Processos de salto com memória de alcance variável

Douglas Rodrigues Pinto

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Doutor em Ciências

Programa: Estatística

Orientador: Prof. Dr. J. Antonio Galves

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES. Esta tese foi produzida no Núcleo de Apoio à Pesquisa em Modelagem Estocástica e Complexidade/MaCLinC como parte das atividades do Centro de Pesquisa, Inovação e Difusão em Neuromatemática, financiado pela FAPESP (processo 2013/07699-0), e do Programa CAPES-NUFFIC (processo 038/12).

29 de janeiro de 2016

Processos de salto com memória de alcance variável

Esta versão da dissertação/tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 26/01/2016. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. J. Antonio Galves (orientador) - IME-USP
- Prof. Dr. Anatoli Iambartsev - IME-USP
- Prof. Dr. Christophe Pouzat - Université de Paris - Descartes
- Prof. Dr. Jacob Ricardo Fraimann Maus - Universidad de la República
- Prof^{a}. Dr^{a}. Chang Chung Yu Dorea - UNB

Resumo

Resumo

Rodrigues, D. **Processos de salto com memória de alcance variável**. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.
Nessa tese apresentamos uma nova classe de modelos, os processos de saltos com memória de alcance variável, uma generalização a tempo contínuo do processo introduzido em Galves e Löcherbach (2013). Desenvolvemos um novo estimador para a árvore de contexto imersa no processo de salto com memória de alcance variável, considerando mais parâmetros fornecidos pela amostra. Obtivemos também uma cota superior da taxa de convergência da árvore estimada para árvore real, provando a convergência quase certa do estimador.

Palavras-chave: Seleção de modelo, processos de salto com memória de alcance variável, algoritmo contexto.

Abstract

Abstract

Rodrigues, D. **Jump process with memory of variable length**. Thesis (PhD) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.
In this work we deal with a new class of models: the jump processes with variable length memory. This is a continuous-time generalization of the process introduced in Galves and Löcherbach (2013). We present a new estimator for the tree context embedded in this process, considering all information provided by the sample. We also present an exponential upper bound for the rate of convergence, proving then the almost sure convergence of the estimator.
Keywords: model selection, jump processes with variable length memory, algorithm context.

Sumário

1 Introdução **1**

1.1 Apresentação do trabalho 1

1.2 Principais contribuições científicas desta tese 2

1.3 Organização do trabalho 2

2 Cadeias estocásticas com memória de alcance variável **5**

2.1 Cadeias estocásticas com memória de alcance variável 6

2.2 Autômato Probabilístico 9

2.3 Seleção de árvores de contexto 10

2.3.1 Algoritmo contexto 11

2.3.2 Critério da informação bayesiana (BIC) 12

3 Processos de saltos a tempo contínuo com memória de alcance variável **13**

3.1 Definição do Processo 13

4 Seleção Estatística de Modelos **19**

4.1 Inferência nos processos de salto com memória de alcance variável 19

4.2 Estimação das taxas de salto do processo 20

4.3 Estimação da árvore de contexto do processo 23

4.4 Taxa de Convergencia do estimador das taxas de saltos 25

4.5 Taxa de Convergência do estimador da árvore de contexto 41

5 Simulação **47**

5.1 Pseudo-Algoritmo de simulação 47

5.2 Pseudo-Agoritmo de estimação 48

6 Conclusões **51**

Referências Bibliográficas **53**

Capítulo 1

Introdução

1.1 Apresentação do trabalho

Na construção de um algoritmo universal para otimizar a compressão de dados, Rissanen (1983) introduziu a ideia de considerar apenas uma parte relevante do passado - os contextos - para estimar as probabilidades de transição de uma cadeia de símbolos. Assim, os processos são descritos por um conjunto de contextos, que respeita a propriedade do sufixo, e uma família de probabilidades associadas aos contextos, formando assim uma árvore probabilística de contextos. Além de considerar um número menor de variáveis que a cadeia de Markov usual, esse modelo também considera as dependências estruturais presentes nos dados.

Em 2013, Galves e Löcherbach publicaram um inovador artigo, apresentando uma nova classe de modelos a tempo discreto, buscando modelar probabilisticamente a atividade de uma rede neural. Nesse modelo, cada partícula do sistema representa um neurônio, cujos estados possíveis são 0 (repouso) ou 1 (disparo). Enquanto está em repouso, a probabilidade de "disparo"de cada partícula depende da atividade acumulada pelo sistema desde seu último disparo, ou seja, as probabilidades de transição do processo dependem de uma porção variável do passado. Essa característica motivou Galves e Locherbach a modelar esse sistema neuronal como uma cadeia com memória de alcance variável, classe de modelos introduzida por Jorma Rissanen. O artigo de Galves e Löcherbach (2013) inspirou diversos trabalhos na área de neuromatemática, como Duarte e Ost (2015), Hodara e Löcherbach (2015), entre outros, sendo a principal motivação do modelo apresentado nesta tese.

Do ponto de vista da modelagem neuronal, no processo de saltos com memória de alcance variável apresentado nessa tese, a evolução do sistema se dá a tempo contínuo, e o sistema é composto por um conjunto finito de neurônios, com taxa de disparos exponencial, cujo parâmetro depende da configuração do passado do sistema. No modelo de Galves e Löcherbach, os contextos, ou seja, a porção do passado que precisamos conhecer para definir as probabilidades de transição, são dados pelo último instante em que houve o disparo de todas as partículas do sistema. Nesta tese apresentamos uma abstração desse modelo: as taxas de disparo dependem de uma árvore probabilística de contextos qualquer.

Uma questão natural que surge é: dada uma amostra gerada por um processo se salto com memória de alcance variável, como inferir a árvore probabilística de contextos imersa no processo?

Rissanen não apenas introduziu a classe de modelos com memória de alcance variável. Ele também apresentou um método - o algoritmo contexto - para estimar as árvores probabilísticas de contexto. A principal ideia do algoritmo contexto é construir uma árvore completa até uma certa altura e, utilizando uma função custo, podar as folhas da árvore consideradas irrelevantes, até obter uma árvore minimal cujas folhas são os menores contextos que

atendem à função custo.

Uma desvantagem em aplicar o algoritmo contexto para estimar a árvore imersa no processo de saltos com memória de alcance variável é que há perda de informação. Isso ocorre porque ele leva em conta apenas os estados do processo, desconsiderando os intervalos de tempo entre as transições.

No presente trabalho, propomos um novo estimador para a árvore de contextos imersa, levando em conta as taxas de transição do sistema. A ideia principal do estimador é, dada uma sequência de símbolos da amostra, estimar as taxas de transição para essa sequência e obter seu contexto através de uma função custo gerada pelas taxas estimadas. Este método tende a se tornar mais preciso, uma vez que usa uma quantidade maior de informação gerada pela amostra.

A motivação de trabalhar com os intervalos dos tempos de salto do processe surgiu durante a análise de dados de EEG, cujos estímulos são gerados por uma cadeia com memória de alcance variável. O objetivo principal era provar que os sinais de EEG emitidos pelo cérebro durante esses estímulos preservam a mesma estrutura probabilística. Galves, Duarte e Ost tratam da existência de uma cadeia com memória de alcance variável oculta nesse processo gerado pelos EEG's, no artigo *Retrieving context tree from EEG data*, a ser publicado em breve.

1.2 Principais contribuições científicas desta tese

- Introdução de uma nova classe de modelos: os processos de saltos com memória de alcance variável.
- Desenvolvimento de um novo estimador para a árvore de contexto imersa no processo de salto com memória de alcance variável, considerando uma quantidade maior de informações fornecidas pela amostra que os estimadores usuais.
- Obtenção de uma cota superior para a velocidade de convergência da árvore estimada para a árvore real.
- Demonstração da convergência quase certa da árvore calculada pelo estimador proposto para a árvore real.

1.3 Organização do trabalho

No capítulo 2, apresentamos as cadeias estocásticas com memória de alcance variável e suas principais propriedades. Além disso, apresentamos também algumas notações que serão utilizadas ao longo de todo o trabalho.

No capítulo 3, apresentamos o processo de saltos com memória de alcance variável e sua aplicação na modelagem a tempo contínuo da comunicação de um sistema de neurônios e em processos markovianos de saltos binários observados de forma intermitente.

No capítulo 4, apresentamos um estimador para as taxas de salto do processo de salto com memória de alcance variável e sua velocidade de convergência para o valor real da taxa. Em seguida, apresentamos um novo estimador para a árvore de contextos imersa no processo de salto com memória de alcance variável. É também nesse capítulo que apresentamos o principal resultado dessa tese, a convergência quase certa da árvore estimada pelo novo método para a árvore real.

No capítulo 5, construímos os pseudo-algoritmos para simulação de um processo de salto com memória de alcance variável e para o estimador da sua árvore de contextos.

Capítulo 2

Cadeias estocásticas com memória de alcance variável

O objetivo desta tese é introduzir uma nova classe de processo estocástico a tempo contínuo: os processos de salto com memória de alcance variável.

A inspiração desse modelo veio das notas de aula do professor Antonio Galves e da versão a tempo contínuo do problema de modelagem da comunicação neuronal, apresentado em Galves e Löcherbach (2013).

Quando um neurônio emite um potencial de ação, ele perde carga e portanto volta para um estado de repouso, reiniciando o acumulo de carga, para eventualmente emitir um novo potencial de ação. Dessa forma, não é possível fixar a porção do passado que é relevante para inferimos sobre a dinâmica do processo. Esse comportamento não-markoviano sugere uma modelagem considerando o processo do tipo cadeia com memória de alcance variável, introduzido em Rissanen (1983). Isso se deve ao fato da probabilidade de disparo de um neurônio depende da carga acumulada desde o seu último disparo, ou seja, depende de uma porção variável do passado.

Dado um conjunto finito de neurônios $\mathcal{I}$, consideramos que os intervalos entre disparos de um neurônio possuem distribuição exponencial, com taxa dependendo da quantidade de carga recebida desde seu último disparo. Note que, a cada disparo de um neurônio do sistema, as taxas de disparo de todos os elementos que se comunicam com ele são atualizadas.

Sejam $(T_n)_{n\in\mathbb{Z}}$ os instantes de disparos dos neurônios em $\mathcal{I}$. Definimos $(X_n)_{n\in\mathbb{Z}}$ um processo estocástico assumindo valores em $\mathcal{I}$, onde X_n indica o neurônio que emitiu um potencial de ação no instante T_n. Então, $(X_n)_{n\in\mathbb{Z}}$ é uma cadeia com memória de alcance variável, onde a porção do passado que devemos conhecer para definir as probabilidades de transição é o último instante em que observamos o disparo de todos os neurônios.

Assim, tomando $Y_t = X_n$ se $T_n \leq t < T_{n+1}$, temos que $(Y_t)_{t\in\mathbb{R}}$ é um processo de salto a tempo contínuo com memória de alcance variável. Esse modelo é exposto formalmente no capítulo 3.

Nesta tese consideramos uma generalização deste modelo, onde o processo é definido num alfabeto enumerável e a cadeia imersa é uma árvore probabilística de contextos qualquer. Estamos particularmente interessados em realizar inferência e seleção estatística nesta classe de modelos.

A seguir, apresentamos as definições e notações das cadeias com memória de alcance variável e dos autômatos probabilísticos. Tais conceitos serão usados ao longo de todo o texto.

2.1 Cadeias estocásticas com memória de alcance variável

Esses modelos, que são a base da classe de sistemas aqui apresentada, foram introduzidos em Rissanen (1983). A seguir, apresentamos uma breve definição dessas cadeias.

A partir de agora A será sempre um conjunto finito de símbolos, chamado alfabeto. Para quaisquer inteiros $m \leq n$, denotamos $\omega_m^n = (\omega_m, \omega_{m+1}, \cdots, \omega_n)$ como uma sequência de símbolos de A.

Seja $A^{\mathbb{N}}$ o conjunto de todas as sequências finitas de símbolos de A, isto é

$$A^{\mathbb{N}} = \bigcup_{k=1}^{\infty} A^{\{-k,\cdots,-1\}},$$

e $A^{\infty} = A^{\{\cdots,-n,\cdots,-2,-1\}}$ o conjunto de todas as sequências semi-infinitas de símbolos de A. Denotamos, então, $A^* = A^{\mathbb{N}} \cup A^{\infty}$.

Dadas duas sequências u e ω de elementos de A, onde ω é uma sequência com finitos termos, $u\omega$ representa a sequência obtida a partir da concatenação das duas sequências.

Definição 2.1.1. Para $k \in \mathbb{N} \cup \{\infty\}$, dizemos que a sequência $u = u_{-j}^{-1}$ é sufixo próprio de uma sequência $\omega = \omega_{-k}^{-1}$ se $j < k$ e $u_i = \omega_i$ para $i = -1, -2, \cdots, -j$. Denotamos essa relação por $u \prec \omega$.

Se ω é uma sequência finita, então chamamos de $suf(\omega)$ o maior sufixo próprio de ω. Formalmente,

$$suf(\omega_{-k}^{-1}) = \omega_{-(k-1)}^{-1}.$$

Definição 2.1.2. Um subconjunto τ de sequências elementos de A é uma árvore irredutível se satisfaz às seguintes condições:

i) Propriedade do sufixo: Se $\omega \in \tau$, então nenhum sufixo próprio de ω pertence a τ;

ii) Irredutibilidade: Nenhum elemento em τ pode ser substituído por um sufixo próprio sem violar a propriedade de sufixo.

O item *ii*) da definição 2.1.2 diz que, dado $\omega \in \tau$, se substituirmos ω por $suf(\omega)$ em τ, então existe um $v \in \tau$ tal que $suf(\omega) \prec v$, violando a propriedade do sufixo.

Exemplo 2.1.1. Sejam as estruturas a seguir.

A estrutura (a) corresponde a $\tau = \{00, 10, 0, 1\}$. Não é uma árvore irredutível, pois viola a propriedade do sufixo: como $00 \in \tau$ e 0 é um sufixo de 00, então 0 não poderia pertencer a τ.

A estrutura (b) corresponde a $\tau = \{00, 11, 01\}$. Não é uma árvore irredutível, pois viola a irredutibilidade: podemos substituir 00 por 0 sem violar a propriedade de sufixo.

A estrutura (c) corresponde a $\tau = \{00, 10, 1\}$. É uma árvore irredutível, pois respeita a propriedade do sufixo e a irredutibilidade.

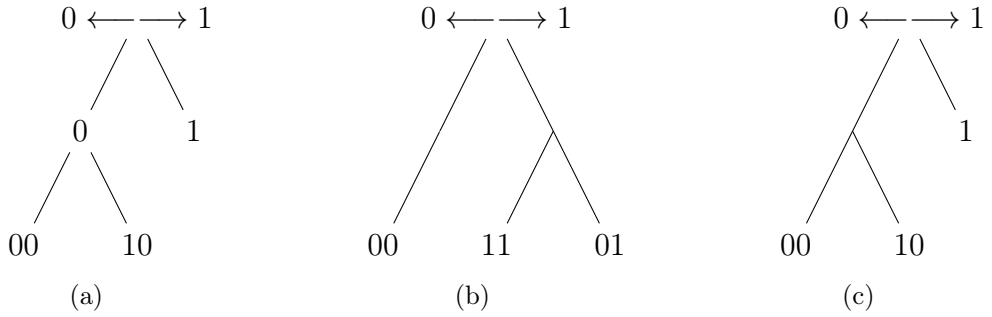

Figura 2.1

Dessa forma, se τ é uma árvore irredutível, então, na linguagem combinatória, τ pode ser interpretada como as folhas de uma árvore com raiz.

Para cada sequências ω de elementos de A , definimos $|\omega|$ como o comprimento da sequência ω, ou seja, é a quantidade de símbolos de A que ela possui. Se τ é uma árvore irredutível, então

$$|\tau| = \max_{\omega \in \tau} |\omega|$$

é a profundidade da árvore. Nessa tese, consideraremos o caso generalizado, onde $|\tau|$ pode ser infinito.

Dada uma árvore irredutível τ e um número natural k, $\tau|_k$ representa a árvore τ truncada no nível k, ou seja,

$$\tau|_k = \{\omega \in \tau \,:\, |\omega| \leq k\} \cup \{\omega \in A^k \,:\, w \prec u \text{ para algum } u \in \tau,\, |u| > k\}.$$

Exemplo 2.1.2. Consideramos a seguinte árvore irredutível

$$\tau = \{0100,\, 1100,\, 0111,\, 1111.\, 000,\, 011,\, 10,\, 01\}$$

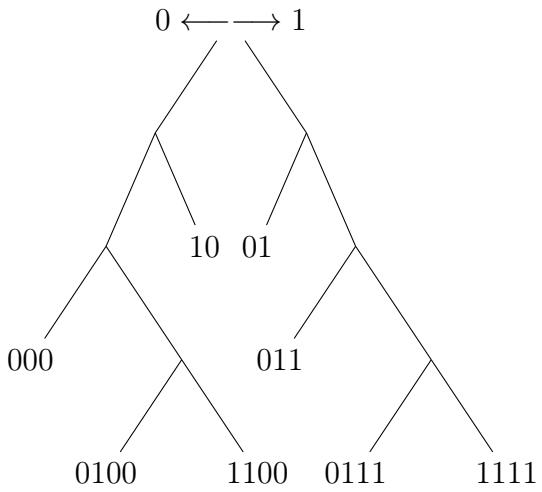

Figura 2.2

Se $k = 3$, por exemplo, para construir $\tau|_3$ devemos truncar a árvore no nível 3, ou seja,

1) Consideramos todas as folhas com comprimento menor ou igual a 3: $\{000, 10, 01, 011\}$.

2) Substituímos as sequencias de comprimento maior que 3 pelo seu sufixo próprio de comprimento igual a 3: $\{100, 111\}$.

Dessa forma, obtemos a seguinte árvore irredutível.

$$\tau|_3 = \{000,\ 100,\ 011,\ 111,\ 10,\ 01\}$$

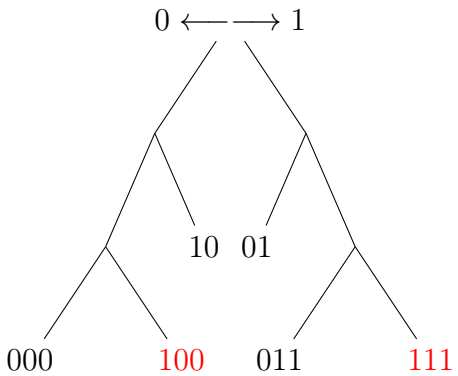

Figura 2.3

Definição 2.1.3. Uma árvore probabilística de contextos em A é um par $(\tau,\, p)$, tal que

i) τ é uma árvore irredutível;

ii) $\{p\,(\cdot|\,\omega)\ ;\ \omega \in \tau\}$ é uma família de probabilidades de transição em A

Seja $(X_n)_{n\in\mathbb{Z}}$ um processo estocástico estacionário em A. Para cada $\omega \in A^k$ e todo $m \in \mathbb{Z}$, definimos a probabilidade estacionária da sequência ω por

$$p\,(\omega) = \mathbb{P}\left(X_{m-1}^{m-k} = \omega\right)$$

Se $p\,(\omega) > 0$, $a \in$ A, então

$$p\,(a|\,\omega) = \mathbb{P}\left(X_m = a|\, X_{m-k}^{m-1} = \omega\right)$$

Definição 2.1.4. Uma sequência ω de elementos de A é um contexto do processo $(X_t)_{t\in\mathbb{Z}}$ se $p(\omega) > 0$ e se

i) para toda sequência u de elementos de A tal que $\omega \prec u$, temos que

$$\mathbb{P}\left(X_m = a|\, X_{m-|u|}^{m-1} = u\right) = p(a|\,\omega),\ \text{ para todo } a \in A,\ m \in \mathbb{Z}.$$

Nesse caso, dizemos também que ω é o contexto da sequência u para o processo $(X_t)_{t\in\mathbb{Z}}$.

ii) se $v \prec w$, então v não satisfaz i).

Definição 2.1.5. Seja (τ, p) uma árvore probabilística de contextos em A. Dizemos que o processo estocástico $(X_t)_{t\in\mathbb{Z}}$ é uma cadeia de alcance variável compatível com (τ, p) se

i) $\omega \in \tau \Leftrightarrow \omega$ é um contexto para $(X_t)_{t\in\mathbb{Z}}$.

ii) para cada $\omega \in \tau$, $p(a|\,\omega) = \mathbb{P}\left(X_m = a|\, X_{m-|\omega|}^{m-1} = \omega\right)$, para todo $a \in A$.

A seguir, apresentamos uma importante ferramenta que utilizaremos no resultado central dessa tese, que é a *função de comprimento do contexto.*

Definição 2.1.6. Sejam $(X_t)_{t\in\mathbb{Z}}$ uma cadeia de alcance variável compatível com (τ, p) assumindo valores em A. O funcional $\ell : A^* \to \mathbb{N} \cup \{\infty\}$ é uma função de comprimento do contexto se, para cada sequência $u \in A^*$,

$$\ell(u) = |C_\tau(u)|,$$

onde

$$C_\tau(u) = \begin{cases} \omega, & \text{se existe } \omega \in \tau \text{ tal que } \omega \prec u \\ \infty, & \text{caso contrário.} \end{cases}$$

Para cada sequência u, $\ell(u)$ nos fornece o tamanho do contexto de u. Intuitivamente, dada uma sequência $x_{-\infty}^{-1}$, a função $\ell(x_{-\infty}^{-1})$ nos diz até que ponto do passado devemos voltar para definir as probabilidades de transição. Quando ℓ é igual a ∞, então devemos voltar mais no passado para definir as probabilidades de transição.

2.2 Autômato Probabilístico

Ron at al (1996) definem autômato probabilístico para um conjunto finito de estados Q. Aqui, apresentamos uma definição de autômatos para um conjunto enumerável de estados. Então, um autômato probabilístico M é uma 5-upla $(Q, A, \Gamma, \delta, \pi)$, onde

Q é um conjunto de estados enumerável;
A é um alfabeto finito;
$\Gamma : Q \times A \to Q$ é uma função de transição;
$\delta : Q \times A \to [0, 1]$ é a função de probabilidade do próximo símbolo;
$\pi : Q \to [0, 1]$ é a distribuição inicial;

onde as funções δ, π e Γ devem satisfazer às seguintes condições:

i) para cada $e \in Q$, $\sum_{a\in A} \delta\,(e, a) = 1$;

ii) $\sum_{e \in Q} \pi(e) = 1$;

iii) Γ é definida somente nos estados $e \in Q$ e $a \in A$ tais que $\delta(e, a) > 0$.

Proposição 2.2.1. Seja (τ, p) uma árvore de contexto probabilística satisfazendo à seguinte propriedade:

$$\text{Para todos } \omega \in \tau,\ a \in A, \text{ se } p(a|\omega) > 0, \text{ então } \omega a \text{ tem um sufixo em } \tau.$$

Então, considerando $Q = \tau$, $\delta = p$, $\Gamma(w, a) = C_\tau(wa)$ e π arbitrário, temos que $(\tau, A, p, C_\tau, \pi)$ é um autômato probabilístico.

Suponha que $(X_n)_{n \in \mathbb{Z}}$ é uma cadeia estocástica com memória de alcance variável associada a (τ, p) satisfazendo a propriedade de autômato probabilístico.

Isso implica que se

$$C_\tau\left(X_{-\infty}^{n-1}\right) = \omega \in \tau \text{ e } X_n = a,$$

então $C_\tau\left(X_{-\infty}^{n}\right) = C_\tau\left(X_{-\infty}^{n-1} a\right)$ é sufixo de $C_\tau\left(X_{-\infty}^{n-1}\right) a$.

Ou seja, temos que $C_\tau\left(X_{-\infty}^{n}\right)$ depende apenas de $C_\tau\left(X_{-\infty}^{n-1}\right)$ e do valor de X_n. Isso motiva a seguinte proposição.

Proposição 2.2.2. Sejam $(X_n)_{n \in \mathbb{Z}}$ uma cadeia estocástica com memória de alcance variável compatível com (τ, p), satisfazendo a propriedade de autômato probabilístico, e $C_n = C_\tau\left(X_{-\infty}^{n}\right)$. Então $(C_n)_{n \in \mathbb{Z}}$ é uma cadeia de Markov assumindo valores nas folhas de τ.

De fato, pela definição temos que C_n depende apenas de C_{n-1} e do valor de X_n, ou seja

$$C_n = g(C_{n-1}, X_n).$$

Temos que $X_n = f(C_{n-1}, U_n)$, onde $U_n \sim U[0, 1]$. Então, temos que $C_n = \tilde{g}(C_{n-1}, U_n)$. Logo,

$$\mathbb{P}\left(C_n | (C_k)_{k=-\infty}^{n-1}\right) = \mathbb{P}(C_n | C_{n-1}).$$

□

2.3 Seleção de árvores de contexto

Apresentamos a seguir dois métodos bastante utilizados para selecionar a árvore de contexto de uma sequência de elementos de um alfabeto finito A: o algoritmo contexto e o critério da informação bayesiana (BIC).

2.3.1 Algoritmo contexto

A ideia principal do algoritmo contexto é a seguinte: dada uma amostra produzida por uma cadeia com memória de alcance variável, consideramos como candidata a árvore de contextos da amostra a árvore completa com uma determinada altura. Utiliza-se então uma função custo, gerada pelas probabilidades de transição empíricas, para podar a árvore candidata, iniciando pelas primeiras folhas, em direção à raiz, até obter uma árvore minimal bem adaptada à amostra. A função custo verifica o quão relevante é voltarmos até um determinado ponto do passado para determinar as probabilidades de transição.

A seguir, apresentamos uma variação do algoritmo contexto introduzido por Rissanen.

Sejam $X_1, X_2, \cdots, X_n$ uma amostra gerada por uma cadeia estacionária $(X_i)_{i\in\mathbb{Z}}$ compatível com a árvore probabilística de contextos (τ, p). Para qualquer natural $k < n$, definimos $N_n(x_{-k}^{-1})$ como o número de ocorrências da sequência x_{-k}^{-1} na amostra, ou seja,

$$N_n\left(x_{-k}^{-1}\right) = \sum_{i=1}^{n-k+1} \mathbb{I}_{\{X_i^{i+k-1}=x_{-k}^{-1}\}},$$

e para todo $a \in A$, para toda sequência ω de elementos de A tal que $N_n(\omega) > 0$, o estimador da probabilidade de transição $p(a|\omega)$ é dado por

$$\hat{p}_n(a|\omega) = \frac{N_n(\omega a)}{N_n(\omega)}. \tag{2.1}$$

A função $\Delta_n(\omega)$ é definida como a maior diferença entre a probabilidade de transição empírica associada à sequência ω e a associada à sequência $suf(\omega)$.

$$\Delta_n(\omega) = \max_{a\in A} \left|\hat{p}_n(a|\omega) - \hat{p}_n(a|suf(\omega)\,)\right|.$$

Seja δ um número positivo dado. Para cada ω, verificamos se $\Delta_n(\omega) > \delta$. Caso positivo, mantemos ω como um contexto. Caso contrário, podamos as folhas, e repetimos o procedimento com as demais folhas.

Formalmente, definimos a árvore obtida pelo algoritmo contexto da seguinte forma.

Definição 2.3.1. Para $\delta > 0$, $1 < d < n$, a árvore de contextos estimada pelo algoritmo contexto é dada por

$$\begin{aligned}\hat{\tau}_n^{\delta,d} \;=\; & \{\omega : |\omega| \le d;\ N_n(a\omega) > 0,\ \Delta_n(a\,suf(\omega)\,) > \delta \text{ para algum } a \in A;\ \text{e } \Delta_n(u\omega) \le \delta \\ & \text{para todo } u \text{ tal que } |u| < d - |\omega|,\ \text{com } N_n(u\omega) \ge 1\}.\end{aligned} \tag{2.2}$$

Em Galves e Leonardi (2008) é apresentado uma cota superior exponencial para o algoritmo contexto no caso em que são admitidas árvores infinitas.

Para o mesmo caso é provado também que, sob certas condições, a árvore de contextos estimada truncada converge quase certamente para a árvore geradora da amostra, truncada no mesmo nível.

Teorema 2.3.1 (Galves e Leonardi, 2008)**.** Seja (τ, p) uma árvore probabilística de contextos satisfazendo:

i) $\alpha_0 := \sum_{a\in A} \inf_{\omega\in\tau} p(a|\omega) > 0.$

ii) $\alpha := \sum_{k\in\mathbb{N}} \left(1 - \inf_{u\in A^k} \sum_{a\in A} \inf_{\omega\in\tau:\omega\succ u} p(a|\omega)\right) < \infty.$

Seja $(X_n)_{n\in\mathbb{Z}}$ uma cadeia estocástica estacionária compatível com (τ, p). Então, para qualquer inteiro k, qualquer d satisfazendo

$$d > \max_{u\notin\tau, |u|\prec k} \min\{l \,:\, \exists\, \omega \in \mathcal{C}_l, \omega \succ u\},$$

onde $\mathcal{C}_l = \{u \in \tau|_l \,:\, p(a|u) \neq p(a|suf(u))$ para algum $a \in A\}$; qualquer $\delta < D_d$ e qualquer n tal que

$$n > \frac{2(|A|+1)}{\min(\delta, D_d - \delta)\,\epsilon_d} + d,$$

onde $D_d = \min_{u\in C_d} \max_{a\in A}\{|p(a|u) - p(a|suf(u))|\}$, e $\epsilon_d = \min\{p(\omega) \,:\, |\omega| \leq d$ e $p(\omega) > 0\}$, temos que

$$\hat{\tau}_n^{\delta,d}|_k \to^{q.c.} \tau|_k, \tag{2.3}$$

2.3.2 Critério da informação bayesiana (BIC)

Em 2006, Imre Csiszár e Zsolt Talata propuseram um método consistente para estimar a árvore de contextos, em um tempo linear, utilizando o critério da informação bayesiana (BIC). Dentro de um conjunto de árvores candidatas, que são especificadas por um limite máximo de tamanho, procura-se pela árvore que maximiza a função de verossimilhança penalizada. Essa árvore é a árvore de contextos estimada pelo método BIC.

Para uma amostra $X_1, X_2, \cdots, X_n$, a função de verossimilhança penalizada BIC para cada árvore candidata τ é dada por

$$BIC_\tau(X_1^n) = -\log\left(\hat{\mathbb{P}}_\tau\left(X_1^n\right)\right) - \frac{(|A|-1)\,|\tau|\log(n)}{2},$$

onde $\hat{\mathbb{P}}_\tau\left(X_1^n\right)$ é a máxima verossimilhança da amostra quando consideramos a árvore de contextos τ.

A árvore de contextos estimada pelo método BIC é dada por

$$\hat{\tau}(X_1^n) = arg\max\left\{BIC_\tau(X_1^n) \,:\, \tau \in \mathcal{J}_n\right\}, \tag{2.4}$$

onde $\mathcal{J}_n$ é o conjunto de todas as árvores indexadas por símbolos do alfabeto A e de altura menor ou igual a $\log_{|A|}(n)$.

Em Csiszár e Talata (2006) é demonstrado que a árvore estimada pelo método BIC truncada converge quase certamente para a árvore geradora da amostra no caso em que a árvore cresce em velocidade logaritmica.

Teorema 2.3.2 (Csiszár e Talata, 2006)**.** Se $d(\tau) < \infty$ e $\hat{\tau}(X_1^n)$ é o estimador da árvore de contextos definido em (2.4), então

$$\hat{\tau}(X_1^n) \to^{q.c.} \tau.$$

No caso generalizado, então para qualquer inteiro k,

$$\hat{\tau}(X_1^n)|_k \to^{q.c.} \tau|_k.$$

Garivier (2006) generaliza o resultado de consistência forte do estimador sem a necessidade de hipóteses sobre a velocidade de crescimento da árvore.

Capítulo 3

Processos de saltos a tempo contínuo com memória de alcance variável

Neste capítulo vamos apresentar uma nova classe de processos estocásticos, que será o objeto principal de estudo nesta tese: os processos de salto a tempo contínuo com memória de alcance variável.

Enquanto os processos markovianos de saltos possuem uma cadeia de Markov como cadeia imersa, nosso processo se caracteriza por possuir uma árvore probabilística de contextos qualquer imersa, inclusive infinitas, ocasionando ao processo uma dependência de uma porção variável do passado. As taxas de transição do processo não são são constantes: a cada transição a taxa é atualizada através de uma função que também depende de uma parte variável dos estados passados do processo.

Um importante exemplo de aplicação desse processo é na modelagem da comunicação de um sistema de neurônios. Apresenta,os no exemplo 3.1.1 uma modelagem a tempo contínuo do processo apresentado por Galves e Löcherbach (2013).

3.1 Definição do Processo

No que segue, A é um alfabeto finito, (τ, p) é uma árvore probabilística de contextos em A e $q : \tau \to (0, +\infty)$ é uma função arbitrária.

Definição 3.1.1. Seja $(Y_t)_{t\in\mathbb{R}}$ um processo estocástico à tempo contínuo, assumindo valores no alfabeto finito A. O processo $(Y_t)_{t\in\mathbb{R}}$ é dito um processo de saltos com memória de alcance variável associado a (τ, p) e com taxa q se existirem um processo pontual $\cdots < T_{-1} < T_0 \leq 0 < T_1 < T_2 < \cdots$ e uma cadeia $(X_n)_{n\in\mathbb{Z}}$, $X_n \in A$, tais que

i) $Y_t = X_n$, se $T_n \leq t < T_{n+1}$;

ii) $(X_n)_{n\in\mathbb{Z}}$ é uma cadeia estocástica de alcance variável associado a (τ, p);

iii) $\mathbb{P}\left(T_{n+1} - T_n > t | X_{-\infty}^n\right) = \exp\left[-\frac{t}{q\left(c_\tau(X_{-\infty}^n)\right)}\right]$, onde $c_\tau(X_{-\infty}^n)$ é o único sufixo de $X_{-\infty}^n$ que é folha de τ.

Nosso primeiro exemplo trata da modelagem da comunicação entre um sistema de neurônios de tamanho N.

Exemplo 3.1.1 (Sistema de N neurônios em tempo contínuo)**.** A atividade neural é manifestada pela emissão de potenciais de ação (disparos). Essa atividade é realizada por cada um dos neurônios, em um sistema constituído por uma grande quantidade dessas partículas, que interagem entre si.

A sucessão de disparos que um neurônio é capaz de emitir depende de seu estado, dos estímulos provenientes de outros neurônios e de estímulos externos ao sistema (como o efeito "*leakage*", por exemplo).

O potencial de ação ocorre com uma probabilidade que é uma função crescente da concentração de íons positivos no interior da célula . Após a emissão de um potêncial de ação, o neurônio perde toda sua carga e volta a um estado de repouso, reiniciando o processo de acumulo de carga.

Galves e Löcherbach (2013) consideram um sistema cuja probabilidade de disparo de neurônio depende da atividade acumulada do sistema desde o instante de seu último disparo; e os neurônios podem exercer influência positiva, negativa ou neutra nessa probabilidade de disparo.

Dessa forma, apresentamos a seguinte modelagem a tempo contínuo para a comunicação entre um sistema de neurônios.

Sejam $A = \{1, 2, \cdots, N\}$ e $\phi : \mathbb{R} \to \mathbb{R}$ uma função mensurável crescente tal que $\phi(0) = \delta > 0$.

Seja a árvore probabilística de contextos (τ, p), tal que $C_\tau\left(x_{-\infty}^{-1}\right) = x_{L(x_{-\infty}^{-1})}^{-1}$, onde

$$L(x_{-\infty}^{-1}) = \max\{m \leq -1 : \cup_{k=m}^{-1}\{x_k\} = A\}$$

e cuja probabilidade de transição é dada por

$$p\left(j \,|\, C_\tau\left(x_{-\infty}^{-1}\right)\right) = \frac{\phi\left(\mathcal{U}(j, C_\tau(x_{-\infty}^{-1}))\right)}{\sum_{i\in A} \phi\left(\mathcal{U}(i, C_\tau(x_{-\infty}^{-1}))\right)},$$

onde para

$$L^i(x_{-\infty}^{-1}) = \sup\{m \leq -1 : x_m = i\},$$

$$\omega_{\cdot\to i} : A \to \mathbb{R} \text{ tal que } \omega_{i\to i} = 0,$$

definimos

$$\mathcal{U}\left(i, C_\tau\left(x_{-\infty}^{-1}\right)\right) = \sum_{l\in A} \omega_{l\to i} \sum_{m=L^i(x_{-\infty}^{-1})}^{-1} \mathbb{I}_{\{X_m = l\}}$$

Definimos agora $(X_n)_{n\in\mathbb{Z}}$ uma cadeia estocástica com memória de alcance variável assumindo valores em A, compatível com a árvore probabilista de contextos (τ, p).

Seja $(T_n)_{n\in\mathbb{Z}}$ um processo pontual tal que

$$\mathbb{P}\left(T_{n+1} - T_n > t | X_{-\infty}^n\right) = \exp\left[-\frac{t}{q\left(C_\tau\left(X_{-\infty}^n\right)\right)}\right],$$

onde $q\left(C_\tau\left(X_{-\infty}^n\right)\right) = \sum_{i\in A} \phi\left(\mathcal{U}(i, C_\tau(X_{-\infty}^n)\right)$.

Então, defino o processo a tempo continuo $(Y_t)_{t\in\mathbb{R}}$ por

$$Y_t = X_n, \text{ se } T_n \leq t < T_{n+1}. \tag{3.1}$$

Assim, o processo $(Y_t)_{t\in\mathbb{R}}$ definido em (3.1) é um processo de salto a tempo continuo com memória de alcance variável.

Do ponto de vista da modelagem de um sistema neuronal, o conjunto A representa um conjunto de neurônios, para cada i, $i \in A$, $\omega_{j\to i}$ são os pesos sinápticos, ou seja, representa uma quantificação da influência que o disparo do neurônio j exerce na probabilidade de disparo do neurônio i. Essa influência pode ser excitatória ($\omega_{j\to i} > 0$), inibitória ($\omega_{j\to i} < 0$) ou neutra ($\omega_{j\to i} = 0$).

Para cada t, Y_t nos fornece a informação do último neurônio a disparar num tempo menor ou igual a t, T_n representa o n-ésimo instante de disparo de algum neurônio no sistema e X_n informa qual o neurônio que disparou no instante T_n.

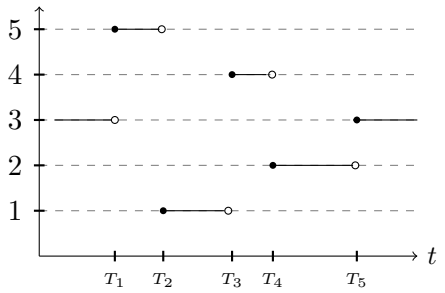

Figura 3.1: *Exemplo de evolução do processo $(Y_t)_{t\in\mathbb{R}}$ para um sistema com 5 neurônios*

Supomos agora que um sinal é emitido por uma fonte e deve percorrer dois transmissores até chegar ao receptor final. Se algum dos transmissores está ocupado, então a mensagem não chega ao receptor final. Supomos que os transmissores são independentes entre sí e que possuem um comportamento markoviano. Então, definindo $Y_t = 1$ quando os dois transmissores estão livres no tempo t e $Y_t = 0$ caso contrário, temos que $(Y_t)_{t\in\mathbb{R}}$ é um processo de salto a tempo contínuo com memória de alcance variável. Formalizamos essa modelagem no exemplo a seguir.

Exemplo 3.1.2 (Processo markoviano de saltos binários observados de forma interminente)**.** Sejam $A = \{0, 1\}$ e $(W_n)_{n\in\mathbb{Z}}$ uma cadeia de Markov, assumindo valores em A, com matriz de probabilidade de transição r, tal que $r(0|1) = r(1|0) = 1$. Sejam $\alpha(0) \in (0, +\infty)$ e $\alpha(1) \in (0, +\infty)$ dois parâmetros reais quaisquer fixados.

Definimos $(Z_t)_{t\in\mathbb{R}}$ como o processo markoviano de saltos com $(W_n)_{n\in\mathbb{Z}}$ como cadeia de Markov imersa e com taxas de espera $\alpha(0)$ e $\alpha(1)$, ou seja, $(Z_t)_{t\in\mathbb{R}}$ é o processo de saltos cujo gerador infinitesimal é, para toda função contínua e limitada, dado por

$$\mathcal{L}f(x) = \sum_{y\in A} \alpha(x) r(y|x) \left[f(y) - f(x)\right]. \tag{3.2}$$

Seja agora $\left(\overline{Z}_t\right)_{t\in\mathbb{R}}$ um outro processo markoviano de saltos assumindo valores em $A = \{0, 1\}$ com cadeia imersa $\left(\overline{W}_n\right)_{n\in\mathbb{Z}}$ possuindo a mesma matriz de probabilidades de transição r e com taxas de espera $\overline{\alpha}(0) \in (0, +\infty)$ e $\overline{\alpha}(1) \in (0, +\infty)$.

Supomos $(Z_t)_{t\in\mathbb{R}}$ e $\left(\overline{Z}_t\right)_{t\in\mathbb{R}}$ independentes entre si e definimos o processo

$$Y_t = Z_t \cdot \overline{Z}_t.$$

Esse é um processo de saltos com memória de alcance variável associado à árvore $\tau = \{10^k : k = 0, 1, \cdots\} \cup \{0^\infty\}$.

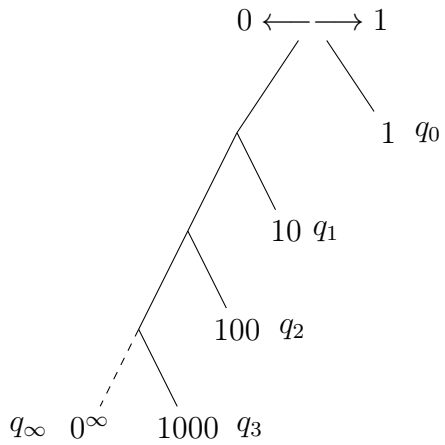

Figura 3.2: *árvore de contextos do processo markoviano de saltos binários observados de forma interminente*

Observe que se observarmos no tempo t um ímpar de zeros no contexto, então $(Z_t, \overline{Z}_t) = (0, 1)$ ou $(1, 0)$; e se observarmos um número par de zeros, então $(Z_t, \overline{Z}_t) = (0, 0)$.

Dessa forma, as taxas $q : \tau \to (0, \infty)$ são dadas por

$$\begin{aligned}
&q(1) = \alpha(1) + \overline{\alpha}(1)\\
&q(10) = (\alpha(0) + \overline{\alpha}(1)) \cdot \frac{\alpha(1)}{\alpha(1) + \overline{\alpha}(1)} + (\alpha(1) + \overline{\alpha}(0)) \cdot \frac{\overline{\alpha}(1)}{\alpha(1) + \overline{\alpha}(1)}\\
&q(10^{2k}) = \alpha(0) + \overline{\alpha}(0)\\
&q(10^{2k+1}) = (\alpha(1) + \overline{\alpha}(0)) \cdot \frac{\alpha(0)}{\alpha(0) + \overline{\alpha}(0)} + (\alpha(0) + \overline{\alpha}(1)) \cdot \frac{\overline{\alpha}(0)}{\alpha(0) + \overline{\alpha}(0)},
\end{aligned}$$

onde $k = \{1, 2, \cdots\}$.

As probabilidades de transição p da cadeia imersa $(X_n)_{n\in\mathbb{Z}}$ em $(Y_t)_{t\in R}$ são dadas por

$$\begin{aligned}
&p(0|1) = 1\\
&p(0|01) = \frac{\overline{\alpha}(1)}{\alpha(1) + \overline{\alpha}(1)} \cdot \frac{\alpha(1)}{\alpha(1) + \overline{\alpha}(0)} + \frac{\alpha(1)}{\alpha(1) + \overline{\alpha}(1)} \cdot \frac{\overline{\alpha}(1)}{\alpha(0) + \overline{\alpha}(1)}\\
&p(0|0^{2k}1) = 1\\
&p(0|0^{2k+1}1) = \frac{\overline{\alpha}(1)}{\alpha(0) + \overline{\alpha}(1)} \cdot \frac{\overline{\alpha}(0)}{\alpha(0) + \overline{\alpha}(0)} + \frac{\alpha(1)}{\alpha(1) + \overline{\alpha}(0)} \cdot \frac{\alpha(0)}{\alpha(0) + \overline{\alpha}(0)},
\end{aligned}$$

onde $k = \{1, 2, \cdots\}$.

Como

$$\sum_{k=0}^{\infty} p(1\,|0^{2k+1}1) = +\infty,$$

podemos afirmar pelo Lema de Borel-Cantelli que há uma ocorrencia infinita de 1's, logo o processo é não trivial e toda sequência $x_{-\infty}^{-1}$ possui um contexto em τ.

Proposição 3.1.1. As cadeias imersas dos exemplos 3.1.2 e 3.1.1 são automatas probabilistas.

Demonstração:

No exemplo 3.1.2 $\tau = \{10^k : k = 0, 1, \cdots\} \cup \{0^\infty\}$. Seja $C_\tau\left(X_{-\infty}^{-1}\right) = \mathrm{w}$.

Então, para $a \in \{0, 1\}$,

$$C_\tau\left(X_{-\infty}^{0}\right) = C_\tau\left(X_{-\infty}^{-1}a\right) = \begin{cases} a, \text{ se } a = 1 \\ wa, \text{ se } a = 0. \end{cases}$$

Logo, $C_\tau\left(X_{-\infty}^{0}\right)$ é sufixo de $C_\tau\left(X_{-\infty}^{-1}\right)a$.

No exemplo 3.1.1 τ são todas as sequências $x_{-k}x_{-k+1}\cdots x_{-1}$, $k \geq |A|$, tais que $x_{-k} \neq x_{-i}$, $\forall\, i = 1, \cdots, k-1$, e $\cup_{j=-k}^{-1}\{X_k\} = A$.

Seja $C_\tau\left(X_{-\infty}^{-1}\right) = X_{-L^i(X_{-\infty}^{-1})}^{-1}$, para $L^i(X_{-\infty}^{-1}) < \infty$.

Então, para $X_0 = a \in A$, temos que

$$C_\tau\left(X_{-\infty}^{0}\right) = C_\tau\left(X_{-\infty}^{-1}a\right) = \begin{cases} X_{-L^i(X_{-\infty}^{-1})}^{-1}a, \text{ se } a \neq X_{-L^i(X_{-\infty}^{-1})} \\ X_{-L^i(X_{-\infty}^{-1})+l}^{-1}a, \text{ se } a = X_{-L^i(X_{-\infty}^{-1})}, \end{cases}$$

para algum $l < L^i(X_{-\infty}^{-1})$. Logo, $C_\tau\left(X_{-\infty}^{0}\right)$ é sufixo de $C_\tau\left(X_{-\infty}^{-1}a\right)$.

No caso $L^i(X_{-\infty}^{-1}) = \infty$ fica ainda mais clara a propriedade de automata probabilista.

Capítulo 4

Seleção Estatística de Modelos

Nesse capítulo apresentamos um novo estimador para a árvore probabilística de contextos imersa no processo de saltos com memória de alcance variável. A ideia principal do estimador é a seguinte. Obtemos primeiramente uma estimativa para as taxas de transição de uma sequência de símbolos observadas na amostra e, através de de uma função custo gerada pelas taxas estimadas, estimamos o contexto da sequência. A árvore construída por esse algoritmo é uma árvore irredutível e seu truncamento em qualquer altura finita converge quase certamente para a árvore geradora da amostra truncada na mesma altura, Esse é o conteúdo do teorema 4.3.1, principal resultado desta tese.

4.1 Inferência nos processos de salto com memória de alcance variável

Seja $(X_i, T_i)_{i=1}^n$ uma amostra gerada por um processo de saltos com memória de alcance variável $(Y_t)_{t\in\mathbb{R}}$ com parâmetros $(\tau,\, p,\, q)$, como na Definição 3.1.1, onde

- (τ, p) é uma árvore probabilística de contextos;
- $q : \tau \to (0, +\infty)$ e uma função real positiva definida nas folhas de τ, representando as taxas de salto do processo;
- $(X_i)_{i=1}^n$ é uma cadeia com memória de alcance variável compatível com (τ, p), assumindo valores num alfabeto finito A, onde, para cada i, X_i indica o estado do processo após o i-ésimo salto;
- $(T_i)_{i=1}^n$ é um processo pontual crescente, onde, para cada i, T_i representa o instante do i-ésimo salto do processo.

Vamos considerar, ainda, que o processo $(Y_t)_{t\in\mathbb{R}}$ satisfaz às seguintes hipóteses:

Hipótese 1: $\alpha_0 := \sum_{a\in A} \inf_{\omega\in\tau} p(a|\omega) > 0.$

Hipótese 2: $\alpha := \sum_{k\in\mathbb{N}} \left(1 - \inf_{u\in A^k} \sum_{a\in A} \inf_{\omega\in\tau:\omega\succ u} p(a|\omega)\right) < \infty.$

Hipótese 3: Para todo $\omega \in \tau$, $0 < q(\omega) < +\infty$.

Hipótese 4: Para todo $av \in \tau$, onde $a \in A$, temos que

$$q(av) \neq \sum_{u:uv\in\tau} \frac{p(uv)}{p(v)} q(uv).$$

A hipótese 4 nos diz que , para cada elemento $av \in \tau$, sua taxa de salto não pode ser escrita como uma média ponderada das taxas dos descendentes de seu sufixo v, com ponderação dada por $\left(\frac{p(uv)}{p(v)}\right)_{u:uv\in\tau}$.

Na demonstração do teorema 4.4.3 ficará mais claro a necessidade dessa hipótese.

No caso especial onde temos $A = \{a, b\}$ e, para alguma sequência v, temos que av e bv pertencem a τ, ocorre que

$$\begin{aligned} q(av) \neq \frac{p(av)}{p(v)} q(av) + \frac{p(bv)}{p(v)} q(bv) \quad &\Leftrightarrow \quad \left(1 - \frac{p(av)}{p(v)}\right) q(av) \neq \frac{p(bv)}{p(v)} q(bv) \\ &\Leftrightarrow \quad \frac{p(bv)}{p(v)} q(av) \neq \frac{p(bv)}{p(v)} q(bv) \\ &\Leftrightarrow \quad q(av) \neq q(bv). \end{aligned}$$

A partir de agora, sempre que nos referirmos a uma amostra do processo $(Y_t)_{t\in\mathbb{R}}$, vamos supor que ela satisfaz às hipóteses 1-4.

Supondo que os parâmetros (τ, p, q) são desconhecidos, o objetivo principal dessa tese é apresentar um novo estimador para o conjunto de contextos τ, baseado nas taxas de saltos do processo.

Dessa forma, nosso primeiro objetivo é construir um estimador $\hat{q}_n(\cdot)$ para as taxas de salto $q(\cdot)$ do processo.

Na próxima seção apresentamos um estimador consistente e não-viciado para $q(\omega)$, para todo $\omega \in \tau$. Além disso, demonstramos importantes propriedades de $\hat{q}_n(\cdot)$, essenciais para para sua utilização como principal elemento na estimação de τ.

4.2 Estimação das taxas de salto do processo

Dada uma amostra $(X_i, T_i)_{i=1}^n$, seja $\omega \in A^k$ uma subsequência de $(X_i)_{i=1}^n$, tal que $N_n(\omega) > 0$, onde $N_n(\omega)$ é o número de vezes que a sequência ω é observada na amostra, ou seja,

$$N_n(\omega) = \sum_{i=k}^{n} \mathbb{I}_{\{X_{i-k+1}^i = \omega\}}.$$

Então, definimos o estimador $\hat{q}_n(\cdot)$ como a média dos intervalos dos tempos de salto após observarmos a ocorrência da sequência ω, ou seja,

$$\hat{q}_n(\omega) = \frac{1}{N_n(\omega)} \sum_{i=k}^{n-1} (T_{i+1} - T_i) \, \mathbb{I}_{\{X_{i-k}^{i-1} = \omega\}}. \tag{4.1}$$

Apresentamos a seguir o principal teorema dessa seção, enunciando a consistência forte do estimador $\hat{q}_n(\cdot)$ definido acima.

Teorema 4.2.1. Dada uma amostra $(X_i, T_i)_{i=1}^n$, para toda sequência ω de elementos de A tal que $\omega \in \tau$,

$$\hat{q}_n(\omega) \to^{q.c} q(\omega). \tag{4.2}$$

Demonstração: A demonstração será apresentada na seção 4.4. A ideia principal será provar que, para todo $\omega \in \tau$, $\hat{q}_n(\omega)$ converge para $q(\omega)$ com velocidade exponencial, e daí obtemos a consistência forte de $\hat{q}_n(\cdot)$.

Um corolário importante desse teorema diz que, se tomarmos um sequência que contém um contexto como sufixo, então o resultado de convergência acima também é valido.

Corolário 4.2.1. Dada uma amostra $(X_i, T_i)_{i=1}^n$, sejam $\omega \in \tau$ e u uma sequência de elementos de A, tal que $N_n(u\omega) > 0$. Então

$$\hat{q}_n(u\omega) \to^{q.c} q(\omega).$$

Demonstração: A demonstração será apresentada na seção 4.4.

Não é difícil ver que, para todo $\omega \in \tau$, $\hat{q}_n(\omega)$ é um estimador não viciado de $q(\omega)$, como demonstramos na proposição a seguir.

Proposição 4.2.1. Para todo $\omega \in \tau$,

$$\mathbb{E}\Big(\hat{q}_n(\omega)\Big) = q(\omega).$$

Demonstração: Dada uma amostra $(X_i, T_i)_{-(k-1)}^n$ e uma sequência $\omega = x_{-k}^{-1}$ tal que $N_n(\omega) > 0$. Sejam

$$\begin{aligned}
&n_1 = \inf\big\{j = 1, 2, \cdots n \,:\, X_{j-k+1}^j = \omega\big\}\\
&n_2 = \inf\big\{j > n_1 \,:\, X_{j-k+1}^j = \omega\big\}\\
&\cdots\\
&n_{N_n(\omega)} = \inf\big\{j > n_{N_n(\omega)-1} \,:\, X_{j-k+1}^j = \omega\big\}.
\end{aligned}$$

Para $i = 1, 2, \cdots, N_n(\omega)$, definimos $Z_i = (T_{n_i+1} - T_{n_i})$. Assim,

$$\sum_{i=k+1}^{n} (T_{i+1} - T_i)\, \mathbb{I}_{\{X_{i-k}^{i-1} = \omega\}} = \sum_{i=1}^{N_n(\omega)} Z_i,$$

ou seja,

$$\hat{q}_n(\omega) = \frac{1}{N_n(\omega)} \sum_{i=k}^{N_n(\omega)} Z_i.$$

Por hipótese, temos que, para cada i, as variáveis aleatórias $Z_1, Z_2, \cdots Z_i$ são condicionalmente independentes, dado $X_{-(k-1)}^{i-1}$. Então, para cada $j = 1, \cdots, n-k$, temos que

$$\mathbb{E}\Big(\hat{q}_n(\omega)|N_n(\omega)=j\Big)=\mathbb{E}\left(\frac{1}{j}\sum_{i=1}^{j}Z_i\right)=\frac{1}{j}\sum_{i=1}^{j}\mathbb{E}(Z_i).$$

Como, para cada i, $\mathbb{E}(Z_i)=q(\omega)$, temos que

$$\mathbb{E}\Big(\hat{q}_n(\omega)|N_n(\omega)=j\Big)=q(\omega)$$

Logo, concluímos que

$$\mathbb{E}\left(\hat{q}_n(\omega)\right)=\mathbb{E}\left[\mathbb{E}\left(\hat{q}_n(\omega)|N_n(\omega)\right)\right]=\mathbb{E}\left[q(\omega)\right]=q(\omega).$$

□

Embora as taxas se saltos $\{q(\omega),\,\omega\in\tau\}$ estejam definidas apenas nos elementos de τ, o estimador $\hat{q}_n(\cdot)$ fornece estimativas de taxas para qualquer sequência de elementos de A.

Dessa forma, o estimador $\hat{q}_n(\cdot)$ será interessante se, dado um elemento $\omega\in\tau$, e uma sequência qualquer u de elementos de A, obtivermos que, para um n suficientemente grande, $\hat{q}_n(u\omega)$ tenha um valor arbitrariamente próximo de $\hat{q}_n(\omega)$.

O teorema 4.2.2 a seguir garante que, se ω é um elemento de τ, então tomando qualquer sequência maior que possua ω como sufixo, a diferença do valor estimado das suas taxas converge quase certamente a zero.

Teorema 4.2.2. Dada uma amostra $(X_i,T_i)_{i=1}^n$, sejam $\omega\in\tau$ e u uma sequência de elementos de A, tal que $N_n(u\omega)>0$. Então, temos que

$$|\hat{q}_n(\omega)-\hat{q}_n(u\omega)|\to^{q.c}0$$

quando $n\to\infty$.

Demonstração: Queremos mostrar que, para todo $\epsilon>0$,

$$\mathbb{P}\left(|\hat{q}_n\left(\omega\right)-\hat{q}_n(u\omega)|\geq\epsilon\text{ i.v.}\right)=0.$$

De fato, temos que, para $\epsilon>0$,

$$\begin{aligned}&\mathbb{P}\left(|\hat{q}_n(\omega)-\hat{q}_n(u\omega)|\geq\epsilon\right)=\mathbb{P}\left(|\hat{q}_n(\omega)-q(\omega)+q(\omega)-\hat{q}_n(u\omega)|\geq\epsilon\right)\\&\leq\mathbb{P}\left(|\hat{q}_n(\omega)-q(\omega)|+|q(\omega)-\hat{q}_n(u\omega)|\geq\epsilon\right)\\&\leq\mathbb{P}\left(|\hat{q}_n(\omega)-q(\omega)|\geq\frac{\epsilon}{2}\right)+\mathbb{P}\left(|q(\omega)-\hat{q}_n(u\omega)|\geq\frac{\epsilon}{2}\right)\end{aligned}$$

Pelo teorema 4.2.1, temos que

$$\mathbb{P}\left(|\hat{q}_n(\omega)-q(\omega)|\geq\frac{\epsilon}{2}\text{ i.v.}\right)=0,$$

e pelo corolário 4.2.1, obtemos

$$\mathbb{P}\left(|q(\omega)-\hat{q}_n(u\omega)|\geq\frac{\epsilon}{2}\text{ i.v.}\right)=0.$$

Logo, concluímos que

$$\mathbb{P}\left(|\hat{q}_n(u) - \hat{q}_n(u\omega)| \geq \epsilon \text{ i.v.}\right) = 0.$$

□

O corolário 4.2.1 nos diz que, se ω é um contexto de τ, então para qualquer sequência finita u de elementos de A, $\hat{q}_n(u\omega)$ converge quase certamente para $q(\omega)$. Uma questão natural que surge é, se v é sufixo próprio de algum contexto de τ, qual é o comportamento assintótico de $\hat{q}_n(v)$?

O teorema a seguir mostra que, se τ é uma árvore de contextos finita e v é sufixo próprio de algum contexto de τ, então $\hat{q}_n(v)$ converge quase certamente para uma média ponderada das taxas de salto dos contextos dos quais v é sufixo próprio. Lembremos que, pela irredutibilidade da árvore de contextos τ, se v é sufixo próprio de algum elemento de τ, então existem pelo menos dois contextos dos quais v é sufixo próprio.

Teorema 4.2.3. Dada uma amostra $(X_i, T_i)_{i=1}^n$, supomos que a árvore de contextos τ é finita, com $|\tau| = d$. Seja v uma sequência de elementos de A, tal que v é sufixo próprio de algum elemento de τ, isto é, existe uma outra sequência finita u, $|u| \geq 1$, tal que $uv \in \tau$. Então

$$\hat{q}_n(v) \to^{q.c} \sum_{u\,:\,uv\in\tau} \frac{p(uv)}{p(v)} q(uv).$$

Demonstração: A demonstração será apresentada na seção 4.4.

Seja v uma sequência de elementos de A, e uma sequência u, com $|u| \geq 1$, tal que $uv \in \tau$. Então, por abuso de notação, definimos

$$q(v) := \sum_{u:uv\in\tau} \frac{p(uv)}{p(v)} q(uv). \tag{4.3}$$

4.3 Estimação da árvore de contexto do processo

Dada uma amostra $(X_i, T_i)_{i=1}^n$, vamos apresentar um novo estimador para a árvore de contextos τ, baseado nas taxas de saltos $q(\cdot)$ do processo.

Inicialmente, estimamos as taxas de saltos do processo, utilizando o estimador $\hat{q}_n(\cdot)$, definido em (4.1). No teorema 4.2.2 vimos que, se ω é uma sequência pertencente a τ, então para toda sequência u de elementos de A, $\hat{q}_n(u\omega)$ converge para $\hat{q}_n(\omega)$ com probabilidade 1. Além disso, o corolário 4.2.1 prova que a estimação da taxa de salto de qualquer sequência arbitrariamente grande converge para para a taxa do seu contexto.

Baseado nesses resultados, para um $\delta > 0$, propomos um estimador $\hat{\tau}_n^\delta$ para a árvore de contextos τ, que funciona da seguinte forma:

1) Construímos uma árvore completa até uma certa altura d, e definimos $\hat{\tau}_n^\delta$ como o conjunto das folhas dessa árvore.

2) Em cada ramo (conjunto de folhas com mesmo sufixo próprio) de τ, estimamos as taxas de saltos $\hat{q}_n(\cdot)$ e comparamos com o valor da taxa estimada de seu sufixo próprio.

3) Se o menor valor do módulo da diferença entre o valor estimado nas folhas e no seu sufixo for menor que um valor δ fixado, então excluímos aquele conjunto de folhas de $\hat{\tau}_n^\delta$ e as substituímos por seu sufixo.

4) Se, em algum ramo, a diferença entre suas taxas estimadas em relação com as taxas estimadas de seu sufixo for maior que δ, então o ramo permanece em $\hat{\tau}_n^\delta$ e não pode mais ser excluído.

5) Repetimos o procedimento nos demais ramos até não ser mais possível realizar exclusões.

Formalmente, temos a seguinte definição.

Definição 4.3.1. Para u uma sequência tal que $N_n(au) > 0$, para todo $a \in A$, seja

$$\Delta_n(u) = \max_{a \in A} |\hat{q}_n(u) - \hat{q}_n(au)|. \tag{4.4}$$

Para d fixado, $0 < \delta$, definimos o estimador de τ como

$$\hat{\tau}_n^\delta = \left\{ \omega \in A_1^d \,:\, \Delta_n\left(suf(\omega)\right) > \delta \text{ e } \Delta_n\left(u\omega\right) \leq \delta \;\forall\; u \in A_1^{d-|\omega|} \text{ tal que } N_n(u\omega) > 0 \right\}, \tag{4.5}$$

onde $A_r^s = \bigcup_{i=r}^{s} A^i$ é o conjunto das sequências de tamanho pelo menos r e no máximo s.

Assim, obtemos um novo estimador para a árvore de contextos quando a amostra é gerada por um processo de saltos com memória de alcance variável.

Apresentamos o principal resultado dessa tese.

Teorema 4.3.1. Seja $(X_i,\, T_i)_{i=1}^n$ uma amostra gerada por um processo de saltos com memória de alcance variável com parâmetros $(\tau,\, p,\, q)$, assumindo valores num alfabeto A. Se $\hat{\tau}_n^\delta$ é o estimador definido em (4.5), então, com probabilidade 1, temos que

$$\hat{\tau}_n^\delta|_k = \tau|_k$$

para todo n suficientemente grande.

Demonstração: A demonstração será apresentada na seção 4.5. Vamos provar que $\hat{\tau}_n^\delta|_k$ converge para $\tau|_k$ com velocidade exponencial, e então concluir a consistência forte de $\hat{\tau}_n^\delta$.

A proposição a seguir mostra que $\hat{\tau}_n^\delta$ preserva a propriedade de árvore irredutível de τ.

Proposição 4.3.1. Dada uma amostra $(X_i, T_i)_{i=1}^n$, a árvore estimada $\hat{\tau}_n^\delta$ é uma árvore irredutível.

Demonstração: Vimos na definição 2.1.2 que uma árvore é irredutível se ela satisfaz as propriedades do sufixo e de irredutibilidade.

i) *Prova da propriedade do sufixo:* Seja $\omega \in \hat{\tau}_n^\delta$. Se existisse um $v \in \hat{\tau}_n^\delta$ tal que $v \prec \omega$, então teríamos

$$\Delta_n(uv) \geq \delta, \text{ para qualquer } u \text{ tal que } |u| \leq d - |\omega|.$$

Em especial, como $\omega \succ v$, teríamos $\Delta_n\left(suf(\omega)\right) \geq \delta$, o que é absurdo, pois $\omega \in \hat{\tau}_n^\delta$.

Logo, temos que vale a propriedade do sufixo.

ii) *Prova da irredutibilidade:* Vamos mostrar que, se $\omega \in \hat{\tau}_n^\delta$, então não podemos substituir ω por nenhum de seus sufixos próprios sem violar a propriedade do sufixo.

De fato, seja $a \in A$ tal que $a\, suf(\omega) \neq \omega$. Então

$$\Delta_n\left(suf(a\, suf(\omega))\right) = \Delta_n\left(suf(\omega)\right) > \delta.$$

Logo, existe u, com $|u| \leq d - |\omega|$, tal que $u\, suf(\omega) \neq \omega$ e $u\, suf(\omega) \in \hat{\tau}_n^\delta$.

Assim, substituindo ω por qualquer um de seus sufixos próprios viola a propriedade do sufixo.

Concluímos que $\hat{\tau}_n^\delta$ é uma árvore irredutível.

□

4.4 Taxa de Convergencia do estimador das taxas de saltos

O teorema a seguir apresenta um majorante para a probabilidade de erro entre o valor estimado das taxas $q(\cdot)$ e o valor real. Esse teorema é fundamental para a prova do teorema 4.3.1.

Teorema 4.4.1. Seja $(X_i, T_i)_{i=-(k-1)}^n$ uma amostra de um processo de saltos com memória de alcance variável $(Y_t)_{t\in\mathbb{R}}$ compatível com $(\tau,\, p,\, q)$ satisfazendo as hipóteses 1, 2 e 3 e ω tal que $|\omega| = k$. Então, para qualquer $\epsilon > 0$,

$$\mathbb{P}\left(|\hat{q}_n(\omega) - q(\omega)| > \epsilon\right) \leq 4e^{-n\cdot h(p(\omega), q(\omega), \epsilon)}, \tag{4.6}$$

onde h é uma função real positiva definida por

$$h\left(p(\omega), q(\omega), \epsilon\right) = \frac{p(\omega)}{2} \cdot \min\left\{\frac{\epsilon}{q(\omega)} - \log\left(1 + \frac{\epsilon}{q(\omega)}\right),\, \frac{c}{2}\right\},$$

onde $c = \frac{\alpha_0}{8e(\alpha+\alpha_0)}$.

Demonstração: Dada uma amostra $(X_i, T_i)_{-(k-1)}^n$ e uma sequência $\omega = x_{-k}^{-1}$ tal que $N_n(\omega) > 0$. Sejam

$$
\begin{aligned}
&n_1 = \inf\left\{j \geq 1 \,:\, X_{j-k+1}^{j} = \omega\right\} \\
&n_2 = \inf\left\{j > n_1 \,:\, X_{j-k+1}^{j} = \omega\right\} \\
&\cdots \\
&n_{N_n(\omega)} = \inf\left\{j > n_{N_n(\omega)-1} \,:\, X_{j-k+1}^{j} = \omega\right\}.
\end{aligned}
$$

Para $i = 1, 2, \cdots, N_n(\omega)$, definimos $Z_i = (T_{n_i+1} - T_{n_i})$. Assim,

$$
\sum_{i=k+1}^{n} (T_{i+1} - T_i)\, \mathbb{I}_{\{X_{i-k}^{i-1} = x_{-k}^{-1}\}} = \sum_{i=1}^{N_n(\omega)} Z_i.
$$

Então, temos que

$$
\begin{aligned}
\mathbb{P}\left(|\hat{q}_n(\omega) - q(\omega)| > \epsilon\right) &= \mathbb{P}\left(\left|\frac{1}{N_n(\omega)} \sum_{i=1}^{N_n(\omega)} Z_i - q(\omega)\right| > \epsilon\right) \\
&= \mathbb{P}\left(\left|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega) q(\omega)\right| > N_n(\omega)\epsilon\right).
\end{aligned}
$$

Para algum $m > 0$, vamos condicionar o termo da direita em dois eventos disjuntos: $\{N_n(\omega) \geq m\}$ e $\{N_n(\omega) < m\}$.

Dessa forma , temos que

$$
\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega) q(\omega)\right| > N_n(\omega)\epsilon\right) \\
= \;& \mathbb{P}\left(\left|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega) q(\omega)\right| > N_n(\omega)\epsilon\,,\; N_n(\omega) \geq m\right) \\
+ \;& \mathbb{P}\left(\left|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega) q(\omega)\right| > N_n(\omega)\epsilon\,,\; N_n(\omega) < m\right). \qquad (4.7)
\end{aligned}
$$

Como $\left\{\left|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega) q(\omega)\right| > N_n(\omega)\epsilon\,,\; N_n(\omega) < m\right\} \subset \left\{N_n(\omega) < m\right\}$, substituindo o segundo termo em (4.7), obtemos que

$$
\begin{aligned}
\mathbb{P}\left(|\hat{q}_n(\omega) - q(\omega)| > \epsilon\right) &\leq \mathbb{P}\left(\left|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega) q(\omega)\right| > N_n(\omega)\epsilon\,,\; N_n(\omega) \geq m\right) \\
&+ \mathbb{P}\left(N_n(\omega) < m\right) \qquad (4.8)
\end{aligned}
$$

Vamos considerar $m = n[p(\omega) - \beta]$, para algum $0 < \beta < p(\omega)$. Aplicando o teorema 3.1 de Galves e Leonardi (2008), temos que uma cota superior para o segundo termo de (4.8) é dada por

$$\begin{aligned}\mathbb{P}\big(N_n(\omega) < n(p(\omega)-\beta)\big) &= \mathbb{P}\big(np(\omega) - N_n(\omega) > n\beta\big)\\ &\leq \mathbb{P}\big(|N_n(\omega) - np(\omega)| > n\beta\big) \leq e^{\frac{1}{e}} \exp\left\{\frac{-n\beta^2 c}{p(\omega)}\right\}, \quad (4.9)\end{aligned}$$

onde $c = \frac{\alpha_0}{8e(\alpha+\alpha_0)}$.

Para encontrarmos um limitante superior para o primeiro termo de (4.8), nós o reescrevemos da seguinte forma.

$$\begin{aligned}&\mathbb{P}\Bigg(\Bigg|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega)\Bigg| > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Bigg)\\ &= \mathbb{P}\Bigg(\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega) > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Bigg)\\ &+\mathbb{P}\Bigg(N_n(\omega)q(\omega) - \sum_{i=1}^{N_n(\omega)} Z_i > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Bigg). \quad (4.10)\end{aligned}$$

Para o primeiro termo a direita da igualdade em (4.10), temos que

$$\begin{aligned}&\mathbb{P}\Bigg(\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega) > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Bigg)\\ &= \mathbb{P}\Bigg(\sum_{i=1}^{N_n(\omega)} Z_i > N_n(\omega)[q(\omega)+\epsilon]\,,\, N_n(\omega) \geq m\Bigg)\\ &= \sum_{j=m}^{n} \mathbb{P}\Big(\sum_{i=1}^{j} Z_i > j[q(\omega)+\epsilon]\,,\, N_n(\omega) = j\Big). \quad (4.11)\end{aligned}$$

Como, para cada i, Z_i e $N_n(\omega)$ são independentes, temos que

$$\begin{aligned}&\sum_{j=m}^{n} \mathbb{P}\Big(\sum_{i=1}^{j} Z_i > j[q(\omega)+\epsilon]\,,\, N_n(\omega) = j\Big)\\ &= \sum_{j=m}^{n} \mathbb{P}\Big(\sum_{i=1}^{j} Z_i > j[q(\omega)+\epsilon]\Big) \cdot \mathbb{P}\Big(N_n(\omega) = j\Big) \quad (4.12)\end{aligned}$$

Para cada $j \in \{m, m+1, \cdots, n\}$, temos que

$$\mathbb{P}\Big(\sum_{i=1}^{j} Z_i > j[q(\omega)+\epsilon]\Big) \leq \sup_{s\in[m,n]} \mathbb{P}\Big(\sum_{i=1}^{s} Z_i > s[q(\omega)+\epsilon]\Big). \quad (4.13)$$

Ainda, como $\{N_n(\omega) = j\,,\, j = 1, \cdots, n\}$ são eventos disjuntos, temos que

$$\sum_{j=m}^{n} \mathbb{P}\Big(N_n(\omega) = j\Big) \leq \sum_{j=1}^{n} \mathbb{P}\Big(N_n(\omega) = j\Big) = 1 \tag{4.14}$$

Assim, com (4.13) e (4.14), obtemos uma cota superior para (4.12).

$$\begin{aligned}
&\sum_{j=m}^{n} \mathbb{P}\Big(\sum_{i=1}^{j} Z_i > j[q(\omega)+\epsilon]\Big) \cdot \mathbb{P}\Big(N_n(\omega) = j\Big) \\
&\leq \sup_{s\in[m,n]} \mathbb{P}\Big(\sum_{i=1}^{s} Z_i > s[q(\omega)+\epsilon]\Big) \sum_{j=m}^{n} \mathbb{P}\Big(N_n(\omega) = j\Big) \\
&\leq \sup_{s\in[m,n]} \mathbb{P}\Big(\sum_{i=1}^{s} Z_i > s[q(\omega)+\epsilon]\Big). \qquad (4.15)
\end{aligned}$$

Sejam s uma constante inteira, $1 \leq s \leq n$, $x = q(\omega)+\epsilon$ e θ um número real, $0 < \theta < \frac{1}{q(\omega)}$. Pela desigualdade de Markov, temos que

$$\begin{aligned}
\mathbb{P}\Big(\sum_{i=1}^{s} Z_i > sx\Big) &= \mathbb{P}\left(e^{\theta \sum_{i=1}^{s} Z_i} > e^{\theta s x}\right) \leq \left[\mathbb{E}(e^{\theta Z})\right]^s e^{-\theta s x} \\
&= \exp\left\{-s\left[\theta x + \log(1-\theta q(\omega))\right]\right\} \qquad (4.16)
\end{aligned}$$

A desigualdade (4.16) vale para todo $0 < \theta < \frac{1}{q(\omega)}$ Vamos calcular o $\tilde{\theta}$ que maximiza $g(\theta) = \theta x + \log\left(1 - \theta q(\omega)\right)$.

Derivando $g(\theta)$, obtemos

$$g'(\theta) = x - \frac{q(\omega)}{1-\theta q(\omega)}.$$

Calculando o valor de $\tilde{\theta}$ que iguala $g'(\theta)$ a zero, obtemos

$$\begin{aligned}
g'(\theta) = 0 \quad \Leftrightarrow \quad & x - \frac{q(\omega)}{1-\theta q(\omega)} = 0 \Leftrightarrow x = \frac{q(\omega)}{1-\theta q(\omega)} \\
& x - \theta\, x\, q(\omega) = q(\omega) \Leftrightarrow \theta = \frac{x - q(\omega)}{x\, q(\omega)}.
\end{aligned}$$

Logo, obtemos que

$$g'(\tilde{\theta}) = 0 \Leftrightarrow \tilde{\theta} = \frac{1}{q(\omega)} - \frac{1}{x} \tag{4.17}$$

Nota que, como $x = q(\omega) + \epsilon$, então $0 < \tilde{\theta} < \frac{1}{q(\omega)}$.

Para verificar se o $\tilde{\theta}$ definido em (4.17) é um ponto de máximo de $g(\theta)$, calculamos sua segunda derivada, obtendo

$$g''(\theta) = \frac{-q^2(\omega)}{[1-\theta\, q(\omega)]^2}.$$

Logo, para todo $0 < \theta < \frac{1}{\theta}$, $g''(\theta) < 0$, o que implica que $\tilde{\theta}$ é ponto de máximo de $g(\theta)$.

Assim, substituindo θ por $\tilde{\theta}$ em (4.16), temos que, para para um número inteiro determinístico s,

$$\mathbb{P}\left(\sum_{i=1}^{s} Z_i > s\,[q(\omega)+\epsilon]\right) \leq \exp\left\{-s\left[\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)+\epsilon}{q(\omega)}\right)\right]\right\} \tag{4.18}$$

Para que a desigualdade (4.18) não seja trivial, é necessário que

$$\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)+\epsilon}{q(\omega)}\right) > 0. \tag{4.19}$$

Para verificar a desigualdade acima, utilizamos o lema a seguir.

Lema 4.4.1. Para todo $x \in (0, +\infty) \setminus \{1\}$, temos que

$$x - 1 - \log(x) > 0$$

Demonstração: Temos que

$$\begin{aligned} x - 1 - \log(x) > 0 \quad &\Leftrightarrow \quad x - 1 > \log(x) \Leftrightarrow e^{x-1} > x \\ &\Leftrightarrow \quad e^x > e \cdot x, \end{aligned}$$

o que ocorre para todo $x \in (0, +\infty) \setminus \{1\}$.

De fato, sejam $g_1(x) = e^x$ e $g_2(x) = e \cdot x$. Então,

$$g_1(x) = g_2(x) \Leftrightarrow x = 1.$$

Como $g_1(0) = 1 > 0 = g_2(0)$, então

$$e^x > e \cdot x, \text{ para todo } x < 1.$$

Para $x > 1$, basta reparar que $g_1(1) = g_2(1)$ e

$$g_1'(x) = e^x > e = g_2'(x), \text{ para todo } x > 1.$$

Logo, $e^x > e \cdot x$ também para todo $x > 1$, e está provado o lema.

Como $q(\omega)$, $\epsilon > 0$, temos que $\frac{q(\omega)+\epsilon}{q(\omega)} > 1$. Aplicando o lema 4.4.1, temos que a desigualdade (4.19) é válida, o que implica que o limitante superior (4.18) não é trivial.

Logo, por (4.18),

$$\begin{aligned} \sup_{s\in[m,n]} \mathbb{P}\left(\sum_{i=1}^{s} Z_i > s(\omega)\,[q(\omega)+\epsilon]\right) \quad &\leq \quad \sup_{s\in[m,n]} \exp\left\{-s\left[\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)+\epsilon}{q(\omega)}\right)\right]\right\} \\ &\leq \quad \exp\left\{-m\left[\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)+\epsilon}{q(\omega)}\right)\right]\right\}. \end{aligned}$$

Como $m = n\big(p(\omega) - \beta\big)$, $0 < \beta < p(\omega)$, e pela desigualdade acima, concluímos que

$$\begin{aligned}
&\mathbb{P}\Bigg(\Bigg\{\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega) > N_n(\omega)\epsilon\Bigg\} \cap \Big\{N_n(\omega) \geq m\Big\}\Bigg) \\
&\leq \exp\Bigg\{-n\big(p(\omega) - \epsilon\big)\Bigg[\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\Bigg(\frac{q(\omega)+\epsilon}{q(\omega)}\Bigg)\Bigg]\Bigg\}, \qquad (4.20)
\end{aligned}$$

e temos um limitante superior para o primeiro termo à direita da desigualdade em (4.8).

Para o segundo termo a direita da igualdade em (4.10), temos que

$$\begin{aligned}
&\mathbb{P}\Bigg(N_n(\omega)q(\omega) - \sum_{i=1}^{N_n(\omega)} Z_i > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Bigg\}\Bigg) \\
&= \mathbb{P}\Bigg(-\sum_{i=1}^{N_n(\omega)} Z_i > -N_n(\omega)[q(\omega) - \epsilon]\,,\, N_n(\omega) \geq m\Bigg) \\
&= \sum_{j=m}^{n} \mathbb{P}\Big(-\sum_{i=1}^{j} Z_i > -j[q(\omega) - \epsilon]\,,\, N_n(\omega) = j\Big).
\end{aligned}$$

Como no caso anterior, pela independência de Z_i e $N_n(\omega)$, para todo i, temos que

$$\begin{aligned}
&\sum_{j=m}^{n} \mathbb{P}\Big(-\sum_{i=1}^{j} Z_i > -j[q(\omega) - \epsilon]\,,\, N_n(\omega) = j\Big) \\
&= \sum_{j=m}^{n} \mathbb{P}\Big(-\sum_{i=1}^{j} Z_i > -j[q(\omega) - \epsilon]\Big) \cdot \mathbb{P}\Big(N_n(\omega) = j\Big).
\end{aligned}$$

Aplicando raciocínio análogo ao utilizado em (4.13) e (4.14), obtemos que

$$\begin{aligned}
&\sum_{j=m}^{n} \mathbb{P}\Big(-\sum_{i=1}^{j} Z_i > -j[q(\omega) - \epsilon]\Big) \cdot \mathbb{P}\Big(N_n(\omega) = j\Big) \\
&\leq \sup_{s\in[m,n]} \mathbb{P}\Big(-\sum_{i=1}^{s} Z_i > -s[q(\omega) - \epsilon]\Big) \sum_{j=m}^{n} \mathbb{P}\Big(N_n(\omega) = j\Big) \\
&\leq \sup_{s\in[m,n]} \mathbb{P}\Big(-\sum_{i=1}^{s} Z_i > -s[q(\omega) - \epsilon]\Big) \qquad (4.21)
\end{aligned}$$

Para $\epsilon > 0$, s inteiro, $x = q(\omega) - \epsilon$ e $0 < \theta$, aplicando a desigualdade de Markov, temos que

$$\begin{aligned}\mathbb{P}\Big(-\sum_{i=1}^{s} Z_i > -sx\Big) &= \mathbb{P}\left(e^{-\theta\sum_{i=1}^{s} Z_i} > e^{-\theta sx}\right) \leq \left[\mathbb{E}(e^{-\theta Z})\right]^s e^{\theta sx} \\ &= \exp\left\{-s\left[-\theta x + \log\left(1+\theta q(\omega)\right)\right]\right\} \qquad (4.22)\end{aligned}$$

Novamente, vamos calcular o $\tilde{\theta}$ que maximiza $g(\theta) = -\theta x + \log\left(1+\theta q(\omega)\right)$. Derivando $g(\theta)$, obtemos que

$$g`(\theta) = -x + \frac{q(\omega)}{1+\theta\, q(\omega)}. \qquad (4.23)$$

Igualando $g`(\theta)$ a zero, temos

$$\begin{aligned} g`(\theta) = 0 \quad &\Leftrightarrow \quad -x + \frac{q(\omega)}{1+\theta\, q(\omega)} = 0 \Leftrightarrow x = \frac{q(\omega)}{1+\theta\, q(\omega)} \\ &\Leftrightarrow \quad x + x\,\theta\, q(\omega) = q(\omega) \Leftrightarrow \tilde{\theta} = \frac{1}{x} - \frac{1}{q(\omega)}. \end{aligned}$$

Como $g''(\theta) = -\left(\dfrac{q(\omega)}{1+\theta\, q(\omega)}\right)^2 < 0$, então $\tilde{\theta} = \frac{1}{q(\omega)-\epsilon} - \frac{1}{q(\omega)}$ é ponto de máximo de $g(\theta)$. Substituindo θ por $\tilde{\theta}$ em (4.22), temos que, para um inteiro determinístico s,

$$\mathbb{P}\Big(-\sum_{i=1}^{s} Z_i > -s\left[q(\omega)-\epsilon\right]\Big) \leq \exp\left\{-s\left[\frac{q(\omega)-\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)-\epsilon}{q(\omega)}\right)\right]\right\}. \qquad (4.24)$$

Pelo lema 4.4.1, temos que

$$\frac{q(\omega)-\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)-\epsilon}{q(\omega)}\right) > 0,$$

e a desigualdade (4.24) não é trivial.

Como $m = n(p(\omega)-\beta)$, por (4.21), temos um limitante superior para o segundo termo a direita da desigualdade (4.8), dado por

$$\begin{aligned}&\mathbb{P}\Big(N_n(\omega)q(\omega) - \sum_{i=1}^{N_n(\omega)} Z_i > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Big) \\ &\leq \exp\left\{-n(p(\omega)-\beta)\left[\frac{q(\omega)-\epsilon}{q(\omega)} - 1 - \log\left(\frac{q(\omega)-\epsilon}{q(\omega)}\right)\right]\right\}. \qquad (4.25)\end{aligned}$$

Aplicando os limites superiores (4.20) e (4.25) em (4.10), obtemos que

$$
\begin{aligned}
&\mathbb{P}\Big(\Big|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega)\Big| > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Big) \\
&\leq \exp\Big\{-n(p(\omega)-\beta)\Big[\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\Big(\frac{q(\omega)+\epsilon}{q(\omega)}\Big)\Big]\Big\} \\
&+\exp\Big\{-n(p(\omega)-\beta)\Big[\frac{q(\omega)-\epsilon}{q(\omega)} - 1 - \log\Big(\frac{q(\omega)-\epsilon}{q(\omega)}\Big)\Big]\Big\}. \qquad (4.26)
\end{aligned}
$$

Para concluir o teorema, seja o lema a seguir.

Lema 4.4.2. Para $x > 0$, seja a função $f : (0, +\infty) \to \mathbb{R}$, dada por

$$f(x) = x - 1 - \log(x).$$

então, para todo $\xi \in (0,1)$, temos que

$$f(1-\xi) \geq f(1+\xi)$$

Demonstração: para $\xi \in (0,1)$, temos que

$$
\begin{aligned}
f(1-\xi) \geq f(1+\xi) \quad &\Leftrightarrow \quad (1-\xi) - 1 - \log(1-\xi) \geq (1+\xi) - 1 - \log(1+\xi) \\
&\Leftrightarrow \quad \log\Big(\frac{1+\xi}{1-\xi}\Big) - 2 \geq 0. \qquad (4.27)
\end{aligned}
$$

Seja $g(\xi) = \log\Big(\frac{1+\xi}{1-\xi}\Big) - 2$.
No ponto $\xi = 0$, temos $g(0) = 0$.
Para $\xi \in (0,1)$,

$$g`(\xi) = \Big(\frac{1}{1+\xi} + \frac{1}{1-\xi}\Big) - 2 = \frac{2}{(1+\xi)(1-\xi)} - 2.$$

Nota que, para $\xi \in (0,1)$,

$$0 < (1+\xi)(1-\xi) = 1 - \xi^2 < 1,$$

o que implica que

$$g`(\xi) > 0, \text{ para } \xi \in (0,1),$$

ou seja, $g(0) = 0$ e $g(\xi)$ crescente em $(0,1)$.
Logo, (4.27) é verdadeiro, e está provado o lema.

Aplicando o lema 4.4.2, temos que

$$\frac{q(\omega)-\epsilon}{q(\omega)} - 1 - \log\Big(\frac{q(\omega)-\epsilon}{q(\omega)}\Big) \geq \frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\Big(\frac{q(\omega)+\epsilon}{q(\omega)}\Big), \qquad (4.28)$$

o que implica que, de (4.26) e (4.28), obtemos um limitante superior para o primeiro termo de (4.8), dado por

$$
\begin{aligned}
&\mathbb{P}\Big(\Big|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega)\Big| > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Big) \\
&\leq 2\exp\Big\{-n(p(\omega)-\beta)\Big[\frac{q(\omega)+\epsilon}{q(\omega)} - 1 - \log\Big(\frac{q(\omega)+\epsilon}{q(\omega)}\Big)\Big]\Big\}. \\
&= 2\exp\Big\{-n(p(\omega)-\beta)\Big[\frac{\epsilon}{q(\omega)} - \log\Big(1+\frac{\epsilon}{q(\omega)}\Big)\Big]\Big\}. \qquad (4.29)
\end{aligned}
$$

Enfim, aplicando as desigualdades (4.9) e (4.29), concluímos que

$$
\begin{aligned}
\mathbb{P}\left(|\hat{q}_n(\omega) - q(\omega)| > \epsilon\right) &\leq \mathbb{P}\Big(\Big|\sum_{i=1}^{N_n(\omega)} Z_i - N_n(\omega)q(\omega)\Big| > N_n(\omega)\epsilon\,,\, N_n(\omega) \geq m\Big) \\
&+ \mathbb{P}\left(N_n(\omega) < m\right) \\
&\leq 2\exp\Big\{-n(p(\omega)-\beta)\left[\frac{\epsilon}{q(\omega)} - \log\Big(1+\frac{\epsilon}{q(\omega)}\Big)\right]\Big\} \\
&+ e^{\frac{1}{e}}\exp\Big\{\frac{-n\beta^2 c}{p(\omega)}\Big\} \qquad (4.30)
\end{aligned}
$$

Como a desigualdade vale para todo $0 < \beta < p(\omega)$, tomamos $\beta = p(\omega)/2$, e obtemos que

$$
\mathbb{P}\left(|\hat{q}_n(\omega) - q(\omega)| > \epsilon\right) \leq 4 \cdot e^{-n\cdot h(p(\omega),q(\omega),\epsilon)}, \qquad (4.31)
$$

onde $h\left(p(\omega), q(\omega), \epsilon\right) = \frac{p(\omega)}{2} \cdot \min\Big\{\frac{\epsilon}{q(\omega)} - \log\Big(1+\frac{\epsilon}{q(\omega)}\Big)\,,\, \frac{c}{2}\Big\}$.

□

O corolário a seguir generaliza o resultado do teorema anterior, no caso em que consideramos uma sequência que possui um contexto como sufixo.

Corolário 4.4.1. Sob as mesmas condições do teorema 4.4.1, sejam $\omega \in \tau$, u uma sequência de elementos de A, tal que $N_n(u\omega) > 0$, e $k = |u\omega|$. Então, para todo $\epsilon > 0$,

$$
\mathbb{P}\big(|\hat{q}_n(u\omega) - q(\omega)| > \epsilon\big) \leq 4e^{-n\cdot h(p(u\omega),q(\omega),\epsilon)}.
$$

Demonstração: Na demonstração do teorema 4.4.1, substituindo $N_n(\omega)$ por $N_n(u\omega)$ e $p(\omega)$ por $p(u\omega)$ em (4.8), temos que

$$
\begin{aligned}
\mathbb{P}\left(|\hat{q}_n(u\omega) - q(\omega)| > \epsilon\right) &\leq \mathbb{P}\Big(\Big|\sum_{i=1}^{N_n(u\omega)} Z_i - N_n(u\omega)q(\omega)\Big| > N_n(u\omega)\epsilon\,,\, N_n(u\omega) \geq m\Big) \\
&+ \mathbb{P}\left(N_n(u\omega) < m\right).
\end{aligned}
$$

Considerando $m = n\big(\frac{p(u\omega)}{2}\big)$, de forma análoga obtemos que

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{N_n(u\omega)} Z_i - N_n(u\omega)q(\omega)\Big| > N_n(u\omega)\epsilon\,,\; N_n(u\omega) \geq m\Big)$$
$$\leq 2\exp\left\{-n\frac{p(u\omega)}{2}\left[\frac{\epsilon}{q(\omega)} - \log\left(1+\frac{\epsilon}{q(\omega)}\right)\right]\right\},$$

e

$$\mathbb{P}\left(N_n(u\omega) < m\right) \leq e^{\frac{1}{e}}\exp\left\{\frac{-np(u\omega)c}{2}\right\},$$

e concluímos que

$$\mathbb{P}\big(|\hat{q}_n(u\omega) - q(\omega)| > \epsilon\big) \leq 4e^{-n\cdot h(p(u\omega),q(\omega),\epsilon)}.$$

□

Com o teorema 4.4.1 possuímos ferramentas suficientes para demonstrar o teorema 4.2.1, que enuncia consistência forte do estimador $\hat{q}_n(\cdot)$, e do corolário 4.2.1.

Demonstração do teorema 4.2.1: Pelo teorema 4.4.1, temos que, para $\omega \in \tau$, existem $a, b > 0$ tais que

$$\mathbb{P}(|\hat{q}_n(\omega) - q(\omega)| > \epsilon) \leq b \cdot e^{-n\,a}.$$

Como, para todo $a > 0$,

$$\sum_{n=0}^{+\infty} e^{-na} < +\infty,$$

então, aplicando o Lema de Borel-Cantelli, temos que

$$\mathbb{P}\left(|\hat{q}_n(\omega) - q(\omega)| > \epsilon \text{ i.v. }\right) = 0,$$

ou seja, $\hat{q}_n(\omega) \to^{q.c} q(\omega)$, para todo $\omega \in \tau$.

□

Demonstração do corolário 4.2.1: Basta aplicar o resultado do corolário 4.4.1 de forma análoga à demonstração do teorema acima, e temos o resultado desejado.

□

O teorema a seguir mostra que, para uma árvore finita τ, a taxa de salto estimada de um sufixo próprio de um elemento de τ converge, com velocidade exponencial, para a média ponderada das taxas dos contextos que contém esse sufixo. Esse resultado será essencial para a demonstração do teorema 4.2.3.

Teorema 4.4.2. Dada uma amostra $(X_i, T_i)_{i=1}^n$, supomos que a árvore de contextos τ é finita, com $|\tau| = d$. Seja v uma sequência de elementos de A, tal que v é sufixo próprio de algum elemento de τ, isto é, existe uma outra sequência finita u, $|u| \geq 1$, tal que $uv \in \tau$.

Para todo $\epsilon > 0$, sejam

$$a = \frac{\epsilon}{2A^{d-|v|} \cdot \sup\limits_{u\,:\,uv\in\tau} \{q(uv)\}},$$

e, para algum inteiro $K > \frac{\epsilon}{p(v)}$,

$$\eta(a, p(v)) = \min\left\{\frac{a}{K}\,,\, p(v)\Big(p(v) - \frac{\epsilon}{k}\Big)\epsilon\right\}.$$

Então, para algum u^*, tal que $u^*v \in \tau$, temos que

$$\mathbb{P}\left(\left|\hat{q}_n(v) - \sum_{u\,:\,uv\in\tau} \frac{p(uv)}{p(v)} \cdot q(uv)\right| > \epsilon\right) \leq 10 \cdot e^{-n\cdot\min\left\{h\left(p(u^*v), q(u^*v), \frac{\epsilon}{2A^{d-|v|}}\right),\, \frac{c\cdot\eta^2(a,p(v))}{p(v)}\right\}}.$$

Demonstração: Primeiramente, vamos reescrever $\hat{q}_n(v)$ como

$$\begin{aligned}
\hat{q}_n(v) &= \frac{1}{N_n(v)} \sum_{i=|v|}^{n-1} (T_{i+1} - T_i)\, \mathbb{I}_{\{X_{i-|v|}^{i-1} = v\}} \\
&= \frac{1}{N_n(v)} \sum_{u\,:\,uv\in\tau} \sum_{i=|uv|}^{n-1} (T_{i+1} - T_i)\, \mathbb{I}_{\{X_{i-|uv|}^{i-1} = uv\}}.
\end{aligned}$$

Multiplicando e dividindo por $N_n(uv)$ e n no interior do primeiro somatório, temos que

$$\begin{aligned}
\hat{q}_n(v) &= \sum_{u\,:\,uv\in\tau} \frac{N_n(uv)}{n} \cdot \frac{n}{N_n(n)} \cdot \frac{1}{N_n(uv)} \sum_{i=|uv|}^{n-1} (T_{i+1} - T_i)\, \mathbb{I}_{\{X_{i-|uv|}^{i-1} = uv\}} \\
&= \sum_{u\,:\,uv\in\tau} \frac{\hat{p}_n(uv)}{\hat{p}_n(v)} \cdot \hat{q}_n(uv).
\end{aligned}$$

Então, temos que

$$\begin{aligned}
& \mathbb{P}\left(\left|\hat{q}_n(v) - \sum_{u\,:\,uv\in\tau} \frac{p(uv)}{p(v)} q(uv)\right| > \epsilon\right) \\
= \; & \mathbb{P}\left(\left|\sum_{u\,:\,uv\in\tau} \frac{\hat{p}_n(uv)}{\hat{p}_n(v)} \hat{q}_n(uv) - \sum_{u\,:\,uv\in\tau} \frac{p(uv)}{p(v)} q(uv)\right| > \epsilon\right) \\
\leq \; & \mathbb{P}\left(\sum_{u\,:\,uv\in\tau} \left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} \hat{q}_n(uv) - \frac{p(uv)}{p(v)} q(uv)\right| > \epsilon\right). \qquad (4.32)
\end{aligned}$$

Somando e subtraindo $\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} q(uv)$ no interior do módulo em (4.32), temos que

$$
\begin{aligned}
& \mathbb{P}\Big(\sum_{u\,:\,uv\in\tau} \Big| \frac{\hat{p}_n(uv)}{\hat{p}_n(v)}\hat{q}_n(uv) - \frac{p(uv)}{p(v)}q(uv) \Big| > \epsilon \Big) \\
= \; & \mathbb{P}\Big(\sum_{u\,:\,uv\in\tau} \Big| \frac{\hat{p}_n(uv)}{\hat{p}_n(v)}\hat{q}_n(uv) - \frac{\hat{p}_n(uv)}{\hat{p}_n(v)}q(uv) + \frac{\hat{p}_n(uv)}{\hat{p}_n(v)}q(uv) - \frac{p(uv)}{p(v)}q(uv) \Big| > \epsilon \Big) \\
= \; & \mathbb{P}\Big(\sum_{u\,:\,uv\in\tau} \Big| \frac{\hat{p}_n(uv)}{\hat{p}_n(v)}\big(\hat{q}_n(uv) - q(uv)\big) + q(uv)\Big(\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\Big) \Big| > \epsilon \Big). \qquad (4.33)
\end{aligned}
$$

Como $0 \leq \frac{\hat{p}_n(uv)}{\hat{p}_n(v)} \leq 1$ e $|\tau| = d$, de (4.33) obtemos que

$$
\begin{aligned}
& \mathbb{P}\Big(\sum_{u\,:\,uv\in\tau} \Big| \frac{\hat{p}_n(uv)}{\hat{p}_n(v)}\big(\hat{q}_n(uv) - q(uv)\big) + q(uv)\Big(\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\Big) \Big| > \epsilon \Big) \\
\leq \; & \mathbb{P}\Big(\sum_{u\,:\,uv\in\tau} \Big| \hat{q}_n(uv) - q(uv) \Big| > \frac{\epsilon}{2} \Big) + \mathbb{P}\Big(\sum_{u\,:\,uv\in\tau} q(uv) \Big| \frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)} \Big| > \frac{\epsilon}{2} \Big) \\
\leq \; & \mathbb{P}\Big(A^{d-|v|} \sup_{u\,:\,uv\in\tau} \Big\{ \Big| \hat{q}_n(uv) - q(uv) \Big| \Big\} > \frac{\epsilon}{2} \Big) \\
+ \; & \mathbb{P}\Big(A^{d-|v|} \sup_{u\,:\,uv\in\tau} \{q(uv)\} \cdot \sup_{u\,:\,uv\in\tau} \Big\{ \Big| \frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)} \Big| \Big\} > \frac{\epsilon}{2} \Big). \qquad (4.34)
\end{aligned}
$$

Pelo teorema 4.4.1, temos que um limitante superior para o primeiro termo de (4.34) é dado por

$$
\mathbb{P}\Big(\sup_{u\,:\,uv\in\tau} \Big\{ \Big| \hat{q}_n(uv) - q(uv) \Big| \Big\} > \frac{\epsilon}{2A^{d-|v|}} \Big) \leq 4e^{-n\cdot h(p(u'v), q(u'v), \frac{\epsilon}{2A^{d-|v|}})}, \qquad (4.35)
$$

para algum u', tal que $u'v \in \tau$.

Para o segundo termo de (4.34) vamos precisar dos lemas a seguir.

Lema 4.4.3. Sob as mesmas hipóteses do teorema 4.4.1, seja v uma sequência de elementos de A tal que $p(v) > 0$. Então, para todo $\epsilon > 0$, todo inteiro K tal que $K > \frac{\epsilon}{p(v)}$ e um n suficientemente grande, temos que

$$
\mathbb{P}\Big(\Big| \frac{1}{p(v)} - \frac{1}{\hat{p}_n(v)} \Big| > \epsilon \Big) \leq 2e^{\frac{1}{e}} e^{-n\cdot c\cdot \eta^2(\epsilon, p(v))},
$$

onde

$$
\eta\left(\epsilon, p(v)\right) = \min\Big\{ \frac{\epsilon}{K}\,,\, p(v)\Big(p(b) - \frac{\epsilon}{K}\Big)\epsilon \Big\}.
$$

Demonstração: Dado um $\epsilon > 0$, temos que

$$
\mathbb{P}\Big(\Big| \frac{1}{p(v)} - \frac{1}{\hat{p}_n(v)} \Big| > \epsilon \Big) = \mathbb{P}\big(|p(v) - \hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon \big). \qquad (4.36)
$$

Para um inteiro K tal que $K > \frac{\epsilon}{p(v)}$, vamos particionar (4.36) nos eventos

$$
\Big\{ |p(v) - \hat{p}_n(v)| \geq \frac{\epsilon}{K} \Big\} \text{ e } \Big\{ |p(v) - \hat{p}_n(v)| < \frac{\epsilon}{K} \Big\}
$$

. Então,

$$\begin{aligned}
& \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\right) \\
= \; & \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\,,\, |p(v)-\hat{p}_n(v)| \geq \frac{\epsilon}{K}\right) \\
+ \; & \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\,,\, |p(v)-\hat{p}_n(v)| < \frac{\epsilon}{K}\right). \qquad (4.37)
\end{aligned}$$

No primeiro termo de (4.37) temos que

$$\mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\,,\, |p(v)-\hat{p}_n(v)| \geq \frac{\epsilon}{K}\right) \leq \mathbb{P}\left(|p(v)-\hat{p}_n(v)| \geq \frac{\epsilon}{K}\right). \quad (4.38)$$

Ainda, observe que

$$|p(v)-\hat{p}_n(v)| < \frac{\epsilon}{K} \Rightarrow \hat{p}_n(v) > p(v) - \frac{\epsilon}{K},$$

ou seja

$$p(v)\hat{p}_n(v)\epsilon > p(v)\left(p(v) - \frac{\epsilon}{K}\right)\epsilon.$$

Logo, do segundo termo de (4.37), temos que

$$\begin{aligned}
& \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\,,\, |p(v)-\hat{p}_n(v)| < \frac{\epsilon}{K}\right) \\
\leq \; & \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\right) \\
\leq \; & \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\Big(p(v) - \frac{\epsilon}{K}\Big)\epsilon\right). \qquad (4.39)
\end{aligned}$$

Substituindo as desigualdades obtidas em (4.38) e (4.39) em (4.37), obtemos que

$$\begin{aligned}
\mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\hat{p}_n(v)\epsilon\right) \; & \leq \; \mathbb{P}\left(|p(v)-\hat{p}_n(v)| \geq \frac{\epsilon}{K}\right) \\
& + \; \mathbb{P}\left(|p(v)-\hat{p}_n(v)| > p(v)\Big(p(v) - \frac{\epsilon}{K}\Big)\epsilon\right). \quad (4.40)
\end{aligned}$$

Aplicando o teorema 3.1 de Galves e Leonardi (2008) em (4.40), concluímos que

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{p(v)} - \frac{1}{\hat{p}_n(v)}\right| > \epsilon\right) \; & \leq \; e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\left(\frac{\epsilon}{K}\right)^2}{p(v)}\right\} \\
& + \; e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\left[p(v)\left(p(v) - \frac{\epsilon}{K}\right)\epsilon\right]^2}{p(v)}\right\} \\
& \leq \; 2e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\eta^2\left(\epsilon, p(v)\right)}{p(v)}\right\},
\end{aligned}$$

onde $\eta(\epsilon, p(v)) = \min\left\{\frac{\epsilon}{K}\,,\, p(v)\left(p(v) - \frac{\epsilon}{K}\right)\epsilon\right\}$.

Lema 4.4.4. Sob as mesmas hipóteses do teorema 4.4.1, sejam u e v sequências finitas de elementos de A, tal que $p(v) > 0$. Então, para todo $\epsilon > 0$, todo inteiro K tal que $K > \frac{\epsilon}{p(v)}$ e

um n suficientemente grande, temos que,

$$\mathbb{P}\left(\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\right| > \epsilon\right) \leq 3e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c \cdot \eta^2\left(\frac{\epsilon}{2}, p(v)\right)}{p(v)}\right\}$$

Demonstração:

$$\begin{aligned}
& \mathbb{P}\left(\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\right| > \epsilon\right) \\
= \; & \mathbb{P}\left(\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv) - \hat{p}_n(uv) + \hat{p}_n(uv)}{p(v)}\right| > \epsilon\right) \\
= \; & \mathbb{P}\left(\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)} + \frac{\hat{p}_n(uv)}{p(v)} - \frac{\hat{p}_n(uv)}{p(v)}\right| > \epsilon\right) \\
= \; & \mathbb{P}\left(\left|\hat{p}_n(uv)\left(\frac{1}{\hat{p}_n(v)} - \frac{1}{p(v)}\right) + \frac{1}{p(v)}\left(\hat{p}_n(uv) - p(uv)\right)\right| > \epsilon\right).
\end{aligned}$$

Como $0 \leq \hat{p}_n(uv) \leq 1$, obtemos que

$$\begin{aligned}
& \mathbb{P}\left(\left|\hat{p}_n(uv)\left(\frac{1}{\hat{p}_n(v)} - \frac{1}{p(v)}\right) + \frac{1}{p(v)}\left(\hat{p}_n(uv) - p(uv)\right)\right| > \epsilon\right) \\
\leq \; & \mathbb{P}\left(\left|\frac{1}{\hat{p}_n(v)} - \frac{1}{p(v)}\right| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\left|\hat{p}_n(uv) - p(uv)\right| > \frac{p(v)\epsilon}{2}\right).
\end{aligned}$$

Aplicando o lema 4.4.3 e o teorema 3.1 de Galves e Leonardi (2008), e como $p(uv) < p(v)$ e $0 < p(v) - \frac{\epsilon}{2K} < 1$, temos que

$$\begin{aligned}
& \mathbb{P}\left(\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\right| > \epsilon\right) \\
\leq \; & 2e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\eta^2\left(\frac{\epsilon}{2}, p(v)\right)}{p(v)}\right\} + e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\left(\frac{p(v)\epsilon}{2}\right)^2}{p(uv)}\right\} \\
\leq \; & 3e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\eta^2\left(\frac{\epsilon}{2}, p(v)\right)}{p(v)}\right\},
\end{aligned}$$

onde $\eta\left(\frac{\epsilon}{2}, p(v)\right) = \min\left\{\frac{\epsilon}{2K}, p(v)\left(p(v) - \frac{\epsilon}{2K}\right)\frac{\epsilon}{2}\right\}$.

Do lema 4.4.4 e por (4.35), obtemos que

$$\begin{aligned}
& \mathbb{P}\left(\left|\hat{q}_n(v) - \sum_{u\,:\,uv\in\tau} \frac{p(uv)}{p(v)} \cdot q(uv)\right| > \epsilon\right) \\
\leq \; & \mathbb{P}\left(\sup_{u\,:\,uv\in\tau}\left\{\left|\hat{q}_n(uv) - q(uv)\right|\right\} > \frac{\epsilon}{2A^{d-|v|}}\right) \\
+ \; & \mathbb{P}\left(\sup_{u\,:\,uv\in\tau}\left\{\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\right|\right\} > \frac{\epsilon}{2A^{d-|v|}\sup_{u\,:\,uv\in\tau}\{q(uv)\}}\right)
\end{aligned} \tag{4.41}$$

Então, para algum u' tal que $u'v \in \tau$ e para

$$a = \frac{\epsilon}{2A^{d-|v|} \cdot \sup\limits_{u:uv\in\tau} \{q(uv)\}},$$

concluímos que

$$\begin{aligned}
& \mathbb{P}\left(A^{d-|v|} \sup_{u\,:\,uv\in\tau} \left\{\left|\hat{q}_n(uv) - q(uv)\right|\right\} > \frac{\epsilon}{2}\right) \\
+ \quad & \mathbb{P}\left(A^{d-|v|} \sup_{u\,:\,uv\in\tau} \{q(uv)\} \cdot \sup_{u\,:\,uv\in\tau} \left\{\left|\frac{\hat{p}_n(uv)}{\hat{p}_n(v)} - \frac{p(uv)}{p(v)}\right|\right\} > \frac{\epsilon}{2}\right) \\
\leq \quad & 4\exp\left\{-n \cdot h(p(u'v), q(u'v), \frac{\epsilon}{2A^{d-|v|}})\right\} \\
+ \quad & 3e^{\frac{1}{e}} \cdot \exp\left\{\frac{-n \cdot c\eta^2\left(a, p(v)\right)}{p(v)}\right\} \\
\leq \quad & 10 \cdot e^{-n\cdot\min\left\{h\left(p(u^*v), q(u^*v), \frac{\epsilon}{2A^{d-|v|}}\right), \frac{c\cdot\eta^2(a,p(v))}{p(v)}\right\}}.
\end{aligned}$$

□

Demonstração do teorema 4.2.3: Basta aplicar o lema de Borel-Cantelli, utilizando o resultado do teorema 4.4.2, e temos o resultado desejado.

□

Para que o estimador $\hat{\tau}_n$ seja funcional, é necessário que , dado $u \in \tau$, para um n suficientemente grande, tenhamos que $\hat{q}_n(u)$ e $\hat{q}_n(suf(u))$ assumam valores suficientemente distintos, sob pena de excluirmos indevidamente de $\hat{\tau}_n$ um elemento que pertence a τ.

O teorema a seguir mostra que , para um δ conveniente, $\mathbb{P}\left(\Delta_n(suf(u)) \leq \delta\right)$ converge exponencialmente a zero.

Teorema 4.4.3. Para $|\tau| = d$, seja $u \in \tau$ e $v = suf(u)$. Então, existe um $\delta > 0$ tal que

$$\mathbb{P}\left(\Delta_n(suf(u)) \leq \delta\right) \leq 14 \cdot e^{-n\cdot\min\left\{h\left(p(u^*v), q(u^*v), \frac{\delta}{2A^{d-|v|}}\right), \frac{c\cdot\eta^2(a,p(v))}{p(v)}, h(p(u),q(u),\delta)\right\}},$$

para algum u^* tal que $u^*v \in \tau$.

Demonstração:

$$\begin{aligned}
\mathbb{P}\left(\Delta_n(v) \leq \delta\right) &= \mathbb{P}\left(\max_{a\in A} |\hat{q}_n(v) - \hat{q}_n(av)| \leq \delta\right) \\
&\leq \mathbb{P}\left(|\hat{q}_n(v) - \hat{q}_n(u)| \leq \delta\right),
\end{aligned}$$

pois $\left\{\max\limits_{a\in A} |\hat{q}_n(v) - \hat{q}_n(av)| \leq \delta\right\} \subset \left\{|\hat{q}_n(v) - \hat{q}_n(u)| \leq \delta\right\}$.

Seja $q(v) = \sum\limits_{z:zv\in\tau} \frac{p(zv)}{p(v)} q(zv)$. Então, temos que

$$\begin{aligned}
\mathbb{P}\left(|\hat{q}_n(v) - \hat{q}_n(u)| \leq \delta\right) &= \mathbb{P}\left(|\hat{q}_n(v) - q(v) + q(v) - \hat{q}_n(u)| \leq \delta\right) \\
&\leq \mathbb{P}\left(|q(v) - \hat{q}_n(u)| - |\hat{q}_n(v) - q(v)| \leq \delta\right) \\
&= \mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq \delta + |\hat{q}_n(v) - q(v)|\right).
\end{aligned}$$

Vamos particionar nos eventos disjuntos $\{|\hat{q}_n(v) - q(v)| \leq \delta\}$ e $\{|\hat{q}_n(v) - q(v)| > \delta\}$. Dessa forma, obtemos

$$\begin{aligned}
& \mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq \delta + |\hat{q}_n(v) - q(v)|\right) \\
= \; & \mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq \delta + |\hat{q}_n(v) - q(v)|, |\hat{q}_n(v) - q(v)| \leq \delta\right) \\
+ \; & \mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq \delta + |\hat{q}_n(v) - q(v)|, |\hat{q}_n(v) - q(v)| > \delta\right) \\
\leq \; & \mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq 2\delta\right) + \mathbb{P}\left(|\hat{q}_n(v) - q(v)| > \delta\right). \qquad (4.42)
\end{aligned}$$

Pelo teorema 4.4.2, temos que um majorante para o segundo termo de (4.42) é dado por

$$\mathbb{P}\left(|\hat{q}_n(v) - q(v)| > \delta\right) \leq 10 \cdot e^{-n \cdot \min\left\{h\left(p(u^*v), q(u^*v), \frac{\delta}{2A^{d-|v|}}\right), \frac{c \cdot \eta^2(a, p(v))}{p(v)}\right\}}, \qquad (4.43)$$

para alguma sequência u^* tal que $u^*v \in \tau$.

Para o primeiro termo de (4.42), vamos precisar do lema a seguir.

Lema 4.4.5. Seja $u \in \tau$ e $v = suf(u)$. Então, existe um $\delta > 0$ tal que

$$\mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq 2\delta\right) \leq 4 \cdot e^{-n \cdot h(p(u), q(u), \delta)}.$$

Demonstração:

$$\begin{aligned}
\mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq 2\delta\right) &= \mathbb{P}\left(|q(v) - \hat{q}_n(u) - q(u) + q(u)| \leq 2\delta\right) \\
&\leq \mathbb{P}\left(|q(v) - q(u)| - |\hat{q}_n(u) - q(u)| \leq 2\delta\right) \\
&= \mathbb{P}\left(|q(v) - q(u)| \leq 2\delta + |\hat{q}_n(u) - q(u)|\right).
\end{aligned}$$

Vamos particionar nos eventos disjuntos $\{|\hat{q}_n(u) - q(u)| \leq \delta\}$ e $\{|\hat{q}_n(u) - q(u)| > \delta\}$. Dessa forma, obtemos

$$\begin{aligned}
& \mathbb{P}\left(|q(v) - q(u)| \leq 2\delta + |\hat{q}_n(u) - q(u)|\right) \\
= \; & \mathbb{P}\left(|q(v) - q(u)| \leq 2\delta + |\hat{q}_n(u) - q(u)|, |\hat{q}_n(u) - q(u)| \leq \delta\right) \\
+ \; & \mathbb{P}\left(|q(v) - q(u)| \leq 2\delta + |\hat{q}_n(u) - q(u)|, |\hat{q}_n(u) - q(u)| > \delta\right) \\
\leq \; & \mathbb{P}\left(|q(v) - q(u)| \leq 3\delta\right) + \mathbb{P}\left(|\hat{q}_n(u) - q(u)| > \delta\right). \qquad (4.44)
\end{aligned}$$

Como $u \in \tau$, pelo teorema 4.4.1, temos que um majorante para o segundo termo de (4.44) é dado por

$$\mathbb{P}\left(|\hat{q}_n(u) - q(u)| > \delta\right) \leq 4 \cdot e^{-n \cdot h(p(u), q(u), \delta)}.$$

Para o primeiro termo de (4.44), temos que, pela hipótese 4, $|q(v) - q(u)| > 0$. Logo, considerando

$$\delta \leq \frac{|q(v) - q(u)|}{3},$$

temos que

$$\mathbb{P}\left(|q(v) - q(u)| \leq 3\delta\right) = 0$$

Logo, concluímos que existe δ positivo tal que

$$\mathbb{P}\left(|q(v) - \hat{q}_n(u)| \leq 2\delta\right) \leq 4 \cdot e^{-n \cdot h(p(u), q(u), \delta)}.$$

Assim, concluindo a demonstração do teorema, temos que

$$\begin{aligned}
\mathbb{P}\left(\Delta_n(suf(u)) \leq \delta\right) &\leq 10 \cdot e^{-n \cdot \min\left\{h\left(p(u^*v), q(u^*v), \frac{\delta}{2A^{d-|v|}}\right), \frac{c \cdot \eta^2(a, p(v))}{p(v)}\right\}} + 4 \cdot e^{-n \cdot h(p(u), q(u), \delta)} \\
&\leq 14 \cdot e^{-n \cdot \min\left\{h\left(p(u^*v), q(u^*v), \frac{\delta}{2A^{d-|v|}}\right), \frac{c \cdot \eta^2(a, p(v))}{p(v)}, h(p(u), q(u), \delta)\right\}},
\end{aligned}$$

para algum u^* tal que $u^*v \in \tau$.

□

4.5 Taxa de Convergência do estimador da árvore de contexto

Dada uma amostra $(X_i, T_i)_{i=1}^n$, sejam $\hat{q}_n(\cdot)$ o estimador das taxas de disparo do processo gerador da amostra, como definido em (4.1) e $\hat{\tau}_n^\delta$ o estimador da árvore de contextos geradora da amostra, definido em (4.5).

Nessa seção, vamos provar que a árvore estimada, truncada em uma altura finita qualquer, converge com probabilidade 1 para a árvore geradora da amostra, truncada na mesma altura.

Inicialmente, vamos apresentar dois lemas, que serão essenciais para provar que a probabilidade da árvore estimada truncada diferir da árvore geradora da amostra truncada decai a zero com velocidade exponencial.

No primeiro lema, vamos calcular um limite superior para a probabilidade de, dado um elemento $\omega \in \tau$ e uma sequência u de elementos de A qualquer, obtemos que $u\omega$ é um candidato a elemento de $\hat{\tau}_n^\delta$, isto é, $\Delta_n\big(suf(u\omega)\big) > \delta$, para algum $\delta > 0$.

Lema 4.5.1. Sejam $\omega \in \tau$ tal que $|\omega| < k$, para algum inteiro k, e u uma sequência de elementos de A. Então, para $v = suf(u\omega)$, $\delta > 0$, temos que

$$\mathbb{P}\left(\Delta_n(v) > \delta\right) \leq 8|A|e^{-n \cdot \bar{h}(A,\, v, \delta/2)}, \tag{4.45}$$

onde

$$\bar{h}(A,\, v, \delta/2) = \min_{a \in A \cup \{\emptyset\}} \left\{h\big(p(av), q(av), \delta/2\big)\right\}. \tag{4.46}$$

Demonstração:

$$\begin{aligned}
\mathbb{P}\left(\Delta_n(v) > \delta\right) &= \mathbb{P}\left(\max_{a\in A} |\hat{q}_n(v) - \hat{q}_n(av)| > \delta\right) \\
&\leq \sum_{a\in A} \mathbb{P}\left(|\hat{q}_n(v) - q(v) + q(v) - \hat{q}_n(av)| > \delta\right) \\
&\leq \sum_{a\in A} \left\{\mathbb{P}\left(|\hat{q}_n(v) - q(v)| > \frac{\delta}{2}\right) + \mathbb{P}\left(|\hat{q}_n(av) - q(av)| > \frac{\delta}{2}\right)\right\}. \quad (4.47)
\end{aligned}$$

Pelo teorema 4.4.1, obtemos as desigualdades

$$\mathbb{P}\left(|\hat{q}_n(v) - q(v)| > \frac{\delta}{2}\right) \leq 4e^{-n\cdot h(p(v),q(v),\delta/2)},$$

e

$$\mathbb{P}\left(|\hat{q}_n(av) - q(av)| > \frac{\delta}{2}\right) \leq 4e^{-n\cdot h(p(av),q(av),\delta/2)}.$$

Como temos que

$$\begin{aligned}
\sum_{a\in A} \mathbb{P}\left(|\hat{q}_n(av) - q(av)| > \frac{\delta}{2}\right) &\leq |A| \max_{a\in A}\left\{\mathbb{P}\Big(|\hat{q}_n(av) - q(av)| > \frac{\delta}{2}\Big)\right\} \\
&\leq |A| \max_{a\in A}\left\{4e^{-n\cdot h(p(av),q(av),\delta/2)}\right\} \\
&= 4|A| e^{-n\cdot \min_{a\in A}\{h(p(av),q(av),\delta/2)\}},
\end{aligned}$$

concluímos de (4.47) que

$$\begin{aligned}
\mathbb{P}\left(\Delta_n(v) > \delta\right) &= |A|\mathbb{P}\left(|\hat{q}_n(v) - q(v)| > \frac{\delta}{2}\right) + \sum_{a\in A} \mathbb{P}\left(|\hat{q}_n(av) - q(av)| > \frac{\delta}{2}\right) \\
&\leq 4|A|e^{-n\cdot h(p(v),q(v),\delta/2)} + 4|A|e^{-n\cdot \min_{a\in A}\{h(p(av),q(av),\delta/2)\}} \\
&\leq 8|A|e^{-n\cdot \bar{h}(A,\,v,\delta/2)},
\end{aligned}$$

onde

$$\bar{h}(A,\, v, \delta/2) = \min_{a\in A\cup\{\emptyset\}} \left\{h\big(p(av), q(av), \delta/2\big)\right\}.$$

□

No lema seguinte vamos calcular um limitante superior para a probabilidade de, dado um elemento $v \in \tau$, obtermos que $\Delta_n(v) \leq \delta$, para algum $0 < \delta < D_d$, onde

$$D_d = \min_{\substack{u\in\tau|_d : q(suf(u)) \\ \neq q(a\, suf(u))}} \max_{a\in A} |q(suf(u)) - q(a\, suf(u))|.$$

Lema 4.5.2. Seja $0 < \delta < D_d$. Para $\omega \in \hat{\tau}_n^\delta$ tal que $|\omega| < k$, seja $u \in \tau|_d$ tal que $u \succ \omega$ e

$q(suf(u)) \neq q(a\, suf(u))$. Então, se $v = suf(u)$,

$$\mathbb{P}\left(\Delta_n(v) \leq \delta\right) \leq 4(|A|+1)e^{-n\cdot\bar{h}(A,v,(D_d-\delta)/2)}.$$

Demonstração:

Para qualquer $a \in A$,

$$\begin{aligned} |\hat{q}_n(v) - \hat{q}_n(av)| &= |\hat{q}_n(v) - q(v) + q(v) - q(av) + q(av) - \hat{q}_n(av)| \\ &\geq |q(v) - q(av)| - |\hat{q}_n(v) - q(v)| - |\hat{q}_n(av) - q(av)|. \end{aligned}$$

Nota que se existem funções reais tais que $r(a) \geq f(a) - g(a)$ para todo $a \in A$, então

$$r(a) \geq f(a) - \max_{x\in A} g(x).$$

Tomando o máximo em ambos os lados, temos que

$$\max_{a\in A} r(a) \geq \max_{a\in A}\{f(a) - \max_{x\in A} g(x)\} \Rightarrow \max_{a\in A} r(a) \geq \max_{a\in A} f(a) - \max_{a\in A} g(a)$$

.

Logo,

$$\max_{a\in A} |\hat{q}_n(v) - \hat{q}_n(av)| \geq \max_{a\in A} |q(v) - q(av)| - \max_{a\in A}\Big\{|\hat{q}_n(v) - q(v)| + |\hat{q}_n(av) - q(av)|\Big\}. \quad (4.48)$$

Dessa forma,

$$\Delta_n(v) \geq D_d - \max_{a\in A}\Big\{|\hat{q}_n(v) - q(v)| + |\hat{q}_n(av) - q(av)|\Big\},$$

onde $D_d = \min\limits_{\substack{u\in\tau|_d : q(suf(u)) \\ \neq q(a\, suf(u))}} \max\limits_{a\in A} |q(suf(u)) - q(a\, suf(u))|$ é uma constante.

Assim,

$$\begin{aligned} \mathbb{P}\left(\Delta_n(v) \leq \delta\right) &\leq \mathbb{P}\left(D_d - \max_{a\in A}\{|\hat{q}_n(v) - q(v)| + |\hat{q}_n(av) - q(av)|\} \geq \delta\right) \\ &= \mathbb{P}\left(\max_{a\in A}\{|\hat{q}_n(v) - q(v)| + |\hat{q}_n(av) - q(av)|\} \geq D_d - \delta\right) \\ &= \mathbb{P}\left(|\hat{q}_n(v) - q(v)| + \max_{a\in A}\{|\hat{q}_n(av) - q(av)|\} \geq D_d - \delta\right). \end{aligned}$$

Pela desigualdade

$$\mathbb{P}\Big(\max_{a\in A}\left\{|\hat{q}_n(av) - q(av)| \geq \frac{D_d - \delta}{2}\right) \leq \sum_{a\in A} \mathbb{P}\left(|\hat{q}_n(av) - q(av)| \geq \frac{D_d - \delta}{2}\right),$$

temos que

$$\mathbb{P}\left(\Delta_n(v) \leq \delta\right) \leq \mathbb{P}\left(|\hat{q}_n(v) - q(v)| \geq \frac{D_d - \delta}{2}\right) + \sum_{a\in A} \mathbb{P}\left(|\hat{q}_n(av) - q(av)| \geq \frac{D_d - \delta}{2}\right) \quad (4.49)$$

Pelo teorema 4.4.1, temos que

$$\mathbb{P}\left(|\hat{q}_n(v) - q(v)| \geq \frac{D_d - \delta}{2}\right) \leq 4\exp\left\{-n \cdot h(p(v), q(v), (D_d - \delta)/2)\right\}$$

e

$$\sum_{a \in A} \mathbb{P}\left(|\hat{q}_n(av) - q(av)| \geq \frac{D_d - \delta}{2}\right) \leq 4|A| \max_{a \in A} \exp\left\{-n \cdot h(p(av), q(av), (D_d - \delta)/2)\right\}.$$

Logo, concluímos que

$$\begin{aligned} \mathbb{P}\left(\Delta_n(v) \leq \delta\right) &\leq 4\exp\left\{-n \cdot h(p(v), q(v), (D_d - \delta)/2)\right\} \\ &+ 4|A| \max_{a \in A} \exp\left\{-n \cdot h(p(av), q(av), (D_d - \delta)/2)\right\} \\ &\leq 4(|A| + 1)e^{-n \cdot \bar{h}(A, v, (D_d - \delta)/2)}, \end{aligned} \tag{4.50}$$

e está provado o lema.

□

Apresentamos agora um teorema provando que a probabilidade da árvore estimada truncada $\hat{\tau}_n^\delta|_k$ ser diferente da árvore geradora da amostra truncada $\tau|_k$ converge com velocidade exponencial para zero.

Para demonstrar esse resultado, vamos dividir o evento $\left\{\hat{\tau}_n^\delta|_k \neq \tau|_k\right\}$ em dois eventos, e provar que cada um deles converge exponencialmente a zero.

Teorema 4.5.1. Dada uma amostra $(X_i, T_i)_{i=1}^n$, para um inteiro k, $0 < \delta < D_d$, sejam

$$h_1^* = \min_{\substack{\omega \in \hat{\tau}_n^\delta: \\ |\omega| < k}} \{\bar{h}(A, suf(\bar{u}\omega), (D_d - \delta)/2\};$$

$$h_2^* = \min_{\substack{\omega \in \tau: \\ |\omega| < k}} \min_{u\omega \in \hat{\tau}_n^\delta} \{\bar{h}(A, suf(u\omega), \delta/2)\};$$

$$h^* = \min\{h_1^*, h_2^*\}.$$

Então,

$$\mathbb{P}\left(\hat{\tau}_n^\delta|_k \neq \tau|_k\right) \leq 12|A|^{d+1}e^{-nh^*}. \tag{4.51}$$

Demonstração:

Vamos reescrever o evento $\left\{\hat{\tau}_n^\delta|_k \neq \tau|_k\right\}$ como a união de dois eventos:

$U_n = \{\hat{\tau}_n^\delta|_k$ contém um sufixo de algum elemento de $\tau|_k\}$;

$O_n = \{\tau|_k$ contém um elemento que é sufixo de algum elemento de $\hat{\tau}_n^\delta|_k\}$.

Em linguagem de teoria de conjuntos, sejam

$$U_n = \bigcup_{\substack{\omega\in\hat{\tau}_n^\delta:\\ |\omega|<k}} \bigcap_{u\omega\in\tau|_d} \{\Delta_n(suf(u\omega)) \leq \delta\};$$

$$O_n = \bigcup_{\substack{\omega\in\tau:\\ |\omega|<k}} \bigcup_{u\omega\in\hat{\tau}_n^\delta} \{\Delta_n(suf(u\omega)) > \delta\};$$

Então, temos que

$$\left\{\hat{\tau}_n^\delta|_k \neq \tau|_k\right\} = U_n \cup O_n.$$

Seja $\omega \in \hat{\tau}_n^\delta$ tal que $|\omega| < k$, $a \in A$ e u uma sequência tal que $|u| \leq k - |\omega| - 1$. Então,

$$\mathbb{P}\left(\bigcap_{u\omega\in\tau|_d} \{\Delta_n(suf(u\omega)) \leq \delta\}\right) \leq \mathbb{P}\left(\Delta_n(v) \leq \delta\right), \tag{4.52}$$

onde $v = suf(\bar{u}\omega)$, para algum $\bar{u}\omega \in \tau|_d$ tal que $\bar{u}\omega \succ \omega$ e $q(suf(\bar{u}\omega)) \neq q(a\, suf(\bar{u}\omega))$.

Então, por (4.52) e pelo lema 4.5.2,

$$\begin{aligned}
\mathbb{P}(U_n) &= \mathbb{P}\left(\bigcup_{\substack{\omega\in\hat{\tau}_n^\delta:\\ |\omega|<k}} \bigcap_{u\omega\in\tau|_d} \{\Delta_n(suf(u\omega)) \leq \delta\}\right) \leq \sum_{\substack{\omega\in\hat{\tau}_n^\delta:\\ |\omega|<k}} \mathbb{P}\left(\bigcap_{u\omega\in\tau|_d} \{\Delta_n(suf(u\omega)) \leq \delta\}\right) \\
&\leq \sum_{\substack{\omega\in\hat{\tau}_n^\delta:\\ |\omega|<k}} \mathbb{P}\left(\Delta_n(suf(\bar{u}\omega)) \leq \delta\right) \leq |A|^k \cdot 4(|A|+1)e^{-nh_1^*}.
\end{aligned}$$

onde $h_1^* = \min_{\substack{\omega\in\hat{\tau}_n^\delta:\\ |\omega|<k}} \{\bar{h}(A, suf(\bar{u}\omega), (D_d - \delta)/2\}$.

Por outro lado, aplicando o lema 4.5.1 temos que

$$\begin{aligned}
\mathbb{P}(O_n) &= \mathbb{P}\left(\bigcup_{\substack{\omega\in\tau:\\ |\omega|<k}} \bigcup_{u\omega\in\hat{\tau}_n^\delta} \{\Delta_n(suf(u\omega)) > \delta\}\right) \leq \sum_{\substack{\omega\in\tau:\\ |\omega|<k}} \sum_{u\omega\in\hat{\tau}_n^\delta} \mathbb{P}\left(\Delta_n(suf(u\omega)) > \delta\right) \\
&\leq \sum_{\substack{\omega\in\tau:\\ |\omega|<k}} \sum_{u\omega\in\hat{\tau}_n^\delta} 8|A|e^{-n\bar{h}(A, suf(u\omega), \delta/2)} \leq 8|A|^{d+1}e^{-nh_2^*},
\end{aligned}$$

onde $h_2^* = \min_{\substack{\omega\in\tau:\\ |\omega|<k}} \min_{u\omega\in\hat{\tau}_n^\delta} \{\bar{h}(A, suf(u\omega), \delta/2)\}$.

Então, temos que

$$\begin{aligned}
\mathbb{P}(\hat{\tau}_n^\delta|_k \neq \tau|_k) &= \mathbb{P}(U_n) + \mathbb{P}(O_n) \leq 4|A|^k(|A|+1)e^{-nh_1^*} + 5|A|^{d+1}e^{-nh_s^*} \\
&\leq 12|A|^{d+1}e^{-nh^*},
\end{aligned} \tag{4.53}$$

onde $h^* = \min\{h_1^*, h_2^*\}$.

□

Enfim, demonstramos a seguir o principal resultado dessa tese, a consistência forte do estimador $\hat{\tau}_n^\delta$, apresentada no teorema 4.3.1.

Demonstração do teorema 4.3.1:

Para d, δ e k fixados, temos que

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(\hat{\tau}_n^{\delta}|_k \neq \tau|_k\right) \leq \sum_{n\in\mathbb{N}} 12|A|^{d+1}e^{-nh^*} < \infty.$$

Aplicando o lema de Borel-Cantelli, temos que

$$\mathbb{P}\left(\hat{\tau}_n^{\delta}|_k \neq \tau|_k \text{ i.v. }\right) = 0,$$

ou seja, $\hat{\tau}_n^{\delta}|_k \to^{qc} \tau|_k$.

$\square$

Capítulo 5

Simulação

Neste capítulo vamos construir o pseudo-algoritmo de simulação de uma amostra de um processo de salto com memória de alcance variável e o pseudo algoritmo que gera uma estimativa para a árvore de contextos de uma amostra advinda do mesmo tipo de processo.

5.1 Pseudo-Algoritmo de simulação

Para gerar uma amostra de um processo de salto com memória de alcance variável, precisamos definir o tamanho da amostra, o alfabeto em que o processo está definido, a árvore probabilística de contextos imersa do processo e as taxas de salto para cada contexto.

Dados esses elementos, o primeiro passo do algoritmo de simulação consiste em definir algum dos contextos finitos como esta inicial do processo. Em seguida, calculamos o tempo até o próximo salto, baseada numa variável aleatória exponencial, com taxa definida pelo contexto observado. Para definir o novo estado do processo, utilizamos a árvore probabilística de contextos. O processo é repetido até obter o tamanho da amostra definido inicialmente.

As seguintes variáveis serão utilizadas:

- T_i representa o i-ésimo instante de transição do processo.
- X_i representa o estado do processo no instante T_i.
- A é o alfabeto da amostra.
- n é um inteiro positivo, indicando o tamanho da amostra.
- (τ, p) é uma árvore probabilística de contextos.
- x^0_{-k} é o estado inicial do processo.

Algorithm 1

Input: ${x^0}_{-k}$, n, A, (τ, p), $\{q(\omega), \omega \in \tau\}$.
Output: Uma amostra $(X_i, T_i)_{i=1}^n$.

```
T_0 ← 0
ω ← x_0^{-k}
for i in 1 to n do
    tempo ← Exponencial(ω)
    T_i ← T_{i-1} + tempo
    Sorteio um elemento de A, utilizando as probabilidades {p(a|ω) : ainA}, e nomeio ele X_i
    ω ← x_l^i, o de l é tal que x_l^i ∈ τ.
end for
Return (X_i, T_i)_{i=1}^n
```

5.2 Pseudo-Agoritmo de estimação

Nessa seção, vamos apresentar o pseudo-algoritmo para o estimador da árvore de contextos (4.5).

As seguintes variáveis serão utilizadas:

- T_i representa o i-ésimo instante de transição do processo.
- X_i representa o estado do processo no instante T_i.
- A é o alfabeto da amostra.
- k é um inteiro positivo, indicando o nível que queremos truncar a árvore.
- δ é um número real positivo, utilizado como critério de poda da árvore.
- $B_\tau(s)$ são os filhos de uma sequência s.
- $\Delta_n(s)$ é a função definida em (4.4).
- $\hat{\tau}_n$ é uma árvore de contextos.

Algorithm 2

Input: Uma amostra $(X_i, T_i)_{i=1}^n$, A, k e δ.
Output: Uma árvore de contextos $\hat{\tau}_n|_k$.

```
d ← log_{|A|}(n)
τ ← T_n
Flag(s) ← "não visitado" para todos nós internos s de τ tais que |B_τ(s)| ≥ 2
for i in d to 1 do
    while ∃s: ℓ(s) = i, Flag(s) = "não visitado" e |B_τ(s)| ≥ 2 do
        Tome s tal que ℓ(s) = i, Flag(s) = "não visitado" e com |B_τ(s)| ≥ 2
        Realizamos o teste estatístico Δ_n(s) to test H_0^(s)
        if Δ_n(s) > c_n then
            Flag(u) ← "visited" ∀u ⪯ s such that |B_τ(u)| ≥ 2
        else
            Flag(s) ← "visited"
            τ ← τ \ B_τ(s) ∪ {s}
        end if
    end while
end for
τ̂_n = τ.
Return τ̂_n|_k
```

Capítulo 6

Conclusões

Introduzimos neste trabalho uma nova classe de processo estocástico a tempo contínuo, o processo de salto com memória de alcance variável. Apresentamos um exemplo de aplicação dessa classe de modelos, modelando o problema de comunicação neuronal apresentado em Galves e Löcherbach (2013). Consideramos na aplicação apenas alguns aspectos da dinâmica de iteração entre as partículas do sistema, devendo essa questão ser aprofundada em estudos futuros, considerando outras variantes do comportamento neuronal.

Propusemos um estimador para a árvore de contextos do processo de salto com memória de alcance variável , baseado nas taxas de transição empíricas da amostra, considerando uma quantidade maior de informação da amostra em comparação com o algoritmo contexto e o método BIC.

Demonstramos a convergência com velocidade exponencial da taxa de transição empírica para a taxa real, o que foi essencial para provar a convergência quase certa da árvore estimada truncada para a árvore real geradora da amostra, truncada no mesmo nível, nos casos em que a árvore real é infinita.

Estudos de simulação devem ser realizados futuramente, bem como um estudo mais aprofundado da aplicação do estimador proposto na compreensão da dinâmica de interação do sistema neuronal.

Referências Bibliográficas

[CFF02] F. Comets, R. Fernández e P. A. Ferrari. Processes with long memory: Regenerative construction and perfect simulation. *Ann. Appl. Probab.*, 12(3):921–943, 08 2002.

[CT06] I. Csiszar e Z. Talata. Context tree estimation for not necessarily finite memory processes, via bic and mdl. *Information Theory, IEEE Transactions on*, 52(3):1007–1016, March 2006.

[DGG06] D. Duarte, An. Galves e N. L. Garcia. Markov approximation and consistent estimation of unbounded probabilistic suffix trees. *Bulletin of the Brazilian Mathematical Society*, 37(4):581–592, 2006.

[DO15] A. Duarte e G. Ost. A model for neural activity in the absence of external stimuli. *a ser publicado em Markov Processes and Related Fields*, 2015.

[FGF01] R. Fernández, A. Galves e P. Ferrari. *Coupling, Renewal and Perfect Simulations of Chain of Infnite Order*. Lecture Notes for the 5a. Escola Brasileira de Probabilidade, Ubatuba, Brasil. 2001.

[Gar06] Aurélien Garivier. Consistency of the unlimited bic context tree estimator. *IEEE Transactions on Information Theory*, 52(10):4630–4635, Oct. 2006.

[GGG+12] A. Galves, C. Galves, J. García, N. L. Garcia e F. Leonardi. Context tree selection and linguistic rhythm retrieval from written texts. *Ann. Appl. Stat.*, 6(1):186–209, 03 2012.

[GL08a] A. Galves e F. Leonardi. Exponential inequalities for empirical unbounded context trees. Em Vladas Sidoravicius e MariaEulália Vares, editors, *In and Out of Equilibrium 2*, volume 60 of *Progress in Probability*, páginas 257–269. Birkhäuser Basel, 2008.

[GL08b] A. Galves e E. Löcherbach. Stochastic chains with memory of variable length. *ArXiv e-prints*, Abril 2008.

[GL13] A. Galves e E. Löcherbach. Infinite systems of interacting chains with memory of variable length - a stochastic model for biological neural nets. *Journal of Statistical Physics*, 151(5):896–921, 2013.

[HL14] P. Hodara e E. Löcherbach. Hawkes processes with variable length memory and an infinite number of components. *ArXiv e-prints*, Outubro 2014.

[Ris83] J. Rissanen. A universal data compression system. *IEEE Transactions on Information Theory*, 29(5):656–663, 1983.

[RST96] D. Ron, Y. Singer e N. Tishby. The power of amnesia: Learning probabilistic automata with variable memory length. *Machine Learning*, 25(2-3):117–149, 1996.

[Spi70] F. Spitzer. Interaction of markov processes. *Adv. Math.*, 5:290–290, 1970.