

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Cadeias estocásticas de memória ilimitada com aplicação na neurociência

Ricardo Felipe Ferreira

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Ricardo Felipe Ferreira

Cadeias estocásticas de memória ilimitada com aplicação na neurociência

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

Área de Concentração: Estatística

Orientador: Prof. Dr. Alessandro Giacomo Grimbert Gallo

USP – São Carlos
Abril de 2019

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

F383c Ferreira, Ricardo Felipe
 Cadeias estocásticas de memória ilimitada com
aplicação na neurociência / Ricardo Felipe Ferreira;
orientador Alessandro Giacomo Grimbert Gallo. --
São Carlos, 2019.
 80 p.

 Tese (Doutorado - Programa Interinstitucional de
Pós-graduação em Estatística) -- Instituto de Ciências
Matemáticas e de Computação, Universidade de São
Paulo, 2019.

 1. Cadeias estocásticas de memória ilimitada. 2.
Medida invariante. 3. Inferência em processos
estocásticos. 4. Limites da concentração . 5.
Neurociência. I. Gallo, Alessandro Giacomo
Grimbert, orient. II. Título.

Ricardo Felipe Ferreira

Stochastic chains with unbounded memory applied in
neuroscience

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

Concentration Area: Statistics

Advisor: Prof. Dr. Alessandro Giacomo Grimbert Gallo

USP – São Carlos
April 2019

*À minha mãe,
por tudo.*

AGRADECIMENTOS

A conclusão de qualquer trabalho científico envolve a participação de pessoas que, com seu tempo, conhecimento, talento ou paciência me ajudaram a elaborar um trabalho muito melhor do que eu poderia ter feito sozinho.

Em primeiro lugar, agradeço ao meu orientador Sandro Gallo por aceitar dividir comigo um pouco de sua sabedoria, pelas inúmeras horas de discussões elucidadoras, companheirismo e incentivo constante. De fato, é um exímio orientador e pesquisador e, definitivamente, uma inspiração.

Por comentar alguns manuscritos intermediários e, com toda paciência que lhe é inerente, explicar conceitos chaves para o desenvolvimento da teoria aqui apresentada, sinto-me em débito com Frédéric Paccaut, o qual, com toda sua boa vontade, viabilizou minha estadia na *Université de Picardie Jules Verne* durante uma visita técnica. Sou grato também a Daniel Takahashi por apresentar-me à Teoria da Informação e despertar-me um interesse por essa área. A ele ainda agradeço os conselhos sobre o uso da Teoria da Informação na Neurociência.

Meus sinceros agradecimentos a Rodrigo Pena e Guilherme Ost que disponibilizaram notas, algoritmos e trabalhos com respeito ao funcionamento da atividade cerebral e sua modelagem matemática. Suas contribuições à pesquisa aqui realizada foram decisivas e estão presentes em várias páginas que compõem esta tese.

Agradeço a todos do Departamento de Estatística da Universidade Federal de São Carlos, principalmente Daiane Zuanetti por incentivar-me a seguir na vida acadêmica e com quem sempre nutri um profícuo diálogo que estimulou constantes revisões na minha pesquisa e na minha maneira de fazer ciência. Reconheço toda sua ajuda com a leitura de mais de um racunho e a paciência em me ouvir despejando centenas de detalhes e preocupações teóricas. Não me esqueço de Daniela Marinello e Maria Isabel Araujo que sempre estiveram disponíveis para resolver meus muitos problemas administrativos e de computador. Destaco também a inestimável ajuda de Vera Tomazella, antiga coordenadora da pós-graduação, que dentro de suas possibilidades sempre buscou proporcionar um ambiente agradável de produtividade e estimulação intelectual.

Aos bons amigos que fizeram da minha vida mais feliz nesse período, pela presença e por me lembrarem de que nem só de matemática vive uma alma. Infelizmente, citá-los individualmente demandaria uma memória formidável, então, para poupar-me do risco do esquecimento, opto por agradecer-los sobre os alicerces da generalidade. Todavia, estaria sendo injusto, se não destacasse a presença fundamental de Fellipe Bruno durante esses dez anos de

academia. Membro da família e irmão de coração sou grato pelos debates filosóficos, políticos e culturais que ele proporciona durante nossos encontros. Além dos momentos insanos, repletos de seu humor espirituoso. A sua disposição na leitura, releitura e correções de tantos manuscritos ao longo da minha formação é de caráter indelével.

O inglês atualmente é fundamental para a divulgação científica e esse trabalho não teria tomado as devidas proporções se não fosse o ensinamento cuidadoso e paciente de Arthur Médice que, com seu profissionalismo impecável, ajudou a verificar e dar forma a mais de um trabalho oriundo desta tese.

Devo profundos agradecimentos à minha família. À minha amada mãe Helena, a qual este trabalho é dedicado, pelo amor incondicional e, sobretudo, pelo infinito estímulo aos estudos. À minha irmã Patrícia, por toda disponibilidade e paciência incomensurável durante esse período. Ao meus avós Ovídio e Dona Maria por toda inspiração que emanam. Principalmente minha avó que, embora analfabeta, sempre se comportou como uma cientista nata. Roubei tempo de todos eles para poder aperfeiçoar e rever obsessivamente este trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) processo #2016/12918 – 0, vinculado ao Centro de Pesquisa, Inovação e Difusão em Neuromatemática - CEPID Neuromat, processo #2013/07699 – 0. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP. À esses órgãos sou extremamente grato pela inestimável ajuda na viabilização deste projeto e por também tornar possível o amadurecimento desta pesquisa ao ser divulgada em congressos e encontros científicos.

“Todos começamos com o ‘realismo ingênuo’, isto é, a doutrina de que as coisas são aquilo que parecem ser. Achamos que a grama é verde, que as pedras são duras e que a neve é fria. Mas a física nos assegura que o verdejar da grama, a dureza das pedras e a frieza da neve não são o verdejar da grama, a dureza das pedras e a frieza da neve que conhecemos em nossa experiência própria, e sim algo muito diferente.”
(Bertrand Russell)

RESUMO

FERREIRA, R. F. **Cadeias estocásticas de memória ilimitada com aplicação na neurociência**. 2019. 80 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

As cadeias estocásticas de memória ilimitada são uma generalização natural das cadeias de Markov, no caso em que as probabilidades de transição podem depender de todo o passado da cadeia. Estas cadeias, introduzidas, independentemente, por Onicescu e Mihoc em 1935 e Doeblin e Fortet em 1937, vêm recebendo uma atenção crescente na literatura probabilística, não só por serem uma classe mais rica que a classe das cadeias de Markov, como por suas capacidades práticas de modelagem de dados científicos em diversas áreas, indo da biologia à linguística. Neste trabalho, as utilizamos para modelar a interação entre sequências de disparos neuronais. Nosso objetivo principal é desenvolver novos resultados matemáticos acerca das cadeias de memória ilimitada. Inicialmente, estudamos as condições que garantem a existência e a unicidade de cadeias estacionárias compatíveis com uma família de probabilidades de transição descontínua. Em seguida, tratamos do entendimento da fenomenologia dos trens de disparos neuronais e usamos da informação dirigida para modelar a informação que flui de uma sequência de disparos a outra. Nesta ocasião, fixamos limites da concentração para estimação da informação dirigida.

Palavras-chave: Cadeias Estocásticas de Memória Ilimitada, Medida Invariante, Inferência em processos estocásticos, Limites da Concentração, Neurociência.

ABSTRACT

FERREIRA, R. F. **Stochastic chains with unbounded memory applied in neuroscience**. 2019. 80 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Stochastic chains with unbounded memory are a natural generalization of Markov chains, in the sense that the transition probabilities may depend on the whole past. These process, introduced independently by Onicescu and Mihoc in 1935 and Doeblin and Fortet in 1937, have been receiving increasing attention in the probabilistic literature, because they form a class richer than the Markov chains and have practical capabilities modelling of scientific data in several areas, from biology to linguistics. In this work, we use them to model interactions between spike trains. Our main goal is to develop new mathematical results about stochastic chains with unbounded memory. First, we study conditions that guarantee the existence and uniqueness of stationary chains compatible with a discontinuous family of transition probabilities. Then, we address the understanding of the phenomenology of spike trains and we propose to use directed information to quantify the information flow from one neuron to another. In this occasion, we fix concentration bounds for directed information estimation.

Keywords: Stochastic chains with unbounded memory, Invariant Measure, Inference in stochastic processes, Concentration bounds, Neuroscience.

LISTA DE ILUSTRAÇÕES

Figura 1 – Árvore completa associada ao alfabeto $A = \{0, 1, 2\}$	28
Figura 2 – Árvore esparsa.	29
Figura 3 – Terceiro nível da árvore esparsa.	41
Figura 4 – Árvore de contexto correspondente ao Exemplo 17 com o nível 3 delimitado.	41
Figura 5 – Propagação dos disparos ao longo da estrutura de um neurônio (MOYA, 2011).	52
Figura 6 – Atividade neuronal sobre um tempo discretizado com $\Delta t = 25ms$	53

SUMÁRIO

1	INTRODUÇÃO	19
2	PRELIMINARES	23
2.1	Cadeias estacionárias	23
2.2	Cadeias estacionárias de memória ilimitada	25
2.3	Cadeias estacionárias de memória de alcance variável	27
3	A ESTACIONARIEDADE EM CADEIAS DE MEMÓRIA ILIMITADA E DE ALCANCE VARIÁVEL	31
3.1	Existência	31
3.2	Unicidade	39
3.3	Demonstrações	42
4	LIMITES DA CONCENTRAÇÃO PARA INFORMAÇÃO DIRIGIDA EMPÍRICA	51
4.1	Motivação	51
4.2	Medidas baseadas na Entropia de Shannon	54
4.3	Estimadores Empíricos	56
4.4	Propriedades de Concentração do Estimador	57
4.5	Demonstrações	59
5	CONSIDERAÇÕES FINAIS	75
	REFERÊNCIAS	77

INTRODUÇÃO

As cadeias estocásticas de memória ilimitada são generalizações das cadeias de Markov, de forma que a dependência sobre o passado pode ser ilimitada. Em outras palavras, são processos estocásticos indexados por \mathbb{Z} assumindo valores em um conjunto A (alfabeto), construídos a partir de um conjunto de probabilidades de transição, o qual define a probabilidade condicional de um símbolo em qualquer tempo, dado o passado. Sobre este passado é, ainda, fixada uma medida de probabilidade, isto é, uma condição inicial que permite a construção futura do processo.

Na literatura de processos estocásticos, o primeiro estudo sistemático sobre tais objetos data de 1935 e deve-se a Onicescu e Mihoc ([ONICESCU; MIHOC, 1935](#)). Mais tarde, em 1937, aparece no trabalho de Doeblin e Fortet ([DOEBLIN; FORTET, 1937](#)). Em ambos os trabalhos, os autores utilizaram o termo *chaînes à liaisons complètes* (cadeias com conexão completa) para denominar esta nova classe de processos estocásticos. Ainda aparece na literatura sobre o nome de “cadeias de ordem infinita”, nomenclatura proposta por [Harris \(1955\)](#).

Em Teoria Ergódica, [Keane \(1972\)](#) introduziu a denominação g -função para referir-se ao conjunto de probabilidades de transição com dependência sobre um passado infinito e g -medida para referir-se à medida invariante por translação cujas probabilidades condicionais são especificadas por uma dada g -função. Muito embora, as g -funções sejam, geralmente, chamadas de núcleos de probabilidade na teoria dos processos estocásticos; preferimos optar, neste trabalho, pela terminologia da Teoria Ergódica.

Dada uma g -função g , uma importante questão diz respeito às condições que devemos impor sobre g para que exista uma g -medida. Uma vez garantida a existência, um interesse natural são as condições sobre g que nos garantem a unicidade dessa g -medida. As condições comumente utilizadas na literatura para garantir a existência e a unicidade de uma g -medida são baseadas na continuidade de g (ver, por exemplo, [Onicescu e Mihoc \(1935\)](#), [Doeblin e Fortet \(1937\)](#), [Harris \(1955\)](#), [Ledrappier \(1974\)](#), [Johansson e Öberg \(2008\)](#), [Fernandez e Maillard \(2005\)](#)). Mesmo assim a continuidade não é necessária para existência como pode ser observado,

por exemplo, nos trabalhos de Gallo (2011) e Santis e Piccioni (2012) que a obtiveram através de uma abordagem de simulação perfeita e em Cénac *et al.* (2012) que provam a existência para exemplos específicos. No entanto, tais abordagens são restritivas, pois ou são baseadas em exemplos ou são de natureza construtiva, implicando não só a existência de uma g -medida como também a sua unicidade. Recentemente, Gallo e Paccaut (2013) garantem a existência de uma g -medida obtida a partir de condições impostas sobre o conjunto de descontinuidade de g , o qual denotamos por \mathcal{D}_g . É neste contexto que fixamos a primeira contribuição dessa tese, a qual garante a existência de uma g -medida compatível a partir da construção de uma medida de probabilidade via um passado fixado de forma que tal medida não atribui peso para os pontos de descontinuidade. Se para resolver o problema de existência foram necessárias condições sobre o conjunto de descontinuidades, para solucionar a questão da unicidade, serão necessárias condições sobre o conjunto das continuidades. Para formalizar estas condições é prático utilizar a taxa de continuidade uniforme para o caso contínuo (ver Doeblin e Fortet (1937) e Johansson e Öberg (2008)) e para o caso descontínuo as *árvores de contextos* são as que auxiliam na formalização do conceito de continuidade local (ver Gallo (2011), Gallo e Paccaut (2013) e Gallo e Garcia (2013)). Neste caso, a terminologia que utilizamos para os processos compatíveis com árvores de contexto é *cadeias estocásticas de alcance variável*. Neste sentido é que fixamos a segunda contribuição deste trabalho, a qual diz que se o tamanho dos contextos de uma árvore de probabilidade “visto” pela medida construída no resultado de existência é quase-certamente finito, então esta é uma g -medida e, além disso, se o tamanho tem esperança finita, então esta medida é única.

No que diz respeito a parte aplicativa deste projeto, os trabalhos desenvolvidos seguem a seguinte estrutura. Após uma propícia discretização de um intervalo de tempo, a atividade de um neurônio é modelada através de uma sequência de zeros e uns, em que 1 significa que o neurônio dispara – propagação de um sinal elétrico através de um fenômeno biofísico conhecido – e 0 significa que o neurônio está em repouso. Essas atividades neuronais são avaliadas nos instantes de tempo que compõem a discretização e essas sequências de zeros e uns é o que conhecemos como trens de disparos neuronais. Este fenômeno depende do grupo de neurônios que estamos estudando. No contexto das cadeias estocásticas de memória ilimitada, o processo assume valor em $\{0, 1\}^{\mathcal{I}}$, em que \mathcal{I} representa um grupo de neurônios qualquer de interesse. Portanto, a questão da modelagem neste contexto é encontrar a g -função sobre A , ou seja, as probabilidades de um subconjunto de neurônios do conjunto \mathcal{I} dispararem no instante presente dada a história dos disparos de todos os neurônios considerados. Dispomos de alguns modelos na literatura. Bruno Cessac, e seus coautores, desenvolveram um estudo sistemático sobre a modelagem de trens de disparos neuronais usando cadeias estocásticas de longo alcance (ver, por exemplo, Cessac (2008), Cessac *et al.* (2009), Cessac (2011a), Cessac (2011b) e Cessac e Cofre (2013)) e, obtiveram, a partir de considerações fisiológicas, biológicas e físicas, modelos de g -funções. Neste contexto, Galves e Löcherbach (2013) e Galves e Locherbach (2015) consideram um modelo baseado em sistemas possivelmente infinito associados a processos de interação com

memória de tamanho variável. Este modelo vem sendo amplamente utilizado na literatura por suas capacidades matemáticas, práticas e computacionais (ver, por exemplo, [Duarte e Ost \(2014\)](#) e [Duarte, Ost e Rodríguez \(2015\)](#)).

A fim de estudar a informação que flui de uma sequência de disparos neuronal a outra, usamos uma modelagem via cadeias estocásticas de memória ilimitada. [Cai et al. \(2017\)](#) utiliza a informação dirigida – uma medida baseada em entropia – para medir as influências causais entre sequências de disparos neuronais a fim de inferir a conexão sináptica. Esta ferramenta pode ser utilizada para detectar conexões funcionalmente relevantes a partir de uma métrica baseada na informação fornecida pela gravação de disparos neuronais. Dada uma amostra de duas sequências de disparos neuronais, a estimação da informação dirigida é realizada através de estimadores empíricos ([GABRIELLI; GALVES; GUIOL, 2013](#)). Um dos interesses em probabilidade e estatística é mensurar o quanto perdemos no processo de estimação. Neste sentido, propomos limites da concentração para informação dirigida empírica em torno da média populacional e da taxa de entropia, em um contexto em que as probabilidades de transição associadas ao processo bivariado que representa as sequências de disparos neuronais em estudo são dadas por uma g -função contínua com decaimento de memória exponencial. Recentemente, [Galves e Locherbach \(2015\)](#) propuseram um modelo no qual a probabilidade de observarmos um disparo neuronal em um dado instante de tempo depende de toda a evolução temporal do sistema a partir do último disparo. Este modelo, que vem sendo bem aceito pela literatura, é utilizado neste trabalho para ilustrar os resultados sobre os limites da concentração para estimação da informação dirigida.

Este trabalho está organizado da seguinte maneira. No próximo capítulo apresentamos definições e noções básicas sobre g -medidas. No capítulo 3 tratamos das questões de existência e unicidade de g -medidas compatíveis com g -funções descontínuas. No capítulo 4 apresentamos os limites da concentração para estimação da informação dirigida e discutimos sua relação com a neurociência. Por fim, no capítulo 5, apresentamos as considerações finais deste trabalho.

PRELIMINARES

Neste capítulo fixamos as notações básicas e estruturamos o ambiente matemático que utilizamos ao longo desta tese.

2.1 Cadeias estacionárias

Um processo estocástico a tempo discreto é uma sequência de variáveis aleatórias $\mathbf{X} = \{X_n : n \in \mathbb{Z}\}$ definidas sobre um espaço de probabilidade (Ω, Σ, P) com valores em um conjunto, possivelmente infinito, A , o qual denominamos *alfabeto*. Neste trabalho, os processos estocásticos assumem valores sempre em alfabetos finitos.

Denotamos por $A^{\mathbb{Z}}$ o espaço produto cujos elementos são as sequências bi-infinitas $\mathbf{a} = \{a_n : n \in \mathbb{Z}\}$ e por \mathcal{F} a σ -álgebra dos borelianos de $A^{\mathbb{Z}}$. Definimos, então, \mathcal{M} como sendo o conjunto das medidas de probabilidade definidas sobre o espaço mensurável $(A^{\mathbb{Z}}, \mathcal{F})$. Para qualquer $\mathbf{a} \in A^{\mathbb{Z}}$, a sequência $(a_m, a_{m+1}, \dots, a_n)$, em que cada $a_i \in A$, é denotada por a_m^n e o conjunto de todas as sequências finitas a_m^n é denotado por A_m^n , exceto quando $m = 1$, situação na qual preferimos escrever A^n . As sequências infinitas à esquerda (passados) são denotadas por $a_{-\infty}^n := (\dots, a_{n-1}, a_n)$, qualquer que seja $n \in \mathbb{Z}$ e o conjunto dos passados é denotado por $A^{\leq n}$.

A *distribuição conjunta de ordem k* do processo \mathbf{X} é a medida μ_k definida sobre A^k de forma que

$$\mu_k(a_1^k) = P\{X_1^k = a_1^k\}, \quad a_1^k \in A^k.$$

Se o mapa invertível $T : A^{\mathbb{Z}} \rightarrow A^{\mathbb{Z}}$ é uma *transformação de translação*, i.e., para qualquer $\mathbf{a} \in A^{\mathbb{Z}}$, $(T\mathbf{a})_n = a_{n+1}$, para qualquer $n \in \mathbb{Z}$, então a medida $\mu \in \mathcal{M}$ é dita ser *invariante por translação* se, e somente se, $\mu \circ T^{-1} = \mu$, ou seja, $\mu(T^{-1}B) = \mu(B)$ para todo $B \in \mathcal{F}$, em que

$$T^{-1}B = \{\mathbf{a} \in A^{\mathbb{Z}} : T\mathbf{a} \in B\}, \quad B \subseteq A^{\mathbb{Z}}.$$

Note que, se B é um cilindro, i.e., para $m \leq n$,

$$B = [a_m^n] := \left\{ \mathbf{x} \in A^{\mathbb{Z}} : x_i = a_i, \text{ para todo } m \leq i \leq n \right\},$$

então quando $m = 1$ e $n = k$, $\mu_k(a_1^k) = \mu([a_1^k])$, para qualquer $a_1^k \in A^k$. O processo \mathbf{X} com essa propriedade é dito ser *estacionário*. O conjunto das medidas $\mu \in \mathcal{M}$ invariantes pela translação T é denotado por \mathcal{M}_T .

Em teoria das probabilidades, um processo estocástico está bem definido quando conhecemos suas distribuições conjuntas. Portanto, o espaço sobre o qual são definidas as variáveis aleatórias X_n não tem importância, podendo, então, (Ω, Σ, P) ser escolhido da forma que convier, desde que as distribuições conjuntas se mantenham inalteradas. Logo, podemos, sem perda de generalidade, trabalhar exclusivamente sobre o espaço $(A^{\mathbb{Z}}, \mathcal{F}, \mu)$, de forma que

$$\mu(B) := P\{\mathbf{X} \in B\}, \quad \forall B \in \mathcal{F}. \quad (2.1)$$

Os processos estacionários a tempo discreto são normalmente denominados pela literatura como *cadeias estacionárias* e é essa a nomenclatura que preferimos utilizar daqui em diante. Além disso, sempre que falarmos, neste trabalho, de cadeias estacionárias estaremos nos referindo a processos estocásticos com alfabeto finito, a tempo discreto e estacionários.

Exemplo 1 (Cadeias independentes). Cadeias estocásticas com alfabeto finito cujas variáveis aleatórias são independentes e identicamente distribuídas (i.i.d.) são o exemplo mais simples de cadeias estacionárias. Note que, se $\mathbf{X} = \{X_n : n \in \mathbb{Z}\}$ é uma sequência de variáveis aleatórias i.i.d. e $\mu \in \mathcal{M}$, então, quaisquer que sejam $m \leq n$,

$$\mu\left(\left[(T^{-1}a)_m^n\right]\right) = \prod_{i=m}^n \mu([a_i]) = \mu([a_m^n]).$$

Como \mathcal{F} é gerada pelos cilindros, concluímos $\mu \in \mathcal{M}_T$ ou, equivalentemente, \mathbf{X} é estacionária.

Exemplo 2 (Cadeias de Markov). Seja $\mathbf{X} = \{X_n : n \in \mathbb{Z}\}$ uma cadeia de Markov de ordem k ($k \geq 1$) com alfabeto finito e matriz de transição Q . Se \mathbf{X} é homogênea e $\mu Q = \mu$, então

$$\begin{aligned} \mu\left(\left[(T^{-1}a)_1^k\right]\right) &= P\{X_2^{k+1} = a_1^k\} \\ &= \sum_{b \in A} P\{X_{k+2} = b\} Q(a_1^k, b) \\ &= \sum_{b \in A} \mu([b]) Q(a_1^k, b) \\ &= \mu\left(\left[a_1^k\right]\right). \end{aligned}$$

Portanto, uma cadeia de Markov de ordem k é estacionária se é homogênea e $\mu Q = \mu$.

Um dos objetivos deste trabalho é estudar a estacionariedade para o caso de cadeias cujas probabilidades de transição podem depender de todo o seu passado, i.e., cadeias sobre as quais

não há imposição sobre o limite da sua memória. Essa classe de processos estocásticos, a qual denominaremos *cadeias estocásticas de memória ilimitada*, é uma extensão natural das cadeias de Markov como veremos adiante.

2.2 Cadeias estacionárias de memória ilimitada

Para qualquer inteiro n considere $\mathcal{F}^{\leq n}$ a σ -álgebra gerada pela sequência de variáveis aleatórias $\{X_j : j \leq n\}$ com valores em A . Definimos, então, uma g -função como sendo a função $\mathcal{F}^{\leq 0}$ -mensurável $g : A^{\leq -1} \times A \rightarrow \mathbb{R}$ tal que

$$\forall a_{-\infty}^{-1} \in A^{\leq -1}, \quad \sum_{b \in A} g(a_{-\infty}^{-1}b) = \sum_{b \in A} g\left(\left(T^{-1}\mathbf{a}\right)_{-\infty}^0 b\right) = 1. \quad (2.2)$$

De forma geral, $g(a_{-\infty}^{-1}b)$ especifica a probabilidade de transição de $a_{-\infty}^{-1} \in A^{\leq -1}$ para $b \in A$. As g -funções, portanto, podem também ser interpretadas como as regras que regem a evolução temporal de um sistema físico.

Exemplo 3. A g -função associada às transições de uma sequência de variáveis aleatórias independentes e identicamente distribuídas com distribuição $\mu \in \mathcal{M}$ satisfaz $g(a_{-\infty}^{-1}b) = \mu([b])$ para qualquer $a_{-\infty}^{-1} \in A^{\leq -1}$ e $b \in A$.

Exemplo 4. Os mais simples exemplos de g -funções são as matrizes de transição associadas às cadeias de Markov de ordem k ($k \geq 1$). Elas satisfazem $g(a_{-\infty}^{-1}b) = g(c_{-\infty}^{-1}b)$ sempre quando $a_{-k}^{-1} = c_{-k}^{-1}$, $\forall a_{-\infty}^{-1}, c_{-\infty}^{-1} \in A^{\leq -1}$ e $\forall b \in A$.

Exemplo 5. Um exemplo simples de g -função não-Markoviana é dado considerando o alfabeto $A = \{0, 1\}$ e uma sequência $\{q_n : n \in \mathbb{N} \cup \{+\infty\}\}$ de números reais pertencentes ao intervalo $[0, 1]$. Neste caso, g é dada por

$$g(a_{-\infty}^{-1}1) = q_{\ell(a_{-\infty}^{-1})},$$

em que, para qualquer $a_{-\infty}^{-1} \in A^{\leq -1}$, $\ell(a_{-\infty}^{-1}) := \inf\{k \geq 1 : a_{-k} = 1\}$ e $\ell(a_{-\infty}^{-1}) = +\infty$ quando $a_{-k} = 0$ para todo $k \geq 1$. Dessa forma, o valor da g -função depende da distância em que ocorre o último símbolo 1 na sequência $\dots a_{-2}a_{-1}$. De fato, para qualquer $k \geq 1$ e todo $c \in A$, $g(a_{-\infty}^{-1}c)$ não necessariamente é igual a $g(b_{-\infty}^{-1}c)$ quando $a_{-k}^{-1} = b_{-k}^{-1} = 0_{-1}^{-k}$. Portanto, g não se trata de uma matriz de transição.

Dada uma g -função g , uma questão natural, é sobre a existência de uma medida invariante por translação compatível com g . Este é um dos problemas discutidos nesta tese. Formalizamos, assim, o conceito de compatibilidade nas Definições 1 e 2.

Definição 1. Uma cadeia estocástica $\mathbf{X} = \{X_n : n \in \mathbb{Z}\}$ com valores em um alfabeto finito A e que esteja definida sobre um espaço de probabilidade (Ω, Σ, P) é dita ser compatível com uma dada g -função g se, e somente se, para P -quase toda $a_{-\infty}^{-1} \in A^{\leq -1}$ e para todo $b \in A$,

$$P\{X_0 = b \mid X_{-\infty}^{-1} = a_{-\infty}^{-1}\} = g(a_{-\infty}^{-1}b).$$

Neste caso, \mathbf{X} é dita ser uma *cadeia estocástica de memória ilimitada*.

Definição 2. Uma medida $\mu \in \mathcal{M}$ é compatível com g se, e somente se, para μ -quase toda $a_{-\infty}^{-1} \in A^{\leq -1}$ e $b \in A$,

$$E_{\mu} [\mathbb{I}(\{a_0 = b\}) | \mathcal{F}^{\leq -1}] (\mathbf{a}) = \mu ([b] | \mathcal{F}^{\leq -1}) (\mathbf{a}) = g (a_{-\infty}^{-1} b).$$

Se, além disso, μ é invariante pela translação T então μ é dita ser uma g -medida.

Quando uma medida de probabilidade $\mu \in \mathcal{M}$ é compatível com uma dada g -função g , vamos escrever $\mu \in \mathcal{M}(g)$. Se, além disso, μ for invariante por translação, então $\mu \in \mathcal{M}_T(g) := \mathcal{M}_T \cap \mathcal{M}(g)$. De fato, se μ for uma g -medida então a cadeia \mathbf{X} definida segundo essa distribuição é uma cadeia estacionária. Algumas g -funções podem apresentar mais de uma g -medida compatível, é o caso das g -funções definidas nos Exemplos 6 e 7.

Exemplo 6 (Modelo do tipo Bramson e Kalikow(1993)). Definimos uma distância aleatória no passado $N \in \mathbb{N}$ com distribuição de probabilidade

$$P(N = m_j) = p_j, \quad \forall j \in \mathbb{N},$$

em que $\{m_j\}_{j=1}^{+\infty}$ é uma sequência crescente de números inteiros positivos ímpares e $\{p_j\}_{j=1}^{+\infty}$ é uma sequência decrescente de números reais em $[0, 1]$ tal que $p_j \geq 0$, para todo $j \in \mathbb{N}$, $\sum_{j=1}^{+\infty} p_j = 1$ e $p_k \leq \frac{1}{2} \sum_{j=k}^{+\infty} p_j$, qualquer que seja $k \in \mathbb{N}$. Note que, conforme j cresce os m_j 's tornam-se cada vez maiores, enquanto a probabilidade de N assumir tais valores diminui.

Seja $A = \{-1, +1\}$ e $0 < \varepsilon < 1/2$, então a partir de N definimos, para toda sequência $a_{-N}^{-1} \in A^N$, a variável aleatória W de forma que

$$W(a_{-N}^{-1}) = \begin{cases} \varepsilon, & \text{se } \sum_{j=1}^N a_{-j} > 0, \\ 1 - \varepsilon, & \text{se } \sum_{j=1}^N a_{-j} < 0, \end{cases}$$

i.e., o valor de W depende da quantidade de $+1$'s e -1 's existentes em uma dada sequência de tamanho N formada por símbolos de A . Neste contexto, podemos definir a seguinte g -função

$$g(a_{-\infty}^{-1} 1) = \sum_{j=1}^{+\infty} p_j w(a_{-m_j}^{-1}), \quad \forall a_{-\infty}^{-1} \in A^{\leq -1},$$

em que $w(\cdot)$ é o valor observado da variável aleatória W para um valor de N conhecido. Observe que, para todo $j \in \mathbb{N}$ fixo, $w(a_{-m_j}^{-1})$ é uma matriz de transição Markoviana, que assume os valores ε ou $1 - \varepsilon$ dependendo se maioria dos símbolos da sequência $a_{-m_j}^{-1}$ forem $+1$ ou -1 , respectivamente. Logo, podemos encarar a g -função g como a soma convexa de matrizes de transição Markovianas. Neste caso, se $E_P[N] = +\infty$ não existe uma única g -medida.

Exemplo 7. Seja $A = \{-1, 1\}$ e

$$H = H(a_{-\infty}^{-1}) := - \sum_{n \geq 1} \frac{a_{-n}}{n^{1+\gamma}},$$

em que γ é um constante real positiva. Então, podemos definir uma g -função por

$$g(a_{-\infty}^{-1}1) := \frac{e^{-\beta H}}{e^{\beta H} + e^{-\beta H}},$$

com $\beta > 0$ representando, em *lattice systems*, o inverso da temperatura. Neste caso, se $\gamma \leq 1/2$ então não existe uma única g -medida.

2.3 Cadeias estacionárias de memória de alcance variável

Lembramos que o conjunto A^k representa o conjunto de todas as sequências de tamanho k com símbolos em A . Em particular, A^0 é o conjunto formado somente pela sequência vazia, isto é, $A^0 := \{\emptyset\}$. Neste sentido, a partir de agora denotamos o conjunto de todas as sequências finitas com símbolos em A por $A^* := \cup_{j=0}^{\infty} A^j$. Fixamos também a notação $|\cdot|$ para representar a quantidade de elementos de uma dada sequência ou de um dado conjunto.

Os modelos de árvores de contexto foram introduzidos primeiramente em Teoria da Informação por [Rissanen \(1983\)](#) como uma generalização parcimoniosa dos modelos Markovianos. Os processos estocásticos, definidos sobre um alfabeto finito, que permitem uma representação de árvore de contexto são chamados *cadeias de memória de alcance variável*.

Uma *árvore de contexto* é um conjunto de sequências finitas ou infinitas à esquerda. A ideia básica é que, para cada uma dessas sequências (passado), apenas um sufixo finito do passado, chamado *contexto*, é suficiente para prever o próximo símbolo.

Dizemos que uma sequência de símbolos $s_{-|s|}^{-1} \in A^{|s|}$ é um *sufixo* de uma sequência (finita ou não) $w_{-|w|}^{-1}$ se $|s| \leq |w|$ e existir uma sequência $u_{-|w|}^{-|s|-1}$ tal que $w_{-|w|}^{-1} = u_{-|w|}^{-|s|-1} s_{-|s|}^{-1}$. Neste caso, escrevemos $s_{-|s|}^{-1} \prec w_{-|w|}^{-1}$. Agora, quando $s_{-|s|}^{-1} \prec w_{-|w|}^{-1}$ ou $s_{-|s|}^{-1} = w_{-|w|}^{-1}$, escrevemos $s_{-|s|}^{-1} \preceq w_{-|w|}^{-1}$.

Definição 3 (Propriedade do sufixo). Um subconjunto τ de $A^* \cup A^{\leq -1}$ é uma árvore se nenhuma sequência $s_{-|s|}^{-1} \in \tau$ é um sufixo de outra sequência $w_{-|w|}^{-1} \in \tau$, $|s| \leq |w|$. Esta propriedade é chamada propriedade do sufixo.

A *altura* da árvore τ é definida como

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

Quando $h(\tau) < +\infty$, a árvore τ tem um número finito de sequências. Neste caso, dizemos que τ é *limitada* e denotamos o seu número de sequências por $|\tau|$. Por outro lado, se $h(\tau) = +\infty$, então τ é um conjunto infinito de sequências. Neste caso, dizemos que τ é *ilimitada*.

Cada sequência $w_{-|w|}^{-1} \in \tau$ pode ser vista como um ramo que liga a raiz da árvore (desenhada no topo) com uma folha. Por exemplo, a sequência w_{-k}^{-1} é representada por k arestas rotuladas, de cima para baixo, sequencialmente pelos símbolos w_{-1}, \dots, w_{-k} . Por esta razão,

w_{-k}^{-1} também é chamada de *folha* da árvore. Para o caso de uma sequência semi-infinita $v_{-\infty}^{-1} \in \tau$, a folha é infinita. Os *nós* de uma árvore τ são todas as sequências $s_{-|s|}^{-1}$, finitas ou infinitas, tais que $s_{-|s|}^{-1} \preceq w_{-|w|}^{-1}$ qualquer que seja $w_{-|w|}^{-1} \in \tau$, com $|s| \leq |w|$, de forma que a raiz é identificada como a sequência vazia \emptyset . Quando os nós são finitos dizemos que eles são *nós internos*. Os *filhos* de um nó interno $s_{-|s|}^{-1}$ são aquelas sequências $as_{-|s|}^{-1}$, $a \in A$, tal que $as_{-|s|}^{-1} \preceq w_{-|w|}^{-1}$, para qualquer $w_{-|w|}^{-1} \in \tau$.

Definição 4 (Árvore de Contexto). A árvore τ é *completa* se, e somente se, qualquer sequência $v_{-\infty}^{-1} \in A^{\leq -1}$ possui um sufixo pertencente a τ . Pela propriedade do sufixo temos que esse sufixo é único. Tal sufixo é chamado de *contexto* e o denotamos por $c_{\tau}(v_{-\infty}^{-1})$. Uma árvore completa é chamada de *árvore de contexto*.

Exemplo 8. Um dos mais simples exemplos de árvore é a árvore completa. Seja $A = \{0, 1, 2\}$ e $\tau = \{00, 10, 20, 01, 11, 21, 02, 12, 22\}$. Neste caso, τ é limitada, pois $h(\tau) = 2$ e sua representação pictorial é dada pela Figura 1.

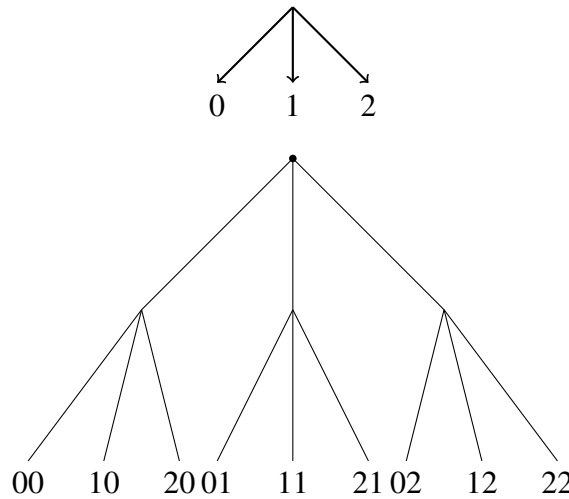


Figura 1 – Árvore completa associada ao alfabeto $A = \{0, 1, 2\}$.

Exemplo 9 (Árvore esparsa). Um exemplo de árvore ilimitada é a árvore esparsa. Essa árvore é uma representação pictorial da dinâmica definida no Exemplo 5, que atribui pesos de acordo com a distância do símbolo 1 na sequência 10_{-i}^{-1} , $i \geq 1$. Neste caso, $A = \{0, 1\}$ e $\tau := \{0_{-\infty}^{-1}\} \cup \cup_{i \geq 1} \{10_{-i}^{-1}\}$ e sua representação pictorial pode ser visto na Figura 2.

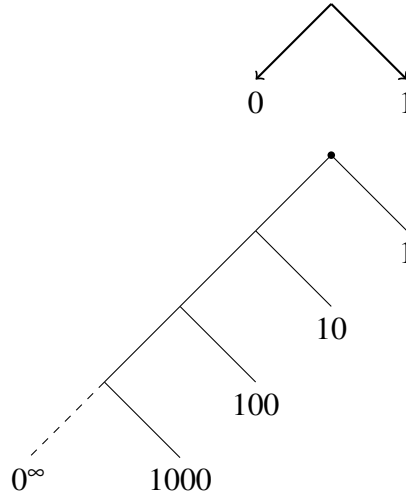


Figura 2 – Árvore esparsa.

Observação 1. Podemos estender a noção de contexto para seqüências finitas, i.e., para qualquer $w_{-|w|}^{-1} \in A^{|w|}$ com $|w| < +\infty$ dizemos que $c_\tau(w_{-|w|}^{-1}) = s_{-|s|}^{-1}$, com $|s| \leq |w|$, se $s_{-|s|}^{-1} \sqsupseteq w_{-|w|}^{-1}$ e $s_{-|s|}^{-1} \in \tau$.

Definição 5 (Árvore de contexto probabilística). Uma árvore de contexto probabilística com elementos em A é um par ordenado (τ, p) tal que

1. τ é uma árvore de contexto;
2. $p := \left\{ p\left(b \middle| s_{-|s|}^{-1}\right) : b \in A \text{ e } s_{-|s|}^{-1} \in \tau \right\}$ é uma família de probabilidades de transição sobre A .

Definição 6 (Cadeia estocástica de alcance variável). Dizemos que a cadeia estocástica \mathbf{X} definida sobre (Ω, \mathcal{F}, P) é compatível com uma árvore de contextos probabilística (τ, p) se para P -quase toda $a_{-\infty}^{-1} \in A^{\leq -1}$ e qualquer $b \in A$ temos

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

Neste caso, dizemos que \mathbf{X} é uma cadeia de memória de alcance variável

Observe que utilizando a g -função do Exemplo 5, a árvore esparsa definida do Exemplo 9 é uma árvore de contexto probabilística.

É importante salientar que se τ é tal que $h(\tau) = k < +\infty$, então o processo \mathbf{X} é uma cadeia de Markov de ordem k . Neste caso, a árvore de contexto probabilística leva a uma descrição mais econômica da cadeia, porque uma coleção de $(|A| - 1)|\tau|$ probabilidades de transição é suficiente para descrever a cadeia, ao invés de $(|A| - 1)|A|^k$. Logo, em um processo de estimação a quantidade de parâmetros a serem estimados na abordagem Markoviana é superior do que na modelagem via árvores de contexto, principalmente para valores de k muito grandes.

Em um contexto de inferência estatística assumimos que as cadeias de memória de alcance variável ou de tamanho ilimitado são estacionárias, pois, do ponto de vista físico, é interessante analisar os resultados estocásticos de um experimento sem uma dependência explícita da escolha do instante inicial e , além disso, o valor médio relativo a um dado vetor aleatório não dependeria do tempo.

A ESTACIONARIEDADE EM CADEIAS DE MEMÓRIA ILIMITADA E DE ALCANCE VARIÁVEL

Diferente do que ocorre com as cadeias de Markov de alfabeto finito para as quais sempre existe pelo menos uma medida invariante por translação compatível com uma dada matriz de transição, ou equivalentemente, uma cadeia estacionária; a existência de uma cadeia estacionária compatível com uma dada g -função não é garantida. A unicidade dessa g -medida é outra questão amplamente discutida na literatura. No contexto Markoviano sabemos que a irreduzibilidade da matriz de transição implica em unicidade, porém isso não é sempre verdade para g -medida. Este capítulo, portanto, pretende, além de outras coisas, esclarecer o conceito de continuidade e explorar a sua relação com a existência e a unicidade de uma cadeia estacionária de memória ilimitada com particular enfoque em cadeias de alcance variável.

3.1 Existência

Dada uma g -função g , o problema de existência consiste em encontrar condições sobre g que garantam $\mathcal{M}_T(g) \neq \emptyset$. Uma condição suficiente para a existência de uma medida $\mu \in \mathcal{M}_T(g)$ é assumir que g seja contínua em todo ponto $a_{-\infty}^{-1} \in A^{\leq -1}$ (ver, [Keane \(1972\)](#)), isto é,

$$g\left(c_{-\infty}^{-k-1} a_{-k}^{-1} b\right) \rightarrow g\left(a_{-\infty}^{-1} b\right), \quad \text{quando } k \rightarrow +\infty,$$

qualquer que sejam $c_{-\infty}^{-1} \in A^{\leq -1}$ e $b \in A$. Pela compacidade de $\mathcal{M}(g)$ é equivalente dizer que para todo $\varepsilon > 0$, existe um inteiro positivo k tal que

$$|g(c_{-\infty}^{-1} b) - g(a_{-\infty}^{-1} b)| < \varepsilon,$$

sempre quando $c_{-\infty}^{-1}$ e $a_{-\infty}^{-1}$ pertencem a $A^{\leq -1}$ com $a_{-k}^{-1} = c_{-k}^{-1}$. Este conceito pode, também, ser formalizado em termos da noção da variação de g de k -ésima ordem utilizada, em geral, para quantificar a taxa de continuidade de g .

Definição 7 (Variação de g de k -ésima ordem). Se g é uma g -função, então a variação de g de k -ésima ordem no ponto $a_{-\infty}^{-1} \in A^{\leq -1}$ é definida por

$$\text{var}_k^g(a_{-\infty}^{-1}) := \sup \{ |g(a_{-\infty}^{-1}b) - g(c_{-\infty}^{-1}b)| : a_{-k}^{-1} = c_{-k}^{-1} \},$$

em que $c_{-\infty}^{-1} \in A^{\leq -1}$ e $b \in A$.

Note que $\text{var}_k^g(a_{-\infty}^{-1})$ converge para 0 se, e somente se, $a_{-\infty}^{-1}$ é um ponto de continuidade de g . Portanto, se uma medida de probabilidade é compatível com uma g -função contínua, então “a dependência do presente sobre um passado remoto é fraca”. Para fins de ilustração, verificamos que as g -funções definidas nos Exemplos 4, 6 e 7 são contínuas.

Exemplo 10 (Cadeias de Markov). Como vimos no Exemplo 4, para cadeias de Markov de ordem m ($m \geq 1$) a g -função g é dada pela sua matriz de transição, a qual satisfaz $g(a_{-\infty}^{-1}b) = g(c_{-\infty}^{-1}b)$ sempre quando $a_{-m}^{-1} = c_{-m}^{-1}$, qualquer que seja $b \in A$. Dessa forma, se $k \geq m$ então $\text{var}_k^g(a_{-\infty}^{-1}) = 0$.

Exemplo 11 (Modelo do tipo Bramson e Kalikow (1993)). No Exemplo 6 a g -função g é definida, para todo $a_{-\infty}^{-1} \in A^{\leq -1}$, como

$$g(a_{-\infty}^{-1}1) := \sum_{j \geq 1} p_j w(a_{-m_j}^{-1}),$$

em que

$$w(a_{-m_j}^{-1}) = \begin{cases} \varepsilon, & \text{se } \sum_{i=1}^{m_j} a_{-i} > 0, \\ 1 - \varepsilon, & \text{se } \sum_{i=1}^{m_j} a_{-i} < 0, \end{cases}$$

com $0 < \varepsilon < 1/2$ e $\{m_j\}_{j=1}^{+\infty}$ uma sequência crescente de números inteiros positivos ímpares. Dessa forma,

$$\begin{aligned} \text{var}_k^g(a_{-\infty}^{-1}) &:= \sup_{a_{-k}^{-1} = c_{-k}^{-1}} \left| \sum_{j \geq 1} p_j w(c_{-m_j}^{-1}) - \sum_{j \geq 1} p_j w(a_{-m_j}^{-1}) \right| \\ &= \sup_{a_{-k}^{-1} = c_{-k}^{-1}} \left| \sum_{\substack{j \geq j^* \\ m_{j^*} > k}} p_j w(c_{-j}^{-1}) - \sum_{\substack{j \geq j^* \\ m_{j^*} > k}} p_j w(a_{-j}^{-1}) \right| \\ &= |1 - 2\varepsilon| \sum_{\substack{j > j^* \\ m_{j^*} > k}} p_j \\ &\leq \sum_{\substack{j > j^* \\ m_{j^*} > k}} p_j, \end{aligned}$$

a qual converge a 0 quando $k \rightarrow +\infty$.

Exemplo 12. No Exemplo 7, consideramos o alfabeto $A = \{-1, 1\}$ e a g -função g foi definida como

$$g(a_{-\infty}^{-1}1) := \phi(-\beta H(a_{-\infty}^{-1})),$$

em que $\beta > 0$, $a_{-\infty}^{-1} \in A^{\leq -1}$, $H(a_{-\infty}^{-1}) = -\sum_{n \geq 1} \frac{a_{-n}}{n^{1+\gamma}}$ com $\gamma > 0$ e

$$\phi(t) = \frac{e^t}{e^t + e^{-t}}, \quad t \in \mathbb{R}.$$

Note que $0 \leq \phi' \leq 1/2$, então

$$\text{var}_k^g(a_{-\infty}^{-1}) \leq \beta \sum_{n > k} \frac{1}{n^{1+\gamma}},$$

a qual converge para 0 quando $k \rightarrow +\infty$, pois $\sum_{n > k} \frac{1}{n^{1+\gamma}} < +\infty$.

Note que se g for uma matriz de transição irredutível no Exemplo 4, então uniformemente no passado, existe sempre uma probabilidade positiva de ocorrer uma transição para qualquer outro símbolo de A . Uma condição que comumente vamos assumir é mais forte que irredutibilidade e trata-se da não-nulidade uniforme, a qual é satisfeita pelos Exemplos 11 e 12 e cuja definição é apresentada a seguir.

Definição 8 (g -função uniformemente não-nula). Uma g -função g é uniformemente não-nula se ela é uniformemente limitada no intervalo $[0, 1]$, ou seja, se existe $\varepsilon \in (0, \frac{1}{2})$ tal que $\varepsilon \leq g(a_{-\infty}^{-1}b) \leq 1 - \varepsilon$, quaisquer que sejam $a_{-\infty}^{-1} \in A^{\leq -1}$ e $b \in A$.

De agora em diante, sempre que uma g -função for contínua e uniformemente não-nula diremos que ela é *regular*. A regularidade de uma g -função somada a uma taxa de convergência rápida são características que nos garantem a unicidade da cadeia estacionária. Essa questão será tratada na próxima seção, por agora nos limitamos à questão da existência. Retomemos, portanto, ao Exemplo 5 e foquemos nossa atenção sobre a existência da medida invariante compatível com a g -função lá definida. É importante destacar, antes de prosseguirmos, que em Cénac *et al.* (2012) encontramos um trabalho detalhado com respeito à estacionariedade de cadeias compatíveis com essa g -função e que muito dos cálculos apresentados a seguir são inspirados por esse trabalho.

Exemplo 13. Lembremos que a g -função g definida no Exemplo 5 é dada por

$$g(a_{-\infty}^{-1}1) := q_{\ell(a_{-\infty}^{-1})},$$

em que $\ell(a_{-\infty}^{-1}) := \inf\{k \geq 1 : a_{-k} = 1\}$ e $\{q_n : n \in \mathbb{N} \cup \{+\infty\}\}$ é uma sequência de números reais em $[0, 1]$. Suponha que exista uma g -medida μ e seja \mathbf{X} a respectiva cadeia estacionária. A árvore de contexto com a qual \mathbf{X} é compatível é a árvore esparsa $\mathcal{T} = \{0_{-\infty}^{-1}\} \cup \cup_{i \geq 1} \{10_{-i}^{-1}\}$ (ver o Exemplo 9). Neste sentido, ao verificarmos como o peso da medida de probabilidade μ é distribuído ao longo dos contextos e de seus nós, temos um conhecimento probabilístico completo com relação a cadeia \mathbf{X} .

Primeiramente, usamos o teorema da probabilidade total para concluir que

$$\mu([1] \cap [0_{-k}^{-1}]) = \mu([1]) \prod_{i=0}^{k-1} (1 - q_{i+1}), \quad (3.1)$$

sempre quando $k \geq 1$. Se $k = 0$ definimos $\mu([1] \cap [0_{-k}^{-1}]) := \mu([1])$. Agora, como μ é aditiva para conjuntos disjuntos, temos $\mu([0_{-k}^{-1}]) = \mu([0] \cap [0_{-k}^{-1}]) + \mu([1] \cap [0_{-k}^{-1}])$, equivalentemente, $\mu([0_{-k-1}^{-1}]) = \mu([0_{-k}^{-1}]) - \mu([1] \cap [0_{-k}^{-1}])$. Assim, por recursividade em k ,

$$\mu([0_{-k}^{-1}]) = 1 - \sum_{i=0}^{k-1} \mu([1] \cap [0_{-i}^{-1}]) = 1 - \mu([1]) \sum_{i=0}^{k-1} \prod_{j=0}^{i-1} (1 - q_{j+1}). \quad (3.2)$$

Logo,

$$\mu([0_{-\infty}^{-1}]) = 1 - \mu([1]) \sum_{i \geq 0} \prod_{j=0}^{i-1} (1 - q_{j+1}). \quad (3.3)$$

No caso em que $q_{\infty} > 0$ temos, pela definição de probabilidade condicional, $\mu([0_{-\infty}^{-1}]) = \mu([0_{-\infty}^{-1}]) q_{\infty}$, então devemos ter, necessariamente, $\mu([0_{-\infty}^{-1}]) = 0$. Por (3.3) segue que $\mu([1]) \neq 0$ e

$$\sum_{i \geq 0} \prod_{j=0}^{i-1} (1 - q_{j+1}) < +\infty. \quad (3.4)$$

Reciprocamente, quando assumimos que essa série é convergente as equações (3.1), (3.2) e (3.3) estão bem definidas. Portanto, quando $q_{\infty} > 0$, uma g -medida existe se, e somente se, (3.4) se mantem.

Desta forma, se $q_{\infty} > 0$ e $\sum_{i \geq 0} \prod_{j=0}^{i-1} (1 - q_{j+1}) = +\infty$, então **não existe** a g -medida.

Exemplo 14. Novamente considere a g -função definida no Exemplo 5. Agora, de forma que

$$q_i := \begin{cases} 1 - \varepsilon, & \text{se } i \text{ é par,} \\ \varepsilon, & \text{se } i \text{ é ímpar,} \end{cases}$$

em que $0 < \varepsilon < 1/2$ e $i \in \mathbb{N}$. Vamos supor que $q_{\infty} > 0$, então g apresenta uma descontinuidade em $0_{-\infty}^{-1}$, pois $g(a_{-\infty}^{-i-1} 10_{-i+1}^{-1} 1)$ é uma sequência que oscila entre ε e $1 - \varepsilon$ conforme i cresce, qualquer que seja $a_{-\infty}^{-1} \in A^{\leq -1}$.

No entanto,

$$\sum_{i \geq 0} \prod_{j=0}^{i-1} (1 - q_{j+1}) \leq \sum_{i=0}^{\infty} (1 - \varepsilon)^i < +\infty,$$

pois é uma série geométrica. Assim, como estamos no cenário em que $q_{\infty} > 0$, o Exemplo 13 nos garante que a g -medida existe.

Continuidade, portanto, não é necessária para existência (ver, para mais exemplos, Gallo (2011)). Apenas recentemente, Gallo (2011), Cénac *et al.* (2012), Santis e Piccioni (2012) e Gallo

e Paccaut (2013) estudaram g -medidas com g -funções que não são necessariamente contínuas. Nesses trabalhos, uma consequência das condições utilizadas para garantir a existência é que a medida não atribui peso às descontinuidades. Neste contexto, é natural pensar que se uma medida de probabilidade é compatível com uma g -função g e não atribui peso às descontinuidades, então $\mathcal{M}_T(g) \neq \emptyset$. No entanto, não temos, a priori, uma medida inicial, fazendo-se necessária a construção dessa medida de referência.

Pelo teorema de Ionescu-Tulcea (Ionescu (1949)), sabemos que dado qualquer passado $a_{-\infty}^{-1} \in A^{\leq -1}$, existe uma única medida $\mu^a \in \mathcal{M}$ satisfazendo $\mu^a(C) = \mathbb{I}\{a_{-\infty}^{-1} \in C\}$ qualquer que seja $C \in \mathcal{F}^{\leq -1}$ e, para todo, $n \geq 1$ e $b_0^{n-1} \in A^n$ temos

$$\mu^a(b_0^{n-1}) = \prod_{i=0}^{n-1} g(a_{-\infty}^{-1} b_0^{i-1} b_i),$$

em que \mathbb{I} denota a função indicadora. Em palavras, esta é a medida iniciada a partir da sequência $a_{-\infty}^{-1}$ e com a qual construímos o futuro iterativamente utilizando g .

Sejam $\mu^{a,0} := \mu^a$, $\mu^{a,i} := \mu^a \circ T^{-i}$ para qualquer $i \geq 1$ e

$$\bar{\mu}^{a,k} := \frac{1}{k} \sum_{i=0}^{k-1} \mu^{a,i}, \quad k \geq 1. \quad (3.5)$$

Pela compacidade de \mathcal{M} existe pelo menos uma subsequência de $\{\bar{\mu}^{a,k} : k \geq 1\}$ que é convergente. A nossa medida de referência será, então, o limite desta subsequência, a qual denotamos por $\bar{\mu}^a$. Como veremos nas demonstrações a média de Césaro $\bar{\mu}^{a,k}$ é invariante por translação qualquer que seja $k \geq 1$, portanto é natural construir a medida de referência $\bar{\mu}^a$ a partir dessa média.

Observação 2. Em termos de processos estocásticos, consideramos, para cada $a_{-\infty}^{-1} \in A^{\leq -1}$, a cadeia $\mathbf{X}^a := \{X_t^a : t \in \mathbb{Z}\}$ definida sobre (Ω, Σ, P) com valores em um alfabeto A de forma que $(X_t^a)_{t \leq -1} = a_{-\infty}^{-1}$ e para $t \geq 0$ a cadeia é construída iterativamente utilizando uma dada g -função g da seguinte forma

$$P(X_0^a = b_0, X_1^a = b_1, \dots, X_{n-1}^a = b_{n-1}) = \prod_{i=0}^{n-1} g(a_{-\infty}^{-1} b_0^{i-1} b_i),$$

qualquer que seja a sequência b_0^{n-1} de símbolos de A e qualquer que seja $n \in \mathbb{N}$.

Então, dada a sequência $a_{-\infty}^{-1} \in A^{\leq -1}$, nossa cadeia de referência é o processo $\mathbf{Y}^a := \{Y_t^a : t \in \mathbb{Z}\}$ definido sobre (Ω, Σ, P) com valores em A construído probabilisticamente como sendo o limite

$$P(Y_0^a = b_0, Y_1^a = b_1, \dots, Y_{n-1}^a = b_{n-1}) = \lim_{k \rightarrow +\infty} \frac{1}{k} \sum_{i=0}^{k-1} P(X_i^a = b_0, X_{i+1}^a = b_1, \dots, X_{i+n-1}^a = b_{n-1}),$$

qualquer que seja $b_0^{n-1} \in A^n$.

As observações anteriores nos levam ao nosso primeiro resultado, o qual fornece uma condição geral para o conjunto dos pontos de descontinuidade de g

$$\mathcal{D}_g := \left\{ c_{-\infty}^{-1} \in A^{\leq -1} : \lim_{k \rightarrow +\infty} \text{var}_k^g(c_{-\infty}^{-1}) \neq 0 \right\},$$

de forma que exista pelo menos uma g -medida. Este é o conteúdo do Teorema 1.

Teorema 1. Sejam g uma g -função e $\bar{\mu}^a$ uma medida limite com $a_{-\infty}^{-1} \in A^{\leq -1}$. Se para l suficientemente grande $\bar{\mu}^a([c_{-l}^{-1}]) > 0$ qualquer que seja $c_{-\infty}^{-1} \in \mathcal{D}_g^c := A^{\leq -1} \setminus \mathcal{D}_g$ e $\bar{\mu}^a(\mathcal{D}_g) = 0$, então $\bar{\mu}^a \in \mathcal{M}_T(g)$ e, conseqüentemente, $\mathcal{M}_T(g) \neq \emptyset$.

Observação 3. Em Oliveira (2015) o autor introduz uma classe de cadeias estocásticas com alfabeto finito, chamada processos de contextos aleatórios, cuja probabilidade de transição de uma sequência $a_{-\infty}^{-1} \in A^{\leq -1}$ para um símbolo $b \in A$ é dada por uma combinação convexa de probabilidades condicionais, em que cada uma dessas probabilidades está associada com uma árvore de contexto τ_k , $k \geq 1$, e representa a transição de um contexto de $a_{-\infty}^{-1}$ pertencente à árvore τ_k para o símbolo $b \in A$. Gallo e Garcia (2013) utilizaram uma ideia similar, a qual denominaram combinação convexa de árvores de contexto, para mostrar a existência da cadeia estacionária via simulação perfeita, isto é, um algoritmo que permite construir uma amostra distribuída segundo a lei estacionária de uma cadeia. Oliveira (2015) afirma que a cadeia estacionária de contextos aleatórios existe se, e somente se, a sua família de probabilidade de transição é quase certamente contínua. Desta forma, o Teorema 1 nos garante que a cadeia estacionária compatível com uma dada g -função $\bar{\mu}^a$ -quase certamente contínua possui uma representação de contextos aleatórios.

Muito embora exigir que $\bar{\mu}^a(\mathcal{D}_g) = 0$ seja uma condição natural, pois dessa forma a medida de referência não “enxerga” as descontinuidades, ela não é completamente explícita. Mesmo assim, duas situações, que ainda não foram contempladas em resultados de existência na literatura, são alcançados com o Teorema 1 e nos fornecem afirmações mais explícitas, as quais enunciamos nos Corolários 1 e 2.

Corolário 1. Se g é uma g -função uniformemente não-nula e \mathcal{D}_g é enumerável, então $\mathcal{M}_T(g) \neq \emptyset$.

Exemplo 15. Vimos no Exemplo 13 que existe pelo menos uma g -medida compatível com a g -função g definida no Exemplo 5 se, e somente se, $q_\infty > 0$ e $\sum_{l \geq 0} \prod_{j=0}^{l-1} (1 - q_{l+1}) < +\infty$. O conjunto dos pontos de continuidade de g é $\mathcal{D}_g^c = \{10_{-l}^{-1} : l \leq 0\}$, de forma que para $l = 1$ temos $10_{-1}^{-1} = 1$. Assim, se $\bar{\mu}^a$ é uma medida limite com $a_{-\infty}^{-1} \in A^{\leq -1}$ segue que

$$\bar{\mu}^a([1]) = \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \mu^a([T^{-i}1]). \quad (3.6)$$

Tomando $a_{-\infty}^{-1} \in A^{\leq -1}$ tal que $a_{-1} = 1$ temos, pelo Teorema da Renovação e pelo fato que (3.6) é uma média de Césaró,

$$\bar{\mu}^a([1]) = \frac{1}{1 + \sum_{l \geq 0} \prod_{j=0}^{l-1} (1 - q_{l+1})}. \quad (3.7)$$

Assumindo $\sum_{l \geq 0} \prod_{j=0}^{l-1} (1 - q_{l+1}) < +\infty$ segue que $\bar{\mu}^a([1]) > 0$. Supondo que $0 < q_j < 1$ para todo $j \geq 1$ vem que $\bar{\mu}^a([c_{-l}^{-1}]) > 0$ para qualquer $c_{-\infty}^{-1} \in \mathcal{D}_g^c$ e $l \geq 1$. De fato, qualquer que seja $c_{-\infty}^{-1} \in \mathcal{D}_g^c$ com $l \geq 1$ existe uma sequência $\{i_n : 1 \leq n \leq k\}$, $k \in \mathbb{Z} \cap (0, +\infty)$ tal que

$$\begin{aligned} \bar{\mu}^a([c_{-l}^{-1}]) &= \bar{\mu}^a\left([0_{-l}^{i_1}] \cap [1] \cap [0_{i_1+1}^{i_2}] \cap [1] \cap \dots \cap [0_{i_k+1}^{-1}]\right) \\ &= \prod_{i=1}^{i_1-l} (1 - q_i) \frac{1}{q_k + 1} \prod_{l=1}^k \left[q_{i_l+1} \prod_{j=1}^{i_l} (1 - q_j) \right] \\ &> 0. \end{aligned}$$

Como $\mathcal{D}_g = \{0_{-\infty}^{-1}\}$ é enumerável temos $\bar{\mu}^a(\mathcal{D}_g) = 0$ e, portanto, pelo Teorema 1 concluímos que $\mathcal{M}_T(g) \neq \emptyset$.

Dessa forma, os resultados anteriores contemplam a questão da existência associada a g -função definida no Exemplo 5 e tratada no Exemplo 13.

Exemplo 16 (De Santis e Piccioni (2012)). Considere ℓ a função definida no Exemplo 5 e g uma g -função definida sobre o alfabeto $A = \{0, 1\}$ de forma que $g(0_{-\infty}^{-1}1) = \varepsilon > 0$ e para qualquer $a_{-\infty}^{-1} \neq 0_{-\infty}^{-1}$ e $b \in A$ tem-se

$$g(a_{-\infty}^{-1}b) = \varepsilon + (1 - 2\varepsilon) \sum_{n \geq 1} \left[\mathbb{I}\{a_{-\ell(a_{-\infty}^{-1})-n} = b\} q_n^{\ell(a_{-\infty}^{-1})} \right],$$

em que, para cada $\ell \geq 0$, $\{q_n^\ell : n \geq 1\}$ é uma distribuição de probabilidade sobre os inteiros.

Note que $0_{-\infty}^{-1}$ é um ponto de descontinuidade de g , pois para todo $i \in \mathbb{N}$,

$$g(1_{-\infty}^{-i-1}0_{-i}^{-1}1) = \varepsilon + (1 - 2\varepsilon) \sum_{n \geq 1} q_n^{i+1} = 1 - \varepsilon.$$

Então, $g(1_{-\infty}^{-i-1}0_{-i}^{-1}1) \rightarrow 1 - \varepsilon \neq \varepsilon = g(0_{-\infty}^{-1}1)$ quando $i \rightarrow +\infty$. Todavia, para todos os outros pontos g é contínua, uma vez que para qualquer $a_{-\infty}^{-1} \in A^{\leq -1}$ tal que $\ell(a_{-\infty}^{-1}) = \ell < +\infty$ e para todo $c_{-\infty}^{-1}$ e $i > \ell$,

$$g(c_{-\infty}^{-i-1}a_{-i}^{-1}1) = \varepsilon + (1 - 2\varepsilon) \left[\sum_{j=1}^{i-\ell} \mathbb{I}\{a_{-\ell-j} = 1\} + \sum_{j \geq i-\ell+1} \mathbb{I}\{c_{-\ell-j} = 1\} q_j^\ell \right],$$

a qual converge para $g(a_{-\infty}^{-1}1) = \varepsilon + (1 - 2\varepsilon) \sum_{j \geq 1} \mathbb{I}\{a_{-\ell-j} = 1\} q_j^\ell$ quando $i \rightarrow +\infty$.

Assim, $\mathcal{D}_g = \{0_{-\infty}^{-1}\}$, que é enumerável. Então, pelo Corolário 1 concluímos que $\mathcal{M}_T(g) \neq \emptyset$.

Este é um resultado interessante, pois Santis e Piccioni (2012) mostraram a existência da g -medida fazendo algumas suposições acerca do conjunto de distribuições $\{q_n^\ell : n \geq 1\}_{\ell \geq 0}$, enquanto usando o Corolário 1 a existência foi garantida sem quaisquer suposições sobre essas sequências de distribuições.

Observação 4. É importante salientar que a exigência de $\bar{\mu}^a$ ser estritamente positiva sobre os pontos de continuidade para um tamanho de cilindro suficientemente grande é uma hipótese mais fraca que a positividade uniforme da g -função (ver a formalização na demonstração do Corolário 1).

Nos resultados anteriores controlamos o “tamanho” de \mathcal{D}_g . Todavia, mesmo para \mathcal{D}_g não-enumerável é possível garantir a existência de uma g -medida. Neste caso, uma alternativa é focar nossa atenção sobre a forma de \mathcal{D}_g ao invés de seu tamanho. Um exemplo é quando existe uma sequência finita $v_{-|v|}^{-1}$ de símbolos de A que não é uma subsequência dos elementos de \mathcal{D}_g , i.e., para qualquer $c_{-\infty}^{-1} \in \mathcal{D}_g$ temos $c_{-i-|v|+1}^{-i} \neq v_{-|v|}^{-1}$, para todo $i \geq 1$. Neste caso, \mathcal{D}_g é dito ser v -livre.

Corolário 2. Seja g uma g -função uniformemente não-nula. Se \mathcal{D}_g é v -livre para alguma sequência finita $v_{-|v|}^{-1}$, então $\mathcal{M}_T(g) \neq \emptyset$.

Exemplo 17. Seja $A = \{0, 1, 2\}$ e ℓ a função definida no Exemplo 5. A g -função g é, então, definida da seguinte maneira

$$\begin{aligned} g(a_{-\infty}^{-1}1) &= g(a_{-\infty}^{-1}0) = 0.3, & \text{se } a_{-\infty}^{-1} \in \{0, 2\}^{\leq -1}, \\ g(a_{-\infty}^{-1}1) &= g(a_{-\infty}^{-1}0) = 0.26 + \sum_{k \geq 1} \theta_k a_{-\ell(a_{-\infty}^{-1})-k}, & \text{se } a_{-\infty}^{-1} \notin \{0, 2\}^{\leq -1}, \end{aligned}$$

em que $\{\theta_i : i \geq 1\}$ é tal que $\theta_i \geq 0$ e $\sum_{i \geq 1} \theta_i < 0.03$.

Observe que, para qualquer $a_{-\infty}^{-1} \in \{0, 2\}^{\leq -1}$, $g(1_{-\infty}^{-i-1} a_{-i}^{-1} 1) < 0.29$, para i suficientemente grande. Neste caso, g não converge para 0.3, portanto $\{0, 2\}^{\leq -1}$ é um ponto de descontinuidade. Por outro lado, se $a_{-\infty}^{-1} \notin \{0, 2\}^{-\mathbb{N}}$ então

$$g(c_{-\infty}^{-i-1} a_{-i}^{-1} 1) = 0.26 + \sum_{k=1}^{i-\ell} \theta_k a_{-\ell-k} + \sum_{k \geq i-\ell+1} \theta_k c_{k-2i-2+\ell},$$

converge, para todo $c_{-\infty}^{-1} \in A^{\leq -1}$, para $0.26 + \sum_{k \geq 1} \theta_k a_{-\ell(a_{-\infty}^{-1})-k}$, quando $i \rightarrow +\infty$. Portanto, $\mathcal{D}_g = \{0, 2\}^{\leq -1}$, o qual é não-enumerável.

Todavia, D_g é 1-livre, i.e., 1 não é subsequência de $\{0, 2\}^{\leq -1}$ e, além disso, g é uniformemente não-nula. Portanto, pelo Corolário 2, concluímos que $\mathcal{M}_T(g) \neq \emptyset$.

Até então, na literatura, a existência não havia sido mostrada nesse nível de generalidade para esse exemplo. No Corolário 1 e Exemplo 4 de Gallo e Paccout (2013), os autores necessitam que o parâmetro ε seja suficientemente grande para compensar o crescimento exponencial da subsequências de tamanho n dos elementos de \mathcal{D}_g . Por outro lado, Gallo e Garcia (2013) obtiveram existência adicionando uma condição sobre o conjunto dos pontos de continuidade $A^{\leq -1} \setminus \mathcal{D}_g$, condição essa mais forte e que implica em unicidade também.

Vamos agora mencionar outra forma de mensurar o quão pequeno é \mathcal{D}_g , a qual é considerada em [Gallo e Paccaut \(2013\)](#). Primeiramente denotamos por \mathcal{D}_g^n o conjunto das seqüências de tamanho n as quais são sufixos de um passado descontínuo de g , i.e.,

$$\mathcal{D}_g^n := \{d_{-n}^{-1} : d_{-\infty}^{-1} \in \mathcal{D}_g\}.$$

Agora, definimos a *pressão topológica* de \mathcal{D}_g como

$$P_g(\mathcal{D}_g) = \lim_{n \rightarrow +\infty} \sup \frac{1}{n} \log \sum_{d_{-n}^{-1} \in \mathcal{D}_g^n} \sup_{a_{-\infty}^{-1} \in A^{\leq -1}} \prod_{i=0}^{n-1} g(a_{-\infty}^{-n-1} d_{-n}^{-n+i}).$$

O Teorema 1 de [Gallo e Paccaut \(2013\)](#) afirma que existe pelo menos uma medida invariante por translação compatível com g se $P_g(\mathcal{D}_g) < 0$. A próxima proposição afirma que esta condição implica na condição do Teorema 1, mostrando que nosso resultado é mais geral.

Proposição 1. Suponha que g seja uma g -função uniformemente não-nula e que $P_g(\mathcal{D}_g) < 0$, então existe um passado $a_{-\infty}^{-1}$ e um limite $\bar{\mu}^a$ tal que $\bar{\mu}^a(\mathcal{D}_g) = 0$.

3.2 Unicidade

Até agora obtivemos resultados sobre a existência, e nada foi estabelecido para a unicidade de $\mathcal{M}_T(g)$, i.e., não foram fornecidas quaisquer condições sobre g a fim de garantir que $|\mathcal{M}_T(g)| = 1$. Para g -funções regulares resultados a cerca da unicidade envolvem restrições sobre a dependência de g com o passado, o que é obtido impondo que a taxa de continuidade convirja para 0 rapidamente (ver, por exemplo, [Doebelin e Fortet \(1937\)](#) e [Johansson e Öberg \(2008\)](#)). Colocando condições sobre o conjunto de continuidade, $A^{\leq -1} \setminus \mathcal{D}_g$, poderíamos obter, como em [Gallo e Paccaut \(2013\)](#) e [Gallo e Garcia \(2013\)](#), condições para unicidade. Todavia, não seguimos essa abordagem neste trabalho e voltamos nossa atenção para as cadeias estocásticas de memória de alcance variável (ver seção 2.3), i.e., formalizamos nossos resultados de unicidade usando a noção de árvore de contexto introduzida na seção 2.3.

Neste sentido, considere g uma árvore de contexto probabilística (τ, p) e defina $\ell_\tau(c_{-\infty}^{-1})$ como o tamanho do sufixo de $c_{-\infty}^{-1} \in A^{\leq -1}$ que pertence a τ . Assim, temos o seguinte resultado.

Teorema 2. Se $\bar{\mu}^a(\ell_\tau < +\infty) = 1$ para algum $a_{-\infty}^{-1} \in A^{\leq -1}$, então $\bar{\mu}^a \in \mathcal{M}_T(g)$. Quando existe $a_{-\infty}^{-1}$ tal que $E_{\bar{\mu}^a}[\ell_\tau] < +\infty$, temos $|\mathcal{M}_T(g)| = 1$.

Em palavras, se existe um passado cujo tamanho do contexto é quase-certamente finito, então a cadeia estacionária associada a este passado é uma cadeia de memória de tamanho variável, sendo a única quando esse tamanho tem esperança finita. Até onde sabemos este é o único resultado na literatura específico para esse tipo de cadeia. Uma consequência imediata desse resultado, no que diz respeito a existência, é dado pelo Corolário 3 a seguir.

Corolário 3. Se $\sup_a \mu^a (\ell_\tau (a_{-\infty}^{-1} X_0^{i-1}) > i) \rightarrow 0$ quando $i \rightarrow +\infty$ para toda $a_{-\infty}^{-1} \in A^{\leq -1}$, então $\bar{\mu}^a \in \mathcal{M}_T(g)$.

O próximo resultado (Teorema 3) afirma que, sob algumas suposições, a única medida compatível é β -misturadora, i.e., intuitivamente, eventos no futuro tornam-se independentes do passado fixado. Neste contexto, a métrica que utilizamos para medir a distância entre as distribuições envolvidas é a *distância em variação total*.

Definição 9 (Distância em variação total). A distância em variação total entre duas medidas de probabilidade μ e ν definidas sobre $(A^{\mathbb{Z}}, \mathcal{F})$ é a função $d_{VT} : \mathcal{M} \times \mathcal{M} \rightarrow [0, 1]$ tal que

$$d_{VT}(\mu, \nu) = \sup_{B \in \mathcal{F}} |\mu(B) - \nu(B)|. \quad (3.8)$$

Essa distância é, portanto, a maior diferença absoluta entre as probabilidades avaliadas em um mesmo evento pelas duas medidas.

Definição 10 (β -mistura). Uma cadeia estacionária $\mathbf{X} := \{X_n : n \in \mathbb{Z}\}$ ou, equivalentemente, uma medida $\mu \in \mathcal{M}_T$ é dita ser β -misturadora se, e somente se,

$$\lim_{n \rightarrow +\infty} d_{VT}(\mu|_{\mathcal{F}^{\leq -1}} \otimes \mu|_{\mathcal{F}^{\geq n}}, \mu|_{\mathcal{F}^{\leq -1} \otimes \mathcal{F}^{\geq n}}) = 0, \quad (3.9)$$

em que, para todo $n \in \mathbb{Z}$, $\mathcal{F}^{\geq n}$ é a σ -álgebra gerada pelas ocorrências após o instante n ; $\mathcal{F}^{\leq -1} \otimes \mathcal{F}^{\geq n}$ é sub- σ -álgebra de \mathcal{F} gerada pelos retângulos $A \times B$, em que $A \in \mathcal{F}^{\leq -1}$ e $B \in \mathcal{F}^{\geq n}$; e $\mu|_{\mathcal{F}^{\leq -1}} \otimes \mu|_{\mathcal{F}^{\geq n}}$ é a medida produto entre $\mu|_{\mathcal{F}^{\leq -1}}$ e $\mu|_{\mathcal{F}^{\geq n}}$.

É importante notar que utilizamos a notação $\mu|_{\mathcal{G}}$ para dizer que uma medida μ definida sobre \mathcal{F} esta restrita aos eventos que ocorrem na sub- σ -álgebra \mathcal{G} . Além disso, observamos que uma medida β -misturadora pode ser encontrada na literatura com outras nomenclaturas tais como *absolutamente regular* ou *weak-Bernoulli* (ver, por exemplo, [Berbee \(1986\)](#) e [Bradley et al. \(2005\)](#)).

Teorema 3. Se para qualquer $a_{-\infty}^{-1} \in A^{\leq -1}$ temos

$$\bar{\mu}^a \left(\sum_{i \geq 1} \mathbb{I}\{\ell_\tau > i\} < +\infty \right) = 1,$$

então existe uma única g -medida compatível e ela é β -misturadora.

O último resultado desta seção envolve uma condição explícita baseada no que chamamos de *taxa de crescimento exponencial* de τ . Essa taxa mensura o quão rápido $|\tau^n|$ diverge, em que τ^n é o n -ésimo nível da árvore de contexto τ , i.e.,

$$\tau^n := \{x_{-n}^{-1} : x_{-n}^{-1} \preceq w \text{ para algum } w \in \tau\}.$$

Por exemplo, o terceiro nível da árvore esparsa é o conjunto $\tau^3 = \{100, 000\}$, cuja representação gráfica é dada pela Figura 3.

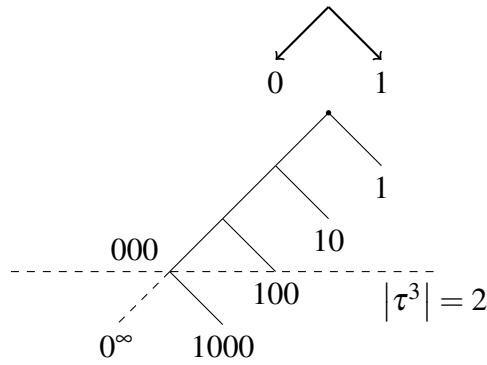


Figura 3 – Terceiro nível da árvore esparsa.

A taxa de crescimento exponencial é, então, definida como

$$\bar{g}r(\tau) := \limsup_n |\tau^n|^{1/n}.$$

Corolário 4. Se $\bar{g}r(\tau) < [1 - (|A| - 1)\varepsilon]^{-1}$, $\varepsilon > 0$, então existe uma única g -medida e ela é β -misturadora.

Este critério foi provado ser uma condição suficiente para existência em [Gallo e Paccaut \(2013\)](#). Aqui, melhoramos este resultado a fim de abranger também a questão da unicidade para o caso de cadeias de memória de tamanho variável.

Exemplo 18. No Exemplo 17, a árvore de contexto τ associada a g -função g é a representada na Figura 4. Neste caso, temos $|\tau^n| = 3 \times 2^{n-1}$ para todo $n \in \mathbb{N}$. Logo, $\bar{g}r(\tau) = 2$ o qual é estritamente menor que $[1 - 2\varepsilon]^{-1}$ se, e somente se, $1/4 < \varepsilon < 1/2$. Portanto, temos a unicidade da g -medida e a β -mistura sempre quando $\varepsilon \in (\frac{1}{4}, \frac{1}{2})$.

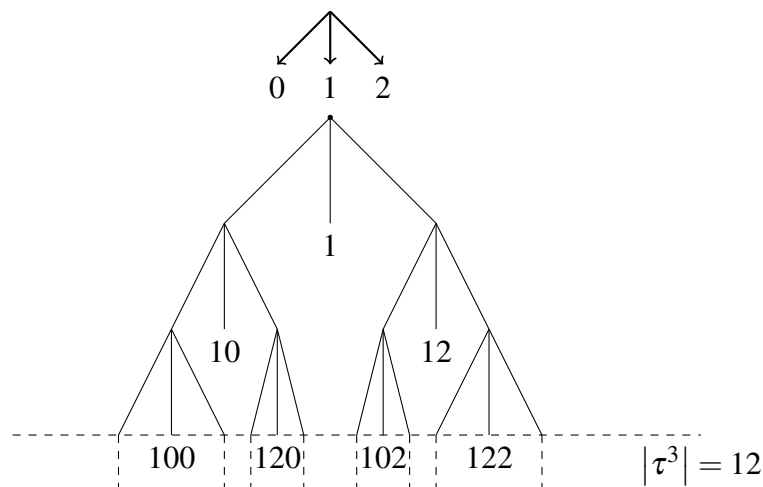


Figura 4 – Árvore de contexto correspondente ao Exemplo 17 com o nível 3 delimitado.

3.3 Demonstrações

Demonstração do Teorema 1. Sejam g uma g -função e $a_{-\infty}^{-1} \in A^{\leq -1}$ um passado fixado. A medida de referência $\bar{\mu}^a$, definida via um limite sobre a média de Césaró (ver, equação (3.5)), é invariante por translação. De fato, para qualquer conjunto mensurável B ,

$$\begin{aligned}
\bar{\mu}^a(T^{-1}B) &= \lim_{k \rightarrow +\infty} \bar{\mu}^{a, n_k}(T^{-1}B) \\
&= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \mu^{a, i}(T^{-1}B) \\
&= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=1}^{n_k} \mu^{a, i}(B) \\
&= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \left(\mu^{a, n_k}(B) - \mu^{a, 0}(B) + \sum_{i=0}^{n_k-1} \mu^{a, i}(B) \right) \\
&= \lim_{k \rightarrow +\infty} \frac{\mu^{a, n_k}(B) - \mu^{a, 0}(B)}{n_k} + \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \mu^{a, i}(B) \\
&= \bar{\mu}^a(B).
\end{aligned}$$

Nos resta, portanto, mostrar que $\bar{\mu}^a \in \mathcal{M}(g)$.

Pelo Teorema Reverso dos Martingales (ver, por exemplo, [Durrett \(2010\)](#)), existe um conjunto $S \subset A^{\leq -1}$ para o qual

$$E_{\bar{\mu}^a} [\mathbb{I}(\{c_0 = b\}) | \mathcal{F}_{-l}^{-1}] (\mathbf{c}) \xrightarrow{l \rightarrow \infty} E_{\bar{\mu}^a} [\mathbb{I}(\{c_0 = b\}) | \mathcal{F}^{\leq -1}] (\mathbf{c}),$$

para $\bar{\mu}^a$ -quase toda $c_{-\infty}^{-1} \in S$. A notação \mathcal{F}_{-l}^{-1} representa a sub- σ -álgebra de \mathcal{F} gerada por todas as ocorrências no intervalo $[-l, -1]$.

Por hipótese, qualquer que seja l suficientemente grande $\bar{\mu}^a([c_{-l}^{-1}]) > 0$ para todo $c_{-\infty}^{-1} \in \mathcal{D}_g^c$. Assim, podemos escrever

$$E_{\bar{\mu}^a} [\mathbb{I}(\{c_0 = b\}) | \mathcal{F}_{-l}^{-1}] (\mathbf{c}) = \frac{\bar{\mu}^a([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^a([c_{-l}^{-1}])},$$

para todo $c_{-\infty}^{-1} \in \mathcal{D}_g^c \cap S$, já que sobre nossas hipóteses $\bar{\mu}^a(\mathcal{D}_g^c \cap S) = 1$.

Portanto,

$$\lim_{l \rightarrow +\infty} \frac{\bar{\mu}^a([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^a([c_{-l}^{-1}])} = E_{\bar{\mu}^a} [\mathbb{I}(\{c_0 = b\}) | \mathcal{F}^{\leq -1}] (\mathbf{c}), \quad \bar{\mu}^a - \text{quase todo } c_{-\infty}^{-1} \in \mathcal{D}_g^c \cap S.$$

Desta forma, para concluirmos a prova do teorema é suficiente mostrar que

$$\lim_{l \rightarrow +\infty} \frac{\bar{\mu}^a([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^a([c_{-l}^{-1}])} = g(c_{-\infty}^{-1}b),$$

para todo $c_{-\infty}^{-1} \in D_g^c \cap S$.

Para $i \geq l$,

$$\begin{aligned} \mu^{a,i}([c_{-l}^{-1}] \cap [b]) &= \mu^a\left(\left[(T^{-i}\mathbf{c})_{i-l}^{i-1}\right] \cap [b]\right) \\ &= \sum_{d_{-i}^{-l-1}} \mu^a\left(\left[(T^{-i}\mathbf{d})_0^{i-l-1}\right] \cap \left[(T^{-i}\mathbf{c})_{i-l}^{i-1}\right] \cap [b]\right) \\ &= \sum_{d_{-i}^{-l-1}} \mu^a\left(\left[(T^{-i}\mathbf{d})_0^{i-l-1}\right]\right) g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-1} b\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \\ &= \sum_{d_{-i}^{-l-1}} \mu^{a,i}\left(\left[d_{-i}^{-l-1}\right]\right) g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-1} b\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \end{aligned}$$

e, como já vimos,

$$\mu^{a,i}([c_{-l}^{-1}]) = \sum_{d_{-i}^{-l-1}} \mu^{a,i}\left(\left[d_{-i}^{-l-1}\right]\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right).$$

Portanto, para k suficientemente grande,

$$\begin{aligned} \bar{\mu}^{a,n_k}([c_{-l}^{-1}] \cap [b]) &= \frac{1}{n_k} \sum_{i=0}^{n_k-1} \mu^{a,i}([c_{-l}^{-1}] \cap [b]) \\ &= \frac{1}{n_k} \left[\sum_{i=l}^{n_k-1} \sum_{d_{-i}^{-l-1}} \mu^{a,i}\left(\left[d_{-i}^{-l-1}\right]\right) g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-1} b\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) + o(n_k) \right] \end{aligned} \quad (3.10)$$

e

$$\bar{\mu}^{a,n_k}([c_{-l}^{-1}]) = \frac{1}{n_k} \left[\sum_{i=l}^{n_k-1} \sum_{d_{-i}^{-l-1}} \mu^{a,i}\left(\left[d_{-i}^{-l-1}\right]\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) + o(n_k) \right]. \quad (3.11)$$

Uma vez que $c_{-\infty}^{-1} \in \mathcal{D}_g^c$ temos

$$\text{var}_l^g(c_{-\infty}^{-1}) = \sup \left\{ |g(d_{-\infty}^{-1} b) - g(c_{-\infty}^{-1} b)| : d_{-l}^{-1} = c_{-l}^{-1} \right\} \xrightarrow{l \rightarrow \infty} 0,$$

e, portanto, para $a_{-\infty}^{-1}$ e d_{-i}^{-l-1}

$$\left| g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-1} b\right) - g\left(c_{-\infty}^{-1} b\right) \right| \leq \text{var}_l^g(c_{-\infty}^{-1}). \quad (3.12)$$

Por simplicidade, vamos introduzir a notação

$$A(k, l; a_{-\infty}^{-1}, c_{-\infty}^{-1}) = \sum_{i=l}^{n_k-1} \sum_{d_{-i}^{-l-1}} \mu^{a,i}\left(\left[d_{-i}^{-l-1}\right]\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right).$$

Aplicando as igualdades (3.10) e (3.11) e a desigualdade (3.12), temos, por um lado,

$$\begin{aligned} \frac{\bar{\mu}^{a,n_k}([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^{a,n_k}([c_{-l}^{-1}])} &= \\ &= \frac{\frac{1}{n_k} \left[\sum_{i=l}^{n_k-1} \sum_{d_{-i}^{-l-1}} \mu^{a,i} \left([d_{-i}^{-l-1}] \right) g \left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-1} b \right) \prod_{k=0}^{l-1} g \left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k} \right) + o(n_k) \right]}{\frac{1}{n_k} [A(k,l; a_{-\infty}^{-1}, c_{-\infty}^{-1}) + o(n_k)]} \\ &\geq \frac{\frac{1}{n_k} [(g(c_{-\infty}^{-1}b) - \text{var}_l^g(c_{-\infty}^{-1})) A(k,l; a_{-\infty}^{-1}, c_{-\infty}^{-1}) + o(n_k)]}{\frac{1}{n_k} [A(k,l; a_{-\infty}^{-1}, c_{-\infty}^{-1}) + o(n_k)]} \end{aligned}$$

e, por outro lado,

$$\frac{\bar{\mu}^{a,n_k}([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^{a,n_k}([c_{-l}^{-1}])} \leq \frac{\frac{1}{n_k} [(g(c_{-\infty}^{-1}b) + \text{var}_l^g(c_{-\infty}^{-1})) A(k,l; a_{-\infty}^{-1}, c_{-\infty}^{-1}) + o(n_k)]}{\frac{1}{n_k} [A(k,l; a_{-\infty}^{-1}, c_{-\infty}^{-1}) + o(n_k)]}.$$

Assim, ao tomarmos $k \rightarrow +\infty$ obtemos

$$g(c_{-\infty}^{-1}b) - \text{var}_l^g(c_{-\infty}^{-1}) \leq \frac{\bar{\mu}^a([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^a([c_{-l}^{-1}])} \leq g(c_{-\infty}^{-1}b) + \text{var}_l^g(c_{-\infty}^{-1}).$$

Logo, para $c_{-\infty}^{-1} \in \mathcal{D}_g^c \cap \mathcal{S}$ temos

$$\frac{\bar{\mu}^a([c_{-l}^{-1}] \cap [b])}{\bar{\mu}^a([c_{-l}^{-1}])} \xrightarrow{l \rightarrow \infty} 0.$$

Por hipótese, $\bar{\mu}^a(\mathcal{D}_g \cap \mathcal{S}) = 1$, portanto para $\bar{\mu}^a$ -quase toda $c_{-\infty}^{-1} \in A^{\leq -1}$ e $b \in A$

$$E_{\bar{\mu}^a} [\mathbb{I}(\{c_0 = b\} | \mathcal{F}^{\leq -1})] (\mathbf{c}) = \mu(b | \mathcal{F}^{\leq -1}) (\mathbf{c}) = g(c_{-\infty}^{-1}b).$$

□

Demonstração do Corolário 1. Uma vez que g é uniformemente não-nula, existe $\varepsilon \in (0, \frac{1}{2})$ tal que $\varepsilon \leq g \leq 1 - \varepsilon$. Portanto para qualquer $c_{-\infty}^{-1} \in A^{\leq -1}$ temos, para $i > l$,

$$\begin{aligned} \mu^{a,i}([c_{-l}^{-1}]) &= \mu^a \left(\left[(T^{-i} \mathbf{c})_{i-l}^{i-1} \right] \right) \\ &= \sum_{d_{-i}^{-l-1}} \mu^a \left(\left[(T^{-i} \mathbf{d})_0^{i-l-1} \right] \cap \left[(T^{-i} \mathbf{c})_{i-l}^{i-1} \right] \right) \\ &= \sum_{d_{-i}^{-l-1}} \mu^a \left(\left[(T^{-i} \mathbf{d})_0^{i-l-1} \right] \right) \prod_{k=0}^{l-1} g \left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k} \right) \\ &\geq \varepsilon^l \sum_{d_{-i}^{-l-1}} \mu^{a,i} \left([d_{-i}^{-l-1}] \right) = \varepsilon^l > 0. \end{aligned} \tag{3.13}$$

e

$$\begin{aligned}
\mu^{a,i}([c_{-l}^{-1}]) &= \mu^a\left(\left[(T^{-i}\mathbf{c})_{i-l}^{i-1}\right]\right) \\
&= \sum_{d_{-i}^{-\ell-1}} \mu^a\left(\left[(T^{-i}\mathbf{d})_0^{i-l-1}\right] \cap \left[(T^{-i}\mathbf{c})_{i-l}^{i-1}\right]\right) \\
&= \sum_{d_{-i}^{-\ell-1}} \mu^a\left(\left[(T^{-i}\mathbf{d})_0^{i-l-1}\right]\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \\
&\leq (1-\varepsilon)^l \sum_{d_{-i}^{-\ell-1}} \mu^{a,i}\left(\left[d_{-i}^{-l-1}\right]\right) = (1-\varepsilon)^l,
\end{aligned}$$

o qual implica em

$$0 < \varepsilon^l \leq \bar{\mu}^{a,n_k}([c_{-l}^{-1}]) := \frac{1}{n_k} \sum_{i=0}^{n_k-1} \mu^{a,i}([c_{-l}^{-1}]) \leq (1-\varepsilon)^l.$$

Assim,

$$0 < \varepsilon^l \leq \bar{\mu}^a([c_{-l}^{-1}]) := \lim_{k \rightarrow +\infty} \bar{\mu}^{a,n_k}([c_{-l}^{-1}]) \leq (1-\varepsilon)^l$$

e, portanto, $\bar{\mu}^a([c_{-\infty}^{-1}]) = 0$.Como $\mathcal{D}_g \subset A^{\leq -1}$ e, por hipótese, \mathcal{D}_g é enumerável, podemos escrever

$$\bar{\mu}^a(\mathcal{D}_g) = \sum_{c_{-\infty}^{-1} \in \mathcal{D}_g} \bar{\mu}^a([c_{-\infty}^{-1}]) = 0.$$

□

Demonstração do Corolário 2. Vamos provar este resultado para o caso em que a sequência finita $v_{-|v|}^{-1}$ é composta por um único símbolo, isto $v \in A$. A extensão para o caso de sequências com mais símbolos é similar. Observe que, para qualquer $l \geq 1$, $\mathcal{D}_g^l \subset (A \setminus \{v\})^l$, em que \mathcal{D}_g^l é o conjunto das sequências finitas de tamanho l composta por símbolos de $A \setminus \{v\}$. Logo,

$$\bar{\mu}^a(\mathcal{D}_g) = \lim_{l \rightarrow +\infty} \bar{\mu}^a(\mathcal{D}_g^l) \leq \lim_{l \rightarrow +\infty} \sum_{c_{-l}^{-1} \in (A \setminus \{v\})^l} \bar{\mu}^a([c_{-l}^{-1}]). \quad (3.14)$$

Observe que,

$$\begin{aligned}
\sum_{c_{-l}^{-1} \in (A \setminus \{v\})^l} \bar{\mu}^a([c_{-l}^{-1}]) &:= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \sum_{c_{-l}^{-1} \in (A \setminus \{v\})^l} \mu^{a,i}([c_{-l}^{-1}]) \\
&= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \sum_{c_{-l}^{-1} \in (A \setminus \{v\})^l} \sum_{d_{-i}^{-\ell-1}} \mu^a\left(\left[(T^{-i}\mathbf{d})_0^{i-l-1}\right]\right) \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \\
&= \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \left[\sum_{d_{-i}^{-\ell-1}} \mu^a\left(\left[d_{-i}^{-l-1}\right]\right) \sum_{c_{-l}^{-1} \in (A \setminus \{v\})^l} \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \right],
\end{aligned}$$

em que

$$\begin{aligned}
\sum_{c_{-l}^{-1} \in (|A| \setminus \{v\})^l} \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) &= \tag{3.15} \\
&= \sum_{c_{-l}^{-2} \in (|A| \setminus \{v\})^{l-1}} \prod_{k=0}^{l-2} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \sum_{c_{-1} \in A \setminus \{v\}} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-1}\right) \\
&= \sum_{c_{-l}^{-2} \in (|A| \setminus \{v\})^{l-1}} \prod_{k=0}^{l-2} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \left[1 - g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-2} v\right)\right] \\
&\leq (1 - \varepsilon) \sum_{c_{-l}^{-2} \in (|A| \setminus \{v\})^{l-1}} \prod_{k=0}^{l-2} g\left(a_{-\infty}^{-i-1} d_{-i}^{-l-1} c_{-l}^{-l+k}\right) \\
&\leq (1 - \varepsilon)^l.
\end{aligned}$$

A última desigualdade é obtida aplicando a mesma estratégia l vezes. Desta forma,

$$\sum_{c_{-l}^{-1} \in (|A| \setminus \{v\})^l} \bar{\mu}^a(c_{-l}^{-1}) \leq \lim_{k \rightarrow +\infty} \frac{1}{n_k} \sum_{i=0}^{n_k-1} (1 - \varepsilon)^l \leq (1 - \varepsilon)^l.$$

Portanto, por (3.14), $\bar{\mu}^a(\mathcal{D}_g) = 0$.

□

Demonstração da Proposição 1 Procedendo exatamente como na demonstração do Corolário 2 é suficiente provar que $\sum_{d_{-l}^{-1} \in \mathcal{D}_g^l} \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-1} b_0^{j-l} c_{-l}^{-l+k}\right)$ é zero quando l diverge para algum $a_{-\infty}^{-1} \in A^{\leq -1}$. Para qualquer δ , temos para qualquer $a_{-\infty}^{-1}$ e l suficientemente grande

$$\begin{aligned}
\sum_{d_{-l}^{-1} \in \mathcal{D}_g^l} \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-1} b_0^{j-l} c_{-l}^{-l+k}\right) &\leq \sum_{d_{-l}^{-1} \in \mathcal{D}_g^l} \sup_{a_{-\infty}^{-1}} \prod_{k=0}^{l-1} g\left(a_{-\infty}^{-1} b_0^{j-l} c_{-l}^{-l+k}\right) \\
&\leq \left(e^{P_g(\mathcal{D}_g) + \delta}\right)^l.
\end{aligned}$$

Usando a suposição da proposição nós sabemos que $P_g(\mathcal{D}_g) < 0$, portanto, escolhendo $\delta < -P_g(\mathcal{D}_g)$ obtemos

$$\lim_l \left(e^{P_g(\mathcal{D}_g) + \delta}\right)^l = 0$$

concluindo a prova da Proposição.

□

Demonstração do Teorema 2. Note que $\mathcal{D}_g \subset \{c_{-\infty}^{-1} \in A^{\leq -1} : \ell_\tau(c_{-\infty}^{-1}) = +\infty\}$. Como, por hipótese, $\bar{\mu}^a(\ell_\tau < +\infty) = 1$, temos

$$\bar{\mu}^a(\mathcal{D}_g) \leq \bar{\mu}^a(\ell_\tau = +\infty) = 1 - \bar{\mu}^a(\ell_\tau < +\infty) = 0,$$

o que prova que $\bar{\mu}^a \in \mathcal{M}_T(g)$.

Agora, para provar a unicidade de $\bar{\mu}^a$, observamos que para uma árvore de contexto τ , temos

$$\text{var}_n^g(c_{-\infty}^{-1}) \leq \mathbb{I}(\{\ell_\tau(c_{-\infty}^{-1}) > n\}),$$

qualquer que seja $c_{-\infty}^{-1} \in A^{\leq -1}$. Johansson e Öberg (2008) provaram que, se para alguma g -medida μ temos

$$E_\mu \left[\sum_{n \geq 1} (\text{var}_n^g)^2 \right] < +\infty,$$

então essa é a única g -medida. Assim, pelo teorema da convergência monótona (ver, por exemplo, Durrett (2010))

$$\begin{aligned} E_{\bar{\mu}^a} \left[\sum_{n \geq 1} (\text{var}_n^g)^2 \right] &= \lim_{k \rightarrow +\infty} \sum_{n=0}^k E_{\bar{\mu}^a} \left[(\text{var}_n^g)^2 \right] \\ &\leq \lim_{k \rightarrow +\infty} \sum_{n=0}^k E_{\bar{\mu}^a} [\mathbb{I}(\{\ell_\tau > n\})] \\ &= \lim_{k \rightarrow +\infty} \sum_{n=0}^k \bar{\mu}^a(\ell_\tau > n) \\ &= E_{\bar{\mu}^a}[\ell_\tau] < +\infty. \end{aligned}$$

Portanto, $|\mathcal{M}_T(g)| = 1$.

□

Demonstração do Corolário 3. Seja $a_{-\infty}^{-1} \in A^{\leq -1}$ tal que $\mu^a(\ell_\tau > i) \xrightarrow{i \rightarrow \infty} 0$. Então,

$$\begin{aligned} \bar{\mu}^a(\mathcal{D}_g) &\leq \bar{\mu}^a(\ell_\tau = +\infty) \\ &= \lim_{i \rightarrow +\infty} \lim_{k \rightarrow +\infty} \frac{1}{m_k} \sum_{j=0}^{m_k-1} \mu^{a,j}(\ell_\tau > i) \\ &= \lim_{i \rightarrow +\infty} \lim_{k \rightarrow +\infty} \frac{1}{m_k} \sum_{j=0}^{m_k-1} \sup_{a_{-\infty}^{-1} \in A^{\leq -1}} \mu^a(\ell_\tau > i) \xrightarrow{i \rightarrow \infty} 0. \end{aligned}$$

Portanto, pelo Teorema 1, $\mathcal{M}_T(g) \neq \emptyset$.

□

Demonstração do Teorema 3. Supor que para qualquer sequência infinita a esquerda $a_{-\infty}^{-1}$ de símbolos de A temos

$$\bar{\mu}^a \left(\sum_{i \geq 1} \mathbb{I}\{\ell_\tau > i\} < +\infty \right) = 1$$

implica que para algum $a_{-\infty}^{-1}$, $\bar{\mu}^a(\ell_\tau > i) \xrightarrow{i \rightarrow +\infty} 0$ e, portanto, pelo Corolário 3, $\mathcal{M}_T(g) \neq \emptyset$.

Agora, para provar que μ^a é β -misturadora, começamos observando que para qualquer $a_{-\infty}^{-1}$ e $c_{-\infty}^{-1}$ em $A^{\leq -1}$ temos

$$|g(a_{-\infty}^{-1}d_0^{i-1}b) - g(c_{-\infty}^{-1}d_0^{i-1}b)| \leq \mathbb{I}\{\ell_\tau(x_{-\infty}^{-1}) > n\}, \quad (3.16)$$

para todo $x_{-\infty}^{-1} \in A^{\leq -1}$ tal que $x_{-i}^{-1} = d_0^{i-1}$. Portanto, se $\sum_{i \geq 1} \mathbb{I}\{\ell_\tau(x_{-\infty}^{-1}) > i\} < +\infty$ para $\bar{\mu}^a$ -quase toda $x_{-\infty}^{-1}$, a desigualdade (3.16) nos garante que

$$\sum_{i \geq 1} \sum_{b \in A} |g(a_{-\infty}^{-1}d_0^{i-1}b) - g(c_{-\infty}^{-1}d_0^{i-1}b)| < +\infty,$$

o que implica, pelo Teorema 2(iii) e Teorema 1(i) de Gallesco, Gallo e Takahashi (2018) que existe uma única medida invariante por translação, a qual denotamos por μ , e além disso,

$$\lim_{i \rightarrow +\infty} \sup_{B \in \mathcal{F}^{\geq i}} |\mu^a(B) - \mu^c(B)| = 0$$

para todo $a_{-\infty}^{-1}$ e $c_{-\infty}^{-1}$ em $A^{\leq -1}$ em que $\mathcal{F}^{\geq i}$ é a σ -álgebra gerada pelo processo \mathbf{X} após o instante i . Logo,

$$\lim_{i \rightarrow +\infty} \sup_{B \in \mathcal{F}^{\geq i}} |\mu^a(B) - \mu(B)| = 0,$$

desde que

$$\sup_{B \in \mathcal{F}^{\geq i}} |\mu^a(B) - \mu(B)| \leq \int_{A^{\leq -1}} \sup_{B \in \mathcal{F}^{\geq i}} |\mu^a(B) - \mu^b(B)| \mu(db).$$

Portanto, pelo Corolário 2.8(ii) de Tong, Handel *et al.* (2014) concluímos que μ é também β -misturadora.

□

Demonstração do Corolário 4. De acordo com o Corolário 1 de Gallo e Paccaut (2013) se $\bar{g}r(\tau) < [1 - (|A| - 1)\varepsilon]^{-1}$, $0 < \varepsilon < 1/2$, então $\mathcal{M}_T(g) \neq \emptyset$. Nos resta então provar as outras afirmações, i.e., unicidade e β -mistura da medida invariante compatível com g . Neste sentido, observamos que $\tau^n := \{x_{-|x|}^{-1} : |x| \leq n\} \cup \{x_{-n}^{-1} : \ell_\tau(x_{-\infty}^{-1}) = +\infty\} \subset \mathcal{D}_g^n$, então é suficiente provar que

$$\limsup_n |\tau^n|^{1/n} \leq [1 - (|A| - 1)\varepsilon]^{-1} \Rightarrow \sum_{i \geq 1} \sup_{a_{-\infty}^{-1}} \bar{\mu}^a(\ell_\tau(x_{-\infty}^{-1}) > i) < +\infty, \quad (3.17)$$

uma vez que o último termo implica, pelo Teorema de Borel-Cantelli (ver, por exemplo, Durrett (2010)), em $\sum_{i \geq 1} \mathbb{I}\{\ell_\tau(x_{-\infty}^{-1}) > i\} < +\infty$ para $\bar{\mu}^a$ -quase toda $x_{-\infty}^{-1} \in A^{\leq -1}$. Assim, pelo Teorema 3, concluiríamos que existe uma única medida invariante compatível com g e essa medida é β -misturadora.

Neste sentido, vamos provar (3.17). Observe que, por hipótese, para qualquer $0 < \alpha < 1$ existe N_α tal que para $i > N_\alpha$ temos $|\mathcal{T}^n| \leq [1 - (|A| - 1)\varepsilon]^{-n(1-\alpha)}$. Portanto,

$$\begin{aligned}
\bar{\mu}^a(\ell_\tau(x_{-\infty}^{-1}) > i) &= \sum_{n>i} \sum_{v \in \mathcal{T}^n} \bar{\mu}^a(v) \\
&\leq \sum_{n>i} \sum_{v \in \mathcal{T}^n} [1 - (|A| - 1)\varepsilon]^n \\
&= \sum_{n>i} |\mathcal{T}^n| [1 - (|A| - 1)\varepsilon]^n \\
&\leq \sum_{n>i} [1 - (|A| - 1)\varepsilon]^{-n(1-\alpha)} [1 - (|A| - 1)\varepsilon]^n \\
&= \sum_{n>i} [1 - (|A| - 1)\varepsilon]^{n\alpha} \\
&= \left(\frac{1-\varepsilon}{\varepsilon}\right) [1 - (|A| - 1)\varepsilon]^{i\alpha}.
\end{aligned}$$

Observe que a última quantidade é independente de $a_{-\infty}^{-1}$ e, portanto,

$$\sum_{i \geq 1} \sup_{a_{-\infty}^{-1}} \bar{\mu}^a(\ell_\tau(x_{-\infty}^{-1}) > i) \leq N_\alpha + \sum_{i > N_\alpha} \left(\frac{1-\varepsilon}{\varepsilon}\right) [1 - (|A| - 1)\varepsilon]^{i\alpha} < +\infty.$$

□

LIMITES DA CONCENTRAÇÃO PARA INFORMAÇÃO DIRIGIDA EMPÍRICA

Entender como os sistemas neuronais codificam, computam e transmitem informações é fundamental para entender a função cerebral. Portanto, é natural analisar a interação entre componentes do cérebro usando a teoria da informação, uma ferramenta estatística projetada para quantificar informações. Inspirados pelo trabalho de *Cai et al. (2017)* que utilizam a informação dirigida para medir as influências causais entre sequências de disparos neuronais a fim de inferir conexões sinápticas, é que apresentamos, neste capítulo, a terceira contribuição desta tese: limites da concentração para a estimação da informação dirigida no contexto de cadeias estacionárias, com valores em alfabetos finitos, cuja distribuição é uma g -medida. O método utilizado para obter os limites da concentração é baseado na desigualdade exponencial para funções Lipschitz.

4.1 Motivação

Neurociência é o estudo científico do sistema nervoso. Pesquisas nesta área tem crescido significativamente durante a segunda metade do século XX, principalmente devido aos avanços em biologia molecular e neurociência computacional. Este fato produziu uma grande quantidade de dados e uma série de novos problemas foram revelados, os quais tem permitido estudar o sistema nervoso em todos os seus aspectos.

Neurônios são células excitáveis especializadas em comunicação, i.e., elas são capazes de se comunicar com outros neurônios e outros tipos de células através de junções especializadas chamadas sinapses, nas quais sinais elétricos podem ser transmitidos de uma célula para outra. Esta atividade é manifestada por variações locais do seu potencial de membrana, em outras palavras, variações da diferença em potencial elétrico entre o ambiente intracelular e extracelular. Essas variações são chamadas potenciais de ação e são causadas pela mudança de ions, que se movem através da membrana, da região mais ionizada para a menos ionizada. Potenciais de

ação em neurônios são também conhecidos como impulsos nervosos ou disparos, e a sequência temporal dos potenciais de ação gerados por um neurônio é chamada de trem de disparo neuronal. Na Figura 5 (de Moya (2011)) é possível ver como a propagação de disparos ocorre ao longo da estrutura de um neurônio, dos dendritos até o soma (corpo celular), e do soma até a sinapse, via o axon. O sistema nervoso, conseqüentemente, emerge dessas conexões entre neurônios.

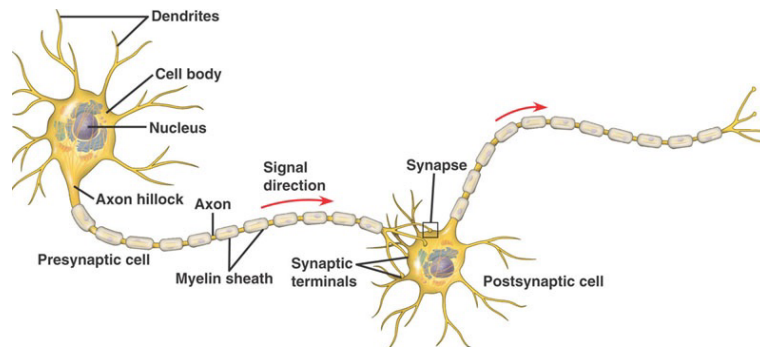


Figura 5 – Propagação dos disparos ao longo da estrutura de um neurônio (MOYA, 2011).

Ferramentas matemáticas foram empenhadas para entender e acompanhar o progresso da neurociência. Neste sentido, um modelo de neurônio é uma equação ou um conjunto de equações que descrevem a evolução do potencial de membrana do neurônio e, possivelmente, variáveis adicionais, de forma que a modelagem é biologicamente plausível, matematicamente tratável e numericamente eficiente (CESSAC, 2010). Aparentemente, dados neuronais assumem certa aleatoriedade. De acordo com Buesing *et al.* (2011), dados experimentais sugerem que neurônios, sinapses e sistemas neurais são inerentemente estocásticos, uma vez que os estímulos externos interferem apenas na atividade altamente estocástica do disparo espontâneo de redes corticais de neurônios. 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 a descrição matemática e o tratamento de fenômenos neuronais.

A atividade neuronal é manifestada pela emissão de sequências de disparos elétricos geradas por excitações vindo de outros neurônios ou de fontes externas. Neste sentido, podemos considerar uma sequência temporal determinística e finita $\{t_0, t_1, \dots, t_n\}$ para os instantes de ocorrência dos disparos em um dado intervalo de tempo $(0, T] \subset \mathbb{R}$, em que $0 \leq t_0 < t_1 < \dots < t_n \leq T$. Além disso, é útil considerar os tempos de ocorrência de disparos com um certo grau de acurácia Δt . Para um Δt suficientemente pequeno, geralmente da ordem de milissegundos (*ms*), a sequência de disparos pode ser representada por uma sequência binária \mathbf{x} de tamanho $T/\Delta t$ composta de 0's e 1's, tal que, para cada $i = 0, 1, \dots, n$, $x_i = 1$ se um disparo no intervalo $(i\Delta t, (i+1)\Delta t]$ foi observado, e $x_i = 0$ se não observou-se disparo. Esta sequência de 0's e 1's é o que chamamos de trem de disparos neuronal. Na Figura 6 (de London, Shreibman e Segev (2002)) é possível ver a atividade neuronal representada sobre um tempo discretizado com $\Delta t = 25ms$, de forma que a sequência binária representa um trem de disparos neuronal.

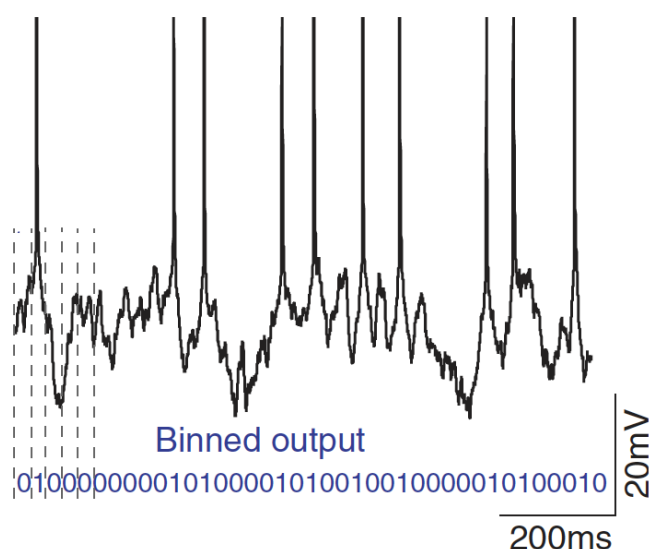


Figura 6 – Atividade neuronal sobre um tempo discretizado com $\Delta t = 25ms$.

Neste contexto, seja \mathcal{S} um conjunto enumerável de neurônios, assim para cada neurônio $i \in \mathcal{S}$ consideramos uma variável aleatória binária $X_t^{(i)}$ para indicar se o neurônio i dispara ou não em Δt . Neste trabalho, estamos interessados em quantificar a informação que flui de um neurônio para outro, então vamos considerar $|\mathcal{S}| = 2$ e denotamos os processos estocásticos associados a cada uma das sequências de disparos neuronais como $\mathbf{X} := \{X_t : t \in \mathbb{Z}\}$ e $\mathbf{Y} := \{Y_t : t \in \mathbb{Z}\}$, respectivamente. Na neurociência, tais processos estocásticos tem sido utilizados envolvendo uma descrição Markoviana ou como uma cadeia de memória de tamanho variável. A primeira descrição tem se mostrado inadequada, como mostram os trabalhos de [Friston et al. \(2010\)](#); [Truccolo, Hochberg e Donoghue \(2010\)](#) e [Cessac \(2011b\)](#), enquanto modelos com memória de tamanho variável tem sido bastante usados recentemente, como é afirmado em [Galves e Locherbach \(2015\)](#)

Um sistema neuronal biológico tem a seguinte característica. É um sistema com um alto número de componentes interagindo (aproximadamente 10^{11}), os neurônios [...]. A probabilidade de disparo de um dado neurônio depende da atividade acumulada do sistema após seu último tempo de disparo. Isto implica que o sistema é não-Markoviano. A evolução temporal de cada neurônio parece uma cadeia estocástica de memória de tamanho variável, mesmo a influência do passado sendo, na verdade, de ordem infinita.

As medidas baseadas em entropia começam a ser aplicadas em neurociência logo após o trabalho de [Shannon \(1948\)](#) e vem sendo utilizadas desde então. No recente trabalho de [Cordeiro et al. \(2019\)](#), os autores apresentam um estudo sistemático sobre as principais contribuições da teoria da informação na neurociência, as quais permitem desde o estudo da capacidade de

transmissão de informação entre as células nervosas até a inferência da conectividade de redes neurais biológicas através da gravação dos impulsos elétricos.

Neste trabalho, vamos, então, considerar \mathbf{X} e \mathbf{Y} como sendo com memória de tamanho variável e a Informação Dirigida, que é uma medida baseada em entropia, será utilizada para inferir a quantidade de informação que flui entre dois disparos neuronais, de forma que com *informação* queremos dizer tudo que reduz a incerteza. Em outras palavras, quanto mais informação, mais padronizada ou estruturada a sequência tende a ser.

Enquanto a atividade neuronal pode ser observada diretamente, as interações entre as estruturas neuronais podem somente ser inferidas a partir dos dados. [Cai et al. \(2017\)](#) propõem estimar a informação dirigida em pequenas redes neurais a partir de árvores de contexto maximizadas (ver, [O'Neill et al. \(2012\)](#)) com base no estimador KT das probabilidades condicionais (ver, [Krichevsky e Trofimov \(1981\)](#)). Neste trabalho, entretanto, realizamos a estimação da informação dirigida a partir de estimadores empíricos. Primeiro, por ser um estimador simples e intuitivo, segundo porque a informação dirigida é uma função da entropia cujas propriedades estatísticas foram bastante estudadas (ver, por exemplo, [Chazottes e Gabrielli \(2005\)](#) e [Chazottes e Maldonado \(2011\)](#)).

4.2 Medidas baseadas na Entropia de Shannon

Entropia é uma medida que quantifica a desordem de um sistema e/ou a incerteza de um evento. Do ponto de vista físico, o conceito de entropia está associado tanto ao grau de desorganização da matéria quanto à tendência dessa desorganização. Shannon, por outro lado, define a entropia como a medida da informação de uma dada mensagem, a qual está relacionada com a frequência dos símbolos que foram transmitidos ([SHANNON, 1948](#)).

Seja μ uma medida invariante por translação compatível com uma g -função g e \mathbf{X} o processo estacionário correspondente. A *entropia de \mathbf{X} associada a blocos de tamanho k* é definida por

$$H\left(X_1^k\right) := - \sum_{a_1^k \in A^k} \mu\left(\left[a_1^k\right]\right) \log \mu\left(\left[a_1^k\right]\right),$$

Essa medida captura quantitativamente as correlações de alcances menores que k , em contraste com a entropia simples $H(X) = H(X_1)$ a qual é sensível apenas à frequências dos diferentes símbolos de A .

Na literatura, um dos parâmetro que temos interesse é a *taxa da entropia*, cuja existência, para processos estacionários, é garantida a partir de propriedades da *entropia condicional*, a qual fornece uma ferramenta conveniente para estabelecer alguns resultados essenciais à teoria desenvolvida neste capítulo. A *entropia condicional de \mathbf{X} associada a blocos de tamanho k* é

dada por

$$H\left(X_k \mid X_1^{k-1}\right) := - \sum_{a_1^k \in A^k} \mu\left(\left[a_1^k\right]\right) \log \frac{\mu\left(\left[a_1^k\right]\right)}{\mu\left(\left[a_1^{k-1}\right]\right)}.$$

Note que $H\left(X_k \mid X_1^{k-1}\right) = H\left(X_1^k\right) - H\left(X_1^{k-1}\right)$ para todo $k \geq 2$. Como μ é uma g -medida e $0 \leq H\left(X_k \mid X_1^{k-1}\right) \leq H\left(X_{k-1} \mid X_1^{k-2}\right)$, a existência da taxa de entropia $h(X)$ é garantida como sendo o limite

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

Um objetivo é entender, a partir da gravação dos impulsos elétricos gerados por dois neurônios, as relações casuais existentes entre os disparos observados. Segundo [Wiener \(1956\)](#) uma melhoria na previsão do futuro de um processo estocástico \mathbf{Y} a partir de informações do passado de um processo \mathbf{X} é vista como um indicativo de uma interação casual de \mathbf{X} para \mathbf{Y} . Essas interações casuais através das estruturas cerebrais são chamadas de *conectividades efetivas* ([FRISTON, 1994](#)).

Várias medidas de relações casuais existem. Estamos interessados na *informação dirigida*, uma medida da Teoria da Informação proposta por [Massey \(1990\)](#) que mensura a conectividade efetiva. Se \mathbf{X} e \mathbf{Y} são dois processos estocásticos, então a *informação dirigida de \mathbf{X} para \mathbf{Y} associada a blocos de tamanho k* é definida como

$$I\left(X_1^k \rightarrow Y_1^k\right) := \sum_{j=1}^k \left[H\left(Y_j \mid Y_1^{j-1}\right) - H\left(Y_j \mid Y_1^{j-1} X_1^j\right) \right],$$

em que $H\left(Y_1 \mid Y_1^0\right) := H\left(Y_1\right)$, $H\left(Y_1 \mid Y_1^0 X_1\right) = H\left(Y_1 \mid X_1\right)$ e

$$H\left(Y_j \mid Y_1^{j-1}, X_1^j\right) := - \sum_{\substack{a_1^j \in A^j \\ b_1^j \in A^j}} \nu\left(\left[a_1^j\right] \cap \left[b_1^j\right]\right) \log \frac{\nu\left(\left[a_1^j\right] \cap \left[b_1^j\right]\right)}{\nu\left(\left[a_1^j\right] \cap \left[b_1^{j-1}\right]\right)},$$

com ν sendo a distribuição estacionária do processo bivariado $(\mathbf{Y}, \mathbf{X}) := \{(Y_n, X_n) : n \in \mathbb{N}\}$. Portanto, $I\left(X_1^k \rightarrow Y_1^k\right)$ mensura a quantidade de informação que flui de uma sequência de disparos neuronais à outra e, naturalmente, incorpora informações direcionais e dinâmicas, pois é assimétrica e baseada em probabilidades de transição.

Uma das contribuições deste trabalho é a proposta de um limite superior para probabilidade de concentração de um estimador da informação dirigida em torno da sua taxa. A *taxa da entropia de informação dirigida* é dada por

$$\lim_{k \rightarrow \infty} \frac{I\left(X_1^k \rightarrow Y_1^k\right)}{k} = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{j=1}^k \left[H\left(Y_j \mid Y_1^{j-1}\right) - H\left(Y_j \mid Y_1^{j-1} X_1^j\right) \right] = h(Y) - h(Y \mid X) := I(X \rightarrow Y),$$

em que

$$h(Y \mid X) = \lim_{k \rightarrow \infty} H\left(Y_k \mid Y_1^{k-1}, X_1^{k-1}\right),$$

é a *taxa de entropia condicional* (ver, por exemplo, [Haruna e Nakajima \(2013\)](#)).

4.3 Estimadores Empíricos

Dada uma amostra $x_1^n := (x_1, \dots, x_n) \in A^n$ de um processo estacionário a tempo discreto $\mathbf{X} := \{X_n : n \in \mathbb{Z}\}$, definimos as medidas empíricas por

$$\hat{\mu}_k(a_1^k | x_1^n) := \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} I(x_i^{i+k-1} = a_1^k), \quad k \in \mathbb{N}.$$

Em outras palavras, $\hat{\mu}_k$ é obtida deslizando uma janela de tamanho k ao longo de x_1^n e contando o número de vezes que a_1^k é visto nesta janela e depois dividindo esse valor pelo número máximo de vezes em que a_1^k pode ser observada em x_1^n . Se x_1^n e $k \leq n$ estão fixados, então a frequência relativa define uma medida de probabilidade $\hat{\mu}_k(\cdot | x_1^n)$ sobre A^k , chamada *distribuição empírica da sobreposição de blocos de tamanho k* .

De forma análoga, dada uma amostra $\mathbf{z}_1^n = (x_1^n, y_1^n) = ((x_1, y_1), \dots, (x_n, y_n)) \in A^n \times A^n$ de um processo bivariado estacionário a tempo discreto $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) := \{(X_n, Y_n) : n \in \mathbb{Z}\}$, definimos as medidas empíricas por

$$\hat{\nu}_{k,k}(a_1^k b_1^k | x_1^n y_1^n) := \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} I(x_i^{i+k-1} = a_1^k, y_i^{i+k-1} = b_1^k), \quad k \in \mathbb{N}.$$

No caso em que \mathbf{z}_1^n e $k \leq n$ estão fixados, $\hat{\nu}_{k,k}(\cdot | \mathbf{z}_1^n)$ define uma medida de probabilidade sobre $A^k \times A^k$, o qual denominamos *distribuição empírica da sobreposição de dois blocos de tamanho k* .

Agora podemos definir os seguintes estimadores empíricos.

Definição 11. Seja $x_1^n \in A^n$ e $y_1^n \in A^n$ amostras dos processos estacionários a tempo discreto $\mathbf{X} := \{X_n : n \in \mathbb{Z}\}$ e $\mathbf{Y} := \{Y_n : n \in \mathbb{Z}\}$, respectivamente. A *entropia empírica associada a blocos de tamanho k* é definida como

$$\hat{H}_{x_1^n}(X_1^k) := - \sum_{a_1^k \in A^k} \hat{\mu}_k(a_1^k | x_1^n) \log \hat{\mu}_k(a_1^k | x_1^n).$$

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

$$\hat{H}_{x_1^n}(X_k | X_1^{k-1}) := - \sum_{a_1^k \in A^k} \hat{\mu}_k(a_1^k | x_1^n) \log \frac{\hat{\mu}_k(a_1^k | x_1^n)}{\hat{\mu}_{k-1}(a_1^{k-1} | x_1^n)}.$$

A *informação dirigida empírica associada a dois blocos de tamanho k* é definida como

$$\hat{I}_{z_1^n}(X_1^k \rightarrow Y_1^k) := \sum_{j=1}^k \left[\hat{H}_{y_1^n}(Y_j | Y_1^{j-1}) - \hat{H}_{(y_1^n, x_1^n)}(Y_j | Y_1^{j-1}, X_1^j) \right],$$

em que

$$\hat{H}_{z_1^n}(Y_k | Y_1^{k-1}, X_1^{k-1}) := - \sum_{\substack{a_1^{k-1} \in A^{k-1} \\ b_1^k \in A^k}} \hat{\nu}_{k,k-1}(a_1^{k-1} b_1^k | x_1^n y_1^n) \log \frac{\hat{\nu}_{k-1,k}(a_1^{k-1} b_1^k | x_1^n y_1^n)}{\hat{\nu}_{k-1,k-1}(a_1^{k-1} b_1^{k-1} | x_1^n y_1^n)}.$$

A entropia empírica associada a blocos de tamanho k converge para a taxa de entropia conforme o tamanho dos blocos e da amostra aumentam (ver, por exemplo, [Shields \(1996\)](#)), ou seja,

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{\hat{H}_{x_1^n}(X_1^k)}{k} = \lim_{k \rightarrow \infty} \frac{H(X_1^k)}{k} = h(X). \quad (4.1)$$

Podemos, entretanto, tomar um limite único em (4.1) ao escrevermos k como uma função de n e impormos que, para todo $n \geq 1$, $k(n) \leq \frac{1}{h(X)}$. Assim, se $k(n) \rightarrow \infty$ quando $n \rightarrow \infty$, temos

$$\lim_{n \rightarrow \infty} \frac{\hat{H}_{x_1^n}(X_1^{k(n)})}{k(n)} = h(X), \quad (4.2)$$

para μ -quase toda \mathbf{x} ([ORNSTEIN; WEISS, 1990](#)). A entropia $H(X_1^k)$ é côncava em μ e atinge seu valor máximo em $\log |A|$, logo $h(X) \leq \log |A|$ e, conseqüentemente, podemos tomar

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

Como \mathbf{X} é estacionário, pode-se mostrar ([Shields, 1996](#)) que, se $k(n) \leq \frac{1-\varepsilon}{\log |A|} \log n$, para qualquer $\varepsilon \in (0, 1)$, então

$$\lim_{n \rightarrow \infty} \hat{H}_{x_1^n}(X_{k(n)} | X_1^{k(n)-1}) = h(X), \quad (4.3)$$

para μ -quase toda \mathbf{x} .

Desta forma, como \mathbf{X} e \mathbf{Y} , são processos estacionários se $k(n) \leq \frac{1-\varepsilon}{\log |A|} \log n$, para qualquer $\varepsilon \in (0, 1)$ temos, por (4.2) e (4.3),

$$\lim_{n \rightarrow \infty} \frac{\hat{I}_{z_1^n}(X_1^{k(n)} \rightarrow Y_1^{k(n)})}{k(n)} = h(Y) + h(Y|X) = I(X \rightarrow Y), \quad (4.4)$$

para ν -quase toda \mathbf{z} .

4.4 Propriedades de Concentração do Estimador

Sejam A um alfabeto finito e $F : A^n \times A^n \rightarrow \mathbb{R}$ uma função mensurável. Vamos definir a *variação de g em i* por

$$\delta_i F := \sup \{ |F(\mathbf{z}_1^n) - F(\mathbf{w}_1^n)| : z_j = w_j, \forall j \neq i \}$$

e, então, a norma $\ell^2(\mathbb{N}^2)$ desse operador é dada por

$$\|\delta F\|_{\ell^2(\mathbb{N}^2)}^2 = \sum_{i=1}^n (\delta_i F)^2.$$

Quando $\|\delta F\|_{\ell^2(\mathbb{N}^2)}^2 < +\infty$, F é dita ser *Lipschitz separável*. Podemos dizer que a concentração de F em torno da sua média é do tipo exponencial se para todo $t > 0$

$$\mu \left\{ |F - E_\mu[F]| \geq t \right\} \leq 2 \exp \left(-\frac{2t^2}{D \|\delta F\|_{\ell^2(\mathbb{N}^2)}^2} \right), \quad (4.5)$$

em que a constante D é obtida através do supremo da norma $\ell^2(\mathbb{N}^2)$ da matriz de acoplamento. A matriz de acoplamento de \mathbf{X} é uma matriz triangular superior $D^{\mathbf{X}}$ definida para $i, j \in \mathbb{N}$ por

$$D_{ii}^{\mathbf{X}} := 1, \\ D_{i,i+j}^{\mathbf{X}} := \max_{a,b \in A} P_{i,a,b}^{\mathbf{X}} \left\{ X_{i+j}^{(a)} \neq X_{i+j}^{(b)} \right\},$$

em que $X_k^{(a)} = X_k^{(b)}$ para todo $k < i$, $X_i^{(a)} := a \neq b := X_i^{(b)}$ e $P_{i,a,b}^{\mathbf{X}}$ é o acoplamento máximo (ver, Lindvall (2002)) entre $\mathbf{X}^{(a)}$ e $\mathbf{X}^{(b)}$. Neste trabalho, assumimos que

$$D := \|\bar{D}\|_{\ell^2(\mathbb{N})}^2 = \sup_{\substack{u \in \ell^2(\mathbb{N}) \\ \|u\|_{\ell^2(\mathbb{N})} = 1}} \|\bar{D}u\|_{\ell^2(\mathbb{N})}^2 < +\infty,$$

em que $\bar{D}_{i,j} := \sup_{\mathbf{X}} D_{i,j}^{\mathbf{X}}$ para todo natural i e j .

Agora estamos prontos para afirmar os principais resultados deste capítulo.

Teorema 4. Seja $\mathbf{Z} = \{(X_n, Y_n) : n \in \mathbb{Z}\}$ um processo bivariado estacionário compatível com uma g -função regular g com taxa de continuidade somável e $D > 0$ a constante que aparece em (4.5). Desde que $k(n) \leq \frac{\varepsilon}{2 \log |A|} \log n$ e μ seja uma g -medida, tem-se, para todo $\varepsilon \in (0, 1)$, $t > 0$ e $n \geq 2$,

$$\mu \left\{ \left| \frac{\hat{I}_{z_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} - E_\mu \left[\frac{\hat{I}_{z_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} \right] \right| \geq t \right\} \leq 2 \exp \left(-\frac{C(n) \Gamma t^2}{\log^4 n} \right), \quad (4.6)$$

em que $C(n) := \left(1 - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-2\varepsilon} - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n^{2\varepsilon}} + \frac{1}{n^{2\varepsilon}} \right)$ e $\Gamma := \frac{\log^2 |A|}{36D}$.

A informação dirigida é uma medida baseada em entropia que tem muita semelhança com a informação mútua sendo que uma das diferenças é levar em conta uma parte casual de X ao longo da história da sequência Y em análise (CAI *et al.*, 2017). Um interesse natural é estudar como o estimador empírico da informação dirigida se concentra em torno da sua taxa, já que essa é uma medida comumente usada na prática por ser limitada pela máxima entropia.

Teorema 5. Seja $\mathbf{Z} = \{(X_n, Y_n) : n \in \mathbb{Z}\}$ um processo bivariado estacionário compatível com uma g -função regular g com taxa de continuidade somável e $D > 0$ a constante que aparece em

(4.5). Desde que $\theta < \frac{1}{|A|}$, $k(n) < \frac{\log n}{2 \log |A|}$ e μ é uma g -medida, existe $c > 0$, tal que para todo $t > 0$ e para todo n suficientemente grande

$$\mu \left(\left| \frac{\hat{I}_{z_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} - I(X \rightarrow Y) \right| \geq t + \frac{c}{n^{1-\gamma} - \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma}} \right) \leq 2 \exp \left\{ -\frac{C(n)\Gamma t^2}{\log^4 n} \right\}, \quad (4.7)$$

em que

$$C(n) := \left(1 - \frac{1}{\log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-\gamma} - \frac{1}{\log |A|} \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma} \right),$$

$$\gamma := \frac{1}{1 + \frac{\log |A|}{\log(1/\theta)}} \quad \text{e} \quad \Gamma := \frac{\log^2 |A|}{36D}.$$

Observação 5. Com os limites de concentração podemos obter intervalos de confiança com probabilidade de cobertura maior ou igual $1 - \alpha$ fixando em $\alpha \in (0, 1)$ os limites superiores dos limites de concentração. Neste caso, necessitamos de uma amostra das sequências de disparos dos neurônios em análise e a partir deles estimar o valor de D , o qual, quando (\mathbf{X}, \mathbf{Y}) pode ser simulado perfeitamente é, pelo Corolário 1 de Gallo e Takahashi (2014), menor ou igual a $(1 + E[\theta])^2$ se o tempo de espera θ até o acoplamento dos processos no procedimento de simulação perfeita é finito. Uma vantagem do uso de técnicas de inferência baseadas na informação dirigida no contexto da neurociência esta associada com sua robustez com relação a não-linearidade dos sinais elétricos trocados entre sequências de disparos neuronais.

4.5 Demonstrações

Demonstração do Teorema 4. Para a demonstração do Teorema 4 serão necessárias algumas proposições e lemas, as quais enunciamos a seguir e demonstramos no final desta seção.

Proposição 2 (Propriedade da Informação Dirigida). Para qualquer $k \geq 2$ temos

$$I \left(X_1^k \rightarrow Y_1^k \right) = \sum_{j=1}^k I \left(X_1^{j-1}; Y_1^j \right) - \sum_{j=1}^k I \left(X_1^{j-1}; Y_1^{j-1} \right), \quad (4.8)$$

em que

$$I \left(X_1^j, Y_1^j \right) := H \left(X_1^j \right) + H \left(Y_1^j \right) - H \left(X_1^j, Y_1^j \right) \text{ e}$$

$$I \left(X_1^j, Y_1^{j-1} \right) := H \left(X_1^j \right) + H \left(Y_1^{j-1} \right) - H \left(X_1^j, Y_1^{j-1} \right)$$

são, respectivamente, a *informação mútua* entre as sequências X_1^j e Y_1^j e a *informação mútua* entre as sequências X_1^j e Y_1^{j-1} . Aqui, $H \left(X_1^j, Y_1^j \right)$ e $H \left(X_1^j, Y_1^{j-1} \right)$ representam a *entropia conjunta* entre duas sequências, cuja definição é dada por

$$H \left(X_1^m Y_1^n \right) := - \sum_{x_1^m, y_1^n} \mu \left([x_1^m] \cap [y_1^n] \right) \log \mu \left([x_1^m] \cap [y_1^n] \right),$$

quaisquer que sejam os inteiros positivos m e n .

Lema 1 (Limite da concentração para entropia conjunta empírica). Seja $\mathbf{Z} = \{(X_n, Y_n) : n \in \mathbb{Z}\}$ um processo bivariado estacionário compatível com uma g -função regular g com taxa de continuidade somável e $D > 0$ a constante que aparece em (4.5). Desde que $k(n) \leq \frac{\varepsilon}{2 \log |A|} \log n$ e μ seja uma g -medida, tem-se, para todo $\varepsilon \in (0, 1)$, $t > 0$ e $n \geq 2$,

$$\mu \left\{ \left| \frac{\hat{H}_{z_1^n} \left(X_1^{k(n)} Y_1^{k(n)} \right)}{k(n)} - E_\mu \left[\frac{\hat{H}_{z_1^n} \left(X_1^{k(n)} Y_1^{k(n)} \right)}{k(n)} \right] \right| \geq t \right\} \leq 2 \exp \left(-\frac{C(n)t^2}{4D \log^2 n} \right), \quad (4.9)$$

em que $C(n) := \left(1 - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-2\varepsilon} - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n^{2\varepsilon}} + \frac{1}{n^{2\varepsilon}} \right) e$

$$\hat{H}_{z_1^n} \left(X_1^k Y_1^k \right) := - \sum_{a_1^k, b_1^k} \hat{\mu}_{k,k} \left(a_1^k b_1^k \right) \log \hat{\mu}_{k,k} \left(a_1^k b_1^k \right).$$

Demonstração. Dado qualquer inteiro $k \geq 1$, considere a função $F : A^n \times A^n \rightarrow \mathbb{R}$ definida por

$$F \left(\mathbf{z}_1^n \right) = F \left(x_1^n, y_1^n \right) := \hat{I}_{z_1^n} \left(X_1^k \rightarrow Y_1^k \right).$$

Pela Proposição 2 sabemos que

$$\hat{I}_{z_1^n} \left(X_1^k \rightarrow Y_1^k \right) = \sum_{j=1}^n \hat{I}_{z_1^n} \left(X_1^j; Y_1^j \right) - \sum_{j=1}^n \hat{I}_{z_1^n} \left(X_1^j; Y_1^{j-1} \right).$$

Assim, definindo $F_1 \left(\mathbf{z}_1^n \right) := \sum_{j=1}^n \hat{I}_{z_1^n} \left(X_1^j; Y_1^j \right)$ e $F_2 \left(\mathbf{z}_1^n \right) := \sum_{j=1}^n \hat{I}_{z_1^n} \left(X_1^j; Y_1^{j-1} \right)$ temos, para quaisquer amostras $\mathbf{z}_1^n \in A^n \times A^n$ e $\mathbf{w}_1^n \in A^n \times A^n$ e para todo $i = 1, \dots, n$, que

$$\delta_i F := \sup \left\{ |F \left(\mathbf{z}_1^n \right) - F \left(\mathbf{w}_1^n \right)| : \mathbf{z}_j = \mathbf{w}_j, \forall j \neq i \right\},$$

em que $\mathbf{z}_j = (x_j, y_j)$ e $\mathbf{w}_j = (u_j, v_j)$ para qualquer $j \in \mathbb{Z}$.

Nos cálculos a seguir utilizamos a notação \mathcal{H}_i para representar a hipótese sobre a qual tomamos o supremo na igualdade anterior e usamos a identidade $\hat{I}_{z_1^n} \left(X_1^k; Y_1^k \right) = \hat{H}_{y_1^n} \left(Y_1^k \right) + \hat{H}_{x_1^n} \left(X_1^k \right) - \hat{H}_{z_1^n} \left(X_1^k Y_1^k \right)$.

$$\begin{aligned}
\delta_i F &\leq \sup_{\mathcal{H}_i} \{|F_1(\mathbf{z}_1^n) - F_1(\mathbf{w}_1^n)|\} + \sup_{\mathcal{H}_i} \{|F_2(\mathbf{z}_1^n) - F_2(\mathbf{w}_1^n)|\} \\
&= \sup_{\mathcal{H}_i} \{|[F_1(x_1^n) - F_1(y_1^n)] + [F_2(y_1^n) - F_2(x_1^n)]|\} \\
&\leq \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{I}_{\mathbf{z}_1^n}(X_1^j; Y_1^j) - \hat{I}_{\mathbf{w}_1^n}(X_1^j; Y_1^j) \right| + \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{I}_{\mathbf{z}_1^n}(X_1^j; Y_1^{j-1}) - \hat{I}_{\mathbf{w}_1^n}(X_1^j; Y_1^{j-1}) \right| \\
&= 2 \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{x_1^n}(X_1^j) - \hat{H}_{y_1^n}(X_1^j) \right| + \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{y_1^n}(Y_1^j) - \hat{H}_{x_1^n}(Y_1^j) \right| \\
&\quad + \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{\mathbf{z}_1^n}(X_1^j; Y_1^j) - \hat{H}_{\mathbf{w}_1^n}(X_1^j; Y_1^j) \right| + \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{y_1^n}(Y_1^{j-1}) - \hat{H}_{x_1^n}(Y_1^{j-1}) \right| \\
&\quad + \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{\mathbf{z}_1^n}(X_1^j; Y_1^{j-1}) - \hat{H}_{\mathbf{w}_1^n}(X_1^j; Y_1^{j-1}) \right| \\
&\leq 5|A|^{2k} \frac{k^2}{n-k+1} \log n + \sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{\mathbf{z}_1^n}(X_1^j; Y_1^{j-1}) - \hat{H}_{\mathbf{w}_1^n}(X_1^j; Y_1^{j-1}) \right|,
\end{aligned}$$

em que a última desigualdade segue do Teorema 4.1 de [Chazottes e Maldonado \(2011\)](#), da demonstração do Lema 1 e do fato que $j \leq k$.

Nosso objetivo agora é encontrar um limitante superior para

$$\sum_{j=1}^k \sup_{\mathcal{H}_i} \left| \hat{H}_{\mathbf{z}_1^n}(X_1^j; Y_1^{j-1}) - \hat{H}_{\mathbf{w}_1^n}(X_1^j; Y_1^{j-1}) \right|$$

independente do valor i . Todavia, é suficiente encontrarmos esse limitante para

$$\sum_{j=1}^k \sum_{a_1^j b_1^{j-1}} \sup_{\mathcal{H}_i} \left| \hat{\mu}(a_1^j b_1^{j-1} | x_1^n y_1^n) \log \hat{\mu}(a_1^j b_1^{j-1} | x_1^n y_1^n) - \hat{\mu}(a_1^j b_1^{j-1} | u_1^n v_1^n) \log \hat{\mu}(a_1^j b_1^{j-1} | u_1^n v_1^n) \right|, \quad (4.10)$$

em que

$$\hat{\mu}_{j,j-1}(a_1^j b_1^{j-1} | x_1^n y_1^n) := \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} \mathbb{I}(x_i^{i+j-1} = a_1^j; y_i^{i+j-2} = b_1^{j-1})$$

e

$$\hat{\mu}_{j,j-1}(a_1^j b_1^{j-1} | u_1^n v_1^n) := \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} \mathbb{I}(u_i^{i+j-1} = a_1^j; v_i^{i+j-2} = b_1^{j-1}).$$

Neste sentido, segmentamos o problema em três casos e em seguida, a partir de constatações algébricas, encontramos um limitante uniforme em i .

1. Para $1 \leq i \leq j$: As parcelas que podem ser diferentes nas somatórias que definem $\hat{\mu}_{j,j-1} \left(a_1^j b_1^{j-1} | x_1^n y_1^n \right)$ e $\hat{\mu}_{j,j-1} \left(a_1^j b_1^{j-1} | u_1^n v_1^n \right)$ são aquelas que dependem dos i -ésimos elementos das sequências x_1^n e u_1^n ou das sequências y_1^n e v_1^n , i.e.,

$$\begin{aligned} & \mathbb{I} \left(x_1^j = a_1^j; y_1^{j-1} = b_1^{j-1} \right) + \dots + \mathbb{I} \left(x_i^{i+j-1} = a_1^j; y_i^{i+j-2} = b_1^{j-1} \right) \\ & \qquad \qquad \qquad \neq \\ & \mathbb{I} \left(u_1^j = a_1^j; v_1^{j-1} = b_1^{j-1} \right) + \dots + \mathbb{I} \left(u_i^{i+j-1} = a_1^j; v_i^{i+j-2} = b_1^{j-1} \right). \end{aligned}$$

Como $x_i \neq y_i$, as somas acima diferem no mínimo em 1 unidade e no máximo em i unidades. Logo, a maior diferença absoluta entre tais somas é i . Desde que as parcelas remanescentes são as mesmas tanto para amostra \mathbf{z}_1^n quanto para a amostra \mathbf{w}_1^n , podemos definir como α o valor da soma restante. Neste contexto, concluímos que a expressão (4.10) pode ser limitada por

$$\sum_{j=1}^k |A|^{2j-1} \left| \frac{\alpha}{n-j+1} \log \left(\frac{\alpha}{n-j+1} \right) - \frac{\alpha+j}{n-j+1} \log \left(\frac{\alpha+j}{n-j+1} \right) \right|.$$

2. Para $j+1 \leq i \leq n-j+1$: As parcelas que podem ser diferentes nas somatórias que definem $\hat{\mu}_{j,j-1} \left(a_1^j b_1^{j-1} | x_1^n, y_1^n \right)$ e $\hat{\mu}_{j,j-1} \left(a_1^j b_1^{j-1} | u_1^n v_1^n \right)$ são aquelas que dependem dos i -ésimos elementos das sequências x_1^n e u_1^n ou y_1^n e v_1^n , i.e.,

$$\begin{aligned} & \mathbb{I} \left(x_{i-j+1}^i = a_1^j; y_{i-j+1}^{i-1} = b_1^{j-1} \right) + \dots + \mathbb{I} \left(x_i^{i+j-1} = a_1^j; y_i^{i+j-2} = b_1^{j-1} \right) \\ & \qquad \qquad \qquad \neq \\ & \mathbb{I} \left(u_{i-j+1}^i = a_1^j; v_{i-j+1}^{i-1} = b_1^{j-1} \right) + \dots + \mathbb{I} \left(u_i^{i+j-1} = a_1^j; v_i^{i+j-2} = b_1^{j-1} \right) \end{aligned}$$

Como $x_i \neq y_i$, as somas acima diferem no mínimo em 1 unidade e no máximo em j unidades. Logo, a maior diferença absoluta entre tais somas é j . Desde que as parcelas remanescentes são as mesmas tanto para amostra \mathbf{z}_1^n quanto para a amostra \mathbf{w}_1^n , podemos definir como β o valor da soma restante. Neste contexto, concluímos que a expressão (4.10) pode ser limitada por

$$\sum_{j=1}^k |A|^{2j-1} \left| \frac{\beta}{n-j+1} \log \left(\frac{\beta}{n-j+1} \right) - \frac{\beta+j}{n-j+1} \log \left(\frac{\beta+j}{n-j+1} \right) \right|.$$

3. Para $n-j+2 \leq i \leq n$: As parcelas que podem ser diferentes nas somatórias que definem $\hat{\mu}_{j,j-1} \left(a_1^j b_1^{j-1} | x_1^n y_1^n \right)$ e $\hat{\mu}_{j,j-1} \left(a_1^j b_1^{j-1} | u_1^n v_1^n \right)$ são aquelas que dependem dos i -ésimos elementos das sequências x_1^n e u_1^n ou y_1^n e v_1^n , i.e.,

$$\begin{aligned} & \mathbb{I}\left(x_{i-j+1}^i = a_1^j; y_{i-j+1}^{i-1} = b_1^{j-1}\right) + \dots + \mathbb{I}\left(x_{n-j+1}^n = a_1^j; y_{n-j+1}^{n-1} = b_1^{j-1}\right) \\ & \neq \\ & \mathbb{I}\left(u_{i-j+1}^i = a_1^j; v_{i-j+1}^{i-1} = b_1^{j-1}\right) + \dots + \mathbb{I}\left(u_{n-j+1}^n = a_1^j; v_{n-j+1}^{n-1} = b_1^{j-1}\right) \end{aligned}$$

Como $x_i \neq y_i$, as somas acima diferem no mínimo em 1 unidade e no máximo em $n - i + 1$ unidades. Logo, a maior diferença absoluta entre tais somas é $n - i + 1$. Desde que as parcelas remanescentes são as mesmas tanto para amostra \mathbf{z}_1^n quanto para a amostra \mathbf{w}_1^n , podemos definir como γ o valor da soma restante. Neste contexto, concluímos que a expressão (4.10) pode ser limitada por

$$\sum_{j=1}^k |A|^{2j-1} \left| \frac{\gamma}{n-j+1} \log\left(\frac{\gamma}{n-j+1}\right) - \frac{\gamma+j}{n-j+1} \log\left(\frac{\gamma+j}{n-j+1}\right) \right|.$$

Em cada caso, α , β e γ representam a menor soma remanescente, i.e.,

$$\begin{aligned} \alpha &= (n-j+1) - (i) \leq n-j \\ \beta &= (n-j+1) - (j) \leq n-j \\ \gamma &= (n-j+1) - (n-i+1) \leq n-j. \end{aligned}$$

Além disso, para quaisquer par positivos de inteiros ℓ e j tal que $0 \leq \ell \leq n-j$ temos

$$\begin{aligned} & \left| \frac{\ell}{n-j+1} \log\left(\frac{\ell}{n-j+1}\right) - \frac{\ell+j}{n-j+1} \log\left(\frac{\ell+j}{n-j+1}\right) \right| \\ &= \left| \frac{\ell}{n-j+1} \log\left(\frac{\ell}{n-j+1}\right) + \frac{\ell+j}{n-j+1} \log\left(\frac{n-j+1}{\ell+j}\right) \right| \\ &= \frac{\ell+j}{n-j+1} \log\left(\frac{n-j+1}{\ell+j}\right) + \frac{\ell}{n-j+1} \log\left(\frac{\ell}{n-j+1}\right) \\ &= \frac{\ell+j}{n-j+1} \log\left(\frac{n-j+1}{\ell+j}\right) - \frac{\ell}{n-j+1} \log\left(\frac{n-j+1}{\ell}\right). \end{aligned}$$

Como $\log\left(\frac{n-j+1}{\ell+j}\right) \leq \log\left(\frac{n-j+1}{\ell}\right)$ segue que

$$\begin{aligned} & \left| \frac{\ell}{n-j+1} \log\left(\frac{\ell}{n-j+1}\right) - \frac{\ell+j}{n-j+1} \log\left(\frac{\ell+j}{n-j+1}\right) \right| \\ &= \frac{\ell+j}{n-j+1} \log\left(\frac{n-j+1}{\ell+j}\right) - \frac{\ell}{n-j+1} \log\left(\frac{n-j+1}{\ell}\right) \\ &\leq \frac{\ell+j}{n-j+1} \log\left(\frac{n-j+1}{\ell}\right) - \frac{\ell}{n-j+1} \log\left(\frac{n-j+1}{\ell}\right) \\ &\leq \frac{j}{n-j+1} \log(n). \end{aligned}$$

Portanto, para todo $i = 1, \dots, n$, a expressão (4.10) pode ser majorada por

$$\sum_{j=1}^k |A|^{2j-1} \frac{j}{n-j+1} \log(n),$$

o qual é tal que

$$\sum_{j=1}^k |A|^{2j-1} \frac{j}{n} \log(n) \leq |A|^{2k-1} \frac{k^2}{n-k+1} \log n.$$

pois $j \leq k$.

Desta forma, se definirmos $F_3 := \hat{H}_{\mathbf{s}_1^n} \left(X_1^j; Y_1^{j-1} \right)$ com $\mathbf{s}_1^n \in A^n \times A^n$ temos que

$$\delta_i F_3 \leq |A|^{2k-1} \frac{k^2}{n-k+1} \log n,$$

para todo $i = 1, \dots, n$. Logo,

$$\begin{aligned} \delta_i F &\leq 5|A|^{2k} \frac{k^2}{n-k+1} \log n + |A|^{2k-1} \frac{k^2}{n-k+1} \log n \\ &\leq 6|A|^{2k} \frac{k^2}{n-k+1} \log n. \end{aligned}$$

Portanto,

$$\|\delta F\|_{\ell^2(\mathbb{N})}^2 = \delta_1^2 F + \delta_2^2 F + \dots + \delta_n^2 F \leq 36|A|^{4k} \frac{n}{(n-k+1)^2} k^4 \log^2 n.$$

Aplicando o Teorema 5 de [Chazottes et al. \(2006\)](#) temos

$$\begin{aligned} &\mu \left\{ \left| \frac{\hat{I}_{\mathbf{z}_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} - E_\mu \left[\frac{\hat{I}_{\mathbf{z}_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} \right] \right| \geq t \right\} \\ &\leq 2 \exp \left\{ -\frac{k^2(n)t^2}{4D \|\delta F\|_{\ell^2(\mathbb{N})}^2} \right\} \\ &\leq 2 \exp \left\{ -\frac{t^2}{144D |A|^{4k(n)} \frac{n}{(n-k(n)+1)^2} k^2(n) (\log n)^2} \right\}. \end{aligned}$$

Tomando $k(n) \leq \frac{\varepsilon}{2 \log |A|} \log n$, com $\varepsilon \in (0, 1)$, segue que $|A|^{4k(n)} \leq n^{2\varepsilon}$, $k^2(n) \leq \frac{\log^2 n}{4 \log^2 |A|}$ e $-\frac{(n-k(n)+1)^2}{n^{2\varepsilon+1}} \leq -\left(1 - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n} + \frac{1}{n}\right) \left(n^{1-2\varepsilon} - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n^{2\varepsilon}} + \frac{1}{n^{2\varepsilon}}\right) := -C(n)$. Portanto, para todo $t > 0$,

$$\mu \left\{ \left| \frac{\hat{I}_{\mathbf{z}_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} - E_{\mu_\phi} \left[\frac{\hat{I}_{\mathbf{z}_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} \right] \right| \geq t \right\} \leq 2 \exp \left(-\frac{C(n)\Gamma t^2}{\log^4 n} \right),$$

em que $\Gamma := \frac{\log^2 |A|}{36D}$. □

Demonstração do Teorema 5. Para a demonstração do Teorema 5 serão necessários dois lemas, os quais enunciamos a seguir. O Lema 3 demonstramos antes de apresentar a prova do Teorema 5, pois elementos de sua demonstração serão necessários durante a prova deste teorema. O Lema 2, por outro lado, é demonstrado ao final desta seção.

Lema 2. Seja $\mathbf{Z} = \{(X_n, Y_n) : n \in \mathbb{Z}\}$ um processo bivariado estacionário compatível com uma g -função regular g com taxa de continuidade somável e $D > 0$ a constante que aparece em (4.5). Desde que $k(n) \leq \frac{\varepsilon}{2 \log |A|} \log n$ e μ seja uma g -medida, tem-se, para todo $\varepsilon \in (0, 1)$, $t > 0$ e $n \geq 2$,

$$\mu \left\{ \left| \hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) - E_\mu \left[\hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) \right] \right| \geq t \right\} \leq 2 \exp \left(- \frac{C(n) \Gamma t^2}{\log^4 n} \right), \quad (4.11)$$

em que $C(n) := \left(1 - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-2\varepsilon} - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n^{2\varepsilon}} + \frac{1}{n^{2\varepsilon}} \right)$ e $\Gamma := \frac{\log^2 |A|}{4D}$ e

$$\hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) = \hat{H}_{z_1^n} \left(Y_1^{k(n)} X_1^{k(n)} \right) - \hat{H}_{z_1^n} \left(Y_1^{k(n)} X_1^{k(n)-1} \right).$$

Lema 3 (Limite da concentração para entropia condicional a dois blocos diferentes). Seja $\mathbf{Z} = \{(X_n, Y_n) : n \in \mathbb{Z}\}$ um processo bivariado estacionário compatível com uma g -função regular g com taxa de continuidade somável e $D > 0$ a constante que aparece em (4.5). Desde que $\theta < \frac{1}{|A|}$, $k(n) < \frac{\log n}{2 \log |A|}$ e μ seja uma g -medida, existe $c > 0$, tal que para todo $t > 0$ e para todo n suficientemente grande

$$\mu \left\{ \left| \hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) - h(Y|X) \right| \geq t + \frac{c}{n^{1-\gamma} - \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma}} \right\} \leq 2 \exp \left\{ - \frac{C(n) \Gamma t^2}{\log^4 n} \right\}, \quad (4.12)$$

em que

$$C(n) := \left(1 - \frac{1}{\log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-\gamma} - \frac{1}{\log |A|} \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma} \right),$$

$$\gamma := \frac{1}{1 + \frac{\log |A|}{\log(1/\theta)}} \quad \text{e} \quad \Gamma := \frac{\log^2 |A|}{4D}.$$

Demonstração. Dado qualquer inteiro $k \geq 1$, considere a função $F : A^n \times A^n \rightarrow \mathbb{R}$ definida por

$$F(z_1^n) = F(x_1^n, y_1^n) := \hat{H}_{z_1^n} \left(Y_k \mid Y_1^{k-1} X_1^k \right).$$

Pelo Lema 2, temos

$$\delta_i F \leq 2|A|^{2k} \frac{k}{n-k+1} \log n, \quad (4.13)$$

para todo $i = 1, \dots, n$.

Além disso,

$$\begin{aligned}
F(\mathbf{z}_1^n) &:= - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right)}{\hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} \mid x_1^n y_1^n \right)} \\
&= - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right)}{\hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} \mid x_1^n y_1^n \right)} \\
&\quad + \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\mu \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} \\
&\quad - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\mu \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)},
\end{aligned}$$

Definindo,

$$\begin{aligned}
\hat{\Delta}_k(x_1^n; y_1^n) &:= - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right)}{\hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} \mid x_1^n y_1^n \right)} \\
&\quad + \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\mu \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)},
\end{aligned}$$

obtemos

$$F(\mathbf{z}_1^n) = - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\mu \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} + \hat{\Delta}_k(x_1^n; y_1^n).$$

Podemos reescrever $\hat{\Delta}_k(x_1^n; y_1^n)$ da seguinte maneira

$$\begin{aligned}
\hat{\Delta}_k(x_1^n; y_1^n) &= - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right)}{\mu \left([a_1^k b_1^k] \right)} \\
&\quad + \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k \mid x_1^n y_1^n \right) \log \frac{\hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} \mid x_1^n y_1^n \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} \\
&:= -H_{k,k} \left(\hat{\mu}(\cdot \mid x_1^n y_1^n) \mid \mu_\phi \right) + H_{k,k-1} \left(\hat{\mu}(\cdot \mid x_1^n y_1^n) \mid \mu_\phi \right),
\end{aligned}$$

em que

$$H_{k,k}(\eta \mid \mu) = \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \eta \left([a_1^k b_1^k] \right) \log \frac{\eta \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^k] \right)}$$

e

$$H_{k,k-1}(\eta | \mu) = \sum_{\substack{a_1^k \in A^k \\ b_1^{k-1} \in A^{k-1}}} \eta \left([a_1^k b_1^{k-1}] \right) \log \frac{\eta \left([a_1^k b_1^{k-1}] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)},$$

são as entropias relativas de η com respeito a μ associadas, respectivamente, a blocos de ordem $k \times k$ e $k \times k - 1$.

Note que,

$$\begin{aligned} & \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \log \frac{\hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} | x_1^n y_1^n \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} \\ &= \sum_{\substack{a_1^k \in A^k \\ b_1^{k-1} \in A^{k-1}}} \left[\log \frac{\hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} | x_1^n y_1^n \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} \sum_{b_k \in A} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \right] \\ &= \sum_{\substack{a_1^k \in A^k \\ b_1^{k-1} \in A^{k-1}}} \hat{\mu}_{k,k-1} \left(a_1^k b_1^{k-1} | x_1^n y_1^n \right) \log \frac{\hat{\mu} \left(a_1^k b_1^{k-1} | x_1^n y_1^n \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} \\ &= H_{k,k-1} \left(\hat{\mu}(\cdot | x_1^n y_1^n) | \mu \right). \end{aligned}$$

Pela fórmula (4.16) de [Gabrielli, Galves e Guiol \(2013\)](#) temos

$$\left| E_{\mu_\phi} \left[H_{k,k} \left(\hat{\mu}(\cdot | x_1^n y_1^n) | \mu \right) \right] \right| \leq \frac{M_1 |A|^{2k}}{n-k+1} \quad \text{e} \quad \left| E_{\mu_\phi} \left[H_{k,k-1} \left(\hat{\mu}(\cdot | x_1^n y_1^n) | \mu \right) \right] \right| \leq \frac{M_2 |A|^{2k}}{n-k+1},$$

em que M_1 e M_2 são constantes positivas. Logo,

$$\left| E_\phi \left[\hat{\Delta}_k(x_1^n; y_1^n) \right] \right| \leq \frac{M}{n-k+1} |A|^{2k}, \quad (4.14)$$

de forma que $M := M_1 + M_2$.

Agora, definindo $\phi(\mathbf{b} | \mathbf{a}) := \log g \left(b_k \left| b_{-\infty}^{k-1} a_{-\infty}^k \right. \right)$ e $\phi_k(\mathbf{b} | \mathbf{a}) := \log \frac{\mu \left([b_1^k] [a_1^k] \right)}{\mu \left([b_1^{k-1}] [a_1^k] \right)}$,

segue, pela proposição 3.2 de [Parry e Pollicott \(1990\)](#) que existe uma constante $L > 0$ tal que

$$\|\phi - \phi_k\|_\infty \leq \text{var}_k(g) \leq L\theta^k,$$

com $\theta \in (0, 1)$. Assim, para todo $\mathbf{a} \in A^{\mathbb{N}}$ e $\mathbf{b} \in A^{\mathbb{N}}$,

$$\begin{aligned} |\phi(\mathbf{b} | \mathbf{a}) - \phi_k(\mathbf{b} | \mathbf{a})| &\leq L\theta^k \Rightarrow \frac{|\phi(\mathbf{b} | \mathbf{a}) - \phi_k(\mathbf{b} | \mathbf{a})|}{\theta^k} \leq L \\ &\Rightarrow |\phi(\mathbf{b} | \mathbf{a}) - \phi_k(\mathbf{b} | \mathbf{a})| := \mathcal{O}(\theta^k) \\ &\Rightarrow \phi_k(\mathbf{b} | \mathbf{a}) = \phi(\mathbf{b} | \mathbf{a}) - \mathcal{O}(\theta^k). \end{aligned}$$

Logo,

$$\begin{aligned} - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \log \frac{\mu \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} &= - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \phi_k \left(\mathbf{b} | \mathbf{a} \right) \\ &= - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \phi \left(\mathbf{b} | \mathbf{a} \right) + \mathcal{O} \left(\theta^k \right). \end{aligned}$$

Observe que,

$$\begin{aligned} - \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \phi \left(\mathbf{b} | \mathbf{a} \right) &= - \frac{1}{n-k+1} \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \left[\sum_{i=1}^{n-k+1} \mathbb{I} \left(x_i^{i+k-1} = a_1^k; y_i^{i+k-1} = b_1^k \right) \right] \phi \left(\mathbf{b} | \mathbf{a} \right) \\ &= - \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \left[\mathbb{I} \left(x_i^{i+k-1} = a_1^k; y_i^{i+k-1} = b_1^k \right) \phi \left(\mathbf{b} | \mathbf{a} \right) \right], \\ &= - \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \left[\mathbb{I} \left(x_i^{i+k-1} = a_1^k; y_i^{i+k-1} = b_1^k \right) \log g \left(b_k | b_{-\infty}^{k-1} a_{-\infty}^k \right) \right] \\ &= - \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \log g \left(y_{i+k-1} | y_{-\infty}^{i+k-2} x_{-\infty}^{i+k-1} \right) \\ &= - \frac{1}{n-k+1} \sum_{i=1}^n \phi \left(T^i \mathbf{y} | T^i \mathbf{x} \right). \end{aligned}$$

Portanto,

$$- \sum_{\substack{a_1^k \in A^k \\ b_1^k \in A^k}} \hat{\mu}_{k,k} \left(a_1^k b_1^k | x_1^n y_1^n \right) \log \frac{\mu \left([a_1^k b_1^k] \right)}{\mu \left([a_1^k b_1^{k-1}] \right)} = - \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \phi \left(T^i \mathbf{y} | T^i \mathbf{x} \right) + \mathcal{O} \left(\theta^k \right). \quad (4.15)$$

Com base em tais constatações, concluímos que

$$F \left(\mathbf{z}_1^n \right) = \hat{H}_{z_1^n} \left(Y_k | Y_1^{k-1} X_1^k \right) = - \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \phi \left(T^i \mathbf{y} | T^i \mathbf{x} \right) + \hat{\Delta}_k \left(x_1^n; y_1^n \right) + \mathcal{O} \left(\theta^k \right),$$

e, conseqüentemente,

$$E_\mu \left[F \left(\mathbf{z}_1^n \right) \right] = h \left(Y | X \right) + E_\mu \left[\hat{\Delta}_k \left(X_1^n; Y_1^n \right) \right] + \mathcal{O} \left(\theta^k \right).$$

Assim,

$$\left| E_\mu \left[\hat{H}_{z_1^n} \left(Y_k | Y_1^{k-1} X_1^k \right) \right] - h \left(Y | X \right) \right| = \left| E_\mu \left[\hat{\Delta}_k \left(x_1^n; y_1^n \right) \right] + \mathcal{O} \left(\theta^k \right) \right|$$

Seja $k(n) := q \frac{\log n}{\log |A|}$, em que $q \in (0, 1)$ é alguma constante fixa e conhecida. Para tanto, escolhamos

$$q = \frac{\log |A|}{\log |A| - \log \theta},$$

pois $q < 1/2$ e, conseqüentemente, $k(n) < \frac{\log n}{2 \log |A|}$ como é pressuposto pelas hipóteses desse Teorema.

Lembrando que $\theta < \frac{1}{|A|}$ e utilizando (4.14), temos

$$\begin{aligned} \left| E_{\mu} \left[\hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) \right] - h(Y|X) \right| &\leq \frac{M}{n - k(n) + 1} |A|^{2k(n)} + \mathcal{O} \left(\theta^{k(n)} \right) \\ &\leq \frac{M}{n - k(n) + 1} |A|^{2k(n)} + L \theta^{k(n)} \\ &\leq \frac{M}{n - k(n) + 1} |A|^{2k(n)} + L \frac{1}{|A|^{k(n)}} \\ &\leq \frac{M}{n - k(n) + 1} |A|^{2k(n)} + L |A|^{2k(n)} \\ &= \frac{c}{n - k(n) + 1} |A|^{2k(n)} \\ &\leq \frac{c}{n - \log n + 1} |A|^{2k(n)}. \end{aligned}$$

em que $c := M + L$.

Uma vez que $\gamma := \frac{\log \theta}{\log \theta - \log |A|}$, temos $|A|^{2k(n)} < n^{\gamma}$ com $0 < \gamma < 1$ e $0 < 1 - \gamma < 1/2$.

Portanto,

$$\left| E_{\mu} \left[\hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) \right] - h(Y|X) \right| \leq \frac{c}{n^{1-\gamma} - \frac{\log n}{n^{\gamma}} + \frac{1}{n^{\gamma}}}. \quad (4.16)$$

Agora, aplicando o Corolário 2 e usando (4.16), concluímos que

$$\mu \left(\left| \hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) - h(Y|X) \right| \geq t + \frac{c}{n^{1-\gamma} - \frac{\log n}{n^{\gamma}} + \frac{1}{n^{\gamma}}} \right) \leq 2 \exp \left\{ -\frac{C(n) \Gamma t^2}{\log^4 n} \right\},$$

em que $C(n) := \left(1 - \frac{1}{\log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-\gamma} - \frac{1}{\log |A|} \frac{\log n}{n^{\gamma}} + \frac{1}{n^{\gamma}} \right)$ e $\Gamma := \frac{\log^2 |A|}{4D}$. \square

Demonstração. Para qualquer $k \geq 1$,

$$\hat{I}_{z_1^n} \left(X_1^k \rightarrow Y_1^k \right) = \sum_{j=1}^k \hat{H}_{y_1^n} \left(Y_j \mid Y_1^{j-1} \right) - \sum_{j=1}^k \hat{H}_{z_1^n} \left(Y_1 \mid Y_1^{j-1} X_1^j \right).$$

Pelas demonstrações do Lema 3 e do Teorema 4.2 de Chazottes e Maldonado (2011), temos

$$\hat{H}_{y_1^n} \left(Y_j \mid Y_1^{j-1} \right) = -\frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \phi \left(T^i \mathbf{y} \right) + \hat{\Delta}_j \left(y_1^n \right) + \mathcal{O} \left(\theta^j \right) \quad (4.17)$$

$$\hat{H}_{z_1^n} \left(Y_j \mid Y_1^{j-1} X_1^j \right) = -\frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \phi \left(T^i \mathbf{y} \mid T^i \mathbf{x} \right) + \hat{\Delta}_j \left(x_1^n; y_1^n \right) + \mathcal{O} \left(\theta^j \right), \quad (4.18)$$

para todo $j = 1, \dots, k$.

Logo,

$$\hat{I}_{z_1^n} \left(X_1^k \rightarrow Y_1^k \right) = -\frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \phi(T^i \mathbf{y}) + \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \phi(T^i \mathbf{y} | T^i \mathbf{x}) + \sum_{j=1}^k \hat{\Delta}_j(y_1^n) - \sum_{j=1}^k \hat{\Delta}_j(x_1^n; y_1^n).$$

Tomando o valor esperado sobre μ em ambos os lados da igualdade anterior, obtemos

$$E_\mu \left[\frac{\hat{I}_{z_1^n} \left(X_1^k \rightarrow Y_1^k \right)}{k} \right] \leq h(Y) - h(Y|X) + \frac{1}{k} \sum_{j=1}^k E_\mu [\hat{\Delta}_j(y_1^n)] - \frac{1}{k} \sum_{j=1}^k E_\mu [\hat{\Delta}_j(x_1^n; y_1^n)].$$

Usando o fato de que $I(X \rightarrow Y) = h(Y) - h(Y|X)$, podemos escrever

$$\left| E_\mu \left[\frac{\hat{I}_{z_1^n} \left(X_1^k \rightarrow Y_1^k \right)}{k} \right] - I(X \rightarrow Y) \right| \leq \left| \frac{1}{k} \sum_{j=1}^k E_\mu [\hat{\Delta}_j(y_1^n)] - \frac{1}{k} \sum_{j=1}^k E_\mu [\hat{\Delta}_j(x_1^n; y_1^n)] \right|.$$

Seja $k(n) := q \frac{\log n}{\log |A|}$, em que $q \in (0, 1)$ é alguma constante fixa e conhecida. Para tanto, escolhamos

$$q = \frac{\log |A|}{\log |A| - \log \theta},$$

pois $q < 1/2$ e, conseqüentemente, $k(n) < \frac{\log n}{2 \log |A|}$ como é pressuposto pelas hipóteses desse Teorema.

As demonstrações do Teorema 4.2 de [Chazottes e Maldonado \(2011\)](#) e do Lema 3 nos garantem que existem constantes positivas M_1 e M_2 tais que

$$\left| E_\mu [\hat{\Delta}_j(y_1^n)] \right| \leq \frac{M_1 |A|^k}{n-k+1} \quad \text{e} \quad \left| E_\mu [\hat{\Delta}_j(x_1^n; y_1^n)] \right| \leq \frac{M_2 |A|^{2k}}{n-k+1},$$

qualquer que seja $j = 1, \dots, k(n)$. Assim,

$$\begin{aligned} \left| E_\mu \left[\frac{\hat{I}_{z_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} \right] - I(X \rightarrow Y) \right| &\leq \left| \frac{1}{k(n)} \sum_{j=1}^{k(n)} E_\mu [\hat{\Delta}_j(y_1^n)] - \frac{1}{k(n)} \sum_{j=1}^{k(n)} E_\mu [\hat{\Delta}_j(x_1^n; y_1^n)] \right| \\ &\leq \left| E_\mu [\hat{\Delta}_{k(n)}(y_1^n)] - E_\mu [\hat{\Delta}_{k(n)}(x_1^n; y_1^n)] \right| \\ &\leq \frac{c}{n-k(n)+1} |A|^{2k(n)}, \end{aligned}$$

em que $c := M_1 + M_2$.

Uma vez que $\gamma := \frac{\log \theta}{\log \theta - \log |A|}$, temos $|A|^{2k(n)} < n^\gamma$ com $0 < \gamma < 1$ e $0 < 1 - \gamma < 1/2$.

Portanto,

$$\left| E_\mu \left[\frac{\hat{I}_{z_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} \right] - I(X \rightarrow Y) \right| \leq \frac{c}{n^{1-\gamma} - \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma}}. \quad (4.19)$$

Agora, aplicando o Teorema 4 e usando (4.19), concluímos que

$$\mu \left(\left| \frac{\hat{I}_{z_1^n} \left(X_1^{k(n)} \rightarrow Y_1^{k(n)} \right)}{k(n)} - I(X \rightarrow Y) \right| \geq t + \frac{c}{n^{1-\gamma} - \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma}} \right) \leq 2 \exp \left\{ -\frac{C(n)\Gamma t^2}{\log^4 n} \right\},$$

em que $C(n) := \left(1 - \frac{1}{\log |A|} \frac{\log n}{n} + \frac{1}{n} \right) \left(n^{1-\gamma} - \frac{1}{\log |A|} \frac{\log n}{n^\gamma} + \frac{1}{n^\gamma} \right)$ e $\Gamma := \frac{\log^2 |A|}{36D}$. \square

Demonstração da Proposição 2. De fato,

$$\begin{aligned} I \left(X_1^k \rightarrow Y_1^k \right) &:= \sum_{j=1}^k \left[H \left(Y_j \mid Y_1^{j-1} \right) - H \left(Y_j \mid Y_1^{j-1} X_1^j \right) \right] \\ &= \sum_{j=1}^k \left[H \left(Y_1^j \right) - H \left(Y_1^{j-1} \right) - H \left(X_1^j Y_1^j \right) + H \left(X_1^j Y_1^{j-1} \right) \right] \\ &= \sum_{j=1}^k \left[H \left(Y_1^j \right) + H \left(X_1^j \right) - H \left(X_1^j Y_1^j \right) - H \left(Y_1^{j-1} \right) - H \left(X_1^j \right) + H \left(X_1^j Y_1^{j-1} \right) \right] \\ &= \sum_{j=1}^k \left[I \left(X_1^j; Y_1^j \right) - I \left(X_1^j; Y_1^{j-1} \right) \right] \\ &= \sum_{j=1}^n I \left(X_1^j; Y_1^j \right) - \sum_{j=1}^n I \left(X_1^j; Y_1^{j-1} \right). \end{aligned}$$

\square

Demonstração do Lema 1. Dado qualquer inteiro $k \geq 1$, considere a função $F : A^n \times A^n \rightarrow \mathbb{R}$ definida por

$$F \left(z_1^n \right) = F \left(x_1^n y_1^n \right) = \hat{H}_{z_1^n} \left(X_1^k Y_1^k \right) = \hat{H}_{x_1^n y_1^n} \left(X_1^k Y_1^k \right),$$

a qual é Lipschitz separável, pois $\|\delta F\|_{\ell^2(\mathbb{N})}^2 < \infty$. De fato, para quaisquer amostras $z_1^n \in A^n \times A^n$ e $w_1^n \in A^n \times A^n$ e para todo $i = 1, \dots, n$, segue que

$$\begin{aligned} \delta_i K &:= \sup \left\{ \left| F \left(z_1^n \right) - F \left(w_1^n \right) \right| : z_j = w_j, \forall j \neq i \right\} \\ &\leq \sum_{c_1^k \in A^k \times A^k} \sup \left\{ \left| \hat{\mu} \left(c_1^k \mid z_1^n \right) \log \hat{\mu} \left(c_1^k \mid z_1^n \right) - \hat{\mu} \left(c_1^k \mid w_1^n \right) \log \hat{\mu} \left(c_1^k \mid w_1^n \right) \right| : z_j = w_j, \forall j \neq i \right\}, \end{aligned}$$

em que para qualquer amostra $w_1^n \in A^n \times A^n$ e para toda sequência $c_1^k = \{(a_j, b_j) : j = 1, \dots, k\} \in A^k \times A^k$, temos

$$\hat{\mu} \left(c_1^k \mid z_1^n \right) := \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \mathbb{I} \left(\omega_i^{i+k-1} = c_1^k \right) = \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} \mathbb{I} \left(\tilde{x}_i^{i+k-1} = a_1^k, \tilde{y}_i^{i+k-1} = b_1^k \right),$$

em que $\omega_j = (\tilde{x}_j, \tilde{y}_j)$, para todo $j \in \mathbb{Z}$.

Então, procedendo como no Teorema 4.1 de [Chazottes e Maldonado \(2011\)](#), concluímos que

$$\delta_i F \leq |A|^{2k} \frac{k}{n-k+1} \log(n) \Rightarrow \|\delta F\|_{\ell^2(\mathbb{N})}^2 = \delta_1^2 F + \delta_2^2 F + \dots + \delta_n^2 F \leq |A|^{4k} \frac{n}{(n-k+1)^2} k^2 (\log n)^2.$$

Aplicando o Teorema 5 de [Chazottes et al. \(2006\)](#) e tomando $k(n) \leq \frac{\varepsilon}{2 \log |A|} \log n$, com $\varepsilon \in (0, 1)$, segue que $|A|^{4k(n)} \leq n^{2\varepsilon}$ e

$$-\frac{(n-k(n)+1)^2}{n^{2\varepsilon+1}} \leq -\left(1 - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n} + \frac{1}{n}\right) \left(n^{1-2\varepsilon} - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n^{2\varepsilon}} + \frac{1}{n^{2\varepsilon}}\right) := -C(n).$$

Portanto, para todo $t > 0$,

$$\mu \left\{ \left| \frac{\hat{H}_{\mathbf{z}_1^n} \left(X_1^{k(n)} Y_1^{k(n)} \right)}{k(n)} - E_\mu \left[\frac{\hat{H}_{\mathbf{z}_1^n} \left(X_1^{k(n)} Y_1^{k(n)} \right)}{k(n)} \right] \right| \geq t \right\} \leq 2 \exp \left(-\frac{C(n)t^2}{4D \log^2 n} \right)$$

□

Demonstração do Lema 2. Dado qualquer inteiro $k \geq 1$, considere a função $F : A^n \times A^n \rightarrow \mathbb{R}$ definida por

$$F(\mathbf{z}_1^n) = F(x_1^n y_1^n) = \hat{H}_{\mathbf{z}_1^n} \left(Y_k \mid Y_1^{k-1} X_1^k \right).$$

Sabemos que

$$\hat{H}_{\mathbf{z}_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) = \hat{H}_{\mathbf{z}_1^n} \left(X_1^k Y_1^k \right) - \hat{H}_{\mathbf{z}_1^n} \left(X_1^k Y_1^{k-1} \right).$$

Assim, definindo $F_1(\mathbf{z}_1^n) := \hat{H}_{\mathbf{z}_1^n} \left(X_1^k Y_1^k \right)$ e $F_2(\mathbf{z}_1^n) := \hat{H}_{\mathbf{z}_1^n} \left(X_1^k Y_1^{k-1} \right)$ temos, para quaisquer amostras $\mathbf{z}_1^n \in A^n \times A^n$ e $\mathbf{w}_1^n \in A^n \times A^n$ e para todo $i = 1, \dots, n$, que

$$\begin{aligned} \delta_i F &:= \sup \{ |F(\mathbf{z}_1^n) - F(\mathbf{w}_1^n)| : \mathbf{z}_j = \mathbf{w}_j, \forall j \neq i \} \\ &\leq \sup \{ |F_1(\mathbf{z}_1^n) - F_1(\mathbf{w}_1^n)| : \mathbf{z}_j = \mathbf{w}_j, \forall j \neq i \} + \sup \{ |F_2(\mathbf{z}_1^n) - F_2(\mathbf{w}_1^n)| : \mathbf{z}_j = \mathbf{w}_j, \forall j \neq i \} \\ &= \delta_i F_1 + \delta_i F_2 \\ &\leq |A|^{2k} \frac{k}{n-k+1} \log n + |A|^{2k-1} \frac{k}{n-k+1} \log n. \end{aligned}$$

em que a última desigualdade segue da demonstração do Teorema 1.

Portanto,

$$\|\delta_i F\|_{\ell^2(\mathbb{N})}^2 = \delta_1^2 F + \delta_2^2 F + \dots + \delta_n^2 F \leq 4|A|^{4k} \frac{n}{(n-k+1)^2} k^2 \log^2 n.$$

Aplicando o Teorema 5 de [Chazottes et al. \(2006\)](#) temos

$$\begin{aligned} & \mu \left\{ \left| \hat{H}_{z_1^n} \left(Y_k \mid Y_1^{k-1} X_1^k \right) - E_{\mu_\phi} \left[\hat{H}_{z_1^n} \left(Y_k \mid Y_1^{k-1} X_1^k \right) \right] \right| \geq t \right\} \\ & \leq 2 \exp \left\{ -\frac{t^2}{4D \|\delta K\|_{\ell^2(\mathbb{N})}^2} \right\} \\ & \leq 2 \exp \left\{ -\frac{nt^2}{16D |A|^{4k(n)} k(n)^2 (\log n)^2} \right\}. \end{aligned}$$

Tomando $k(n) \leq \frac{\varepsilon}{2 \log |A|} \log n$, com $\varepsilon \in (0, 1)$, segue que $|A|^{4k(n)} \leq n^{2\varepsilon}$, $k^2(n) \leq \frac{\log^2 n}{4 \log |A|}$ e $-\frac{(n-k(n)+1)^2}{n^{2\varepsilon+1}} \leq -\left(1 - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n} + \frac{1}{n}\right) \left(n^{1-2\varepsilon} - \frac{\varepsilon}{2 \log |A|} \frac{\log n}{n^{2\varepsilon}} + \frac{1}{n^{2\varepsilon}}\right) := -C(n)$. Portanto, para todo $t > 0$

$$\mu \left\{ \left| \hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) - E_{\mu} \left[\hat{H}_{z_1^n} \left(Y_{k(n)} \mid Y_1^{k(n)-1} X_1^{k(n)} \right) \right] \right| \geq t \right\} \leq 2 \exp \left(-\frac{C(n) \Gamma t^2}{\log^4 n} \right),$$

em que $\Gamma := \frac{\log^2 |A|}{4D}$.

□

CONSIDERAÇÕES FINAIS

Nesta tese, estudamos as cadeias estocásticas de memória ilimitada a partir de dois pontos de vista. O primeiro está relacionado com as características estruturais desse objeto matemático, sobre as quais apresentamos novos resultados a cerca da existência e unicidade de g -medidas compatíveis com g -funções descontínuas. O segundo é motivado por sua capacidade de modelagem, em que as utilizamos para modelar sequências de disparos neuronais. Neste sentido, estudamos as propriedades da interação neuronal, situação na qual utilizamos a informação dirigida para quantificar a informação trocada entre dois neurônios e, então, propomos resultados relativos a desigualdades da concentração para informação dirigida empírica.

Neste contexto, tópicos interessantes para estudos futuros são:

1. Generalizar os resultados de limite concentração obtidos para o caso de g -funções descontínuas.
2. Propor um Teorema Central do Limite para a informação dirigida empírica.
3. Confrontar os resultados obtidos e, os que possivelmente venhamos obter, com dados reais.

Reiteramos que o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) processo #2016/12918 – 0, vinculado ao Centro de Pesquisa, Inovação e Difusão em Neuromatemática - CEPID Neuromat, processo #2013/07699 – 0. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade dos autores e não necessariamente refletem a visão da FAPESP.

REFERÊNCIAS

- BERBEE, H. Periodicity and absolute regularity. **Israel Journal of Mathematics**, Springer, v. 55, n. 3, p. 289–304, 1986. Citado na página 40.
- BRADLEY, R. C. *et al.* Basic properties of strong mixing conditions. a survey and some open questions. **Probability surveys**, The Institute of Mathematical Statistics and the Bernoulli Society, v. 2, p. 107–144, 2005. Citado na página 40.
- BUESING, L.; BILL, J.; NESSLER, B.; MAASS, W. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. **PLoS computational biology**, Public Library of Science, v. 7, n. 11, p. e1002211, 2011. Citado na página 52.
- CAI, Z.; NEVEU, C. L.; BAXTER, D. A.; BYRNE, J. H.; AAZHANG, B. Inferring neuronal network functional connectivity with directed information. **Journal of neurophysiology**, American Physiological Society Bethesda, MD, v. 118, n. 2, p. 1055–1069, 2017. Citado nas páginas 21, 51, 54 e 58.
- CÉNAC, P.; CHAUVIN, B.; PACCAUT, F.; POUYANNE, N. Variable length markov chains and dynamical sources. **Séminaire de Probabilités XLIV, Lecture Notes in Math**, 2012. Citado nas páginas 20, 33 e 34.
- CESSAC, B. A discrete time neural network model with spiking neurons. **Journal of Mathematical Biology**, Springer, v. 56, n. 3, p. 311–345, 2008. Citado na página 20.
- _____. A view of neural networks as dynamical systems. **International Journal of Bifurcation and Chaos**, World Scientific, v. 20, n. 06, p. 1585–1629, 2010. Citado na página 52.
- _____. A discrete time neural network model with spiking neurons: II: Dynamics with noise. **Journal of mathematical biology**, Springer, v. 62, n. 6, p. 863–900, 2011. Citado na página 20.
- _____. Statistics of spike trains in conductance-based neural networks: Rigorous results. **The journal of Mathematical Neuroscience**, Springer, v. 1, n. 1, p. 8, 2011. Citado nas páginas 20 e 53.
- CESSAC, B.; COFRE, R. Spike train statistics and gibbs distributions. **Journal of Physiology-Paris**, Elsevier, v. 107, n. 5, p. 360–368, 2013. Citado na página 20.
- CESSAC, B.; ROSTRO, H.; VASQUEZ, J.-C.; VIÉVILLE, T. How gibbs distributions may naturally arise from synaptic adaptation mechanisms. a model-based argumentation. **Journal of Statistical Physics**, Springer, v. 136, n. 3, p. 565–602, 2009. Citado na página 20.
- CHAZOTTES, J.; GABRIELLI, D. Large deviations for empirical entropies of g-measures. **Nonlinearity**, IOP Publishing, v. 18, n. 6, p. 2545, 2005. Citado na página 54.
- CHAZOTTES, J.; MALDONADO, C. Concentration bounds for entropy estimation of one-dimensional gibbs measures. **Nonlinearity**, IOP Publishing, v. 24, n. 8, p. 2371, 2011. Citado nas páginas 54, 61, 69, 70 e 72.

- CHAZOTTES, J.-R.; COLLET, P.; KÜLSKE, C.; REDIG, F. Concentration inequalities for random fields via coupling. **Probability Theory and Related Fields**, Springer, v. 137, n. 1-2, p. 201–225, 2006. Citado nas páginas 64, 72 e 73.
- CORDEIRO, V. L.; PENA, R. F. de O.; CEBALLOS, C. A. C.; SHIMOURA, R. O.; ROQUE, A. C. Aplicações da teoria da informação à neurociência. **Revista Brasileira de Ensino de Física**, v. 41, n. 2, p. e20180197, 2019. Citado na página 53.
- DOEBLIN, W.; FORTET, R. Sur des chaînes à liaisons complètes. **Bulletin de la Société Mathématique de France**, v. 65, p. 132–148, 1937. Citado nas páginas 19, 20 e 39.
- DUARTE, A.; OST, G. A model for neural activity in the absence of external stimuli. **arXiv preprint arXiv:1410.6086**, 2014. Citado na página 21.
- DUARTE, A.; OST, G.; RODRÍGUEZ, A. A. Hydrodynamic limit for spatially structured interacting neurons. **Journal of Statistical Physics**, Springer, v. 161, n. 5, p. 1163–1202, 2015. Citado na página 21.
- DURRETT, R. **Probability: theory and examples**. [S.l.]: Cambridge university press, 2010. Citado nas páginas 42, 47 e 48.
- FERNANDEZ, R.; MAILLARD, G. Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. **Journal of statistical physics**, Springer, v. 118, n. 3-4, p. 555–588, 2005. Citado na página 19.
- FRISTON, K. J. Functional and effective connectivity in neuroimaging: a synthesis. **Human brain mapping**, Wiley Online Library, v. 2, n. 1-2, p. 56–78, 1994. Citado na página 55.
- FRISTON, K. J.; DAUNIZEAU, J.; KILNER, J.; KIEBEL, S. J. Action and behavior: a free-energy formulation. **Biological cybernetics**, Springer, v. 102, n. 3, p. 227–260, 2010. Citado na página 53.
- GABRIELLI, D.; GALVES, A.; GUIOL, D. Fluctuations of the empirical entropies of a chain of infinite order. **Math. Phys. Electron**, v. 9, p. 5–17, 2013. Citado nas páginas 21 e 67.
- GALLESCO, C.; GALLO, S.; TAKAHASHI, D. Dynamic uniqueness for stochastic chains with unbounded memory. **Stochastic Processes and their applications**, v. 128, p. 689–706, 2018. Citado na página 48.
- GALLO, S. Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme. **Advances in Applied Probability**, Cambridge Univ Press, v. 43, n. 03, p. 735–759, 2011. Citado nas páginas 20 e 34.
- GALLO, S.; GARCIA, N. L. Perfect simulation for locally continuous chains of infinite order. **Stochastic Processes and their Applications**, Elsevier, v. 123, n. 11, p. 3877–3902, 2013. Citado nas páginas 20, 36, 38 e 39.
- GALLO, S.; PACCAUT, F. On non-regular g -measures. **Nonlinearity**, v. 26, n. 3, p. 763–776, 2013. ISSN 0951-7715. Citado nas páginas 20, 35, 38, 39 e 48.
- GALLO, S.; TAKAHASHI, D. Y. Attractive regular stochastic chains: perfect simulation and phase transition. **Ergodic Theory and Dynamical Systems**, Cambridge University Press, v. 34, n. 5, p. 1567–1586, 2014. Citado na página 59.

- GALVES, A.; LÖCHERBACH, E. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. **J. Stat. Phys.**, v. 151, n. 5, p. 896–921, 2013. ISSN 0022-4715. Citado na página [20](#).
- GALVES, A.; LOCHERBACH, E. Modeling networks of spiking neurons as interacting process with memory of variable length. **arXiv preprint arXiv:1502.06446**, 2015. Citado nas páginas [20](#), [21](#) e [53](#).
- HARRIS, T. E. On chains of infinite order. **Pacific J. Math.**, v. 5, p. 707–724, 1955. ISSN 0030-8730. Citado na página [19](#).
- HARUNA, T.; NAKAJIMA, K. Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic markov processes. **The European Physical Journal B**, Springer, v. 86, n. 5, p. 230, 2013. Citado na página [55](#).
- IONESCU, T. Measures dans les espaces produits. **Att. Acad. Naz. Lincei. Rend**, v. 7, p. 208–211, 1949. Citado na página [35](#).
- JOHANSSON, A.; ÖBERG, A. Square summability of variations and convergence of the transfer operator. **Ergodic Theory Dynam. Systems**, v. 28, n. 4, p. 1145–1151, 2008. ISSN 0143-3857. Citado nas páginas [19](#), [20](#), [39](#) e [47](#).
- KEANE, M. Strongly mixing g -measures. **Invent. Math.**, v. 16, p. 309–324, 1972. ISSN 0020-9910. Citado nas páginas [19](#) e [31](#).
- KRICHEVSKY, R.; TROFIMOV, V. The performance of universal encoding. **IEEE Transactions on Information Theory**, IEEE, v. 27, n. 2, p. 199–207, 1981. Citado na página [54](#).
- LEDRAPPIER, F. Principe variationnel et systemes dynamiques symboliques. **Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete**, Springer, v. 30, n. 3, p. 185–202, 1974. Citado na página [19](#).
- LINDVALL, T. **Lectures on the coupling method**. [S.l.]: Courier Corporation, 2002. Citado na página [58](#).
- LONDON, M.; SHCREIBMAN, A.; SEGEV, I. Estimating information theoretic quantities of spike-trains using the context tree weighting algorithm. **Nature neuroscience**, v. 5, 2002. Citado na página [52](#).
- MASSEY, J. Causality, feedback and directed information. In: CITESEER. **Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)**. [S.l.], 1990. p. 303–305. Citado na página [55](#).
- MOYA, J. G. Integration of the information in complex neural networks with noise. Universitat Politècnica de Catalunya, 2011. Citado na página [52](#).
- OLIVEIRA, R. I. Stochastic processes with random contexts: A characterization and adaptive estimators for the transition probabilities. **IEEE Transactions on Information Theory**, IEEE, v. 61, n. 12, p. 6910–6925, 2015. Citado na página [36](#).
- O’NEILL, A.; HUTTER, M.; SHAO, W.; SUNEHAG, P. Adaptive context tree weighting. In: IEEE. **Data Compression Conference (DCC), 2012**. [S.l.], 2012. p. 317–326. Citado na página [54](#).

- ONICESCU, O.; MIHOC, G. Sur les chaines de variables statistiques. **Bull Sci., Math**, v. 59, p. 174–92, 1935. Citado na página 19.
- ORNSTEIN, D. S.; WEISS, B. How sampling reveals a process. **The Annals of Probability**, JSTOR, p. 905–930, 1990. Citado na página 57.
- PARRY, W.; POLLICOTT, M. Zeta functions and the periodic orbit structure of hyperbolic dynamics. **Astérisque**, v. 187, n. 188, p. 1–268, 1990. Citado na página 67.
- RISSANEN, J. A universal data compression system. **IEEE Transactions on information theory**, IEEE, v. 29, n. 5, p. 656–664, 1983. Citado na página 27.
- SANTIS, E. D.; PICCIONI, M. Backward coalescence times for perfect simulation of chains with infinite memory. **Journal of Applied Probability**, Cambridge University Press, v. 49, n. 2, p. 319–337, 2012. Citado nas páginas 20, 34 e 37.
- SHANNON, A. Shannon a. **Mental Health for the Whole Child: Moving Young Clients from Disease & Disorder to Balance & Wellness**, 1948. Citado nas páginas 53 e 54.
- SHIELDS, P. C. The ergodic theory of discrete sample paths. graduate studies in mathematics. **American Mathematical Society**, 1996. Citado na página 57.
- TONG, X. T.; HANDEL, R. V. *et al.* Conditional ergodicity in infinite dimension. **The Annals of Probability**, Institute of Mathematical Statistics, v. 42, n. 6, p. 2243–2313, 2014. Citado na página 48.
- TRUCCOLO, W.; HOCHBERG, L. R.; DONOGHUE, J. P. Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. **Nature neuroscience**, Nature Publishing Group, v. 13, n. 1, p. 105, 2010. Citado na página 53.
- WIENER, N. The theory of prediction. **Mathematics for Engineers**. McGraw-Hill, New York, 1956. Citado na página 55.

