & e & \hat{Q}_{1000}(1|00) = \frac{114}{232} \\ \hat{Q}_{1000}(0|10) &= \frac{274}{343} & e & \hat{Q}_{1000}(1|10) = \frac{69}{343} \\ \hat{Q}_{1000}(0|01) &= \frac{113}{342} & e & \hat{Q}_{1000}(1|01) = \frac{229}{342} \\ \hat{Q}_{1000}(0|11) &= \frac{68}{82} & e & \hat{Q}_{1000}(1|11) = \frac{14}{82} \end{aligned}$$

Portanto, a matriz de probabilidades de transição é dada por  $\hat{Q}_{1000}$

$$\hat{Q}_{1000} := \begin{matrix} & 0 & 1 \\ \begin{matrix} 00 \\ 01 \\ 10 \\ 11 \end{matrix} & \begin{pmatrix} .508 & .492 \\ .798 & .202 \\ .330 & .670 \\ .829 & .171 \end{pmatrix} \end{matrix}.$$

### 4.1.2 Seleção da ordem da cadeia

Dada uma amostra observada, a seleção de modelos é a tarefa de selecionar um modelo estatístico a partir de um conjunto de modelos candidatos de acordo com algum critério pré-estabelecido. No contexto de cadeias de Markov, a seleção de modelos se resume a determinar a ordem da cadeia, ou seja, temos o conhecimento que as cadeias possuem ordem finita, mas desconhecemos tal ordem. Assim, com base na comparação das probabilidades estimadas para algum  $k$  com as estimadas para  $k + 1$ , vamos escolher como ordem da cadeia o menor valor de  $k$  para o qual as probabilidades de transição em comparação são suficientemente próximas.

Seja  $X := \{X_t : t \in \mathbb{N}\}$  uma cadeia de Markov homogênea de ordem finita  $k$  definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P_Q)$  com valores em um espaço de estados finito  $\mathcal{S}$  e com probabilidades de transição

$$Q := \{Q(b|a_0^{k-1}) : b \in \mathcal{S} \text{ e } a_0^{k-1} \in \mathcal{S}^k\}.$$

Dada uma amostra  $x_0^{n-1} \in \mathcal{S}^n$  da cadeia de Markov  $X$  de ordem finita  $k$ ,  $k$  desconhecido e tal que  $k \ll n$ , o problema de estimar a ordem  $k$  da cadeia de Markov consiste em determinar o menor valor  $\hat{k}_n$  para o qual  $\hat{Q}_n^{(k)}$  seja suficientemente próxima de  $\hat{Q}_n^{(k+1)}$ , ou seja,

$$\hat{k}_n = \min\{k \geq 0 : \hat{Q}_n^{(k)} \approx \hat{Q}_n^{(k+1)}\},$$

em que  $\hat{Q}_n^{(k)}$  denota a matriz de transição estimada quando assumimos que a ordem da cadeia é  $k$ . Na definição acima,  $\hat{k}_n \ll n$  significa  $\hat{k}_n < \lfloor \frac{\log_{|\mathcal{S}|} n}{2} \rfloor$  (ver Rissanen (1983)) e  $\hat{Q}_n^{(k)} \approx \hat{Q}_n^{(k+1)}$  significa que

$$\max_{b \in \mathcal{S}} \max_{c \in \mathcal{S}} |Q_n^{(k)}(b|a_{-k}^{-1}) - Q_n^{(k+1)}(b|a_{-k-1}^{-1}c)| \leq \delta, \quad \forall a_{-k}^{-1} \in \mathcal{S}^k,$$

em que  $\delta \in (0,1)$  é o limiar fixado. Em outras palavras, queremos testar sequencialmente a hipótese nula

$$H_0 : Q_n^{(k)}(b|a_{-k}^{-1}) = Q_n^{(k+1)}(b|a_{-k-1}^{-1}), \quad \forall b \in \mathcal{S} \text{ e } \forall a_{-k-1}^{-1} \in \mathcal{S}^{k+1}$$

utilizando a estatística de teste

$$\Delta_n(a_{-k}^{-1}) = \max_{b \in \mathcal{S}} \max_{c \in \mathcal{S}} |Q_n^{(k)}(b | a_{-k}^{-1}) - Q_n^{(k+1)}(b | a_{-k}^{-1}c)|,$$

até encontrarmos um valor  $\hat{k}_n$  para o qual a hipótese nula seja rejeitada. Neste caso, rejeitamos  $H_0$  quando  $\Delta_n(a_{-k}^{-1}) > \delta$  para algum  $a_{-k}^{-1} \in \mathcal{S}^k$ .

Para que a estimação da ordem da cadeia seja factível, precisamos considerar amostras para as quais a matriz de transição estimada seja estocástica. Neste sentido, fixamos a Definição 4.6.

**Definição 4.6** *Sejam  $n$  um número inteiro positivo e  $x_0^{n-1} \in \mathcal{S}^n$  uma amostra de uma cadeia de Markov  $X$  de ordem finita  $k$  desconhecida,  $k < n$ . A amostra  $x_0^{n-1}$  é  $k$ -admissível quando para toda sequência  $a_0^{k-1} \in \mathcal{S}^k$  temos  $\mathcal{N}_n(a_0^{k-1}) \geq 1$ .*

Dada uma amostra  $k$ -admissível  $x_0^{n-1}$  de uma cadeia de Markov  $X$  em que  $k$  é desconhecido, podemos solucionar o problema de estimação da ordem da cadeia a partir do seguinte algoritmo.

**Algoritmo:**

1. Catalogar todas as sequências de tamanho  $d(n) = \lfloor \frac{\log_{|\mathcal{S}|} n}{2} \rfloor$  que aparecem na amostra  $x_0^{n-1}$ .
2. Definir  $\delta \in (0,1)$  pequeno.
3. Para cada sequência  $a_{-d(n)}^{-1}$  catalogada defina  $k = d(n) - 1$  e calcule

$$\Delta_n(a_{-k}^{-1}) = \max_{b \in \mathcal{S}} \max_{c \in \mathcal{S}} |Q_n^{(k)}(b | a_{-k}^{-1}) - Q_n^{(k+1)}(b | a_{-k}^{-1}c)|.$$

4. Se  $\max_{a_{-k}^{-1} \in \mathcal{S}^k} \Delta_n(a_{-k}^{-1}) < \delta$ , então descarta-se a ordem  $k + 1$  da cadeia e definimos  $d(n) = k$  e repetimos o algoritmo a partir do passo (3). Caso contrário, definimos  $\hat{k}_n = k + 1$ , ou seja,

$$\hat{k}_n = \max \left\{ k : \max_{a_{-k}^{-1} \in \mathcal{S}^k} \Delta_n(a_{-k}^{-1}) \geq \delta \right\}.$$

No Exemplo 4.7 realizamos a simulação de uma cadeia de Markov de ordem 2 e aplicamos o algoritmo apresentado a uma amostra admissível a partir das estimativas

de máxima verossimilhança das matrizes de transição e concluimos que a ordem estimada é, de fato, 2.

**Exemplo 4.7** *Suponha que estejamos interessados em monitorar a atividade de um neurônio ao longo dos  $\mathbb{N}$  e que modelamos a evolução desse sistema através de uma cadeia de Markov homogênea  $X := \{X_t : t \in \mathbb{N}\}$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P_Q)$ , com valores no espaço de estados  $\mathcal{S} = \{0,1\}$  e com matriz de transição  $Q$ . Seja  $(x_0, \dots, x_{29})$  uma amostra da cadeia  $X$ , de ordem desconhecida  $k$ , em que o número de ocorrências das sequências de tamanhos 1, 2 e 3 foi registrado, respectivamente, conforme a Tabela 4.2*

Tabela 4.2: Número de aparições das sequências de tamanhos 1, 2 e 3 na amostra.

$a_0$	$\mathcal{N}_{30}$	$\mathcal{N}_{29}$
0	18	17
1	12	12
$a_0 a_1$	$\mathcal{N}_{30}$	$\mathcal{N}_{29}$
00	12	11
01	5	5
10	5	5
11	7	7
$a_0 a_1 a_2$	$\mathcal{N}_{30}$	$\mathcal{N}_{29}$
000	8	7
001	3	3
010	2	2
100	3	3
011	3	3
101	2	2
110	3	3
111	4	4

Com o intuito de seguir o algoritmo apresentado anteriormente, precisamos, primeiramente, encontrar o valor de  $d(n)$ , ou seja,

$$\begin{aligned}
 d(n) &= \left\lfloor \frac{\log_{|\mathcal{S}|} n}{2} \right\rfloor \\
 &= \left\lfloor \frac{\log_2 30}{2} \right\rfloor \\
 &\cong 2.
 \end{aligned}$$

Sendo  $d(30) = 2$  e considerando  $\delta = 10^{-1}$  estimamos as matrizes de transição até a ordem  $d(30) = 2$ .

$$\hat{Q}_{30}^0 := \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} .6 & .4 \end{pmatrix} \end{array} \quad \hat{Q}_{30}^1 := \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} .706 & .294 \\ .417 & .583 \end{pmatrix} \end{array}$$

$$\hat{Q}_{30}^2 := \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \begin{pmatrix} .725 & .275 \\ .4 & .6 \\ .6 & .4 \\ .428 & .572 \end{pmatrix} \end{array}$$

Como, para toda sequência  $a_0^{n-1} \in \mathcal{S}^n$ , temos  $N_n(a_0^{d(30)-1}) \geq 1$ , concluímos que a amostra  $(x_0, \dots, x_{29})$  é admissível para  $d(30) = 2$ . Assim, definimos  $k = d(30) - 1$  e calculamos

$$\Delta_{30}(a) = \max_{b \in \mathcal{S}} \max_{c \in \mathcal{S}} |Q^{(1)}(b|a) - Q^{(2)}(b|ac)|, \quad a \in \{0,1\}.$$

Neste sentido, temos

$$\begin{aligned} & \max\{|\hat{Q}_{30}^1(0|0) - \hat{Q}_{30}^2(0|00)|, |\hat{Q}_{30}^1(0|1) - \hat{Q}_{30}^2(0|10)|, |\hat{Q}_{30}^1(0|0) - \hat{Q}_{30}^2(0|01)|, |\hat{Q}_{30}^1(0|1) - \hat{Q}_{30}^2(0|11)|\} \\ &= \max\{|0.706 - 0.725|, |0.294 - 0.4|, |0.706 - 0.6|, |0.294 - 0.428|\} \\ &= \max\{0.19, 0.106, 0.106, 0.134\} \\ &= 0.134. \end{aligned}$$

e

$$\begin{aligned} & \max\{|\hat{Q}_{30}^1(1|0) - \hat{Q}_{30}^2(1|00)|, |\hat{Q}_{30}^1(1|1) - \hat{Q}_{30}^2(1|10)|, |\hat{Q}_{30}^1(1|0) - \hat{Q}_{30}^2(1|01)|, |\hat{Q}_{30}^1(1|1) - \hat{Q}_{30}^2(1|11)|\} \\ &= \max\{|0.294 - 0.375|, |0.583 - 0.6|, |0.294 - 0.4|, |0.583 - 0.572|\} \\ &= \max\{0.081, 0.017, 0.106, 0.011\} \\ &= 0.106. \end{aligned}$$

Portanto,

$$\max\{0.134, 0.106\} = 0.134.$$

Como  $0.134 > 0.1$ , concluímos que  $\hat{k}_{30} = 2$  é a ordem da cadeia.

## 4.2 Inferência em cadeias de memória com alcance variável

É importante notar que os estimadores de máxima verossimilhança das probabilidades de transição para o caso das cadeias de memória com alcance variável são análogos aos do caso das cadeias de Markov de ordem finita. Em outras palavras, se  $X := \{X_t : t \in \mathbb{Z}\}$  é uma cadeia de memória com alcance variável definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$ , compatível com uma árvore de contexto probabilística  $(\tau, Q)$ , e com valores em um espaço de estados  $\mathcal{S}$ , então

$$P\{X_0 = b | X_{-\infty}^{-1} = a_{-\infty}^{-1}\} = Q(b | c_\tau(a_{-\infty}^{-1})).$$

para uma dada amostra  $x_0^{n-1} \in \mathcal{S}^n$  de  $X$ , o estimador de máxima verossimilhança de  $Q$  é  $\hat{Q}_n := \left\{ \hat{Q}_n(b | a_{-|a|}^{-1}) : b \in \mathcal{S} \text{ e } a_{-|a|}^{-1} \in \tau \right\}$ , em que

$$\hat{Q}_n(b | a_{-|a|}^{-1}) = \frac{\mathcal{N}_n(a_{-|a|}^{-1} b)}{\sum_{b \in \mathcal{S}} \mathcal{N}_n(a_{-|a|}^{-1} b)}; \quad \forall a_{-|a|}^{-1} \in \tau \text{ e } \forall b \in \mathcal{S}.$$

Além disso, Para estimar a árvore de contexto  $\tau$  podemos usar o mesmo método utilizado para a estimação da ordem de uma cadeia de Markov, como mostramos na próxima subseção.

### 4.2.1 Algoritmo contexto

O algoritmo contexto para estimação de árvores de contextos compatíveis com cadeias de memória com alcance variável segue a mesma ideia do algoritmo que apresentamos para a seleção de ordem de cadeias de Markov. O algoritmo contexto foi introduzido por Rissanen (1983) e, calcula, para cada nó da árvore uma medida de discrepância entre as probabilidades de transição associada a este contexto e àquela associada ao contexto acrescido de um passo a mais no passado. Iniciamos com uma árvore candidata que possui todas as folhas até um nível  $d(n)$  fixado; se a medida de discrepância é maior que um limiar fixado, os contextos são mantidos na árvore; caso contrário, eles são podados. Realizamos este procedimento até obter a menor árvore de contexto bem adaptada a amostra. A diferença deste algoritmo com o utilizado para seleção da ordem de cadeias de Markov é que ao invés de comparar todas as probabilidades de um nível  $k + 1$  com as do nível  $k$ ,

analisa-se os nós da árvore de contexto individualmente e compara-se as probabilidades de transição das folhas com seus respectivos nós. Assim, reduzimos o número de parâmetros a serem estimados e produzimos, portanto, um modelo mais parcimonioso, reduzindo também o custo computacional.

**Definição 4.8** *Sejam  $n$  um número inteiro positivo e  $x_0^{n-1} \in \mathcal{S}^n$  uma amostra de uma cadeia de memória com alcance variável  $X$ . Uma árvore de contexto admissível com altura  $k$ , sendo  $k$  um inteiro positivo, é qualquer árvore  $\tau$  satisfazendo*

i)  $a_{-|a|}^{-1} \in \tau$ , quando,  $h(a_{-|a|}^{-1}) \leq k$  e  $\mathcal{N}_n(a_{-|a|}^{-1}) \geq 1$ ,

ii) Qualquer sequência  $a_{-|a|}^{-1} \in \mathcal{S}^*$  com  $\mathcal{N}_n(a_{-|a|}^{-1}) \geq 1$  é sufixo de algum  $l_{-|l|}^{-1} \in \tau$  ou tem um sufixo  $l_{-|l|}^{-1} \in \tau$ ,  $|a| < |l|$ .

Dada uma amostra  $x_0^{n-1} \in \mathcal{S}^n$  de uma cadeia de memória com alcance variável  $X$  compatível com a árvore de contexto probabilística  $(\tau, Q)$ , em que  $\tau$  é uma árvore admissível com altura  $d(n)$ , podemos solucionar o problema de estimação da árvore de contexto a partir do seguinte algoritmo

**Algoritmo contexto:**

1. Catalogar todas as sequências de tamanho  $d(n) = \lfloor \frac{\log_{|\mathcal{S}|} n}{2} \rfloor$  que aparecem na amostra  $x_0^{n-1}$ .
2. Definir  $\delta \in (0,1)$  pequeno.
3. Para cada sequência catalogada defina  $k = d(n) - 1$  e calcule

$$\Delta_n(a_{-k}^{-1}) = \max_{b \in \mathcal{S}} \max_{c \in \mathcal{S}} \left| \hat{Q}_n^{(k)}(b | a_{-k}^{-1}) - \hat{Q}_n^{(k+1)}(b | a_{-k}^{-1}c) \right|.$$

4. Se  $\Delta_n(a_{-k}^{-1}) < \delta$ , então poda-se a folha, i.e., descarta-se o nível  $k + 1$  associado ao nó  $a_{-k}^{-1}$ . Caso contrário, não realizamos a poda e tomamos  $ba_{-k}^{-1} \in \hat{\tau}$ . Vamos dizer, neste caso, que o nó  $a_{-k}^{-1}$  foi marcado com o símbolo  $\mathcal{T}$ .
5. Realizamos o passo 4 para toda sequência catalogada.
6. Definimos  $d(n) = k$  e, para todo nó do nível  $d(n) - 1$  que não possui folha marcada com o símbolo  $\mathcal{T}$  repetimos o procedimento a partir do passo 3.

7. Ao final deste processo definimos  $\hat{\tau}$  como a árvore de contexto estimada.

No Exemplo 4.9 realizamos a simulação de uma cadeia de memória com alcance variável e aplicamos o algoritmo apresentado a uma árvore admissível e a partir das estimativas de máxima verossimilhança das probabilidades de transição aplicamos o algoritmo contexto para encontrar a árvore de contextos irredutível.

**Exemplo 4.9** *Suponha que estejamos interessados em monitorar a atividade de um neurônio ao longo dos  $\mathbb{Z}$  e que modelamos a evolução desse sistema através de uma cadeia com memória de alcance variável  $X := \{X_t : t \in \mathbb{Z}\}$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P_Q)$ , com valores no espaço de estados  $\mathcal{S} = \{0,1\}$  e compatível com uma árvore de contexto probabilística  $(\tau, Q)$ . Seja  $(x_0, \dots, x_{255})$  uma amostra da cadeia  $X$  em que o número de ocorrências das sequências de tamanhos 1,2,3,4 e 5 foi registrado, respectivamente, conforme as Tabelas 4.3,4.4 e 4.5.*

Com o intuito de seguir o algoritmo apresentado anteriormente, precisamos, primeiramente, encontrar o valor de  $d(256)$ , ou seja,

$$\begin{aligned} d(256) &= \lfloor \frac{\log_2 256}{2} \rfloor \\ &= 4. \end{aligned}$$

Considerando  $\delta = 10^{-1}$  construímos a árvore candidata  $\hat{\tau}$  com todas as folhas de tamanho  $d(256) = 4$  (ver Figura 4.9).

Posteriormente, estimamos as matrizes de transição até a ordem  $d(256) = 4$ .

$$\hat{Q}_{256}^0 := \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{cc} 0 & 1 \end{array} & \begin{pmatrix} .51 & .49 \end{pmatrix} \end{array} \quad \hat{Q}_{256}^1 := \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{cc} 0 & 1 \end{array} & \begin{pmatrix} .346 & .654 \\ .672 & .328 \end{pmatrix} \end{array}$$

$$\hat{Q}_{256}^2 := \begin{array}{ccc} & & \begin{array}{cc} 0 & 1 \end{array} \\ & \begin{array}{cc} 00 & 01 \end{array} & \begin{pmatrix} .24 & .76 \\ .6 & .4 \end{pmatrix} \\ \begin{array}{cc} 10 & 11 \end{array} & \begin{pmatrix} .4 & .6 \\ .85 & .15 \end{pmatrix} & \end{array}$$

Como, para toda sequência  $a_{-4}^{-1} \in \hat{\tau}$ , temos  $\mathcal{N}_n(a_{-4}^{-1}) \geq 1$ , concluímos que a árvore candidata  $\hat{\tau}$  é admissível para  $d(256) = 4$ .



permitindo podar as folhas. O mesmo processo foi realizado para todos os ramos e o resultado obtido é apresentado na Figura 4.1.

Contudo, não existe mais nenhum nó não associado a uma folha com símbolo  $\mathcal{T}$ , portanto, a árvore de contextos irredutível estimada é

$$\hat{\tau} = \{0000, 1000, 100, 0010, 1010, 110, 0001, 1001, 110, 0001, 1001, 101, 011, 0111, 1111\}.$$

Tabela 4.3: Número de aparições das sequências de tamanhos 1, 2 e 3 na amostra.

$a_0$	$\mathcal{N}_{256}$	$\mathcal{N}_{255}$
0	130	130
1	126	125
$a_0a_1$	$\mathcal{N}_{256}$	$\mathcal{N}_{255}$
00	45	45
01	85	85
10	84	84
11	41	40
$a_0a_1a_2$	$\mathcal{N}_{256}$	$\mathcal{N}_{255}$
000	11	11
001	34	34
010	50	50
100	34	34
011	35	35
101	50	50
110	34	34
111	6	6

Tabela 4.4: Número de aparições das sequências de tamanho 4 na amostra.

$a_0a_1a_2a_3$	$\mathcal{N}_{256}$	$\mathcal{N}_{255}$
0000	1	1
0001	10	10
0010	17	17
0100	20	20
1000	11	11
1001	23	23
1010	31	31
1100	14	14
1101	20	20
1110	5	5
1111	1	1
0011	17	16
0110	28	28
1011	19	19
0101	30	30
0111	6	6

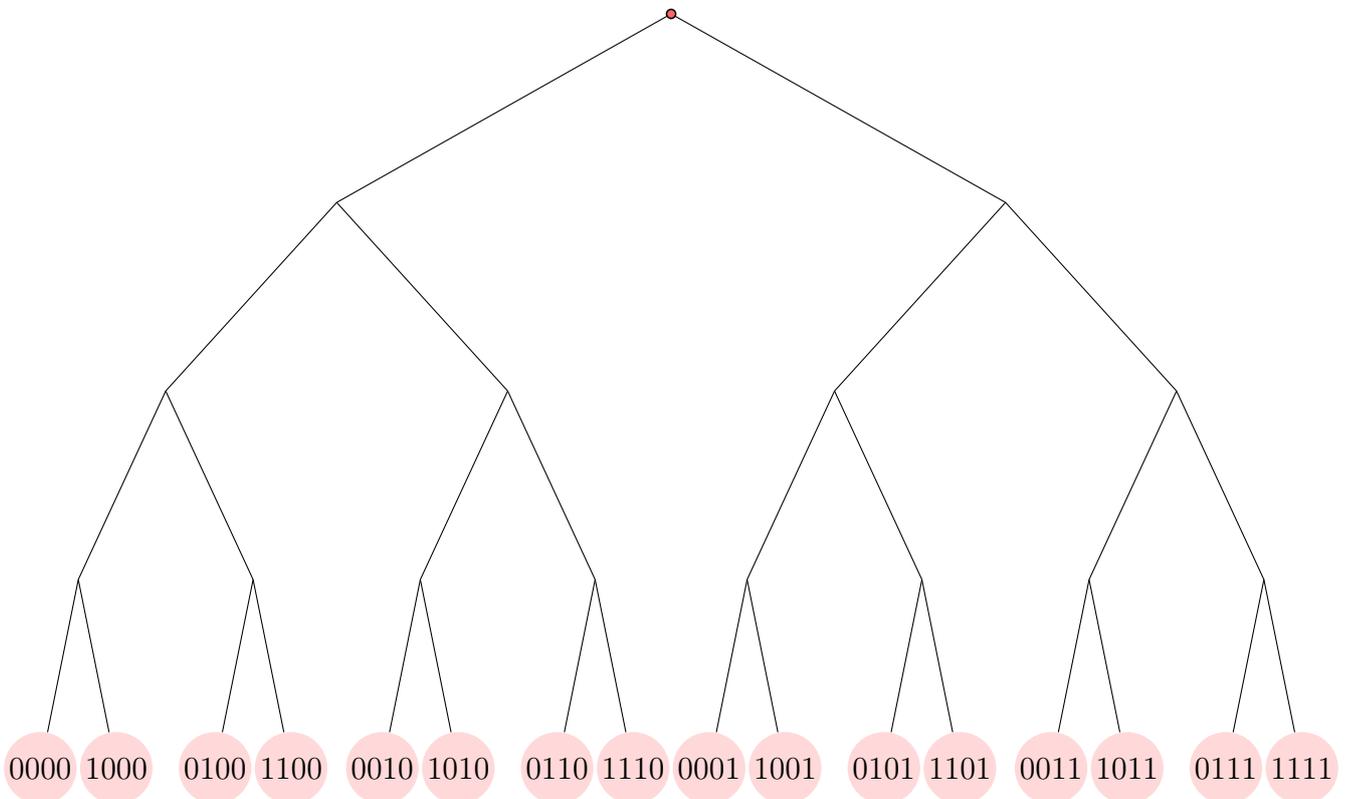


Tabela 4.5: Número de aparições das sequências de tamanhos 5 na amostra.

$a_0a_1a_2a_3a_4$	$\mathcal{N}_{256}$	$\mathcal{N}_{255}$
00000	0	0
00001	1	1
00010	5	5
00011	5	5
00100	7	7
00101	10	10
01000	5	5
01001	15	15
10000	1	1
10001	10	10
10010	13	13
10011	10	9
10100	16	16
10101	15	15
11000	7	7
11001	7	7
11010	10	10
11011	10	10
11100	2	2
11101	3	3
00110	9	9
00111	7	7
01100	21	21
01101	7	7
10110	13	13
10111	6	6
01010	21	21
01011	9	9
01110	5	5
01111	1	1
11110	1	1
11111	0	0

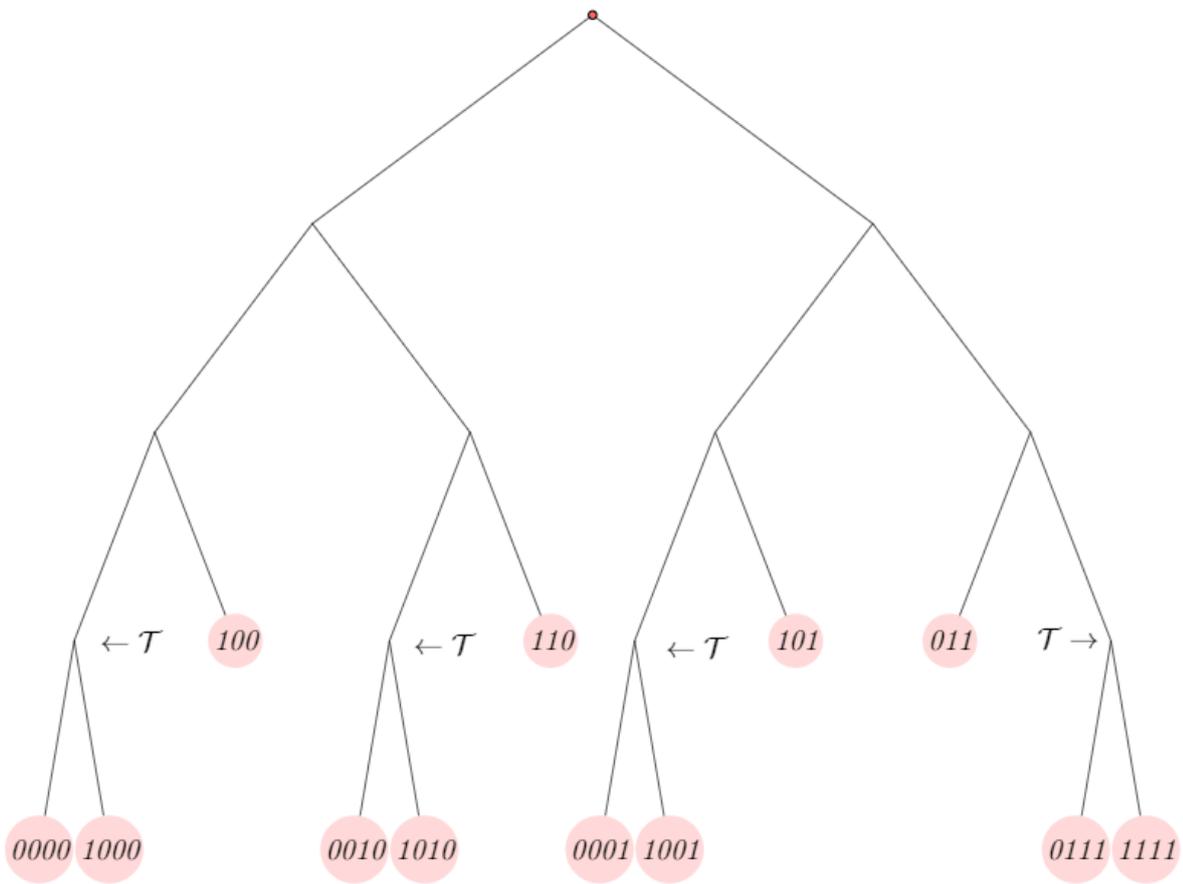


Figura 4.1: Árvore de contextos obtidas após uma interação do algoritmo.

## Capítulo 5

# Medidas da Teoria da Informação

A descrição de como diferentes estruturas do cérebro estão relacionadas umas com as outras têm um importante papel no entendimento das funções cerebrais. A Teoria da Informação é uma candidata natural para fixar medidas acerca da conectividade neuronal. Essa teoria foi introduzida por Shannon (1948) e vem sendo amplamente utilizada em neurociência para, por exemplo, quantificar a transmissão de informação entre neurônios e inferir os diferentes níveis de conectividade do sistema neural. As medidas da informação são modelos livres, os quais não precisamos assumir quaisquer hipóteses a cerca de uma estrutura específica de interação entre variáveis em um conjunto de dados. Esses métodos tendem a ser mais simples do que os métodos baseados em modelos paramétricos, além de quantificarem uma gama maior de interações e fenômenos por não estarem limitados a hipóteses do modelo assumido. O principal objetivo deste trabalho é inferir a transmissão de informação entre neurônios ao longo de um intervalo de tempo finito com base em um estudo de simulação e em um conjunto de dados reais. Neste sentido, assumimos que a sequência de disparos neuronais é gerada a partir de uma cadeia estocástica com memória de alcance variável e utilizamos medidas da informação para entender como a informação é carregada através de momentos distintos de um neurônio e quais propriedades podem ser verificadas entre tais neurônios e certos estímulos. Portanto, apresentamos, neste capítulo, alguns conceitos da teoria da informação que serão fundamentais para compreensão do estudo de simulação e da aplicação em questão.

## 5.1 Informação

Sem levar em conta as características particulares dos eventos observáveis de um determinado experimento, a teoria desenvolvida por Shannon busca definir informação através de uma abordagem axiomática. A primeira premissa considerada por ele como verdadeira é que a informação extraída da observação de um evento é uma função da probabilidade do evento ocorrer ou não. A segunda propriedade axiomática é que a informação é uma função positiva da probabilidade. Assim, o terceiro axioma fixado por Shannon é que um evento com probabilidade 1 resulta em nenhuma informação, pois é nada surpreendente. Mais especificamente, quanto mais provável é um evento, menos surpreendente ele é e, portanto, menos informação ele contém. Por fim, ele fixou que a quantidade total de informação obtida a partir da observação de  $n$  eventos independentes é a soma das informações de cada um desses eventos.

Neste sentido, informação é uma função  $I : (0,1] \rightarrow [0, +\infty)$  tal que

1.  $I(p) \geq 0, \forall p \in (0,1]$ ;
2.  $I(1) = 0$ ;
3.  $I(p) \rightarrow 0$ , quando  $p \rightarrow 1$ ;
4.  $I(\prod_{i=1}^n p_i) = \sum_{i=1}^n I(p_i)$ .

A única função que possui as propriedades axiomáticas de 1 a 4, especificadas anteriormente, é uma função logarítmica (ver, por exemplo, pág.84 , Dieudonné 1969). Assim, a informação extraída da observação de um evento que possui probabilidade  $p$  de ocorrer é

$$I(p) = -K \log_b p,$$

em que  $K > 0$  é uma constante de proporcionalidade, portanto assumimos  $K = 1$ ; e  $b$  é uma base arbitrária que nos fornece a unidade da informação, assumimos  $b = 2$ , i.e., bits.

**Definição 5.1** *Sejam  $(\Omega, \mathcal{F}, P)$  um espaço de probabilidade e  $A$  um evento de  $\mathcal{F}$  tal que  $P(A) = p$  com  $0 < p \leq 1$ . A informação própria de  $A$  é dada por*

$$I_A(p) := -\log_2 p.$$

**Exemplo 5.2** *Suponha que estejamos interessados em monitorar a atividade de um neurônio que dispara com probabilidade  $\frac{1}{2}$  e queremos saber qual a informação própria associada ao evento  $A$ : “o neurônio disparou”. Neste caso,*

$$I_A(p) = -\log_2 p = -\log_2 \frac{1}{2} = \log_2 2 = 1 \text{ bit}$$

Note que, neste caso,  $I_A(p)$  é a quantidade de perguntas do tipo “sim ou não” ou o tamanho do código binário necessário para codificar todas as respostas (casos) possíveis.

**Exemplo 5.3** *Suponha que estejamos interessados em monitorar a atividade de um neurônio que pode receber um estímulo excitatório (1) com probabilidade  $\frac{2}{5}$ , um estímulo inibitório (2) com probabilidade  $\frac{2}{5}$  e o estímulo pode falhar (3) com probabilidade  $\frac{1}{5}$ . Seja  $X$  a variável aleatória que indica o tipo de estímulo recebido pelo neurônio.*

*A função de distribuição de probabilidade  $\mu$  de  $X$  é dada por*

$x$	1	2	3	$\Sigma$
$\mu(x)$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	1

*A informação própria dos eventos elementares de  $X$  é dada por*

$x$	1	2	3
$I_{\{X=x\}}(\mu)$	$\cong 1.32$	$\cong 1.32$	$\cong 2.32$

Observe que a informação própria dos eventos elementares associados a uma variável aleatória  $X$  depende da distribuição de probabilidade de  $X$ . Assim, se  $\mu$  é a distribuição de probabilidade de  $X$ , então  $I_X(\mu) := -\log_2 \mu(X)$  é uma variável aleatória. A pergunta natural que surge é: qual a informação média associada à variável aleatória  $X$ ? É o que respondemos na próxima seção.

## 5.2 Entropia, entropia conjunta e entropia condicional

Entropia é uma grandeza que quantifica a desordem de um sistema e/ou a incerteza de um evento. Shannon, define a entropia como a medida da informação de uma dada mensagem que está relacionada com frequência dos diferentes símbolos que foram transmitidos (Shannon 1948).

**Definição 5.4** *Seja  $X$  uma variável aleatória discreta definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\mathcal{X}$  e com função de distribuição*

de probabilidade  $\mu(x) = P(X = x)$ ,  $x \in \mathcal{X}$ . A entropia  $H(X)$  da variável aleatória  $X$  é o valor esperado de  $I_X(\mu)$ , ou seja,

$$H(X) := E[I_X(\mu)] = - \sum_{x \in \mathcal{X}} \mu(x) \log \mu(x).$$

Note que a entropia  $H(X)$  é uma função da distribuição de  $X$  não dependendo, portanto, dos valores que  $X$  assume, mas sim das probabilidades dos eventos elementares associados a esses valores. Desta forma,  $H(X) \geq 0$  e  $H(X) \leq \log |\mathcal{S}|$  (ver, Teorema I.6.3, página 57, Shields 1996).

**Exemplo 5.5** *Suponha que estejamos interessados em monitorar a atividade de um neurônio e modelamos sua atividade em um tempo particular através de uma variável aleatória  $X$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\{0,1\}$  e com função de distribuição de probabilidade  $\mu(x) = P(X = x)$ ,  $x \in \{0,1\}$ . Assim,*

$$X = \begin{cases} 1, & \text{se o neurônio disparar,} \\ 0, & \text{caso contrário.} \end{cases}$$

e

$$\mu(1) = 1 - \mu(0) = p, \quad p \in (0,1].$$

Podemos calcular a entropia da variável aleatória  $X$  da seguinte forma

$$H(X) = -p \log p - (1 - p) \log(1 - p), \quad p \in (0,1].$$

Na Figura 5.1 podemos observar que a entropia é máxima quando  $p = 1/2$  e seu valor é  $H(X) = 1\text{bit}$ . Esse é o cenário com mais incerteza envolvida. Além disso,  $H(X)$  é uma função côncava tal que  $H(X) = 0$  quando  $p = 1$ , cenário em que  $X$  é degenerada e, portanto, não existe incerteza.

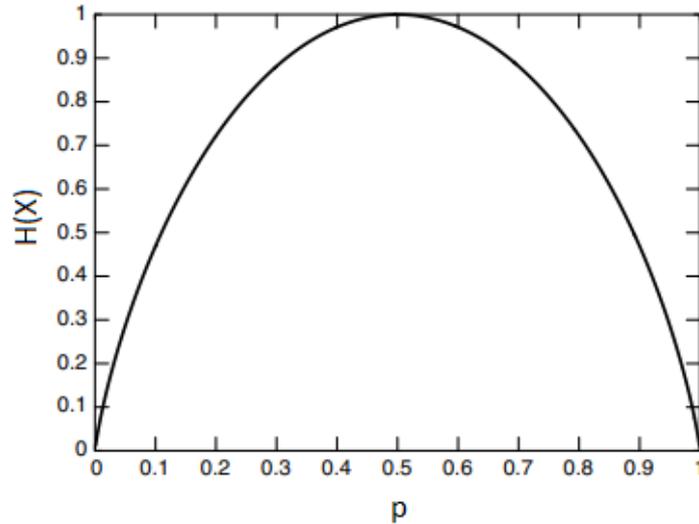


Figura 5.1:  $H(X)$  em função de  $p$  quando  $X$  é uma variável aleatória binária. Adaptado de Cover & Thomas (1991).

A definição de entropia para uma variável aleatória, pode naturalmente ser estendida para um vetor aleatório bidimensional. Neste caso, ela é denominada entropia conjunta.

**Definição 5.6** *Seja  $(X,Y)$  um vetor aleatório discreto definido em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\mathcal{X} \times \mathcal{Y}$  e com função de distribuição de probabilidade conjunta  $\mu(x,y) := P(X = x, Y = y)$ ,  $(x,y) \in \mathcal{X} \times \mathcal{Y}$ . A entropia conjunta  $H(X,Y)$  do vetor aleatório  $(X,Y)$  é definida por*

$$H(X,Y) := E[I_{X,Y}(\mu)] = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mu(x,y) \log \mu(x,y).$$

Analogamente ao caso da entropia conjunta, podemos definir a entropia para distribuições condicionais.

**Definição 5.7** *Seja  $(X,Y)$  um vetor aleatório discreto definido sobre um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\mathcal{X} \times \mathcal{Y}$  e com função de distribuição de probabilidade conjunta  $\mu(x,y) := P(X = x, Y = y)$ ,  $(x,y) \in \mathcal{X} \times \mathcal{Y}$  e distribuição marginal  $\mu(x) = P(X = x)$ ,  $x \in \mathcal{X}$ , com correspondente distribuição condicional  $\mu(y|x) = \mu(x,y)/\mu(x)$ . A entropia condicional  $H(Y|X)$  é definida como*

$$H(Y|X) := E [I_{Y|X}(\mu)] = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mu(x,y) \log \mu(y|x),$$

Além disso, a entropia condicional de  $Y$  dado um evento particular  $X = x$ ,  $x \in \mathcal{X}$  é definida como

$$H(Y|X = x) := \sum_{y \in \mathcal{Y}} \mu(y|x) \log \mu(y|x).$$

Assim, podemos reescrever  $H(Y|X)$  da seguinte forma

$$H(Y|X) = \sum_{x \in \mathcal{X}} \mu(x) H(Y|X = x).$$

Uma identidade importante que relaciona a entropia conjunta com a entropia condicional é a seguinte lei da adição

$$H(X,Y) = H(X) + H(Y|X).$$

A demonstração desse resultado é uma consequência da propriedade aditiva do logaritmo (ver, Teorema 2.2.1, página 16, Cover & Thomas 1991).

**Exemplo 5.8** Suponha que estejamos interessados em monitorar a atividade de dois neurônios e modelamos sua atividade através do par de variáveis aleatórias  $(X,Y)$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\{0,1\} \times \{0,1\}$  e com função de distribuição conjunta  $\mu(x,y) = P(X = x, Y = y)$ ,  $(x,y) \in \{0,1\} \times \{0,1\}$  dada por

$y/x$	0	1	$P(Y=y)$
0	1/4	1/4	1/2
1	1/4	1/4	1/2
$P(X=x)$	1/2	1/2	1

A informação própria dos eventos elementares de  $(X,Y)$ .

$y/x$	0	1	$I_{Y=y}(\mu)$
0	2	2	1
1	2	2	1
$I_{X=x}(\mu)$	1	1	

Logo, a entropia conjunta de  $(X,Y)$  é

$$H(X,Y) = 4 \times \left( 2 \times \frac{1}{4} \right) = 2 \text{ bits.}$$

Além disso, a entropia condicional de  $Y|X$  é

$$H(Y|X) = 2 \times \left( 2 \times \frac{1}{4} \right) = 1 \text{ bits.}$$

e a entropia condicional de  $X|Y$  é

$$H(X|Y) = 2 \times \left( 2 \times \frac{1}{4} \right) = 1 \text{ bits.}$$

Neste caso,  $H(Y|X) = H(X|Y)$ . Todavia, é importante notar que isso nem sempre é verdade.

### 5.3 Entropia de uma cadeia estacionária

Seja  $X = \{X_n : n \in \mathbb{Z}\}$  uma cadeia estacionária definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$  e com distribuição de probabilidade  $\mu$ . Para quaisquer inteiros  $m$  e  $n$  com  $m \leq n$  o cilindro determinado pela sequência  $a_m^n$  de símbolos de  $\mathcal{S}$ , denotado por  $[a_m^n]$ , é o subconjunto de  $\mathcal{S}^{\mathbb{Z}}$  definido por

$$[a_m^n] = \{x \in \mathcal{S}^{\mathbb{Z}} : x_i = a_i, m \leq i \leq n\}.$$

Assim, podemos definir

$$\pi([a_m^n]) = P(X_m^n = a_m^n),$$

qualquer que seja a sequência  $a_m^n$  de símbolos de  $\mathcal{S}$ .

**Definição 5.9** *Seja  $X := \{X_n : n \in \mathbb{Z}\}$  uma cadeia estacionária definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$  e com função de distribuição de probabilidade  $\pi$ . A entropia de  $X$  associada a blocos de tamanho  $k$  é dada por*

$$H(X_1^k) = - \sum_{a_1^k \in \mathcal{S}^k} \pi([a_1^k]) \log \pi([a_1^k]).$$

Essa medida captura quantitativamente as correlações de alcance menores do que  $k$ .

**Definição 5.10** *Seja  $X := \{X_n : n \in \mathbb{Z}\}$  uma cadeia estacionária definida em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um espaço de estados finito  $\mathcal{S}$  e com*

função de distribuição de probabilidade  $\pi$ . A entropia condicional de  $X$  associada a blocos de tamanho  $k$  é dada por

$$H(X_k|X_1^{k-1}) = - \sum_{a_1^k \in \mathcal{X}^k} \pi([a_1^k]) \log \frac{\pi([a_1^k])}{\pi([a_1^{k-1}])},$$

em que, por convenção, tomamos  $H(X_k|X_{k-1}) = H(X_1)$  quando  $k = 1$ . Além disso, para todo  $k \geq 2$ , temos

$$H(X_k|X_1^{k-1}) = H(X_1^k) - H(X_1^{k-1}).$$

Como  $X$  é uma cadeia estacionária o limite

$$h(X) = \lim_{k \rightarrow \infty} H(X_k|X_1^{k-1}) = \lim_{k \rightarrow \infty} \frac{H(X_1^k)}{k}.$$

existe e  $h(X)$  é denominada taxa da entropia. A existência dessa taxa, para processos estacionários, é garantida a partir de propriedades da entropia condicional (veja, página 76, teorema 4.2.1, Cover & Thomas 1991). Neste caso,  $h(X)$  é também a entropia do processo  $X$ .

**Exemplo 5.11** *Suponha que estejamos interessados em monitorar a atividade de um neurônio e registramos as ocorrências de  $k$  disparos como sendo  $X_1, X_2, \dots, X_k$ , uma sequência i.i.d. de variáveis aleatórias, definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\{0,1\}$  e com função de distribuição de probabilidade  $\mu(x) = P(X_0 = x)$ ,  $x \in \{0,1\}$ . O cálculo da taxa de entropia para uma sequência com essas características é dada por*

$$h(X) = \lim_{k \rightarrow \infty} \frac{H(X_0, X_2, \dots, X_{k-1})}{k} = \lim_{k \rightarrow \infty} \frac{kH(X_0)}{k} = H(X_0).$$

**Exemplo 5.12** *Suponha que estejamos interessados em monitorar a atividade de um neurônio e registramos as ocorrências de  $k$  disparos a partir de uma cadeia de Markov  $X = \{X_n : n \in \mathbb{Z}\}$  de ordem 1, a qual é definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\{0,1\}$  e com função de distribuição de probabilidade  $\mu$ .*

O cálculo da taxa de entropia neste caso é dada por

$$h(X) = \lim_{k \rightarrow \infty} H(X_k | X_0^{k-1}) = \lim_{k \rightarrow \infty} H(X_k | X_{k-1}) = H(X_1 | X_0).$$

## 5.4 Medidas da informação

Neste trabalho, utilizamos duas medidas da informação para entender a transmissão de informação entre neurônios ao longo de um intervalo de tempo finito: a informação mútua e a entropia de transferência. Desta forma, apresentamos tais medidas nesta seção.

**Definição 5.13** *Sejam  $X$  e  $Y$  duas variáveis aleatórias definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no produto dos conjuntos finitos  $\mathcal{X} \times \mathcal{Y}$  e com função de distribuição de probabilidade conjunta  $\mu(x,y) = P(X = x, Y = y)$ ,  $x \in \mathcal{X}$  e  $y \in \mathcal{Y}$ . Podemos definir a informação mútua entre  $X$  e  $Y$  por*

$$\mathcal{I}(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mu(x,y) \log \frac{\mu(x,y)}{\mu(x)\mu(y)}.$$

Note que a informação mútua representa a redução de incerteza sobre  $X$  devido ao conhecimento de  $Y$ . Além disso, podemos escrever a informação mútua como função da entropia e da entropia condicional da seguinte forma

$$\mathcal{I}(X; Y) = H(Y) - H(Y|X),$$

$$\mathcal{I}(X; Y) = H(X) - H(X|Y),$$

$$\mathcal{I}(X; Y) = H(X) + H(Y) - H(X,Y)$$

A relação entre a informação mútua com a entropia, entropia condicional e entropia conjunta é expressada pelo diagrama de Venn na Figura 5.2. Neste, percebemos que a informação mútua corresponde a interseção entre as informações médias de  $X$  e  $Y$ .

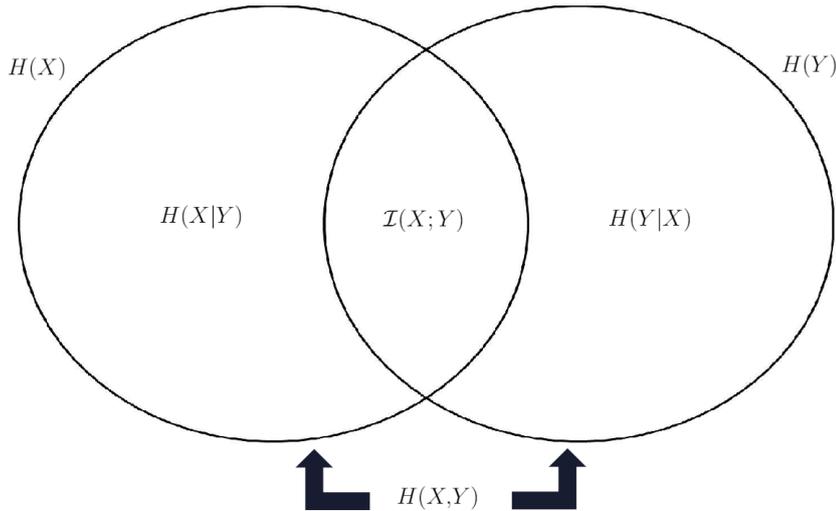


Figura 5.2: Relação entre entropia e informação mútua.

**Exemplo 5.14** *Suponha que estejamos interessados em monitorar a atividade de dois neurônios em um dado instante de tempo e modelamos sua atividade através do par de variáveis aleatórias  $(X,Y)$  definida em um espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores em um conjunto finito  $\{0,1\} \times \{0,1\}$  e com função de distribuição conjunta  $\pi(x,y) = P(X = x, Y = y)$ ,  $(x, y) \in \{0,1\} \times \{0,1\}$  dada por*

$y/x$	0	1	$P(Y=y)$
0	1/8	1/4	3/8
1	1/8	2/4	5/8
$P(X=x)$	1/4	3/4	1

*Temos o interesse de encontrar  $\mathcal{I}(X;Y)$ , logo, a entropia condicional de  $(Y|X)$  é*

$$H(Y|X) = \frac{1}{4}H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{3}{4}H\left(\frac{1}{3}, \frac{2}{3}\right) = \frac{1}{4} \times 1 + \frac{3}{4} \times 0.918 = 0.939 \text{ bits.}$$

*Além disso, a entropia de  $Y$  é*

$$H(Y) = -\frac{3}{8} \log \frac{3}{8} - \frac{5}{8} \log \frac{5}{8} = 0.955 \text{ bits.}$$

*Portanto,*

$$\mathcal{I}(X;Y) = 0.955 - 0.939 = 0.016 \text{ bits,}$$

*que é a informação compartilhada entre  $X$  e  $Y$ .*

Analogamente, podemos definir a informação mútua para blocos de tamanho  $k$ .

**Definição 5.15** *Sejam  $X := \{X_n : n \in \mathbb{Z}\}$  e  $Y := \{Y_n : n \in \mathbb{Z}\}$  duas cadeias estacionárias definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no produto dos conjuntos finitos  $\mathcal{X} \times \mathcal{Y}$  e com função de distribuição de probabilidade conjunta  $\pi$ . A informação mútua associada a blocos de tamanho  $k$  é definida como*

$$\mathcal{I}(X_0^{k-1}; Y_0^{k-1}) = \sum_{a_0^{k-1} \in \mathcal{X}^k} \sum_{b_0^{k-1} \in \mathcal{Y}^k} \pi([a_0^{k-1}, b_0^{k-1}]) \log \frac{\pi([a_0^{k-1}, b_0^{k-1}])}{\pi([a_0^{k-1}])\pi([b_0^{k-1}])}.$$

Note que a informação mútua representa a redução de incerteza sobre  $X_1^k$  devido ao conhecimento de  $Y_1^k$ . Além disso, associado aos termos de entropia e entropia condicional, podemos escrever a informação mútua como

$$\begin{aligned} \mathcal{I}(X_0^{k-1}; Y_0^{k-1}) &= H(Y_0^{k-1}) - H(Y_0^{k-1} | X_0^{k-1}), \\ \mathcal{I}(X_0^{k-1}; Y_0^{k-1}) &= H(X_0^{k-1}) - H(X_0^{k-1} | Y_0^{k-1}), \\ \mathcal{I}(X_0^{k-1}; Y_0^{k-1}) &= H(X_0^{k-1}) + H(Y_0^{k-1}) - H(X_0^{k-1}, Y_0^{k-1}). \end{aligned}$$

**Definição 5.16** *Sejam  $X := \{X_n : n \in \mathbb{Z}\}$  e  $Y := \{Y_n : n \in \mathbb{Z}\}$  duas cadeias estacionárias definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no produto dos conjuntos finitos  $\mathcal{X} \times \mathcal{Y}$  e com função de distribuição de probabilidade conjunta  $\pi$ . Podemos definir a entropia de transferência de  $Y$  para  $X$  associada a blocos de tamanho  $k$  por*

$$T_{Y_0^k \rightarrow X_0^k} = H(Y_k | Y_0^{k-1}) - H(Y_k | Y_0^{k-1}, X_0^{k-1}).$$

A entropia de transferência é a quantificação da passagem de informação de uma variável para outra. Importante notar que a entropia de transferência é sempre um número não negativo, haja vista que, esta é semelhante a uma informação mútua condicional, a qual é positiva por definição. Por fim, trazemos as definições de taxa de entropia de transferência e taxa de informação mútua (para maiores detalhes, veja Haruna & Nakajima 2013), cuja existência é garantida para cadeias estacionárias.

**Definição 5.17** *Sejam  $X := \{X_n : n \in \mathbb{Z}\}$  e  $Y := \{Y_n : n \in \mathbb{Z}\}$  duas cadeias estacionárias definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no produto dos*

conjuntos finitos  $\mathcal{X} \times \mathcal{Y}$ . A taxa de informação mútua entre  $X$  e  $Y$  é definida por

$$i(X,Y) = h(X) + h(Y) - h(X,Y),$$

em que  $h(X)$  é a taxa de entropia de  $X$ ,  $h(Y)$  é a taxa de entropia de  $Y$  e  $h(X,Y)$  é a taxa de entropia conjunta.

**Definição 5.18** *Sejam  $X := \{X_n : n \in \mathbb{Z}\}$  e  $Y := \{Y_n : n \in \mathbb{Z}\}$  duas cadeias estacionárias definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no produto dos conjuntos finitos  $\mathcal{X} \times \mathcal{Y}$ . A taxa de entropia de transferência entre  $X$  e  $Y$  é definida por*

$$t(X|Y) = h(X) - h(X|Y),$$

em que  $h(X)$  é a taxa de entropia de  $X$  e  $h(X|Y)$  é a taxa de entropia condicional.

A taxa da informação mútua e a taxa da entropia de transferência associada a dois processos estacionários  $X$  e  $Y$  são, respectivamente, a informação mútua e entropia de transferência entre os processos  $X$  e  $Y$ .

## 5.5 Estimação das medidas da informação

Sejam  $k$  e  $n$  números inteiros positivos tais que  $k \leq n$  e  $(x_0^{n-1}, y_0^{n-1}) \in \mathcal{X}^n \times \mathcal{Y}^n$  uma amostra de um processo bivariado estacionários a tempo discreto  $(X,Y) := \{(X_n, Y_n) : n \in \mathbb{Z}\}$  com distribuição  $\pi$ . Podemos definir a distribuição empírica de  $\pi$  da seguinte forma

$$\hat{\pi}_{k,k}(a_0^{k-1}b_0^{k-1}|x_0^{n-1}y_0^{n-1}) = \frac{1}{n-k+1} \sum_{i=0}^{n-k} \mathbb{I}(x_i^{i+k-1} = a_0^{k-1}, y_i^{i+k-1} = b_0^{k-1}),$$

$\forall (a_0^{k-1}, b_0^{k-1}) \in \mathcal{X}^k \times \mathcal{Y}^k$ , em que  $k \geq 1$  e  $\mathbb{I}$  é a função indicadora.

Em outras palavras,  $\hat{\pi}_{k,k}$  é obtida deslizando uma janela de tamanho  $k$  ao longo das sequências  $x_0^{n-1}$  e  $y_0^{n-1}$ , com o intuito de contar o número de vezes que  $(a_0^{k-1}, b_0^{k-1})$  aparece nesta amostra e então dividir esse valor pelo número máximo de vezes em que  $(a_0^{k-1}, b_0^{k-1})$  pode ser observada em  $(x_0^{n-1}, y_0^{n-1})$ .

Note que  $\hat{\pi}_{k,k}$  é uma probabilidade em  $\mathcal{X}^k \times \mathcal{Y}^k$ , então suas marginais  $\hat{\pi}_k(\cdot|x_0^{n-1})$  e  $\hat{\pi}_k(\cdot|y_0^{n-1})$  são definidas por

$$\begin{aligned}\hat{\pi}_k(a_0^{k-1}|x_0^{n-1}) &:= \frac{1}{n-k+1} \sum_{b_0^{k-1} \in \mathcal{Y}^k} \hat{\pi}_{k,k}(a_0^{k-1}b_0^{k-1}|x_0^{n-1}y_0^{n-1}), \quad \forall a_0^{k-1} \in \mathcal{X}^k \\ \hat{\pi}_k(b_0^{k-1}|y_0^{n-1}) &:= \frac{1}{n-k+1} \sum_{a_0^{k-1} \in \mathcal{X}^k} \hat{\pi}_{k,k}(a_0^{k-1}b_0^{k-1}|x_0^{n-1}y_0^{n-1}) \quad \forall b_0^{k-1} \in \mathcal{Y}^k.\end{aligned}$$

No caso em que as amostras  $x_0^{n-1}$  e  $y_0^{n-1}$  são geradas a partir das cadeias estocásticas com memória de alcance variável  $X$  e  $Y$  compatíveis, respectivamente, com as árvores de contexto  $\tau_X$  e  $\tau_Y$ , a distribuição empírica  $\hat{\pi}_{k,k}$  é tal que

$$\hat{\pi}_{k,k}(a_0^{k-1}b_0^{k-1}|x_0^{n-1}y_0^{n-1}) = \hat{\pi}_{k,k}(c_{\tau_X}(a_0^{k-1})c_{\tau_Y}(b_0^{k-1})|x_0^{n-1}y_0^{n-1}), \quad \forall (a_0^{k-1}, b_0^{k-1}) \in \mathcal{X}^k \times \mathcal{Y}^k,$$

em que  $k = \max\{h(\tau_X), h(\tau_Y)\}$ .

Agora, podemos definir os estimadores da informação mútua e da entropia de transferência associadas a blocos de tamanho  $k$ .

**Definição 5.19** *Sejam  $x_0^{n-1} \in \mathcal{X}^n$  e  $y_0^{n-1} \in \mathcal{Y}^n$  amostras dos processos estacionários a tempo discreto  $X := \{X_n : n \in \mathbb{Z}\}$  e  $Y := \{Y_n : n \in \mathbb{Z}\}$  definidas em um dado espaço de probabilidade  $(\Omega, \mathcal{F}, P)$  com valores no produto dos conjuntos finitos  $\mathcal{X} \times \mathcal{Y}$ . A entropia empírica associada a blocos de tamanho  $k$  é definida como*

$$\hat{H}_{x_0^{n-1}}(X_0^{k-1}) = - \sum_{a_0^{k-1} \in \mathcal{X}^k} \hat{\pi}_k(a_0^{k-1}|x_0^{n-1}) \log \hat{\pi}_k(a_0^{k-1}|x_0^{n-1}).$$

A entropia condicional empírica associada a blocos de tamanho  $k$  é definida como

$$\hat{H}_{x_0^{n-1}}(X_k|X_0^{k-1}) = - \sum_{a_0^k \in \mathcal{X}^{k+1}} \hat{\pi}_{k+1}(a_0^k|x_0^{n-1}) \log \frac{\hat{\pi}_{k+1}(a_0^k|x_0^{n-1})}{\hat{\pi}_k(a_0^{k-1}|x_0^{n-1})}.$$

A informação mútua associada a blocos de tamanho  $k$  é definida como

$$\begin{aligned}\hat{I}_{(x_0^{n-1}, y_0^{n-1})}(X_0^{k-1}, Y_0^{k-1}) &= \sum_{\substack{a_0^{k-1} \in \mathcal{X}^k \\ b_0^{k-1} \in \mathcal{Y}^k}} \left[ \hat{\pi}_{k,k}(a_0^{k-1}, b_0^{k-1}|x_0^{n-1}, y_0^{n-1}) \right. \\ &\quad \left. \times \log \frac{\hat{\pi}_{k,k}(a_0^{k-1}, b_0^{k-1}|x_0^{n-1}, y_0^{n-1})}{\hat{\pi}_k(a_0^{k-1}|x_0^{n-1}) \hat{\pi}_k(b_0^{k-1}|y_0^{n-1})} \right].\end{aligned}$$

A entropia de transferência associada a blocos de tamanho  $k$  é definida como

$$\hat{T}_{Y_0^k \rightarrow X_0^k} = \hat{H}_{y_0^n}(Y_k | Y_0^{k-1}) - \hat{H}_{(x_0^n, y_0^n)}(Y_k | Y_0^{k-1}, X_0^{k-1}),$$

em que

$$\hat{H}_{(x_0^n, y_0^n)}(Y_k | Y_0^{k-1}, X_0^{k-1}) := - \sum_{\substack{a_0^{k-1} \in \mathcal{X}^k \\ b_0^{k-1} \in \mathcal{Y}^k}} \hat{\pi}_{k, k+1}(a_0^{k-1}, b_0^k | x_0^n, y_0^n) \log \frac{\hat{\pi}_{k, k+1}(a_0^{k-1}, b_0^k | x_0^n, y_0^n)}{\hat{\pi}_{k, k}(a_0^{k-1}, b_0^{k-1} | x_0^n, y_0^n)}.$$

Desde que a cadeia estocástica  $X$  seja estacionária, a entropia empírica associada a passados de tamanho  $k$  converge para a taxa de entropia a medida que aumentamos o tamanho do passado observado (ver, por exemplo, Shields (1996)), ou seja

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\hat{H}_{x_0^{n-1}}(X_0^{k-1})}{k} = \lim_{k \rightarrow \infty} \frac{\hat{H}_{x_0^{n-1}}(X_0^{k-1})}{k} = h(X). \quad (5.20)$$

Ao escrevermos  $k$  como uma função de  $n$  podemos tomar um limite único em (5.20), para todo  $n \geq 0$  e  $k(n) \leq \frac{1}{h(X)}$ . Assim, quando  $k(n) \rightarrow \infty$  quando  $n \rightarrow \infty$ , temos, para  $\mu$ -quase certamente (veja, por exemplo, Ornstein & Weiss (1990))

$$\lim_{n \rightarrow \infty} \frac{\hat{H}_{x_0^{n-1}}(X_0^{k(n)})}{k(n)} = h(X).$$

Além disso, a entropia  $H(X_0^{k-1})$  é côncava em  $\mu$  e seu valor máximo é atingido em  $\log |\mathcal{X}|$ , consequentemente,  $h(X) \leq \log |\mathcal{X}|$ . Portanto, podemos tomar  $k(n)$  como

$$k(n) \leq \frac{\log n}{\log |\mathcal{X}|}.$$

Por fim, como  $X$  é estacionário, ver em Shields (1996), se  $k(n) \leq \frac{1-\epsilon}{\log |\mathcal{X}|} \log n$ , para qualquer  $\epsilon \in (0,1)$ , então

$$\lim_{n \rightarrow \infty} \hat{H}_{x_0^{n-1}}(X_{k(n)} | X_0^{k(n)-1}) = h(X),$$

para  $\mu$ -quase certamente.

Analogamente, quando  $(X, Y)$  é um cadeia estacionária com distribuição  $\mu$ , temos para

$\mu$ -quase certamente

$$\lim_{n \rightarrow \infty} \frac{\hat{I}_{(x_0^{n-1}, y_0^{n-1})}(X_0^{k(n)}, Y_0^{k(n)})}{k(n)} = i(X, Y) \quad \text{e} \quad \lim_{n \rightarrow \infty} \frac{\hat{T}_{Y_0^{k(n)} \rightarrow X_0^{k(n)}}}{k(n)} = t(Y|X).$$

Observe que a estimação das medidas da informação dependem, essencialmente, das probabilidade de transição estimadas. Assim, em um contexto de cadeias de alcance variável, utilizaremos o algoritmo contexto para obter tais estimadores. Na literatura, alguns trabalhos têm estudado o comportamento assintótico da entropia empírica quando a amostra é gerada por cadeias de memória com alcance variável (ver, por exemplo, Cai et al. 2017) ou cadeias de memória de ordem infinita (ver, por exemplo, Chazottes & Gabrielli 2005; Chazottes & Maldonado 2011; Gabrielli et al. 2003). Entretanto, pouco têm sido feito com relação a outras medidas da informação. Até onde nosso conhecimento vai, apenas o artigo de Ferreira (2021) trata de limites de concentração para estimadores empíricos da informação dirigida. Neste sentido, este trabalho também pretende, em um certo sentido, ser uma primeira contribuição a respeito do quão bom são os estimadores empíricos da informação mútua e da entropia de transferência em um ambiente em que a fonte é de memória de alcance variável.



# Capítulo 6

## Aplicação computacional

O problema de inferência da conectividade neuronal é desafiante tanto do ponto de vista neurobiológico quanto estatístico (de Abril et al. 2018). No início deste trabalho, apresentamos uma discussão concisa sobre os diversos mecanismos neurobiológicos que dificultam o estudo da conectividade neural por meio de inferência estatística. Há grande variabilidade presente nos disparos neuronais, consequência da existência de ruídos no sinal elétrico, o que nos leva a considerar um ambiente estocástico para lidar com o problema de conectividade, sendo que, dentro desse ambiente, queremos poder inferir as conexões sinápticas. Neste sentido, propomos a utilização de modelos de cadeias estocásticas de memória com alcance variável para modelar os trens de disparos neuronais e apresentamos uma metodologia de estimação de medidas da informação como solução para o problema de inferência da conectividade neuronal.

O problema de inferência de conectividade é antigo dentro da neurobiologia (ver, por exemplo, y Cajal & Azoulay 1894, Brodmann 1910 e Bota et al. 2005 ). Diversas técnicas como correlações cruzadas, informação mútua e entropia de transferência são aplicadas neste contexto. Mesmo assim, estimar a conectividade cerebral é um problema desafiante do ponto de vista estatístico, tendo em vista que detectar e quantificar conexões em meio a presença de ruídos e flutuações aleatórias não é trivial. Adicionalmente, a maioria dos modelos teóricos ou computacionais que poderiam servir de base para aprimorar as técnicas estatísticas existentes são baseados em registros *in vitro*, ou seja, em experimentos realizados em ambientes controlados e com reduzido sinal externo e ruído (ver, por exemplo, Herrmann & Gerstner 2001). Dessa forma, muitos desses modelos acabam não levando em consideração disparos espontâneos gerados em ambientes com alto grau ruidoso que poderiam carregar informações importantes do ponto de vista neurobiológico.

Além disso, modelos determinísticos como integra-e-dispara (circuito RC com regras de disparo) falham em ter disparos ocorrendo com certa estrutura temporal neurobiológica se não forem acoplados a processos estocásticos. A utilização de cadeias estocásticas de memória com alcance variável como modelo de célula neural traz novas perspectivas para este ponto. Esperamos que este tipo de abordagem consiga reproduzir registros *in vivo* que possuem um alto grau de complexidade (veja, por exemplo, English et al. (2017)). A importância da memória carregada por um neurônio, informação que se mantém ao longo do tempo, na estimação da conectividade neural é avaliada nas próximas seções deste trabalho.

## 6.1 Hipóteses e Cenários

Utilizando as metodologias apresentadas nesta monografia, efetuamos aplicações em cenários computacionais e experimentais. Neste capítulo, abordamos os cenários computacionais e, no próximo, os experimentais. Devido à existência de diversas hipóteses relacionadas a neurônios guardarem ou não informação, sobre a forma que a utilizam, ou mesmo se os estímulos auxiliam ou reduzem a habilidade de um neurônio de trocar informação, ou ainda se são as sinapses que armazenam a informação, iremos propor três cenários que serão utilizados como base para a investigação.

Neste sentido, realizamos a modelagem da sequência de disparos neuronais via cadeias estocásticas de memória com alcance variável e fazemos a inferência da transmissão de informação neuronal ao longo de um intervalo de tempo finito considerando três cenários distintos:

1. Um neurônio, de uma área específica do cérebro, terá sua resposta temporal, disparos e não disparos, observada ao longo de um período de tempo finito  $e$ , então, iremos estimar, para diferentes tamanhos de amostras, a incerteza média esperada com relação à atividade do neurônio quando condicionada ao conhecimento do histórico passado de suas respostas temporais. Neste caso, vamos gerar as amostras de disparos neuronais a partir de uma família de probabilidades de transição, em que a probabilidade do neurônio disparar em um dado instante de tempo condicionada ao seu passado é modelada a partir de uma função de ativação logística  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  dada por

$$\phi(v) = \frac{1}{1 + \exp(-v)}.$$

2. Novamente, vamos considerar o problema de estimar a incerteza média esperada com relação à atividade de um neurônio quando condicionada ao conhecimento do histórico passado de suas respostas. Entretanto, neste caso, as amostras são geradas a partir de uma família de probabilidades de transição, em que a probabilidade do neurônio disparar em um dado instante de tempo condicionada ao seu passado é modelada a partir de uma função de ativação  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  tal que

$$\phi(v) = 1 - \exp(-\exp(v)),$$

cuja inversa  $\phi^{-1}$  é popularmente conhecida na literatura estatística como a função de ligação complemento log-log de uma regressão logística.

3. Dois neurônios N1 e N2, de uma área específica do cérebro, tem sua resposta temporal, disparos e não disparos, observadas ao longo de um período de tempo finito e, então, iremos estimar, para diferentes tamanhos de amostra, a informação que flui de um neurônio ao outro. Em outras palavras, vamos verificar o quanto diminui a incerteza média esperada com relação à atividade do neurônio  $N_i$  quando além do conhecimento do histórico passado das suas respostas temporais, sabemos também o histórico passado da atividade do neurônio  $N_j$ ,  $i, j = 1, 2$  com  $i \neq j$ . Neste caso, vamos gerar as amostras de disparos neuronais a partir de uma cadeia estocástica de alcance variável bivariada, cujas probabilidades de transição são fixadas levando em consideração a Teoria Hebbiana (Hebb, et al. 1949). Esta teoria parte do pressuposto que “células que disparam juntas, permanecem conectadas”. Em outras palavras, quando o neurônio  $N_i$  disparar, o neurônio  $N_j$ , que está conectado a  $N_i$ , possuirá uma probabilidade maior de disparar em sequência,  $i, j = 1, 2$  com  $i \neq j$ . Além disso, quando ambos ficam um tempo suficientemente longo sem disparar, é mais provável que venham a disparar juntos.

O objetivo de cada um desses cenários dentro do escopo deste estudo de simulação é diferente. Nos cenários 1 e 2, estamos interessados em estimar a entropia condicional associada a blocos com um tamanho finito fixado e utilizar réplicas de simulação Monte Carlo (Shreider 1966) para avaliar o desempenho dos métodos de estimação propostos nesta monografia. Em ambos os cenários, as probabilidades de transição que geram as amostras são modeladas a partir de funções de ativação distintas. Assim, pretendemos também verificar se a escolha dessas funções impactam na performance das metodologias de infe-

rência que estão sendo propostas. Por outro lado, no cenário 3, estamos interessados em estimar a entropia de transferência associada a blocos com um tamanho finito fixado a fim de estudar a existência de conectividade entre dois neurônios. Neste último caso, usamos novamente réplicas de simulação Monte Carlo para avaliar os métodos de estimação propostos. Até onde vai nosso conhecimento, não há na literatura trabalho algum que trate da estimação da entropia de transferência utilizando o Algoritmo Contexto neste cenário mais geral de cadeias de alcance variável. Desta forma, este trabalho complementa, em um certo sentido, os estudos sobre estimação da entropia de transferência.

## 6.2 Simulação

As simulações foram feitas em R, versão 8.16, e disponibilizadas publicamente em <https://github.com/Victor-girardi/Comandos-da-Teoria-da-Informa-o> juntamente com a documentação para uso futuro da comunidade científica. Neste contexto, as amostras de disparos neuronais consideradas neste trabalho foram obtidas a partir da simulação de cadeias estacionárias de memória com alcance variável compatíveis com uma árvore de contexto probabilística  $(\tau, Q)$  pré-especificada. Dada uma amostra, a inferência da conectividade neuronal foi realizada a partir da estimação das medidas de informação via método da máxima verossimilhança e do algoritmo Contexto. Contudo, considerar apenas uma amostra não nos fornece informação suficiente a respeito do desempenho do método de estimação proposto nesta monografia. Assim, utilizamos réplicas de simulação Monte Carlo para comparar nossas estimativas com os verdadeiros valores dos parâmetros, o que nos fornece um indicativo a respeito da tendência e da variabilidade dos estimadores considerados. É importante salientar que as simulações mencionadas foram realizadas em cada um dos três cenários descritos na Seção 6.1. A seguir, descrevemos em mais detalhes como cada etapa deste estudo de simulação foi realizada.

### 6.2.1 Amostras dos disparos neuronais

O estudo da conectividade neuronal via simulação é uma viabilização prática para nos ajudar no entendimento de medidas da informação, sem precisar de coletas amostrais de um organismo biológico. Desta forma, o primeiro passo é gerar uma amostra dos disparos neuronais. Nos cenários 1 e 2, estudamos um único neurônio e, portanto, a amostra da sua resposta temporal será gerada a partir de uma cadeia de alcance variável compatível com

uma dada árvore probabilística. Por outro lado, no cenário 3, estamos interessados no estudo da conectividade de dois neurônios. Assim, é necessário considerar uma amostra das respostas temporais de ambos neurônios. Para tal, geramos a amostra a partir de uma cadeia estocástica de alcance variável bivariada. É importante frisar que tanto no caso univariado quanto no caso bivariado, as cadeias em consideração são homogêneas no tempo e estacionárias.

Em todos os cenários, vamos avaliar se conforme aumentamos o tamanho da amostra, as estimativas das medidas da informação tornam-se mais próximas do seu verdadeiro valor. Em outras palavras, estamos interessados em estudar, computacionalmente, a consistência (em média) dos estimadores propostos nesta monografia. Neste sentido, vamos considerar 4 tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Conforme já mencionamos, considerar uma única amostra de cada tamanho não nos fornece informação suficiente para realizar este estudo sobre o desempenho dos estimadores. Assim, para cada tamanho amostral, geramos réplicas de  $M = 100$  amostras.

Outra consideração comum a todos os cenários é a respeito da altura da árvore de contexto a ser utilizada para gerar as amostras. O tempo médio que um neurônio leva para disparar após o seu último disparo depende do neurônio que estamos estudando. Sem perda de generalidade, vamos assumir que estamos trabalhando com neurônios corticais, os quais, em média, levam de 1 a 3 milissegundos (ms) para disparar (ver Softky & Koch (1993)). Na prática, em geral, a discretização temporal utilizada no processo de amostragem das respostas temporais de um neurônio (disparar ou ficar em repouso) leva em consideração intervalos da ordem de 0,30ms. Esta é a ordem que utilizamos para a discretização temporal deste projeto. Desta forma, consideramos a altura da árvore geradora  $\tau$  como sendo  $h(\tau) = 4$ . Note que é razoável considerarmos  $h(\tau) = 4$  pois, em média, precisamos observar a atividade do neurônio em 4 intervalos de tempo de tamanho 0,30ms. Portanto, neste trabalho, a amostra da atividade de um dado neurônio a ser considerada é uma sequência binária  $x_0^{n-1} \in \{0,1\}$  tal que  $x_i = 1$  se, e somente se, observa-se um disparo no  $i$ -ésimo intervalo de tamanho 0,30ms da discretização, em  $i \in \{0,2, \dots, n-1\}$  e  $n$  é um número inteiro positivo.

As amostras de cadeias estocásticas foram simuladas a partir de árvores de contexto probabilísticas conhecidas. A seguir, justificamos a escolha de tais parâmetros de inicialização.

### Cenário 1

Neste cenário, consideramos um único neurônio. Portanto, neste caso, vamos gerar as amostras a partir de uma cadeia estocástica de alcance variável  $X := \{X_n : n \in \mathbb{Z}\}$  compatível com a árvore probabilística  $(\tau_1, Q_1)$  e com valores no alfabeto  $\mathcal{S} = \{0,1\}$ , em que  $\tau_1 = \{1, 10, 100, 1000, 0000\}$  e  $Q_1 = \{Q_1(b|u) : b \in \mathcal{S}, u \in \tau_1\}$  é uma família de probabilidades de transição.

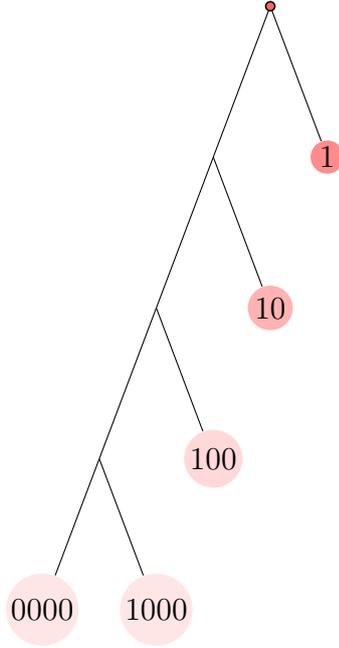


Figura 6.1: Representação pictorial da árvore  $\tau_1 = \{1,10,100,1000,0000\}$ .

Do ponto de vista neurobiológico, um neurônio dispara quando seu potencial de membrana ultrapassa um certo limiar. Em seguida, ele retorna ao seu potencial de repouso e se mantém neste estado até que outro desequilíbrio eletrolítico ocorra. Desta forma, após seu último disparo, o neurônio acumula potencial de membrana até que este seja suficientemente grande para permitir um disparo. Assim, é razoável considerar que a probabilidade de um neurônio disparar condicionada ao conhecimento da sua resposta temporal passada é tão maior quanto mais longe estiver o último disparo deste neurônio no passado em questão. Em outras palavras, fixado um passado  $a_{-\infty}^{-1} \in \{0,1\}^{-\mathbb{N}}$ ,

$$Q_1 \left( 1 \mid a_{-\ell(a_{-\infty}^{-1})}^{-1} \right) \rightarrow 1, \quad \text{quando } \ell(a_{-\infty}^{-1}) \rightarrow \infty,$$

em que  $\ell(a_{-\infty}^{-1}) := \inf\{k \geq 1 : a_{-k} = 1\}$ .

Neste sentido, podemos dizer que fixado um passado  $a_{-\infty}^{-1} \in \{0,1\}^{-\mathbb{N}}$  da atividade de um neurônio, a probabilidade  $Q(1|a_{-\infty}^{-1})$  do neurônio disparar dado o passado  $a_{-\infty}^{-1} \in \{0,1\}^{-\mathbb{N}}$  é uma função do potencial de membrana do neurônio o qual é, a grosso modo, a resultante da sua sequência de disparos passada  $a_{-\infty}^{-1}$ . Esta função é conhecida em neurociência computacional como função de ativação. Neste cenário e no próximo, consideramos funções do tipo sigmóide como sendo as funções de ativação. As funções sigmóides tem formato de “S” (ver Figuras 6.2 e 6.4) e, portanto, são funções não-lineares estritamente crescentes, o que indica que a probabilidade do neurônio disparar aumenta conforme o seu potencial de membrana cresce. No presente cenário, vamos considerar a função logística, que é um exemplo bem comum de função sigmóide e vem sendo extensivamente utilizada em neurociência computacional para fins de modelagem (ver, por exemplo, R. Jolivet & Gerstner (2006)). Outra razão que nos levou a considerar esta função como de ativação é que alguns modelos de regressão utilizados para estudar a conectividade neuronal consideram a inversa desta função como função de ligação (Song 2013).

**Definição 6.1** *Dizemos que a função  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  é a logística padrão quando sua lei é dada por*

$$\phi(x) = \frac{1}{1 + \exp(-x)}.$$

Observe que, o conjunto imagem da função  $\phi$  é o intervalo aberto  $]0,1[$ , portanto é plausível considerá-la para descrever o comportamento de probabilidades. No nosso caso, a variável independente é o potencial de membrana. Portanto, quando o potencial de membrana é 0 segue que a probabilidade do neurônio disparar dado seu passado é  $1/2$  (ver Figura 6.2). Assim, sem perda de generalidade, vamos considerar  $-5\text{mV}$  como sendo o potencial de repouso, embora saibamos que, na prática, esse potencial é de  $-67\text{mV}$ . Além disso, consideramos  $5\text{mV}$  como sendo o limiar de disparo, embora, na prática, esse potencial é próximo de  $0\text{mV}$ . Note que conforme o potencial de membrana aumenta, maior é a probabilidade do neurônio disparar.

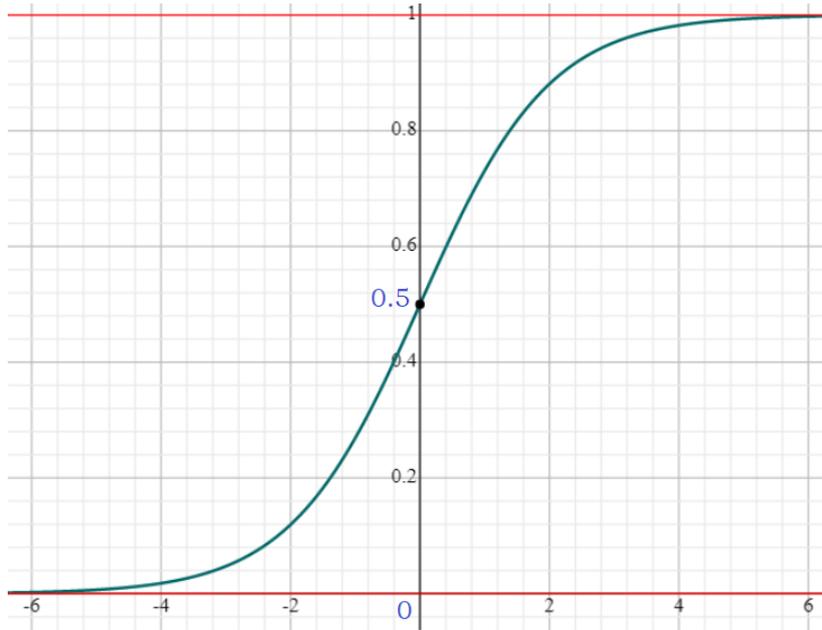


Figura 6.2: Representação no plano cartesiano do gráfico da função logística padrão.

Neste contexto, geramos valores pseudo-aleatórios a partir da função de ativação logística padrão e atribuímos esses valores às probabilidades de transição da família  $Q_1$  levando em consideração a distância do último disparo observado no passado. Desta forma, para cada passado  $u \in \tau_1$ , os valores gerados atribuídos à probabilidade de transição  $Q_1(1|u)$  é maior quanto maior for  $\ell(u)$ . No caso em que  $u = 0000$ , fixamos  $Q_1(1|u) = 1$ . Em outras palavras, o neurônio dispara com probabilidade 1 se o seu último disparo ocorrer em um instante suficientemente longe do tempo de disparo médio. Assim, a matriz de transição considerada neste cenário, a partir dos potenciais de membrana  $(-2,9; -1,7; 0; 0,8; 5)$ , é

$$Q_1 := \begin{matrix} & & 0 & 1 \\ \begin{matrix} 1 \\ 01 \\ 001 \\ 0001 \\ 0000 \end{matrix} & \left( \begin{array}{cc} 0,9478 & 0,0522 \\ 0,8455 & 0,1545 \\ 0,5000 & 0,5000 \\ 0,3100 & 0,6899 \\ 0 & 1 \end{array} \right) \end{matrix}.$$

É importante observar que para gerarmos a amostra de disparos neuronais a partir da matriz de transição  $Q_1$ , precisamos de valores iniciais para cadeia de alcance variável  $X$ . Geramos, portanto,  $k = h(\tau_1) = 4$  valores iniciais a partir de valores pseudo-aleatórios de uma distribuição Bernoulli(1/2). Assim, os valores iniciais a serem considerados para geração da amostra neuronal a partir da árvore de contexto probabilística  $(\tau_1, Q_1)$  é  $x_{-4}^{-1} =$

1010. Vale ressaltar, que a forma como geramos os valores iniciais não importa, pois, vamos excluir 10% das observações do início da amostra a fim de dissipar o efeito desses valores.

A seguir, apresentamos o pseudo-algoritmo utilizado para simular as amostras de disparos neuronais deste cenário.

**Algoritmo 1.** *Pseudo-algoritmo para gerar as amostras de disparos neuronais a partir da árvore de contexto probabilidade  $(\tau_1, Q_1)$ .*

**Entrada:**  $x_0^{k-1}$  (passado) e  $(\tau_1, Q_1)$  (árvore de contexto probabilística)

**Saída:**  $x_0^{n-1}$  (amostra de disparos neuronais de tamanho  $n$ )

Fixar o tamanho  $n$  da amostra a ser gerada

Fixar uma sequência  $x_0^{k-1}$  tal que

- $k = \max_{u \in \tau_1} |u|$ ;
- $k \ll n$ .

Gerar um vetor de números pseudo-aleatórios  $u = (u_k, \dots, u_n)$  gerados a partir de uma distribuição Uniforme(0,1).

**enquanto**  $i \in \{k, \dots, n\}$  **faça** ▷ *repetição*

**se**  $u_i \leq Q_1(1|c_{\tau_1}(x_0^{i-1}))$  **então** ▷ *condição*

$x_i = 1$

**senão**

$x_i = 0$

**fim se**

**fim enquanto**

Por fim, utilizamos  $x_0^{n-1}$  como nossa amostra de disparos neuronais. ▷ *final*

## Cenário 2

Neste cenário, também consideramos um único neurônio. De forma análoga ao cenário 1, vamos gerar as amostras a partir de uma cadeia estocástica de alcance variável  $X := \{X_n : n \in \mathbb{Z}\}$  compatível com a árvore probabilística  $(\tau_2, Q_2)$  e com valores no alfabeto  $A = \{0,1\}$ , em que  $\tau_2 = \{1, 10, 100, 1000, 0000\}$  e  $Q_2 = \{Q_2(b|u) : b \in \mathcal{S}, u \in \tau_2\}$  é uma família de probabilidades de transição. Neste caso, entretanto, utilizamos como função de ativação a inversa da função de ligação complemento log-log de uma regressão

logística. É justamente o fato de na literatura alguns autores considerarem, para estudar a conectividade neuronal, modelos de regressão em que esta função é usada como função de ligação (ver, Song (2013)), que nos motivou a também considerá-la neste trabalho.

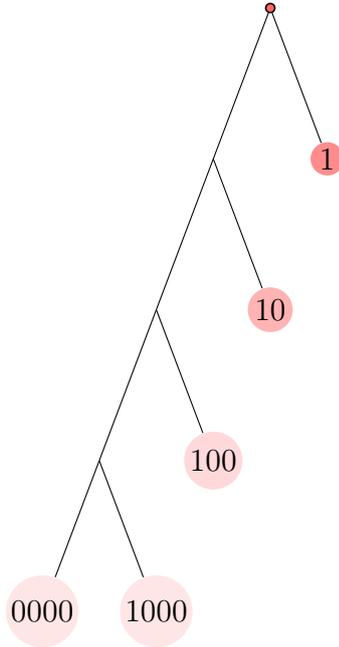


Figura 6.3: Representação pictorial da árvore  $\tau_2 = \{1,10,100,1000,0000\}$ .

**Definição 6.2** Dizemos que a função  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  é a inversa da função de ligação complemento log-log quando sua lei é dada por

$$\phi(x) = 1 - \exp(-\exp(x)).$$

Observe que, o conjunto imagem da função  $\phi$  é o intervalo aberto  $]0,1[$ , portanto é plausível considerá-la para descrever o comportamento de probabilidades. Neste caso, a variável independente é também o potencial de membrana. Portanto, quando o potencial de membrana é 0 segue que a probabilidade do neurônio disparar dado seu passado é 0.64 (ver Figura 6.4). Novamente, vamos considerar  $-5\text{mV}$  como sendo o potencial de repouso e  $-67\text{mV}$  como o limiar de disparo. Note que conforme o potencial de membrana aumenta, maior é a probabilidade do neurônio disparar, isto é, fixado um passado  $a_{-\infty}^{-1} \in \{0,1\}^{-\mathbb{N}}$ ,

$$Q_2 \left( 1 | a_{-\ell(a_{-\infty}^{-1})}^{-1} \right) \rightarrow 1, \quad \text{quando } \ell(a_{-\infty}^{-1}) \rightarrow \infty,$$

em que  $\ell(a_{-\infty}^{-1}) := \inf\{k \geq 1 : a_{-k} = 1\}$ .

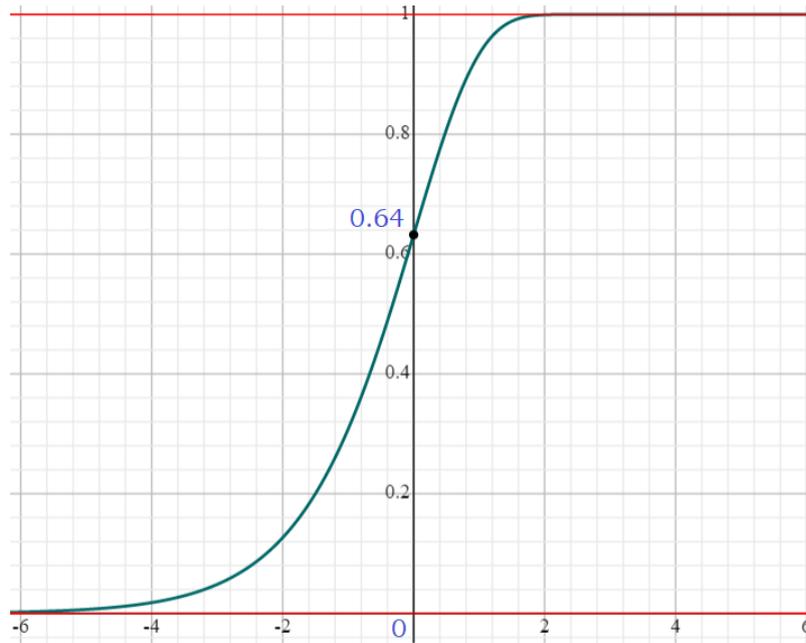


Figura 6.4: Representação no plano cartesiano do gráfico da função inversa da função de ligação complemento log-log.

Neste contexto, geramos valores pseudo-aleatórios a partir da função de ativação como sendo a inversa da função de ligação complemento log-log e atribuímos esses valores às probabilidades de transição da família  $Q_2$ , levando em consideração a distância do último disparo observado no passado. Desta forma, para cada passado  $u \in \tau_2$ , os valores gerados atribuídos à probabilidade de transição  $Q_2(1|u)$  é maior quanto maior for  $\ell(u)$ . No caso em que  $u = 0000$ , fixamos  $Q_2(1|u) = 1$ . Em outras palavras, o neurônio dispara com probabilidade 1 se o seu último disparo ocorrer em um instante suficientemente longe do tempo de disparo médio. Assim, análogo ao Cenário 1, consideramos a matriz de transição, a partir dos potenciais de membrana  $(-2,9; -1,7; 0; 0,8; 5)$ , i.e.,

$$Q_2 := \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{cc} 0 & 1 \\ \left( \begin{array}{cc} 0,9464 & 0,0535 \\ 0,8330 & 0,1669 \\ 0,3678 & 0,6321 \\ 0,1080 & 0,8919 \\ 0 & 1 \end{array} \right) \end{array} .$$

Como já discutimos, para gerarmos a amostra de disparos neuronais a partir da matriz de transição  $Q_2$ , precisamos considerar valores iniciais para a cadeia de alcance variável  $X$ . Geramos, portanto, novamente,  $k = h(\tau_2) = 4$  valores iniciais a partir de valores

pseudo-aleatórios de uma distribuição Bernoulli(1/2). Assim, os valores iniciais a serem considerados para geração da amostra neuronal a partir da árvore de contexto probabilística  $(\tau_2, Q_2)$  é  $x_{-4}^{-1} = 1010$ .

O pseudo-algoritmo utilizado para simular as amostras de disparos neuronais deste cenário é análogo ao Algoritmo 1 utilizado no cenário 1, com a mudança de que a árvore de contexto probabilística de entrada é  $(\tau_2, Q_2)$  ao invés de  $(\tau_1, Q_1)$ .

### Cenário 3

Neste cenário, consideramos dois neurônios. Portanto, neste caso, vamos gerar as amostras a partir de uma cadeia estocástica de alcance variável bivariada  $(X, Y) := \{(X_n, Y_n) : n \in \mathbb{Z}\}$ . A maneira mais comum de fixar uma árvore de contexto probabilística compatível com  $(X, Y)$  é considerar a transformação  $Z_n = X_n + 2Y_n$  para todo  $n \in \mathbb{Z}$  (ver Cai et al. (2017)). A cadeia  $Z := \{Z_n : n \in \mathbb{N}\}$  tem valores no alfabeto  $A = \{0, 1, 2, 3\}$  e assumimos que ela é compatível com a árvore de contexto probabilística  $(\tau_3, Q_3)$ , em que  $\tau_3 = \{1, 2, 3, 10, 20, 30, 100, 200, 300, 1000, 2000, 3000, 0000\}$  e  $Q_3 = \{Q_3(b|u) : b \in A, u \in \tau_3\}$  é uma família de probabilidades de transição.

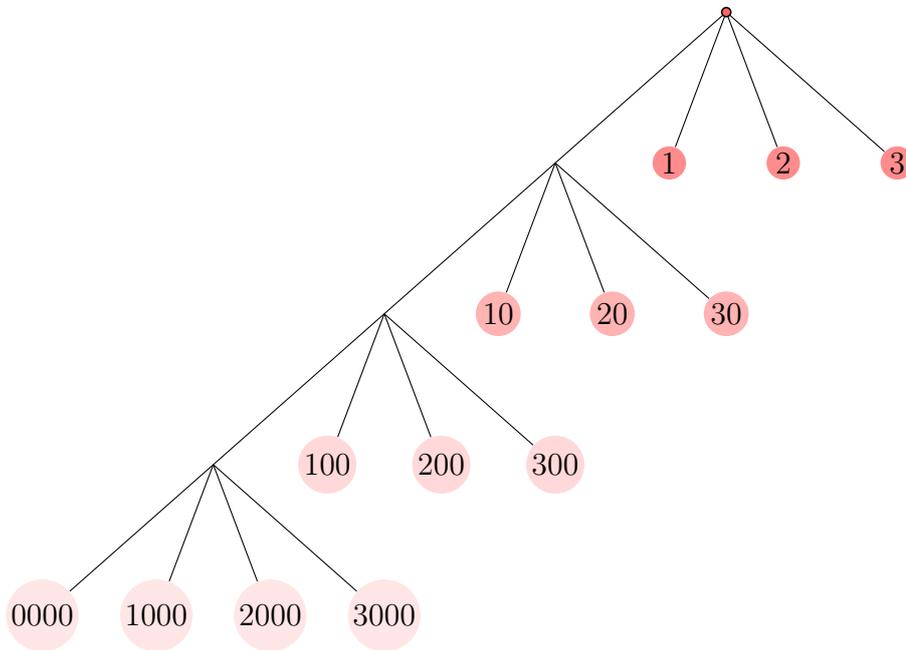


Figura 6.5: Representação pictorial da árvore  $\tau_3$ .

Note que, para qualquer que seja  $n \in \mathbb{Z}$ ,

- se  $Z_n = 0$ , então  $(X_n, Y_n) = (0, 0)$ , isto é, tanto o neurônio 1 quanto o neurônio 2 ficam em repouso no instante  $n$ ;

- se  $Z_n = 1$ , então  $(X_n, Y_n) = (1,0)$ , isto é, o neurônio 1 dispara e o neurônio 2 fica em repouso no instante  $n$ ;
- se  $Z_n = 2$ , então  $(X_n, Y_n) = (0,1)$ , isto é, o neurônio 1 fica em repouso e o neurônio 2 dispara no instante  $n$ ;
- se  $Z_n = 3$ , então  $(X_n, Y_n) = (1,1)$ , isto é, ambos os neurônios disparam no instante  $n$ .

Para fixar a família de probabilidades de transição  $Q_3$ , nos embasamos na teoria Hebbiana (Hebb, et al. 1949), a qual tem como premissa que “células que disparam juntas, permanecem conectadas”. Por exemplo, quando o neurônio 1 dispara, o neurônio 2 possui uma maior probabilidade de disparar, uma vez que o fato de estarem conectados faz com que o potencial de membrana do neurônio 2 aumente após o disparo do neurônio 1. Por outro lado, a partir do momento que ambos os neurônios não disparam por um longo tempo, quando um disparar o incentivo gerado ao outro neurônio é grande o suficiente ao ponto que ambos os neurônios disparam praticamente juntos. Neste contexto, fixamos  $Q_3$  como sendo a seguinte matriz de transição

$$Q_3 := \begin{matrix} & & 0 & 1 & 2 & 3 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 01 \\ 02 \\ 03 \\ 001 \\ 002 \\ 003 \\ 0001 \\ 0002 \\ 0003 \\ 0000 \end{matrix} & \left( \begin{matrix} 0,550 & 0,100 & 0,200 & 0,150 \\ 0,550 & 0,200 & 0,100 & 0,150 \\ 0,650 & 0,100 & 0,100 & 0,150 \\ 0,425 & 0,150 & 0,250 & 0,175 \\ 0,425 & 0,250 & 0,150 & 0,175 \\ 0,450 & 0,150 & 0,150 & 0,250 \\ 0,350 & 0,150 & 0,300 & 0,200 \\ 0,350 & 0,300 & 0,150 & 0,200 \\ 0,250 & 0,200 & 0,200 & 0,350 \\ 0,100 & 0,250 & 0,350 & 0,300 \\ 0,100 & 0,350 & 0,250 & 0,300 \\ 0,100 & 0,250 & 0,250 & 0,400 \\ 0,000 & 0,000 & 0,000 & 1,000 \end{matrix} \right) \end{matrix}.$$

Note que no caso em que  $u = 0000$ , fixamos  $Q_3(3|u) = 1$ . Em outras palavras, ambos os neurônios disparam com probabilidade 1 se o seu último disparo ocorrer em um instante

suficientemente longe do tempo de disparo médio.

É importante observar que para gerarmos a amostra de disparos neuronais a partir da matriz de transição  $Q_3$ , precisamos fixar valores iniciais para cadeia de alcance variável bivariada  $(X,Y)$ . Geramos, portanto,  $k = h(\tau_3) = 4$  valores iniciais para cada uma das cadeias a partir de valores pseudo-aleatórios de uma distribuição Bernoulli(1/2). Assim, os valores iniciais a serem considerados para geração da amostra neuronal a partir da árvore de contexto probabilística  $(\tau_3, Q_3)$  são  $\{(x_i, y_i) : i = 1, 2, 3, 4\} = \{(0,1),(0,1),(1,0),(1,1)\}$ .

A seguir apresentamos o pseudo-algoritmo utilizado para simular as amostras de disparos neuronais deste cenário.

**Algoritmo 2.** *Pseudo-algoritmo para gerar as amostras de disparos neuronais a partir da árvore de contexto probabilidade  $(\tau_3, Q_3)$ .*

**Entrada:**  $x_0^{k-1}$  (passado do neurônio 1),  $y_0^{k-1}$  (Passado do neurônio 2) e  $(\tau_3, Q_3)$  (árvore de contexto probabilística)

**Saída:**  $z_0^{n-1}$  (amostra dos disparos neuronais de tamanho  $n$ )

Fixar o tamanho  $n$  da amostra a ser gerada

Fixar as sequências  $x_0^{k-1}$  e  $y_0^{k-1}$ , relacionadas, respectivamente, aos neurônios N1 e N2, de forma que

- $k = \max_{u \in \tau_3} |u|;$
- $k \ll n.$

Gerar a sequência  $z_0^{k-1}$  tal que  $z_i = x_i + 2y_i$  para todo  $i = 0, 1, \dots, k - 1$ .

Fixar um vetor de números pseudo-aleatórios  $u = (u_1, \dots, u_{n-k})$  gerados a partir de uma distribuição Uniforme(0,1).

**enquanto**  $i \in \{1, \dots, n - k\}$  **faça** ▷ *repetição*

**se**  $u_i \leq Q_3(0|c_{\tau_3}(z_0^{i-1}))$  **então** ▷ *condição*

$z_i = 0$

**fim se**

**se**  $u_i \leq Q_3(0|c_{\tau_3}(z_0^{i-1})) + Q_3(1|c_{\tau_3}(z_0^{i-1}))$  **então** ▷ *condição*

$z_i = 1$

**fim se**

**se**  $u_i \leq Q_3(0|c_{\tau_3}(z_0^{i-1})) + Q_3(1|c_{\tau_3}(z_0^{i-1})) + Q_3(2|c_{\tau_3}(z_0^{i-1}))$  **então** ▷ *condição*

```

     $z_i = 2$ 
senão
     $z_i = 3$ 
fim se
fim enquanto
Por fim, utilizamos  $z_0^{n-1}$  como nossa amostra de disparos neuronais. ▷ final

```

## 6.2.2 Estimação da conectividade neuronal

Com a amostra dos disparos neuronais geradas, podemos realizar a estimação da árvore de contexto probabilística a partir da aplicação do algoritmo contexto. Com isso, conseguimos estimar as medidas da informação a serem utilizadas para estudar a conectividade neuronal. A inferência da conectividade neuronal é importante, pois nos ajuda entender os princípios operacionais do cérebro tais como identificar o mecanismo de codificação e transmissão de informação.

Como mencionamos, vamos utilizar, nos três cenários, o algoritmo contexto para realizar a inferência da árvore de contexto probabilística com base nas amostras geradas e nas estimativas de máxima verossimilhança das matrizes de transição. Como vimos no capítulo 4, o algoritmo contexto calcula, para cada nó da árvore, uma medida de discrepância entre as probabilidades de transição associadas ao referido contexto e àquelas associadas ao contexto acrescido de um passo a mais no passado. Se a medida de discrepância é maior do que um limiar fixado, os contextos são mantidos na árvore; caso contrário, eles são podados. Este procedimento é realizado até se obter a menor árvore de contexto bem adaptada à amostra.

Em todos os cenários, o limiar que fixamos para decidir a respeito do procedimento de poda da árvore foi definido como sendo  $\delta = 0,10$ , o que significa que a poda só ocorre se a distância absoluta entre as probabilidades de transição em consideração não for maior do que  $0,10$ . Assim,  $\delta$  é um valor crítico para um dado nível de significância adotado para testar sequencialmente, para cada nó da árvore, a hipótese nula de que a probabilidade de transição de um contexto para um símbolo  $a \in \{0,1\}$  é igual a probabilidade de transição desse mesmo contexto acrescido de um passo no passado para o mesmo símbolo  $a$ .

Outra consideração comum a todos os cenários é a respeito da exclusão de 10% das amostras geradas. Realizamos esta ação, para dissipar a influência que os valores iniciais

das cadeias possuem sobre a estacionariedade.

A seguir descrevemos em mais detalhes o processo de estimação em cada um dos cenários.

### Cenários 1 e 2

Seja  $x_0^{n-1} \in \{0,1\}^n$  a amostra da atividade de um neurônio que foi gerada a partir da árvore de contexto probabilística  $(\tau_i, Q_i)$ ,  $i \in \{1,2\}$ , conforme o Algoritmo 1, já com a remoção de 10% dos primeiros valores gerados. Para cada  $i \in \{1,2\}$ , calculamos as estimativas de máxima verossimilhança das probabilidades de transição da seguinte forma:

$$\hat{Q}_{i,n}(b|u_{-|u|}^{-1}) = \frac{\mathcal{N}_n(u_{-|u|}^{-1}b)}{\sum_{b=0}^1 \mathcal{N}_n(u_{-|u|}^{-1}b)}; \quad \forall u_{-|u|}^{-1} \in \tau_i \text{ e } \forall b \in \{0,1\},$$

em que  $\mathcal{N}_n(u)$  é o número de vezes que a sequência  $u$  apareceu na amostra  $x_0^{n-1}$ .

Com a estimativa  $\hat{Q}_{i,n} := \left\{ \hat{Q}_{i,n}(b|u_{-|u|}^{-1}) : b \in \{0,1\}, u_{-|u|}^{-1} \in \tau_i \right\}$  da família de probabilidades de transição em mãos, aplicamos o algoritmo contexto para estimar a árvore de contexto  $\hat{\tau}_i$ ,  $i \in \{1,2\}$ . Este algoritmo já foi apresentado e discutido com detalhes no Capítulo 4, portanto, para não nos fazermos repetitivos, decidimos não expô-lo aqui novamente.

Por fim, para cada  $i \in \{1,2\}$ , o estimador da entropia condicional associado a blocos de tamanho  $k$ ,  $k < n$ , é dado por

$$\hat{H}_{x_0^{n-1}}(X_k|X_0^{k-1}) = - \sum_{x_0^{k-1} \in \{0,1\}^k} \left[ \prod_{b=0}^1 \prod_{u \in \tau_i} \hat{Q}_{i,n}(b|u_{-|u|}^{-1})^{\mathcal{N}_n(u_{-|u|}^{-1}b)} \log_2 \hat{Q}_{i,n}(b|u_{-|u|}^{-1}) \right].$$

O tamanho do bloco  $k$  escolhido para este trabalho é da ordem  $\log_2 n$ , pois esta é uma das condições suficientes para obtermos consistência dos estimadores das probabilidades de transição (ver Seção III.2, p. 174, Shields (1996)).

### Cenário 3

Seja  $z_0^{n-1} \in \{0,1,2,3\}^n$  a amostra da atividade de dois neurônios que foi gerada a partir da árvore de contexto probabilística  $(\tau_3, Q_3)$  compatível com a cadeia de alcance variável bivariada  $(X,Y) := \{(X_t, Y_t) : t \in \mathbb{Z}\}$ , conforme o Algoritmo 2, já com a remoção de 10% dos primeiros valores gerados. Para cada  $u \in \tau_3$  e  $b \in \{0,1\}$ , calculamos as estimativas de

máxima verossimilhança das probabilidades de transição da seguinte forma:

$$\hat{Q}_{3,n} \left( b \mid u_{-|u|}^{-1} \right) = \frac{\mathcal{N}_n(u_{-|u|}^{-1} b)}{\sum_{b=0}^1 \mathcal{N}_n(u_{-|u|}^{-1} b)}.$$

em que  $\mathcal{N}_n(u)$  é o número de vezes que a sequência  $u$  apareceu na amostra  $z_0^{n-1}$ .

Com a estimativa  $\hat{Q}_{3,n} := \left\{ \hat{Q}_{3,n} \left( b \mid u_{-|u|}^{-1} \right) : b \in \{0,1\}, u_{-|u|}^{-1} \in \tau_3 \right\}$  da família de probabilidades de transição em mãos, aplicamos o algoritmo contexto para estimar a árvore de contexto  $\hat{\tau}_3$ .

Por fim, o estimador da entropia de transferência associado a blocos de tamanho  $k$  é dado por

$$\hat{T}_{X_0^{k-1} \rightarrow Y_0^{k-1}} = \hat{H}_{y_0^{n-1}}(Y_k \mid Y_0^{k-1}) - \hat{H}_{(x_0^{n-1}, y_0^{n-1})}(Y_k \mid X_0^{k-1} Y_0^{k-1}).$$

Note que,

$$\hat{H}_{y_0^{n-1}}(Y_k \mid Y_0^{k-1}) = - \sum_{y_0^k \in \{0,1\}^{k+1}} P(Y_0^k = y_0^k) \log_2 P(Y_k = y_k \mid X_0^{k-1} = x_0^{k-1}, Y_0^{k-1} = y_0^{k-1}),$$

em que

$$\begin{aligned} P(Y_0^k = y_0^k) &= \sum_{x_0^k \in \{0,1\}^{k+1}} P(Y_0^k = y_0^k, X_0^k = x_0^k) \\ &= \sum_{z_0^k \in \{0,1,2,3\}^{k+1}} P(Z_0^k = z_0^k) \\ &= \sum_{z_0^k \in \{0,1,2,3\}^{k+1}} \prod_{u \in \hat{\tau}_3} \prod_{a=0}^3 \hat{Q}_{3,n}(a \mid u)^{\mathcal{N}_n(ua)}, \end{aligned}$$

e

$$\begin{aligned} P(Y_k = y_k \mid Y_0^{k-1} = y_0^{k-1}) &= \frac{P(Y_0^k = y_0^k)}{P(Y_0^{k-1} = y_0^{k-1})} \\ &= \frac{\sum_{z_0^k \in \{0,1,2,3\}^{k+1}} \prod_{u \in \hat{\tau}_3} \prod_{a=0}^3 \hat{Q}_{3,n}(a \mid u)^{\mathcal{N}_n(ua)}}{\sum_{z_0^{k-1} \in \{0,1,2,3\}^k} \prod_{u \in \hat{\tau}_3} \prod_{a=0}^3 \hat{Q}_{3,n}(a \mid u)^{\mathcal{N}_n(ua)}}, \end{aligned}$$

para cada  $y_0^k \in \{0,1\}^{k+1}$  sendo que  $z_0^k$  é tal que  $z_i = x_i + 2y_i$ , para todo  $i = 0, \dots, k$ . Por outro lado,

$$\hat{H}_{(x_0^{n-1}, y_0^{n-1})} (Y_k | X_0^{k-1} Y_0^{k-1}) = - \sum_{\substack{x_0^{k-1} \in \{0,1\}^k \\ y_0^k \in \{0,1\}^{k+1}}} \left[ P(X_0^{k-1} = x_0^{k-1}; Y_0^k = y_0^k) \right. \\ \left. \times \log_2 P(Y_k = y_k | X_0^{k-1} = x_0^{k-1}, Y_0^{k-1} = y_0^{k-1}) \right],$$

em que

$$P(X_0^{k-1} = x_0^{k-1}; Y_0^k = y_0^k) = \sum_{x_k=0}^1 P(Y_0^k = y_0^k, X_0^k = x_0^k) \\ = \sum_{z_k=0}^3 P(Z_0^k = z_0^k) \\ = \sum_{z_k=0}^3 \prod_{u \in \hat{\tau}_3} \prod_{a=0}^3 \hat{Q}_{3,n}(a|u)^{\mathcal{N}_n(ua)},$$

e

$$P(Y_k = y_k | X_0^{k-1} = x_0^{k-1}, Y_0^{k-1} = y_0^{k-1}) = \frac{P(X_0^{k-1} = x_0^{k-1}; Y_0^k = y_0^k)}{P(X_0^{k-1} = x_0^{k-1}; Y_0^{k-1} = y_0^{k-1})} \\ = \frac{\sum_{z_k=0}^3 \prod_{u \in \hat{\tau}_3} \prod_{a=0}^3 \hat{Q}_{3,n}(a|u)^{\mathcal{N}_n(ua)}}{\prod_{u \in \hat{\tau}_3} \prod_{a=0}^3 \hat{Q}_{3,n}(a|u)^{\mathcal{N}_n(ua)}},$$

para cada  $y_0^k \in \{0,1\}^{k+1}$  sendo que  $z_0^k$  é tal que  $z_i = x_i + 2y_i$ , para todo  $i = 0, \dots, k$ . A fim de garantir a consistência dos estimadores das probabilidade de transição, assumimo, neste cenário, que o tamanho do bloco  $k$  é da ordem  $\frac{\log_2 n}{2}$ , pois, neste caso, o tamanho do alfabeto é 4.

## 6.3 Resultados

Nesta seção, estudamos a performance do método de estimação proposto levando em consideração os três cenários apresentados anteriormente e utilizando os dados simulados e as técnicas de estimação descritas na Seção 6.2. Especificamente, geramos réplicas de amostras para cada cenário descrito previamente, estimamos as árvores de contextos probabilística usando ambos, o método da máxima verossimilhança e o algoritmo Contexto e, por fim, estimamos a entropia condicional nos cenários 1 e 2 e a entropia de transferência no cenário 3. A performance dos estimadores é avaliada, para diferentes tamanhos de

amostra, pelo erro absoluto médio empírico e pelo erro quadrático médio empírico.

### 6.3.1 Cenário 1

Um neurônio, de uma área específica do cérebro, terá sua resposta temporal, disparos e não disparos, observada ao longo de um período de tempo finito e, então, iremos estimar, para diferentes tamanhos de amostras, a incerteza média esperada com relação à atividade do neurônio quando condicionada ao conhecimento do histórico de suas respostas temporais. Neste caso, vamos gerar as amostras de disparos neuronais a partir de uma família de probabilidades de transição, em que a probabilidade do neurônio disparar em um dado instante de tempo condicionada ao seu passado é modelada a partir de uma função de ativação logística.

Como descrito na Seção 6.2, geramos a atividade do neurônio a partir dos valores fixados  $x_{-4}^{-1} = 0101$  e da árvore de contexto probabilística  $(\tau_1, Q_1)$ . Considerando blocos de tamanho  $k = 10$  segue que a entropia condicional teórica é igual a

$$H_1(X_k | X_0^{k-1}) = 0,1529956 \text{ bits.}$$

Portanto, a incerteza média de um neurônio disparar, dados passados de tamanho  $k$ , é aproximadamente 0,153 bits.

A Tabela 6.1 apresenta os valores dos erros absolutos médios empíricos, obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos a Figura 6.6 que descreve o comportamento dos erros absolutos médios conforme os tamanhos das amostras aumentam.

Tabela 6.1: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_1$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_1(1   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,0522	0,0048	0,0009	0,0008	0,0006
c(1, 0)	0,1545	0,0038	0,0026	0,0011	0,0009
c(1, 0, 0)	0,5000	0,0129	0,0134	0,0039	0,0010
c(1, 0, 0, 0)	0,6899	0,0034	0,0023	0,0011	0,0009
c(0, 0, 0, 0)	1,0000	0,0000	0,0000	0,0000	0,0000

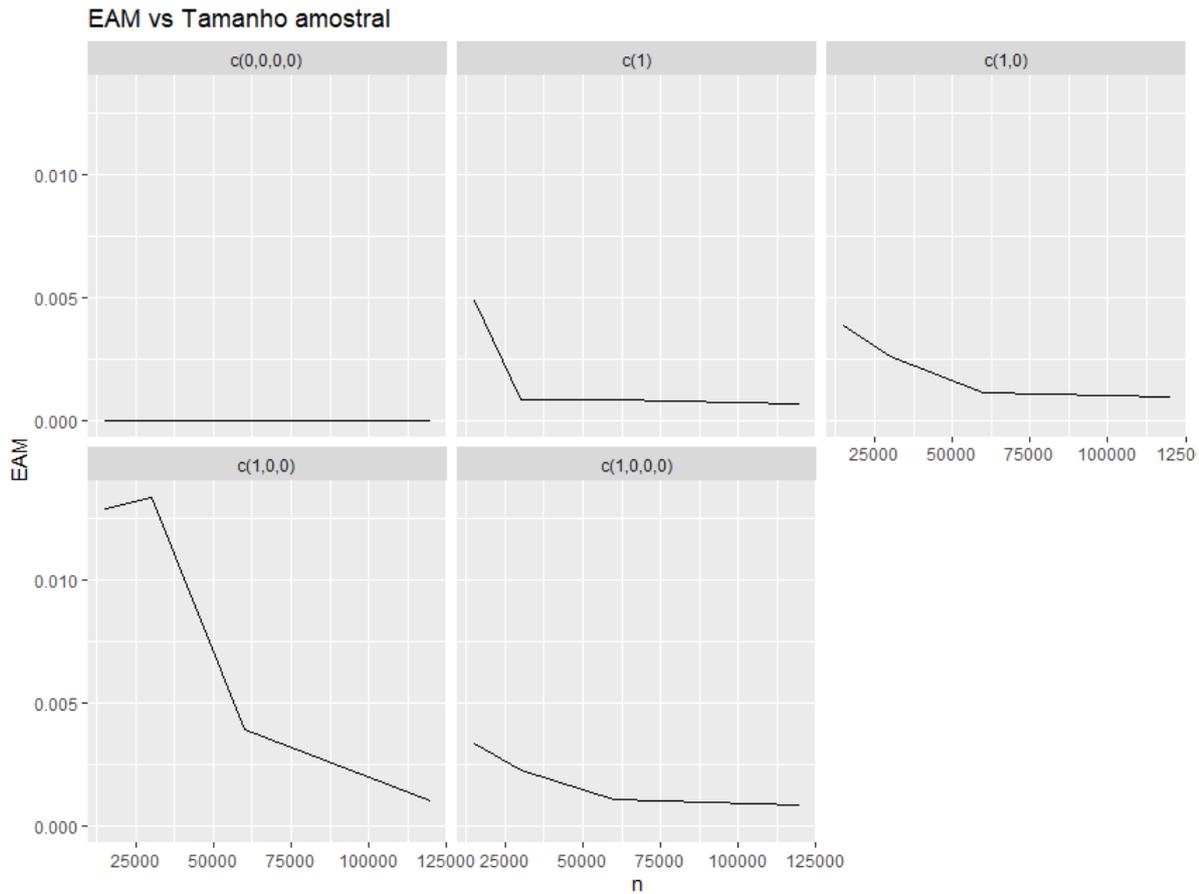


Figura 6.6: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

Ao observarmos tanto a Tabela 6.1 quanto a Figura 6.6, percebemos que conforme aumentamos o tamanho da amostra os erros absolutos tendem a diminuir, o que é um indicativo de que os estimadores são assintoticamente não-viesados.

A Tabela 6.2 apresenta os valores dos erros quadráticos médios empíricos, obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos a Figura 6.7 que descreve o comportamento dos erros quadráticos médios conforme os tamanhos das amostras aumentam.

Tabela 6.2: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_1$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_1(1 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
$c(1)$	0,0522	2,365e-05	7,436e-07	7,762e-06	4,437e-07
$c(1, 0)$	0,1545	1,480e-05	6,935e-06	1,293e-06	9,823e-07
$c(1, 0, 0)$	0,5000	1,666e-04	1,783e-04	1,531e-05	1,067e-06
$c(1, 0, 0, 0)$	0,6899	1,124e-05	5,284e-06	1,210e-06	7,802e-07
$c(0, 0, 0, 0)$	1,0000	0,0000	0,0000	0,0000	0,0000

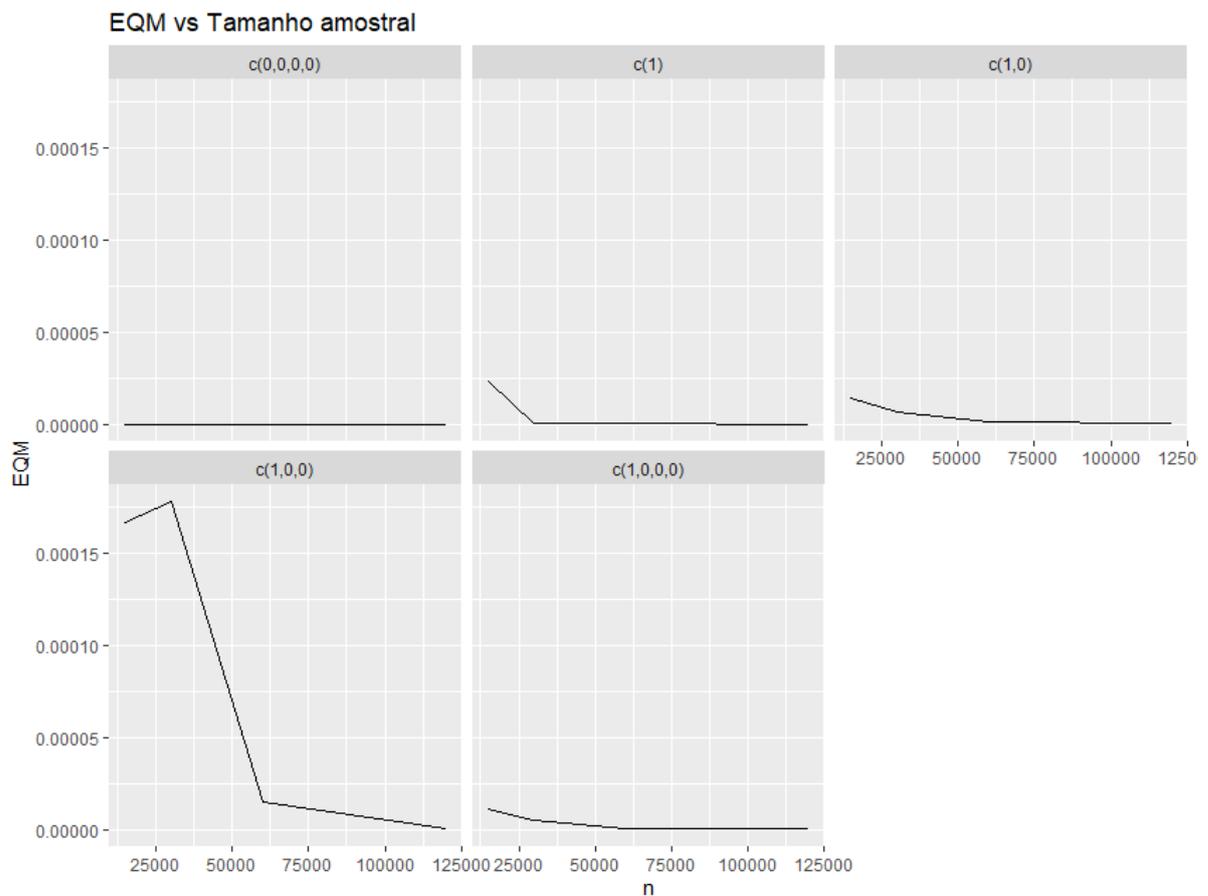


Figura 6.7: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

Ao observarmos tanto a Tabela 6.2 quanto a Figura 6.7, percebemos novamente que conforme aumentamos o tamanho da amostra os erros quadráticos médios empíricos tendem a diminuir, o que é um indicativo de que os estimadores são consistentes.

A Tabela 6.3 apresenta os valores dos erros absolutos médios empíricos (EAM) e dos erros quadráticos médios (EQM), obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos as Figuras 6.6 e 6.7 que descrevem o comportamento dos erros absolutos e quadráticos médios conforme os tamanhos das amostras aumentam.

Tabela 6.3: Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando  $\hat{Q}_1$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Métrica	Entropia Teórica	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
EAM	0,1529956	0,0064	0,0042	0,0014	0,0002
EQM	0,1529956	4,050e-05	1,773e-05	1,944e-06	2,238e-08

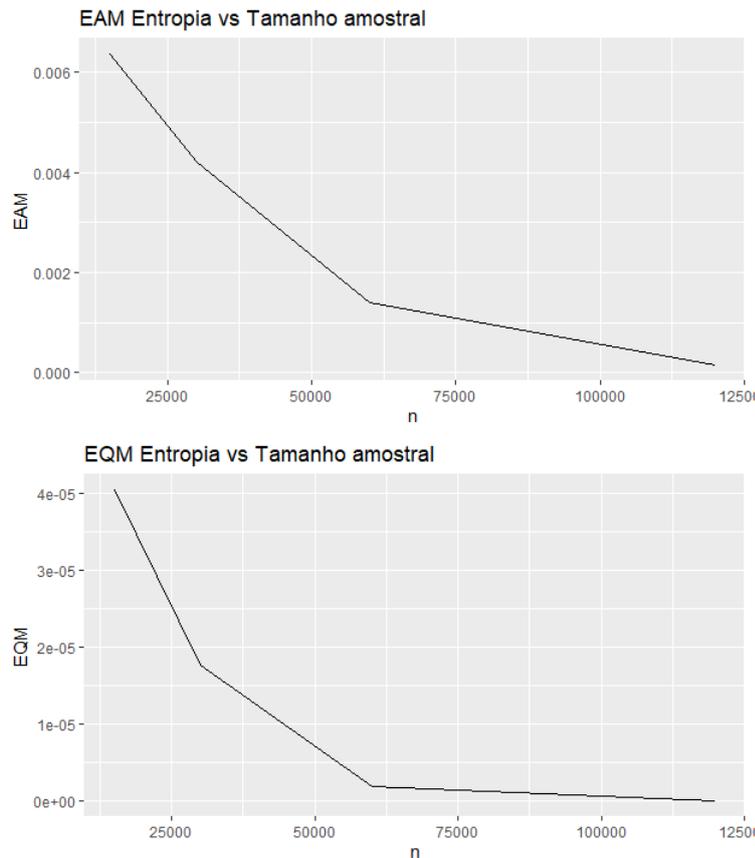


Figura 6.8: Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Ao verificarmos a Tabela 6.3 e as Figuras 6.8, temos um indicativo de que os estimadores da entropia condicional propostos neste trabalho são assintoticamente não-viesados e consistentes.

### 6.3.2 Cenário 2

Novamente, vamos considerar o problema de estimar a incerteza média esperada com relação à atividade de um neurônio quando condicionada ao conhecimento do histórico passado de suas respostas. Entretanto, neste caso, as amostras são geradas a partir de uma família de probabilidades de transição, em que a probabilidade do neurônio disparar em um dado instante de tempo condicionada ao seu passado é modelada a partir de uma função de ativação é a inversa da ligação complemento log-log.

Como descrito na Seção 6.2, geramos a atividade do neurônio a partir dos valores fixados  $x_{-4}^{-1} = 0101$  e da árvore de contexto probabilística  $(\tau_2, Q_2)$ . Considerando blocos de tamanho  $k = 10$  segue que a entropia condicional teórica é igual a

$$H_2(X_k | X_0^{k-1}) = 0,1035425 \text{ bits.}$$

Portanto, a incerteza média de um neurônio disparar, dado os passados de tamanho  $k$ , é aproximadamente 0,104 bits. Quando comparamos esse valor com o valor da entropia condicional teórica calculada para o cenário 1, observamos que, condicionada ao passado, a incerteza média a respeito da atividade do neurônio é menor no cenário 2, i.e., quando utilizamos como função de ativação a inversa da função de ligação complemento log-log.

A Tabela 6.4 apresenta os valores dos erros absolutos médios empíricos, obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos a Figura 6.9 que descreve o comportamento dos erros absolutos médios conforme os tamanhos das amostras aumentam.

Tabela 6.4: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_2$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_2(1 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
$c(1)$	0,0535	0,0003	0,0010	0,0008	0,0003
$c(1, 0)$	0,1669	0,0055	0,0006	0,0009	0,0001
$c(1, 0, 0)$	0,6321	0,0069	0,0029	0,0031	0,0009
$c(1, 0, 0, 0)$	0,8919	0,0038	0,0021	0,0019	0,0005
$c(0, 0, 0, 0)$	1,0000	0,0000	0,0000	0,0000	0,0000

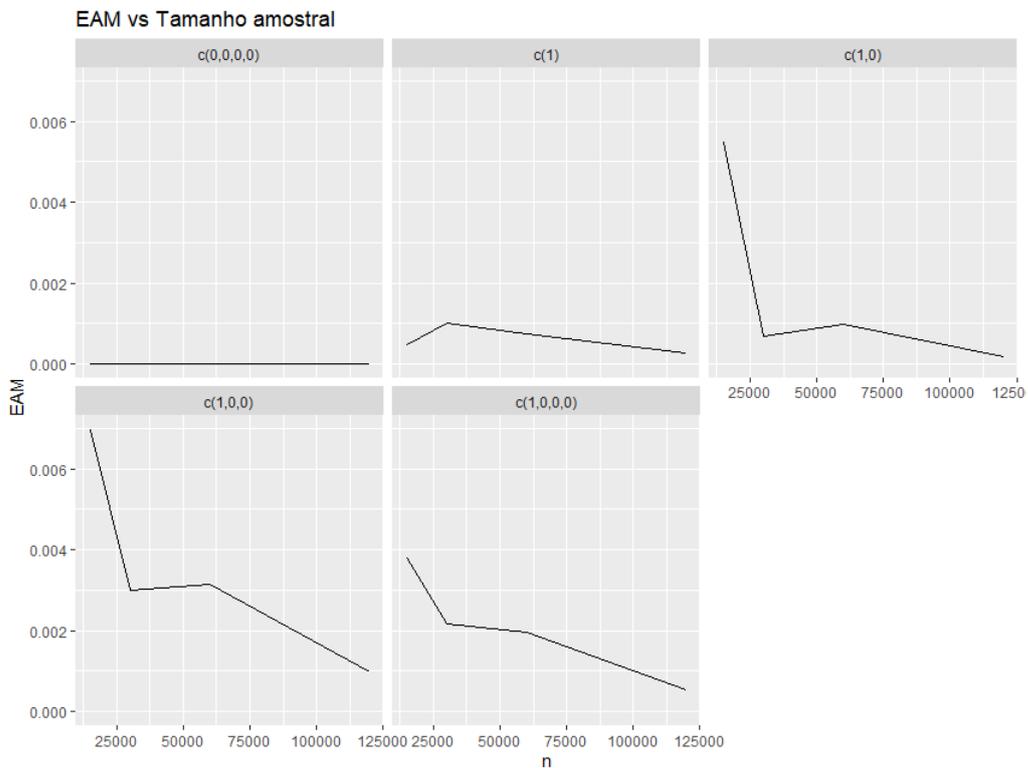


Figura 6.9: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

Ao observarmos tanto a Tabela 6.4 quanto a Figura 6.9, percebemos que conforme aumentamos o tamanho da amostra os erros absolutos tendem a diminuir, o que é um indicativo de que os estimadores são assintoticamente não-viesados.

A Tabela 6.5 apresenta os valores dos erros quadráticos médios empíricos, obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos a Figura 6.10 que descreve o comportamento dos erros quadráticos médios conforme os tamanhos das amostras aumentam.

Tabela 6.5: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_2$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_2(1 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
$c(1)$	0,0535	1,1109e-07	1,0008e-06	5,6567e-07	8,0202e-08
$c(1, 0)$	0,1669	3,0014e-05	4,8400e-07	9,8272e-07	3,0246e-08
$c(1, 0, 0)$	0,6321	4,8665e-05	8,9177e-06	9,8205e-06	9,8667e-07
$c(1, 0, 0, 0)$	0,8919	1,4345e-05	4,7251e-06	3,7920e-06	2,9934e-07
$c(0, 0, 0, 0)$	1,0000	0,0000	0,0000	0,0000	0,0000

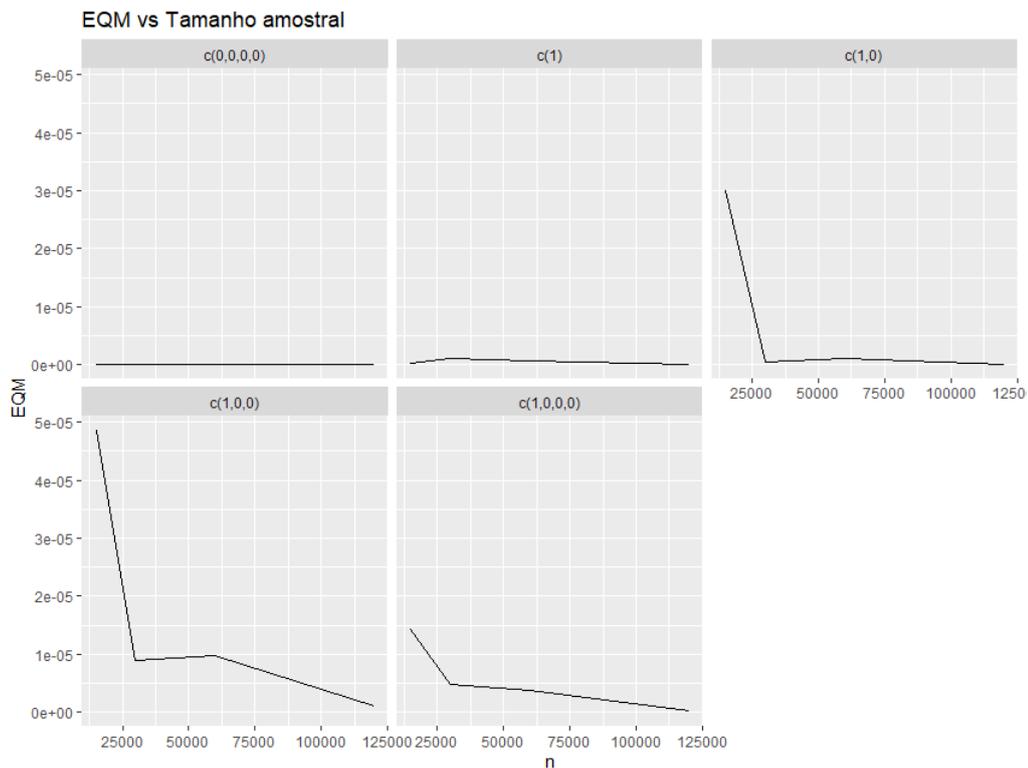


Figura 6.10: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

Ao observarmos tanto a Tabela 6.5 quanto a Figura 6.10, percebemos novamente que conforme aumentamos o tamanho da amostra os erros quadráticos médios empíricos tendem a diminuir, o que é um indicativo de que os estimadores são consistentes.

A Tabela 6.3 apresenta os valores dos erros absolutos médios empíricos (EAM) e dos erros quadráticos médios (EQM), obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos as Figuras 6.6 e 6.7 que descrevem o comportamento dos erros absolutos e quadráticos médios conforme os tamanhos das amostras aumentam.

Tabela 6.6: Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando  $\hat{Q}_2$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Métrica	Entropia Teórica	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
EAM	0,1035425	0,0054908	0,0018539	0,0012917	0,0005955
EQM	0,1035425	3,01489e-05	3,43695e-06	1,66849e-06	3,54620e-07

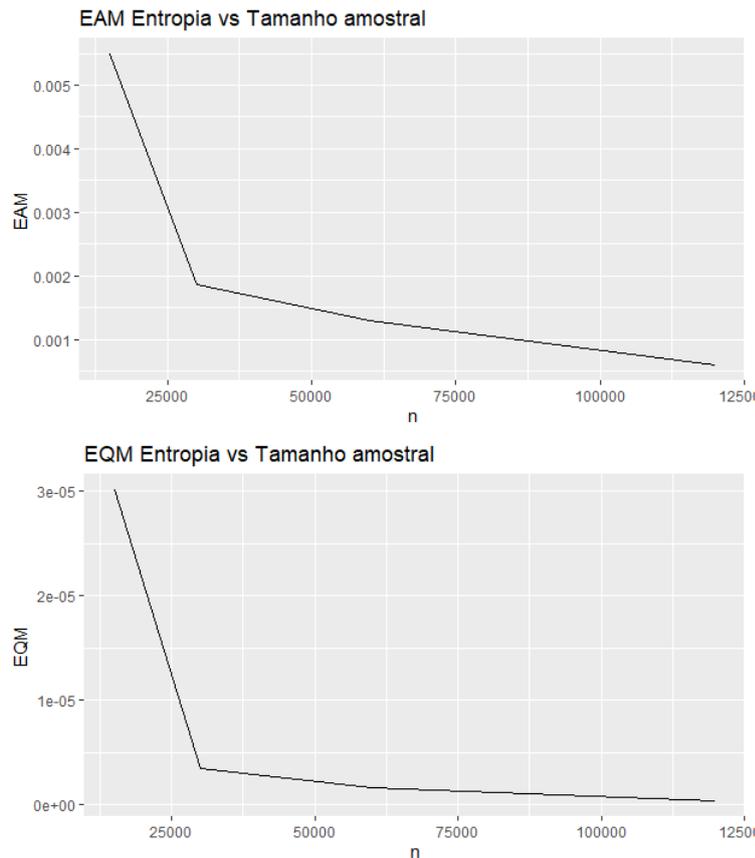


Figura 6.11: Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Ao verificarmos a Tabela 6.6 e as Figuras 6.11, temos um indicativo de que os estimadores da entropia condicional propostos neste trabalho são assintoticamente não-viesados e consistentes.

Ao compararmos os resultados obtidos no cenário 2 com aqueles obtidos no cenário 1, observamos que para cada tamanho de amostra fixado, a ordem dos erros empíricos obtidos são suficientemente próximos em ambos os cenários, tanto na estimação das probabilidades de transição quanto na estimação das medidas da informação. Portanto, temos um indicativo de que a escolha da função de ativação não interfere na performance dos métodos de estimação propostos, os quais alcançam consistência em ambos os cenários.

### 6.3.3 Cenário 3

Dois neurônios,  $N_1$  e  $N_2$ , de uma área específica do cérebro, tem sua resposta temporal, disparos e não disparos, observadas ao longo de um período de tempo finito e, então, iremos estimar, para diferentes tamanhos de amostra, a informação que flui de um neurônio ao outro. Em outras palavras, vamos verificar o quanto diminui a incerteza média esperada com relação à atividade do neurônio  $N_i$  quando além do conhecimento do passado das suas respostas temporais, sabemos também o histórico da atividade do neurônio  $N_j$ ,  $i, j = 1, 2$  com  $i \neq j$ .

Como descrito na Seção 6.2, geramos a atividade dos neurônios a partir dos valores fixados  $\{(x_i, y_i) : i = 1, 2, 3, 4\} = \{(0, 1), (0, 1), (1, 0), (1, 1)\}$  e da árvore de contexto probabilística  $(\tau_3, Q_3)$ . Considerando blocos de tamanho  $k = 6$  segue que a entropia de transferência teórica é igual a

$$\begin{aligned} T_{Y_0^{k-1} \rightarrow X_0^{k-1}} &= H(Y_k | Y_0^{k-1}) - H(Y_k | Y_0^{k-1}; X_0^{k-1}) \\ &= 0,1335908 - 0,0027052 \\ &= 0,1308856 \text{ bits.} \end{aligned}$$

Portanto, quando condicionado ao conhecimento da atividade histórica do neurônio  $N_2$  (modelada pela cadeia  $Y$ ), o conhecimento da atividade histórica do neurônio  $N_1$  (modelada pela cadeia  $X$ ) diminui em aproximadamente 98% a incerteza média relativa a atividade do neurônio  $N_2$ . Portanto, temos um indicativo de que os neurônios em questão estão conectados. Esta conexão já era esperada uma vez que a matriz de transição foi definida de acordo com a teoria Hebbiana.

As Tabelas 6.7 até a 6.12 apresentam os valores dos erros absolutos médios empíricos, obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos também as Figura 6.12 até a 6.15 que descrevem o comportamento dos erros absolutos médios conforme os tamanhos das amostras aumentam.

Tabela 6.7: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(0 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(0   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,5500	0,5500	0,5500	0,0012	0,0007
c(2)	0,5500	0,5500	0,5500	0,0178	0,0078
c(3)	0,6500	0,6500	0,0019	0,0034	0,0014
c(1,0)	0,4250	0,4250	0,4250	0,0174	0,0026
c(2,0)	0,4250	0,0333	0,4250	0,0032	0,0008
c(3,0)	0,4500	0,0230	0,4500	0,0043	0,0015
c(0,2)	-	0,5259	0,5685	-	-
c(1,2)	-	0,5252	0,5153	-	-
c(2,2)	-	0,5469	0,5224	-	-
c(3,2)	-	0,5240	0,5164	-	-
c(0,1)	-	-	0,5321	-	-
c(1,1)	-	0,5403	-	-	-
c(2,1)	-	-	0,5211	-	-
c(3,1)	-	0,5536	0,5274	-	-
c(2,3)	-	0,6098	-	-	-
c(3,3)	-	0,6521	-	-	-
c(1, 0, 0)	0,3500	0,0332	0,0298	0,0009	0,0019
c(2, 0, 0)	0,3500	0,0115	0,0099	0,0014	0,0014
c(3, 0, 0)	0,2500	0,0125	0,0112	0,0080	0,0052
c(0, 1, 0)	-	-	0,4482	-	-
c(1, 1, 0)	-	-	0,4057	-	-
c(2, 1, 0)	-	0,4455	-	-	-
c(3, 1, 0)	-	-	0,4481	-	-
c(0, 1, 1)	-	-	0,5487	-	-
c(1, 1, 1)	-	-	0,5063	-	-
c(2, 1, 1)	-	-	0,5207	-	-
c(3, 1, 1)	-	-	0,5314	-	-
c(0, 2, 0)	-	-	0,4408	-	-
c(1, 2, 0)	-	-	0,3641	-	-
c(3, 2, 0)	-	-	0,3902	-	-
c(0, 2, 1)	-	0,5618	-	-	-
c(1, 2, 1)	-	0,5541	-	-	-
c(2, 2, 1)	-	0,5607	-	-	-
c(3, 2, 1)	-	0,5013	-	-	-
c(1, 0, 1)	-	0,5204	-	-	-

Tabela 6.8: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(0 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(0   \text{Estado})$	n = 13500	n = 27000	n = 54000	n = 108000
c(2, 0, 1)	-	0,5551	-	-	-
c(3, 0, 1)	-	0,5044	-	-	-
c(1, 1, 3)	-	0,4901	-	-	-
c(2, 1, 3)	-	0,5581	-	-	-
c(3, 1, 3)	-	0,5396	-	-	-
c(0, 0, 3)	-	0,6602	-	-	-
c(1, 0, 3)	-	0,6158	-	-	-
c(2, 0, 3)	-	0,6421	-	-	-
c(3, 0, 3)	-	0,6401	-	-	-
c(1, 0, 0, 0)	0,1000	0,0102	0,0088	0,0047	0,0019
c(2, 0, 0, 0)	0,1000	0,0512	0,0095	0,0048	0,0043
c(3, 0, 0, 0)	0,1000	0,0331	0,0251	0,0004	0,0002
c(0, 0, 0, 0)	0,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,4721	-	-	-
c(1, 0, 0, 1)	-	0,4741	-	-	-
c(2, 0, 0, 1)	-	0,4940	-	-	-
c(3, 0, 0, 1)	-	0,4504	-	-	-
c(0, 2, 2, 0)	-	-	0,5143	-	-
c(1, 2, 2, 0)	-	-	0,4168	-	-
c(2, 2, 2, 0)	-	-	0,4187	-	-
c(3, 2, 2, 0)	-	-	0,5155	-	-
c(0, 2, 1, 0)	-	-	0,4673	-	-
c(1, 2, 1, 0)	-	-	0,4365	-	-
c(2, 2, 1, 0)	-	-	0,2415	-	-
c(3, 2, 1, 0)	-	-	0,4988	-	-

Tabela 6.9: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(1 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(1   \text{Estado})$	n = 13500	n = 27000	n = 54000	n = 108000
c(1)	0,1000	0,1000	0,1000	0,0111	0,0011
c(2)	0,2000	0,2000	0,2000	0,0105	0,0055
c(3)	0,1000	0,1000	0,0096	0,0069	0,0049
c(1,0)	0,1500	0,1500	0,1500	0,0041	0,0009
c(2,0)	0,2500	0,0241	0,2500	0,0081	0,0041
c(3,0)	0,1500	0,0117	0,1500	0,0041	0,0013
c(0,2)	-	0,2213	0,2144	-	-
c(1,2)	-	0,2401	0,2377	-	-
c(2,2)	-	0,2044	0,2188	-	-
c(3,2)	-	0,1901	0,2222	-	-
c(0,1)	-	-	0,1140	-	-
c(1,1)	-	0,1174	-	-	-
c(2,1)	-	-	0,1087	-	-

Tabela 6.10: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(1 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(1   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(3,1)	-	0,1290	0,1008	-	-
c(2,3)	-	0,1044	-	-	-
c(3,3)	-	0,1099	-	-	-
c(1, 0, 0)	0,1500	0,0270	0,0302	0,0115	0,0087
c(2, 0, 0)	0,3000	0,0180	0,0088	0,0140	0,0086
c(3, 0, 0)	0,2000	0,0098	0,0105	0,0059	0,0032
c(0, 1, 0)	-	-	0,1841	-	-
c(1, 1, 0)	-	-	0,1402	-	-
c(2, 1, 0)	-	0,2004	-	-	-
c(3, 1, 0)	-	-	0,1544	-	-
c(0, 1, 1)	-	-	0,1240	-	-
c(1, 1, 1)	-	-	0,1189	-	-
c(2, 1, 1)	-	-	0,1904	-	-
c(3, 1, 1)	-	-	0,1305	-	-
c(0, 2, 0)	-	-	0,2614	-	-
c(1, 2, 0)	-	-	0,2430	-	-
c(3, 2, 0)	-	-	0,2331	-	-
c(0, 2, 1)	-	0,0915	-	-	-
c(1, 2, 1)	-	0,0881	-	-	-
c(2, 2, 1)	-	0,1201	-	-	-
c(3, 2, 1)	-	0,1044	-	-	-
c(1, 0, 1)	-	0,1340	-	-	-
c(2, 0, 1)	-	0,1105	-	-	-
c(3, 0, 1)	-	0,1088	-	-	-
c(1, 1, 3)	-	0,1102	-	-	-
c(2, 1, 3)	-	0,0981	-	-	-
c(3, 1, 3)	-	0,1503	-	-	-
c(0, 0, 3)	-	0,1108	-	-	-
c(1, 0, 3)	-	0,1074	-	-	-
c(2, 0, 3)	-	0,1096	-	-	-
c(3, 0, 3)	-	0,1390	-	-	-
c(1, 0, 0, 0)	0,2500	0,0811	0,0273	0,0130	0,0102
c(2, 0, 0, 0)	0,3500	0,0537	0,0121	0,00007	0,00004
c(3, 0, 0, 0)	0,2500	0,0401	0,0368	0,0250	0,0117
c(0, 0, 0, 0)	0,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,1409	-	-	-
c(1, 0, 0, 1)	-	0,1902	-	-	-
c(2, 0, 0, 1)	-	0,1105	-	-	-
c(3, 0, 0, 1)	-	0,1209	-	-	-
c(0, 2, 2, 0)	-	-	0,2306	-	-
c(1, 2, 2, 0)	-	-	0,1705	-	-
c(2, 2, 2, 0)	-	-	0,2036	-	-
c(3, 2, 2, 0)	-	-	0,1503	-	-
c(0, 2, 1, 0)	-	-	0,1805	-	-
c(1, 2, 1, 0)	-	-	0,2047	-	-
c(2, 2, 1, 0)	-	-	0,1639	-	-
c(3, 2, 1, 0)	-	-	0,1501	-	-

Tabela 6.11: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(2 \mid \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(2 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,2000	0,2000	0,2000	0,0042	0,0038
c(2)	0,1000	0,1000	0,1000	0,0114	0,0014
c(3)	0,1000	0,1000	0,0060	0,0048	0,0028
c(1,0)	0,2500	0,2500	0,2500	0,0051	0,0001
c(2,0)	0,1500	0,0301	0,1500	0,0061	0,0021
c(3,0)	0,1500	0,0200	0,1500	0,0132	0,0049
c(0,2)	-	0,1408	0,0905	-	-
c(1,2)	-	0,1026	0,0937	-	-
c(2,2)	-	0,1108	0,1105	-	-
c(3,2)	-	0,1390	0,1703	-	-
c(0,1)	-	-	0,2044	-	-
c(1,1)	-	0,2201	-	-	-
c(2,1)	-	-	0,1899	-	-
c(3,1)	-	0,2061	0,2011	-	-
c(2,3)	-	0,1109	-	-	-
c(3,3)	-	0,0922	-	-	-
c(1, 0, 0)	0,3000	0,0206	0,0221	0,0080	0,0052
c(2, 0, 0)	0,1500	0,0130	0,0107	0,0006	0,0022
c(3, 0, 0)	0,2000	0,0104	0,0097	0,0027	0,0001
c(0, 1, 0)	-	-	0,2601	-	-
c(1, 1, 0)	-	-	0,2804	-	-
c(2, 1, 0)	-	0,2240	-	-	-
c(3, 1, 0)	-	-	0,2344	-	-
c(0, 1, 1)	-	-	0,2014	-	-
c(1, 1, 1)	-	-	0,2168	-	-
c(2, 1, 1)	-	-	0,1984	-	-
c(3, 1, 1)	-	-	0,1993	-	-
c(0, 2, 0)	-	-	0,1420	-	-
c(1, 2, 0)	-	-	0,1620	-	-
c(3, 2, 0)	-	-	0,1337	-	-
c(0, 2, 1)	-	0,2077	-	-	-
c(1, 2, 1)	-	0,2133	-	-	-
c(2, 2, 1)	-	0,1972	-	-	-
c(3, 2, 1)	-	0,2201	-	-	-
c(1, 0, 1)	-	0,2047	-	-	-
c(2, 0, 1)	-	0,2108	-	-	-
c(3, 0, 1)	-	0,2091	-	-	-
c(1, 1, 3)	-	0,1047	-	-	-
c(2, 1, 3)	-	0,1097	-	-	-
c(3, 1, 3)	-	0,1228	-	-	-
c(0, 0, 3)	-	0,0998	-	-	-
c(1, 0, 3)	-	0,1084	-	-	-
c(2, 0, 3)	-	0,1067	-	-	-
c(3, 0, 3)	-	0,0978	-	-	-

Tabela 6.12: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(2 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(2   \text{Estado})$	n = 13500	n = 27000	n = 54000	n = 108000
c(1, 0, 0, 0)	0,3500	0,0307	0,0182	0,0098	0,0070
c(2, 0, 0, 0)	0,2500	0,0228	0,0304	0,0083	0,0029
c(3, 0, 0, 0)	0,2500	0,0187	0,0174	0,0040	0,0011
c(0, 0, 0, 0)	0,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,2041	-	-	-
c(1, 0, 0, 1)	-	0,2204	-	-	-
c(2, 0, 0, 1)	-	0,1756	-	-	-
c(3, 0, 0, 1)	-	0,2005	-	-	-
c(0, 2, 2, 0)	-	-	0,1173	-	-
c(1, 2, 2, 0)	-	-	0,1608	-	-
c(2, 2, 2, 0)	-	-	0,1408	-	-
c(3, 2, 2, 0)	-	-	0,1507	-	-
c(0, 2, 1, 0)	-	-	0,2401	-	-
c(1, 2, 1, 0)	-	-	0,2301	-	-
c(2, 2, 1, 0)	-	-	0,2044	-	-
c(3, 2, 1, 0)	-	-	0,2599	-	-

Tabela 6.13: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(3   \text{Estado})$	n = 13500	n = 27000	n = 54000	n = 108000
c(1)	0,1500	0,1500	0,1500	0,0141	0,0041
c(2)	0,1500	0,1500	0,1500	0,0041	0,0009
c(3)	0,1500	0,1500	0,0091	0,0084	0,0064
c(1,0)	0,1750	0,175	0,1750	0,0082	0,0018
c(2,0)	0,1750	0,0340	0,1750	0,0109	0,0070
c(3,0)	0,2500	0,0108	0,2500	0,0049	0,0021
c(0,2)	-	0,1120	0,1264	-	-
c(1,2)	-	0,1321	0,1572	-	-
c(2,2)	-	0,1379	0,1483	-	-
c(3,2)	-	0,1468	0,0911	-	-
c(0,1)	-	-	0,1495	-	-
c(1,1)	-	0,1222	-	-	-
c(2,1)	-	-	0,1803	-	-
c(3,1)	-	0,1130	0,1707	-	-
c(2,3)	-	0,1749	-	-	-
c(3,3)	-	0,1458	-	-	-
c(1, 0, 0)	0,2000	0,0172	0,0111	0,0044	0,0016
c(2, 0, 0)	0,2000	0,0233	0,0107	0,0107	0,0079
c(3, 0, 0)	0,3500	0,0120	0,0114	0,0047	0,0019

Tabela 6.14: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(3   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(0, 1, 0)	-	-	0,1076	-	-
c(1, 1, 0)	-	-	0,1737	-	-
c(2, 1, 0)	-	0,1306	-	-	-
c(3, 1, 0)	-	-	0,1631	-	-
c(0, 1, 1)	-	-	0,1259	-	-
c(1, 1, 1)	-	-	0,1580	-	-
c(2, 1, 1)	-	-	0,0905	-	-
c(3, 1, 1)	-	-	0,1388	-	-
c(0, 2, 0)	-	-	0,1558	-	-
c(1, 2, 0)	-	-	0,2309	-	-
c(3, 2, 0)	-	-	0,2430	-	-
c(0, 2, 1)	-	0,1390	-	-	-
c(1, 2, 1)	-	0,1445	-	-	-
c(2, 2, 1)	-	0,1411	-	-	-
c(3, 2, 1)	-	0,1742	-	-	-
c(1, 0, 1)	-	0,1409	-	-	-
c(2, 0, 1)	-	0,1236	-	-	-
c(3, 0, 1)	-	0,1777	-	-	-
c(1, 1, 3)	-	0,2950	-	-	-
c(2, 1, 3)	-	0,2341	-	-	-
c(3, 1, 3)	-	0,1873	-	-	-
c(0, 0, 3)	-	0,1292	-	-	-
c(1, 0, 3)	-	0,1684	-	-	-
c(2, 0, 3)	-	0,1416	-	-	-
c(3, 0, 3)	-	0,1231	-	-	-
c(1, 0, 0, 0)	0,3000	0,0813	0,0398	0,0275	0,0191
c(2, 0, 0, 0)	0,3000	0,0457	0,0207	0,0130	0,0072
c(3, 0, 0, 0)	0,4000	0,0176	0,0257	0,0203	0,0122
c(0, 0, 0, 0)	1,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,1829	-	-	-
c(1, 0, 0, 1)	-	0,1153	-	-	-
c(2, 0, 0, 1)	-	0,2199	-	-	-
c(3, 0, 0, 1)	-	0,2282	-	-	-
c(0, 2, 2, 0)	-	-	0,1384	-	-
c(1, 2, 2, 0)	-	-	0,2519	-	-
c(2, 2, 2, 0)	-	-	0,2369	-	-
c(3, 2, 2, 0)	-	-	0,1835	-	-
c(0, 2, 1, 0)	-	-	0,1121	-	-
c(1, 2, 1, 0)	-	-	0,1287	-	-
c(2, 2, 1, 0)	-	-	0,3902	-	-
c(3, 2, 1, 0)	-	-	0,0912	-	-

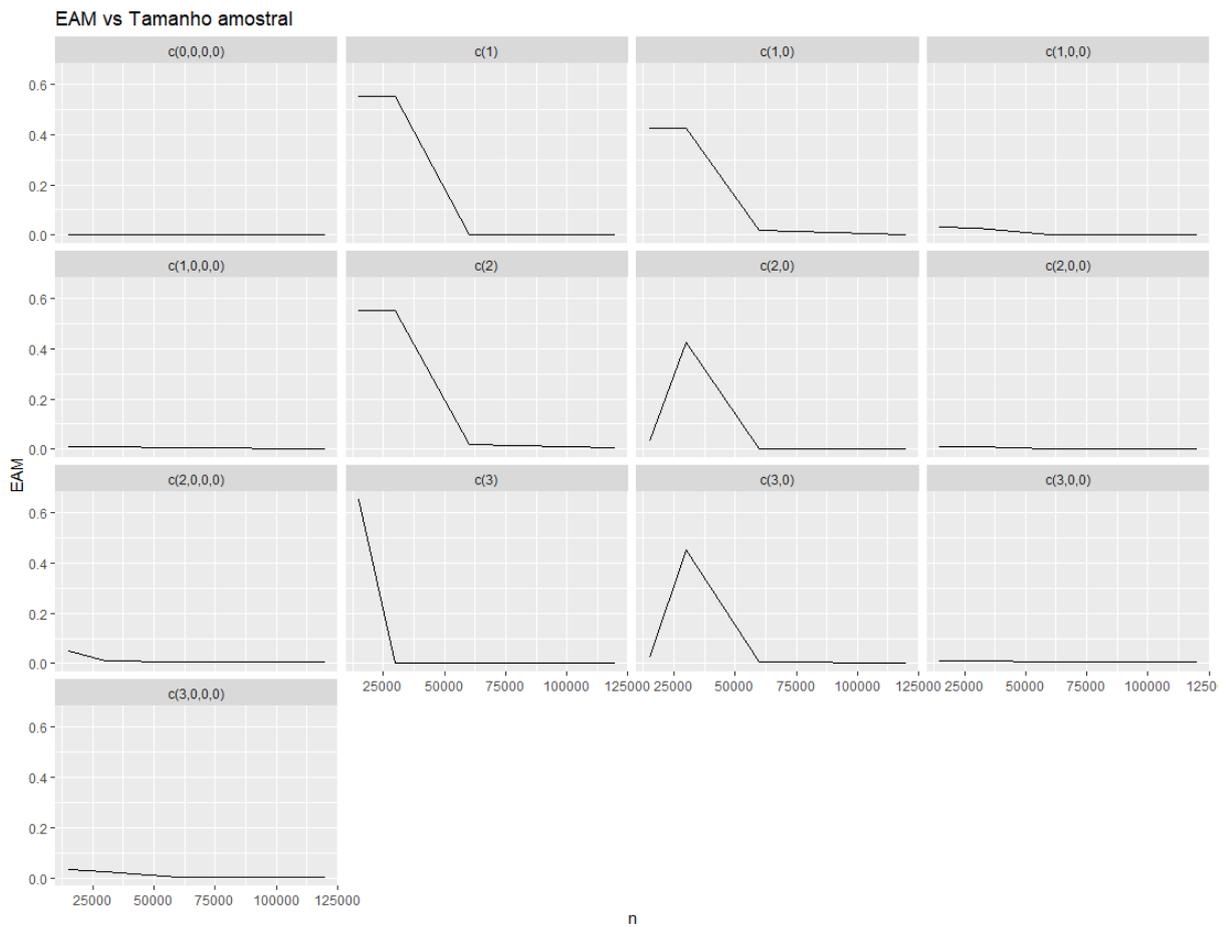


Figura 6.12: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(0 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

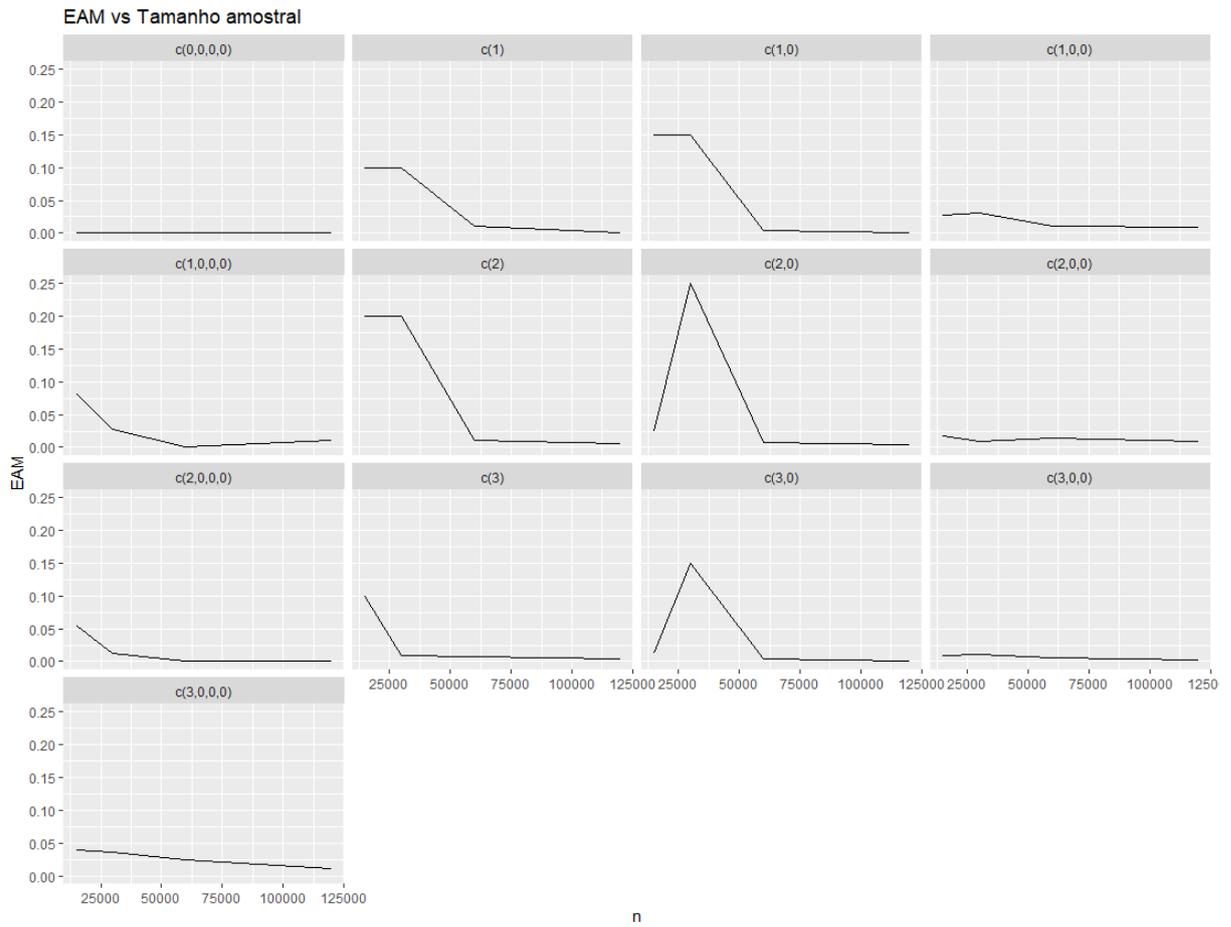


Figura 6.13: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(1 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

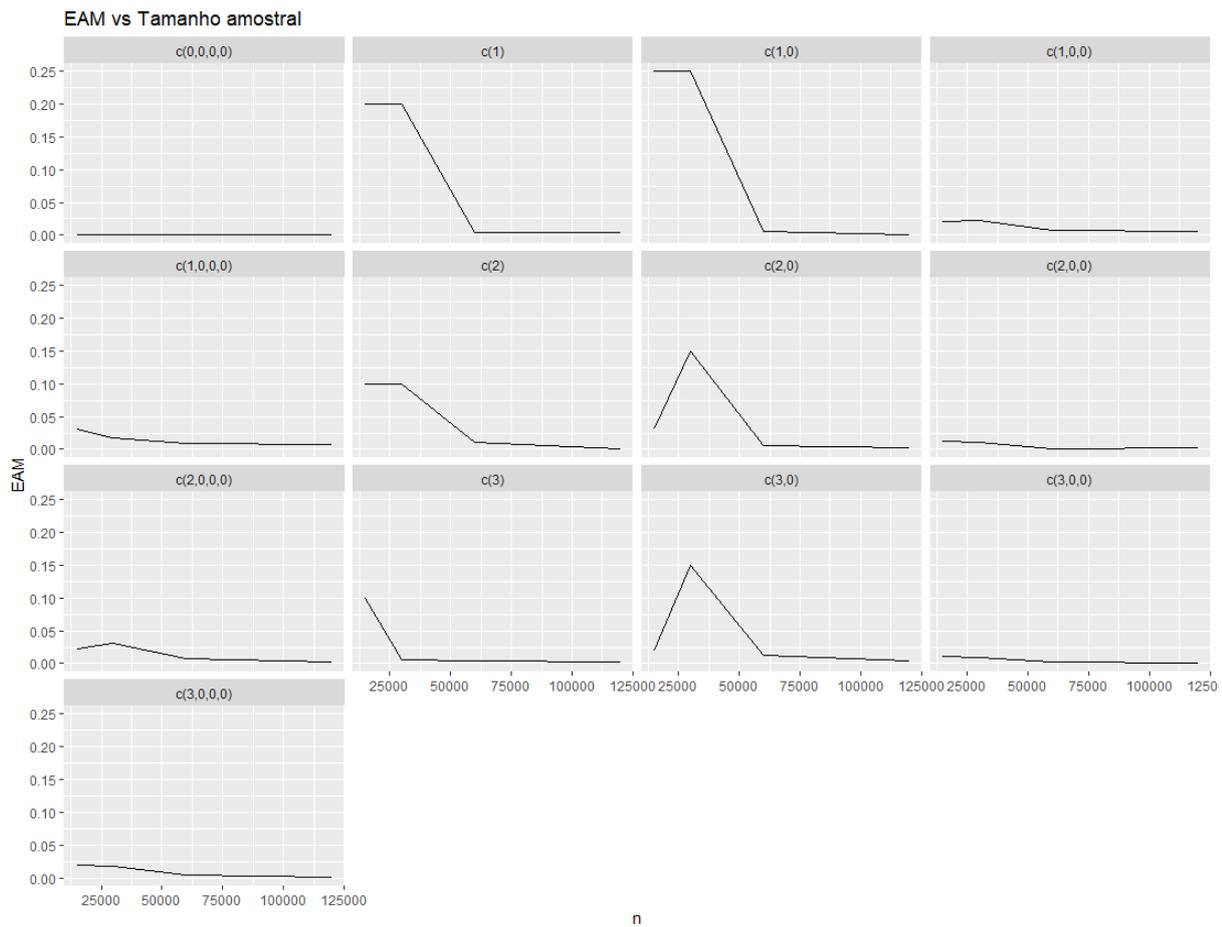


Figura 6.14: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(2 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

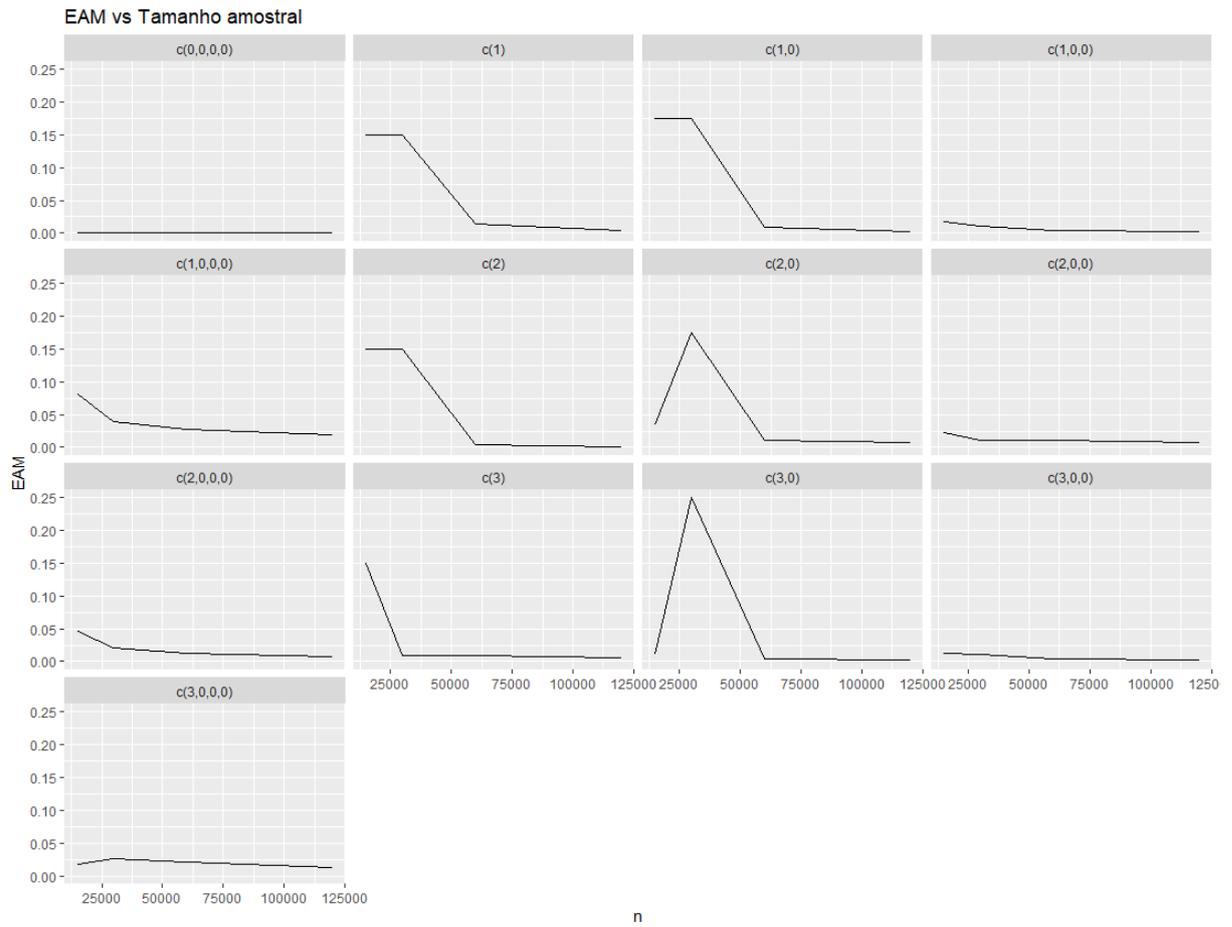


Figura 6.15: Erros absolutos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(3 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

Ao observarmos as Tabelas 6.7 até a 6.14 e as Figuras 6.12 até a 6.15, percebemos que conforme aumentamos o tamanho da amostra os erros absolutos médios empíricos tendem a diminuir, o que é um indicativo de que os estimadores são assintoticamente não-viesados.

As Tabela 6.15 até a 6.22 a seguir apresentam os valores dos erros quadráticos médios empíricos, obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos também as Figuras 6.16 até a 6.19 que descreve o comportamento dos erros quadráticos médios conforme os tamanhos das amostras aumentam.

Tabela 6.15: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(0 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(0   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,5500	0,3025	0,3025	1,44e-06	4,90e-07
c(2)	0,5500	0,3025	0,3025	3,17e-04	6,08e-05
c(3)	0,6500	0,4225	3,61e-06	1,16e-05	1,96e-06
c(1,0)	0,4250	0,1806	0,1806	3,03e-04	6,76e-06
c(2,0)	0,4250	1,11e-03	0,1806	1,02e-05	6,4e-07
c(3,0)	0,4500	5,29e-04	0,2025	1,85e-05	2,25e-06
c(0,2)	-	0,2766	0,3232	-	-
c(1,2)	-	0,2758	0,2655	-	-
c(2,2)	-	0,2991	0,2729	-	-
c(3,2)	-	0,2746	0,2667	-	-
c(0,1)	-	-	0,2831	-	-
c(1,1)	-	0,2919	-	-	-
c(2,1)	-	-	0,2715	-	-
c(3,1)	-	0,3065	0,2782	-	-
c(2,3)	-	0,3719	-	-	-
c(3,3)	-	0,4252	-	-	-
c(1, 0, 0)	0,3500	1,10e-03	8,88e-04	8,1e-07	3,61e-06
c(2, 0, 0)	0,3500	1,32e-04	9,80e-05	1,96e-06	1,96e-06
c(3, 0, 0)	0,2500	1,56e-04	1,25e-04	6,4e-05	2,70e-05
c(0, 1, 0)	-	-	0,2009	-	-
c(1, 1, 0)	-	-	0,1646	-	-
c(2, 1, 0)	-	0,1985	-	-	-
c(3, 1, 0)	-	-	0,2008	-	-
c(0, 1, 1)	-	-	0,3011	-	-
c(1, 1, 1)	-	-	0,2563	-	-
c(2, 1, 1)	-	-	0,2711	-	-
c(3, 1, 1)	-	-	0,2824	-	-
c(0, 2, 0)	-	-	0,1943	-	-
c(1, 2, 0)	-	-	0,1326	-	-
c(3, 2, 0)	-	-	0,1523	-	-

Tabela 6.16: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(0 \mid \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(0 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(0, 2, 1)	-	0,3156	-	-	-
c(1, 2, 1)	-	0,3070	-	-	-
c(2, 2, 1)	-	0,3144	-	-	-
c(3, 2, 1)	-	0,2513	-	-	-
c(1, 0, 1)	-	0,2708	-	-	-
c(2, 0, 1)	-	0,3081	-	-	-
c(3, 0, 1)	-	0,2544	-	-	-
c(1, 1, 3)	-	0,2402	-	-	-
c(2, 1, 3)	-	0,3115	-	-	-
c(3, 1, 3)	-	0,2912	-	-	-
c(0, 0, 3)	-	0,4359	-	-	-
c(1, 0, 3)	-	0,3792	-	-	-
c(2, 0, 3)	-	0,4123	-	-	-
c(3, 0, 3)	-	0,4097	-	-	-
c(1, 0, 0, 0)	0,1000	1,04e-04	7,74e-05	2,21e-05	3,61e-06
c(2, 0, 0, 0)	0,1000	2,62e-03	9,03e-05	2,30e-05	1,85e-05
c(3, 0, 0, 0)	0,1000	1,1e-03	6,3e-04	1,6e-07	4,0e-08
c(0, 0, 0, 0)	0,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,2229	-	-	-
c(1, 0, 0, 1)	-	0,2248	-	-	-
c(2, 0, 0, 1)	-	0,2440	-	-	-
c(3, 0, 0, 1)	-	0,2029	-	-	-
c(0, 2, 2, 0)	-	-	0,2645	-	-
c(1, 2, 2, 0)	-	-	0,1737	-	-
c(2, 2, 2, 0)	-	-	0,1753	-	-
c(3, 2, 2, 0)	-	-	0,2657	-	-
c(0, 2, 1, 0)	-	-	0,2184	-	-
c(1, 2, 1, 0)	-	-	0,1905	-	-
c(2, 2, 1, 0)	-	-	0,0583	-	-
c(3, 2, 1, 0)	-	-	0,2488	-	-

Tabela 6.17: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(1 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(1   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,1000	0,1000	0,1000	0,0111	0,0011
c(2)	0,2000	0,2000	0,2000	0,0105	0,0055
c(3)	0,1000	0,1000	0,0096	0,0069	0,0049
c(1,0)	0,1500	0,1500	0,1500	0,0041	0,0009
c(2,0)	0,2500	0,0241	0,2500	0,0081	0,0041
c(3,0)	0,1500	0,0117	0,1500	0,0041	0,0013
c(0,2)	-	0,2213	0,2144	-	-
c(1,2)	-	0,2401	0,2377	-	-
c(2,2)	-	0,2044	0,2188	-	-
c(3,2)	-	0,1901	0,2222	-	-
c(0,1)	-	-	0,1140	-	-
c(1,1)	-	0,1174	-	-	-
c(2,1)	-	-	0,1087	-	-
c(3,1)	-	0,1290	0,1008	-	-
c(2,3)	-	0,1044	-	-	-
c(3,3)	-	0,1099	-	-	-
c(1, 0, 0)	0,1500	0,0270	0,0302	0,0115	0,0087
c(2, 0, 0)	0,3000	0,0180	0,0088	0,0140	0,0086
c(3, 0, 0)	0,2000	0,0098	0,0105	0,0059	0,0032
c(0, 1, 0)	-	-	0,1841	-	-
c(1, 1, 0)	-	-	0,1402	-	-
c(2, 1, 0)	-	0,2004	-	-	-
c(3, 1, 0)	-	-	0,1544	-	-
c(0, 1, 1)	-	-	0,1240	-	-
c(1, 1, 1)	-	-	0,1189	-	-
c(2, 1, 1)	-	-	0,1904	-	-
c(3, 1, 1)	-	-	0,1305	-	-
c(0, 2, 0)	-	-	0,2614	-	-
c(1, 2, 0)	-	-	0,2430	-	-
c(3, 2, 0)	-	-	0,2331	-	-
c(0, 2, 1)	-	0,0915	-	-	-
c(1, 2, 1)	-	0,0881	-	-	-
c(2, 2, 1)	-	0,1201	-	-	-
c(3, 2, 1)	-	0,1044	-	-	-
c(1, 0, 1)	-	0,1340	-	-	-
c(2, 0, 1)	-	0,1105	-	-	-
c(3, 0, 1)	-	0,1088	-	-	-
c(1, 1, 3)	-	0,1102	-	-	-
c(2, 1, 3)	-	0,0981	-	-	-
c(3, 1, 3)	-	0,1503	-	-	-
c(0, 0, 3)	-	0,1108	-	-	-
c(1, 0, 3)	-	0,1074	-	-	-
c(2, 0, 3)	-	0,1096	-	-	-
c(3, 0, 3)	-	0,1390	-	-	-

Tabela 6.18: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(1 \mid \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(1 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1, 0, 0, 0)	0,2500	0,0811	0,0273	0,0130	0,0102
c(2, 0, 0, 0)	0,3500	0,0537	0,0121	0,00007	0,00004
c(3, 0, 0, 0)	0,2500	0,0401	0,0368	0,0250	0,0117
c(0, 0, 0, 0)	0,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,1409	-	-	-
c(1, 0, 0, 1)	-	0,1902	-	-	-
c(2, 0, 0, 1)	-	0,1105	-	-	-
c(3, 0, 0, 1)	-	0,1209	-	-	-
c(0, 2, 2, 0)	-	-	0,2306	-	-
c(1, 2, 2, 0)	-	-	0,1705	-	-
c(2, 2, 2, 0)	-	-	0,2036	-	-
c(3, 2, 2, 0)	-	-	0,1503	-	-
c(0, 2, 1, 0)	-	-	0,1805	-	-
c(1, 2, 1, 0)	-	-	0,2047	-	-
c(2, 2, 1, 0)	-	-	0,1639	-	-
c(3, 2, 1, 0)	-	-	0,1501	-	-

Tabela 6.19: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(2 \mid \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(2 \mid \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,2000	0,0400	0,0400	1,76e-05	1,44e-05
c(2)	0,1000	0,0100	0,0100	1,3e-04	1,96e-06
c(3)	0,1000	0,0100	3,60e-05	2,30e-05	7,84e-06
c(1,0)	0,2500	0,0625	0,0625	2,60e-05	1,00e-08
c(2,0)	0,1500	9,06e-04	0,0225	3,72e-05	4,41e-06
c(3,0)	0,1500	4,00e-04	0,0225	1,74e-04	2,40e-05
c(0,2)	-	0,0198	8,19e-03	-	-
c(1,2)	-	0,0105	8,78e-03	-	-
c(2,2)	-	0,0123	0,0122	-	-
c(3,2)	-	0,0193	0,0290	-	-
c(0,1)	-	-	0,0418	-	-
c(1,1)	-	0,0484	-	-	-
c(2,1)	-	-	0,0361	-	-
c(3,1)	-	0,0425	0,0404	-	-
c(2,3)	-	0,0123	-	-	-
c(3,3)	-	8,50e-03	-	-	-
c(1, 0, 0)	0,3000	4,24e-04	4,88e-04	6,40e-05	2,70e-05
c(2, 0, 0)	0,1500	1,69e-04	1,14e-04	3,60e-07	04,84e-06
c(3, 0, 0)	0,2000	1,08e-04	9,41e-05	7,29e-06	1,0e-08

Tabela 6.20: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(2 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(2   \text{Estado})$	n = 13500	n = 27000	n = 54000	n = 108000
c(0, 1, 0)	-	-	0,0677	-	-
c(1, 1, 0)	-	-	0,0786	-	-
c(2, 1, 0)	-	0,0502	-	-	-
c(3, 1, 0)	-	-	0,0549	-	-
c(0, 1, 1)	-	-	0,0406	-	-
c(1, 1, 1)	-	-	0,0470	-	-
c(2, 1, 1)	-	-	0,0394	-	-
c(3, 1, 1)	-	-	0,0397	-	-
c(0, 2, 0)	-	-	0,0202	-	-
c(1, 2, 0)	-	-	0,0262	-	-
c(3, 2, 0)	-	-	0,0179	-	-
c(0, 2, 1)	-	0,0431	-	-	-
c(1, 2, 1)	-	0,0455	-	-	-
c(2, 2, 1)	-	0,0389	-	-	-
c(3, 2, 1)	-	0,0484	-	-	-
c(1, 0, 1)	-	0,0419	-	-	-
c(2, 0, 1)	-	0,0444	-	-	-
c(3, 0, 1)	-	0,0437	-	-	-
c(1, 1, 3)	-	0,0110	-	-	-
c(2, 1, 3)	-	0,0120	-	-	-
c(3, 1, 3)	-	0,0151	-	-	-
c(0, 0, 3)	-	9,96e-03	-	-	-
c(1, 0, 3)	-	0,0118	-	-	-
c(2, 0, 3)	-	0,0114	-	-	-
c(3, 0, 3)	-	9,56e-03	-	-	-
c(1, 0, 0, 0)	0,3500	9,42e-04	3,31e-04	9,60e-05	4,90e-05
c(2, 0, 0, 0)	0,2500	5,20e-04	9,24e-04	6,89e-05	8,41e-06
c(3, 0, 0, 0)	0,2500	3,50e-04	3,03e-04	1,60e-05	1,21e-06
c(0, 0, 0, 0)	0,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,0417	-	-	-
c(1, 0, 0, 1)	-	0,0486	-	-	-
c(2, 0, 0, 1)	-	0,0308	-	-	-
c(3, 0, 0, 1)	-	0,0402	-	-	-
c(0, 2, 2, 0)	-	-	0,0138	-	-
c(1, 2, 2, 0)	-	-	0,0259	-	-
c(2, 2, 2, 0)	-	-	0,0198	-	-
c(3, 2, 2, 0)	-	-	0,0227	-	-
c(0, 2, 1, 0)	-	-	0,0576	-	-
c(1, 2, 1, 0)	-	-	0,0529	-	-
c(2, 2, 1, 0)	-	-	0,0418	-	-
c(3, 2, 1, 0)	-	-	0,0675	-	-

Tabela 6.21: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(3   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1)	0,1500	0,0225	0,0225	1,99e-04	1,68e-05
c(2)	0,1500	0,0225	0,0225	1,68e-05	8,10e-07
c(3)	0,1500	0,0225	8,28e-05	7,06e-05	4,10e-05
c(1,0)	0,1750	0,0306	0,0306	6,72e-05	3,24e-06
c(2,0)	0,1750	1,16e-03	0,0225	1,19e-04	4,90e-05
c(3,0)	0,2500	1,17e-04	0,0625	2,40e-05	4,41e-06
c(0,2)	-	0,0125	0,0160	-	-
c(1,2)	-	0,0175	0,0247	-	-
c(2,2)	-	0,0190	0,0220	-	-
c(3,2)	-	0,0216	8,30e-03	-	-
c(0,1)	-	-	0,0224	-	-
c(1,1)	-	0,0149	-	-	-
c(2,1)	-	-	0,0325	-	-
c(3,1)	-	0,0128	0,0291	-	-
c(2,3)	-	0,0306	-	-	-
c(3,3)	-	0,0213	-	-	-
c(1, 0, 0)	0,2000	2,96e-04	1,23e-04	1,94e-05	2,56e-06
c(2, 0, 0)	0,2000	5,43e-04	1,14e-04	1,14e-04	6,24e-05
c(3, 0, 0)	0,3500	144e-04	1,30e-04	2,21e-05	3,61e-06
c(0, 1, 0)	-	-	0,0116	-	-
c(1, 1, 0)	-	-	0,0302	-	-
c(2, 1, 0)	-	0,0171	-	-	-
c(3, 1, 0)	-	-	0,0266	-	-
c(0, 1, 1)	-	-	0,0159	-	-
c(1, 1, 1)	-	-	0,0250	-	-
c(2, 1, 1)	-	-	0,0082	-	-
c(3, 1, 1)	-	-	0,0193	-	-
c(0, 2, 0)	-	-	0,0243	-	-
c(1, 2, 0)	-	-	0,0533	-	-
c(3, 2, 0)	-	-	0,0590	-	-
c(0, 2, 1)	-	0,0193	-	-	-
c(1, 2, 1)	-	0,0209	-	-	-
c(2, 2, 1)	-	0,0199	-	-	-
c(3, 2, 1)	-	0,0303	-	-	-
c(1, 0, 1)	-	0,0199	-	-	-
c(2, 0, 1)	-	0,0153	-	-	-
c(3, 0, 1)	-	0,0316	-	-	-
c(1, 1, 3)	-	0,0870	-	-	-
c(2, 1, 3)	-	0,0548	-	-	-
c(3, 1, 3)	-	0,0351	-	-	-
c(0, 0, 3)	-	0,0167	-	-	-
c(1, 0, 3)	-	0,0284	-	-	-
c(2, 0, 3)	-	0,0201	-	-	-
c(3, 0, 3)	-	0,0152	-	-	-

Tabela 6.22: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição, simulados a partir do Algoritmo 1, considerando  $\hat{Q}_3(3 | \text{Estado})$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Estado	$Q_3(3   \text{Estado})$	$n = 13500$	$n = 27000$	$n = 54000$	$n = 108000$
c(1, 0, 0, 0)	0,3000	0,0066	0,0016	7,56e-04	3,64e-04
c(2, 0, 0, 0)	0,3000	2,08e-03	4,28e-04	1,69e-04	5,63e-05
c(3, 0, 0, 0)	0,4000	3,10e-04	6,60e-04	4,12e-04	1,49e-04
c(0, 0, 0, 0)	1,0000	0,0000	0,0000	0,0000	0,0000
c(0, 0, 0, 1)	-	0,0335	-	-	-
c(1, 0, 0, 1)	-	0,0133	-	-	-
c(2, 0, 0, 1)	-	0,0484	-	-	-
c(3, 0, 0, 1)	-	0,0521	-	-	-
c(0, 2, 2, 0)	-	-	0,0192	-	-
c(1, 2, 2, 0)	-	-	0,0635	-	-
c(2, 2, 2, 0)	-	-	0,0561	-	-
c(3, 2, 2, 0)	-	-	0,0337	-	-
c(0, 2, 1, 0)	-	-	0,0126	-	-
c(1, 2, 1, 0)	-	-	0,0166	-	-
c(2, 2, 1, 0)	-	-	0,1523	-	-
c(3, 2, 1, 0)	-	-	8,31e-03	-	-

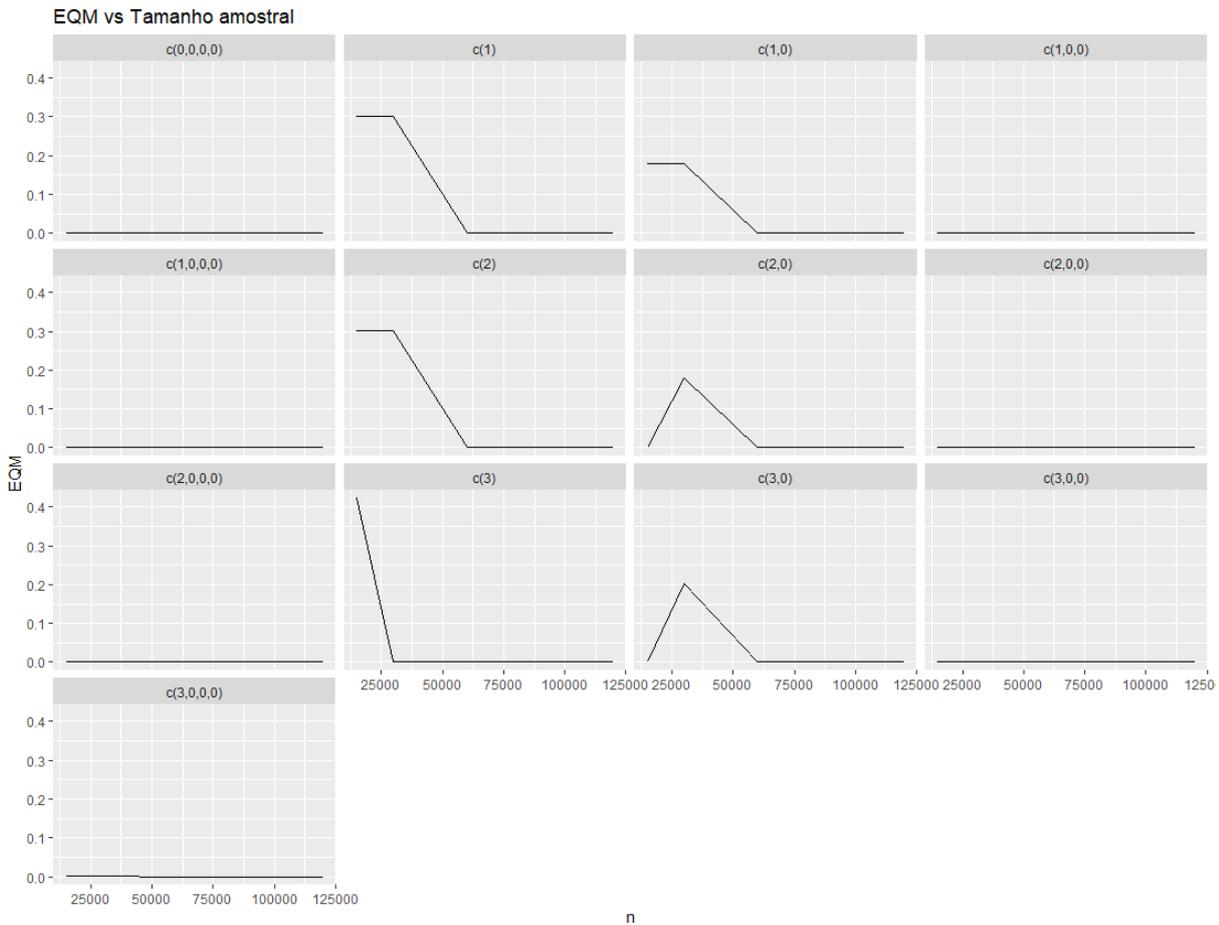


Figura 6.16: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(0 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

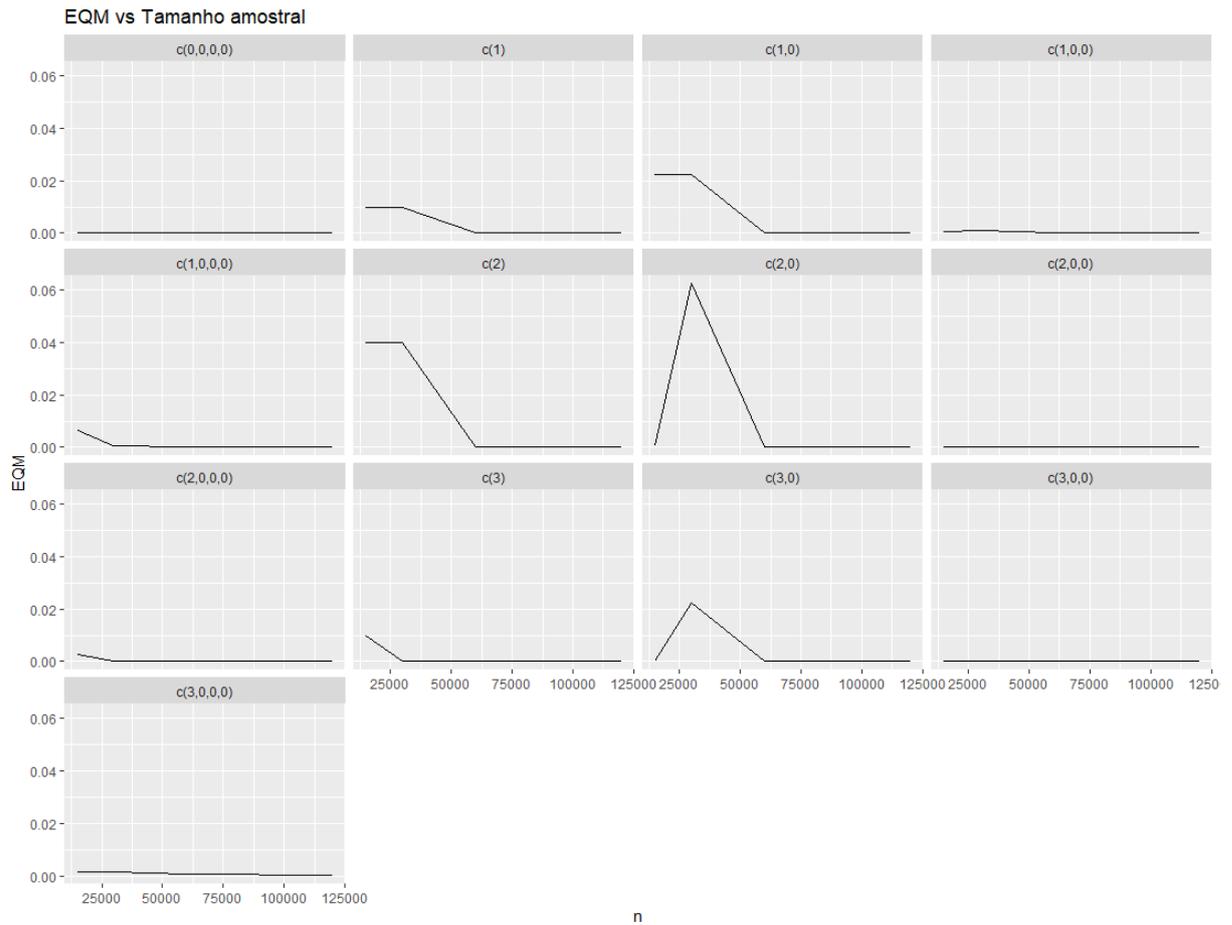


Figura 6.17: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(1 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

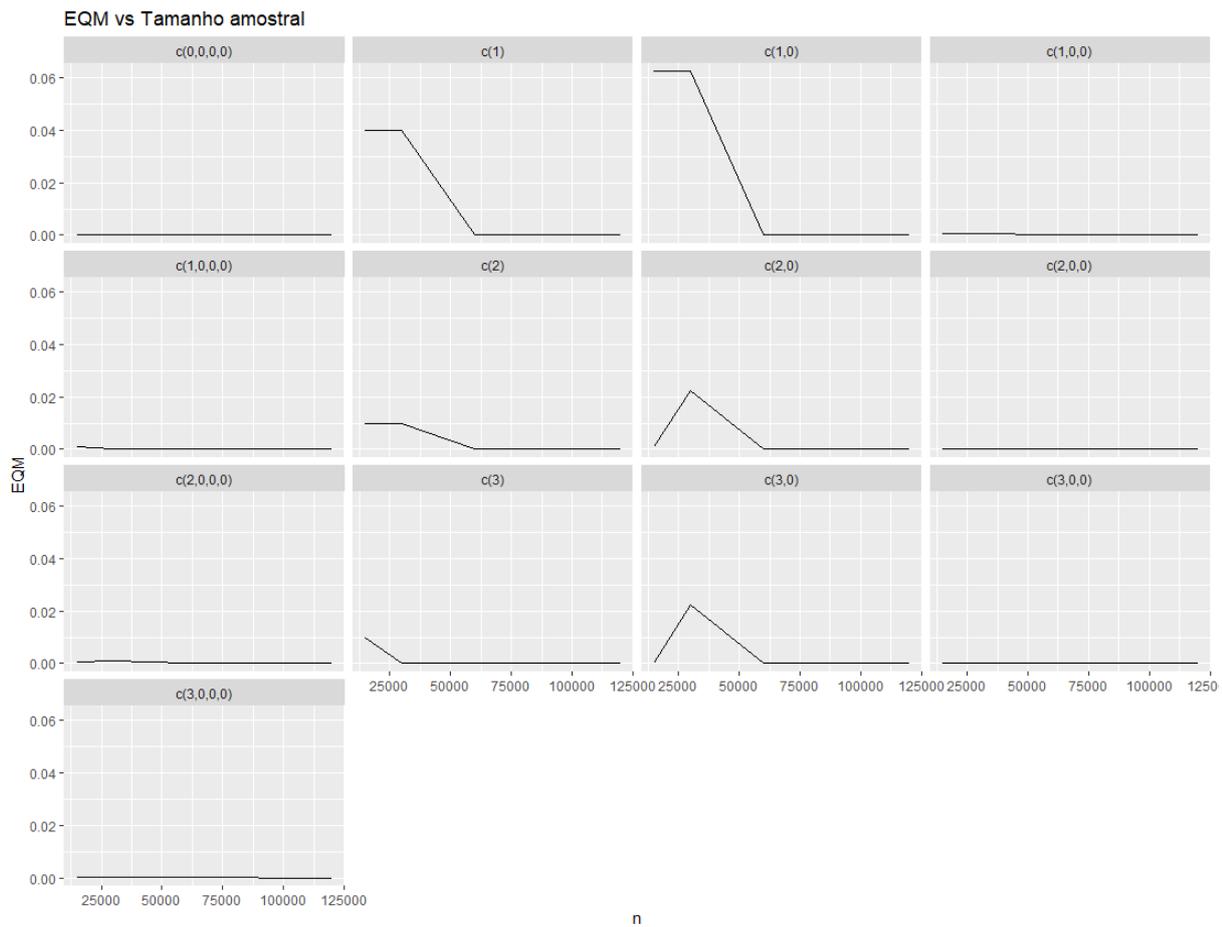


Figura 6.18: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(2 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

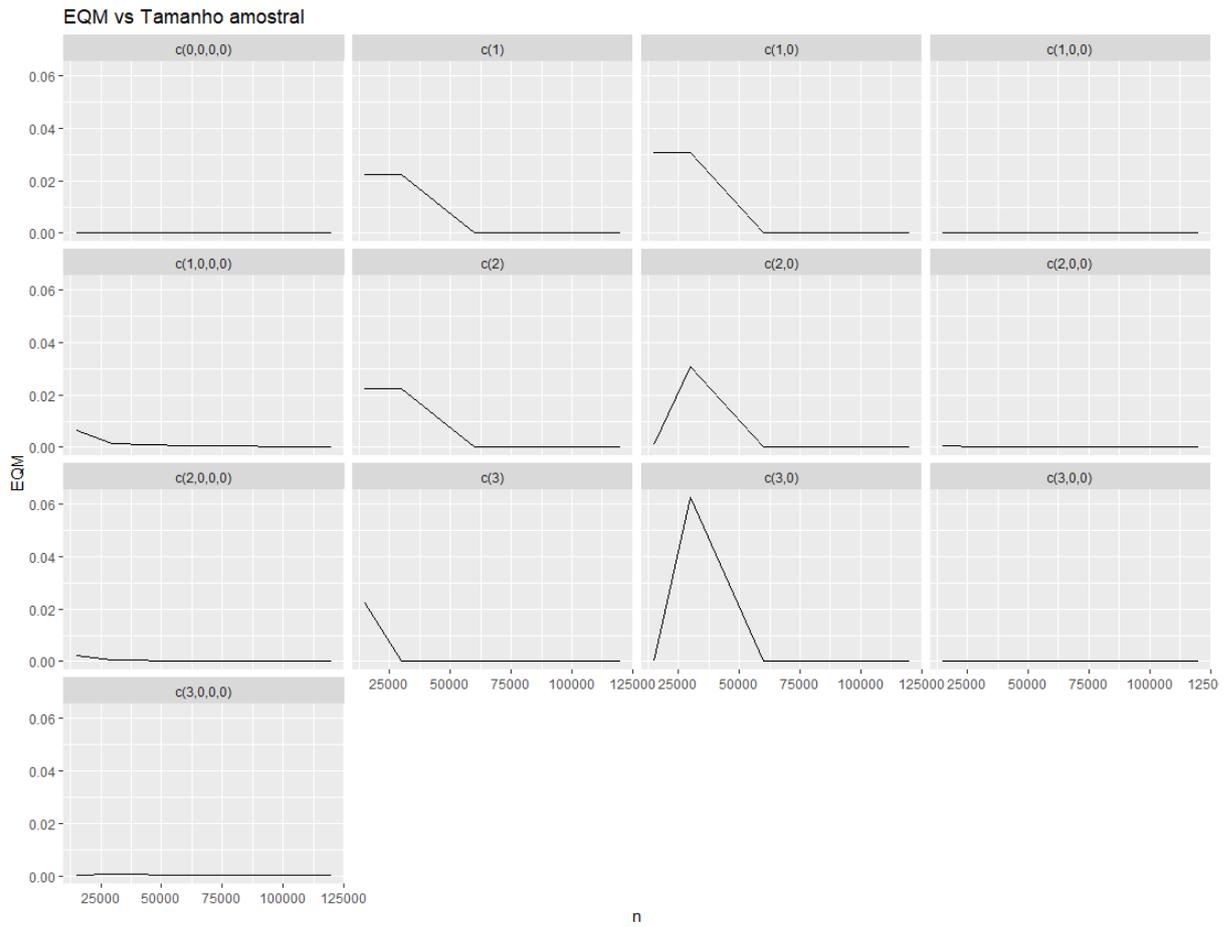


Figura 6.19: Erros quadráticos médios empíricos de 100 estimadores das probabilidade de transição  $\hat{Q}_3(2 | \text{Estado})$ , simulados a partir do Algoritmo 1, conforme o tamanho das amostras aumenta.

Ao observarmos as Tabela 6.15 até a 6.22 e as Figuras 6.16 até a 6.19, percebemos novamente que conforme aumentamos o tamanho da amostra os erros quadráticos médios empíricos tendem a diminuir, o que é um indicativo de que os estimadores são consistentes.

Além disso, notamos que o algoritmo contexto estima corretamente a árvore a partir de um tamanho de amostra de  $n = 54.000$ , diferente do que ocorre nos cenários 1 e 2, em que as árvores já são estimadas corretamente para  $n = 13.500$ . Este resultado é esperado, pois no cenário 3 temos um alfabeto de cardinalidade maior e, conseqüentemente, mais ramos a serem podados.

A Tabela 6.23 apresenta os valores dos erros absolutos médios empíricos (EAM) e dos erros quadráticos médios (EQM), obtidos a partir de 100 réplicas amostrais, para quatro tamanhos de amostras:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ . Neste contexto, apresentamos também a Figuras 6.20 que descreve o comportamento dos erros absolutos e quadráticos médios conforme os tamanhos das amostras aumentam.

Ao analisarmos a Tabela 6.23 e as Figuras 6.20, temos um indicativo de que os estimadores da entropia de transferência propostos neste trabalho são assintoticamente não-viesados e consistentes.

Por fim, é importante notar que mesmo neste cenário os estimadores propostos neste trabalho aparentam ser consistentes, mesmo que de forma mais lenta do que a observado nos cenários 1 e 2.

Tabela 6.23: Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia condicional, considerando  $\hat{Q}_3$  e quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .

Métrica	Ent. de Transferência	n = 13500	n = 27000	n = 54000	n = 108000
EAM	0,1308856	0,724410	0,3317225	0,112334	0,02177084
EQM	0,1308856	0,5247698	0,110039817	0,012618928	4,73968e-04

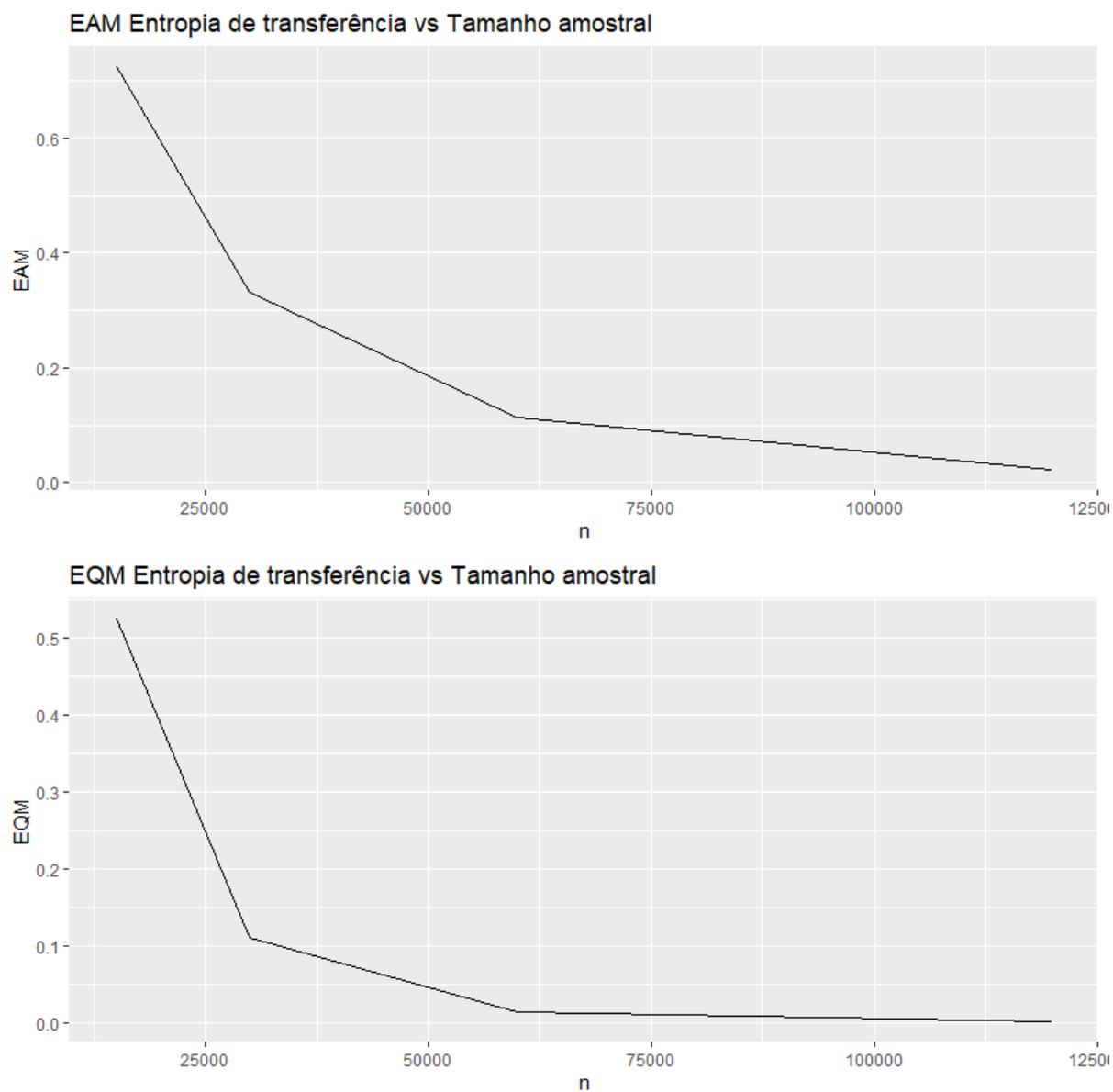


Figura 6.20: Erros absolutos e quadráticos médios empíricos de 100 estimadores da entropia de transferência, considerando quatro diferentes tamanhos de amostra:  $n = 13.500$ ,  $n = 27.000$ ,  $n = 54.000$  e  $n = 108.000$ .



# Capítulo 7

## Aplicação em dados reais

De forma a complementar o estudo de todos os capítulos anteriores, optou-se por uma aplicação em dados extraídos de neurônios reais. Neste capítulo, iremos contrapor as simulações computacionais, bem como os diferentes cenários, com séries de disparos de neurônios. Os dados utilizados para implementação dos diferentes cenários propostos foram extraídos, em forma de dados brutos, do banco de dados públicos disponibilizado pelo Buzsáki Lab (Petersen 2018). Optou-se por este tipo de aplicação, pois estes são dados eletrofisiológicos registrados com eletrodos extracelulares em ratos e camundongos desempenhando atividades livres. Registros foram feitos em diversas áreas do cérebro como hipocampo, tálamo e córtex cerebral. Abaixo justificaremos a escolha de neurônios do hipocampo para análise. Outra característica do conjunto de dados é a divisão de observações pelo ciclo sono-vigília que representam momentos em que o animal está acordado ou dormindo. Durante estes ciclos, áreas distintas como córtex e hipocampo se comunicam para consolidação de memória, ver Mizuseki & Miyawaki (2017). Assim, os disparos desses neurônios bem como a forma que se comunicam tornam esse sistema e o conjunto de dados muito interessantes para a aplicação dos estudos aqui realizados.

### 7.1 Conjunto de dados

Como apresentado anteriormente, os dados reais são públicos e estão sendo utilizados neste projeto para demonstrar a implementação das técnicas discutidas a partir da prática em situações reais. Utilizamos registros de dois neurônios do hipocampo que foram classificados como pré- e pós-sináptico seguindo a categorização feita em English et al. (2017). Note que os dados são os mesmos utilizados em English et al. (2017) : neurô-

nios piramidais e interneurônios do hipocampo. Justifica-se esta escolha, pois neurônios piramidais (excitatórios) do hipocampo são similares aos encontrados no córtex cerebral (Fishell & Hanashima 2008) que foram as escolhas de capítulos anteriores. A similaridade é encontrada tanto em morfologia quanto em número médio de disparos. Além disso, a arquitetura do hipocampo é menos sujeita a conexões recorrentes como no córtex. Assim, distinguir conexões diretas de não-diretas no hipocampo é mais simples do que no córtex.

Comparativamente ao capítulo anterior, e especificamente ao cenário 3, o par de neurônios estudados devem estar conectados. Experimentalmente, mesmo utilizando dados do hipocampo, identificar neurônios conectados não é uma tarefa simples. O experimento que identificou o par de neurônios utilizados aqui o faz por meio de estimulação externa por optogenética. Em optogenética, um vírus é adicionado às células que começam a codificar um canal iônico sensível à luz. Uma fibra ótica é então introduzida no cérebro do animal e pulsos luminosos são enviados ao neurônio assim os excitando. Se ao excitar um determinado neurônio ocorre a verificação do aumento no disparo de um vizinho, há uma forte evidência de conexão. Mesmo assim, em termos gerais, a excitação pode ser indireta o que traria falsos positivos para os resultados do experimento. Na prática, portanto, a excitação é forçada a níveis mais altos. Assim, é razoável que o conjunto de dados possua um neurônio pré com número de disparos médio menor do que o pós que estará disparando em níveis muito maiores. A seguir, descrevemos em mais detalhes especificações sobre os dois neurônios:

- **Dados do neurônio pré-sináptico:** Neste conjunto de dados, encontra-se o registro eletrofisiológico que contém os tempos de disparos (em segundos) de um neurônio do hipocampo de camundongo. Este neurônio é o primeiro neurônio do trem de disparos observado, ou seja, ele é o responsável por excitar o neurônio seguinte a partir de seu disparo. Por conta disso, possui uma quantidade menor de disparos. Seus dados tem as seguintes características:

- Número de observações: 52.712 disparos;
- Intervalo de tempo observado: De 0,00 segundos até 15.632,00 segundos (4,34 horas);
- Tempo médio para um disparo: 0,29 segundos;
- Notação:  $N_{pre}$ .

- **Dados do neurônio pós-sináptico:** Neste conjunto de dados, encontra-se o registro eletrofisiológicos do tempo de disparo (em segundos) de um neurônio do hipocampo de camundongo. Este neurônio é o neurônio seguinte do trem de disparo observado, ou seja, ele é excitado pelo neurônio anterior a partir de seu disparo. Por conta disso, possui uma quantidade maior de disparos. Seus dados tem as seguintes características:

- Número de Observações: 449.463 disparos;
- Intervalo de tempo observado: De 0,00 segundos até 15.632,00 segundos (4,34 horas);
- Tempo médio para um disparo: 0,03 segundos;
- Notação:  $N_{pos}$ .

Outra característica notável é que durante este experimento não há meios de identificar a natureza do neurônio pós-sináptico. Acredita-se que o neurônio pós-sináptico seja um interneurônio que naturalmente dispara mais que o piramidal pré-sináptico. O pré-sináptico é facilmente detectado, pois é o único que reage com o vírus, porém o neurônio pós-sináptico somente é escolhido porque disparou mais ao ser induzido pelo pré mesmo que o tempo de registro tenha sido igual.

## 7.2 Discretização

Os dados coletados já vem discretizados devido ao fato de que o aparelho de aquisição possui um tempo de amostragem, o que é razoável na transformação analógica-digital. Como resultado, eles não são exatamente contínuos. Mesmo assim, o período de monitoramento da atividade neural atribuída nessa discretização é pequeno, o qual iremos denotar por  $T$ , e a sequência de disparos observadas  $\{t_0, t_1, \dots, t_n\}$ , em que  $0 \leq t_0 < t_1 < \dots < t_n \leq T$ . Portanto, para que os dados experimentais tenham nível de discretização similar aos da simulação, no processo de tratamento podemos assumir uma janela de ação maior e repartir a mesma em intervalos que denominamos janelas temporais, cuja notação é  $\Delta t$  (MacKay & McCulloch 1952; Rieke 1997). Em cada janela temporal, a cadeia estocástica utilizada para representar o disparo do neurônio terá valor igual a 1, caso ocorra pelo menos um disparo no período determinado e 0 caso o contrário. Portanto, para um

$\Delta t$  suficientemente pequeno, o trem de disparo pode ser representado por uma sequência binária, de 0's e 1's, de comprimento  $T/\Delta t$ .

Com o intuito de ilustrar o conceito de janela temporal, trazemos a Figura 7.1 com a divisão de uma janela de 0,03 segundos. Note que este exemplo é propositalmente simplificado somente para demonstrar o processo de discretização. Na prática, os potenciais de ação ocorrem de forma extremamente imprevisível e em meio a muito ruído. Desta forma, um algoritmo de detecção destes picos será desenvolvido para lidar com essa questão.

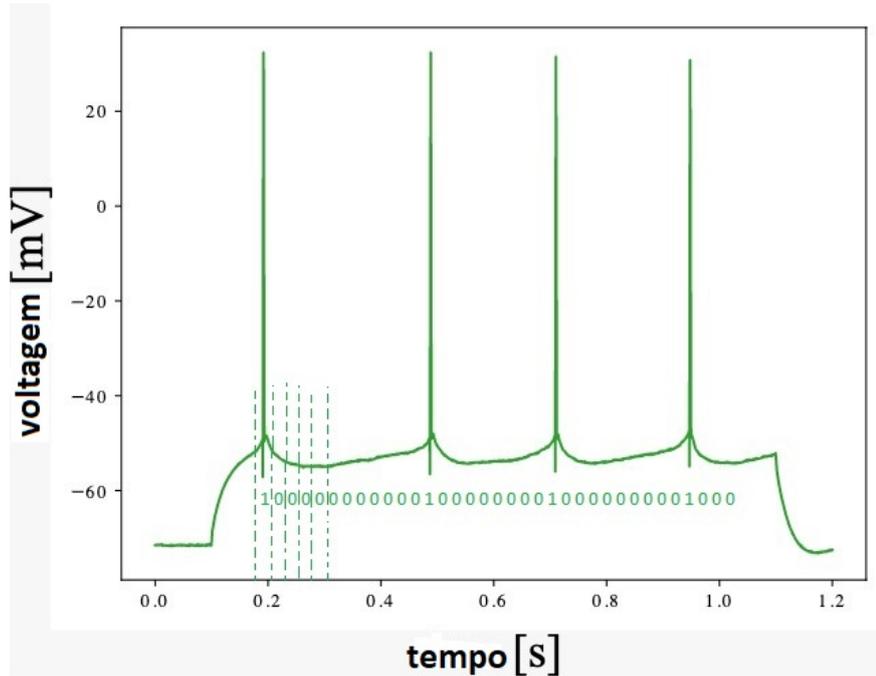


Figura 7.1: Registro eletrofisiológico de neurônio piramidal no hipocampo do rato. Trazemos o registro da membrana de um roedor por aproximadamente 1s. Discretizamos o tempo em janelas temporais de  $\Delta t = 0.03s$  (linhas tracejadas verticais). A cada janela temporal será atribuído o valor de 1 caso menos um disparo seja observado e 0 caso contrário.

Além disso, a amostra que utilizamos é de ordem finita. Apesar de possuir o conhecimento de que a dinâmica neuronal é capaz de produzir dados de forma praticamente ilimitada, é mais eficaz considerar uma árvore  $\tau$  limitada para este trabalho, o que implica em uma amostra de tamanho finito. Dessa forma, a profundidade máxima da árvore de contexto que estimamos é conhecida.

### 7.3 Resultados

Nesta seção, aplicamos o método de estimação proposto neste trabalho, em um conjunto de dados reais. Especificamente, geramos janelas temporais para os neurônios obser-

vados individualmente, sendo a janela temporal (ou bin) considerado igual à 0,29 segundos e para a observação conjunta utilizamos um bin de 0,16 segundos, que é a média entre o tempo de disparo médio do neurônio pré-sináptico (0,29) e a média do tempo de disparo do neurônio pós-sináptico (0,03). O primeiro valor escolhido de 0,29 é o tempo médio entre os disparos do neurônio pré-sináptico. A partir da observação de registros de disparos reais, estimamos as árvores de contextos probabilística usando ambos, o método da máxima verossimilhança e o algoritmo Contexto e, por fim, estimamos a entropia condicional para o neurônio pré-sináptico e o neurônio pós-sináptico, e também a entropia de transferência no neurônio pré-sináptico para o neurônio pós-sináptico. Os parâmetros utilizados são o limiar de corte  $\delta = 0.10$  e altura da árvore de contexto  $h(\tau) = 4$ .

### 7.3.1 Entropia condicional

Primeiramente, apresentamos os resultados obtidos para o banco de dados do neurônio pré-sináptico e, posteriormente, os resultados obtidos para o banco de dados do neurônio pós-sináptico. Para o banco de dados relacionado ao neurônio pré-sináptico, estimamos a seguinte árvore de contexto  $\tau_{pre}$ :

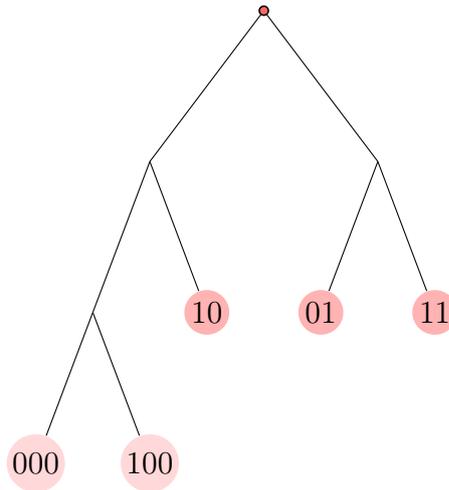


Figura 7.2: Representação pictorial da árvore  $\tau_{pre} = \{10,01,11,100,000\}$ .

Note que a árvore para o neurônio pré-sináptico possui uma leve semelhança com as observadas nos cenários 1 e 2. Apesar de começarmos com uma árvore de altura  $h(\tau_{pre}) = 4$ , os elementos do nível 4 foram todos podados. Entretanto, ao adentrar no nível 3, os elementos (0,0,0) e (1,0,0) foram mantidos, fato que se assemelha aos

primeiros cenários de simulação. Além disso, para essa árvore, obtemos a seguinte matriz de transição estimada:

$$Q_{pre} := \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{cc} 0 & 1 \\ \left( \begin{array}{cc} 0,4705236 & 0,5294764 \\ 0,5361175 & 0,4638825 \\ 0,3787067 & 0,6212933 \\ 0,3084607 & 0,3915393 \\ 0,7550837 & 0,2449163 \end{array} \right) \end{array}.$$

A partir disso, a entropia condicional empírica associada ao neurônio pré-sináptico e a blocos de tamanho  $k = 10$ , obtida a partir da árvore de contexto probabilística estimada  $(\tau_{pre}; Q_{pre})$  é

$$H_{pre}(X_k | X_0^{k-1}) = 0,1596173 \text{ bits.}$$

Portanto, condicionado a atividade histórica do neurônio pré-sináptico, a incerteza média relativa a sua atividade em um dado instate temporal é aproximadamente 0,16 bits. Quando comparamos este valor com o obtido no cenário 1 do Capítulo 6, percebemos que eles são bem próximos, isso deve-se à semelhança das árvores de contexto probabilísticas. Portanto, temos um indicativo da aderência da função de ativação proposta no cenário 1 aos dados do neurônio pré-sináptico.

Analogamente, para o banco de dados relacionado ao neurônio pós-sináptico estimamos a árvore de contexto  $\tau_{pos}$ :

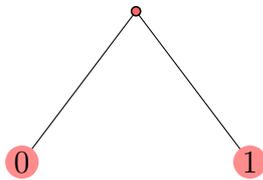


Figura 7.3: Representação pictorial da árvore  $\tau_{pos} = \{0,1\}$ .

Ao analisar a árvore apresentada na Figura 7.3, percebemos que apesar de começarmos com uma árvore de altura  $h(\tau_{pre}) = 4$  terminamos com uma árvore de altura 1. Isto ocorre, devido a alta proporção de disparos do neurônio pós-sináptico, logo, com apenas 1 passo no passado é possível descrever o comportamento transitório do neurônio. Além disso, para esta árvore, obtemos a seguinte matriz de transição:

$$Q_{pos} := \begin{matrix} & 0 & 1 \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0,17513778 & 0,8248622 \\ 0,02637401 & 0,9736260 \end{pmatrix} \end{matrix}.$$

A partir disso, a entropia condicional empírica associada ao neurônio pós-sináptico e a blocos de tamanho  $k = 10$ , obtida a partir da árvore de contexto probabilística estimada  $(\tau_{pos}; Q_{pos})$  é

$$H_{pos}(X_k | X_0^{k-1}) = 0,001335799 \text{ bits.}$$

Portanto, a incerteza média de um neurônio disparar, considerando blocos de tamanho  $k = 10$ , é aproximadamente 0,00134 bits. Note que devido ao fato da proporção de disparos alta, a incerteza média é consideravelmente menor do que a obtida pro neurônio pré-sináptico. De fato, neurônios que recebem estímulos constantes disparam com mais frequência então temos menos incerteza com relação a sua atividade.

### 7.3.2 Entropia de transferência

Como apontado previamente, o par de neurônios exposto aqui estão conectados sinap-ticamente. Experimentalmente, esta tipo verificação se dá por meio de uma indução do neurônio pós-sináptico para disparar mais, sendo que assim o mesmo possui um intervalo de silêncio menor. Para contornar tal fato, utilizamos uma janela temporal (bin) que leva em consideração o tempo médio de disparo tanto do neurônio pré-sináptico quanto do neurônio pós-sináptico. Assim, consideramos um bin de 0,16 segundos para a discretiza-ção das observações de disparos neuronais. Após a discretização, aplicamos ao conjunto de dados a transformação  $N = N_{pre} + 2N_{pos}$ , que é análoga à transformação utilizada no cenário 3. Dessa forma, a árvore de contexto estimada  $\tau$  é

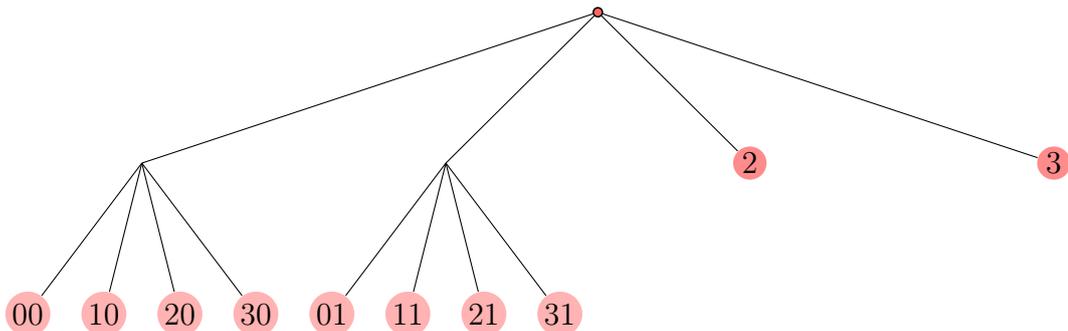


Figura 7.4: Representação pictorial da árvore  $\tau = \{2,3,00,10,20,30,01,11,21,31\}$ .

Além disso, para essa árvore obtemos a seguinte matriz de transição:

$$Q := \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 2 \\ 3 \\ 00 \\ 01 \\ 02 \\ 03 \\ 10 \\ 11 \\ 12 \\ 13 \end{matrix} & \begin{pmatrix} 0,078 & 0,005 & 0,668 & 0,248 \\ 0,042 & 0,005 & 0,506 & 0,447 \\ 0,368 & 0,013 & 0,499 & 0,121 \\ 0,240 & 0,030 & 0,560 & 0,170 \\ 0,259 & 0,0122 & 0,585 & 0,143 \\ 0,195 & 0,027 & 0,509 & 0,269 \\ 0,260 & 0,008 & 0,496 & 0,236 \\ 0,077 & 0,000 & 0,462 & 0,461 \\ 0,139 & 0,032 & 0,545 & 0,282 \\ 0,157 & 0,013 & 0,509 & 0,320 \end{pmatrix} \end{matrix}.$$

Note que, um pouco diferente do cenário 3, o neurônio pós-sináptico (representado pelo valor 2) possui uma maior probabilidade de permanecer disparando. A partir disso, a entropia de transferência do neurônio pré-sináptico para o neurônio pós-sináptico associada a blocos de tamanho  $k = 6$ , obtida a partir da árvore de contexto probabilística estimada  $(\tau; Q)$  é

$$\begin{aligned} T_{N_{pre} \rightarrow N_{pos}} &= H(N_{pos}|N_{pos}) - H(N_{pos}|N_{pos}; N_{pre}) \\ &= 3,981538 - 0,073653 \\ &= 3,907885 \text{ bits.} \end{aligned}$$

Portanto, quando condicionado ao conhecimento da atividade histórica do neurônio pós-sináptico, o conhecimento da atividade histórica do neurônio pré-sináptico diminui em aproximadamente 99% a incerteza média relativa a atividade do neurônio pós-sináptico. Portanto, temos um indicativo de que os neurônios em questão estão conectados.

Os resultados obtidos neste capítulo são coerentes aos resultados obtidos nas simulações e também à maneira com a qual o experimento foi conduzido. Assim, o método de estimação proposto neste trabalho é também suficientemente robusto para verificar conexões em um contexto de dados reais.

# Capítulo 8

## Considerações Finais

Nesta monografia, estudamos a modelagem da atividade de células neuronais ao longo de um intervalo tempo finito via cadeias estocásticas de alcance variável e também como algumas medidas da teoria da informação podem ser utilizadas para entender o comportamento da atividade neuronal e para identificar conexões sinápticas entre neurônios. Uma vez observado o registro fisiológico de tais células, investigamos como estimar as probabilidades de transição da cadeia estocástica subjacente que rege a dinâmica do fenômeno em estudo, e como encontrar a menor árvore de contexto que melhor explica os dados observados. Neste sentido, consideramos os estimadores de máxima verossimilhança para a família de probabilidades de transição e estimamos a árvore de contexto através do algoritmo contexto. Utilizamos, então, a árvore de contexto probabilística estimada para estimar a entropia condicional e a entropia de transferência, que são as medidas da informação que utilizamos para estudar a conectividade neuronal.

Do ponto de vista neurobiológico, a metodologia proposta mostrou-se robusta o suficiente para detectar conexões neuronais em um contexto de dados reais. Como os registros fisiológicos desse conjunto de dados são realizados dentro do tecido de um organismo vivo, a inferência da conectividade neuronal se torna mais difícil, pois verifica-se um alto grau de ruído. O bom desempenho da metodologia para dados de um experimento realizado *in vivo* é uma contribuição importante desta monografia, uma vez que muitos dos modelos teóricos e/ou computacionais da literatura são baseados em experimentos conduzidos em ambientes controlados e fechado de laboratório, fora do organismo vivo.

No que diz respeito à aplicação computacional, cujo o código está disponível em <https://github.com/Victor-girardi/Comandos-da-Teoria-da-Informa-o>, os métodos de estimação propostos tiveram uma boa performance nos três cenários considerados

neste trabalho. Além de ilustrarem os resultados teóricos, já bem difundidos na literatura, a respeito da consistência dos estimadores de máxima verossimilhança das probabilidade de transição e da entropia condicional empírica, nosso estudo de simulação trás evidências a respeito da consistência da entropia de transferência empírica. Até onde vai nosso conhecimento, não há na literatura trabalho algum que trata da consistência da entropia de transferência empírica neste contexto de cadeias de alcance variável. Portanto, tais evidências empíricas complementam, em um certo sentido, os estudos sobre a estimação da entropia de transferência.

O próximo passo natural deste trabalho é mostrar matematicamente a consistência da entropia de transferência empírica para este ambiente mais geral de cadeias de alcance variável bivariada. Um outro estudo interessante seria considerar outras funções de ativação e verificar como elas afetam a performance da metodologia de estimação proposta. Em particular, julgamos interessante que tais funções de ativação sejam propostas através de observações empíricas, pois, desta forma, elas se tornam mais factíveis do ponto de vista neurobiológico. Diante do que expomos nesta monografia, acreditamos que ambos os trabalhos futuros mencionados são altamente exequíveis. Portanto, esperamos que, no futuro, tenhamos a oportunidade de conduzir tais estudos ou vê-los publicados.

# Glossário

**Canal AMPA:** é o canal de cátions com receptores de alfa-amino-3-hidroxi-metil-5-4-isoxazolpropiónico.

**Capacitor:** conjunto de dois ou mais condutores isolados entre si por meio de dielétricos e que tem como função armazenar carga e energia elétrica no campo eletrostático que se estabelece entre os condutores.

**Cognição humana:** é uma função psicológica atuante na aquisição do conhecimento e se dá através de alguns processos, como a percepção, a atenção, associação, memória, raciocínio, juízo, imaginação, pensamento e linguagem.

**Condutância:** dimensão da medida que caracteriza a capacidade de um material condutor, numa corrente contínua, para ocasionar a eletricidade e se equivale ao inverso da resistência.

**Córtex cerebral:** é formado pela substância cinzenta (que contém o corpo celular do neurônio), e também é o local do processamento neural mais sofisticado e distinto. É uma fina camada, a mais externa do cérebro dos vertebrados, que possui uma espessura que pode variar de 1 a 4 mm.

**Flutuação:** variação dos seres vivos que se manifesta com um valor que oscila, de um para outro indivíduo, à volta de um valor médio.

**Canal  $GABA_A$ :** O receptor  $GABA_A$  é um dos dois canais iônicos ativado por ligante, responsável por mediar os efeitos do ácido gama-aminobutírico.

**Canal  $GABA_B$ :** são receptores, ativados pelo ácido gama-aminobutírico (GABA), ligados a uma proteína G em canais de potássio.

**Hipocampo:** é uma estrutura localizada nos lobos temporais do cérebro humano, considerada a principal sede da memória e importante componente do sistema límbico. Além disso é relacionado com a navegação espacial.

***In vitro*:** é uma expressão latina que designa todos os processos biológicos que têm lugar fora dos sistemas vivos, no ambiente controlado e fechado de um laboratório e que

são feitos normalmente em recipientes de vidro.

***In vivo:*** é uma expressão latina que refere-se à experimentos realizados dentro ou no tecido de organismos vivos.

**Neurobiologia:** é o estudo das células do sistema nervoso e da organização dessas células dentro de circuitos funcionais que processam a informação e medeiam o comportamento.

**Neurônio:** também chamados de células nervosas, são células do sistema nervoso que estão relacionadas com a propagação do impulso nervoso, sendo consideradas as unidades básicas desse sistema.

**Canal NMDA:** é o canal de cátions com receptores de N-metil-D-aspartato.

**Resistor:** dispositivo dotado de resistência, usado em circuitos elétricos para proteção, operação ou controle do circuito.

**Ruído:** designação genérica de todos os defeitos que perturbam ou impedem uma transmissão.

**Tálamo:** é uma estrutura localizada no diencéfalo, entre o córtex cerebral e o mesencéfalo, formada fundamentalmente por substância cinzenta (núcleos de neurônios) do encéfalo. São duas massas neuronais situadas na profundidade dos hemisférios cerebrais.

# Referências Bibliográficas

- Anderson, P. G. (1987), ‘Webster’s medical desk dictionary’, JAMA **258**(9), 1244–1244.
- Bota, M., Dong, H.-W. & Swanson, L. W. (2005), ‘Brain architecture management system’, Neuroinformatics **3**(1), 15–47.
- Brémaud, P. (2013), ‘Markov chains: Gibbs fields, monte carlo simulation, and queues’, Springer Science & Business Media **31**.
- Brodmann, K. (1910), Feinere anatomie des grosshirns, in ‘Handbuch der Neurologie’, Springer, pp. 206–307.
- Buesing, L., Bill, J., Nessler, B. & Maass, W. (2011), ‘Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons’, Public Library of Science **7**(11), e1002211.
- Bühlmann, P., Wyner, A. J. et al. (1999), ‘Variable length markov chains’, The Annals of Statistics **27**(2), 480–513.
- Cai, Z., Neveu, C. L., Baxter, D. A., Byrne, J. H. & Aazhang, B. (2017), ‘Inferring neuronal network functional connectivity with directed information’, Journal of Neurophysiology **118**(2), 1055–1069.
- Cessac, B. (2008), ‘A discrete time neural network model with spiking neurons’, Journal of Mathematical Biology **56**(3), 311–345.
- Cessac, B. (2011a), ‘A discrete time neural network model with spiking neurons: Ii: Dynamics with noise’, Journal of Mathematical Biology **62**(6), 863–900.
- Cessac, B. (2011b), ‘Statistics of spike trains in conductance-based neural networks: Rigorous results’, The journal of Mathematical Neuroscience **1**(1), 8.

- Cessac, B. & Cofre, R. (2013), ‘Spike train statistics and gibbs distributions’, Journal of Physiology-Paris **107**(5), 360–368.
- Cessac, B., Rostro, H., Vasquez, J.-C. & Viéville, T. (2009), ‘How gibbs distributions may naturally arise from synaptic adaptation mechanisms. a model-based argumentation’, Journal of Statistical Physics **136**(3), 565–602.
- Chazottes, J. & Gabrielli, D. (2005), ‘Large deviations for empirical entropies of g-measures’, Nonlinearity **18**(6), 2545.
- Chazottes, J. & Maldonado, C. (2011), ‘Concentration bounds for entropy estimation of one-dimensional gibbs measures’, Nonlinearity **24**(8), 2371.
- Cover, T. M. & Thomas, J. A. (1991), ‘Information theory and statistics’, Elements of Information Theory **1**(1), 279–335.
- de Abril, I. M., Yoshimoto, J. & Doya, K. (2018), ‘Connectivity inference from neural recording data: Challenges, mathematical bases and research directions’, Neural Networks **102**, 120–137.
- Dieudonné, J. (1969), ‘Foundations of modern analysis’, Read Books Ltd .
- Duarte, A. & Ost, G. (2014), ‘A model for neural activity in the absence of external stimuli’, arXiv preprint arXiv:1410.6086 .
- Duarte, A., Ost, G. & Rodríguez, A. A. (2015), ‘Hydrodynamic limit for spatially structured interacting neurons’, Journal of Statistical Physics **161**(5), 1163–1202.
- English, D. F., McKenzie, S., Evans, T., Kim, K., Yoon, E. & Buzsáki, G. (2017), ‘Pyramidal cell-interneuron circuit architecture and dynamics in hippocampal networks’, Neuron **96**(2), 505–520.
- Faisal, A. A., Selen, L. P. & Wolpert, D. M. (2008), ‘Noise in the nervous system’, Nature Reviews Neuroscience **9**(4), 292–303.
- Ferrari, P. A. & Galves, A. (2000), ‘Construction of stochastic processes, coupling and regeneration’, XIII Escuela Venezolana de Matemática .
- Ferreira, R. F. (2021), ‘Concentration bounds for plug-in estimator of directed information’, IEEE Transactions on Information Theory .

- Ferreira, R. F., Gallo, S. & Paccaut, F. (2020), ‘Non-regular g-measures and variable length memory chains’, Nonlinearity **33**(11), 6026.
- Fishell, G. & Hanashima, C. (2008), ‘Pyramidal neurons grow up and change their mind’, Neuron **57**(3), 333–338.
- Foik, A. T., Popiołkiewicz, J., Żeber, K., Urban, P., Bekisz, M., Wróbel, A. & Waleszczyk, W. J. (2014), ‘Analysis of neural networks in subcortical visual structures using correlation methods’, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2014 **9290**, 92901W.
- Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010), ‘Action and behavior: a free-energy formulation’, Biological cybernetics **102**(3), 227–260.
- Gabrielli, D., Galves, A. & Guiol, D. (2003), ‘Fluctuations of the empirical entropies of a chain of infinite order’, arXiv preprint cond-mat/0308508 .
- Galves, A. & Leonardi, F. (2008), ‘Exponential inequalities for empirical unbounded context trees’, In and Out of Equilibrium 2 pp. 257–269.
- Galves, A. & Löcherbach, E. (2008), ‘Stochastic chains with memory of variable length’, arXiv preprint arXiv:0804.2050 .
- Galves, A. & Löcherbach, E. (2013), ‘Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets’, Journal of Statistical Physics **151**(5), 896–921.
- Galves, A. & Löcherbach, E. (2015), ‘Modeling networks of spiking neurons as interacting processes with memory of variable length’, arXiv preprint arXiv:1502.06446 .
- Gerstner, W., Kistler, W. M., Naud, R. & Paninski, L. (2014), ‘Neuronal dynamics: From single neurons to networks and models of cognition’, Cambridge University Press .
- González-Montoro, A., Cao, R., Espinosa, N., Marino, J. & Cudeiro, J. (2011), ‘Autocorrelation measures and independence tests in spike trains’, Modern Mathematical Tools and Techniques in Capturing Complexity pp. 471–484.
- Gripp, M. (2019), ‘Inferência em cadeias de alcance variável’, Trabalho de Conclusão de Curso. Departamento de Estatística, Universidade Federal Fluminense - RJ .

- Haruna, T. & Nakajima, K. (2013), ‘Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic markov processes’, The European Physical Journal B **86**(5), 230.
- Herculano-Houzel, S. (2009), ‘The human brain in numbers: a linearly scaled-up primate brain’, Frontiers in Human Neuroscience **3**, 31.
- Herrmann, A. & Gerstner, W. (2001), ‘Noise and the psth response to current transients: I. general theory and application to the integrate-and-fire neuron’, Journal of Computational Neuroscience **11**(2), 135–151.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., of Biochemistry, D., Jessell, M. B. T., Siegelbaum, S. & Hudspeth, A. (2000), ‘Principles of neural science’, McGraw-hill New York **4**.
- Karnup, S. V. & Zhadin, M. N. (1983), ‘Autocorrelation analysis of spontaneous discharges of cortical neurons during learning.’, Neuroscience and Behavioral Physiology .
- Linderman, S. W., Miller, A. C., Adams, R. P., Blei, D. M., Paninski, L. & Johnson, M. J. (2016), ‘Recurrent switching linear dynamical systems’, arXiv preprint arXiv:1610.08466 .
- London, M., Shreiberman, A. & Segev, I. (2002), ‘Estimating information theoretic quantities of spike-trains using the context tree weighting algorithm’, Nature Neuroscience **5**.
- MacKay, D. M. & McCulloch, W. S. (1952), ‘The limiting information capacity of a neuronal link’, The Bulletin of Mathematical Biophysics **14**(2), 127–135.
- Mizuseki, K. & Miyawaki, H. (2017), ‘Hippocampal information processing across sleep/wake cycles’, Neuroscience research **118**, 30–47.
- Mood, A. M.; Graybill, F. A. e. B. D. C. (1974), ‘Introduction to the theory of statistics’, MacGraw-Hill .
- Ornstein, D. S. & Weiss, B. (1990), ‘How sampling reveals a process’, The Annals of Probability pp. 905–930.

- Petersen, P. C. . G. (2018), ‘Introduction to stochastic processes’, Public access to electrophysiological datasets collected in the Buzsaki Lab. Zenodo. <http://doi.org/10.5281/zenodo.3629881> .
- Port, P. H. S. & Stone, C. J. (1972), ‘Introduction to stochastic processes’, Boston, MA: HoughtonMifflin .
- R. Jolivet, A. Rauch, H. L. & Gerstne, W. (2006), ‘J. comput. neurosci’, pp. 35–49.
- Rieke, F.;Ruyter van Steveninck, R. (1997), ‘Spikes: exploring the neural code’, MIT Press .
- Rissanen, J. (1983), ‘A universal data compression system’, IEEE Transactions on Information Theory **29**(5), 656–664.
- Roque, P. A. C. (n.d.), ‘Notas de aula neurociência computacional’, <https://docplayer.com.br/34095855-Introducao-a-neurociencia-computacional-graduacao-prof-antonio-roque-aula-6.html> .
- Shannon, C. E. (1948), ‘A mathematical theory of communication’, The Bell System Technical Journal **27**(3), 379–423.
- Shields, P. C. (1996), ‘The ergodic theory of discrete sample paths’, American Mathematical Soc. **13**.
- Shreider, I. A., B. N. (1966), ‘The monte carlo method of statistical trials’, Pergamon Press. **1**.
- Softky, W. R. & Koch, C. (1993), ‘The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps’, Journal of neuroscience **13**(1), 334–350.
- Song, D., W. . B. (2013), ‘Identification of sparse neural functional connectivity using penalized likelihood estimation and basis functions’, Journal of computational neuroscience **35**(3), 335–357.
- Truccolo, W., Hochberg, L. R. & Donoghue, J. P. (2010), ‘Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes’, Nature Neuroscience **13**(1), 105–111.
- Viana, P. R. L. (n.d.), ‘Notas de aula eletrodinâmica clássica’, , C. Reinwald.